ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES 60, 306–325 (1994)

SMARTS and SMARTER: Improved Simple Methods for Multiattribute Utility Measurement

WARD EDWARDS

Social Science Research Institute, University of Southern California

AND

F. HUTTON BARRON

Department of Management Science and Statistics, University of Alabama

This paper presents two approximate methods for multiattribute utility measurement, SMARTS and SMARTER, each based on an elicitation procedure for weights. Both correct an error in SMART, originally proposed by Edwards in 1977, and in addition SMARTER is simpler to use. SMARTS uses linear approximations to single-dimension utility functions, an additive aggregation model, and swing weights. The paper proposes tests for the usability of these approximations. SMARTER, based on a formally justifiable weighting procedure developed by Barron and Barrett, uses the same procedures as SMARTS except that it omits the second of two elicitation steps in swing weights, substituting calculations based on ranks. It can be shown to perform about 98% as well as SMARTS does, without requiring any difficult judgments from elicitees.

This paper presents two methods of multiattribute utility measurement, each based on an elicitation procedure for weights. Both are derived from the spirit and the techniques of SMART (Simple Multi-attribute Rating Technique), originally sketched by Edwards in 1971, and more fully presented and first named in 1977. SMARTS (SMART using Swings) remedies an intellectual error of SMART by using an invention called swing weights; as presented here, it has some other improvements also.

SMARTER (SMART Exploiting Ranks) uses Barron and Barrett's (under review) formally justifiable rank weights to eliminate the most difficult judgmental step in SMARTS. A decision based on these weights, on average, gains 98 to 99% of the utility obtainable by using full elicitation of weights.

SMART should be dead; SMARTS replaced it some time ago.

Edwards' work on this paper was supported in part by the Wood Kalb Foundation. Address reprint requests to Ward Edwards, Department of Psychology, University of Southern California, Los Angeles, CA 90089-1061.

0749-5978/94 $6.00

SMARTER is a dramatic improvement on SMARTS in ease of elicitation. A returnable postcard can hold a SMARTER elicitation for prespecified attributes; interviews are not needed. We think SMARTER is likely to appeal to market researchers, public involvement specialists, and others for whom easy remote elicitation is useful.

This paper assumes a single decision maker throughout; extension to the case of an organization with reasonably agreed-on values is straightforward.

The next section of this paper is a succinct step-by-step description of SMARTS and SMARTER, intended as a how-to-do-it checklist. The checklist is not self-contained; in particular, Steps 7 and 8 are listed in it but described later in the paper. After that, we discuss the two key ideas underlying the paper: multiattribute utility and the strategy of heroic approximation. Then come detailed discussions of technical and how-to-do-it issues, keyed to the step in the procedures for SMARTS and SMARTER to which each is relevant. A reader unfamiliar with these ideas should read it from start to finish. A user, familiar with the ideas but wanting to be reminded of the sequence of steps in the course of an application, should find the checklist sufficient.

SMARTS AND SMARTER: TWO CHECKLISTS

SMARTS

Step 1: Purpose and decision makers. Identify the purpose of the value elicitation, and the individual, organization, or organizations whose values should be elicited. This procedure is complete when you can do two things. (A) Make an explicit and exhaustive list of elicitees, or specify a procedure for identifying elicitees that is guaranteed to produce an acceptable list. (B) Prepare explicit instructions specifying both the nature of the structure and numbers to be elicited and the way in which they will be used. These instructions may be intended for elicitees, but more often are records and/or reminders for elicitor and client.

Step 2: Value tree. Elicit a structure (an objectives hierarchy or value tree; see Keeney & Raiffa, 1976, or von Winterfeldt & Edwards, 1986, for details) or a list of attributes potentially relevant to the purpose of the value elicitation from each elicitee, or from face-to-face groups selected to represent classes of elicitees. If possible, all elicitees should come to agree on the structure and labels (not ranking or weights) of relevant attributes. An approach to obtaining agreement is to include all suggested attributes in an initial structure and then hold a group session that eliminates duplicates and inappropriate proposed attributes, relabels ambiguously labeled attributes, restructures to correct partial overlaps, etc. Try to avoid having too many attributes. If you have 12 or more, try to reduce the number (e.g., by combining related attributes; by redefining too-specific attributes; by omitting unimportant attributes that, if retained, would receive low weight). This advice pertains to attributes that are actually scored (von Winterfeldt & Edwards, 1986, calls them twigs; a more orthodox term is leaves). In a value tree, higher-order attributes that are not directly scored help define and explicate those that are and are useful in sensitivity analyses. Use as many of these as needed to make the value tree make sense.

Step 3: Objects of evaluation. If the purpose of the elicitation did not specify the objects of evaluation, use the attribute structure from Step 2 to invent some. As Keeney (1992) has pointed out, values define options. Options, or outcomes of taking them, are normally the objects of evaluation. The output of Step 3 should be either a full list of objects of evaluation, or a real or hypothetical sample of such objects at least as large as the proposed number of attributes. In contexts such as competitions that use multiattribute utilities as scores, the scoring rules must be well defined before any entries are submitted; if so, only hypothetical entries can be used in this step. In preparing hypothetical objects of evaluation, try to anticipate the full range of scores you will later encounter for each attribute; a range that is a bit too wide is preferable to one that is substantially too narrow—though a too-narrow range is not a disaster. In most other contexts the objects of evaluation, and so the ranges of scores, are knowable in advance.

