REVISED

THE EVALUATION OF CRIMINAL JUSTICE PROGRAMS:

AN APPROACH TO THE USE OF MULTIATTRIBUTE UTILITY TECHNOLOGY

Ward Edwards and J. Robert Newman

Social Science Research Institute
University of Southern California

with the collaboration of

Kurt Snapper

Planning Systems Associates, Inc.
Herndon, Virginia

and

David Seaver

Decision Science Consortium, Inc.
Falls Church, Virginia

November, 1980

l was developed under a grant from the National Institute of Law
t and Criminal Justice, Law Enforcement Assistance Administration,
of Justice (LEAA Grant No. 79-NI-AX-0002). Points of view or
tated in this document are those of the authors and do not
y represent the official position or policies of the U.S. Department

ABSTRACT

This document describes an approach to social program evaluation, developed from decision analysis, called Multiattribute Utility Technology (MAUT). The document is designed to be used by evaluators. Every attempt is made to describe MAUT as simply as possible so that it can be used without reference to other technical literature. The document contains seven chapters and four appendices. Chapter 1 gives an overview of what program evaluation is or should be from one perspective and illustrates how MAUT fits into a broad spectrum of evaluation techniques. Chapter 2 gives an example application of MAUT. It is self contained; readers of this document can understand what MAUT is all about by just reading this chapter. Chapters 3 through 6 contain all the technical detail of MAUT amply illustrated with actual program evaluations that have been carried out to various stages of completion. All the arithmetic necessary for the analysis is described and where possible actual forms used for collecting the necessary data for a MAUT analysis are included for possible use by other evaluators. Chapter 7, entitled Sensitivity Analysis, describes how to evaluate the MAUT technology itself. The four technical appendices contain listings and descriptions of programs that can be used with standard hand held calculators that make the arithmetical tasks of a MAUT analysis easy to accomplish.

EXECUTIVE SUMMARY

The purpose of this document is to present one approach to evaluation of social programs: Multiattribute Utility Technology (MAUT). MAUT, as described in this document, is not offered as a substitute for other modes of evaluation but rather as a widely applicable method for organizing and presenting evaluative information. MAUT proceeds along a series of steps: (1) Identification of the objects of evaluation and the function or functions that the evaluation is intended to perform. (2) Identification of stakeholders, the people or groups who have a stake or interest in the program being evaluated. (3) Elicitation from the stakeholders the relevant value dimensions or attributes. A value attribute is something that is important to the process of the program and/or its output. These value attributes are usually organized into a hierarchical structure called a value tree. (4) The assessment by each stakeholder of the relative importance of the value attributes identified in step 3. (5) The measurement of how well each object of evaluation serves each value attribute at the lowest level of the value tree. These are called single attribute utilities or location measures. (6) The aggregation of the location measures with measures of importance. This yields an overall measure of the "worth" of the program being evaluated. (7) The conduct of a sensitivity analysis which essentially tests to see whether different numerical inputs to the MAUT analysis will lead to different conclusions. In a sense sensitivity analysis is an evaluation of the MAUT technique itself when it is applied to the evaluation of any program.

The goal of this document is to describe a version of MAUT that is simple and straightforward so that any person who wishes to conduct a MAUT evaluation can do so without reference to any other source describing the technical details of MAUT. All the arithmetic necessary for the conduct of MAUT is illustrated by many examples and can be done by hand or better, with the help of a programmable hand calculator, the programs for which are supplied in the appendix to the document. With the exception of the first chapter no references to the technical literature are given but every technical detail of MAUT given in the document is backed up by research published in the technical literature. The following are brief summaries of each chapter in the document: Chapter 1 discusses the nature of evaluation in general and how MAUT relates to other approaches to evaluation. The different classes and purposes of evaluation are discussed. Emphasis is placed on the relation between evaluation and decision making, in particular decisions about how successful or unsuccessful a program may be or decisions about how to monitor an ongoing program and make recommendations for its improvement. Several examples from the criminal justice area are given and the chapter ends with an actual demonstration of how an application of MAUT to a particular criminal justice program led to a decision about that program.

Chapter 2 presents a completely worked out example of the conduct of a MAUT evaluation. The example is simple enough to be completely understandable yet complex enough to illustrate all the technical ideas. For those wanting to know what MAUT is all about this chapter is self contained and no further reading of the document is necessary unless one wants to see how the actual technical details are developed and applied in various contexts.

Chapter 3 discusses how to identify the relevant stakeholders of any program and elicit from them the value attributes to be used in the evaluation. The importance of recognizing the organizations and groups who might have generated the program as well as the administrators of the program and the **clients** of the program is emphasized. Each group might have different concepts and

therefore values about the program. It is important that the list of attributes be as complete as possible but also kept as short as possible. Several examples are given from evaluations of a Community Anti-Crime Program, a Dispute Resolution program, a Juvenile Delinquent Control program and a selection of a school desegregation program. How to construct value trees is described in detail using these examples.

Chapter 4 discusses and illustrates the major techniques to elicit importance weights from the stakeholders. Each technique is illustrated with a numerical example from actual evaluations that have been conducted. The possible use of the value tree in simplifying the weighting process is described. Forms that have been found useful in obtaining weight judgments from people are also included for possible use by other evaluators.

Chapter 5 describes and illustrates how location measures for each value attribute are determined for each of the options or entities being evaluated. A simple graphical method is described and its algebraic equivalent is also given. The importance of obtaining all the location measures on a common scale is emphasized. Example of all the possibilities that are apt to be met in practical evaluations are given and forms for obtaining the location measures are also included for possible use by other evaluators.

Chapter 6 describes and illustrates how the two sets of numbers, the importance weights and the location measures, are combined into an aggregated composite that reflects the overall evaluation of the program. All the arithmetic is illustrated by actual examples. How to present the results of an evaluation using MAUT techniques is also described and the importance of presenting an "evaluation profile" of the program being evaluated is demonstrated.

Chapter 7 describes how a sensitivity analysis is conducted on the MAUT analysis itself. Sensitivity analysis consists of changing some of the numbers that went into the initial MAUT analysis and doing it over again to see if the conclusions change and if so by how much. In most practical situations the final result of MAUT is not affected in any significant way by reasonable changes in the input. The chapter also discusses how sensitivity analysis can enhance the acceptability of the result of a MAUT evaluation. Chapter 7 is somewhat more technical than the other chapters and can be glossed over on first reading. It is strongly recommended, however, that every MAUT evaluation be accompanied by a sensitivity analysis of the MAUT technique itself.

The document concludes with four appendices containing descriptions of computer programs that can be keyed into a hand-held calculator to help an evaluator using MAUT to do all the required arithmetic.

CHAPTER 1 SUMMARY

Chapter 1 begins by defining the purpose of the document: to present a version of Multiattribute Utility Technology (MAUT). The version chosen for presentation emphasizes multiple stakeholder, multiple program objectives, wholehearted acceptance of subjectivity, and linkage of evaluation to decision. The Chapter distinguishes four reasons for evaluation: curiosity, monitoring, fine tuning, and several forms of programmatic choice. MAUT is useful to them all because it implies comparison of something with something else with respect to multiple objectives. MAUT is not a mode of evaluation in itself; instead, it is a way of organizing and aggregating evaluative efforts. The Chapter briefly lists the 7 steps of a MAUT, discusses the relationship between evaluation and decision, and makes suggestions about how evaluative efforts can be made more likely to influence decisions. It concludes with an instance of a MAUT that led to a decision.

CHAPTER 1

EVALUATION IN CRIMINAL JUSTICE AND OTHER SOCIAL PROGRAM CONTEXTS

Evaluation is rapidly becoming Big Business. Questions like "Is this plan wise?" "Should I choose option A or option B?" "At what funding level should this program be supported?" "How well is this program doing?" have been asked of social programs since long before we were born. But the idea that one could answer such questions systematically and in a manner other than simply looking at the object of evaluation and making an intuitive judgment is a development of the 1960's and 1970's. As inflated costs and less-inflated program budgets come into steadily escalating conflict, the task of weeding out the programs worthy of support from those that are not, and of providing guidance for programs in existence, will continue to grow in importance -- as will the resources and attention devoted to developing satisfying methods of performing that task.

What Is Evaluation?

The literature of evaluation is already huge, and grows daily. The purpose of this document is not academic, and we do not intend more than the most cursory of references even to the literature on the method of evaluation that is our topic. For a recent and very scholarly presentation of evaluation methods and results from a broad spectrum of viewpoints, including our own, see Klein and Teilmann (1980), Handbook of Criminal Justice Evaluation. Edwards's chapter in that Handbook will be of particular interest to scholars who find the ideas presented in this document stimulating and potentially useful to them, since it discusses the same ideas in a far more technical way, reviews a significant amount of literature, and cites the literature of this and of other methods.

The purpose of this document is to present one approach to evaluation: Multiattribute Utility Technology (MAUT). (For those interested in the sequence of acronyms that have been applied to this and similar ideas, the

following bit of history might be helpful. The term "multiattribute utility measurement" is long and clumsy; an acronym was inevitable. Two have competed in the literature during the 1970's: MAUM and MAUT. In the latter, the T originally stood for Theory. We prefer MAUT to MAUM, which sounds too much like "mom", but see little theoretical content in what we have to say, and so have substituted Technology for Theory. In various other publications, Edwards has proposed a version of MAUT that he called Simple MultiAttribute Rating Technique (SMART). The methods presented in this document are in some ways different from and simpler than those included in SMART; the differences seem substantial enough to us so that we prefer not to use that acronym.)

The goal of this document is to make a version of MAUT simple and straightforward enough so that the reader can, with diligence and frequent re-examinations of it, conduct relatively straightforward MAUT evaluations him- or herself, with no more help than a programmable hand calculator and some programs that we provide. In so doing, we will frequently resort to techniques that professional decision analyst will recognize as approximations and/or assumptions. The literature justifying those approximations is extensive and complex; to review it here would blow to smithereens our goal of being nontechnical.

What is MAUT, and how does it relate to other approaches to evaluation? Edwards, Guttentag, and Snapper (1975) discussed that question in 1975, and we have little to add. MAUT depends on a few key ideas:

1. When possible, evaluations should be comparative.
2. Programs normally serve multiple constituencies.
3. Programs normally have multiple goals, not all equally important.
4. Judgments are inevitably a part of any evaluation.
5. Judgments of magnitude are best when made numerically.

6. Evaluations typically are, or at least should be, relevant to decisions.

Some of the six points above are less innocent than they seem. If programs serve multiple constituencies, evaluations of them should normally be addressed to the interests of those constituencies; different constituencies can be expected to have different interests. If programs have multiple goals, evaluations should attempt to assess how well they serve them; this implies multiple measures and comparisons. The task of dealing with multiple measures of effectiveness (which may well be simple subjective judgments in numerical form) makes less appealing than might otherwise be the case the notion of social programs as experiments or quasi-experiments. While the tradition that programs should be thought of as experiments, or at least as quasi-experiments, has wide currency and wide appeal in evaluation research, its implementation becomes more difficult as the number of measures needed for a satisfactory evaluation increases. When experimental or other hard data are available, they can easily be incorporated in a MAUT evaluation.

Finally, the willingness to accept subjectivity into evaluation, combined with the insistence that judgments be numerical, serves several useful purposes. First, it partly closes the gap between intuitive and judgmental evaluations and the more quantitative kind; indeed, it makes coexistence of judgment and objective measurement within the same evaluation easy and natural. Second, it opens the door to easy combination of complex concatenations of values. For instance, evaluation researchers often distinguish between process evaluations and outcome evaluations. Process and outcome are different, but if a program has goals of both kinds, its evaluation can and should assess its performance on both. Third, use of subjective inputs can, if need be, greatly shorten the time required for an evaluation to be carried out. A MAUT evaluation can be carried out from original definition of the evaluation

problem to preparation of the evaluation report in as little as a week of concentrated effort. The inputs to such an abbreviated evaluative activity will obviously be almost entirely subjective. But the MAUT technique at least produces an audit trail such that the skeptic can substitute other judgments for those that seem doubtful, and can then examine what the consequences for the evaluation are. We know of no MAUT social program evaluation that took less than two months, but in some other areas of application we have participated in execution of complete MAUT evaluations in as little as two days -- and then watched them be used as the justification for major decisions. Moreover, we heartily approved; time constraints on the decision made haste necessary, and we were very pleased to have the chance to provide some orderly basis for decision in so short a time.

One decision analysis consulting firm has adopted two procedural rules that they now routinely use to facilitate major decisions. One is that the actual decision makers must participate, full time and away from home base. The second is that no one may bring a briefcase; the goal of the MAUT procedures they use is to capture and organize the intuitions of these decision makers, rather than to collect and aggregate detailed facts. Judged by user satisfaction, these procedures are a big success.

Classes of Purposes for Evaluations

Evaluations can be done for various reasons; different reasons can and do lead to different forms of evaluative activities.

The most common reason for evaluation is that it is required; perhaps by mandate from Congress or from a sponsor or perhaps by rules internal to the program organization. Such formal requirements for evaluation are becoming more common; if the so-called "sunset laws" become more widespread, the requirements will be built into them.

The organizational requirement for an evaluation is normally based on the supposition that decisions need to be made. Sometimes the question is whether the program should be continued, modified, or scrapped. Sometimes it is simply what relatively minor changes, if any, should be made in program design, management, or functioning to improve its effectiveness. Sometimes no specific decisions are behind such mandated evaluations; the spirit of such evaluations is somewhat similar to the spirit that leads to annual external audit of corporate books.

Major evaluations are often required as a basis for potential major programmatic changes -- up to and including the most major of all changes: the birth or death of a program. Sometimes such decisions are pure life-or-death choices; at least equally often, some social problem requires attention, and the decision problem is which of several alternative approaches to dealing with it looks most promising. Funding-level decision are also programmatic choices; the same program at two substantially different funding levels is really two different programs.

From this welter of considerations, we think we can distinguish four different classes of reasons for evaluations: curiosity, monitoring, fine tuning, and programmatic choice. Curiosity in itself is seldom a basis for wisely performed evaluations, since most programs are too specific in character for the kinds of generalizations to which wisely applied curiosity can lead, and generalized curiosity is a poor guide to choice of evaluative methods or measures.

Monitoring is both an appropriate and a necessary function for any program, and we believe MAUT offers useful tools for monitoring. Monitoring shades over into fine-tuning; the same tools are relevant to both. Programmatic choice is the most important use to which evaluative information can be put, and the tools of MAUT are most directly relevant to it.

These reasons for evaluation share two common characteristics that make MAUT applicable to them all. The first is that, implicitly or explicitly, all require comparison of something with something else. This is most obvious in the case of programmatic choice. But even monitoring has the characteristic, since one normally wonders whether or not some minor change would change significantly one of the monitored values. An important implication of the comparative nature of virtually every evaluation is that some of the comparisons are inevitably between the program as it is and the program as it might be -- that is, between real and imaginary programs or programmatic methods. The necessity of comparing real with imaginary objects is one of the problems that most approaches to evaluation find very difficult to solve. The normal approach of traditional methods is to make the comparison object real, typically by embodying it in an experimental (or control) group, locus, or program. We admire such comparisons when they can be made (e.g., in drug trials), but consider them impractical for most social program evaluations. MAUT deals with this problem by accepting data and judgments on equivalent footings; judgment is the most generally useful tool we know for assessing the consequences of nonexistent programs. (Such judgments, of course, are best when based on relevant data, e.g., from other programs in other places.)

The second characteristic that the various reasons for evaluation share is that programs virtually always have multiple objectives, and consequently that evaluations should assess as many of these as seem important.

We use the word "program" in a sense broader than has been common; we are concerned with many social programs other than social service delivery programs. We consider arms procurements, treaties among nations, labor contracts, choices made by businesses about such questions as where to locate new plants, and other similar public decisions with major impacts on people to be "programs", and to deserve evaluation. One version or another of the methods we discuss ha been used for purposes as diverse as deciding whether to expand a Community Anti-Crime Program area, evaluating the Office of the Rentalsman in Vancouver as a dispute resolution mechanism, evaluating alternative school desegregation plans for Los Angeles, choosing among alternative sites for dams and nuclear power plants, evaluating competing bids for various kinds of military hardware, formulating U.S. negotiating positions in international negotiations, and assessing the combat readiness of Marine Corps brigades. For a bit more information and a number of references to such applications, see Edwards (1980).

Since we claim that MAUT can be applied to evaluative problems of each of the kinds we can identify, are we asserting that it is a universally applicable mode of evaluation -- perhaps a substitute for alternative modes? No. MAUT is, we believe, a very widely applicable method of organizing and presenting evaluative information. As such, it is compatible with any other evaluative activity designed to yield numbers as outputs. Since the ideas of MAUT do not limit the sources of the evaluative information, they can be combined with whatever data sources the evaluator finds satisfying and relevant to his or her problem.

Is MAUT an evaluative method at all? Without an answer to the question about where the evaluative information it must use will come from, the answer is no. However, Chapter 6 of this document presents some ideas about answers to that question. Whether those answers are a part of MAUT or external to it is obviously only a question of definition; the reader can choose.

Steps in a MAUT Evaluation and the Content of this Document

It may be helpful at this point to summarize concisely the steps involved in any MAUT evaluation. This will (a) summarize the remainder of the document, (b) provide a brief procedural guide, (c) identify, but not define,

the technical terms; they are defined one by one in the remainder of the document, and (d) provide a guide to the content of the remainder of the document.

First, a note about technical terms. There are a lot of them, and many will seem non-standard to those familiar with the MAUT literature. In every case that we can identify, use of a non-standard term corresponds to a shading of difference between what this document discusses and what previous publications (including many of which Edwards was an author) have discussed. Many more versions of MAUT exist than researchers active in developing it. While all depend on the same basic ideas, details of implementation change, and such changes produce corresponding changes in jargon. Many non-technical readers will wish to skip this section and go on to the next.

Step 1: Identify the objects of evaluations and the function or functions that the evaluation is intended to perform. Normally there will be several objects of evaluation, at least some of them imaginary, since evaluations are comparative. The functions of the evaluation will often control the choice of objects of evaluation. We have argued that evaluations should help decision-makers to make decisions. If the nature of those decisions is known, the objects of evaluation will often be controlled by that knowledge. Step 1 is outside the scope of this document. Some of the issues inherent in it have already been discussed in this Chapter. Chapter 2, devoted to setting up an example that will be carried through the document, illustrates Step 1 for that example.

Step 2: Identify the stakeholders (technical terms to be explained later are set in italics). Chapter 3 discusses this in detail.

Step 3: Elicit from stakeholder representatives the relevant value dimensions or attributes, and (often) organize them into a hierarchical structure called a value tree. Chapter 3 both explains how to do this and presents a real example.

Step 4: Assess for each stakeholder group the relative importances of each of the values identified at Step 3. Such judgments can, of course, be expected to vary from one stakeholder group to another; methods of dealing with such value conflicts are important. Chapter 4 presents assessment techniques and introduces some discussion of value differences; Chapter 7 returns to the issue of value differences.

Step 5: Ascertain how well each object of evaluation serves each value at the lowest level of the value tree. Such numbers, called single-attribute utilities or (in our preferred lingo) location measures, ideally report measurements, expert judgments, or both. If so, they should be independent of stakeholders and so of value disagreements among stakeholders; however, this ideal is not always met. Location measures need to be on a common scale, in order for Step 4 to make sense. Chapter 5, which is so far as we know unique in this literature in its emphasis on simplicity of methods, discusses both how to obtain location measures and how to put them on a common scale.

Step 6: Aggregate location measures with measures of importance. This is the topic of Chapter 6.

Step 7: Perform sensitivity analyses. The question underlying any sensitivity analysis is whether a change in the analysis, e.g., using different numbers as inputs, will lead to different conclusions. While conclusions may have emerged from Step 6, they deserve credence as a basis for action only after their sensitivity is explored in Step 7. Chapter 7 shows how some fairly simple sensitivity analyses can be performed.

Steps 6 and 7 will normally produce the results of a MAUT evaluation. Chapter 7 also has suggestions about how such results can be presented.

The Relation between Evaluation and Decision

The tools of MAUT are most useful for guiding decisions; they grow out of a broader methodological field called decision analysis. The relation of evaluation to decision has been a topic of debate among evaluation re-seachers -- especially the academic evaluation researchers who wonder whether or not their evaluations are used, and if so, appropriately used. Some evaluators take the position that their responsibility is to provide the relevant facts; it is up to someone else to make the decisions. "We are not elected officials." This position is sometimes inevitable, of course; the evaluator is not the decision-maker as a rule, and cannot compel the decision maker to attend to the result of the evaluation, or to base decisions on it. But it is unattractive to many evaluators; certainly to us.

We know of three devices that make evaluations more likely to be used in decisions. The first and most important is to involve the decision makers heavily in the evaluative process; this is natural if, as is normally the case, they are among the most important stakeholders. The second is to make the evaluation as directly relevant to the decision as possible, preferably by making sure that the options available to the decision maker are the objects of evaluation. The third is to make the product of the evaluation useful -- which primarily means making it readable and short. Exhaustive scholarly documents tend to turn busy decision makers off. Of course nothing in these obvious devices guarantees success in making the evaluation relevant to the decision. However, non-use of these devices comes close to guar-anteeing failure.

By "decisions" we do not necessarily mean anything apocalyptic; the process of fine tuning a program requires decisions too. So this document unabashedly assumes that either the evaluator or the person or organization commissioning the evaluations has the options, or alternative courses of action, in mind, and proposes to select among them in part on the basis of the evaluation -- or else that the information is being assembled and aggregated because of someone's expectation that that will be the case later on.

Whether or not this assumption fits your circumstances, it certainly does fit many in the criminal justice field. Examples:

Should this jurisdiction adopt a no-bail pre-trial release program for some offenses? If it does, the value best served is fairness to indigent defendants. The values possibly ill-served are certainty of appearance for trial, staff time spent in screening candidates for no-bail release, and danger to the community.

Should this jurisdiction adopt a full or partial work-release program? The values served are that the prisoners are self-supporting and so the program saves money both for their own support and for that of their families, and that the program may facilitate reintegration of released prisoners into employment and the community. Values possibly ill-served include staff time for management of the program, danger to the community (perhaps with special emphasis on employers), and tension among those not selected for work release.

Should misdemeanors be handled by citations and mail-in fines instead of arrest and booking? The values served are saving of time and cost for police, reduced interference with freedom of the accused misdemeanants, and increased income from fines. The values possibly ill-served include presumed deterrant effect, avoidance of recidivism, numbers of cases in which an accused innocent misdemenant chooses to plead guilty and pay rather than fight, and public respect for the laws so treated.

None of these decision questions have, so far as we know, been attacked with MAUT tools. They could be.

A Real Example of a MAUT Analysis that Helped Make a Decision

The Office of Community Anti-Crime Programs (OCAP) of the Law Enforcement Assistance Administration (LEAA) has been funding a number of community-based anti-crime projects throughout the country. The Decision Science Consortium, Inc. has been hired to perform a large MAUT analysis of this whole program; the key people in that evaluation have been Dr. Kurt Snapper and Dr. David Seaver. A more detailed discussion of the evaluation as a whole appears in Chapter 3 of this manual.

The following discussion of a specific decision within that evaluation program is condensed from Snapper and Seaver (in press). One of the community projects within OCAP's program is that of the Midwood-Kings Highway Development Corporation (MKDC) in Brooklyn. The objectives, called attributes in this document, of that particular project, and the weights given to them by its Director, are given in Exhibit 1. Note that all attributes are approximately equally important -- a quite unusual finding.

Those attributes and weights were elicited in the first year of the MKDC project. The project was quite successful in improving on the pre-project scores on these objectives in its area.

In 1979, a decision problem arose. The city of New York adopted a "coterminality" policy; police and other service delivery areas were to become aligned or "coterminous" with community districts. Since MKDC served a part of the area served by the Midwood Civic Action Council (MCAC), the problem was whether to expand MKDC's area of service to include all of MCAC's area -- a 50% expansion. No additional LEAA funds were expected for MKDC, so the concern was that expansion of the service area would lead to dilution of service quality and effectiveness. On the other hand, political considerations of various sorts argued for the expansion.

Working with Dr. Seaver and Dr. Snapper, the MKDC Project Director did a MAUT analysis of the two extreme options: to expand or not. The result is

EXHIBIT 1 OF CHAPTER 1

MKDC CAC VALUE ATTRIBUTES

Number	Title of Attribute	Importance Weight
1	Reduce Crime	.141
2	Reduce Fear of Crime	.140
3	Increase Police Responsiveness	.119
4	Serve Community Ombudsman Rule	.126
5	Increase Resident Involvement	.149
6	Institutionalize Organization	.111
7	Provide Technical Assistance	.104
8	Integrate Other Social Services	.110
		1.000

presented in Exhibit 2. It is important to note that the measures on which Exhibit 2 are based on judgments of the MKDC Project Director, and refer to the MKDC area alone. The baseline or zero point on each attribute is pre-MKDC Project measures. The 100 point on each dimension is the Project Director's judgment of the best that could be expected to be accomplished by the project. The weights used to combine the various utilities on each attribute into aggregate utilities come from Exhibit 1. The aggregate utility serves as one basis for the evaluation, the higher these values, the better the option. Note that both Exhibits 1 and 2 are sets of judgments by the Project Director. A less abbreviated MAUT would have included other stakeholders.

The Project Director was relatively surprised by the results presented in Exhibit 2; he had expected that expansion of the service area would lead to much more degradation of service than Exhibit 2 shows. He therefore chose to go ahead and expand the area, since he felt that in the presence of such a relatively minor effect on service, the political considerations were compelling.[1]

Political events in New York City have delayed implementation of coterminality, and there is some doubt about whether it will ever be implemented. However, MKDC is now considering petitioning LEAA to expand its target area to all of MCAC's area.

One reason for that decision is yet another version of the analysis. Recall that Exhibit 2 is based only on predicted measures within the original MKDC area. If the area were to be expanded, it would be appropriate to take

[1]Attribute 6, Option 1, in Exhibit 2 shows a value of 105 on a 0 to 100 scale. This simply means that the Project Director judged 1981 performance on this dimension to be better than the best he thought could be expected when he defined end points of the dimension. While such violations of the 0-100 range can occur, they should be rare.

EXHIBIT 2 OF CHAPTER 1

A MAUT ANALYSIS OF THE MKDC EXPANSION DECISION

Value Attributes	1979	1980	1981
Option 1: Expand to include all the MCAC area			
1. Reduce Crime	68	78	85
2. Reduce Fear of Crime	43	64	90
3. Increase Police Responsiveness	63	83	98
4. Serve Ombudsman Role	25	42	83
5. Increase Resident Involvement	28	69	95
6. Institutionalize Organization	46	70	105
7. Give Technical Assistance	25	40	80
8. Integrate Social Services	75	88	97
Aggregate Utility	46	67	92
Option 2: Do not expand at all			
1. Reduce Crime	68	81	89
2. Reduce Fear of Crime	43	71	97
3. Increase Police Responsiveness	63	84	100
4. Serve Ombudsman Role	25	50	100
5. Increase Resident Involvement	28	85	100
6. Institutionalize Organization	46	66	100
7. Give Technical Assistance	25	50	100
8. Integrate Social Services	75	90	100
Aggregate Utility	46	73	98

those measures over the whole MCAC area instead. Exhibit 3 shows the result of a MAUT analysis based on predicted measures covering the whole MCAC area. Note that expansion of the area leads to severe initial degradation (for the year 1979) of the project effectiveness measures, since the new area includes a substantial region within which the old MKDC project, which had been very successful, has not been operating. However, the forecast leads to the conclusion that, although the figures are not as high as either of those in Exhibit 2 for the MKDC area alone, they show major improvement with time. This invites the idea that "the greatest good of the greatest number" is well served by expanding, even in the presence of constant funding.

The Director also judged that a funding difference of only $60,000 would make the difference between leaving the original MKDC project ineffectual and giving it the necessary resources to serve all of the MCAC area as well as it was then serving MKDC. This is obviously an interesting assessment to report to LEAA in connection with any application to expand the MKDC area.

This is an example of a MAUT analysis carried out in a day. In spite of its brevity and omissions (e.g., of other stakeholders and of assessments of the political consequences of expanding or not expanding the area), it led a decision maker in a criminal justice project to change his mind, and provided him with the necessary information and analysis to defend that change of mind to sponsors, peers, and those he serves.

EXHIBIT 3 OF CHAPTER 1

PROJECT EFFECTIVENESS IN THE FULL MCAC AREA, ASSUMING EXPANSION

Value Attributes	1979	1980	1981
1. Reduce Crime	-5	63	76
2. Reduce Fear of Crime	10	53	81
3. Increase Police Responsiveness	0	63	84
4. Serve Ombudsman Role	10	35	60
5. Increase Resident Involvement	15	43	90
6. Institutionalize Organization	NA	66	70
7. Give Technical Assistance	0	25	50
8. Integrate Social Services	0	75	90
Aggregate Utility	5	53	76

REFERENCES

EDWARDS, W. (1971). "Social utilities", The Engineering Economist, 6:119-129.

________ GUTTENTAG, M., and SNAPPER, K (1975). "Effective evaluation: A decision theoretic approach." In C.A. Bennett and A. Lumsdaine (eds.), Evaluation and Experiment: Some Critical Issues in Assessing Social Programs. New York: Academic Press.

________ (1977). "How to use multiattribute utility measurement for social decision making," IEEE Transactions on Systems, Man, and Cybernetics, 7:326-340.

________ (1980). "Multiattribute utility for evaluation: Structures, uses, and problems." In M. Klein and K. Teilmann (eds.), Handbook of Criminal Justice Evaluation. Beverly Hills: Sage Publications.

SNAPPER, K. and SEAVER, D. (1978). "Application of Decision Analysis to Program Planning and Evaluation," Decision Science Consortium, Inc., Reston, Virginia, Technical Report 78-1.

CHAPTER 2 SUMMARY

Chapter 2 presents an example in detail. A social service center needs to move; six sites are available. Using staff weights applied to a value tree with 12 twigs, the Director of the Center is able to eliminate three of the six sites and to reach a conclusion among the other three.

Various technical problems arise and are discussed in presentation of the example. One is cost. The analysis treats cost as an evaluative attribute but keeps it separate from all other attributes until the end. Dominance techniques are used to eliminate options based on aggregated utilities and cost. An illustration is given of how judgments or tradeoffs between cost and all other attributes can be used as a basis for a single multi-attributed evaluation of what option is best. A second problem is how the nature of the context affects detailed definitions of values. A third is how to deal with options that fall outside anticipated ranges on one or more values. A fourth is how to go about operationalizing some values in order to obtain location measures. The last is what to do about ties in value, cost, or both.

