Judgment and Decision Making

An Interdisciplinary Reader

Second Edition

Edited by

Terry Connolly

University of Arizona

Hal R. Arkes

Ohio University

Kenneth R. Hammond

University of Colorado

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK http://www.cup.cam.ac.uk
40 West 20th Street, New York, NY 10011-4211, USA http://www.cup.org
10 Stamford Road, Oakleigh, Melbourne 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain

First published 1999

Printed in the United States of America

Typeface Palatino 9.75/12 pt. *System* LaTeX 2$_\varepsilon$ [TB]

A catalog record for this book is available from the British Library.

Library of Congress Cataloging-in-Publication Data
Judgment and decision making : an interdisciplinary reader / edited by
Terry Connolly, Hal R. Arkes, Kenneth R. Hammond. – Rev. ed.
p. cm. – (Cambridge series on judgment and decision making)
Includes bibliographical references and index.
ISBN 0-521-62355-3 (hardcover). – ISBN 0-521-62602-1 (pbk.)
1. Decision making. 2. Judgment. I. Connolly, Terry.
II. Arkes, Hal R., 1945– . III. Hammond, Kenneth R. IV. Series.
BF441.J79 1999
302.3 – dc21 98-51484
CIP

ISBN 0 521 62355 3 hardback
ISBN 0 521 62602 1 paperback

Contents

Part I

Introduction and Overview

The four chapters in this section lay out some of the topics that have been of central interest to JDM researchers and practitioners. The first, taken from Edwards and Newman's 1982 book, describes in some detail the evaluation process one might go through to select a new office location. The process is explicitly normative in the decision tree approach described in the Introduction: It aims to advise the decision maker what she should do. The problem here is, in fact, simpler in one important sense than in the earlier example, in that no uncertainty is involved. The decision maker knows, for sure, how large, convenient, attractive, and so on each available site is: If she chooses Option 1, she will get a known package of features, if she chooses Option 2, she will get another known package of features, and so on. The problem is in making the trade-offs between the good and bad features of each. The MAUT procedure the authors describe helps the decision maker to set up and make the various judgments required to arrive at a "best" choice. The example is worth following through in detail both to understand what the MAUT technology offers and to better understand the sense in which "best" is used in this approach. Von Winterfeldt and Edwards (1986) describe more complex applications, whereas Edwards and Barron (1994) describe important progress in the weight-assignment problem.

The second chapter, by Tversky and Kahneman, represents an early high-water mark of an important line of research called the "Heuristics and Biases Program." This program, which stimulated an enormous body of research and discussion in the 1970s and 1980s, turned on two fairly simple ideas. First, it is reasonably easy to devise judgment tasks in which many people behave in ways that seem to violate relevant normative standards: They "make mistakes." Second, Tversky and Kahneman proposed an account of many of these errors as manifestations of a small set of cognitive "heuristics" or rules of thumb that, though generally effective enough for most situations, would lead to predictable errors in some carefully constructed tasks. The "error," in this sense, was evidence for the existence of the "heuristic," the existence of which was, in turn, assumed to depend on its general (though

not invariable) adequacy. The "representativeness heuristic," for example, is the expectation that a sample of some process will look roughly like the underlying process. However, our intuitions about how closely samples match processes are imperfect, and can sometimes trip up our judgments. Because many of the studies in this line turned on demonstrating deviations between actual behavior and a normative model, many observers interpreted the results as casting a gloomy light on human cognitive abilities – as suggesting, in one memorable reading, that we are "cognitive cripples" subject to dozens of debilitating biases. Counterattacks and reinterpretations (see, for example, Chapter 35, by Gigerenzer & Goldstein; Chapter 36, by Hilton & Slugowski; and Chapter 31, Juslin, in this volume) argued that the models themselves were not compellingly normative or that the tasks were "fixed" in various ways. Some of these counterattacks have tried to make the case that, far from crippled, we are cognitive heroes, tuned by evolution to superb inferential performance. There is, in fact, almost no evidence on which to estimate an overall human judgmental batting average, even if one could define such a number. A balanced view (see, for example, Jungermann, Chapter 33, this volume) would probably conclude that it is not close to either 0 or 1.0.