Step 4: Objects-by-attributes matrix. Formulate a matrix of objects of evaluation by attributes (like Table 1 of this paper). Its entries should be scores, physical value-related measures, if available. If scores are not available, its entries can be judged single-dimension utilities.

Step 5: Dominated options. Eliminate ordinally dominated options. Ordinal dominance can usually be recognized by visual inspection (see e.g., von Winterfeldt & Edwards, 1986, pp. 388–399). If you happen to notice one or more cardinally dominated options, eliminate them also; this further reduces the total number of options but is unlikely to affect the range of any attribute. Check to make sure that elimination of a dominated option has not substantially reduced any attribute ranges (by raising the lowest available value). If elimination of an option has substantially reduced a range, consider whether the attribute is still worth using. If not, return to Step 2 to eliminate the attribute.[1]

[1] Elimination of dominated options is not necessary; they fall out in the course of subsequent analysis if not eliminated at this point. But reconsideration of ranges as a result of such eliminations can be useful. Ranges can shrink to zero, or to values so near zero that the attribute is no longer worth considering. This is by no means guaranteed, but happens often enough to justify at least elimination of ordinally dominated options.

Step 6: Single-dimension utilities. Reformulate the entries of the objects-by-attributes matrix as single-dimensional utilities. To do so, first test the linearity of single-dimension utilities for each dimension for which physical scores are available. If use of linearity as an approximation is justified, use the ranges of the scores, or a larger range if the actual range seems too small and the full list of objects of evaluation is not available, to specify upper and lower bounds for single-dimension utility functions. Calculate single-dimensional utilities from linear equations for these functions, or draw them as graphs and read off the points. If a linear approximation is usable, this is a purely computational step. If scores are available but the test for linearity fails, you can use any of the single-dimension utility elicitation methods spelled out in von Winterfeldt & Edwards (1986).

If no physical measure relevant to the attribute is available, this step (or its equivalent in Step 4) requires elicitation. Elicitees may be those who will judge weights (at Steps 7 and 8) or may be individuals to whom the weighters are willing to delegate the responsibility for providing single-dimension utilities. (An example is a clothing manufacturer considering which items of apparel to include in next year's line. She might delegate assessment of how well an item conforms to current imperatives of fashion to an expert in her employ and might delegate assessment of the marketability of that item to another expert. But she might retain for herself the weighting task of judging how important relative to each other fashion and ease of marketing are in choice among items. The latter task seems to us to be the essence of value judgments and so the essence of multiattribute utility measurement.)

At the end of this Step, all needed single-dimension utilities (except for directly judged utilities for objects of evaluation not yet available) should be known.

The final task in this Step is to test for conditional monotonicity (see the technical discussion in this paper). If it is present, an additive model should be an acceptable approximation. If not, nonadditive models explained in Keeney and Raiffa (1976) and von Winterfeldt and Edwards (1986) can be used. What follows assumes an additive model. It also assumes either that linearity of single-dimensional utilities is acceptable as an approximation or that single-dimension utilities have been directly elicited.

Step 7. Do Part 1 of swing weighting. Elicitation methods are described below.

Step 8. Do Part 2 of swing weighting; elicitation methods are described below. Calculate all multiattribute utilities.

Step 9. Decide.

SMARTER

Steps 1–7 and 9 of SMARTER are identical with those same steps in SMARTS. Step 8 is: use Table 2 or Eq. (2) directly to calculate weights. Calculate all multiattribute utilities.

Basic Ideas underlying SMARTS and SMARTER

We next review the two key ideas underlying SMARTS and SMARTER.

Multiattribute utility. Howard Raiffa presented the fundamental insight underlying multiattribute utility in 1968 and expanded on it a very influential Technical Report in 1969. That insight is that if anything is valued at all, it is valued for more than one reason. That is, any outcome of a decision is most naturally described by a vector of numbers that relate to value. The task facing the analyst who wishes to use those numbers to guide decisions is to aggregate that vector into a scalar that the decision maker wishes to maximize—a single number measured at least on an interval scale. The definitive exposition of formally justified procedures for doing this appears in Keeney and Raiffa's (1976) book.

The theoretical literature on utility makes a major distinction between values, appropriate to decision making in riskless situations, and utilities, appropriate to decision making in contexts involving risk. We consider that distinction spurious and so ignore it in this paper. The issues are examined in detail in von Winterfeldt and Edwards (1986, see especially pp. 211–215) and need not be reexamined here.

The strategy of heroic approximation. Two beliefs motivated SMART and motivate SMARTS, SMARTER, and this paper. One is that simpler tools are easier to use and so more likely to be useful. The second is that the key to appropriate selection of methods is concern about the trade-off between modeling error and elicitation error. Edwards originally invented SMART in part because the judgments of indifference between pairs of hypothetical options required by Keeney and Raiffa (1976) seemed difficult and unstable. He believed and we believe that more nearly direct assessments of the desired quantities are easier and less likely to produce elicitation errors. See Edwards, von Winterfeldt, and Moody (1988) for an earlier presentation of the same view.