Chapter 2

An Example

In this chapter we present a fairly simple example of how to use multiattribute utility technology for evaluation. The example is intended to be simple enough to be understandable yet complex enough to illustrate all of the technical ideas necessary for the analysis. Every idea introduced and illustrated is discussed in more detail in subsequent chapters. The example itself also reappears in later chapters.

Unfortunately, we cannot structure our discussion around the real example that we presented in the last chapter. It does not have all of the features of MAUT that we need to examine. So we have invented an example that brings out all the properties of the method, and that will, we hope, be sufficiently realistic to fit with the intuitions of those who work in a criminal justice environment. Please do not judge our lack of realism too harshly. The example in this chapter is complex enough; if we had worked hard to achieve full realism the example would have bogged down in too many details.

<u>The problem: how to evaluate new locations for a drug counseling center.</u>

The Drug-Free Center is a private non-profit contract center that gives counseling to clients sent to it by the courts of its city as a condition of their parole. It is a walk-in facility with no beds or other special space requirements; it does not use methadone. It has just lost its lease, and must relocate.

The Director of the Center has screened the available spaces to which it might move. All spaces that are inappropriate because of zoning, excessive neighborhood resistance to the presence of the Center, or inability to satisfy such legal requirements as access for the handicapped have been eliminated, as have spaces of the wrong size, price, or location. The city is in a period of economic recession, and so even after this pre-screening a substantial number of options are available. Six sites are chosen as a result of informal screening for serious evaluation. The Director must, of course, satisfy the sponsor, the Probation Department, and the Courts that the new location is appropriate, and must take the needs and wishes of both employees and clients into account. But as a first cut, the Director wishes simply to evaluate the sites on the basis of values and judgments of importance that make sense internally to the Center.

The Evaluation Process

The first task is to identify stakeholders. They were listed in the previous paragraph. A stakeholder is simply an individual or group with a reason to care about the decision, and with enough impact on the decision maker so that the reason should be taken seriously. Stakeholders are sources of value attributes. An attribute is something that the stakeholders, or some subset of them, care about enough so that failure to consider it in the decision would lead to a poor decision. We discuss the elicitation of attributes from stakeholders in in Chapter 3.

In this case, to get the evaluation started, the Director consulted, as stakeholders, the members of the Center staff. Their initial discussion of values elicited a list of about 50 verbal descriptors of values. A great many of these were obviously the same idea under a variety of different verbal labels. The Director, acting as leader of the discussion, was able to see these instances, and to persuade those who originally proposed these as values to agree on a rephrasing that captured and coalesced these overlapping or duplicating ideas. She did so both because she wanted to keep the list short and because she knew that if the same idea appeared more than once in the final list, she would be "double counting;" that is, including the same value twice. Formally, there is nothing wrong with double counting so long as the weights reflect it. But in practice, it is important to avoid, in part because the weights will often not reflect it, and in part because

the analysis is typically complex, and addition of extra and unnecessary attributes simply makes the complexity worse.

A second step in editing the list was to eliminate values that, in the view of the stakeholders, could not be important enough to influence the decision. Such a value, considered and then eliminated because it was unimportant, was "proximity to good lunching places." The Director was eager to keep the list of values fairly short, and her staff cooperated. In a less collegial situation, elimination of attributes can be much more difficult. Devices that help accomplish it are almost always worthwhile, so long as they do not leave some significant stakeholder feeling that his or her pet values have been summarily ignored.

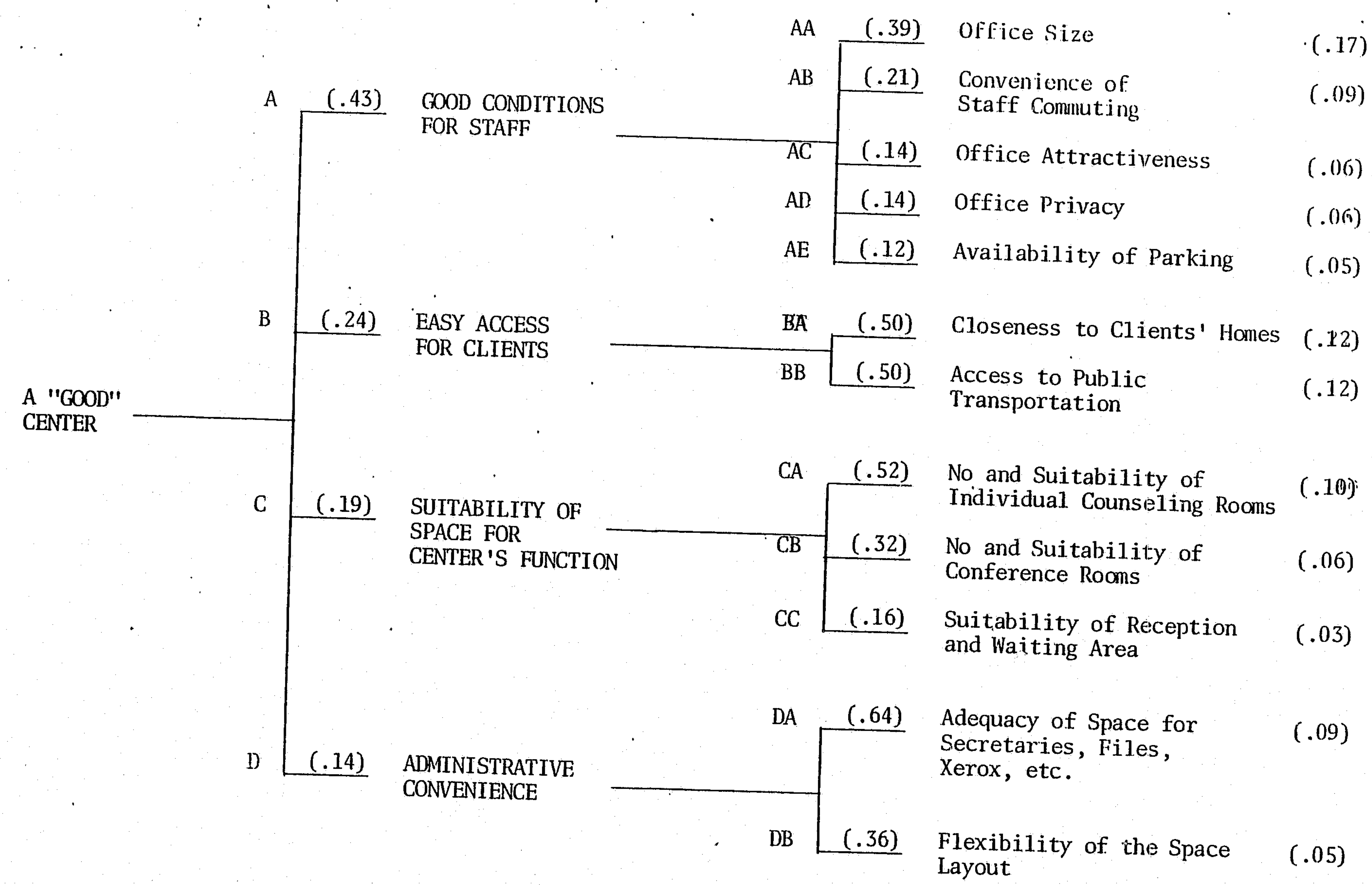
The Director was also able to obtain staff assent to organizing its values into four broad categories, each with subcategories. Such a structure is called a Value Tree. The one that the Director worked with is shown in Exhibit 1. We explain the numbers shortly.

Several questions need review at this stage.

(a) Have all important attributes been listed? Others had been proposed and could obviously have been added. The list does not mention number or location of toilets, proximity to restaurants, presence or absence of other tenants in the same building who might prefer not to have the clients of this kind of organization as frequent users of the corridors, racial-ethnic composition of the neighborhood, area crime rate, and various others. All of these and many more had been included in early lists, and eliminated after discussion. Bases for elimination include not only duplication and unimportance, already discussed, but also that the sites under

EXHIBIT 1 OF CHAPTER 2

A VALUE TREE FOR A DRUG-FREE CENTER



consideration did not vary from one another on that attribute, or varied very little. That is why racial-ethnic composition and crime rate were eliminated. Even an important attribute is not worth considering unless it contributes to discrimination among sites.

For program evaluation purposes, this principle needs to be considered in conjunction with the purpose of the evaluation. If the function of the evaluation is primarily to guide development of the program, then important attributes should be included even if they serve no discriminative function; in such cases, there may be no discriminative function to serve.

The Director was satisfied with the list. It was relatively short, and she felt that it captured the major issues--given the fact that even more major requirements for a new site had been met by prescreening out all options that did not fulfill them.

An obvious omission from the attribute list is cost. For simplicity, we will treat cost as the annual lease cost, ignoring the possibility of other relevant differences among leases.

One possibility would be to treat cost as another attribute, and this is often done, especially for informal or quick evaluations. In such a procedure, one would specify a range of possible costs, assign a weight to that attribute, which essentially amounts to a judgment about how it trades off against other attributes, and then include it in the analysis like any other attribute. We have chosen not to do so in this example, for two reasons. First, some evaluations may not involve cost in any significant way (monitoring, for example), and we wish to illustrate procedures for cost-independent applications of MAUT. Second, we consider the kind of judgment required to trade off cost against utility points to be the least secure and most uncomfortable to make of all those that

go into MAUT. For that reason, we like to use procedures, illustrated later, that permit extremely crude versions of that judgment to determine final evaluation.

While on the topic, we should discuss two other aspects of trading off dollars against aggregated utilities.

The first is budget constraints. If a budget constrains, in this example, the amount of rent the Center can pay, then it is truly a constraint, and sites that fail to meet it must be rejected summarily. More common, however, is the case in which money can be used for one purpose or another. A full analysis would require considering also the loss, in this instance, that would result from spending more on rent and so having less to spend on other things. Such considerations are crucial, but we do not illustrate them here. In order to do so, we would have to provide a scenario about what budget cuts the Director would need to make in other categories to pay additional rents. At the time she must choose among sites, she may not know what these are. Fairly often, the expansion of the analysis required to evaluate all possible ways in which a program might be changed by budget reallocations is very large indeed--far too large to make an easy example. So we prefer to think of this as a case in which the Director's budget is large enough so that, for the range of costs involved, belt-tightening can take care of the difference between smallest and largest. A fuller analysis would consider the programmatic impact of fund reallocation, and could explore the utility consequences of alternative reallocations. This circumscription of the analysis in the interest of making it manageable is very common; relevant issues are and should be left out of every analysis. (An equivalent statement: if it can be avoided, no MAUT analysis should include every attribute judged relevant by any stakeholder. More on this in Chapter 3.) The goal is to enlist stakeholder cooperation in keeping the list of attributes reasonably short.

The other issue having to do with cost but not with the example of this chapter is the portfolio problem. This is the generic name for situations in which a decision maker must choose, not a single option, but a number of them from a larger set. Typically, the limit on the number that can be chosen is specified by a budget constraint. The methods presents in this manual require considerable adaptation to be used formally for portfolio problems, because the decision maker normally wants the portfolio as a whole to have properties such as balance, diversity, or coverage (e.g., of topics, regions, disciplines, problems, etc.) which are not attributes of the individual options themselves. Formally, each possible portfolio is an option, and a value tree relevant to the portfolio, not to the individual options, is needed. But such formal complexity is rarely used. A much more common procedure in portfolio problems is to evaluate the individual elements using methods like those of this Manual, choose from the best so identified, and then examine the resulting set of choices to make sure that it meets the budget constraint and looks acceptable as a portfolio.

You will have encountered such terms as benefit-cost analysis. Such analyses are similar in spirit to what we are doing here, but quite different in detail. Bu introducing into the analysis early assumptions about how non-financial values trade off with money, both benefits and costs can be expressed in dollar terms. We see little merit in doing so for criminal justice or other social programs, since early translation of non-monetary effects into money terms tends to lead to underassessment of the importance of non-financial consequences. The methods we present in this Chapter and in Chapter 7 are formally equivalent to doing it all in money, but do not require an equation between utility and money until the very end of the analysis, if then.

Back to our example. In the initial elicitation of values from the staff, the orderly structure of Exhibit 1, the Value Tree, did not appear. Indeed, it took much thought and trial and error to organize the attributes into a tree structure. Formally, only the attributes at the bottom of the tree, which are called twigs, are essential for evaluation. Exhibit 1 is a two-level Value Tree; that is, all second-level values are twigs. More often, different branches of a Value Tree will vary in how many levels they have. This document later presents a four-level example, and examples with as many as 14 levels exist.

Tree structures are useful in MAUT in three ways. First, they present the attributes in an orderly structure; this helps thought about the problem. Second, the tree structure can make elicitation of importance weights for twigs (which we discuss below) much easier than it would otherwise be, by reducing the number of judgments required. Chapter 4 discusses this further. Finally, Value Trees permit what we call subaggregation. Often a single number is much too compressed a summary of how attractive an option is. Tree structures permit more informative and less compressed summaries. This is further discussed in Chapter 6 and 7.

Exhibit 1 contains a notational scheme we have found useful in value trees. Main branches of the tree are labelled with capital letters, A, B, and so on. Subattributes under each main branch are labelled with double letters, AA, AB, ..., BA, BB..., and so on. This is a two level tree, so only double letters are needed.

Assignment of importance weights.

The numbers in Exhibit 1 are importance weights for the attributes. Note that the weights in Exhibit 1 sum to 1 at each level of the tree. That is, the weights of A, B, C, and D sum to 1. Similarly, the weights of AA through AE sum to 1, as do those of BA and BB and so on. This is a convenient convention, both for elicitation of weights and for their use. The final weights for each attribute at each twig of the tree are easily obtained by "multiplying through the tree." For example, the weight .17 for twig AA (office size) is obtained by multiplying the normalized weight of A (.43) by the normalized weight for AA (.39) to yield .43 x .39 = .17. We discuss multiplying through the tree further in Chapter 4.

The weights presented in Exhibit 1 emerged from a staff meeting in which, after an initial discussion of the idea of weighting, each individual staff member produced a set of weights, using the ratio method described in Chapter 4. Then all the sets of weights were put on the blackboard, the inevitable individual differences were discussed, and afterward each individual once again used the ratio method to produce a set of weights. These still differed, though by less than did the first set. The final set was produced by averaging the results of the second weighting; the average values were acceptable to the staff as representing its value system.

The Director had some reservations about what the staff had produced, but kept them to herself. She worried about whether the weights associated with staff comfort issues were perhaps too high and those associated with appropriateness to the function of the organization were perhaps too low. (Note that she had no serious reservations about the relative weights within each major branch of the Value Tree; her concerns were about the relative weights of the four major branches of the tree. This illustrates the usefulness of organizing lists of twigs into a tree structure for weighting). The Director chose to avoid argument with her staff by reserving her concerns about those weights for the sensitivity analysis phase of the evaluation. She realized, as did the staff, that other stakeholders would also have to be pleased, and that the Probation Department and the Courts would be much less concerned with staff comfort and much more concerned with suit-

ability to function than was the staff. So if the sensitivity analysis should show that the final decision was sensitive to these weights, she planned to elicit some weights from these other stakeholders, to combine them with those of the staff, and thus to bring the weights of twigs bearing on staff comfort down.

Although a common staff set of weights was obtained by averaging (each staff member equally weighted), the individual weights were not thereafter thrown away. Instead, they were kept available for use in the later sensitivity analysis. In general, averaging may be a useful technique if a consensus position is needed, especially for screening options, but it is dangerous, exactly because it obliterates individual differences in weighting. When stakeholders disagree, it is usually a good idea to use the judgments of each separately in evaluation; only if these judgments lead to conflicting conclusions must the sometimes difficult task of reconciling the disagreements be faced. If it is faced, arithmetic is a last resort, if usable at all; discussion and achievement of consensus is much preferable. Often such discussions can be helped by a sensitivity analysis; it will often turn out that the decision is simply insensitive to the weights.

The assessment of location measures or utilities

With a Value Tree to guide the choice of measures to take and judgments to make, the next task was to make detailed assessments of each of the six sites that had survived initial screening. Such assessments directly lead to the utilities in multiattribute utility measurement. The word "utility" has a 400-year-old history and conveys a very explicit meaning to contemporary decision analysts. The techniques for obtaining such numbers that we present in this manual deviate in some ways from those implicit in that word. So we prefer to call these numbers location measures, since they simply report the location of each object of evaluation on each attribute of evaluation.

Inspect Exhibit 1 again. Two kinds of values are listed on it. Office size is an objective dimension, measureable in square feet. Office attractiveness is a subjective dimension; it must be obtained by judgment. Proximity to public transportation might be taken in this example as measured by the distance from the front door of the building to the nearest bus stop, which would make it completely objective. But suppose the site were in New York. Then distance to the nearest bus stop and distance to the nearest subway stop would both be relevant and probably the latter would be more important than the former. It would make sense in that case to add another level to the Value Tree, in which the value "proximity to public transportation" would be further broken down into those two twigs.

As it happens, in Exhibit 1 all attributes are monotonically increasing; that is, more is better than less. That will not always be true. For some attributes, less is better than more; if "crime rate in the area" had survived the process of elimination that led to Exhibit 1, it would have been an example. On some attributes, intermediate values are preferable to either extreme; such attributes have a peak inside the range of the attribute. If "racial composition of the neighborhood" had survived as an attribute, the staff might well have felt that the site would score highest on that attribute if its racial-ethnic mix matched that of its clients. If only two racial-ethnic categories were relevant, that would be expressed by a twig, such as "percentage of whites in the neighborhood" that would have a peak at the percentage of whites among the Center's clients and would tail off from there in both directions. If more than two racial-ethnic categories were relevant, the value would have been further broken down, with percentage of each relevant racial-ethnic category in the neighborhood as a twig underneath it, and for each of those twigs, the location measure would have a peak at some intermediate value. We will discuss these possibilities and explain how to work with them in Chapter 5.

Exhibit 1 presented the Director with a fairly easy assessment task. She chose to make the needed judgments herself. If the problem were more complex and required more expertise, she might well have asked other experts to make some or all of the necessary judgments.

Armed with a tape measure and a notebook, she visited each of the sites, made the relevant measures and counts, and made each of the required judgments. Thus she obtained the raw materials for the location measures.

However, she had to do some transforming on these raw materials. It is necessary for all location measures to be on a common scale, in order for the assessment of weights to make any sense. Although the choice of common scale is obviously arbitrary, we like one in which 0 means horrible and 100 means as well as one could hope to do.

Consider the case of the office size expressed in square feet. It would make no sense to assign the value 0 to 0 sq. ft.; no office could measure 0 sq. ft. After examining her present accommodations and thinking about those of other similar groups, the Director decided that an office 60 sq. ft. in size should have a value of 0, and one of 160 sq. ft. should have a value of 100. She also decided that values intermediate between those two limits should be linear in utility. This idea needs explaining. It would be possible to feel that you gain much more in going from 60 to 80 sq. ft. than in going from 140 to 160 sq. ft., and consequently that the scale relating sq. ft. to desirability should be nonlinear. Indeed, traditional utility theory makes that assumption in almost every case.

Curved functions relating physical measurements to utility are probably more precise representations of how people feel than straight ones. But fortunately, such curvature almost never makes any difference to the decision. If it does, the fact that the difference exists means that the options are close enough so that it scarcely matters which is chosen. For that reason, when an appropriate physical scale exists, we advocate choosing maximum and minimum values on it, and then fitting a straight line between those boundaries to translate those measurements into the 0 to 100 scale. We present a fuller discussion of how to do this in Chapter 5. Formal arguments in support of our use of linearity are far too technical for this document; see Edwards (1980) for citations leading to them.[1]

The Director did the same kind of thing to all the other attributes for which she had objective measures. The attribute "proximity to clients' homes" presented her with a problem. In principle, she could have chosen to measure the linear distance from the address of each current client to each site, average these measures, choose a maximum and minimum value for the average, and then scale each site using the same procedure described for office size. But that would have been much more trouble than it was worth. So instead she looked at a map, drew a circle on it to represent the boundaries of the area that she believed her organization served, and then noted whether each site was close to the center of the area. It would have been possible to use radial distance from that center as an objective measure, but she chose not to do so, since clients' homes were not homogeneously distributed within the circle. Instead, she treated this as a directly judgmental attribute, simply using the map as an aid to judgment.

Of course, for all judgmental dimensions, the scale is from 0 to 100. For both judgmental and objective attributes, it is important that the scale be realistic. That is, it should be easy to imagine that some of the sites being considered might realistically score 0 or 100 on each attribute.

In this example, since the six sites were known, that could have been assured by assigning a value of 0 to the worst site on a given attribute and a value of 100 to the best on that attribute, locating the others in between. This was not done, and we recommend that it not be done in general. Suppose one of the sites had been rented to someone else, or that a new one turned up. Then if the evaluation scheme were so tightly tied to the specific options available, it would have to be revised. We prefer a procedure in which one attempts to assess realistic boundaries on each relevant attribute with less specific reference to the actual options available. Such a procedure allows the evaluation scheme to remain the same as the option set changes. And the procedure is obviously necessary if the option set is not known, or not fully known, at the time the evaluation scheme is worked out.

[1]The reference is given at the end of Chapter 1.

It can, of course, happen that a real option turns up that is more extreme than a boundary assigned to some attribute. If that happens, the evaluation scheme can still be used. Two possible approaches exist. Consider, for example, the attribute "access to public transportation" operationalized as distance to the nearest bus stop. One might assign 100 to half a block and 0 to 4 blocks. Now, suppose two new sites turn up. For one, the bus stop is right in front of the building entrance; for the other, it is five blocks away. The Director might well judge that it scarcely matters whether the stop is in front of the building entrance or half a block away, and so assign 100 to all values of half a block or closer. However, she might also feel that five blocks is meaningfully worse than four. She could handle the five-block case in either of two ways. She might simply disqualify the site on the basis of that fact. Or, if she felt that the site deserved to be evaluated in spite of this disadvantage, she could assign a negative score (it would turn out to be -29; see Chapter 5) to that site on that dimension. While such scores outside the 0 to 100 range are not common, and the ranges should be chosen with enough realism to avoid them if possible, nothing in the logic or formal structure of the method prevents their use. It is more important that the range be realistic, so that the options are well spread out over its length, than it is to avoid an occasional instance in which options fall outside it.

Exhibit 2 represents the location measures of the six sites that survived initial screening, transformed onto the 0 to 100 scale. As the Director looked at this table, she realized an important point. No matter what the weights, site 6 would never be best in utility. The reason why is that site 2 is at least as attractive as site 6 on all location measures, and definitely better on some. In technical language, site 2 dominates site 6. But the table omits one important issue: cost. Checking cost, she found that site 6 was in fact less expensive than site 2, so she kept it in. If it had been as expensive as site 2 or more so, she would have been justified in summarily rejecting it, since it could never beat site 2. No other option dominates or is dominated by another. (Although she might have dropped site 6 if it had not been cheaper than site 2, she would have been unwise to notify the rental office of site 6 that it was out of contention. If for some reason site 2 were to become unavailable, perhaps because it was rented to someone else, then site 6 would once more be a contender.)

Aggregation of location measures and weights

The Director now had weights provided by her staff and location measures provided either directly by judgment or by calculations based on measurements. Now her task was to aggregate these into measures of the aggregate utility or each site.

Exhibit 2 of Chapter 2

Location Measures for the Six Sites

Site Number	Twig Label											
	AA	AB	AC	AD	AE	BA	BB	CA	CB	CC	DA	DB
1	90	50	30	90	10	40	80	10	60	50	10	0
2	50	30	80	30	60	30	70	80	50	40	70	40
3	10	100	70	40	30	0	95	5	10	50	90	50
4	100	80	10	50	50	50	50	50	10	10	50	95
5	20	5	95	10	100	90	5	90	90	95	50	10
6	40	30	80	30	50	30	70	50	50	30	60	40

The aggregation procedure is the same regardless of the depth of the Value Tree. Simply take the final weight for each twig, multiply it by the location measure for that twig, and sum the products. This is illustrated in Exhibit 3 for site 1. In this case, the sum is 48.79, which is the aggregate utility of site 1. It would be possible but tedious to do this for each site. All calculations like that in Exhibit 3 were done with hand calculator programs; the discrepancy between the 48.79 for site 1 or Exhibit 3 and the 48.80 of Exhibit 4 is caused by a rounding process in the program.

Exhibit 4 shows the aggregate utilities and the costs for each of the six sites. The costs are given as annual rents.

The procedures we are about to describe for dealing with outputs like Exhibit 4 are computationally tedious. For that reason, we have prepared a hand-held calculator program that will do them all, taking you directly from Exhibit 4 to Exhibit 6 or Exhibit 8. User instructions for it appear as Appendix D of this Manual.

Now a version of the idea of dominance can be exploited again. In Exhibit 4, the utility values can be considered as measures of desirability and the rents are costs. Obviously, you would not wish to pay more unless you got an increase in desirability. Consequently, options that are inferior to others in both cost and desirability need not be considered further.

On utility, the rank ordering of the sites from best to worst is 425163. On cost, it is 162345. Obviously sites 1 and 4 will be contenders, since 4 is best in utility (with these weights) and 1 is best in cost. Site 5 is dominated, in this aggregated sense, by site 4, and so is out of the race. Sites 3 and 6 are dominated by site 1, and are also out. So sites 1, 2, and 4 remain as contenders; 2 is intermediate between 1 and 4 in both utility and cost. This result is general. If a set of options is described by aggregated utilities and costs, and dominated

Exhibit 3 of Chapter 2

Calculation of the Aggregate Utility of Site 1

Twig Label	Weight	Location Measure	Weight x Location Measure
AA	.168	90	15.12
AB	.090	50	4.50
AC	.060	30	1.80
AD	.060	90	5.40
AE	.052	10	0.52
BA	.120	40	4.80
BB	.120	80	9.60
CA	.099	10	.099
CB	.061	60	3.66
CC	.030	50	1.50
DA	.090	10	0.90
DB	.050	0	0.00
SUMS	1.000		48.79

Exhibit 4 of Chapter 2

Aggregate Utilities and Rents

Site	Utility	Cost (Rent per year)
1	48.80	$ 48,000
2	53.26	53,300
3	43.48	54,600
4	57.31	60,600
5	48.92	67,800
6	46.90	53,200

options are removed, then all of the remaining options, if listed in order of increasing utility, will turn out also to be listed in order of increasing cost. This makes the decision problem simpler; it reduces to whether each increment in utility gained from moving from an option lower to one higher in such a list is worth the increase in cost. Note that this property does not depend on any numerical properties of the method that will eventually be used to aggregate utility with cost.

A special case arises if two or more options tie in utility, cost, or both. If the tie is in utility; then the one that costs least among the tied options dominates the others; the others should be eliminated. If they tie in cost, the one with the greatest utility dominates the others; the others should be eliminated. If they tie in both utility and cost, then only one of them need be examined for dominance. If one is dominated, all are; if one is undominated, all are. So either all should be eliminated or all should survive to the next stage of the analysis. Note that a tie in aggregate utility can occur in two different ways: by accident of weighting, or because all location measures are equal. If all location measures are equal, the lower cost will always be preferable to the higher one regardless of weights, so the higher cost can be eliminated not only from the main analysis, but from all sensitivity analyses. If they tie in aggregate utility by accident of weighting, changes in weight will ordinarily untie them, and so the tied options must be included in the sensitivity analysis.

If two or more options tie in both aggregate utility and cost (a very rare event indeed!), then only one of them should be carried into the next stage of the analysis. Whatever happens to that one will happen also to its twins; they are indistinguishable (for the current set of weights). If the option that represents the tie emerges from the next stage of the analysis looking best, the only way to discriminate it from its twins is by sensitivity analysis, by considering other attributes or both.

Nothing guarantees that the dominance analysis we just performed will eliminate options. If the ordering in utility had been 123456 and the ordering in cost had been 654321 (just the opposite) no option would have dominated any other, and none could have been eliminated. Such perfect relationships between cost and utility are rare, except perhaps in the marketplace, in which dominated options may be eliminated by market pressure.

The decision about whether to accept an increase in cost in order to obtain an increase in utility is often made intuitively, and that may be an excellent way to make it. But arithmetic can help. In this example, consider Exhibit 5. It lists the three contending sites, 1, 2, and 4, in order of increasing utility and cost. In the second column, each entry is the utility of that site minus the utility of the site just above it. Thus, for example, the 4.05 utility difference associated with site 4 is obtained by substracting the aggregate utility of 2 from that of 4 in Exhibit 4: 57.31 - 53.26 = 4.05. Similarly, the cost difference of $7,300 for site 4 is obtained from Exhibit 4 in the same way: $60, 600 - $53,300 = $7,300. The other numbers in the second and third columns are calculated similarly. The fourth column is simply the number in the third column divided by the number in the second.

The numbers in the fourth column increase from top to bottom. This means that all three sites are true contenders. This is not necessarily the case. In Chapter 7, we present what happens and what to do if that column does not increase continuously in tables like Exhibit 5.

Exhibit 5 of Chapter 2

Incremental Utilities and Costs for the Siting Example

Site No.	Utility Differences (Increment)	Cost Differences (Increment)	Cost Incr./Utility Incr.
1	0	0	
2	4.46	$ 5300	$ 1188
4	4.05	$ 7300	$ 1802

The last column of Exhibit 5 also serves another purpose. Since it is the increase in cost divided by the increase in utility, it is a dollar value for one utility point. Specifically, it is the dollar value for one utility point that would be just large enough to cause you to prefer the higher cost site to the lower cost one. If the dollar value of a utility point is less than $1,188, you should choose site 1; if it is between $1,188 and $1,802, you should choose site 2; and if it is above $1,802, you should choose site 4.

But how can you know the dollar value of a utility point, for yourself or for other stakeholders? The judgment obviously need not be made with much precision--but it is, if formulated in that language, an impossible judgment to make. But it need not be formulated in that language. Consider instead the following procedure. Refer back to Exhibit 1. First pick a twig that you have firm and definite opinions about. Suppose it is DA, availability and suitability of space for secretaries, files, Xerox, etc. Now, ask of yourself and of the other stakeholders "How much money would it be worth to improve that twig by so many points". The typical number of points to use in such questions is 100, so the question becomes "How much would it be worth to improve the availability and suitability of space for secretaries, files, xerox, etc. from the minimum acceptable state, to which I have assigned a location measure of 0, to a state to which I would assign a location measure of 100?"