The third chapter in this section is an extract from a recent book by Hammond. Hammond suggests that much of the argument over the heuristics and biases research can be resolved by considering the two distinct "metatheories" that have underlain JDM research for years. One, the "correspondence" metatheory, focuses on how well someone's judgments and decisions connect to the real world: Does the doctor diagnose the right disease, does the bettor back the winning horse? The other dominant metatheory focuses on the internal "coherence" of someone's judgments and decisions: Do they hang together rationally, are they internally consistent, are they logically reasonable? Obviously, in a finished, mature science, good theories pass both tests. Newtonian physics was, for several centuries, consistent both internally and with the facts of the world as they were known. JDM research, in contrast, is far less developed, and researchers have tended to emphasize one or the other test, reaching opposite conclusions about human competence. Hammond ties the emphasis on coherence or correspondence thinking to a continuum of different ways of thinking, running from purely analytic to purely intuitive.

Regardless of our conclusions about human competence in general, most of us would be grateful for help when we have to make difficult, high-stakes judgments. Swets, in the fourth chapter in this section, sketches one approach, known (for slightly obscure reasons) as the Theory of Signal Detection (TSD). TSD has not been widely used by JDM researchers (though see Getty et al., Chapter 18, this volume, for a fascinating application in radiology), though it provides a powerful framework for integrating judgments and choices. TSD considers situations in which a decision maker must act in some way (for example, to investigate further in face of a suspicious-looking

X-ray, or to abort an airliner landing because of bad weather) on the basis of an uncertain judgment (of the risk of cancer or of windshear). Either action might be mistaken, so the costs of these errors as well as the uncertainty of the underlying judgment must be considered in planning how to act. Note, incidentally, that Connolly (Chapter 14, this volume) takes a decision analytic approach to such a situation in his discussion of reasonable doubt. The two approaches yield exactly equivalent results, despite the apparently quite different frameworks.

References

Edwards, W., & Barron, F. H. (1994). SMARTS and SMARTER: Improved simple methods for multiattribute measurement. *Organizational Behavior and Human Decision Processes, 60*, 306–325.

Hammond, K. R. (1996). *Human judgment and social policy*. New York, Oxford University Press.

von Winterfeldt, D., & Edwards, W. (1986). *Decision analysis and behavioral research*. New York: Cambridge University Press.

1 Multiattribute Evaluation

Ward Edwards and J. Robert Newman

The purpose of this chapter is to present one approach to evaluation: Multiattribute Utility Technology (MAUT). We have attempted to make a version of MAUT simple and straightforward enough so that the reader can, with diligence and frequent reexaminations of it, conduct relatively straightforward MAUT evaluations him- or herself. In so doing, we will frequently resort to techniques that professional decision analysts 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? 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 the notion of social programs as

This chapter originally appeared in Edwards, W., & Newman, J. R. (1982). *Multiattribute Evaluation*. Beverly Hills, CA: Sage.

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 the 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.

Steps in a MAUT Evaluation

Step 1. Identify the objects of evaluation 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 chapter. Some of the issues inherent in it have already been discussed in this chapter. The next section, devoted to setting up an example that will be carried through the document, illustrates Step 1 for that example.

Step 2. Identify the *stakeholders*. . . .

Step 3. Elicit from stakeholder representatives the relevant *value dimensions* or *attributes*, and (often) organize them into a hierarchical structure called a *value tree*. . . .

Step 4. Assess for each stakeholder group the *relative importance* 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. . . .

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 *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. . . .

Step 6. Aggregate location measures with measures of importance. . . .

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. . . .

Steps 6 and 7 will normally produce the results of a MAUT evaluation. . . .

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 researchers – 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, nonuse of these devices comes close to guaranteeing failure.

By "decisions" we do not necessarily mean anything apocalyptic; the process of fine tuning a program requires decisions too. This chapter unabashedly assumes that either the evaluator or the person or organization commissioning the evaluation 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.

An Example

We now 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. . . . 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 social program environment.

The Problem: How to Evaluate New Locations for a Drug Counseling Center

The Drug-Free Center is a private nonprofit contract center that gives counseling to clients sent to it by the courts of its city as a condition of their probation. 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 prescreening 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. . . .

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 duplications 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. An example of this type of 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 Figure 1.1. We explain the numbers shortly.

Several questions need review at this stage.