We call that view the strategy of heroic approximation. Users of that strategy do not identify formally justifiable judgments and then figure out how to elicit them. Rather they identify the simplest possible judgments that have any hope of meeting the underlying requirements of multiattribute utility measurement, and try to determine whether they will lead to substantially suboptimal choices *in the problem at hand.* If not, they try to avoid elicitation errors by using those methods.

Whenever possible, we like to provide checks on the potential for error in the methods we propose. Sensitivity analysis is, of course, the most general of these. But it is not simple. This paper offers rules of thumb about when *not* to use the methods we propose because the potential for errors of significant size is unacceptably large. We believe, but have not proved, that when these rules of thumb are satisfied, the potential for error is small.

SMARTS uses the strategy of heroic approximation to justify linear approximations of single-dimensional utility functions and use of an additive aggregation model. For each, we offer a rule of thumb about when not to use the approximation. SMARTER adds a third use: justification of rank weights. We have not found a rule of thumb about when not to use rank weights, but speculate that the potential for error from using the version of them presented here is never large. A sensitivity analysis tool for assessing error potential is described in the Appendix.

Technical Issues

What follows is a series of technical discussions linked to specific steps of SMARTS, SMARTER, or both. Among other things, these discussions spell out in detail the procedures we advocate for the more technical Steps.

Single-dimension utilities (Links to Step 6). Step 4 in SMARTS or SMARTER is to list some or all of the objects of evaluation along with their scores on physical or judgmental measures related to their values or utilities. A convenient structure for doing so is a rectangular matrix like Table 1. At Step 4 these scores do not need to be (though they are allowed to be) single-dimensional cardinal utilities. They only need to be numbers such that a higher number is preferable to a lower one, in a value or utility sense, i.e., ordinal utilities.

Step 6 consists of rewriting the scores table that is the output of Step 4 so that its entries are single-dimension cardinal utilities, not physical scores. A single-dimension cardinal utility is an interval-scale (or better) measure of the value or desirability of an outcome to a decision maker. The difference between it and an ordinal utility is that, on an interval scale of value or utility, numerically equal differences in magnitude represent equal differences in value or utility. In this paper, the unmodified words utility or value always refer to a cardinal, not ordinal, quantity. A single-dimension utility need not be a function of any physical or judged quantity, but often it is. Such a function relates the utility or value or desirability of some physical or judged quantity, $u(x)$, to its magnitude, x.

Elicitation of the details of utility functions can be tedious and demanding. The contribution of those details to wiser or more valuable choices isoften negligible. Invoking the strategy of heroic approximation, we therefore examine the obvious approach to ignoring them: treat utility functions as linear in x.

From this point of view, four cases often arise; three utility functions and one more case in which utilities are assessed judgmentally without specification of a physical variable. Figure 1 displays all four.

Consider the task of choosing which new car to buy. Assume that you have carefully examined the options and your preferences, and that you have reduced the set of possible purchases to a limited number worth considering. The cars you are considering differ in engine power, record of frequency of trips to the shop for this make of car in past years, amount of crushable steel, and styling; you have determined that only these four dimensions are value-relevant. They do not differ in price.

We chose these dimensions to illustrate the four cases shown in Fig. 1. For you, more engine power is better than less throughout the range of

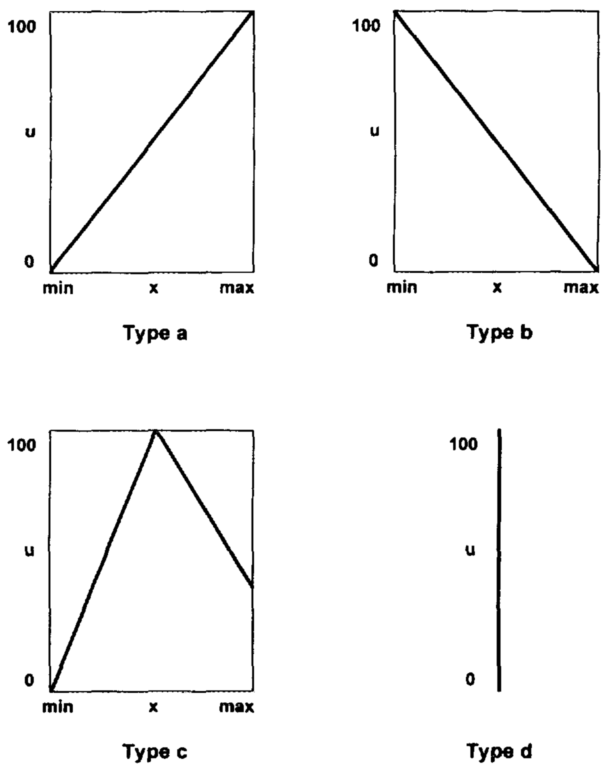

FIG. 1. Four classes of single-dimensional utility functions. Type a are functions in which more of x is better than less. Type b are functions in which less of x is better than more. Type c are functions containing an interior maximum. Type d are direct judgmental utilities for which no underlying single physical variable exists.

engines available. So your utility function is of Type a. Fewer trips to the shop are better than more (Type b). Crushable steel has an internal maximum in your value system (Type c); too little of it heightens your chances of being injured if you have an accident, but too much leads to an oversized, sluggish, hard-to-control car. Finally, styling, though linked to many physical variables, is best thought of for your purposes as a direct judgment of preference rather than as a function of some physical measure (Type d).