Such a question, asked of various stakeholders, will elicit various answers; a compromise or agreed-on number should be found. Suppose, in this example, that it turned out to be $13,500. Now, refer to Exhibit 3 and note that the twig weight for DA is .090. Consequently, a 100-point change in DA will change aggregate utility by 100 x .090 = 9 points--for this particular set of weights. Note, incidentally, that while the 9 point

number depends on the weights, the judgment of the dollar value of a 100-point change in DA does not. Consequently, if you choose to change weights (as we will in Chapter 7 on sensitivity analysis), you will need to recalculate the value of a utility point, but will not need to obtain a new dollar value judgment of this kind from anyone.

If a 9 point change in utility is worth $13,500 then a 1 point change in utility is worth $13,500/9 = $1500. So, using the weights on which this Chapter is based, site 2 is clearly preferable to sites 1 and 4 since $1500 is between $1188 and $1802.

Let us verify that statement. One way to do so is to penalize the more expensive sites by a number of utility points appropriate for their increase in cost. Thus, if utility is worth $1,500 per point, and site 2 costs $5,300 more than site 1, then site 2 should be penalized 5,300/1,500 = 3.53 utility points in order to make it comparable to site 1. Similarly, if utility is worth $1,500 per point, than site 4 should be penalized by the increment in its costs over site 1, $5,300 + $7,300 = $12, 600, divided by the dollar value of a point; 12,600/1,500 = 8.40 utility points. This makes all three sites comparable, by correcting each of the more expensive ones by the utility equivalent of the additional expense. So now the choice could be based on utility alone.

Exhibit 6 makes the same calculation for all three sites and for three different judgments of how much a 9 point swing in aggregate utility is worth: $9,000, $13,500, and $18,000; these correspond, with the weights used in this chapter, to utility values per point of $1,000, $1,500 and $2,000 respectively. Exhibit 6 is included here not because it is a calculation that the Director would ever need to make, but because it demonstrates that the choices made on the basis of Exhibit 5, which is a calculation she might well need to make, are appropriate.

Exhibit 6 of Chapter 2

Aggregate Utilities after Subtracting Penalties for Excess Cost

Site No.	Value of a 100 point swing in DA (weight = .09)		
	$9,000	$13,500	$18,000
1	48.80	48.80	48.80
2	47.96	49.73	50.61
4	44.71	48.91	51.01

As illustrated in Exhibit 6, a utility value of $1000 per point makes site 1 best, a utility value of $1500 per point makes site 2 best, and a utility value of $2000 per point makes site 4 best. Note, however, that the differences in corrected utilities are relatively small. This is normal, and is one reason why we make no strong case for using such calculations to go from Exhibit 4 to Exhibit 6. Elimination of non-contenders is usually both more important and easier to do than selection among those that survive the elimination process, since the surivivors are likely to be close enough to one another in attractiveness so that no choice will be disastrous.

Sensitivity analysis

The Director of the Center had some doubt about the weights her staff had given her. She therefore considered various other weights. We present the details in Chapter 7 on sensitivity analysis. She found a set of weights that make site 5 best in utility, and another for which site 2 is best.

Chapter 7 also presents a minor example of exploring sensitivity to location measures. But the Director was relatively well satisfied with the location measures she was using, and felt no need to change them--and also felt that there were so many that she was unsure which ones to change.

At this point the Director felt she had enough information and analysis to make her recommendation of site 2. Details of how the sensitivity analysis convinced her that site 2 was best are presented in Chapter 7.

CHAPTER 3 SUMMARY

Chapter 3 concentrates on the problem of identifying the stakeholders and eliciting the value attributes from them. A distinction is made between actors, people who make decisions about programs and perhaps take direct action to change programs, and people who are affected or impinged on by the program, either directly or indirectly. Both are important stakeholders but have different roles. The actors are concerned with the relevance of the evaluation to the final decision making process. Thus they can best explicate what values should be considered in their decisions. People who are affected by the program are more concerned with how the programs affects people and thus should concern themselves with the values of these affected people. Techniques for eliciting the value attributes from stakeholders are discussed and illustrated. The structuring of the value attributes into a value tree is demonstrated and the problem of multiple stakeholders with perhaps different value structures is discussed. The advantage of having a common value structure is emphasized, but a common structure is not a necessity to carry out a successful MAUT. Often differences between stakeholders can be described as differences in weights assigned to the attributes. When value attributes are formulated as program objectives then measures have to be defined of how well these objectives have been obtained. A specific example is given for the Community Anti-Crime (CAC) Program. Three other examples of identifying stakeholders, eliciting value attributes from them, and structuring those attributes are given from evaluations in civil justice, juvenile justice, and school desegregation.

CHAPTER 3

THE SOURCE OF THE VALUE ATTRIBUTES: THE CONCEPT OF STAKEHOLDERS, AND THE STRUCTURE OF VALUES

The approach to evaluation, Multiattribute Utility Technology (MAUT), advocated in this document relies heavily on the measurement of utility (subjective value of the entity or project being evaluated). But values tend to be personal and therefore are usually associated with individuals or groups. The generic name for such individuals and/or groups is stakeholders, people who have an interest or a stake in the program or the entity being evaluated, and who are important enough so that their interests should be considered. Stakeholders are often at the policy level of decision making, concerned with the program goals and objectives and the consequences of program operations. The main tasks performed by the stakeholders in evaluation is to identify and structure the value attributes important to the evaluation, and to assign importance weights to these attributes, the topic discussed in Chapter 4.

This type of evaluation also used experts who may or may not be stakeholders. They are experts in the sense they are knowledgeable about a program or entity to be evaluated. They know why the program is in existence; what it is supposed to be doing; and how it is doing it. Experts are often administrators or staff members of a program or somebody who is a close observer of a program. The main task for experts is to assign location measures, the topic discussed in Chapter 5, for those attributes for which objective measures are not readily available.

In the site location example, the stakeholder may be a single individual, the director of the drug counseling center, or it may be a group or committee charged with the responsibility of relocating to a new site such as staff members of the center. Other stakeholders include the sponsor, the Court, the Probation Department, and the Center's clients.

In this chapter we concentrate on the problem of identifying the stakeholders and eliciting the value attributes from them. We also illustrate how these attributes might be structured or organized. All the examples are in the area of social program evaluation, primarily the evaluation of criminal and civil justice programs.

Who are the stakeholders?[1]

We distinguish between actors -- people who may make decisions about the program and perhaps take direct action to change it, and those people who are affected or impinged upon by the program, either directly or indirectly. Both of these types of stakeholders should be involved in an evaluation, although their roles are different. The actors make the decisions to which the evaluation should be relevant. Therefore, they can best explicate what values should be considered in their decisions.

People who are affected by a program have a different role. Their values enter less directly -- but not necessarily less importantly -- into decisions. Among the values relevant to decisions (presumably high among them for many decisions) is how the program affects people. Thus, the evaluation can enhance programmatic decisions by explicitly and accurately representing the values of these affected people.

[1]Much of the material in this chapter was prepared by Dave Seaver and Kurt Snapper of Decision Science Consortium and Planning Systems Associates, respectively.

Criminal justice programs, like any social programs, will have many stakeholders. The primary actors are usually the most easily identified. Begin by examining the organizational structure of the program. If the program has multiple local projects, are these projects homogeneous enough to be considered as involving a single set of stakeholders, or is the program an "umbrella" under which are varied projects whose decision makers will have different values and objectives? Single local projects may, in fact, have two sets of decision makers: those responsible for managing the project, and those responsible for the general administration of the organization(s) conducting the project.

In rare instances, the program being evaluated will be a unitary program with a single level of decision-making authority. In the more common case, the next level up in the decision making hierarchy depends on the type of program being evaluated. For umbrella-type programs, it may be the program office in the non-governmental or governmental agency administering the program. For other programs, it may be a local agency, either governmental or private.

If local programs are funded with block grant money, the State Planning Agency (SPA) which administers the block grants may also be a relevant stakeholder. Within the Federal sponsoring agency, there may be stakeholders in addition to the program office. The administration may be a stakeholder, and the office funding the evaluation (not necessarily the same as the program office) may be a stakeholder. Then, of course, there is the relevant legislature, one of the ultimate organizational decision makers, which will probably need to be considered a stakeholder if the program has been legislatively mandated.

The above discussion does not fit all criminal justice programs. It is intended only to illustrate how an evaluator might begin to identify just who the relevant decision makers are. Stakeholders who are not decision makers, but rather are people or organizations affected by the program, may be more difficult to identify. The clients of the program are such stakeholders, as are other members of the target population who for one reason or another are not actual clients. If the target population does not coincide with the population of the target area, the remainder of the target area population may be affected by the program. Juvenile programs would create this category of stakeholder, such as parents and teachers of the juvenile offenders.

Various components of the criminal justice system should also be considered as stakeholders. Police agencies, courts, prosecuting attorneys' offices, probation agencies, and correction agencies may be affected by a program for which they have no direct decision making authority.

Beyond these populations and agencies, stakeholders may include special interest groups that have a particular concern with the program being evaluated. The nature of these groups would depend on the type of program being evaluated. The research community might also be considered a stakeholder because information could be produced by the evaluation that enhances the knowledge about a particular type of criminal justice program.

Most evaluations will have neither the resources nor the need to work closely with all identified potential stakeholders. Enough interaction with the less significant stakeholders is necessary to ensure their representation in the evaluation. However, any stakeholder likely to use evaluation in making decisions requires careful attention from and extensive interaction with the evaluation staff. Such stakeholders need to feel that the evaluation, or at least a particular part of the evaluation, is being done for them -- and it should be. Identifying such decisions and the stakeholders who make them may easily tax the knowledge and political skills of evaluators up to or beyond their limits.

Identification of stakeholders will depend in part, obviously, on the purpose of the evaluation. Stakeholders for an evaluation of the feasibility of a program before it is installed will include the legislators or others, typically public officials in the case of criminal justice programs, who are responsible for the fact that the program is under consideration. They will also include speakers for the organizations (e.g., police, courts) that may be influenced by the new program if it is implemented. If at least some program staff are already selected, they are stakeholders. In this as in all other criminal justice programs, stakeholders include representatives of the public interest and of the clients whom the program may affect.

Essentially the same list of stakeholders applies to evaluations being conducted during the early stages of a program to see if it is on track -- sometimes called formative evaluations. However, the emphasis is somewhat different. Program people and those directly impacted by the program within the criminal justice system are especially important in this kind of evaluation. The same is true for monitoring. In both of these cases, too, the sponsor(s) are important.

The most traditional idea of evaluation is that it is concerned with measuring the external impact of a program -- to find out whether or not it is fulfilling its goals. For such an evaluation, again the list of stakeholders is much the same, but the emphasis changes. In evaluating the consequences of a program, stakeholders from outside the program have much more importance than is the case for monitoring. In some such cases, it is useful to treat independent academics and others as though they were stakeholders, as is the case in use of review panels. Those external to the program who are affected by it, including criminal justice agencies of various kinds, representatives of public interests, and representatives of client interests, are especially important. If the program is a topic of debate, obviously the sides to that debate are stakeholders.

Although everyone wants to have his or her finger in public pies, the implication of the preceding paragraphs is that evaluations intended basically to guide the internal workings of a program need less heavy involvement of outside stakeholders than evaluations intended to guide major programmatic decisions. But the other implication is that, to whatever limits money, time, and cooperativeness make necessary, it is always better to include too many stakeholders than too few. Normally, the evaluator will be in conflict; well aware of omissions from the list of participating stakeholders, he or she will still find that list inconveniently long. As usual, difficult choices must be made.

Eliciting Value Structures from Stakeholders

Involvement, and getting people to pay attention to results (regardless of whether they agree or disagree with them), is enhanced by communicating with both critical actors and representatives of affected groups at the outset of the evaluation. It is crucial to determine what types of decisions may be made, and at least crudely to consider what factors may critically affect them. Political issues of concern to legislators, for example, are of at best indirect concern to project managers and clients. The first step, therefore, is to query groups and individuals about what values from their perspective the program may affect.

One should perhaps begin with staff (administrators and managers) and clients of the program, but should also ensure that critical actors (e.g., legislative figures), with a view of the topic are queried. One advantage of beginning with staff and clients is that the evaluator needs detailed knowledge about the program, and this is a good source. The elicitation of relevant values can begin with statements about the objectives of the program, e.g., reduce crime, or reduce recidivism. Presumably, the program is expected to produce a change on these value attributes. But care must be taken not to consider only the programmatic objectives. Other values may also be affected by the program, so care must be taken to discuss possible "spinoff" effects with stakeholders. This is particularly true of value attributes that the program may affect negatively, since stakeholders, especially if interviewed early, may be predominantly supporters of the program.

One pattern, which we have observed, is that legislators will express objectives about performance by the agency implementing the program, the agency will have objectives both about its performance and about the performance of individual projects funded through the program, and the individual projects will have their own objectives. More generally, everyone seems to have objectives for that part of that program for which they are immediately responsible. Given the complex administrative and management arrangements of typical programs, this implies several sets of values than can be expected to coincide to some degree but certainly not completely.

We illustrate this by the evaluation of the Community Anti-Crime (CAC) program administered by the Office of Community Anti-Crime Programs (OCAP) within the Law Enforcement Assistance Administration (LEAA). As implemented, its basic idea was that local community organizations interested in adopting some mix of strategies for reducing the incidence of

EXHIBIT 1 OF CHAPTER 3

ILLUSTRATIVE STATEMENTS OF OBJECTIVES (VALUE ATTRIBUTES)
COMMUNITY ANTI-CRIME PROGRAM (CAC)

Community Anti-Crime Program (Office of Community Anti-Crime Programs)

- What is the actual degree of participation by community residents in this project?
- If participation is low, what accounts for this?
- How many new individuals are being mobilized by the project?
- What are community residents' perceptions of projects?
- How effective are projects in "leveraging" other local or community groups into cooperative and/or compatible anti-crime or neighborhood development efforts? What catalytic effects do projects have?

Congress

- What innovative program strategies have been developed to facilitate and maximize community involvement?
- How much autonomy do community groups and projects have from local power brokers and political pressures?

Community Work Group

- What role do neighborhood residents play in the development, planning, and implementation of community crime prevention activities that have received federal funding?

NILECJ -- National Institute of Law Enforcement and Criminal Justice

- To establish a Community Anti-Crime Project within communities that is linked directly with the residents and in which the views and actions represent those of the residents.
- To create or support anti-crime organizations having officials who come from the community and who are representative of the community in their demographic characteristics.
- To produce and implement anti-crime activities through these projects that have their origins in the community and reflect the physical, social, and psychological needs of the community.
- To establish a closer and abiding communication link with the residents to create and maintain feedback on their attitudes and behavior towards anti-crime activities so as to change or modify them as the needs change or are modified.
- To determine and establish a means to enlarge the knowledge of the residents of their roles in preventing crime and for educating them on the crime prevention roles of others, such as the police.

crime could apply for money to LEAA's OCAP. One goal of the program, a goal encouraged by Congress, was to reduce red tape. Consequently grants went directly from LEAA to the applying organization--which might not even have been sufficiently highly structured to meet normal Federal standards of stability and of capability for financial accounting. Another goal, also encouraged by its Congressional supporters, was to encourage the invention of innovative anti-crime strategies. This goal, in turn gave OCAP every reason to seek a wide variety of different activities among its grantees. Moreover, since another program objective was community involvement, any simple measure such as crime statistics for the relevant communities would miss a major point of the program. Also, the National Institute for Law Enforcement and Criminal Justice (NILECJ) was a distinct unit from OCAP within LEAA; it was concerned with research questions and the evaluation process itself. The expressed concerns of Congress, OCAP, NILEJC, and any Community Work Group involved with crime prevention necessarily reflected their differences in perspective concerning what an anti-crime program is all about, but these were not necessarily inconsistent for developing the evaluation from the viewpoint of a MAUT model. The problem was a practical one of culling out, from each group, those objectives that were in fact relevant to the intended purpose of the evaluation model. If the purpose is to assess program effectiveness, only values related directly to the effects of the program should be included. Unless the various organizational entities are working at explicit cross-purposes, it should be possible to structure the values into a single, internally-consistent MAUT model. Also, it will often be possible to develop a MAUT model structure that more-or-less parallels organizational

structures themselves. We have two specific recommendations about how to do this.

1. <u>Separate value attributes from topics people are merely curious about</u>. If one is not careful, asking stakeholders for attributes will yield a hodge-podge of topics. Included will be topics the respondent is curious about ("It would be nice to know what kind of communities are especially likely to get involved in this type of program.") but that are clearly not value attributes as well as topics that seem plausible candidates for value attributes ("We want to make better decisions about which applicants to fund in the future") but on reflection are not appropriate. In this latter case, the quality of future decisions about applicants has no bearing on how the effectiveness of current projects themselves should be assessed.

Two kinds of extraneous topics are likely to arise as pseudo-attributes The first, often of interest to researchers, is information one would like to have about the program or related social phenomena, that is really quite irrelevant to assessing the utility or value of the program itself. The second has to do with the utilization of the evaluation itself. People want to disseminate evaluation results, either to make better decisions or in order to argue some case before the decisions are made. So they want to determine what data (and in what form) the evaluation should feed into the decision process. Elsewhere we discuss in more detail the dynamic nature of program policy-making, planning, and management; but at the present it is sufficient to note that programmatic monitoring (and relatively informal decisions about whether the program is doing "well" or not) typically will rely much more heavily on the programmatic data than will many fundamental policy decisions. Policy decisions, for example, will typically involve comparisons among other programs which are competing for money or other support, and often get tied up in political processes and negotiations. The actors may not have articulated for themselves what decisions may be made, and the task of probing them may (and often will) fall to the evaluator.

To help ferret out legitimate programmatic objectives from extraneous topics, it is useful to ask repeatedly whether a given attribute actually relates to the value of the program, that is, if it varies so does one's opinion about the value of the program. Another useful device is to list some relevant measures or statistics pertaining to the "value attributes", and ask whether these are valid indicators of programmatic value. Often, consideration of specific statistics (i.e., operationalization of the value attribute) will help sharpen perceptions about what a given value attribute actually means. It is useful for the evaluators to reflect for a while on the lists of candidate value attributes to attempt to identify extraneous items. And of course the final check is always to go back to the stakeholders with some specific questions about which items should or should not be counted as true value attributes for the program. This often helps to make previously unstated values explicit.

2. <u>Standardize terminology: eliminate "distinctions without a difference."</u> Essentially the same value attribute can be stated in many ways. When querying stakeholders standardizing terminology is useful, and most commonly held value attributes can be identified in this manner. Sometimes, however, it is useful to consider specific measures or

statistics, and ask whether these in fact seem to reflect fundamentally different value attributes. That is, one wishes to avoid "distinctions without a difference."

A specific example may help make the point. Exhibit 1 is a partial listing of "objectives" for the Community Anti-Crime (CAC) Program, as specified by four different groups who participated as stakeholders in the CAC evaluation.[1] Note that some of these "objectives" are stated as questions in which there is presumably some interest; note also that the phrasing of these questions seems to reflect somewhat different perspectives.

Only the objectives as stated by the fourth group, NILECJ, in fact, sound obviously like programmatic objectives. Nevertheless, underlying each of the group's objectives was a common value: <u>projects should be representative of neighborhoods and residents</u>. Thus, projects "controlled" by a local leader not responsive to concerns of residents, or ones which failed to reflect the "corporate" objectives of the communities themselves should be scored low in regard to representativeness.

For the CAC Program, the term "representative" was agreed to as appropriate to capture the meaning of each group, and passed into common usage rather quickly. Our point here is that there may be obvious semantic differences when there is, in fact, underlying consensus such that it is inappropriate to make a distinction in value attributes.

It is very important to clarify definitions before an attempt is made to structure the value attributes into a MAUT model. Otherwise,

---

[1] This evaluation was conducted by Decision Sciences Consortium under the direction of Dave Seaver and Kurt Snapper.

the model is unlikely to include some value attributes that produce the severest form of double-counting: they are just plain redundant.

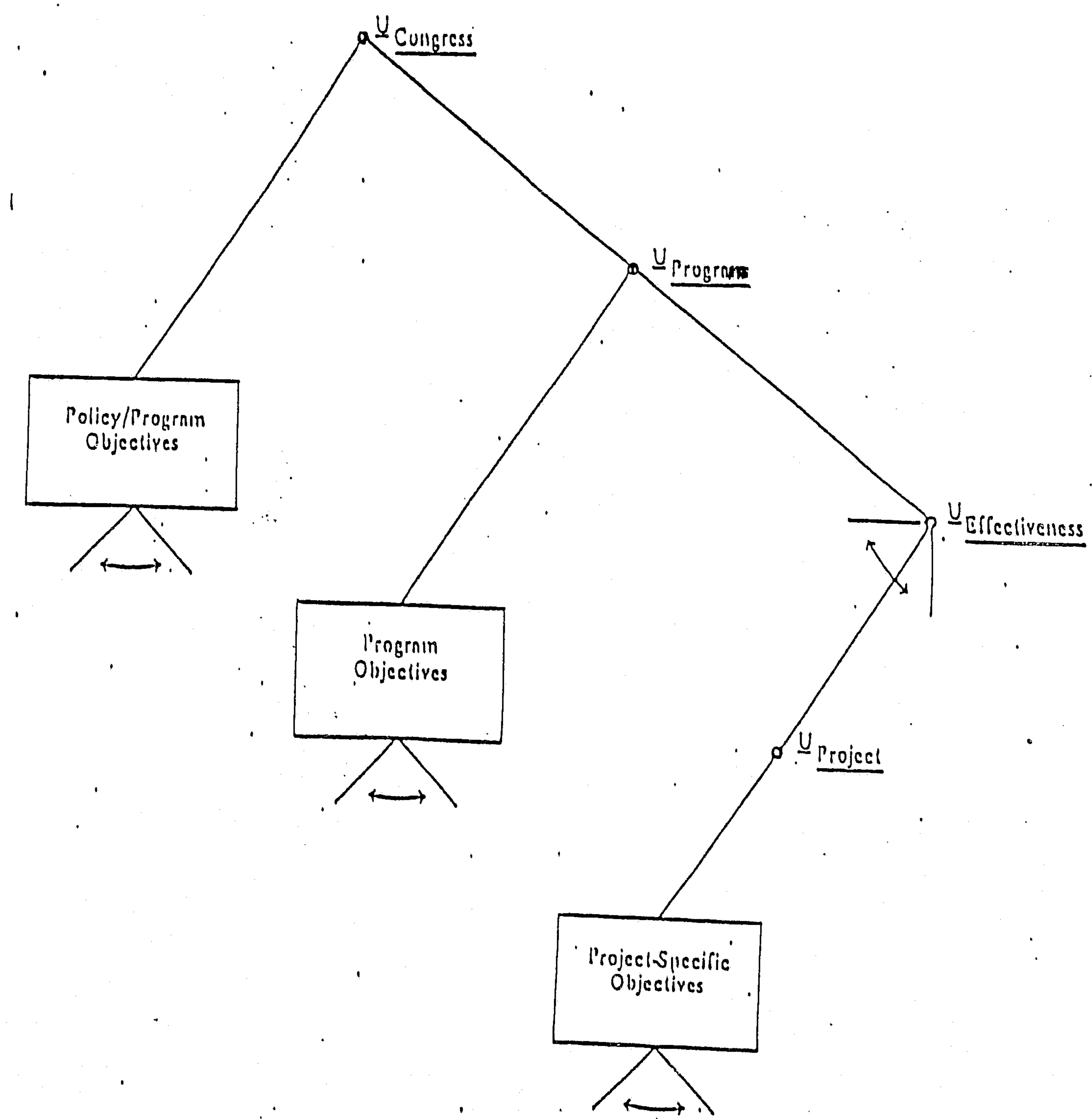
Structuring the MAUT Model

Models should involve a decomposition of program overall value into its component attributes. Earlier we suggested that there may be a tendency for stakeholders to articulate objectives pertaining to that aspect of the program with which they are most directly involved. This has two implications. First, each group will give only part of the total set of value attributes, and one must amalgamate across stakeholders to get the whole. Second, the structure of the MAUT model itself will often parallel the organizational relationships among the stakeholder groups. Specifically, it will often be possible to organize the MAUT model so that different clusters of value attributes may be more-or-less imputed to identifiable groups.

This situation is shown in Exhibit 2 for the CAC program. For this evaluation the value attributes were formulated as "objectives." "Policy/Program Objectives" are those Congress had for the program whereas "Program Objectives" are those the program office set both for itself and for constituent projects funded under the program. Finally, "Project-Specific Objectives" were those the local projects set for themselves. This general model, in fact, described accurately the MAUT model developed for the CAC evaluation, which is shown in Exhibit 3. In this Exhibit the Office of Community Anti-Crime Programs (OCAP) was the program office, and it had "Results Sought" objectives in two main areas. Similarly, there were Program/Policy Objectives specified by Congressional staff. Exhibit 3 does not show them, but submodels would have to be

EXHIBIT 2 OF CHAPTER 3

THE CLUSTER OF OBJECTIVES (VALUE ATTRIBUTES) FOR CAC



developed for various specific projects. These would be represented by further branching under "Project-Specific Objectives" (BB).

Multiple Stakeholders and a Common Value Tree

Nothing guarantees that the values elicited from different stakeholders will be similar enough so that they can all be arrayed in the same value tree. Indeed, attempting to do just that is one of the most demanding tasks that a user of MAUT may have to face.

So far, we have not encountered any instances in which, with enough hard thought and time to discuss the matter with stakeholders, this could not be done. It is obviously extremely useful to do it. While different stakeholders using different Value Tree structures can perform, or have performed for them, the same MAUT evaluation of the same entities, there is no obvious way of relating one evaluation to another unless the tree structures are the same. If the tree structures are the same, then the differences among stakeholders can be described as differences in weights and perhaps also in preferred location measures. Differences in preferred location measures can also be described as differences in weights, since various different measures that purport to measure the same kind of value can be considered as simply another level of the value tree. We consider conflicts about weights to be easier both to interpret and to discuss and perhaps resolve than conflicts about structure, mostly because they lend themselves so easily to compromises. Some unpublished experience backs this up.

Among the devices that an imaginative evaluator can use to make one larger value tree out of those elicited from several stakeholders, the most obvious (requiring, however, cooperation from the stakeholders) is combining categories from different stakeholders by relabeling or otherwise recognizing intellectual similarity behind verbal difference. Another, helpful in persuading a stakeholder to include values that seem unimportant, is the point that that stakeholder can make any value utterly unimportant simply by assigning it zero weight. Still

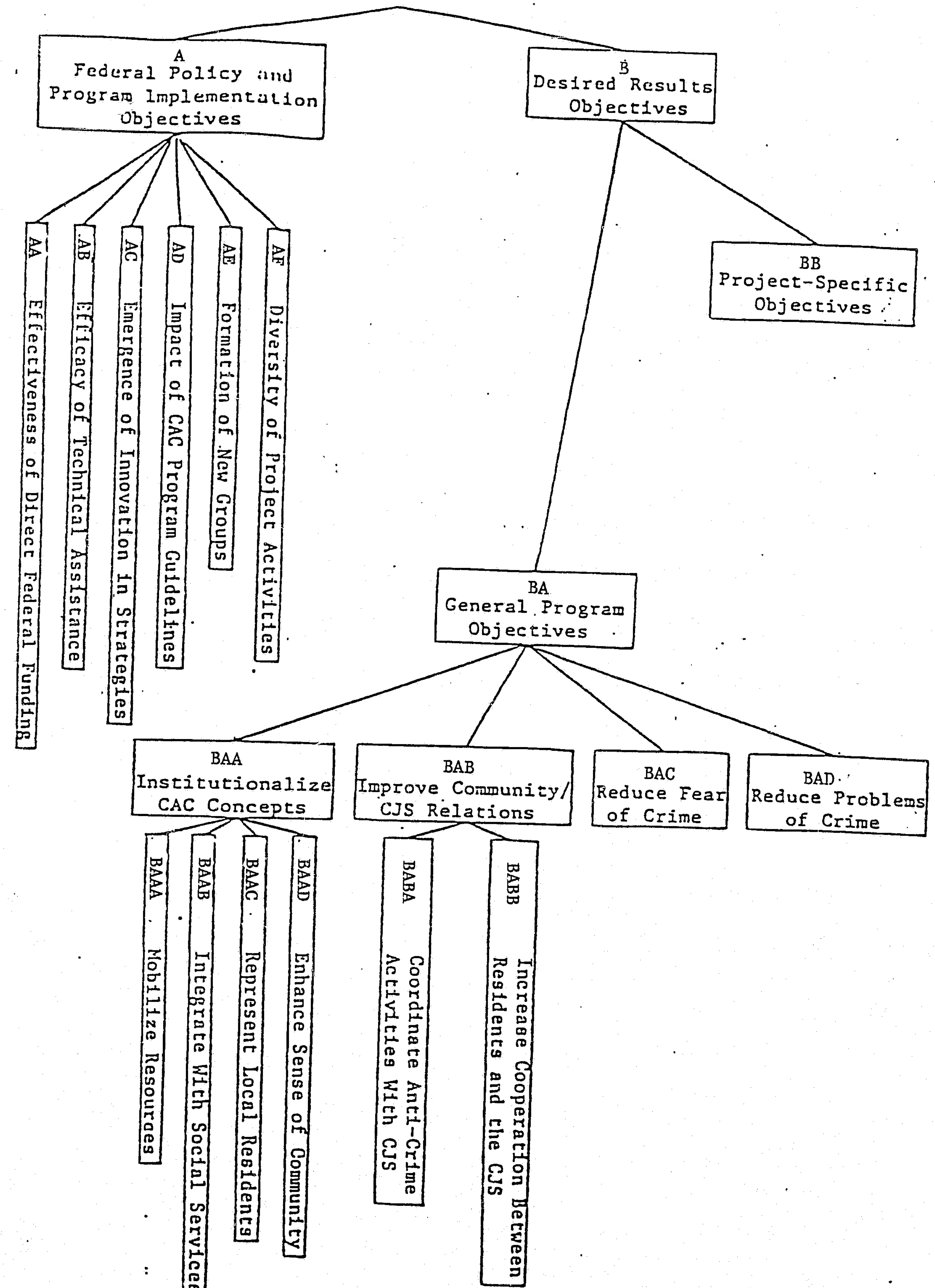
another, illustrated in the CAC example, is to include the values of stakeholders at various levels of an organizational hierarchy by making the hierarchical structure of the Value Tree correspond more or less to that of the organization. Indeed, so blunt a device as simply listing values of different stakeholders as different branches of a Value Tree is always available as a last resort. This Manual will hereafter assume that the evaluator, though working with different weights from different stakeholders, has managed to boil the values down to a common Value Tree.