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

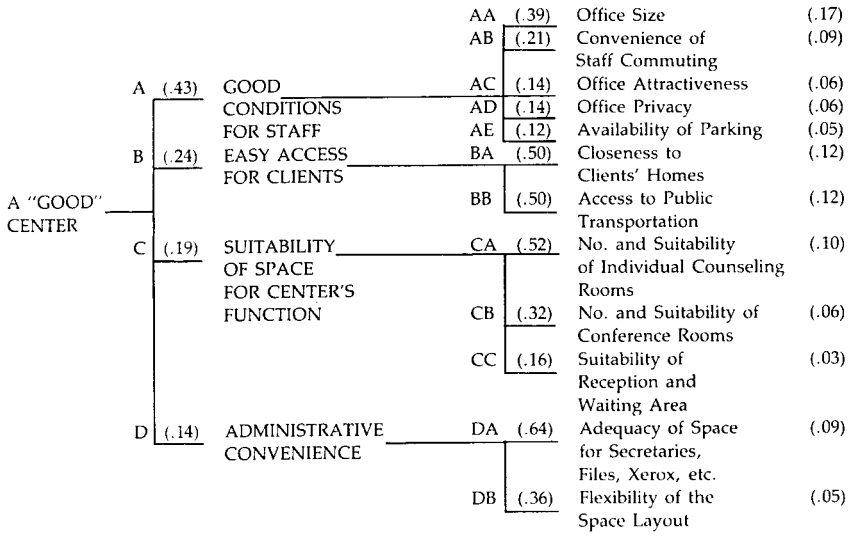

Figure 1.1. A value tree for the Drug-Free Center.

of 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 earlier lists, and eliminated after discussion. Bases for elimination include not only duplication and unimportance, but also that the sites under 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 rent. 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. The 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.) . . . 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 options from a larger set. Typically, the limit on the number that can be chosen is specified by a budget constraint. The methods presented 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) that 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 chapter, 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. By introducing into the analysis early assumptions about how nonfinancial values trade off with money, both benefits and costs can be expressed in dollar terms. We see little merit in doing so for social programs, since early translation of nonmonetary effects into money terms tends to lead to underassessment of the importance of nonfinancial consequences. The methods we present in this section . . . 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 Figure 1.1, 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. Figure 1.1 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. . . . Examples with as many as fourteen 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. . . . 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. . . .

Figure 1.1 contains a notational scheme we have found useful in value trees. Main branches of the tree are labeled with capital letters, A, B, and so on. Subattributes under each main branch are labeled 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 Figure 1.1 are *importance weights* for the attributes. Note that the weights in Figure 1.1 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 $.43 \times .39 = .17$. . . .

The weights presented in Figure 1.1 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*. . . . 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 weights 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.

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 preferred. 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 or utility of each object of evaluation on each attribute of evaluation.

Inspect Figure 1.1 again. Two kinds of values are listed on it. Office size is an objective dimension, measurable 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 Figure 1.1 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 Figure 1.1, 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. . . .

Figure 1.1 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 square footage 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. . . . Formal arguments in support of our use of linearity are far too technical for this chapter. . . .

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 how close each site was 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 to 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.

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 four 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 distances 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 $-29 \ldots$) to that site on that attribute. 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.

Table 1.1 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,

Table 1.1. *Location Measures for Six Sites*

	Twig Label											
Site Number	*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

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 Table 1.1 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 of each site. 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 location measure for that twig, and sum the products. This is illustrated in Table 1.2 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 Table 1.2 were done with hand calculator programs;

Table 1.2. *Calculation of the Aggregate Utility of Site 1*

Twig Label	Weight	Location Measure	Weight × 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	.99
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

Table 1.3. *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

the discrepancy between the 48.79 for site 1 of Table 1.2 and the 48.80 of Table 1.3 is caused by a rounding process in the program. Table 1.3 shows the aggregate utilities and the costs for each of the six sites. The costs are given as annual rents.

Now a version of the idea of dominance can be exploited again. In Table 1.3, 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 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

Table 1.4. *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

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 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 Table 1.4. 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 subtracting the aggregate utility of 2 from that of 4 in Table 1.3: $57.31 - 53.26 = 4.05$. Similarly, the cost difference of $7,300 for site 4 is obtained from Table 1.3 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. . . .

The last column of Table 1.4 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 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 $1188, you should choose site 1; if it is between $1188 and $1802, you should choose site 2; and if it is above $1802, 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 Figure 1.1. 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, and the like. 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, and the like 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 Table 1.2 and note that the twig weight for DA is .090. Consequently, a 100-point change in DA will change aggregate utility by $100 \times .090 = 9$ 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 . . . 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 $1500 per point, and site 2 costs $5300 more than site 1, then site 2 should be penalized 5300/1500 = 3.53 utility points in order to make it comparable to site 1. Similarly, if utility is worth $1500 per point, then site 4 should be penalized by the increment in its costs over site 1, $5300 + $7300 = $12,600, divided by the dollar value of a point; 12,600/1500 = 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.

Table 1.5 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: $9000, $13,500, and $18,000; these correspond, with the weights used in this