An imaginable fifth case would have an internal minimum rather than an internal maximum; we have never encountered an example and so do not discuss the possibility. We ignore logical possibilities such as multiple maxima or minima and utility functions with gaps in them for the same reason.

When the linear approximation is usable, the task of eliciting single-dimensional utilities for utility functions of Types a and b reduces to assessing two extreme values of x, its maximum and its minimum in the context at hand. This is trivial for contexts in which the objects of evaluation can be exhaustively listed in advance. Even if they cannot be, it makes little difference; any choice of two values will do, provided the user recognizes that too small a range can lead to utilities above 100 or below 0. In the case of Type c functions, the extreme values of x must be supplemented by the best value of x and by judgments that specify which branch of the function reaches 0 utility and by how much the other branch does not. For Type d functions, single-dimension utilities must be directly assessed for each object of evaluation. The overwhelming preponderance of instances will be of Types a, b, or d; type c is rare.

When are linear approximations inappropriate? A first thought would be that they are inappropriate if the utility function is non-monotonic. But functions of type c are exactly such a case, and the approach of using two (or more) lines will work well whenever the location of the internal maximum is known or easy to elicit.

The greater the curvature of the function, the less satisfactory will be a linear approximation. Consider a function whose slope changes smoothly from highly positive to 0. This and its mirror image are the two worst imaginable cases, given the constraints of strict monotonicity and restricted range. They can be pretty bad; it is quite possible to produce a discrepancy of 50 (on the arbitrary 0–100 single-dimension utility scale) between the approximation and the "true" function.

Fortunately, human judgment can be used fairly easily to check the adequacy of the approximation. For example, the following line of questioning could be used to check whether it is appropriate to use a linear approximation for the utility of engine power. "Think about small improvements in power at various points within its range. Specifically, think about a fixed improvement of 10 horsepower. Would that improvement be more appealing to you if it fell near the bottom of the scale, or in the middle, or near the top? Or doesn't it matter?" If it doesn't matter, the linear approximation is acceptable. Suppose the respondent prefers the 10 hp improvement near the bottom of the scale. "Now, where does the 10 hp improvement help least?" Suppose the respondent finds it least helpful at the top of the scale. "In a ratio sense, how much more desirable is the improvement at the bottom than the improvement at the top?" That ratio, of course, is the ratio of the greatest to the least slope of the function, and so is an index of the amount of curvature of the function. As a rule of thumb, if the ratio is greater than 2:1, the linear approximation should not be used. That rule of thumb is both crude and conservative; a more sophisticated rule of thumb would take weights and option structure into account—and so would violate the strategy of heroic approximation. Since a continuous utility function with one or more inflection points is better approximated by a straight line than an uninflected one, this test is conservative for inflected functions.

When executing the elicitation sketched in the previous paragraph, the elicitor should be attentive to the possibility that the respondent is inappropriately paying attention to end-points of the utility scale, which are arbitrary. "Anything is better than nothing" or "90% seems very good; why sweat the last 10%?" are comments that, if encountered or elicited, would suggest this misconception. To remedy it, the elicitor should remind the respondent that the end-points of the scale were accidental and that different end-points could as appropriately have been used, and would have been, if the options had been different.

If the linear approximation is not usable, the elicitor can fall back on the well-known methods for single-dimension utility elicitation (see von Winterfeldt & Edwards, 1986, Chap. 7, for methods that do not depend on indifference judgments among hypothetical bets).

The additive model (Links to Steps 7, 8, and 9). Supposing that we know $u(x)$ for each relevant value dimension, we must determine how to aggregate the vector of $u(x)$ values into a scalar in order to carry out Raiffa's program. By far the easiest to use and most familiar model for such aggregations is the additive one. If h ($h = 1, 2, \ldots, H$) is an index identifying the objects of evaluation (cars, in our example) and k ($k = 1, 2, \ldots, K$) is an index of the value dimensions, then this model says that

$$U_h = \sum_{k=1}^{K} w_k u_h(x_{hk}). \tag{1}$$

In Eq. (1), the values of u_h (X_{hk}) are the single-dimensional utilities discussed above. The W_k are the weights, one for each value dimension; by convention, they sum to 1.

Additive models may be good approximations, though not precisely correct. Or they may be lousy even as approximations. Fortunately, an easy-to-use test will weed out almost all instances in which an additive model would be really bad. It consists of looking for instances in which, at one level of value attribute x, more of y is better than less, while at another level of x, less of y is better than more. For example, an automatic transmission may appeal to you more than a manual one for city driving in traffic, but may be highly undesirable in a car designed for rough-road or off-road use. If you are considering which car to buy, and your option set includes both luxury sedans and vehicles designed for off-road use, your evaluation of the presence of an automatic transmission in a specific car may depend on which kind of car it is. Such violations of *conditional monotonicity,* usually easy to detect judgmentally, mean that additive models should *not* be used. If the inputs to a multiattribute utility problem are all conditionally monotonic with one another, we have little hesitation about using additive models. Doing so violates some formal rules unless a demanding technical condition called Additive Difference Independence (see von Winterfeldt and Edwards, 1986, Chapters 8 and 9) is satisfied; given conditional monotonicity, we justify such violations as consistent with the strategy of heroic approximation that motivates this paper. Our personal experience has been that violations of conditional monotonicity, though they do occur, are rare.[2]

The definition of conditional monotonicity for utility functions of type c is only slightly more complicated than for the others. The peak in a type c function should not change its location as a result of changes in the values of other dimensions. Again, this property is easy to test judgmentally.