Identification of Measures in a MAUT Model

When value attributes are formulated as program objectives as in the CAC program, a logical next step would be to identify measures of how well these objectives have been obtained. While this is primarily the topic for a later chapter (Chapter 5), it might help clarification if we illustrated this in the context of the CAC evaluation currently underway. In program evaluation, attainment of some objectives must be assessed wholly judgmentally. That is, experts will make direct ratings about how fully objectives are attained. We will confine ourselves here to the usual case, in which there are data which can be used to assess actual degree of attainment.

To illustrate, Exhibit 4 shows measures which were identified as relevant to the objectives shown in Exhibit 3. Two or three measures are shown for each objective, selected from lists which are typically much larger.

In deciding which specific measures (from a list of several alternatives) should be actually included in the model, there is little premium for choosing



MEASURES OF THE CAC OBJECTIVES (VALUE ATTRIBUTES)

AA Effectiveness of Direct Federal Funding Strategy

- Number of organizations receiving specific CAC funding. (This means that the organizations receive a specific amount of money which they effectively control.)
- Average percentage of grant funds spent on administration.
- Average percent of grant funds spent on equipment.

AB Efficacy of Technical Assistance to CAC Projects

- Percentage of projects receiving TA prior to grant award.
- Percentage of projects satisfied with the TA they received.
- Ratings of TA by evaluation staff as part of Levels II and III Observer Reports.

AC Emergence of Innovation in Project Strategies

- Percentage of projects with innovative activities.
- Percentage of projects that do not just continue or expand previous activiti
- Percentage of projects with activities not specified in AIR MIS.

AD Impact of CAC Program Guidelines on Projects

- Percentage of activities not suggested in the Guidelines.
- Ratings by projects of usefulness and restrictiveness of the Guidelines.

AE Formation of New Groups

- Number of new coalitions formed to receive grants.
- Number of new organizations formed.
- Average time in existence for new organizations.
- Average number of members of new organizations.

AF Diversity of Project Activities

- Percentage of projects with components for youth.
- Percentage of projects with components for the elderly.
- Percentage of projects conducting activities addressing both "causal factors" and "opportunity reduction."

BAAB Institutionalize Project Activities: Integrate with Other Social Services

- Percentage of projects being run by organizations that also provide other services.
- Percentage of projects exchanging referrals with organizations providing other social services.
- Ratings in evaluation staff Observer Reports of relationships with other community organizations.

BAAC Represent Local Residents

- Percentage of projects with resident involvement in policy decision-making.
- Percentage of projects with resident involvement in budget decision-making.
- Percentage of projects with resident involvement in staff selection.

BAAD Enhance Sense of Community

- Percentage of residents who think there has been an increase in the past year in people in the neighborhood helping each other.
- Percentage of residents who have increased the number of neighborhood residents with whom they are acquainted in the past year.
- Percentage of residents who have joined a block club or neighborhood organization in the past year.
- Ratings of neighborhood improvement by evaluation staff on Observer Reports.

BABA Coordinate Anti-Crime Activities with CJS

- Percentage of projects establishing a link with the police.
- Percentage of projects with an active role for the police.
- Percentage of projects with activities (other than above) directly involving the CJS.

BABB Improve Resident Attitudes Toward the Police

- Percentage of residents who think police respond faster to calls than they did a year ago.
- Percentage of residents who think police/community relations have improved in the past year.
- Percentage of residents who think police treatment of residents has improved in the past year.

EXHIBIT 4 (continued)

BAC Reduce Fear of Crime

- Percentage of residents who now feel safer out alone in the neighborhood during the day than they did a year ago.
- Percentage of residents who now feel safer out alone in the neighborhood after dark than they did a year ago.
- Percentage of residents who are now less often very worried about personal attacks during the daytime than they were a year ago.

BAD Reduce Problem of Crime and Victimization

- Percentage of residents who think the severity of robbery as a problem in the neighborhood has been reduced in the past year.
- Percentage of residents who think the severity of stranger assault as a problem in the neighborhood has been reduced in the past year.
- Percentage of residents who think the severity of burglary as a problem in the neighborhood has been reduced in the past year.

BB Project Specific Objectives

- Local projects can be expected to have objectives of their own that do not coincide exactly with the general program objectives. Since this Program recognizes the diversity in various communities, attainment of these project-specific objectives that reflect this diversity is desirable. Of course, no general statement of these objectives can be presented. Rather, some individual projects are being selected and their specific objectives will be delineated and evaluated in an appropriate manner.

  The objectives and measures for selected projects enter into the overall model as submodels, branching under "Project-specific objectives" in Exhibit 3.

and including large numbers of measures. Rather the emphasis should be on including those thought to be the most reliable and valid measures. Variables of marginal relevance or suspect validity should generally not be included. This advice is inappropriate, of course, if such measures are the best available for some important twig of the value tree.

A Comment about Different Measures

Often, there will be multiple, imperfect measures of a given objective; alternative measures may, in fact, be proposed as relevant by different stakeholders. Sometimes the disagreements, especially in an area of hot political debate, about appropriate measures may be severe. This is most likely to happen if disagreeing stakeholders are committed to different answers to a policy question, and therefore want to use measures that will make a particular answer look good or bad. We do not have an example of this in criminal justice but a case in point is the evaluation of nuclear power plants. An obvious value relevant to evaluation of such a plant, either prospectively or retrospectively, is accidents. Subdivisions under that value might be number and severity. But subdivisions under each of those headings could easily be topics of intense debate. Is a malfunction that causes the emergency mechanisms of the reactor to shut it down an accident, or an event incidental to normal operation? If a malfunction occurs and the emergency mechanisms fail to work, but alert response by the reactor crew shuts the reactor down without damage or release of radioactive materials, is that an accident? Suppose the same scenario occurs, but some damage to the core occurs and some radioactive materials are released within the pressure container, but not beyond it? Similarly, consider measures of severity of an accident that does release radioactive materials beyond the pressure container. Expected number of fatalities is a familiar measure for industrial accidents. A more common one in the nuclear-reactor field, however, is expected fatalities from the maximum credible accident.

We do not offer any suggestions about how to resolve such disputes, other than the obvious one that whenever more than one measure of the same value is proposed, it seems natural to treat these as complementary to one another, rather than competitive, and therefore to use all. Then the problem becomes one of weighting them for aggregation, rather than of deciding which to include and which to exclude. We have seen little evidence that this suggestion, which seems very practical to us, is likely to be used in the nuclear debate. Pro-nuclear advocates want measures that make nuclear plants look safe; anti-nuclear advocates want measures that make nuclear plants look dangerous. The techniques described in this manual are unlikely to help in such polarized, institutionalized conflicts. They are better suited to more open-minded stakeholders. Fortunately, our experience in using MAUT in both types of conflicts suggests that participants in the nuclear-power controversy occupy positions with uniquely extreme intransigence; even conflicts about forced school busing are milder.

Three Other Examples

We present three more examples of stakeholders and the origin of the value attributes. One is from the area of civil justice, another, juvenile justice, and the other from educational policy implementation. All evaluation programs were designed by the Social Science Research Institute (SSRI) of the University of Southern California (USC).

Example: The Evaluation of the Office of the Rentalsman. The Province of British Columbia, Canada is experimenting with an interesting alternative to the courts in resolving disputes between landlords and tenants: The Office of the Rentalsman (OR). As a dispute resolution mechanism, the OR was unusual in its inception and operation in the sense that it was set up as a distinct alternative to the Court and operated completely separate from the Court. The legislation that created the OR removed jurisdiction of landlord-tenant disputes involving residential property (commercial property was excluded from the act) from the court and gave it exclusively to the OR. The OR operates quite differently from the Court, in resolving landlord-tenant disputes. It provides an extensive information service, including toll free answering service in which landlords and tenants can call in and ask questions about what their rights and limitations are. The Court provided no such function. A dispute can be brought to the OR in many ways such as a telephone call, via the mail, or just walking into an OR office and filing a complaint. The Court, on the other hand, had only one mechanism for filing a complaint. A great deal of mediation goes on in the OR, whereas, the Court only adjudicated. If a case is serious enough, a hearing is held by an OR officer, and these hearings are similar to Court hearings, although considerably less formal.

The Stakeholders and the Attributes

The evaluation of the OR was concerned not only with measuring how effective the OR was or might be in doing its job but also with obtaining some information about how it compared with the previous mechanism for resolving landlord-

tenant disputes, namely the Court. Thus the identification of the stakeholder-experts consisted of seeking out those persons who were knowledgeable about the current operation of the OR and also the Court procedure for dispute resolution. Among the stakeholders selected were judges, representatives of landlord and tenant organizations, legal scholars concerned with landlord-tenant issues, and, of course, OR staff members. Each person was individually asked to list his or her attributes of importance for the operation of the OR. The attributes, 16 in all, that finally emerged from this process along with their respective definitions are listed in Exhibit 5. It should be noted, and this is typical, that there is considerable "overlap" in the attributes. For example, the attributes of FAIRNESS and IMPARTIALITY might be considered the same thing, although there are subtle differences in these two attributes that, in this study at any rate, the experts wished to retain. The attributes given in Exhibit 5 can be placed in a value tree and this is done in Exhibit 6.

There is another point to be noted in Exhibits 5 and 6. Two attributes listed by the stakeholder, COST (to the users) and EXPENSE (to the institution or taxpayer) are not always considered attributes of importance. As mentioned in Chapter 2, cost considerations often do not enter until presentation of the final results. In this particular evaluation, making direct statements about cost of services in terms of dollars was not possible. Almost all of the stakeholders, however, included either COST or EXPENSE in their list of attributes Thus they were retained in the evaluation as attributes.

The listing of attributes and the specification of a value tree can get quite complicated as the next example will illustrate.

EXHIBIT 5 OF CHAPTER 3

THE VALUE ATTRIBUTES AS IDENTIFIED BY THE EXPERTS

FOR THE OFFICE OF THE RENTALSMAN (LISTED ALPHABETICALLY)

Accessibility -- ease of registering requests and complaints; taking into consideration procedural complexity, hours/days available, and physical location.

Consistency -- the degree to which the institution's decision reflect general rules.

Cost -- cost in dollars or dollar-equivalents (e.g., time) for the individuals to secure the services of the institution.

Education -- the degree to which the institution provides information to members of the public about their rights, obligations, and remedies to disputes.

Expense -- cost to the institution for providing the services which it provides.

Expertise -- the degree to which the institution is familiar with the types of disputes and questions generally submitted to it.

Fairness -- the extent to which the process reflects natural justice and general equity.

Flexibility -- the degree to which the institution's decisions reflect general rules of the circumstances of the particular parties involved in disputes.

Impartiality -- the extent to which decisions do not give special consideration to either landlords or tenants or to any irrelevant attributes of any disputants.

Independence/Accountability -- the degree and consistency with which the institution's behavior is directly influenced by other institutions (e.g., electorate, ministry, private associations, public opinion).

Informality -- the degree to which the dispute process does not follow a prescribed pattern, style.

Investigative Power -- the extent to which the institution takes responsibility for investigating facts.

Jurisdiction -- the degree to which jurisdictional limitations on the institution's operations are a problem for the parties involved in disputes.

Power -- the ease with which the institution can enforce the orders which it issues.

Speed -- the length of time the institution takes to provide information and process disputes.

Visibility -- the extent to which citizens are aware of the Office and its functions.

EXHIBIT 6 OF CHAPTER 3

AGGREGATE VALUE: OFFICE OF THE RENTALSMAN

- A Resolution Process
  - AA Fairness
  - AB Impartiality
  - AC Expertise
  - AD Flexibility
  - AE Consistency
  - AF Jurisdiction
  - AG Informality
- B Action Process
  - BA Power (to enforce decisions)
  - BB Investigative Power
  - BC Independence/Accountability
- C Administrative Process
  - CA Accessibility
  - CB Education
  - CC Visibility
  - CD Speed
- D Financial Process
  - DA Cost (to individual
  - DB Expense (to the institution)

Example: The Evaluation of School Desegregation Plans in Los Angeles.

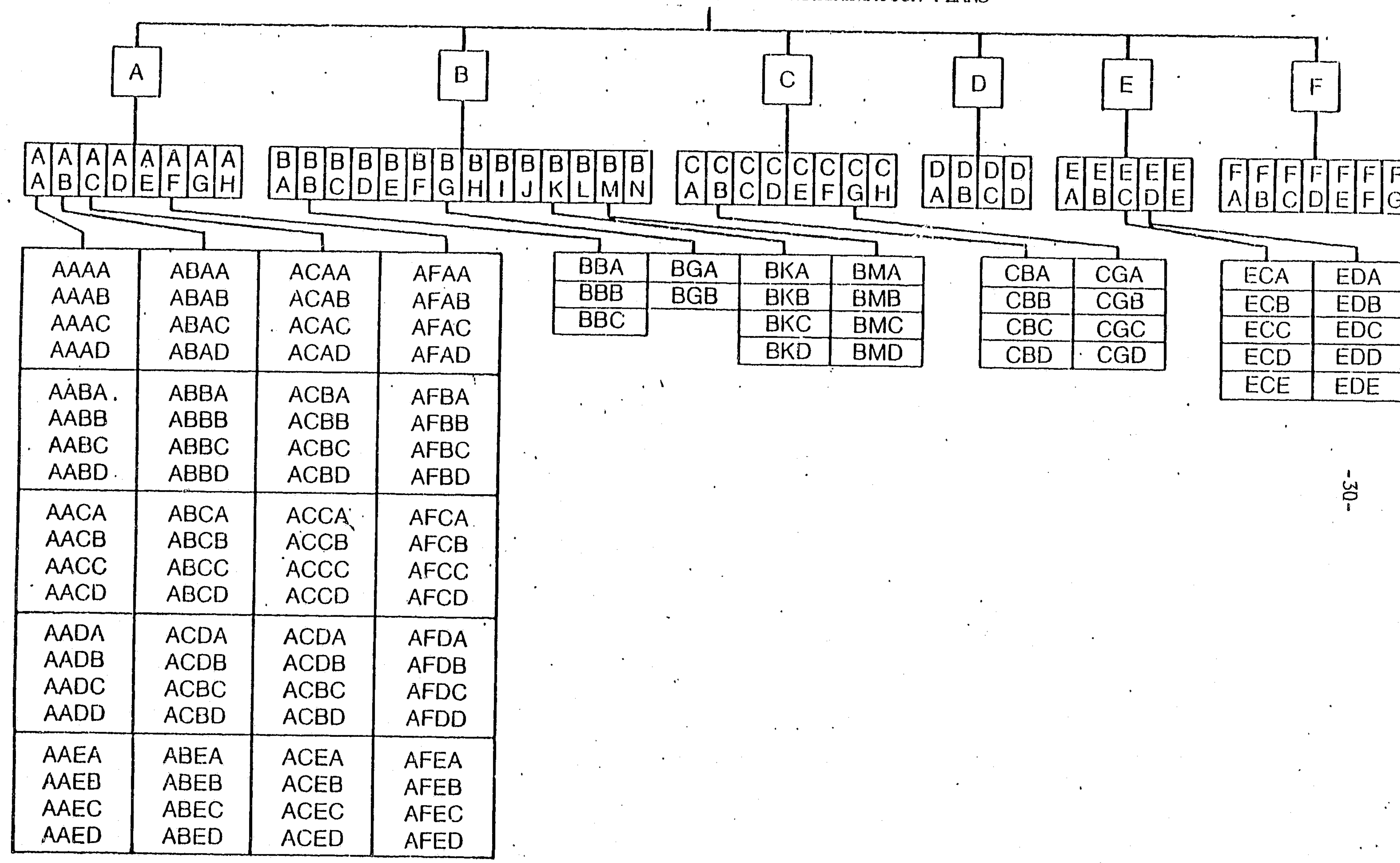
As with other school districts across the nation the Los Angeles Unified School District (LAUSD) was recently forced by a court order to desegregate its schools. Several desegregation/integregation plans had been prepared and submitted to the Los Angeles School Board. A method of evaluating these was needed and a MAUT evaluation was carried out for this purpose.

The Stakeholders and the Attributes

In this case the major stakeholders were the members of the Los Angeles School Board. Since they were elected officials, in principle, these board members represented the "ultimate stakeholders: the citizens of Los Angeles City concerned with public education." However, a preliminary listing of the attributes was done by the evaluator (Dr. Ward Edwards of USC) working closely with several members of the LAUSD staff. Several versions were prepared and presented to members of the Board of Education and to representatives of pro and con desegregation groups and other intervenors. Comments from these groups were elicited; value attributes were changed, added, and dropped (mostly added). The eighth and final version of the value tree is presented in Exhibit 7. This was such a complex tree there is no way to combine the elements and the structure in a single page display. Consequently, Exhibit 7 shows the structure only. Using the notational scheme indicated in the previous exhibits (Exhibits 3 and 6), the letters inside Exhibit 7 refer to the particular value attributes included in the tree broken down from the most general to the most specific. The major attributes are:

EXHIBIT 7 OF CHAPTER 3

AGGREGATE VALUE: LOS ANGELES SCHOOL DESEGREGATION PLANS



A. Effect of a segregation plan on racial-ethnic compositions.

B. Effect of a desegregation plan on educational quality.

C. Community acceptance of a desegregation plan.

D. Implications of a desegregation plan for District personnel.

E. Destabilizing effects of a desegregation plan.

F. Provisions within a desegregation plan for monitoring and evaluation.

Under A of Exhibit 7 we have:

AA. Racial-ethnic proportions of pupils moved from local schools.

AB. Racial-ethnic proportions in resulting schools.

AC. Racial-ethnic proportions of pupils bused. (Note: originally we expected some non-busing plans. None were submitted, so this branch was treated like AA.)

AD. Number of grades affected by reassignments.

AE. Duration in weeks of integrated educational experience.

AF. Numbers of students remaining in isolated schools.

AG. Provisions for reduction of racial-ethnic isolation in still-segregated schools.

AH. Provisions for effectively preventing the resegregation of integrated schools.

The remaining sub-attributes of A and the major attributes B, C, and so on of Exhibit 7 were detailed out in a manner similar to A and will not be presented here. The total number of twigs (bottom location of the Value Tree depicted in Exhibit 7) was 144. This is one of the largest Value Trees we have seen.

In fact, it is far too large. That seems to be a characteristic of most Value Trees built to evaluate important public programs. In part it is appropriate, since different values may be important to different stakeholders. But in part it represents the familiar process of losing perspective when passions are engaged. The larger the set of attributes, the less important each will be as a rule. Consequently, keeping the attribute list as short as the circumstances permit is an aid to perspective -- and also an aid to the intelligibility and simplicity of the analysis.

We have neither had much success ourselves nor seen much in other uses of MAUT in keeping attribute lists short for public decisions -- especially political ones. But if all major stakeholders can sit around a conference table and work in reasonable harmony on pruning excessively bushy Value Trees, the result is often very good. Nice work -- if you can accomplish it.

Example: Diversion of Status Offenders (DSO) Project.

As a result of the Juvenile Justice and Delinquency Prevention (JJDP) Act of 1974, LEAA made funds available to various state jurisdictions to help remove status offenders from detention and correctional institutions. A status offense is an act that is an offense by virtue of the age of the individual committing the act. Five common types of status offenses are runaway, ungovernable, curfew violation, truancy, and minor in possession of alcohol. In a short and somewhat slippery definition, status offenders are "kids" who get caught doing something that would not be an offense if they were "grownup." (It should be noted the exact definition of a status offender varies from jurisdiction to jurisdiction throughout the country, a fact that gives fits to evaluators.)

To achieve the deinstitutionalization of status offenders, the Act called for the development of "advanced techniques" to include: community-based programs and services for the prevention and treatment of juvenile delinquency through the development of foster care and shelter care homes, group homes, halfway houses, homemaker and home health services, and any other designated community based diagnostic, treatment, or rehabilitative service.

The JJDP Act also mandated the evaluation of all federally assisted juvenile delinquency programs awarded to provide for the evaluation of DSO programs in eight geographical regions of the country.

The JJDP evaluation plan for the DSO initiative consisted of awarding separate evaluation grants to evaluators located near each site selected for funding and awarding an overall coordination and national evaluation grant to the Social Science Research Institute, University of Southern California.

The DSO Programs.

Based on a survey of all DSO programs, the national evaluation compiled a listing of seven generic program types. While each of these program types constitutes a distinct category in the overall DSO effort, specific DSO program sites may vary in the extent to which the entire range of categories is included in the local effort. The program types are listed in Exhibit 8.

The Stakeholders and the Attributes.

The identification of the relevant value attributes was somewhat constrained in this application. The national evaluation data collection design had to be completed and implemented before individuals at the DSO sites could be identified as potential stakeholders. Therefore, the national evaluation team served as the stakeholder group and internally reviewed the data projected to be available, the goals of the DSO programs, the structure of the programs, and produced the list of value attributes given in Exhibit 9.

Later, after local stakeholders had been identified, additional attributes relevant to each local program were elicited from them, along with weights applicable to all attributes.

Unfortunately, lateness in availability of evaluation results precluded execution of the remaining steps of a MAUT evaluation.

EXHIBIT 8 OF CHAPTER 3

THE LIST OF TREATMENT PROGRAMS UNDER THE DIVERSION OF STATUS OFFENDERS (DSO) PROJECT

(a) Diversion, diagnostic and evaluation screening unit refers to a unit that 1) makes decision about clients determining which, if any, of various treatment strategies and programs the client will receive and 2) is considered a specific DSO program service that provides a referral for additional service. (DIVE)

(b) Shelter care home refers to temporary residential facilities where placement is 30 days or less. (SHEL)

(c) Group home refers to residential facilities where placement is 31 days or more. (GHOM)

(d) Foster home refers to residential placement in a single family home with the adult male and/or female serving as parent surrogate(s). (FHOM)

(e) Multiple service center refers to non-residential agencies and organizations such as the YMCA, youth service bureaus, and neighborhood drop-in centers where the focus of services is on recreation, handicrafts, character building, employment referrals, advocacy, tutoring, etc., rather than solely psychological counseling or crisis intervention. (MULT)

(f) Outreach intervention refers to short-term, intensive, non-residential, intervention which responds to situational requirements and is designed to effect change in a variety of the client's physical, social, and emotional circumstances. (OUTR)

(g) Counseling only refers to a non-residential program where the sole or primary service is individual or group psychological counseling or therapy, including work with the DSO client's family. (COUN)

EXHIBIT 9 OF CHAPTER 3

THE LIST OF VALUE ATTRIBUTES FOR THE DIVERSION OF STATUS OFFENDERS (DSO) PROJECT

1. The average number of arrests per status offender per year. (NARR)
2. The proportion of status offense arrests that result in court appearance. (COUR)
3. The average number of serious delinquent offenses (such as robbery) that might occur in a six month period, regardless of whether these offenses come to the attention of justice authorities. (DLNQ)
4. The amount of time status offenders spend with family. (FAML)
5. Status offenders' perceptions about the seriousness of an offense such as robbery. (PSER)
6. Status offenders' perceptions of justice system effectiveness. (PJUS)
7. Parental knowledge of whereabouts of status offender. (WHER)
8. Status offenders' attitudes toward observing the law. (ALAW)
9. The average number of minor delinquent offenses (such as truancy) that might occur in a six-month period, regardless of whether these offenses come to the attention of justice authorities. (MDLN)
10. The annual cost in dollars per individual status offender served. (COST)
11. Status offenders' perceptions about the seriousness of an offense such as truancy. (PTRU)
12. Frequency of contact between programs set up to provide services, agencies, or institutions. (FREQ)
13. The percentage of arrested status offenders placed in locked facilities while waiting for a court appearance. (LOCK)
14. Level of the status offender's school performance relative to his school mates. (SCHO)

Comment

This is the longest chapter in this manual, a necessity due to the complexity and importance of the topic. MAUT evaluation begins with the identification of the stakeholders and the listing of the value attributes. We reiterate: if possible, keep the number of attributes small. Our technical experience indicates that 8 would be about right and 15 would already be excessive. However, as some of the examples in this chapter indicate, keeping the number of attributes small is difficult.

The major technical problem that arises with a large number of attributes is that the importance weights to be assigned to the attributes will often get very small and thus blunt the meaningfulness of the weights.

The next chapter discusses the problem of weighting.

CHAPTER 4 SUMMARY

Chapter 4 discusses the problem of weighting the value attributes so that they are arranged in numerical order of importance. Several different techniques for weighting are described in detail and examples of each method are given. Emphasis is on simplicity in weight elicitation from the stakeholders. Special forms that have been used to successfully elicit weights are provided and a way to check on the consistency of weight judgments is given. How to use the value tree as a technique to simplify the weighting process is explained and illustrated by a concrete example. The chapter concludes with a discussion of the problems of assignment of weights by multiple stakeholders.

If possible, it is always desirable to arrive at a consensus on what the weights should be. If the stakeholders can be brought together in a group, they can often arrive at such a consensus. The presentation of the value tree, if one has been constructed, to the group is helpful in aiding this process. But it is not always possible to work with groups. When stakeholders represent different groups with different ideas about what attributes are most important, it is often possible for each group to specify its own weights but no attempt is made to amalgamate the separate group weightings. Replacing individual weights with a single average is not recommended.

CONTINUED

1 OF 3

CHAPTER 4

WEIGHTING VALUE ATTRIBUTES

Why Weight?

Not all attributes are likely to be considered equally important. The function of weights is to express the importance of each attribute relative to all others. The weighting procedures we describe shortly vary in difficulty and in precisio But some form of weighting is usually essential. Weights capture the essence of value judgments. They can be expected to vary from stakeholder to stakeholder; indeed, stakeholders usually contribute only two kinds of judgments to MAUT analyses: attributes and weights. In most evaluations, incidentally, the fact that multiple stakeholders are involved means that you cannot hope to elicit attributes and weights in the same session. Your first elicitation session will be concerned with attributes; after you have elicited all attributes and combined them into (we hope) one value tree, you will need to visit each stakeholder representative again in order to elicit weights on that tree.

Weights should, of course, reflect the purpose of the evaluation. The weights on administrative smoothness and efficiency, for example, might well be higher for an evaluation intended for monitoring or for programmatic fine-tuning than for a full-scale impact evaluation -- though that attribute would be relevant to both. Although the generalization is too simple, it is appealing to think of just two kinds of weights; one relevant to program management decisions, and the other relevant to impact assessment, current or prospective.

Values are reflected in weights, and values change over time. So weights should be re-elicited in situations in which a program is periodically re-evaluated. Since the program is designed around the old weights, it is relevant to evaluate it against both the old weights and the new ones.

In the kind of complex multi-level evaluation of the CAC example in Chapter 3, it would be inappropriate to elicit all weights from one set of respondents.

Different respondents function at different levels of the program. Legislators, for example, are interested in broad programmatic goals, at the top of the Value Tree. Local project administrators, such as the Director of the Midwood-Kings Development Corporation, are naturally concerned with the project-specific values relevant to their own projects. The question of who should assess which weights is a matter of evaluative judgment. The principle is obvious: each stakeholder should judge weights in the level or levels of the tree in which he or she has knowledge, expertise, or interest. Translating that general principle into specific decisions can be sensitive, and is not subject to specific rules that we know of.

The remainder of this chapter discusses various ways of eliciting weights. While we clearly prefer one way, ratio weighting, to its alternatives, we also know that it is a nuisance, and that simple alternatives to it often give essentially equivalent results. That is why we offer simpler alternatives. Those familiar with the very large literature of weighting will recognize that many different weighting procedures, even including equal weights, can often lead to equivalent aggregate utilities, or at least to the same ordering of options. If aggregate utilities are the goal of the analysis, easy procedures may work well, and demand much less effort of respondents. But for such purposes as monitoring and fine-tuning, weights at lower levels of the Value Tree are much more important than for obtaining fully aggregated utilities. Moreover, the weights themselves may be useful information to those concerned with project or program management, since they indicate what stakeholders are most concerned about in a quantitative way.

Equal or Unit Weighting

The easiest weighting scheme is to assign equal or unit weights to each of the attributes, in other words, treat all attributes as equally important. While we do not recommend this scheme, we mention it for the following reasons:

(a) it eliminates the problem of deciding what the weights should be and also eliminates the difficult task of obtaining the importance weight judgments from the stakeholders; (b) if wildly differing weights are obtained from conflicting stakeholders, then assigning equal weights is one way of resolving the disagreement; (c) a MAUT analysis always includes a sensitivity analysis, that is, a study is done to see how the final result is affected by changes in the weights, the location measures, the number of attributes used and so on. One thing to try in a sensitivity analysis is equal weights. Chapter 7 shows how such sensitivity analyses are conducted.

Rank Weights and Rank Reciprocal Weights

The simplest way of assessing differential weights is to arrange the attributes in simple rank order, listing the most important attribute first, the least important attribute last, and the other attributes arranged from high to low between these two extremes. A numerical weight is then assigned to each attribute according to its rank in the list. The two most common ways to assign the numerical weights for the ranked attributes are: (a) Assign the largest rank number to the most important attribute, the next highest number to the second most important attribute and so on down the list until the least important attribute receives the rank of one (1) (such numbers are called inverse ranks). Then add these numbers and divide each by the sum. The procedure of dividing each number by the sum of the numbers is called normalizing. It assures the normalized numbers sum to 1.[1] This final result is called rank weighting. (b) Assign the numerical value of (1) to the most important attribute, two (2) to the next most important attribute, and so on; the least important attribute receives a rank of N where N is the number of attributes. The reciprocal or one (1) divided by each of the numbers so assigned is then taken, and these reciprocals are normalized. This assures that the most important attribute receives the highest numerical weight and the least important attribute receives the lowest weight. This

[1] Appendix A lists a hand held computer program to do normalizing.

is called rank reciprocal weighting. It requires a little more arithmetic than rank weights, but is easy to do.

In any method based on ranks, you must consider the possibility of tied ranks. Suppose that the ranks that would originally have been assigned the numbers 3, 4, and 5 are tied. Then all three of them receive the number 4. That number is used in rank weighting, and its reciprocal is used in rank reciprocal weighting. If 3 and 4 had been tied, each would have received a rank of 3.5.

These two weighting schemes can be illustrated with a numerical example. Refer back to Exhibit 3 of Chapter 2. This lists the twigs and gives their weights for the drug counseling center siting example. We can rearrange them in order of decreasing original weight, assign them ranks, and calculate rank sum and rank reciprocal weights for them. Exhibit 1 does so.

Obviously, any rank weighting method is at best an approximation. Inspection of Exhibit 1 shows that the rank sum weights are far flatter than the rank reciprocal weights. Since the original weights were quite flat themselves, the rank sum procedure produces an excellent approximation to them, while the rank reciprocal procedure does not. Had the original weights been less flat (as they typically are), the rank reciprocal procedure would have produced the better approximation. Unfortunately, if you already know the appropriate weights, there is no point in using an approximation, while if you do not, you cannot be sure which to use.

EXHIBIT 1 OF CHAPTER 4

AN ILLUSTRATION OF RANK WEIGHTING TECHNIQUES

Twig Label	Original Weight	Inverse Rank	Rank Weight	Normal Rank	Reciprocal of Normal Rank	Rank Reciprocal Weight
AA	.168	12	.154	1	1	.326
BA	.120	10.5	.135	2.5	.400	.131
BB	.120	10.5	.135	2.5	.400	.131
CA	.099	9	.115	4	.250	.082
AB	.090	7.5	.096	5.5	.182	.059
DA	.090	7.5	.096	5.5	.182	.059
CB	.061	6	.077	7	.143	.047
AC	.060	4.5	.058	8.5	.118	.039
AD	.060	4.5	.058	8.5	.118	.039
AE	.052	3	.039	10	.100	.033
DB	.050	2	.026	11	.091	.030
CC	.030	1	.013	12	.083	.027
Sums	1.000	78	1.002		3.067	1.003

If the stakeholder has some feeling for whether the proper weights are relatively flat or relatively steep, one might simply choose between these approximations on the basis of that feeling, and accept the results of the approximation. In any case, the stakeholder will have to arrange the things to be weighted in rank order, and to make judgments about ties.

Ratio Weighting

This method begins as with the previous methods, i.e., the attributes are first placed in rank order of importance such that the most important attribute is at the top of the list and the least important is at the bottom of the list. The least important attribute is then assigned a value of ten (10). The stakeholder then assigns numerical weights such that the next in the list (from the bottom) gets a value depicting how much more important that attribute is relative to the least important attribute. Thus if a value of 20 gets assigned that means that attribute is twice as important as the least important which received a value of 10. The stakeholder works up the list of attributes assigning numerical values in a similar fashion. Thus if some other attribute receives a value of 40, it is considered four times as important as the least important attribute which received a value of 10, and twice as important as that attribute which received a value of 20. The stakeholder should be carefully instructed about what the weights mean using this method. Ties are permitted, i.e., if the expert thinks two or more attributes are equal in importance they would receive the same numerical weight. Since this method is more demanding than the rank methods, it is often a good idea to give the stakeholder an example of how the method works before he or she proceeds to make judgments. This is particularly important if the weights are being elicited from the stakeholder via mail questionnaires. The example given in Exhibit 2 is quite useful. Again, we use the siting example, this time confining our attention to its four top-level values. In order of judged importance, they are: (A) Good conditions for staff, (B) Easy access for clients, (C) Suitability of space for the Center's functions, and (D) Administrative convenience.

Exhibit 2 assigns a reference weight of 10 to the least important attribute, administrative convenience (D). The other attributes are then judged relative to that one. The numbers entered in Exhibit 2 are as they would be elicited from individual stakeholders. Thus we will have a set of ratio weights for each stakeholder. Left to their own devices, people tend in this procedure to make judgments that end in 5 or 0, which does little harm, though people should be encouraged to think about their judgments and to make as careful discriminations as their feelings permit.

Consistency Check: Use of the Triangular Table

If possible, the evaluator should work with individual stakeholders, or groups of them if they are making group judgments, when the ratio method is used. If so, the triangular table included in Exhibit 3 is useful in encouraging consistency. The stakeholder first makes the judgments in the first column. If D has a weight of 10, what weight should C have so that

EXHIBIT 2 OF CHAPTER 4

EXAMPLE: EVALUATING SITES

Step 1. Review list of value attributes. (An attribute = a "thing to consider")

Step 2. Rank order the value attributes to reflect their relative importance to you as you evaluate location sites. Ties are acceptable. Enter the letter corresponding to the most important attribute listed in Step 1 on line 1. Enter the second most important on line 2, and so on. If any two attributes are equally important, place both letters on the same line. For three-way ties place three letters on the same line, and so on.

Step 3. Weight the value attributes. Assign 10 points to the least important attribute and then indicate your own opinion about the relative importance of each attribute by assigning weights accordingly. (No upper limit on weights)

Value Attributes	Rank Order			Assign Weights
A. Good conditions for staff	Line 1	A	(Most important)	40
B. Easy access for clients	Line 2	B		30
C. Suitability of space	Line 3	C		20
D. Administrative convenience	Line 4	D	(Least important)	10

Suitability of Space (C) is twice as important as Administrative Convenience (D)

10 points assigned as reference to least important attributes

For this person Good Conditions (A) is twice as important as Suitability of Space (C) and four times as important as Administrative Convenience (D).

EXHIBIT 3 OF CHAPTER 4

THE TRIANGULAR TABLE FOR RATIO WEIGHTING: CHECK FOR CONSISTENCY

Ratio Method Weighting for the Top Level Values in the Siting Example

Attribute Label	1 Ratios to 4	2 Ratios to 3	3 Ratios to 2	4 Weights[1]
A	31	22	18	.43
B	17	12	10	.24
C	14	10		.19
D	10			.14
Sums	72			1.00

[1]These are the normalized weights of the values given in Column 1.

Name ______________________________
(Please Print)

This is the second phase of this process. Part II asks for your views about the relative importance of 16 attributes considered important to the operation and administration of the Rentalsman's office as it is operating in British Columbia.

On Page 3, you will find the 16 attributes listed. We would like you to review this list and then indicate your views as to the relative importance of each attribute on the list as follows:

1. Please consider the 16 attributes (and only these 16) and then RANK ORDER them in decreasing order of importance to you, with number 1 being most importance, and the least important last. (Ties are acceptable.)

2. Once you have rank-ordered them, please reflect on their relative importance to you. How much weight does each attribute carry relative to the other attributes as you would use them to appraise an office such as the Office of the Rentalsman.

Please write the weights you would assign to each attribute to reflect its relative importance to you.

a. Do this by assigning a weight of 10 to the least important attribute (lowest rank) as a common starting point.

b. Next, for the attribute with the next highest rank, assign it a weight to reflect its importance compared to the lowest attribute. For example, it may be half again as important to you as the lowest attribute. If so, it would receive a weight of 15. If it is twice as important, it would receive a weight of 20.

c. Then go to the next most important attribute and compare it to the one just completed and repeat the process.

An attribute with a weight of 40 is twice as important as one with 20 and half as important as one with 80, and so on. An attribute with 50 is as important as one with 20 and one with 30 taken together.

There are no limits to the weight you assign. When finished, you will have weighted all the attributes to reflect their relative importance to you. Page 2 shows a simplified example of how this is done.[1]

Remember, we are interested in your personal preferences, so there are no "right" or "wrong" answers. The definitions of each attributes are given for guidance only. You are free to redefine them in any way you wish.

The example is the one similar to that given in Exhibit 2.

EXHIBIT 5 OF CHAPTER 4

ILLUSTRATION OF THE FORM USED TO ELICIT IMPORTANCE WEIGHTS VIA THE RATIO METHOD

Appraising Programs to Handle Landlord-Tenant Relations

Step 1. Review the following value attributes in terms of their importance in appraising a Rentalsman Office. You may provide your own definitions if you wish.

Step 2. Rank order the value attributes to reflect their relative importance to you as you appraise the Rentalsman's Office. Enter the letter corresponding to the most important attribute listed in Step 1 on Line 1. Enter the second most important on Line 2 and so on. If any two attributes are equally important, place both letters on the same line. For three-way ties, place three letters on the same line, and so on.

Step 3. Weight the value attributes. Assign 10 points to the least important attribute and then indicate your own opinion about the relative importance of each attribute by assigning weights accordingly. (No upper limit on weights.)

ATTRIBUTE Definition

A. SPEED -- The length of time the institution takes to provide information and process disputes.
B. POWER -- The ease with which the institution can enforce the orders which it issues.
C. VISIBILITY -- The extent to which citizens are aware of the Office and its function.
D. FAIRNESS -- The extent to which the process reflects natural justice and general equity.
E. ACCESSIBILITY -- Ease of registering requests and complaints; taking into consideration procedural complexity, hr./days available, and physical location.
F. EDUCATION -- The degree to which the institution provides information to members of the public about their rights, obligations, and remedies to disputes.
G. INVESTIGATIVE POWER -- The extent to which the institution takes responbility for investigating.
H. INFORMALITY -- The degree to which the dispute process does not follow a prescribed pattern, style.
I. INDEPENDENCE/ACCOUNTABILITY -- The degree and consistency with which the institution's behavior is directly influenced by other institution's (e.g., electorate, ministry, private associations, public opinion).
J. FLEXIBILITY -- The degree to which the institution's decisions reflect the circumstances of the particular parties involved in disputes.
K. IMPARTIALITY -- The extent to which decisions do not give special consideration to either landlords or tenants or to any irrelevant attributes of any disputants.
L. EXPERTISE -- The degree to which the institution is familiar with the types of disputes and questions generally submitted to it.
M. JURISDICTION -- The degree to which jurisdictional limitations on the institution's operations are a problem for the parties involved in disputes.
N. COST -- Cost in dollars or dollar-equivalents (e.g., time) for the individuals to secure the services of the institution.
O. EXPENSE -- Cost to the institution for providing the services which it provides.
P. CONSISTENCY -- The degree to which the institution's decisions reflect general rules.

RANK ORDER		ASSIGN WEIGHTS
Line 1	K	170
Line 2	D	160
Line 3	I	120
Line 4	P	110
Line 5	A,E	100
Line 6	G	75
Line 7	N	60
Line 8	L	55
Line 9	J,H	50
Line 10	B	40
Line 11	C	25
Line 12	M	20
Line 13	O,E	10
Line 14		
Line 15		
Line 16		

the ratio of C to D seems appropriate? The answer 15, for example, would mean that C is 1½ times as important as D; the answer 20 would mean that C is twice as important as D, and so on. After that judgment has been made for C, the next judgment is the ratio of B to D. After that, A to D. Then the stakeholder moves over to the next column, which ignores D and assigns a weight of 10 to C, and makes the ratio judgments of A and B to C. Finally, the stakeholder does the same for the third column. The final column is calculated by normalizing the numbers in column 1.

All entries in a table like that of Exhibit 3 should be consistent. If, for example, a stakeholder has made the indicated judgments in column 1, and then in column 2 judged the ratio of B to C to be 20 to 10 (2 to 1), the evaluator would point out the inconsistency between that judgment and the numbers in column 1, and invite the stakeholder to revise either or both judgments to ensure consistency. Only after all the judgments have been made consistently should the final weights be calculated. It is preferable to do that calculation while the stakeholder is still present, so that he or she can consider the weights that result from his or her judgments, decide whether or not they seem appropriate, and if not, go back and revise the ratios. When we use this method, we also quickly calculate normalized weights; the respondent can assess them for consistency at the same time he or she is making revisions for consistency. For rapid normalizing, the program in Appendix A is especially useful.

If the number of values to be compared with one another exceeds 6 or 7, the number of ratio judgments required to complete a full triangular table can get tediously large. (6 values requires 10 judgments, 7 requires 15, 8 requires 21, and so on.) In that case, it may be appropriate to reduce the amount of judgmental labor by using only the first two columns. But at least one column other than the first should ordinarily be filled out, to provide for at least some consistency checking.

If two values are originally judged to be tied, then of course, there is no point in judging ratios in both columns in which they receive 10's; either one will do.

The other function of the triangular table of judgments is to take care of the case in which the least important attribute is considered by the stakeholder to be utterly unimportant, deserving a weight of 0. In that case, the initial set of judgments should be made in the first column in which a non-zero value receives a weight of 10, and all values below that should get weights of 0.

A More Realistic Example

We now give a more realistic example of the ratio method of assigning importance weights, one selected from an actual evaluation study -- the evaluation of a program for resolving landlord-tenant disputes, the Office of the Rentalsman (OR) example alluded to in the previous chapter. In that particular study, 16 attributes were identified. The instruments used to elicit importance weights from the stakeholders are given in Exhibits 4 and 5. Exhibit 4 gives the instructions and Exhibit 5 is the sheet on which the stakeholder gave his/her weights. As in the previous example, the attributes are listed along the side with their respective definitions. The numbers written in under the assigned weights column of Exhibit 5 are the ones actually obtained from one of the participants in this particular evaluation study.

Thus, this particular stakeholder thought that Attribute K, IMPARTIALITY, was the most important attribute and it received an importance weight of 170 which is 17 times as important as the least important attributes of EXPENSE (O) and ACCESSIBILITY (E). The second most important attribute is that of FAIRNESS (D) which received a weight of 160 and so on.

EXHIBIT 6 OF CHAPTER 4[1]

AN EXAMPLE OF RATIO WEIGHTING USING THE VALUE TREE

(.47) A Resolution Process	(.17) B Action Process	(.30) C Administrative Process	(.06) D Financial Process
(.24) AA Fairness (.11)	(.38) BA Power (.07) (to enforce decisions)	(.29) CA Accessibility (.09)	(.63) DA Cost (.04) (to the individual)
(.24) AB Impartiality (.11)	(.31) BB Investigative Power (.05)	(.25) CB Education (.07)	(.37) DB Expense (.02) (to the institution)
(.13) AC Expertise (.06)	(.31) BC Independence/ Accountability (.05	(.23) CC Visibility (.07)	
(.13) AD Flexibility (.06)		(.23) CD Speed (.07)	
(.13) AE Consistency (.06)			
(.11) AF Jurisdiction (.05)			
(.09) AG Information (.04)			

1 The value attributes were those elicited by the expert stakeholders concerned with the evaluation of the Office of the Rentalsman, Vancouver, British Columbia.

An Aid to Ratio Weighting: Use of the Value Tree

If there are many twigs then ratio weighting can be quite time consuming and demanding of the stakeholders. One technique to help the stakeholder is to have him (her) use the Value Tree of the attributes if such a tree has been constructed. Have each stakeholder first judge ratio weights for the main branches of the tree and then under each branch obtain the ratio weights for the sub-attributes, making separate sets of judgments for each lower-level group of values under an upper-level value. The result was illustrated in Exhibit 1 of Chapter 2. When these weights are normalized then the weight for each twig of the tree is easily obtained by multiplication down through the value tree. As an illustration, consider Exhibit 6, which lists the attributes for the value tree of Exhibit 6 of Chapter 3. There are four major attributes: Resolution Process (A), Action Process (B), Administrative Process (C), and Financial Process (D). The experts can be asked to make ratio weight judgments for these four attributes, resulting (as an example) in normalized weights of (.47)A; (.17)B; (.30)C; and (.06)D respectively. Then under each of these the expert can also make the ratio weight judgments, resulting (as an example) under B, in the normalized weights of (.38)BA; (.31)BB; and (.31)BC. To obtain the final weight for each attribute at each twig multiply these two numbers. Thus AA FAIRNESS receives a final weight of .11 (.47 X .24 = .11). As mentioned previously (Chapter 2), **this** is called multiplying through the tree.

Comment

We recommend that if at all possible the ratio weight method be used. Since this requires the stakeholders to rank order the attributes

in order of importance, this will yield 3 possible weighting schemes (rank weights, rank reciprocals, and ratio weights).

Assignment of Weights by Multiple Stakeholders

The use of the MAUT model, which often represents the viewpoints of multiple stakeholders, requires the assignment of weights to the value attributes reflecting each attribute's importance. Using the methods described in this chapter, we arrive at a set of weights for each stakeholder. Should these individual weights be averaged? The answer is yes but we do not recommend replacing the individual weights with the average. Use the average as another set of weights and carry through the analysis to be explained in Chapter 6 using each stakeholder's weights as well as the average weight. The question of interest is whether the average leads to aggregate utilities substantially different in rank order from those of the individual stakeholders. Often, the answer will be "no". Of course it is always desirable, if possible, to arrive at a consensus on what the weights should be. If the stakeholders can be brought together in a group they can often arrive at such a consensus. The presentation of the value tree to the group is often useful in aiding this process. When the stakeholders represent different groups with different ideas about what attributes are most important, it is often possible for each group to specify its own weights but no attempt is made to amalgamate the separate group weightings.

Another approach is for groups essentially to negotiate among themselves to arrive at an "agreed-to" set of weights. For instance, groups may jointly decide that each of them should receive equal weight in an overall sense, so that the weights used in the MAUT model are averages of the weights assigned by each separate group. Such weights normally will not reflect the values of any one individual. Rather, models using such weights are essentially models of policy, jointly formulated by multiple groups or individuals. Thus, just as policies are the result of group consensus, MAUT models which reflect policy also result from group consensus -- in this case, regarding the appropriate weights to use.

CHAPTER 5 SUMMARY

Chapter 5 discusses and illustrates how location measures for each attribute are determined for each option or entity to be evaluated. A location measure, also called single attribute utility, is an assessment of how desirable an option is with respect to each of the value attributes. Simple graphical methods for assigning location measures on a common scale are illustrated. The simple equations giving the same answers as the graphical solution are also provided. All possible cases that are apt to arrive in practical evaluations are demonstrated. A method for eliciting location measures via the mail is given with an example of the type of form that might be used. A somewhat technical but important discussion on the relation between location measures and weights is described. A weight can be interpreted as an exchange rate among location measures. The reason for this is that once a weight is assigned to an attribute it is always for a given range of location measure for that attribute. If the range of such measures should change, then the weights should change accordingly.

The chapter concludes with an example of how to make comparisons among programs when a new program has replaced an older one and the evaluation examines whether the new program is better than the old program.
A technique for assigning location measures for the current and prior programs is illustrated.

CHAPTER 5

THE LOCATION MEASURES

This chapter discusses and illustrates how location measures for each attribute are determined for each of the option or entities to be evaluated.

First, a few reminders about ideas from previous chapters. A location measure is an assessment of how desirable an option is with respect to a particular twig, or bottom node of a Value Tree. This is expressed as a number technically called an utility. Throughout this chapter, we use the words "location measure" and "utility" interchangeably, though our procedures differ from classical ones for utility measurement.

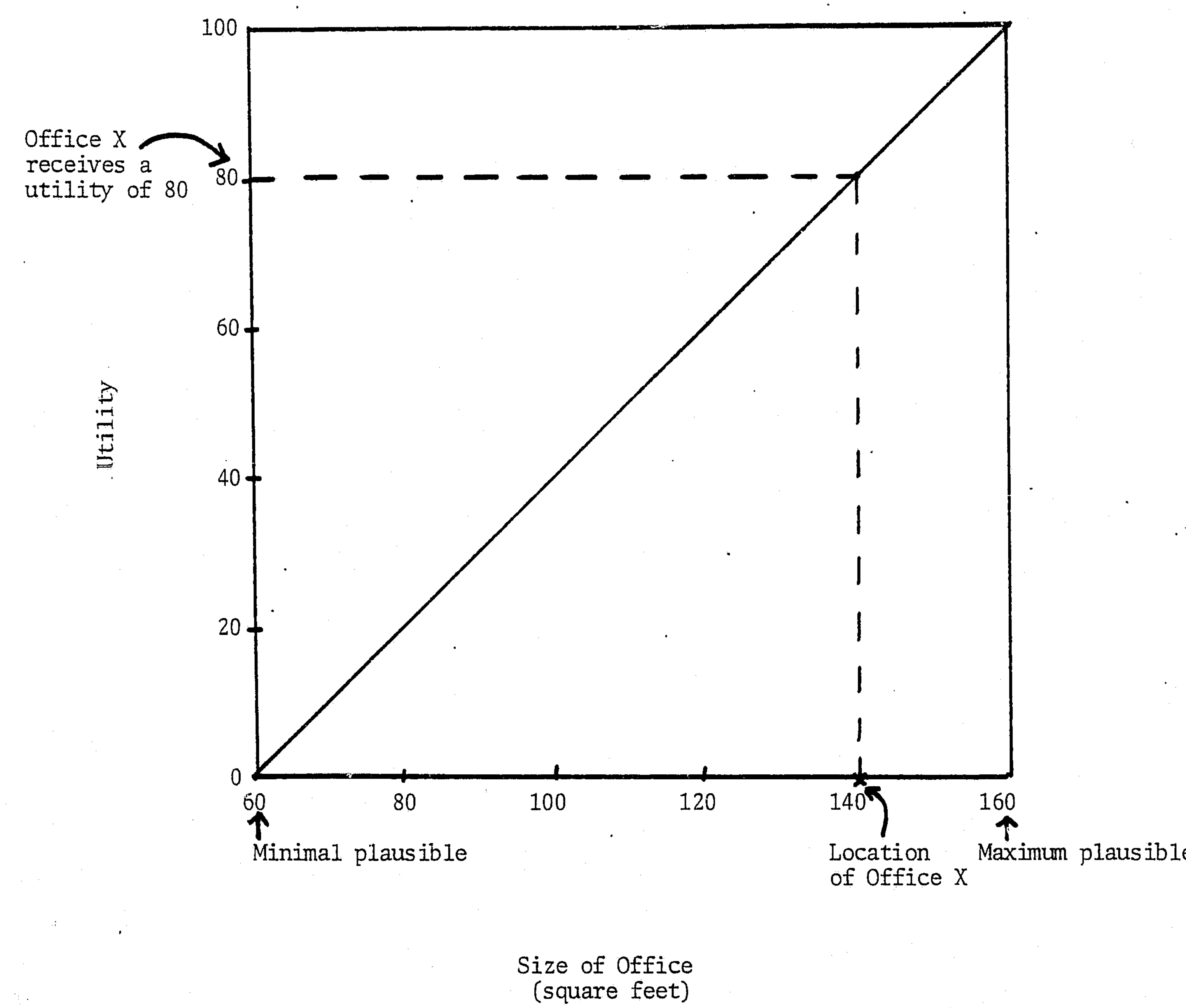
Such numbers, since they are assessments of desirability on single attributes of evaluation, are in principle subjective -- but in fact may often be simple arithmetical transformations on objective measures. We distinguish between the two cases. The first arises when some objective measure captures what you consider the attribute to be concerned with; in that case, your task is to transform the number so that it is comparable in meaning to other numbers. The second arises when the attribute is inherently judgmental, as office attractiveness was in the Drug Free Center siting example. Such judgments, sometimes made by program people or others close to the program and sometimes by independent and presumably impartial experts, ordinarily need no transformations, since they are already on a scale that makes them comparable to all other measures of desirability.

Linear Location Measures

We consider the first kind of instance first. In the Drug Free Center siting example, one attribute was office size. The natural unit in which this attribute is measured is square feet. From Chapter 2, you may recall that the Director assigned a utility of 0 to an office 60 square feet in size, and a utility 100 to an office 160 square feet in size. Since more size is preferable to less,

EXHIBIT 1 OF CHAPTER 5

EXAMPLE OF SIMPLE LINEAR GRAPH RELATING SUBJECTIVE VALUE (UTILITY) TO PHYSICAL ATTRIBUTE LOCATION (OFFICE SIZE IN SQUARE FEET)



and we are confining ourselves to linear functions relating desirability to physical measures whenever, as in this case, desirability either continuously increases or continuously decreases with the physical measure over the whole range, it follows that a 140 square feet office would have a location measure of 80.

The simple graphical representation of this is given in Exhibit 1. The horizontal (X) axis of Exhibit 1 is the range of the attribute in its "natural units", going from the lowest plausible value to the highest plausible value. The vertical (Y) axis goes from 0 to 100. To assign utility simply locate the option on the X axis and "read off" its utility on the Y axis.

If the simple linear relationship between utility and the attribute's natural units as indicated in Exhibit 1 is acceptable, then there is even an easier way to assign utilities. The calculation is simple: for an office size of 140 sq. ft. it is

$$80 = 100\ (140 - 60)/(160 - 60)$$

More generally, if L_A is the actual location measure, L_{min} is the minimum value, and L_{max} is the maximum value, the calculation is:

$$\text{Location of } L_A = 100\ (L_A - L_{min})/(L_{max} - L_{min}). \qquad (1)$$

Although Equation 1 is trivially simple to use, we have provided in Appendix B a simple calculator program that asks for L_{min}, L_{max}, and L_A and then does the arithmetic for you.

Consider another example. In the DSO evaluation reviewed in Chapter 3, the objects of evaluation were programs intended to reduce juvenile crime. One of the attributes might be "A. Average Number of Arrests Per Status Offender Per Year". The natural unit for this attribute is obvious. Its minimum is obviously 0, and that has a utility of 100. Choice of a number to assign utility 0 to is a judgment; it might be 10 or more. In that example, less

is preferable to more, so the equation appropriate to the problem is no longer Equation 1. Instead it is

$$\text{Location of } L_A = 100\ (L_{max} - L_A)/(L_{max} - L_{min}). \qquad (2)$$

To be sure you are with us, use Equation 2 to verify that an average of 7 arrests per status offender per year should get a location measure of 30. The program in Appendix B will do Equation 2 for you also.

Equation 1, then, is appropriate if more is better than less, and Equation 2 is appropriate if less is better than more.

Ranges and Outside-Range Locations

In these examples, some judgmental inputs were necessary because the limits defining utilities of 0 and 100 were judgments. In some cases, no such judgments may be needed. For example, the attribute "B. Percentage of Status Offense Arrests that Result in Court Appearances" seems to have a natural range from 0% to 100%. But such natural ranges may be deceptive. Recall from Chapter 2 that we emphasized the importance of making the boundaries realistic. Is it realistic to expect anything like 100% of status offense arrests to lead to court appearances? If not, you should assign a more realistic upper bound.

In Chapter 2 we discussed the nature of such upper and lower bounds. We said then and reiterate now, that they should be minimum and maximum plausible values, rather than minimum and maximum possible, conceivable, or actual values. The fact that we choose a range not directly controlled by the actual locations we are using (if in fact we know them) means that every now and then an instance will fall outside the range. We gave an example in Chapter 2, in which the Center Director assigned a location measure of 100 to having the bus stop no more than half a block from the Center, and 0 to having it four blocks away. We said in Chapter 2 that on that scale a site with a bus stop 5 blocks away would score -29 in utility. A simple substitution into Equation 2 will now permit you to calculate this number for yourself. The natural unit of measurement is half-blocks; the minimum is 1, and the maximum is 8. So the calculation is:

$$100\ (8 - 10)/(8 - 1) = -100(2/7) = -28.57.$$

While such outside-the-range objects of evaluation can occur for both objectively measureable and judgmental twigs, they should present no special problems so long as the meaning of the ranges is kept clearly in mind. Remember that an alternative approach to outside-the-range cases is to treat them as though they fell at the range boundary; whether to do this or instead to use a number less than 0 or greater than 100 is a judgmental question, and depends on whether you consider the difference between the boundary value and the value observed to make any meaningful difference to the attractiveness of the option.

Judgmental Location Measures

Purely judgmental location measures present no arithmetical problems, since nothing like Equation 1 or 2 need be used. Instead, however, they may present problems because judges may be reluctant to approach the extremes closely, especially the lower extreme. Judges of location measures should keep two things in mind: first, that the location measures serve to differentiate one object of evaluation from another, and so should be well spread out, and second, that assignments of both 0 and 100 should be realistic with respect to the evaluation in hand, not with respect to the dimension in general. Judgmental assessments of self-discipline appropriate to selecting candidates for West Point, for example, would almost certainly be quite inappropriate for selecting candidates for release from a juvenile detention center. The attribute of evaluation (self-discipline) might be the same, but the ranges of it that one would expect to encounter are entirely different for the two examples.

Although in the siting example we had the Director serving as judge of the location measures, that is not generally good practice. Evaluations gain in objectivity and credibility if judgmental assessments are made by experts on the topic of the assessment, preferably experts not too closely tied to the program being evaluated. If different twigs call for differing kinds of expertise, use of more than one outside expert is usually wise. Indeed, it is exactly in assessing judgmental location measures that we consider independent expertise to have its most important role. If those closely associated with a program disagree with externally assessed location measures, they can make independent assessments of their own and then report the consequences of using them instead of the outside ones as an ingredient of the sensitivity analysis.

Bilinear Location Measures

Some location measures do grow out of measurements or counts, but do not have the convenient property either that more is better than less or less is better than more. Sometimes, an intermediate value will be "just right", and deviations from it in either direction will be less attractive. The standard example is the amount of sugar in a cup of coffee. If you like 2 spoonsful, you will find 0 too few and 4 too many. But, in the example, the function is not symmetric for most of us; most people who like sugar in their coffee prefer too much to too little.

We find it convenient to approximate such utility functions by two lines rather than one. One of those two lines will run from 0 at either the minimum or maximum value to 100 at the optimal value. But the other one ordinarily will not descend from the optimal value all the way to zero again. A judge considering a sentencing decision might provide an example. The legally specified boundaries for the decision might be 1 to 5 years. The judge might, for a defendant, feel that 4 years is just right, and that 1 year is so inadequate as to deserve a utility value of 0. So the utility value of any sentence between 1 and 4 years would be given by Equation 1, with 1 as L_{min} and 4 as L_{max}. But the judge might well feel that a 5-year sentence deserves a utility value of, say, 60. In that case, how would the judge assess a 4½ year sentence? Intuition suggests 80, and calculation can confirm it. The appropriate equation is an adaption of Equation 2, with L_{max} being the location measure associated with the maximum sentence, in this case 5 years. The equation is:

$$\text{Location of } L_A = U_{max} + (100 - U_{max})(L_{max} - L_A)/(L_{max} - L_{min}). \quad (3)$$

Note that in this example L_{min} is the location of the peak; in this case 4 years. What Equation 3 does, of course, is to partition the location measure into two parts, U_{max} (60 in this example), and the difference between U_{max} and 100. It automatically awards U_{max}, and then increases it by the proportion of the remainder that corresponds to the proportionate distance between L_{max} and L_A. To make sure that the symbols are all clear, we work the example problem:

$$\text{Location of } L_A = 60 + (100 - 60)(5 - 4.5)/(5 - 4) = 80.$$

It could also happen that the lower, rather than the upper, branch of the bilinear utility function did not hit 0 utility. In that case, the Equation corresponding the Equation 3 would be

$$\text{Location of } L_A = U_{min} + (100 - U_{min})(L_A - L_{min})/(L_{max} - L_{min}). \quad (4)$$

Note that in Equation 4 L_{max} is now the location of the peak.