What was wrong with SMART? (Links to Steps 7 and 8). The values of the weights given in Eq. (1) are related to the values of the single-dimension utilities. To see the point, note that halving each value of u_h (X_{hk}) for some specific value dimension k can be compensated for by doubling the weight for that k and then renormalizing the weight vector; the new utilities are identical to the old ones. Weights reflect the range of the attribute being weighted as well as its importance.

[2] Keeney has said "If additive independence is violated, you probably do not have the appropriate set of fundamental objectives. The reverse is just as important and as accurate. If you do have an appropriate set of fundamental objectives for the context of a decision, additive independence is probably a very reasonable assumption." (Keeney, 1992, p. 167) We agree. Our example of the interaction between whether or not a vehicle is designed for off-road use and whether or not it has manual transmission illustrates Keeney's point. Evaluation of the vehicle for these quite different uses should probably be done separately for each class of use, and then combined by means of some weighting function that reflects probability and importance of performance in each such class.

To obtain weights, Edwards (1977) exploited the great intuitiveness of the notion of importance and the natural and correct idea that in an additive model weights convey the importance of one dimension relative to the others. The procedure was simple. Respondents were asked to judge the ratio of importance of each attribute to all others; such judgments can easily be turned into a set of normalized weights.

But the procedure ignores the fact that range as well as importance must be reflected in any weight. (More specifically, weights must be proportional to a measure of spread times a measure of importance.) For example, in car buying, the price of the car is usually important. But would it still be important if the prices of all cars being considered ranged from $15,000 to $15,100? Obviously the degree of importance of an attribute depends on its spread; that dependence was ignored in SMART weight elicitation. This error is the reason why SMART is not intellectually acceptable. Specifying the range firmly and being careful not to change it, as Edwards and Newman (1982) recommend, does not avoid the intellectual error, though it may, if weight judgments are made appropriately, help prevent it from leading to inappropriate choices.

Swing weights[3] *(Links to Steps 7 and 8)*. Swing weighting does avoid the intellectual error. The word "swing" refers to the operation of changing the score of some object of evaluation on some dimension from one value to a different one (typically from 0 to 100). Suppose, in the car evaluation example, that you have specified exactly four cars that you want to evaluate and have obtained satisfactory single-dimension utilities. The result is presented in Table 1. Casual inspection shows that a 0 and a 100 appear in each column, and so that the full range of each value dimension is used. This property, while not necessary, is pleasant to have. Less casual inspection shows that no option is ordinally or cardinally dominated. Consequently no additional analyses not involving weights can simplify the choice problem, e.g., by making an attribute irrelevant.

Swing weight elicitation proceeds in two Steps. Step 7 yields the rank order of the weights; Step 8 yields the weights themselves.

For Step 7, ask your respondent the following kind of question. "Imagine that there was yet another kind of car, call it the Nometer, and that you were for some strange reason required to buy it. Unfortunately, the

[3] An analyst at Decisions and Designs, Inc., in the 1970's, aware of the ranges-are-weights problem, invented swing weights, but we don't know who it was. Edwards learned about swing weights from Ann Martin (personal communication) and incorporated them in the von Winterfeldt–Edwards book. Many users of SMARTS now cite that source for swing weights. We know of no earlier publication to cite; but neither Martin nor von Winterfeldt nor Edwards invented them.

TABLE 1
SINGLE-DIMENSIONAL UTILITIES FOR THE CAR PURCHASE EXAMPLE

	Value dimensions				
Cars	Power	Shop trips	Crushable steel	Styling	Agg. util.
Anapest	100	90	0	0	76.45
Dactyl	0	100	90	70	44.58
Iamb	70	40	100	40	64.37
Trochee	50	0	40	100	38.12

Note. The entries in the Table are utilities, not physical measures. For all value dimensions, 100 is best and 0 is worst.

Nometer scores 0 on all four dimensions; it is the worst possible car. However, the somewhat kindly deity who makes the rules will allow you to improve just one of the dimensions from its worst value to its best. Which dimension would you choose to improve?'' Suppose the respondent chooses to improve Power. ''Next, imagine that you are stuck with the worst possible car and allowed to improve any dimension *except* Power from its worst value to its best. Which would it be?'' Continue until all dimensions are rank ordered in terms of attractiveness of the 0–100 swing. This completes Step 7. In our example, we shall suppose the ranking was Power, Shop Trips, Crushability, Styling.

Step 8 builds on the ranking obtained at Step 7. It has several presumably equivalent variants.

To elicit swing weights via direct magnitude estimates, one might ask ''Let's call the weight of Power the most important dimension, 100. That is, a swing from 0 to 100 is worth a full 100 points to you. Let's call the weight of some dimension you really don't care about, say size of the ashtray, 0. A 100-point swing on that dimension doesn't matter. Now, on that scale, what is the weight of a 100-point swing on the second most important dimension, Trips to the Shop?'' A similar question can be asked for each dimension. The four resulting judgments, normalized, are the weights.