Since both Equations 3 and 4 are somewhat trickier to use than Equations 1 and 2, Appendix B includes a hand-held calculator program that elicits L_{min}, L_{max}, U_{min}, U_{max}, and the location of the peak. It then determines, for any value of L_A, whether it is above, below, or at the peak, and automatically selects and uses Equation 3 or 4 on the basis of that determination.

Nonlinear Location Measures

Our use of linear and bilinear location measures is, as we pointed out in Chapter 2, an enormous simplification, very much out of the spirit of formal decision analysis. Our justification for doing it is straightforward. If your desirability or utility function increases steadily or decreases steadily, or has one interior maximum, then this approximation will work so well that there is little point in using anything more sophisticated. If the approximation reverses an evaluative ordering as compared with some other form of location measure elicitation, it will be because the options being ordered are so close to each other in attractiveness that fluctuations in weight or location measure judgments would also be enough to change orderings, and consequently no strong conclusions about ordering would be justifiable in any case.

However, if your intuition is severely violated by thinking of your assessments of desirability as expressible by one or at most two straight lines, you have a rather simple alternative. You can simply draw a graph with the physical measure on the X axis, the 0 to 100 scale on the Y axis, and draw whatever function most appeals to you in that graph. Drawing a graph is obviously mandatory in the very rare case in which your utility function has two or more peaks in it. The only such multi-peaked function with some claim to be a potential location measure that we can think of is the relationship between credit-worthiness and age. Bankers tell us that credit applicants in the 20's or in the 40's are preferable as credit risks to applicants in the 30's, other things equal.

Methods and Elicitation

Of various methods available for obtaining location measures, the linear method is probably simplest when appropriate, and the draw-a-graph method is most versatile. But there are many others, and we explain one more. Consider Exhibit 2, which was actually used in the evaluation of the version of Status Offenders (DSO) program described in Chapter 3. In that particular example, respondents are asked to place six reference values of the attribute on a 0-to-100 scale.

They may do so linearly or non-linearly. Then, by interpolation, the location of any other value of the attribute can be determined. Indeed, the method of Exhibit 2 is not limited to attributes in which either more is better than less or less is better than more; it can also deal with attributes in which an intermediate value is optimal.

Results from a mail-and-return use of Exhibit 2 and other forms just like it for other attributes are given for one respondent in Exhibit 3. Note that all of the utility functions presented except that for Delinquency could be exceedingly well approximated by linear or bilinear utility functions. And the approximation for Delinquency, though less than ideal, would in fact not be at all bad: It is unnecessary for most multiattribute utility evaluations to capture every bump and wiggle of such curves -- and indeed experience suggests that such bumps and wiggles are not highly reproducible.

Choice Among Competing Location Measures

Especially in social conflict situations, one common topic of conflict is what location measures to use in an evaluation. An embattled program can live or die depending on whether it is evaluated primarily by looking at measures on which it is doing poorly or on measures on which it is doing well. We discussed the issue in Chapter 3, using the nuclear power debate as an example. Those who wish to defend nuclear power would like the expected

EXHIBIT 2 OF CHAPTER 5

AN EXAMPLE OF THE ALLOCATE 100 POINTS METHOD:

EVALUATING JUVENILE DELINQUENCY TREATMENT PROGRAMS

ATTRIBUTE: AVERAGE NUMBER OF ARRESTS PER STATUS OFFENDER, PER YEAR

STEP 1. CONSIDER THE AVERAGE NUMBER OF ARRESTS PER STATUS OFFENDER, PER YEAR. TYPICALLY, ARRESTS OF INDIVIDUAL STATUS OFFENDERS MIGHT RANGE FROM 0 - 10 OR MORE PER YEAR.

STEP 2. CONSIDER THE FOLLOWING SAMPLE OF POINTS ON THE RANGE GIVEN IN STEP 1:

AVERAGE NUMBER OF ARRESTS PER YEAR

0 2 4 6 8 10 or more

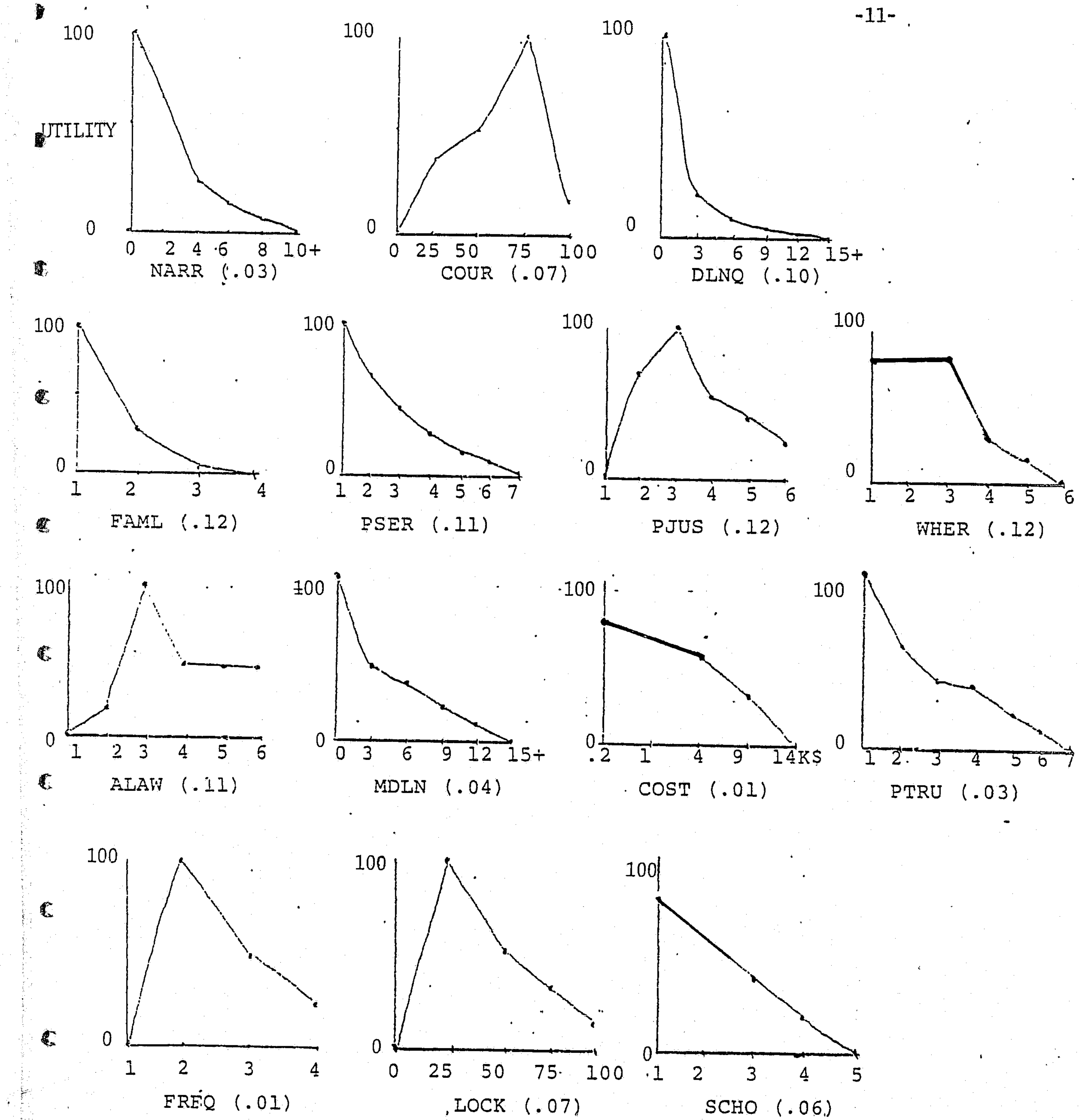
STEP 3. SELECT THE POINT IN STEP TWO THAT YOU CONSIDER "BEST" AND PLACE IT AT 100 ON THE UTILITY POINT SCALE.

STEP 4. SELECT THE POINT IN STEP TWO THAT YOU CONSIDER "WORST" AND PLACE IT AT ZERO ON THE UTILITY POINT SCALE.

STEP 5. PLACE THE REMAINING POINTS IN STEP TWO ON THE UTILITY POINT SCALE (RELATIVE TO THE BEST AND WORST) SO THAT DISTANCES BETWEEN THEIR LOCATIONS REFLECT THEIR RELATIVE WORTH TO YOU.

0....5....10....15....20....25....30....35....40....45....50....55....60....65....70....75....80....85....90....95....100

UTILITY POINT SCALE



An illustration of subjective value utility curves and importance weights (parentheses) for one DSO expert. (See Exhibit 8 of Chapter 3 for a description of the attributes.) The values in the parentheses are the importance weights.

deaths following an accident to be measured as expected fatalities--a quite low number. Those who wish to oppose it want to use deaths following the maximum credible accident for the same purpose--and that is a much larger number.

The only reason we bring this up again here is to reiterate the suggestion of Chapter 3. From a MAUT point of view, there is no good reason for preferring one such measure to another; why not use both? Then the conflict over which measure is appropriate becomes a conflict about which twig of a value tree to weight heavily. This does not eliminate the conflict--but it reduces it to the familiar form that, in our view, characterizes virtually every evaluative conflict we have seen: disagreement, not about values, their structures, or about possible measures, but rather disagreement about weights.

The Relation between Location Measures and Weights

In Chapter 4, on weighting, we never really explained what a weight was. Our reason for not doing so was that, although we could and did use evocative words like "importance", we could give no precise definition until we had discussed location measures. The two concepts are very closely intertwined.

A weight is an exchange rate among location measures. Suppose, for example, that you assign a weight of .5 to attribute A and a weight of .25 to attribute B. A is thus twice as "important" as B (.50/.25 = 2). You are then implicitly saying that, if all other attributes are held constant, you would be willing to assign equal aggregate utilities to option 1, with a location measure of 0 on attribute A and 100 on attribute B, and to option 2, with a location measure of 50 on attribute A and 0 on attribute B. That is, you will pay two utility units of attribute B to gain one of attribute A. Or, to put it the other way around, you would pay half a utility unit of attribute A to gain a full utility unit of attribute B.

It is for this reason that judges of weights need to know the maximum and minimum location values of each attribute. They also should understand the idea that weights are exchange rates. The reason for this is that once a weight is assigned to an attribute it is always for a given range of location measures. If, the range of such measures should change, the weights should change accordingly. Attributes with location measures that are reduced to a very narrow range from minimum to maximum plausible should receive lower weights relative to the other attributes and vice versa.

We recognize that the interpretation of weights as exchange rates among location measures is a subtle idea. In working with stakeholders who are assigning importance weights we explain it by giving examples such as that of the second previous paragraph.

Fortunately, this formal interpretation corresponds quite closely to our intuitive understanding of the word "importance". It seems natural to say that we will pay a lot on an unimportant attribute to gain a little on an important one. This discussion does no more than to give that natural thought quantitative form. It is also fortunate that the relationship between importance weights and location measures will not create a problem if the range of the location measures is chosen carefully in the first place, and then left alone. From a practical standpoint, it is better to have an option with location measures outside the original range receive utilities outside the 0-100 range then to try to figure out how to fix up the weights to adjust for a range change.

Collecting Location Measures

As with most MAUT procedures, we consider it wise to elicit location measures from individual respondents or small groups if possible. Procedures for value and weight elicitation are sufficiently demanding that we are rather skeptical about mail-and-return procedures for carrying them out. But location measures are sometimes easier to elicit, depending on how subjective they are, how willing the respondents are to take time and thought in assessing them, and on the nature and complexity of the dimensions on which they must be assessed. Any of the procedures we have proposed in this Chapter are at least conceivable as mail-and-return procedures -- with the proviso that simple linear procedures are so simple that the respondent may wonder what they are for. We have used all of these methods, but our mail-and-return experience has been confined to the method illustrated in Exhibit 2. It seemed to work well; the responses made sense and the respondents had no undue amount of difficulty in using it.

The Office of the Rentalsman Example

In Chapter 3 we disucssed the evaluation of the Office of the Rentalsman. We return to it now to illustrate one way to deal with the problem of collecting location measures that make comparisons among programs possible when only one program exists.

In such cases, the natural evaluative strategy is to compare the program in place with an alternative to it -- in this case, with the program it replaced, which was to handle landlord-tenant disputes in the courts.

That evaluation identified 16 attributes; they were listed as Exhibit 6 of Chapter 3. The evaluation was intended, among other purposes, to find out how satisfied or dissatisfied the respondents, all expert in such dispute resolution problems and closely familiar with the Office of the Rentalsman experiment, were with the OR itself, and with the Court as an alternative to it.

Each respondent judged how well each attribute was being served by the OR now, how well it should be served, and how well it had been served in the Court system that the OR had replaced. As an illustration of how this was done, consider the attribute of Visibility. Each expert was presented with the following:

VISIBILITY

The extent to which the citizens are aware of the Office and its function.

---

1. How much visibility exists now in the OR?

   (min) 1 2 3 4 5 6 7 (max)

2. How much visibility should exist in the OR?

   (min) 1 2 3 4 5 6 7 (max)

3. How much visibility do you think existed in the Court with respect to resolving landlord-tenant disputes prior to the establishment of the OR?

   (min) 1 2 3 4 5 6 7 (max)

This scale was repeated for all 16 attributes. The expert was requested to give three ratings on a 7-point scale:

1. How much of the attribute did he (she) think actually existed in the operation of the OR at the present time.
2. How much of the attribute did he (she) think should be connected with the operation of the OR.
3. How much of the attribute did he (she) think existed in the Provincial Small Claims Court with respect to resolving landlord-tenant disputes prior to the establishment of the OR.

The answer to the first of these questions can be interpreted as an estimate of how well the expert thought that particular attribute is being served in the OR. The answer to the second question compared to the answer to question 1 can be interpreted as an estimate of how much the expert thought that particular attribute should be increased or decreased in the operation of the OR. For example, if the expert gave a numerical response of 3 to question 1 and a 5 to question 2, that is an indication that that particular attribute's function should be increased.

The answer to the third question can be interpreted as an estimate of how well each expert thought that particular attribute was served in the Court with respect to resolving landlord-tenant disputes. By comparing the answer to question 3 to that of question 1 we can obtain an estimate of satisfaction-dissatisfaction of using the OR versus using the Court for handling landlord-tenant disputes. For example, if the expert gave a numerical response of 3 to question 1 and 5 question 3, that is an indication that the Court was "doing better" on that attribute than the OR. If the converse were true then the expert thought the OR was functioning better on that attribute than the Court.

Note, in this example, the judgments provided by the experts on the 3 scales described above replaced the location-utility values described earlier in this chapter and they are on a 7 point scale, not a 100 point scale. However, the use of a 7 point scale was arbitrary, and it could have been replaced with a 100 point scale. Alternatively, Equation 1 could be used to transform location measures from the 7 point scale to a 0-100 scale. For a score of 3 on the 7 point scale, the arithmetic would be

$$\text{Location} = 100(3 - 1)(7 - 1) = 33.33.$$

The results of this particular evaluation will be given in Chapter 6.

CHAPTER 6 SUMMARY

Chapter 6 describes how to aggregate the two sets of numbers arrived at in the MAUT process: the importance weights, one for each attribute, and the location measures (utilities) assigned to each decision alternative on each of the attributes. Only one aggregation rule is presented since research indicates that it is the most practical and useful. The rule: multiply each location measure on the attribute by the importance weight for that attribute and add up all these products into an aggregate utility U for each decision alternative, entity, or option being evaluated. The larger the numerical value of U the "better", thus whatever decision alternative receives the largest U should be considered "best" under the procedures described in this document. Two examples are given in detail showing all the calculations. An illustration of how to present an evaluation profile for a program is given. With such a profile it is possible to see the strong and weak points of a program and thus guide the decision maker in deciding where program improvements should be made. The chapter concludes with a discussion of sub-aggregation of location measures with the appropriate weights. This is done by aggregating the weights and location measures at different levels of the value tree. For example, each branch of the tree can be treated as a separate MAUT analysis. An actual example of sub-aggregation is left for chapter 7 where it is shown useful for sensitivity analyses of the MAUT procedure.

CHAPTER 6

PUTTING IT ALL TOGETHER: THE AGGREGATION OF WEIGHTS AND VALUES

Through the techniques described in Chapters 4 and 5 we have arrived at two sets of numbers: the importance weights, one for each attribute, usually normalized to sum to 1.0, and the utilities assigned to each decision alternative on each of the attributes usually (but not always) expressed on a scale from 0 to 100. The next step, to be described in this chapter, is to aggregate these two sets of numbers into one composite using an aggregation rule. Atlhough the literature describes very complicated aggregation rules, we use only one because it is by far the simplest. The equation takes the following form:

$$U_j = \sum_{i=1}^{n} w_i u_{ij} \qquad (1)$$

where U_j is the overall or composite utility for the j^{th} option such as a particular site or a particular delinquency treatment program. w_i is the normalized weight assigned to the i^{th} attribute; u_{ij} is the utility of the j^{th} option on the i^{th} attribute. The symbol $\sum_{i=1}^{n}$ means to sum the weighted utilities over all the attributes from the first (1) to the last (n). We illustrated this arithmetic in the site selection of example of Chapter 2. and we refer the reader back to Exhibit 3 in that chapter.

Equation (1) will yield a composite utility for each alternative using a particular set of weights and utilities. These can be averages, of course, but we recommend that they also be kept separate so that each stakeholder's evaluation can be retained for analysis. The larger the numerical value of U the "better", thus whatever decision alternative receives the largest

U should be considered the "best" under the procedure described in this manual. We turn now to examples of this aggregation process.

Example 1. The DSO Project Evaluation

The Diversion of Status Offenders (DSO) Project, evaluation included 14 attributes and 7 different treatment programs. How the importance weights and the utilities can be aggregated into a composite U value is illustrated in Exhibit 1. In Exhibit 1 only two of the programs are listed since we are trying to illustrate the process. Also, the numbers given in Exhibit 1 are hypothetical. Exhibit 1 can be considered a work sheet to accomplish the desired result indicated by equation (1). Consider the program labelled MULT, Multiple Service Center (cf. Exhibit 7 of Chapter 3). On the attribute Number of Arrests (NARR), this particular program had two arrests and this is the program location value for this particular attributes. The location utility graphs (Exhibit 3 of Chapter 5) assigned a utility of 80 to this location value. Or equivalently equation (2) of chapter 5 could be used to assign a utility of 80. When the value is multiplied by the importance weight (.03) for the attribute, the result is 2.4 and that is the weighted utility for that program on that attribute. When all of these weighted utilities are summed, the result is 54.1 which is the composite utility (U) for this particular program. The program labelled GHOM (Group Home), in a similar fashion, receives a composite utility of 49.3. Thus, this particular evaluation technique indicates that MULT is doing better than GHOM.

EXHIBIT 1 OF CHAPTER 6

AN EXAMPLE OF AGGREGATION OF WEIGHTED UTILITIES

Two of the Treatment Programs under the Diversion of Status Offenders (DSO) Project

Attribute[1] / Program[2]	1 NARR	2 COUR	3 DLNQ	4 FAML	5 PSER	6 PJUS	7 WHER	8 ALAW	9 MDLN	10 COST	11 PTRU	12 FREQ	13 LOCK	14 SCHO	Composite Utility
Weight	(.03)	(.07)	(.10)	(.12)	(.11)	(.12)	(.12)	(.11)	(.04)	(.01)	(.03)	(.01)	(.07)	(.06)	
GHOM															
Location Measure	4.0	5.0	12.0	1.0	1.0	4.0	4.0	3.0	9.0	200.0	4.0	1.0	85.0	5.0	
Utility	20.0	6.0	0.0	100.0	100.0	50.0	30.0	100.0	20.0	70.0	50.0	0.0	22.0	0.0	
Weighted Utility	0.7	0.4	0.0	12.2	10.0	5.8	3.7	10.9	0.8	1.0	1.4	0.0	1.5	0.4	49.3
MULT															
Location Measure	2.0	30.0	6.0	1.0	3.0	1.0	1.0	3.0	3.0	1500.0	6.0	1.0	65.0	2.0	
Utility	80.0	34.0	7.0	100.0	45.0	0.0	75.0	100.0	40.0	91.7	10.0	0.0	34.0	100.0	
Weighted Utility	2.4	2.3	0.7	12.2	4.9	0.0	9.2	10.9	1.6	1.2	0.3	0.0	2.3	6.1	54.1

(2.4) — The utility multiplied by the importance weight (.03) yields this value.

(80.0) — The utility assigned for 2 arrests is 80.0. (Cf. Exhibit 3 of Chapter 5.)

(2.0) — There were 2 arrests under this program.

The composite value is obtained by adding up all the weighted utilities.

[1]The attributes are described in Exhibit 9 of Chapter 3.

[2]The program descriptions are given in Exhibit 8 of Chapter 3.

Example 2. The Evaluation of the Office of the Rentalsman (OR).

The Office of the Rentalsman (OR) described in Chapter 3 was set up in Vancouver, British Columbia to handle landlord-tenant disputes. It replaced the Court as the means for resolving such disputes. Thus, in applying MAUT for this evaluation it was not possible to make a direct program comparison as in the previous DSO example. The respondents did, however, identify 16 value attributes considered relevant for resolving disputes. (These are listed as Exhibit 5 of Chapter 3.) This evaluation concentrated on having the experts "rate" the OR as it was operating now on each of the attributes; on how it should be operating; and how the Court operated on each of the attributes prior to the creation of OR. We will present three sets of results of this particular MAUT evaluation ranging from highly summarized (averages) down to individual "expert by attribute" statements.

Calculation of overall weighted values.

An overall weighted value for how the OR as it was operating now and and how it should be operating was obtained. And, in a similar vein we obtained an overall weighted value for the Court when it was being used to resolve landlord-tenant disputes, prior to the advent of the OR. This was done by taking the numerical utility judgment that each expert gave to the sixteen attributes and multiplying each such utility by the importance weight of the attribute. These weighted utilities were then summed over all attributes. This is the familiar additive rule represented by Equation 1. The result can be interpreted as an overall measure of "goodness" or composite utility.

The results are given in Exhibit 2 which presents the summary statistics for the overall weighted utilities, averaged over all twelve

experts. By comparing the overall averages, either mean or median, we see that the OR was considered an improvement on the Court, although the differences are not that large. We also note that the averages for the OR utility Should are always higher than the OR utility Now, indicating that the experts thought there was room for improvement.

The reason why the Court receives an overall weighted utility not too far below that for the OR is that the Court receives higher ratings on the attributes of FAIRNESS and IMPARTIALITY, and these were the attributes that received the highest importance weights (cf. Exhibit 5 of Chapter 4). We turn now to a more detailed comparison of the individual attributes.

Individual values for each attribute.

Exhibit 3 presents the results of the averaged utilities for each value attribute for the OR Now, how it Should be, and how it was Prior to the OR (the Court). By comparing the OR Now column with the Prior (Court) column we see that in six of the sixteen attributes, the Court receives higher average utilities than the OR. These are for the attributes: FAIRNESS, IMPARTIALITY, POWER (to enforce decisions), CONSISTENCY, INDEPENDENCE, and JURISDICTION. For nine of the other attributes, the OR received higher averaged utilities than the Court, the attribute of EXPENSE (to the institution) received the same average utility (4.5), indicating the OR and the Court were considered equivalent on this particular attribute.

Individual satisfaction indices.

As a final result to be presented we did an analysis of how each individual expert rated the OR on how each importance attribute was being represented in the current operation (Now) and how it should be represented (Should). We can state categorically that just about every expert

EXHIBIT 2 OF CHAPTER 6

Overall Weighted Utilities for the OR Now,
How It Should Be and,
How It Was Prior to the OR (Court)[1]

	Utility Now	Utility Should	Utility Prior (Court)
Mean	434.1	599.7	425.3
Median	432.2	603.3	407.4
Standard Deviation	83.7	30.7	82.9
Range			
Minimum	312.7	547.2	275.5
Maximum	526.1	635.5	568.3

1 The numbers in the Table are averaged over all 12 experts.

EXHIBIT 3 OF CHAPTER 6

Average Utilities for Each Value Attribute
For the OR Now, How It Should Be and
How It Was Prior to the OR (Court)

Value Attribute	NOW	SHOULD	PRIOR (COURT)
Fairness	38.1	76.0	53.1
Impartiality	39.9	76.1	57.1
Accessibility	31.9	55.9	23.6
Education	26.9	47.4	11.7
Visibility	27.6	43.1	20.1
Speed	23.1	43.1	18.4
Power	18.1	37.9	41.5
Expertise	28.6	40.6	27.2
Flexibility	25.0	31.4	22.4
Consistency	20.6	32.5	23.6
Independence	19.5	34.7	29.4
Investigative Power	18.3	32.2	11.4
Jurisdiction	18.4	23.9	25.5
Cost[1]	13.8	25.8	8.9
Informality	12.2	14.6	8.0
Expense[1]	4.5	4.8	4.5

1 The numbers assigned to these attributes were transformed such that the higher the number, the "better", i.e., less cost, less expense.

thought the OR could improve on every attribute. Thus, the result is not too interesting and we will not present the data in tabular form. However, a similar analysis on how the OR compared with the Court indicated no such unanimity of opinion. For each expert we compared his (her) numerical rating of the OR on each of the attributes with the rating that same expert gave to the Court. We were thus able to tabulate how "satisfied" each expert was with respect to the OR versus the Court on each attribute and for completeness we included an overall measure of satisfaction. The results are given in Exhibit 4 which depicts the attributes along the columns and the experts along the rows.

A "+" in the body of the table indicates that the expert thought the OR was better than the Court on that attribute, and a "-" indicates the converse, i.e., the Court was better than the OR on that attribute. A "0" means the expert was neutral on that particular attribute. The results in Table 6 are self-explanatory, but a few things should be pointed out. Note, as indicated by the last column, only two experts thought the Court was overall a better way to handle landlord-tenant disputes than the OR was. The other ten experts preferred the OR. On particular attributes, however, the Court was considered much better than the OR. The most notable of these is POWER (to enforce a decision) in which not a single expert yielded a "+" for the OR. On other attributes, on the other hand, the OR was considered much better than the Court. Examples of these are ACCESSIBILITY, EDUCATION, FLEXIBILITY, INVESTIGATIVE POWER, SPEED, and VISIBILITY. Finally we would like to point out that on the attribute of EXPERTISE, seven of the experts thought the Court was better, three thought the OR was better, and two had no opinion or were neutral.

EXHIBIT 4 OF CHAPTER 6

How Each Expert Compared the Office of the Rentalsman with the Court on the Value Attributes.[1]

EXPERT	ACCESSIBILITY	CONSISTENCY	COST	EDUCATION	EXPENSE	EXPERTISE	FAIRNESS	FLEXIBILITY	IMPARTIALITY	INDEPENDENCE	INFORMATION	INVESTIGATIVE POWER	JURISDICTION	POWER	SPEED	VISIBILITY	OVERALL VALUE[2]
1	+	+	-	+	-	+	+	-	0	0	+	+	-	-	+	+	+
2	-	-	-	+	-	-	-	-	-	-	+	-	-	-	-	-	-
3	+	+	+	+	+	0	+	+	+	0	0	+	-	-	+	+	+
4	+	-	-	0	0	-	-	+	-	+	-	+	-	-	+	+	+
5	+	0	+	+	-	0	0	+	-	-	+	+	+	0	+	+	+
6	+	-	+	+	+	-	-	+	+	-	+	+	-	-	+	+	+
7	+	-	0	0	-	+	-	-	-	-	+	+	-	-	+	+	+
8	0	-	0	+	0	-	-	+	-	-	0	+	-	-	-	-	-
9	+	-	+	+	0	-	-	+	-	-	+	0	-	-	+	+	+
10	+	+	+	+	+	+	+	-	0	-	-	+	-	-	+	+	+
11	-	+	+	+	0	-	-	+	+	-	+	0	+	-	0	+	+
12	+	+	+	+	0	-	-	+	+	-	+	+	+	-	+	+	+

1 A "+" means the expert thought the Office of the Rentalsman was better than the Court on that attribute; a "-" means that some experts thought the Court was better on that attribute; a "0" means a neutral or no judgment.

2 The value assigned here is for the overall weighted satisfaction index.

Comment

The evaluation indicates that overall the OR seems to be working successfully. All save two of the experts concur in this. However, the most interesting aspect of the evaluation results is the indication of where the operation of the OR can be improved. This is an example of one way to use MAUT to fine tune a program. There were attributes that the experts thought were better handled by the courts, the most prominent of these being FAIRNESS and IMPARTIALITY, which received the highest importance weights. Also, there is strong evidence that the experts would like to see the OR improve substantially on these two attributes (cf. Exhibit 3). This is one of the advantages of this particular evaluation technique -- it essentially gives a "profile" of the strengths and the weaknesses of a particular social system, in this case the OR. It also indicates a possible dilemma.

In order to improve on the attributes of FAIRNESS and IMPARTIALITY, the procedural practices of the OR may have to get more formal, at least for those disputes requiring a hearing. This may mean adopting some of the formal mechanisms of the Court which it replaced. Again, the MAUT technique indicates that this is probably desirable since the attribute of INFORMALITY receives the second lowest importance weight of 3.3 which is more than three times lower than the most important attributes of FAIRNESS and IMPARTIALITY.

Subaggregation

Equation 1 of this Chapter suggests that the goal of MAUT is to come up with one number, U_j, for each object of evaluation, expressing in highly concentrated form how well that object does on all evaluative dimentions.

But whether that much compression is appropriate depends very much on the purpose of the evaluation. Indeed, the discussion of the Office of the Rentalsman example shows that an aggregate was too compressed even in that instance; much of the discussion looked at individual location measures.

It is not too difficult to compare location measures if, as in that example, there are 16 attributes and two objects of evaluation. But as the numbers of attributes and objects of evaluation increase, the need for aggregation becomes imperative. Fortunately, aggregation need not be an all-or-nothing affair. If a Value Tree has been developed, one can select an appropriate level of higher order value, and aggregate up to it. This is done by using Equation 1 of this chapter, but starting the process of multiplying the weights down through the tree at the level to which you wish to subaggregate, and thus in effect treating each branch of the tree as a separate MAUT analysis. Then the MAUT scores on each branch separately can be presented as a Value Profile -- an aggregated but still informative summary of how each object of evaluation stands with respect to each of the higher-level values considered relevant to its assessment.

We would illustrate the technique here, but it would be a waste of space to do so. Chapter 7 begins with an example of subaggregation applied to the site selection problem of Chapter 2. Each site is characterized by its score on the four top level attributes of that evaluation--and then further arithmetic is done on those already aggregated scores.

Although the idea of subaggregation has been obvious ever since MAUT came into existence, we know of relatively few instances of its application in program evaluation contexts. That surprises us. The techinque seems obvious and appropriate, especially if the purpose of the evaluation is to monitor a program or to guide the process of fine tuning it. The reason, if course, is that it gives information at whatever level of detail seems to be just right for the purpose at hand.

Is it science fictional of us to think that the day might come when every progress report would be accompanied by a subaggregated Value Profile of the project, with the location measures justified if necessary, and with weights agreed on in advance by sponsor and project people? We know of no other way of compressing information into such a clear and sharp display of exactly what one really wants to know.

CHAPTER 7 SUMMARY

Chapter 7 is concerned with an analysis of the MAUT technique itself. It attempts to determine just how sensitive the final result of MAUT is to the numbers and arithmetic that went into the analysis. Sensitivity analysis usually proceeds along a series of steps. Step 1 consists of subaggregation in which the overall utility of each entity being evaluated is recalculated at higher branches of the value tree. Step 2 consists of varying the importance weights on the main branches of the tree involved in the subaggregation. At this stage the effects of dropping value attributes can also be investigated by the simple process of assigning a weight of zero to any attribute one wishes to be eliminated. Step 3 consists of changing the location measures (utilities). The selection of a drug counseling site, first introduced in Chapter 2, is used to illustrate all the arithmetic. With the subaggregated utilities under different weights it is now possible to check again for dominance, i.e., to see if any sites can be eliminated if they are lower in subaggregated utility or worse (higher) in cost than some other site. For any change made in the conduct of a sensitivity analysis the result is always compared with the original MAUT result to see if any major changes take place in the final choice. Often, the final choice is not changed drastically by such changes. When this happens, as it often does in practice, one can have confidence in the MAUT analysis. If this does not happen, i.e., if relatively minor changes in the inputs to MAUT yield quite different results then the MAUT analysis is suspect.

CHAPTER 7

SENSITIVITY ANALYSIS

Sensitivity analysis consists of changing some of the numbers that went into the initial MAUT analysis and doing it over again to see if the conclusions change, and if so by how much.

Obviously, since the initial calculations of MAUT are demanding, any sensitivity analysis will be more so. Indeed, full-blown sensitivity analyses require more in the way of computational support than this document assumes to be available. Consequently, we do not plan to illustrate an elaborate sensitivity analysis. Nothing that we propose in this chapter is beyond the capabilities of the program for working with trees that we provide in Appendices C and D -- in fact, we did the arithmetic using those programs ourselves. Even so, we must warn you that this chapter is tedious and hard to read. If you are not doing a MAUT evaluation, skim it. If you are, get paper and pencil and follow the arithmetic.

Probably the most important kind of sensitivity to look at is sensitivity to weights. This is important both because weights are the essence of value judgments, and because weights, being purely subjective numbers about which people disagree, are more likely to be in dispute than location measures, which may be objective, may depend on the judgments of experts -- or may be in some cases matters of intense controversy. Moreover, if there is some debate about whether a branch or twig belongs in the analysis at all, it can be in effect eliminated in a sensitivity analysis by giving it a weight of 0, or almost that.

We confine our discussion to the drug counseling center siting example.

Step 1: Subaggregating the Location Measure Matrix

The Director, reviewing the original analysis presented in Chapter 2, felt that she was satisfied with the choices of twigs used, and with the lower-level weights. While she might have quibbled with some of the latter, she also knew that changes in lower-level weights will have much less impact on aggregate utility that will changes in higher-level weights. This decision permitted a considerable simplification of the analysis.

Her first step was to subaggregate each of the sites to a level just below the top. This could be done by hand, but she chose to do it instead by using the program in Appendix C separately for each of the four top-level branches of the tree, entering only the appropriate location measures. The result is shown in Exhibit 1. The cost of each site in terms of rent per year is also included in Exhibit 1. The numbers in Exhibit 1 are easy to calculate by hand. Consider the aggregated value of 60.00 for site 1 on Value B. If you refer back to Exhibit 2 of Chapter 2, you will find that it is composed of a location measure of 40 for BA and of 80 for BB. The weights of each of these within the B branch (from Exhibit 1 of Chapter 2) are .5. So the calculation is .5(40) + .5(80) = 60. Exhibit 2 of this chapter shows how the value of 63.60 for site 1 on branch A was calculated. All other numbers in Exhibit 1 are calculated in the same way. The hand-held calculator program of Appendix C makes this easy to do.

The Director next inspected this new table of subaggregated location measures for dominance. Of course, 6 is dominated by 2, as before, but otherwise no new dominated sites appear. Some other site could have become dominated at this stage. If its cost (rent) had also been equal to or higher than that of the dominating site, it could be summarily eliminated from further analysis. (Actually, we would eliminate site 6 at this point if we were doing the analysis "for real"; the price difference between it and 2, which dominates it in utility, is so small that 6 has no chance of ending up the winner. We keep it in because it helps to illustrate some important tools later.)

Next the Director noted which sites were best and worst on each top level value, and what the range between minimum and maximum values were for each. The results appear in Exhibit 3.

EXHIBIT 1 OF CHAPTER 7

Level 1: Subaggregate Utilities of the Six Sites and Cost

Site Number	Attribute Label A	B	C	D	Cost (rent per year)
1	63.60	60.00	32.40	6.40	$48,000
2	48.40	50.00	64.00	59.20	53,300
3	43.90	47.50	13.80	75.60	54,600
4	70.20	50.00	30.80	66.20	60,600
5	35.55	47.50	90.80	35.60	67,800
6	43.30	50.00	46.80	52.80	53,200

EXHIBIT 2 OF CHAPTER 7

Calculation of a Subaggregated Utility (Site 1, Attribute A)

Twig Label	Location Measure	Weight	Location Measure x Weight
AA	90	.39	35.10
AB	50	.21	10.50
AC	30	.14	4.20
AD	90	.14	12.60
AE	10	.12	1.20
Sums		1.00	63.60

EXHIBIT 3 OF CHAPTER 7

Best Sites, Worst Sites, and Range for Top Level Attributes

Attributes	Best Site[1]	Worst Site[1]	Range from Best to Worst
A	4 (70.20)	5 (35.55)	34.65
B	1 (60)	3 & 5 (47.50)	12.50
C	5 (90.80)	3 (13.80)	77.00
D	3 (75.60)	1 (6.40)	69.20

[1]The values in the parentheses are the sub-aggregated utilities from Exhibit 1.

Step 2: Varying Weights

Inspection of Exhibit 3 told the Director that site 5 was most likely to be influenced by the sensitivity analysis -- indeed, it could become a top contender if C, with the widest range, were given a high weight. Changing the weight of B was unlikely to make much difference, since the range of variation in B was so small, relative to the other attributes. Changing weights on C and D would make the most difference in ordering of sites. However, D, administrative convenience, had originally received a weight of .14, and the Director felt that that was plenty. She was most concerned about the weights on A and C. So she decided to change those two weights, holding B and D constant, to see what would happen. (Since the weights must sum to 1, it is impossible to change only one weight.) Since her feeling had initially been that A received too high a weight, she tried only lower weights for A and higher ones for C. Inspection of Exhibits 1 and 3 told her that such changes would help 2 and 5, and hurt 1 and 4. To make these weight changes, she did not need to go back to the original location measures. Instead, she used Exhibit 1 and applied the weights to those aggregated utilities directly, a simple computational task either by hand or with the program given in Appendix C.

First, she decided to explore a radical change, in which B continued to have a weight of .24 and D a weight of .14, but A had a weight of .23 and C a weight of .39. Next she tried an intermediate change, in which A weighed .33 and C weighed .29. The results of both calculations are shown in Exhibit 4. Now she considered old and new rank orderings in utility. The original rank ordering in aggregate utility (from Chapter 2) had been 425163. The first set of weights of Exhibit 4 produces a rank ordering of 524613. The second produces 254613. As expected, the weight changes hurt sites 1 and 4, helped 5 greatly, and helped 2 a little. Inspection of the original location measures or of Exhibit 1 of this chapter will show why this is so; site 5 is outstanding on the twigs under C,

EXHIBIT 4 OF CHAPTER 7

Result of a Sensitivity Analysis of Changing Weights for Attributes A and C

Site No.	(1) Weights A = .23, C = .39 B = .24, D = .14	(2) Weights A = .33, C = .29 B = .24, D = .14	Cost
1	42.56	45.68	$48,000
2	56.38	54.82	53,300
3	37.46	40.47	54,600
4	49.43	53.37	60,600
5	59.98	54.45	67,800
6	47.60	47.26	53,200

but does much less well on most other twigs.

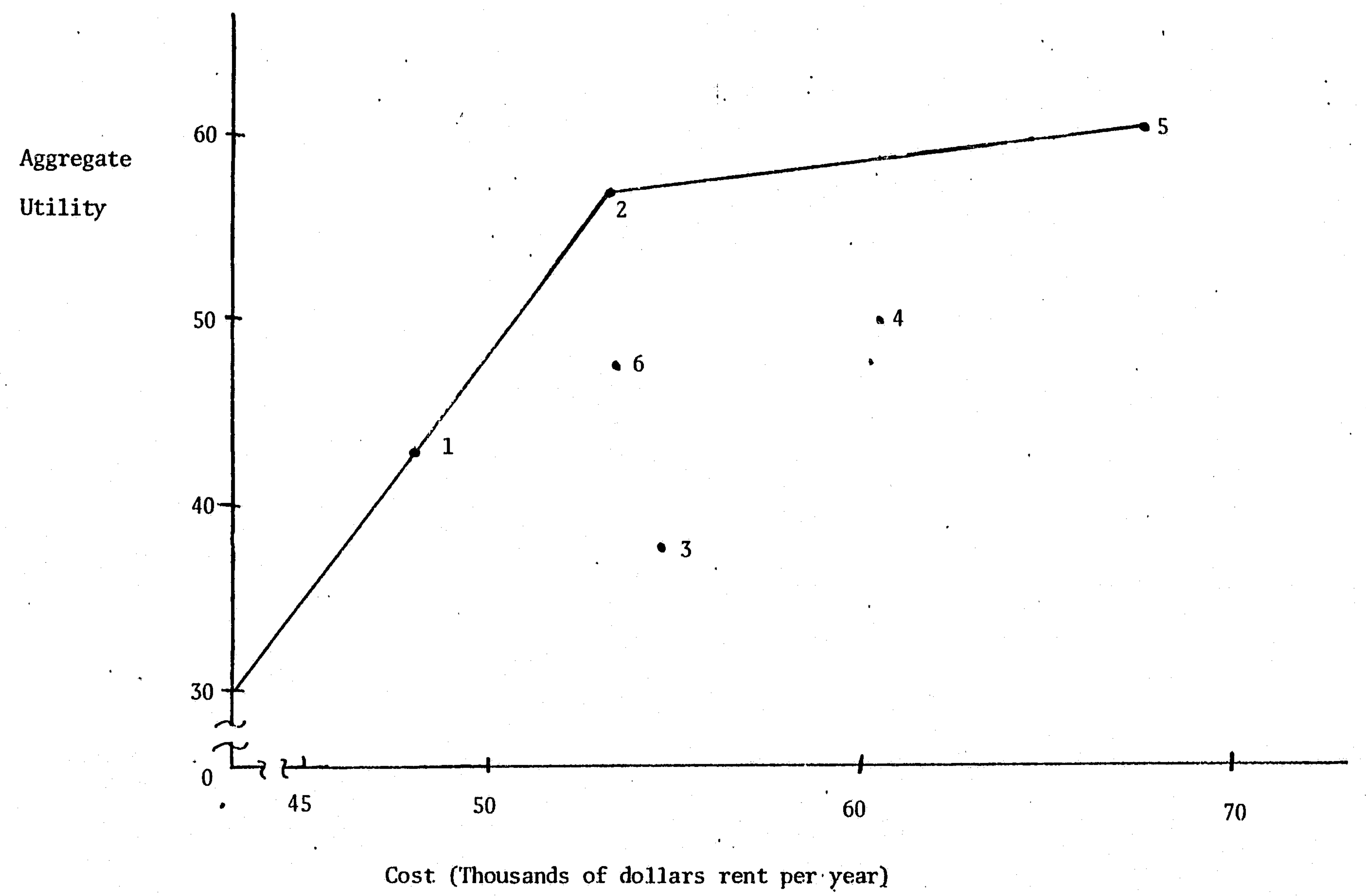
Tha rank ordering from lowest to highest cost (rent) is 162345. For the original weights that left sites 1, 2, and 4 as contenders. For the second set of weights, the first in Exhibit 4, the contenders are 1, 2, 5, and 6. Site 4 is now dominated by site 2. For the third set of weights the contenders are 1, 2, and 6; 5 drops out because it is dominated by 2. In both cases 6 remains a contender because of its relatively low cost. We now illustrate another way of depicting which sites are viable contenders. Simply plot the alternatives in a graph relating aggregated utility (Y axis) to the cost (X axis). This is illustrated in Exhibit 5 for the second set of weights (the first set listed in Exhibit 4) (A = .23, B = .24, C = .39, D = .14). Note the line segment connecting sites 1, 2, and 5. These are clearly the best sites and even though 6 remains in contention because of its low cost, it will not survive. This is a general property of such plots. Any sites in this example that fall below the concave line segment will be eliminated. The converse is also true. Any new alternative plotted on or above the curve would become a contender. Depending on its location, it could cause previous contenders now to be dominated.

At this point, the Director would like to know whether the intermediate possibilities for these two new sets of weights are realistic contenders in view of the relation between cost and utility. Consider first the second set of weights (the first set listed in Exhibit 4). For this set of weights, she needs to prepare a table like Exhibit 5 of Chapter 2, showing successive differences for the four potential contenders in both utility and cost. Exhibit 6 shows that table. The arithmetic is the same as that performedin Exhibit 5 of Chapter 2. Set the utility and cost difference for the top site to 0. Then subtract from each other site's utility and cost the utility and cost of the site just above it. For example, the utility difference of 5.04 for site 6 is obtained by subtracting from the utility for site 6 the utility of site 1 which is just above it (47.50 - 42.56 = 5.04). These are the increments in utility and cost for the undominated sites.

Inspection of the ratio of cost increments to utility increments for site 2 tells us at once that something is wrong. A bit of thought makes the nature

EXHIBIT 5 OF CHAPTER 7

GRAPHICAL REPRESENTATION OF UTILITY VERSUS COST (WEIGHTS: A = .23, B = .24, C = .39, D = .14)



of the problem clear, Obviously, site 6 represents a large increment in cost for a small increment in utility, as compared with site 2. That is why the ratio of increments for site 2 is so low. You may recall from Exhibit 5 of Chapter 2 that such ratios should continuously increase in such a table, if all the contenders are true contenders. In this case they do not. The implication (which is in fact a formal theorem that we will not prove) is that site 6 could never be chosen, no matter what the dollar value of a utility point, from these sites with these weights. Consequently, it is not a contender, and should be eliminated.

Although a quick look at the numbers tells us that it is unlikely and a look at Exhibit 5 tells us that it is not so, it could be the case considering Exhibit 6 alone, that site 2 is also not a contender. To check, we must calculate Exhibit 6 all over again, eliminating site 6. Exhibit 7 shows the result. As expected, this check simply confirms that 2 is a real contender. It also makes clear that, even with these weights, the range between the dollar value of a utility point for which 2 is best and that for which 5 is best is very large indeed. (Note that Exhibit 6 could not have been used to reach this conclusion, since it did not properly reflect the dollar value of a utility point at which 2 becomes preferable to 1.) With Exhibit 7 we know that if the dollar value per utility point is less that $383, site 1 is best. If it is between $383 and $4027, site 2 is best. If over $4027, site 5 is best.

Since 6 was also a contender with the third set of weights (A = .33, C = .29), the same check must be made on its contender status in that case also. Exhibit 8 does so. Again the numbers in the far right column do not increase steadily. And again the option next below the point of the decrease should be deleted -- that is, option 6. In this case, there is no point in redoing Exhibit 8 without site 6 in it to make sure that 1 and 2 are contenders, since the best in utility (site 2) and the lowest in cost (site 1) will always be contenders. It is, however, useful to know what the value of a utility point is for which 2 becomes better than 1 with these new weights -- and Exhibit 8 does not tell us that. Exhibit 9 does.

EXHIBIT 6 OF CHAPTER 7

Increments in Utilities and Costs, Sites 1,2,5,6
(Weights: A = .23, B = .24, C = .39, D = .14)

Site No.	Utility	Cost	Utility Differences (Increment)	Cost Differences (increment)	Cost Incr./ Utility Incr.
1	42.56	$ 48,000	0	0	
6	47.60	53,200	5.04	$ 5200	$ 1032
2	56.38	53,300	8.78	100	11
5	59.98	67,800	3.60	14500	4027

EXHIBIT 7 OF CHAPTER 7

Increments in Utilities and Costs, Sites 1, 2, 5
(Weights: A = .33, B = .24, C = .29, D = .14)

Site No.	Utility Differences (Increment)	Cost Differences (Increment)	Cost Incr./Utility Incr.
1	0	0	
2	13.82	$5300	$383
5	3.60	14500	4027

EXHIBIT 8 OF CHAPTER 7

Increments in Utilities and Costs, A = .33, Sites 1, 2, and 6
(Weights: A = .33, B = .24, C = .29, D = .14)

Site No.	Utility Differences (Increment)	Cost Differences (Increment)	Cost Incr./Utility Incr.
1	0	0	
6	1.58	$5200	$3294
2	7.56	100	13

If th value of a utility point is more than $580, then with these weights site 2 is preferable to site 1.[1]

In the original analysis of Chapter 2, 4 had been better than 2, though substantially more expensive. At this point, it becomes interesting to see for what values of the weights of A and C (holding the weights of B and D constant) 2 and 4 become equal in utility. From Exhibit 1 of this chapter, the larger the weight of A relative to C, the better 4 will do compared with 2. So a weight of A higher than .33 is needed. Trial and error (or solution of a simple linear equation) shows that .36 does the trick. However, even if the weight of A were higher than .36, a second look at Exhibit 5 of Chapter 2 will remind you that with the weights used in that chapter, 2 was still preferable to 4 over a wide range of the dollar values of a utility point. The Director combined that fact with her own feeling that the weight of A should not be greater than .36 -- a feeling that she also checked with her contract monitor, who agreed. So she narrowed the set of contenders down to 1 and 2.

Step 3: Varying Location Measures

This led her to re-examine the location measures for 1 and 2. (With more computational resources, she might have done a more extensive examination of the location measures.) She was particularly concerned with twig CA, the location measures describing the suitability of the individual treatment rooms.

[1] Footnote for technicians only: The tools based on successive differences presented in this section of this chapter serve two purposes. First, they identify any points that lie in concave portions of the function relating aggregate utility to cost, and eliminate them. Second, for the convex function that remains, they identify the critical slopes, or trade-off relations between aggregate utility and money, for which preferences switch from one option to the next. These tools seem to us simpler and more precise than their graphical equivalents, such as Exhibit 5, though it is not possible to deal with derivatives in a completely simple way. If the x axis were treated as the utility of money rather than money itself, the tools of this chapter would be general. This, of course, assumes that the weighted additive utility function holds.

EXHIBIT 9 OF CHAPTER 7

Increments in Utilities and Costs, Sites 1 and 2
(Weights: A = .33, B = .24, C = .29, D = .14)

Site No.	Utility Difference (Increment)	Cost Difference (Increment)	Cost Incr./Utility Incr.
1	0	0	
2	9.14	5300	580

This was a judgmental dimension combining their number, their suitability to their function, their convenience of location to the waiting room, their attractiveness, and so on. Even with a weight of .19 for branch C, CA still receives a weight of (.10) (see Exhibit 1 of Chapter 2) which made it the most important of the judgmental C-branch twigs. Both BA and BB got higher weights, but she was quite content with her assessment of them. Moreover, sites 1 and 2 differed radically on CA; 1 got a location measure of 10, and 2 got one of 80. They differed little on BA and BB (see Exhibit 2 of Chapter 2). She decided to see how much she would need to change the location measures for CA in order to make sites 1 and 2 equal in utility. She of course had to pick a set of upper level weights for the calculation, and chose to use the one in which branch C had a weight of .29. With that upper level weight for C, the twig weight for CA is .29 x .52 = .151. The utility difference between 1 and 2, for that set of weights, is 9.14. So she would need a change in the location measure for CA of 9.14/.151 = 60.5 to make 1 and 2 equal in utility with that set of weights. She quickly concluded that, though she was less secure about that judgment than about some of the others, she could not possibly have been 60 points off. She did not see any need to repeat previous calculations with new location measures for CA. Site 5 already scored very high on it (90), and site 4, though lower (50), was not so low that it would make a substantial difference to the outcome.

As a final step, more from curiosity than because she considered it crucial, she performed the computationally very easy task of trying equal weights. First she tried weighting equally all the original location measures (from Chapter 2). After doing so, she eliminated dominated sites, and found that sites 5, 2, 6, and 1 were left. Further checking showed, as usual, that although site 6 was undominated, it could never be chosen. (In very technical language, the elimination of sites 3 and 4 is by ordinal dominance; the elimination of site 6 is by cardinal dominance.) The utility difference between 2 and 5 was only 2.5 utility points, and the price per utility point

that made 5 preferable to 2 was $5800 -- a very high price per utility point indeed for this example. Next, she weighted equally the subaggregated utilities of Exhibit 1 of this chapter. The ordinal dominance analysis left her with sites 1, 6, and 2, in order of increasing utility. And, as usual, site 6 dropped out for reasons of cardinal dominance -- that is, because any price per utility point that would make it better than 1 would make 2 better than it. In both of these analyses, the price per utility point that would make 2 preferable to 1 was modest: $578 in the first case and $358 in the second. These findings strengthened her already strong feeling not only that site 2 was her best choice, but also that that choice was stable under a variety of different weights.

Step 4: Choose the Final Form of the Analysis

After this (somewhat abbreviated) sensitivity analysis, she invited members of the staff who wished to do so to try their own weights on the contending sites, repeating the analyses already presented. None led to substantially different results. She then decided that she would recommend site 2 to her sponsors and to the other stakeholders. The set of weights that she most believed, and that she based the recommendation on, was the one for which A was weighted .33 and C was weighted .29. Inspection of Exhibit 4 of this chapter shows that sites 1 and 2 differ by 9.14 utility points and by $5300 in price. You will recall from Exhibit 9 of this chapter that for this set of weights the value of a utility point in dollars for which 2 becomes preferable to 1 is $580. To convince herself and others that a utility point was at least that valuable, she included examples like that presented in Chapter 2 in which stakeholders are asked to judge the value of a 100-point swing in a twig. Twig DA (with a weight of .09) was used as the example in Chapter 2. The new weights do not affect that twig, since the weight of D remains unchanged. So a 100-point swing in DA is still a change of 9 points in aggregate utility. (That number would have changed if a twig under the B or the C branch had been chosen for the example.) For any value greater than $5220 of a 100-point swing in DA, a point of utility is worth more than \$580, since 5220/.09 = 580. None of her stakeholders were reluctant to agree that such a swing was worth at least that much. So she recommended site 2.

Her recommendation was accepted -- in part because it was bolstered by such a thorough analysis of the available alternatives to it.

Generalizations about Sensitivity Analysis

The preceding paragraphs imply some generalizations about sensitivity analysis. The first and most important is that careful inspection of the original numbers, and of compressions of them like Exhibit 1 of this Chapter serve to guide exploration; there is no reason to try computations at random.

The second generalization is that if a number of options are to be examined, it is usually desirable and feasible to reduce that number considerably, thus confining the sensitivity analysis to a much smaller set of options and so reducing arithmetic. Dominance permits this to be done formally. But even without dominance, it is usually possible to recognize options to be dropped. Although we did not do it, it is obvious from Exhibit 1 that, unless the weight of administrative convenience is allowed to go much higher than the Director felt was reasonable, site 3 should have been summarily dropped. Every option dropped reduces the arithmetical labor of doing a sensitivity analysis, and so permits a more careful job to be done. Note, however, that utility dominance is not in itself an adequate basis for dropping options, if they are cheap. Site 6 remained in contention almost to the end. This is essentially an accident of this example. Site 2 is so close to site 6 in price that it would have been sensible to drop site 6 intuitively. As it turns out, a final analysis that included sites 1, 2, 4, and 5 would have captured everything that the Director (though not the readers of this manual) would need to know.

Obviously, weight sensitivity should be looked at first. Usually, it is enough to work only with the higher level weights, since the lower level ones have so much less effect. It is also computationally more convenient. The device, illustrated by Exhibit 1 of this Chapter, of aggregating weights

and location measures up to the level just below the one at which the sensitivity analysis is to be done will make the arithmetic easier.

Since there are so many location measures in any MAUT analysis, it is not easy or straightforward to figure out which ones to vary in a sensitivity analysis. The obvious guides are (1) that it makes little sense to vary location measures on low-weighted twigs, (2) that it is equally inappropriate to vary location measures that do not discriminate among the viable contenders, unless there is some reason to believe that one of them is wrong, and (3) that one should think hard about which location measures one trusts, and which are dubious. Without more computational aid than we can offer, exploration of changing location measures is likely to be perfunctory.

For situations like that examined in this document, in which both utility and cost are often relevant, calculation of cost per utility point implied by choice of each undominated option is an indispensable adjunct to analysis based on utilities alone. Formally, these techniques amount to bringing cost in as another attribute of the utility function. We have chosen to treat cost and utility separately until the end of the analysis in this manual because in many evaluation situations choice among options on the basis of both utility and cost is not the issue -- though in many others it is. Consequently, we wanted to provide both methods for dealing with utility alone and methods for combining utility with cost. This chapter has abundantly illustrated how inclusion of cost considerations can affect choices that, in their absence, would be based on utility alone, and has offered methods for exploring the sensitivity of the evaluation process to cost even in the absence of the judgments that establish direct relationships between utility points and dollars. That is because such judgments are often particularly hard to make, and are likely to be more controversial than other judgments that enter into the evaluation process. Fortunately, as this chapter has suggested, rather crude assessments of the value of a utility point will often permit clear-cut choice after dominated options have been eliminated.

No one has yet discovered rules that guide onein making simultaneous changes of the many numbers in a sensitivity analysis. Even with extensive computer support, such large-scale sensitivity analyses are often confusing and frustrating. If the conclusion seems to be relatively stable under changes of weights, as it was in this example, you are usually justified in treating it as valid.

That is how the result will usually turn out. Even in this example, which was designed to be sensitive to weights and turned out to be so much so that we originally wondered whether we had not chosen a poor example, the finding in favor of site 2 ends up seeming quite stable, given a willingness to spend some extra rent for more utility. You are unlikely to encounter a real case more weight-sensitive than this one.

If you do, it will be for one or the other or both of two reasons. One is that two or more options are so close in aggregated utility that is makes virtually no difference which is picked, and so changes in weight switch them back and forth in ranking. This is essentially what happened to sites 2, 4, and 5 of this example. In that kind of situation, other attributes not included in the original analysis should be considered, since the original analysis led to what amounts to a tie. In this analysis, the additional attribute was cost. The other reason for sensitivity to weights is that the options, instead of being relatively homogeneous in location measures like site 2, include many very high and very low ones like site 5. Obviously the larger the range of variation of location measures within an option, the greater the sensitivity to weights will be. Only in this case do we feel that real precision in knowing weights is indispensable. And, in our real world experience, such instances are relatively rare, though they do occur.

A final comment. We have said very little about uncertainty about weights or location measures. Location measures that depend on judgment are likely to vary depending both on who does the judging and on when and how the number is elicited. While

we do pay attention to the magnitude of utility differences in considering whether a utility difference is worth what it costs or not, we do not otherwise pay much attention to such variability. The reason is simply that when a decision must be made, you work with what you have. It makes no difference whether a difference is "significant" in some statistical sense or not. If it is the best guidance you have about what option to choose, you should follow it. And if it is not, then you will be able to incorporate whatever additional guidance you may be able to get into an expanded multiattribute utility analysis. You always leave attributes out to keep the analysis simple. If an analysis leaves you uncomfortably uncertain about what to do or think, and if the problem is important enough to justify another iteration, you can always go back, include more value attributes, reweight, reaggregate, and repeat the analysis. Or, as occasionally happens, if the formal analysis leaves several options very close together in attractiveness, you may choose to consider other attributes informally. This is highly appropriate if they all point in the same direction. If not, then they present you with the kind of tradeoff problem for which MAUT is intended, and an expanded version of the formal analysis becomes the method of choice.

APPENDICES

Appendices A through D contain description of four computer programs for the Texas Instrument (TI) hand held calculators TI 58, 58C, and 59 to help evaluators using MAUT to do the arithmetic required for the analysis. The complete listing of the program is given and can be keyed in by the user directly. The program described in Appendix D is too long to be used in the TI 58 or 58C; the 59 is necessary. Of course, it is not necessary to use these programs in applying MAUT. However, almost all of the arithmetic illustrated in this manual was done using these programs, and we found them quite useful. We welcome comments and criticisms. Users of the TI 59's, which can read programs from magnetic cards, can obtain cards with these programs on them by writing to the authors.

APPENDIX A

A Normalizing Program for the TI 58 (or 58C or 59)

This program simply turns a column of un-normalized numbers into their normalized equivalents, then briefly displays the sum, which is always 1, and then automatically resets with the number of elements in the previous list displayed. It can cope with no more than 44 numbers.

Storage assignments

00	contains N
01	contains N decremented
02	is a prompt for the next number
03	sums the numbers
04.	holds the sum of normalized numbers
05	holds addresses for indirect addressing

Program

INV 2nd Fix	
5	
2nd Op 17	reallocates partitioning to increase number of memories
RCL 00	gets old N back into display
2nd CMs	
2nd Fix 4	
R/S	enter N
STO 00	
STO 01	
5	
STO 05	

2.

2nd Lbl A	
2nd Op 22	adds 1 to 02
2nd Op 25	adds 1 to 05
RCL 02	prompts the next number
R/S	enter a number
SUM 03	
STO 2nd Ind 05	
2nd Dsz 1 A	loops to get another number
RCL 00	
STO 01	
5	
STO 05	
2nd Lbl B	
2nd Op 25	
(	
RCL 2nd Ind 05	
÷	
RCL 03	
)	calculates a normalized number
SUM 04	
R/S	copy the normalized number
2nd Dsz 1 B	loop for another normalized number
RCL 04	
2nd Pause	momentarily displays that the sum of normalized numbers is 1
RST	end of program. Don't forget to press LRN, then RST

APPENDIX B

A Program to Compute Linear or Bilinear Utilities on a 0-100 Scale for the TI 58 (or 58C or 59)

This trivial 146- instruction program makes it convenient to work with linear or bilinear utilities. For linear utilities, it works by asking for an upper bound and a lower bound and then does the appropriate linear calculations depending on whether the user tells it that more is better than less or less is better than more. Cases with a peak in the middle are more complex, because the two bounds may not have the same degree of undesirability.