An alternative approach uses indifference judgments. ''Consider car Nometer, with Styling improved from 0 to 100. Now, you would presumably be indifferent between that one, which we might call Stylish Nometer, and another version of Nometer in which Crushable Steel is somewhat improved with all other dimensions at their worst. But Crushable Nometer presumably need not have Crushability improved to 100 to be exactly as attractive as Stylish Nometer, since you assessed Crushability as more important than Style. For what Crushability utility would you be indifferent between Crushable Nometer and Stylish Nometer?''

This judgment is a direct assessment of the ratio of the weights of

Crushability and Stylishness. Since the other dimensions are set at 0 utility, Eq. (1) says that $u_{Stylish\ Nometer} = u_{Crushable\ Nometer} = 100\,w_4 = Sw_3$, where S is the amount of the swing in Crushability required to equal in attractiveness a 100-point swing in Stylishness. Consequently, $w_3/w_4 = 100/S$.

One could elicit other weight ratios similarly by assessing the amount of swing in each dimension that is as attractive as a 100-point swing in Stylishness. The three weight ratios thus elicited might be called $R(1/4)$, $R(2/4)$, and $R(3/4)$. (Note that we put the weight of the more important dimension over the weight of the less important one; these numbers are therefore all greater than 1.) Since we know (by convention) that the sum of the four weights is 1, we solve for them as follows: $R(1/4) + R(2/4) + R(3/4) = (1 - w_4)/w_4$. Therefore $w_4 = 1/[1 + R(1/4) + R(2/4) + R(3/4)]$. Given w_4, the three ratios give the other three weights.

Using the weight of the least important dimension as the standard may be insecure, since that weight is the smallest of those considered. All possible weight ratios can be recovered by the general procedure described. Any sufficient set can in principle be deciphered into actual weights. For example, suppose $R(1/2)$, $R(2/3)$, and $R(3/4)$ are directly elicited, thus asking the respondent to compare each weight only with its next neighbors in size. Obviously $R(1/2) \times R(2/3) \times R(3/4) = R(1/4)$ and $R(2/3) \times R(3/4) = R(2/4)$. Solution of the system can now proceed as in the previous paragraph. More generally, the redundancy of the information contained in a set of assessed weight ratios permits on-line evaluation and correction of elicitations.

Most of our respondents prefer and have more trust in the result of procedures based on magnitude estimates than those based on indifference judgments; that is why we presented magnitude estimates first. A guess about the reason is that the judgmental task, though in a sense more demanding (assessment of a number with an abstract meaning rather than assessment of a number that makes two options indifferent), is easier both to explain and to do. The result of both judgmental procedures should be the same.

Rank weights (*Links to Step 8 for SMARTER*). Most of the useful numerical information obtained in swing weighting is obtained in Step 7, not Step 8. And Step 7 calls for far easier judgments from the respondent and so is much quicker than Step 8, especially if the elicitor does not wish to use magnitude estimation. Stillwell, Seaver, and Edwards (1981), aware of the literature on equal weights (e.g., Dawes & Corrigan, 1974; Einhorn & Hogarth, 1975; Wainer, 1976) proposed rank weights, which represent preferences better than do equal weights and do not require Step 8. They offered three ways of translating ranks into weights; none had any rationale beyond preserving rankings. Stillwell, Seaver, and Edwards considered all three equally *ad hoc*, but Doyle, Green, and Cook (under review) argue that rank sum weights more closely resemble weights directly elicited from decision makers than do the weights produced by other rank weighting procedures, including ROC weights.

Barron and Barrett's (under review) development of a formally justifiable solution to the task of turning rankings of weights into weights, and even more their demonstration of the quality of the result, is the reason for defining SMARTER and writing this paper. They call their weights Rank Order Centroid, or ROC, weights. The notation of this paper is identical with theirs except that they call the number of attributes n, while we call it K.

The key ideas of the Barron–Barrett derivation are quite simple. If nothing were known about the weights except their sum, set at 1 by convention, then the set of possible non-negative weight vectors would be any that have that sum. If you had no prior reason to prefer one weight vector to another, it would be natural (and error-minimizing) to use equal weights. The point describing equal weights in the hypersurface (simplex) of all possible weights is its centroid.

All that knowing the rank order of weights does to the argument of the preceding paragraph is to change the geometric description of the set of acceptable weights—the simplex. It is straightforward to specify the corner points of the smaller simplex consistent with knowing the ranks, and from them to specify its centroid. Moreover, the equations for the weights have a convenient computational form. If $w_1 \geq w_2 \geq \ldots \geq w_k$, then

$$\begin{aligned}
w_1 &= (1 + 1/2 + 1/3 + \ldots + 1/K)/K \\
w_2 &= (0 + 1/2 + 1/3 + \ldots + 1/K)/K \\
w_3 &= (0 + 0 \quad + 1/3 + \ldots + 1/K)/K \\
w_K &= (0 + \ldots + 0 \quad + 1/K)/K
\end{aligned}$$

More generally, if K is the number of attributes, then the weight of the kth attribute is:

$$w_k = (1/K) \sum_{i=k}^{K} (1/i). \tag{2}$$

Table 2 contains weights calculated from Eq. (2) for values of K from 2 to 16. Partial rank order information (e.g. tied ranks, missing ranks) can be handled, though the computational formulas are less pretty. Barron and Barrett treat such cases, drawing their methods from Kmietowicz and Pearman (1984).

Barron and Barrett checked the error-producing capabilities of ROC

TABLE 2
ROC WEIGHTS FOR INDICATED NUMBER OF ATTRIBUTES

	Number of attributes							
Rank	9	8	7	6	5	4	3	2
1	.3143	.3397	.3704	.4083	.4567	.5208	.6111	.7500
2	.2032	.2147	.2276	.2417	.2567	.2708	.2778	.2500
3	.1477	.1522	.1561	.1583	.1567	.1458	.1111	
4	.1106	.1106	.1085	.1028	.0900	.0625		
5	.0828	.0793	.0728	.0611	.0400			
6	.0606	.0543	.0442	.0278				
7	.0421	.0335	.0204					
8	.0262	.0156						
9	.0123							

Rank	16	15	14	13	12	11	10
1	.2113	.2212	.2323	.2446	.2586	.2745	.2929
2	.1488	.1545	.1608	.1677	.1753	.1836	.1929
3	.1175	.1212	.1251	.1292	.1336	.1382	.1429
4	.0967	.0990	.1013	.1036	.1058	.1079	.1096
5	.0811	.0823	.0834	.0844	.0850	.0851	.0846
6	.0686	.0690	.0692	.0690	.0683	.0670	.0646
7	.0582	.0579	.0573	.0562	.0544	.0518	.0479
8	.0492	.0484	.0471	.0452	.0425	.0388	.0336
9	.0414	.0400	.0381	.0356	.0321	.0275	.0211
10	.0345	.0326	.0302	.0270	.0299	.0174	.0100
11	.0282	.0260	.0230	.0193	.0145	.0083	
12	.0226	.0199	.0165	.0123	.0069		
13	.0173	.0143	.0106	.0059			
14	.0125	.0092	.0051				
15	.0081	.0044					
16	.0039						

weights in extensive simulations; Barron has extended these results further. ROC weights lead to the identification of the best option (defined by assuming SMARTS weights to be true) 75 to 87% of the time, depending on simulation details. But the important calculation is utility loss. (Utility loss is the ratio of the amount of utility lost by the error to a much bigger swing in utility, utility of the optimal strategy minus utility of a random strategy; see von Winterfeldt & Edwards, 1986, Chap. 11, for details and for an argument that such utility losses should be used to evaluate the costliness of errors.) For all conditions (number of alternatives = 5, 10, 15, 20, 25; number of attributes = 3, 6, 9, 12, 15) studied, Barron and Barrett found average utility losses of less than 2%. In short, when ROC weights don't pick the best option, the one they do pick isn't too bad. That is why we are recommending this procedure for routine use.

What is meant by "isn't too bad?" Consider the particular condition in the Barron–Barrett simulations having the largest average value loss. That largest average value loss for ROC weights is 1.9% (1.9 utility units or utiles, using the 0–100 range for utilities that is conventional in this paper). Actual loss was zero for 86.3% of the trials, because SMARTS and SMARTER picked the same option. We know of no way to recognize such cases, or their opposite, without completing SMARTS, in which case SMARTER makes no sense. The average value loss for trials on which a value loss does occur is 13.9, or just under 14% of the 0–100 range. The underlying distribution is severely skewed; most losses are smaller, but a few are quite large. Next, consider the condition, of those studied by Barron and Barrett, in which SMARTS and SMARTER most frequently disagreed. That most frequent loss condition (loss > 0) occurred on 24.9% of the trials, with an average value loss of 1.4 utiles. This implies an average conditional value loss of 5.6 utiles. A sensitivity analysis applicable to value matrices like Table 1, including identification of maximum loss, is shown in the Appendix.

Srivastava, Beach, and Connolly (in press) conducted an experiment intended to compare SMARTS and SMARTER with other ways of eliciting multiattributed values. The stimuli were hypothetical apartments that varied on nine dimensions. Students judged weights by various methods, and in addition rated each apartment on a 9-point scale of desirability. Weights were also recovered statistically from the holistic judgments. The five weighting procedures yielded weights that intercorrelated highly; the highest correlations were among Swing weights, ROC weights, and rank weights by an older procedure. Test–retest reliability of the holistic judgments of attractiveness were fairly low—.64 to .69. Of the weighting procedures, ROC weights produced multiattribute utilities that correlated most highly with holistic judgments, .75.

A Caveat

We close by underlining a point familiar to decision analysts. The most important goal of decision analysis is insight, not numerical treatment. Elicitation and use of such numbers as multiattribute utilities contributes to emergence of insights in important ways. Those insights sometimes emerge from the kind of thinking required to do Part 2 of swing weights. Some analysts with whom we have discussed SMARTER have expressed reservations about the procedure because, by reducing judgmental labor, it reduces the opportunity to have insights. We do not have enough experience with SMARTER to know whether or not this is a serious deficiency that should affect its use. Our guess is that it depends on decision context. But we fully agree that nothing that can be done with multiattribute utilities after they have been elicited is nearly as valuable as the insights that sometimes emerge during the elicitation.

APPENDIX

Analysis of Specific Values and Ranks

In the example presented in Table 1, the ROC weights implied by Step 1 of Swing weighting are .5208 (power) > .2708 (trips to shop) > .1458 (crushable steel) > .0625 (style). The simplex that defines the set of all possible rank weights for four attributes has the following extreme or corner points:

(1,0,0,0)
(1/2,1/2,0,0)
(1/3,1/3,1/3,0)
(1/4,1/4,1/4,1/4).