User instructions. Key in the program, then press RST R/S. The display will show a 0, and asks for the upper bound on the x axis of the utility function. Key it in and press R/S. Next, similarly key in the lower bound on the x axis and press R/S. If you do not intend to use bilinear utilities, simply enter the first x for which you wish a y (utility), and press A if more is better than less, and B if less is better than more. Thereafter, you will retain the same bounds unless you press RST R/S, which resets the program to the top, erases all memory, and starts over again. To get another utility of the same kind, simply enter x and press R/S. You can get to the other kind with the same bounds on the x axis by pressing B if you previously pressed A, or vice versa, but normally you will wish to change bounds when you change directions on the utility scale.

If you wish to work with bilinear utilities, the procedure is the same until you have entered the first x. Then press C. The machine will halt with 0 in the display. At this point you should enter the value, between lower and upper bounds on x, that you consider optimal; it will automatically be assigned a utility of 100. Press R/S. The machine will halt again, with the upper bound on x in the display. Enter the utility, on a 0 to 100 scale, that you consider to correspond to that x. Press R/S again, and the machine will halt with the lower bound on x in the display. Enter, on the 0 to 100 scale, the utility that you consider to correspond to that x. Normally either the upper or the lower bound or both will have a utility of 0. (Example: if you like sugar in your coffee, you may dislike 0 sugar intensely, and it should get a utility of 0. But if your ideal point is 1½ tsp., you might regard 3 as too much, and want to assign it a utility of about 70, since you might feel that too much sugar is much preferable to none at all.)

When you press R/S again, you will get the utility of the x you originally entered. If you enter a new x and press R/S, you will get its utility based on the judgments you just made. If you want to change the optimum point or the utility of either bound, press C. If you want to change the bounds, start over again with RST R/S.

Memory allocations:

00	contains highest value of x
01	contains lowest value of x
02	contains optimum value of x, if it is not an extreme (bilinear case only)
03	contains the utility of the highest value of x (bilinear case only)
04	contains the utility of the lowest value of x (bilinear case only)
05	contains the current value of x whose utility is to be calculated
06-08	working registers for 2nd tan, the subroutine that does the calculating

Label allocations:

A	more is better than less
B	less is better than more
C	elicitation for bilinear utilities
D	return point for bilinear utilities, to skip re-elicitation
2nd cos	label for point in bilinear case to which program goes if x is above optimum x
2nd tan	subroutine that actually does all linear calculations

Program:

2nd CMs	
2nd Fix 1	
0	
R/S	enter high value of x
STO 00	
R/S	enter low value of x
STO 01	
R/S	enter x for which utility is wanted, then press A, B, or C
2nd Lbl A	more is better than less
STO 05	
0	

STO 06	
RCL 01	
STO 07	
RCL 00	
STO 08	initialization completed
2nd Lbl 2nd A'	return point to avoid re-initializing
SBR 2nd tan	calculate and display utility, and get a new x
GTO 2nd A'	loop for another utility
2nd Lbl B	less is better than more
STO 05	
0	
STO 06	
RCL 01	
STO 08	
RCL 00	
STO 07	initialization completed
2nd Lbl 2nd B'	return point to avoid re-initializing
SBR 2nd tan	calculate and display utility, and get a new x
GTO 2nd B'	loop for another utility
2nd Lbl C	bilinear case
STO 05	
0	
R/S	enter optimal x
STO 02	
x⇌t	prepare for later tests
RCL 00	prompt with highest value of x
R/S	enter utility of that value of x
STO 03	
RCL 01	prompt with lowest value of x

```
R/S                 enter utility of that value of x
STO 04
RCL 05              recover x for which utility is needed for test coming up
2nd Lbl D           return point to avoid re-initialization
2nd x ≥ t 2nd cos   go to 2nd cos if x is above optimal x
RCL 04
STO 06
RCL 01
STO 07
RCL 02
STO 08              initialization completed for this case (x below optimal x)
SBR 2nd tan         calculate and display utility; elicit new x
GTO D               loop back for utility of the new x
2nd Lbl 2nd cos     x is greater than optimal x
RCL 03
STO 06
RCL 02
STO 08
RCL 00
STO 07              initialization completed for this case
SBR 2nd tan         calculate and display utility; elicit new x
GTO D               loop back for utility of the new x
2nd Lbl 2nd tan     computational subroutine
(
100
-
RCL 06
```



```
)
X
(
RCL 05
-
RCL 07
)
÷
(
RCL 08
-
RCL 07
)
+
RCL 06
=
R/S                 copy utility; enter new x value
STO 05
INV SBR
RST
```

APPENDIX C

A Program to Form Tree Structures, Accept Weights, Accept Location Measures and Calculate Multiattribute Utilities for the TI 58 (or 58C or 59)

The following program, designed for the TI 59, will perform all computations needed to work with a value tree whose structure and weighting is known ahead of time to its user. It confines itself to the additive case. The 196 instruction program will fit into the TI 58, but the limited memory of the 58 will permit only 23 nodes, which may be too few for some applications. The program repartitions the 59, and so can accommodate up to 83 nodes. The first four lines of the program would be omitted if the 58 were used.

As given, the program cannot sub-aggregate; it only computes total utilities. But it would be easy to use it to sub-aggregate by putting the structure underneath a node up to which sub-aggregation is desired into the program, filling in location measures appropriate to the twigs underneath that node, getting the aggregate utility for that node, and then starting over again with the structure underneath the next node at the same level of sub-aggregation. (This would require slight attention to notation, since the notation would treat each as a separate tree.)

The structure of the program makes use of numbers to the left of the decimal point as labels for nodes, and numbers to the right of the decimal place as weights, though the latter are never seen in the display except as they are entered. It automatically constructs the tree by exploiting the fact that normalized weights below any node sum to 1; it works across from left to right at each successively lower level of the tree, and can accommodate a tree five levels deep, but no deeper. It distinguishes twigs from other nodes by adding 100,000 to each twig, but the number is never seen. The user must, of course, inform the program which nodes are twigs and which are not, since it cannot know that for itself.

A word of caution. The program interprets "summing to 1" as meaning equal to or greater than .999. This has two implications. One is that no node can have a normalized weight as low as .001 if it comes at the right hand end of a list of nodes. The other is that you have no protection from a keying error in keying in the right hand weight for a set of nodes. If you meant .5, but keyed .7, the program will go merrily on, and the result at the end will be screwed up.

Another word of caution. If you happen to have 10 or more nodes within a given group whose weights must sum to 1, the possibility of ambiguous identifiers exists. Thus the identifier 11 could refer to the 11th node at the top level, or to the first node at the second level under node 1 at the top level. In that example, if the top level had exactly 11 nodes, the two identifiers would appear one after the other. More generally, they will be separated. You should be able to cope with the ambiguity by remembering that the program always moves from left to right across the list of nodes at a given level before descending to the next level down. A more serious problem caused by two-digit identifiers is that if a two-

digit identifier appears then the path below it can be only four nodes deep. If two successive two-digit identifiers appear in that path, it can be only three nodes deep. If this should in fact be a problem, you can solve it by relabeling and reordering nodes.

User Instructions

First, either key in the program or (if you are using a TI-59 with a card containing the program available) enter a 1 in the newly turned on machine and then read the card; the program length is 196 instructions. The program will automatically repartition the 59, and must work with the initial partitioning of the 58. If you are using the 58, remember not to key in the first four instructions.
After the program is in, if it has been keyed in press RST, then R/S. If it was read from a card, simply press R/S. A 1 will appear in the display to label the first node. Enter the weight of that node to no more than 4 decimal places and then press R/S. A 2 will appear, labelling the next node. Continue until all nodes and their weights at the top level are entered; the sum of weights must be exactly (not approximately) 1. Then the program will display 11, indicating the first node at the second level; enter its weight. R/S will cause it to store the weight, multiplied by the weight of the next higher node, in this case 1, and to display 12. Continue as before. The program will accept weights of zero, but not weights of 1. It should never encounter a weight smaller than .002 in the righ-hand node of a set that adds to 1.

Whenever you encounter a twig (a node with no other nodes beneath it) the process changes slightly. You key in the weight, and then press A. This tells the program to label this as a twig, and to skip it in subsequent runs across lower levels of the tree. Every branch of the tree must have all of its ends labelled as twigs in this way. If you accidentally miss one, press RST R/S, and start all over again from 1. Any node having a weight of zero should also be labelled as a twig.

After all of the tree and all weights have been entered, the program will automatically display the label of the first twig. Enter the corresponding location measure, and press R/S. The label of the second twig will be displayed next. Twigs are ordered from top to bottom and, within a level, from left to right. When all location measures associated with twigs have been entered (the program doesn't care, but we recommend that they be on a 0-100 scale), the program will automatically calculate and display the aggregate utility, on the same scale as the location measures, of whatever is being evaluated. If you wish, RCL 01 will give the twig count. Another push of R/S, after that has been copied, will momentarily show 1.0000 in the display to prove that the final weights do indeed add to 1, and then will reset to the top, forgetting all about the tree structure in the process. If you should want to re-use the same tree structure and weights with a new set of location measures, then when the final aggregate utility has been displayed, instead of pressing R/S, press B. This will leave the tree structure and weights intact, but prepare the program for interrogating you about a new set of location measures for twigs, starting by displaying the label of the first twig. This feature is quite convenient if several things are to be evaluated by means of the same structure and weights.

The following is a convenient example:

Top level Node				1			2				3
Normalized wt.				.4			.4				.2
2nd level node		11			12	21	22	31		32	33
Normalized wt.		.6			.4*	.2*	.8*	.2*		.4	.4*
Final wt.		.24			.16	.08	.32	.04		.08	.08
3rd level node	111		112						321	322	323
Normalized wt.	.5*		.5*						.5*	.2*	.3*
Final wt.	.12		.12						.04	.016	.024

The asterisks identify normalized weights associated with twigs, for which one would enter the weight and then press A instead of R/S. To check the multiplying and summing feature of the program, simply insert 100 as a location measure at each twig or 100 at some and 0 at others. In the former case, the aggregate will be 100; in the latter, it will be 100 less the sum of 100 times the final weights of those twigs assigned 0 location measures.

Memory assignments

00	is a sequential pointer to memory, used for indirect addressing
01	is used to tell the label former which part of the tree it is in. During location measure elicitation, 01 is a twig counter.
02	sums weights
03	holds the next higher weight in weight elicitation, and the weight by location measure products during location measure elicitation.
04	is another pointer to memory, used in the tree building process. In location measure elicitation, it is a temporary bin for labels.
05	holds the label times 10 during tree building, and is temporary storage for weights during location measure elicitation

The remainder of memory is available for storing nodes, one cell per node. The last cell in memory must be empty, since the program checks that to make sure it has looked at all nodes. This fact was taken into consideration in counting the number of nodes the 58 and 59 could accept.

Labels for program parts

A	sets a flag, to identify twigs
B	is the beginning of the location measure eliciting routine
2nd tan	is a location used for looping with B
B'	is a location within D used when one must return from A
2nd cos	is a location within D used to detour around labelling a node as a twig
C	is the traffic cop routine which tells B which part of the tree to build next
D	is the routine that forms labels and elicits weights
D'	is a location within B used for looping
E	is the output and cleanup routine

Program

INV	
2nd FIX	
9	
2nd Op 17	repartition memory for more memory cells
2nd Fix 4	
2nd CMs	
2nd Op 23	weight of next higher dimension starts being 1
.999	
x⇌t	
5	
STO 00	
STO 04	
GTO 2nd tan	end of initialization
2nd Lbl D	start of label and weight routine
RCL 01	C told 01 where to look
INV 2nd Int	get weight part
STO 03	
(	

```
RCL 01
2nd Int
X
10
)
STO 05                 get all but the units digit of 05 ready
2nd Lbl 2nd tan
2nd Op 25
2nd Op 20
(
RCL 05
+
R/S                    enter weight; if twig, push A instead of R/S
2nd Lbl 2nd B'         place to return from A
SUM 02
X
RCL 03                 multiplies normalized weight times final weight of higher node
INV 2nd if flg 0 2nd cos       sets up detour around twig labeler
+
100000
INV 2nd St flg 0       pulls down the flag
2nd Lbl 2nd cos        end of detour
)
STO 2nd Ind 00         puts label and normalized weight away
RCL 02                 ready for sum test
2nd x ≥ t C            if the sum is 1, go to C
GTO 2nd tan            loop around again for another label and weight
2nd Lbl A              to set flag for twig identification
```



```
2nd St flg 0
GTO 2nd B'
2nd Lbl C              master control
0
STO 02
2nd Op 24
100000
x⇄t                    test for twigs is set up
RCL 2nd Ind 04         get a node to test
2nd x ≥ t C            if it is a twig, skip it
0
x⇄t                    test for end of nodes in memory set up
RCL 2nd Ind 04
2nd x = t B            if all weights are elicited, go to location elicitation
STO 01                 if it failed both tests, start eliciting weights
.999
x⇄t
GTO D
2nd Lbl B              location measure elicitator and user
5
STO 00
0
STO 01
STO 02
STO 03
STO 04                 reinitialization complete
2nd Lbl 2nd D'
0
```

x ⇌ t	
2nd Op 20	
RCL 2nd Ind 00	get a node
2nd x = t E	if it is 0, go to output routine
100000	
x ⇌ t	
RCL 2nd Ind 00	
INV 2nd x ≥ t 2nd D'	if it isn't a twig, get another node
-	
100000	get rid of twig identifier
=	
STO 04	we'll need it again
INV 2nd Int	get weight part
SUM 02	
STO 05	
RCL 04	
2nd Int	get label part for display
R/S	enter a location measure
X	
RCL 05	
=	location measure times weight
SUM 03	
2nd Op 21	twig counter
GTO 2nd D'	
2nd Lbl E	output and cleanup
RCL 03	
R/S	copy aggregate utility
RCL 02	
2nd Pause	display sum of weights
RST	END OF PROGRAM. BE SURE TO PRESS LRN AGAIN, THEN RST.

APPENDIX D

A Program for Eliminating Dominated Options and Finding Trade Offs between Utility and Money for the TI 59

The TI 59 program described here will do the following things. First, it will accept a set of pairs of numbers; the first member of each pair is an aggregate utility, and the second is the corresponding cost in dollars (or whatever unit is appropriate). It will associate a numerical label with each pair, depending on the order in which they are read in. After all pairs in the list have been read in, it will automatically eliminate both ordinally and cardinally dominated options. An ordinally dominated option is one in which you can obtain, by choosing another option, either more utility for the same or less cost, or equal utility for less cost. A cardinally dominated option is one that is not ordinally dominated by the definition given above, but would nevertheless not be chosen because some other more expensive option contributes disproportionately more utility. (Formally, cardinally dominated options lie within concave regions of the function relating aggregate utility to cost.) Its final output is the undominated options, listed in order of increasing utility and cost, and, for each option above the lowest in utility and cost, the price per utility point that is just adequate to make the option with which it is associated preferable to its predecessor.

Program limitations. Aside from an inconveniently long running time and a too-long program, the program has only two

CONTINUED

2 OF 3

operational limitations. One is that it can accept a maximum of 12 options. This limitation is minor. We explain later what to do about it.

The other program limitation has to do with options tied with one another in both aggregate utility and cost. (Such an event should be very rare indeed!) The program eliminates all but the first entered option in any such set of complete ties. If the tied option that is considered is eliminated, then its twins should be also. If the tied option that is considered survives, so do all its twins; they must be reattached to the list by hand. From the point of view of this particular set of aggregate utilities and costs, they are indistinguishable, and there is no basis for choice among them. The only choices you have in such a situation are either to change weights, which may untie the tied aggregate utilities, or to consider other attributes not previously included in the multiattribute utility calculation that led to the aggregate utility.

Unfortunately, the program is 478 instructions long, uses 8 cells of working memory, and requires 4 cells of working memory per option. Even manual performance of some of the tasks that the present program performs automatically cannot reduce it in size enough to fit the TI 58 and leave enough memory space to deal with an acceptable number of options.

User instructions. Since this program is for the TI 59 and cannot be used on the TI 58, these user instructions assume

that you will read it in from a magnetic card. Prepare yourself for using it by listing the options you wish to enter in any order you wish, showing for each its serial number, 1 through N, where N is the number of options to be considered , then its aggregate utility, then its cost. Even if you have more then 12 options to consider, list them all.

Now turn on the machine, put a 1 into the display by pressing 1 on the keyboard, and insert the program card into the card reader from the right in normal orientation, that is, with the writing on the face right side up. The machine should take the card and pass it through; you recover it as it comes out the other side. If the read has been successful, the number 1 will appear in the display. If it has not, the display will flash on and off. In that case, press CLR, re-insert 1 in the display, and repeat the read-in process. Occasionally it may be necessary to repeat it as many as 3 or 4 times before you get a good read. Handle the cards with care, and keep them clean. After 1 has been read, put 2 into the display, rotate the card 180 degrees, so that the writing is visible but upside down, and in this orientation pass it through the machine again. If you don't get a good read on side 2, try again until you do; so long as you do not turn off the machine, side 1 remains intact. Remember to press CLR after an unsuccessful read and to re-insert 2 in the display. After 1 and 2 have both been read, the program is ready to be used.

To use it, press R/S. The program will halt with 1.00 showing in the display. This asks you to insert your first aggregate utility. Such numbers should be 100 or less and not less than 0; they may include decimals. After keying in the number, press R/S. The display will, after a short pause, show 1.50, which prompts you to key in the cost associated with the first utility. Costs may be on any scale you wish, and you may use up to 10 digits. The only constraints are that they may not be negative and that they should be actual costs, not rank orders or similar numbers.

After putting in the cost, press R/S again. The display will prompt you with 2.00, which asks for the second aggregate utility. After keying it in, press R/S again; 2.50 asks you for the second cost. And so on. A whole number always asks for that utility; that number with a .50 after it always asks for the associated cost. You will notice that as you enter more and more utilities, the delay between pressing R/S and seeing the prompt for the next cost increases. This is because the program is rearranging the entries in order of increasing utility as you put them in. It is also checking for ties, and eliminating the appropriate option when it finds one (the one with equal or higher cost).

If, in the course of keying in numbers, you make a key-stroke error, try to correct it by pressing CE. If the display changes to 0, you have corrected it, and you can now enter the correct number. This will not work, and indeed there is no

way to correct the error, if you have entered an incorrect number and pressed R/S. In that case, or if CE does not work, then you have no choice but to press CLR, then RST, then R/S. This causes the program to forget whatever you have entered and to start all over again from the top. You will have to re-enter all previous entries. This procedure will get you out of any trouble you may have in using the program--at the cost of destroying whatever is in memory at the time.

If your list contains no more than 12 options, enter them all. If it contains more than 12 options, then after you have entered 12 the machine will automatically begin processing them, thus saving you from inadvertently trying to enter one more, which would foul up the program. As you are entering options, the display always prompts you for the next utility if it can accept another. It can occur if you have more than 12 options to consider, that more than 12 can be entered. This occurs only if two utilities are tied, in which case the machine rejects the dominated one, if one is dominated, or else rejects the second to be entered, if the tie is in both utility and cost. In any case, the display will prompt you for another only if it can accept it. If not, the machine automatically goes into its processing routine.

In rare cases, you might find yourself with 12 surviving options in the machine, in that case, it can evaluate no more. Try to avoid this by including some options that look likely to dominate them (i.e. provide a lot of utility for not too much cost) in the first 12.

Either on the first pass or on some subsequent pass, you will come to the end of the list of options you plan to process. After you have pressed R/S following the last cost entry, the display will prompt you for the next utility. If, for example, you have six options to consider, and have entered them all, the display will prompt you with 7.00. At this point, instead of pressing R/S, press A. After you have pressed A, wait. The delay is certain to be at least 40 seconds, and can be as much as 4½ minutes. When the program halts again with a number in the display, it will be the serial number of the least-utility least-cost option among the survivors. A press of R/S will next display its utility. Another will display its cost. These are shown to make it easier for you to keep track of exactly where you are, and to permit you to be certain that you in fact made no keystroke errors in those numbers. The next press of R/S will produce the serial number of the next higher survivor in utility and cost. Two more R/S presses will show first its utility and then its cost. The next R/S will show the dollar value of a utility point for which that option is just preferable to its predecessor on the list. The next press of R/S after that will produce the serial number of the next survivor, and the next three after that will produce its utility, cost, and dollar value of a utility point. This continues for as many survivors as there are. If at any time during this process you lose track of where you are or miss a number, you can start the output all over again simply by pressing D.

You can recognize that all survivors have been displayed when the next R/S produces the number of options you originally introduced. Now you have a choice. If all the options you wish to evaluate have already been through the program, press R/S again, and the display will show 1.00. This means that it has reset to the top and is ready for a new set of numbers, having forgotten the old.

If you wish to consider additional options in the same set, simply press E when the number of options already entered is displayed, instead of R/S. If you have forgotten how many options remain as survivors, then press RCL 00 before pressing E; RCL 00 at that point (but not later) will show you the number of current survivors.

After you have pressed E, the display will prompt you with the number next higher than the serial number of the last option entered. In other words, if you entered 12 options on the first pass, the prompt will be 13.00. You can now continue entering options from your original list until you have exhausted that list. If you should enter a number of options sufficient so that they, along with the survivors still in the machine, add to 12, the machine will once more go automatically into processing. If you run out of options before that, press A when the machine prompts you for the utility of the N + 1st option, if you had N to start with. If your option list is so long that you expect to make yet another pass, be sure to keep track of the last option entered. It does no harm to re-enter an option, but can obviously be undesirable to skip over one.

You can continue this process of entering options to replace eliminated ones, which will automatically be compared with the survivors of previous passes, as many times as you wish, so long as the total number of viable options does not reach 12. Such an event should be extremely rare, especially if you have included options likely to be dominators among the original set of 12. If it ever should occur, you will need to prune the option set on some basis other than aggregate utility and cost before you can use the program to examine the remaining options for dominance. The fact that the program considers options in sets of 12 at a time cannot lead it to make a mistake by including or excluding an inappropriate option on the final list, or to miscalculate the dollar value of a utility point that makes the next higher option appropriate. So unless you actually have more than 12 viable options, the final output of this process will be correct and complete.

In order to illustrate the process, the following table of inputs and outputs shows what you should expect. (This is a rearranged version of one of the tables in Chapter 7 of the Manual; rearranged to illustrate the fact that the program reorders its entries in order of increasing utility and cost.)

Inputs

Option Number	Aggregate Utility	Cost
1	47.60	$53,200
2	59.98	67,800
3	49.43	60,600
4	37.46	$54,600
5	56.38	53,300
6	45.26	48,000

Outputs

Option Number	Aggregate Utility	Cost	Value per Utility Point That Makes Option just Preferable to its Predecessor
6	45.26	$48,000	
5	56.38	53,300	$ 476.62
2	59 98	67,800	4027.78

For this example, the delay between pushing A when 7.00 appears in the display and the appearance of 6 (serial number of the first option in the output) is 1 minute 13 seconds.

2nd Fix 02	Initialization
2nd CMs	
08	
STO 00	
STO 01	
.5	
STO 06	
2nd Lbl 2nd sin	To loop to from E
.5	
SUM 06	
56	
x⇌t	Prepares t to test for 12 options
RCL 00	
2nd x≥t A	If there are 12, go to processing
RCL 06	
STO 04	
R/S	Ask for a utility
X	
.001	
=	
SUM 04	Store utility with option no. in 04
2nd Lbl 2nd cos	Hunt for place to put new utility
RCL 2nd Ind 01	Recover a stored utility, if 01 points to it
2nd CP	Set t = 0
2nd x = t 2nd Eng	If not, store utility there
RCL 04	
INV 2nd Int	Recover utility
x⇌t	Put in t
RCL 2nd Ind 01	
INV 2nd Int	Recover stored utility

2nd x = t 2nd tan	If tied go to 2nd tan
2nd x≥ t 2nd Prd	If old is greater than new, go to 2nd Prd
04	
SUM 01	Move pointer
GTO 2nd cos	Loop back
2nd Lbl 2nd Prd	Start of "more options" routine
RCL 01	
x⇌t	Put current location in t
RCL 00	
STO 02	
nd Lbl 2nd Fix	Start of number-moving loop
SBR 2nd Int	Moves numbers four cells down
3	
INV SUM 02	
SBR 2nd Int	
9	
INV SUM 02	
RCL 02	
INV 2nd x ≥ t 2nd Eng	Are all needed moves made? If so, store utility
GTO 2nd Fix	Loop back to move more numbers
2nd Lbl 2nd Eng	Stores utility
RCL 04	
STO 2nd Ind 01	
.5	
SUM 06	
RCL 06	Prepare cost prompt

R/S	Ask for cost
2nd Op 21	
STO 2nd Ind 01	Store it
8	
STO 01	
4	
SUM 00	
GTO 2nd sin	Loop back for next option
2nd Lbl 2nd Int	Subroutine to move utilities (etc) 4 memory cells down
RCL 2nd Ind 02	
STO 03	
4	
SUM 03	
RCL 03	
STO 2nd Ind 02	
INV SBR	
2nd Lbl 2nd tan	Subroutine for tied utilities
2nd Op 21	
RCL 2nd Ind 01	
x ⇄ t	Put current cost in t
.5	
SUM 06	
RCL 06	
R/S	Ask for new cost
2nd x≥t 2nd grad	If new cost is higher than old, discard option
STO 2nd Ind 01	Store new cost
2nd Op 31	
RCL 04	

STO 2nd Ind 01	Stored tied utility with new option number
2nd Lbl 2nd grad	Use old, cheaper option
8	
STO 01	
GTO 2nd sin	Loop back for next option
2nd Lbl A	Start of processing options
9	
STO 01	
13	
STO 02	
2nd Lbl lnx	Start of ordinal dominance eliminator
2nd CP	Zero t
RCL 2nd Ind 02	
2nd x = t B	If 02 points to an empty cell, go to B
x ⇄ t	Otherwise, put it into t
RCL 2nd Ind 01	
2nd x≥ t CE	Is what 01 points to greater than what 02 points to?, then CE
4	
SUM 01	
SUM 02	
GTO lnx	
2nd Lbl CE	
2nd CP	
SBR 2nd Prt	eliminates an option
GTO A	Look for more ordinal dominance
2nd Lbl 2nd Prt	Subroutine to move options up
4	
SUM 01	

2nd Lbl CP	
SBR 2nd log	
3	
SUM 01	
SBR 2nd log	
9	
SUM 01	
RCL 2nd Ind 01	
2nd x = t 2nd Adv	If 01 points to 0, go to 2nd Adv
GTO 2nd CP	Otherwise, move more numbers up
2nd Lbl 2nd Adv	Zeros out bottom options, now moved up
4	
INV SUM 01	
0	
STO 2nd Ind 01	
2nd Op 31	
STO 2nd Ind 01	
INV SBR	
2nd Lbl 2nd log	Subroutine to relocate numbers upward
RCL 2nd Ind 01	
STO 04	
4	
INV SUM 01	
RCL 04	
STO 2nd Ind 01	
INV SBR	
2nd Lbl B	Successive difference calculator
13	
x⇄t	

RCL 02	
2nd x = t D	If 02 is 13, go to D
8	If not, get ready to calculate differences
STO 01	
12	
STO 02	
2nd CP	Zero t
2nd Lbl 2nd List	Difference calculator
RCL 2nd Ind 02	
2nd x = t C	If 02 points to 0, go to C
INV 2nd Int	If not, get difference between it and what 01 points to
-	
RCL 2nd Ind 01	
INV 2nd Int	
=	
STO 04	
2	
SUM 02	
RCL 04	
STO 2nd Ind 02	Store the difference (utility)
2nd Op 21	Get ready to calculate cost differences
2nd Op 32	
RCL 2nd Ind 02	
-	
RCL 2nd Ind 01	
=	
STO 04	
2	
SUM 02	

```
RCL 04
STO 2nd Ind 02          Store the difference (cost)
3
SUM 01
2nd Op 22
GTO 2nd List            Loop back
2nd Lbl C               Cardinal dominance eliminator
14
STO 01
15
STO 02
2nd CP
2nd Lbl 2nd C'          To loop back to
RCL 2nd Ind 02
÷
RCL 2nd Ind 01
=
STO 03
4

SUM 01
SUM 02
2nd Op 31
RCL 2nd Ind 01
2nd x = t D             If 01 points to 0, go to D
2nd Op 21
RCL 2nd Ind 02
÷
RCL 2nd Ind 01
```



```
-
RCL 03
=
INV 2nd ≥ t 2nd Pgm     If option is cardinally dominated, get rid of it
GTO 2nd C'              Loop back to hunt more cardinally dominated options
2nd Lbl 2nd Pgm         Cardinally dominated option disposer
5
INV SUM 01
SBR 2nd Prt             Moves an option up
GTO B                   Go back and get new successive differences
2nd Lbl D               Output displayer
2nd CP
8
STO 01
2nd St. flg. 00
9
STO 02
0
STO 00
RCL 2nd Ind 01
2nd Lbl 2nd D'          To loop to
2nd Int
R/S                     Display surviving label
RCL 2nd Ind 01
INV 2nd Int
X
1000
=
R/S                     Display surviving utility
```

2nd Op 20	
RCL 2nd Ind 02	
R/S	Display corresponding cost
2	
SUM 01	
SUM 02	
2nd If flg 00 2nd Write	Loop around benefit/cost caluclator for 1st option
RCL 2nd Ind 02	Calculate incremental utility/incremental cost ratio
÷	
(	
RCL 2nd Ind 01	
INV 2nd Int	
X	
1000	
)	
=	
R/S	Display ratio
2nd Lbl 2nd Write	
INV 2nd St. flg 00	
2	
SUM 01	
SUM 02	
RCL 2nd Ind 01	
2nd x = t 2nd E'	Are we done; if so, go to E'
GTO 2nd D'	Otherwise, display next set of outputs
2nd Lbl 2nd E'	
2nd Op 36	
RCL 06	

R/S	Display original N
RST	End of program, if all options have been processed
2nd Lbl E	If not, get ready for more
RCL 00	
X	
4	
+	
8	
STO 01	
=	
STO 00	
.5	
SUM 06	
GTO 2nd sin	Loop back to get next option