From a theorem of linear programming it can be shown that the largest possible error produced by using ROC weights to select a car is the maximum difference in multiattribute value between the alternative chosen by ROC weights and the alternative chosen by the weights that define the extreme point, both evaluated using the weights that define the extreme point. Table 3 shows the multiattribute values at each extreme point and at the ROC weights point for the example of Table 1.

Using ROC weights, the Anapest is best. Using equal weights, the Dactyl is best and the Anapest is worst.

How much change in the ROC weights is needed to make the Anapest and the Dactyl equally attractive? We can answer the question by drawing a straight line connecting the ROC weights point with the equal weights

TABLE 3
MULTIATTRIBUTE VALUES[a] FOR ROC AND FOUR EXTREME POINTS

		Weights			
Alternative	ROC	(1,0,0,0)	(½,½,0,0)	(⅓,⅓,⅓,0)	(¼,¼,¼,¼)
Anapest[b]	76.45	100	95	63.33	47.5
Dactyl	44.58	0	50	63.33	65
Iamb	64.38	70	55	70	62.5
Trochee	38.12	50	25	47.5	47.5
Best	76.45	100	95	70	65
ROC[b]	76.45	100	95	63.33	47.5
Error	0	0	0	6.67	17.5

[a] Single-attribute values are on the 0–100 scale conventional in this paper.
[b] With ROC weights, Anapest is the best vehicle.

point in the simplex of possible weights, and then finding the point on that line at which the two cars are equally attractive.

The ray connecting the ROC weights point with equal weights point is defined by Eq. (3). Any value of α between 0 and 1 specifies a point on that ray; the weights that define that point are specified by Eq. (3). For example, for $\alpha = .4$, the weight of Power is .4 (.5208) + .6 (.25) = .3583.

$$Weights(\alpha) = \alpha \begin{pmatrix} .5208 \\ .2708 \\ .1458 \\ .0625 \end{pmatrix} + (1 - \alpha) \begin{pmatrix} .25 \\ .25 \\ .25 \\ .25 \end{pmatrix}, \ 0 \leq \alpha \leq 1 \tag{3}$$

To find the value of α for which the Anapest and the Dactyl are equally attractive, we need only note that the multiattribute utility of each for any intermediate value of α is the convex combination of their multiattribute utilities at the beginning and end of the ray. We therefore solve the following expression (numbers are multiattribute utilities for Anapest and Dactyl, from Table 3) for α:

$$.7645\,\alpha + .475\,(1 - \alpha) = .4458\,\alpha + .65\,(1 - \alpha).$$

In this example, $\alpha = 0.3545$. From Eq. (3), the weight vector at that value of α is (.3460, .2574, .2131, .1835).

Because Anapest is best at (1, 0, 0, 0) and at (1/2, 1/2, 0, 0), we can repeat the analysis of the previous paragraph twice, substituting each of these weight vectors for the ROC weights. Doing so yields two more weight vectors at which Anapest and Dactyl are equally attractive; they are (.3616, .2128, .2128, .2128) and (.32, .32, .18, .18). By identifying points on the boundaries of the region(s) of the simplex in which Anapest is the winner, these weights give some idea of its size and of the closeness of those boundaries to the point specified by ROC weights. In this example, the weight of the most heavily weighted dimension, power, would have to be very considerably lower than the ROC weights value of .5208 to make the conclusion that Anapest is the best car seem insecure.

Since Iamb is best at (1/3, 1/3, 1/3, 0), we can make a similar calculation comparing it with Anapest. The weight vector on the ray connecting ROC weights with (1/3, 1/3, 1/3, 0) that makes Iamb and Anapest equally attractive is (.4000, .3111, .2666, .0223)—again, comfortably distant from the ROC weights vector.

The extreme points that define exactly the weights for which which the alternative that is optimal using ROC weights is also optimal can be determined by using a program (Fukuda & Mizokoshi, 1992) included within the Mathematica Software Package. The appropriate polytope is

$$w_1 \geq w_2 \geq w_3 \geq w_4 \geq 0$$
$$w_1 + w_2 + w_3 + w_4 = 1$$
$$MAV(A) \geq MAV(D)$$
$$MAV(A) \geq MAV(I)$$
$$MAV(A) \geq MAV(T),$$

where $MAV(A) = 100\,(w_1) + 90\,(w_2) + 0\,(w_3) + 0\,(w_4)$ (numbers are from Table 1) and A, D, I, and T are Anapest, Dactyl, Iamb, and Trochee, respectively. Warning: the software may take several hours to execute!

Any convex combination of these extreme points specifies a set of rank weights for which (in this example) Anapest is best.

REFERENCES

Barron, F. H., & Barrett, B. E. *Decision quality using ranked and partially ranked attribute weights*. Under editorial review.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, **81**, 91–106.

Doyle, J. R., Green, R. H., and Cook, W. D. (under review) A priori rank-into-weight schemes in multi-attribute decision making. *Organizational Behavior and Human Decision Making*.

Edwards, W. (1971). Social utilities. *Engineering Economist*, Summer Symposium Series, **6**, 119–129.

Edwards, W. (1977). How to use multiattribute utility measurement for social decision making. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-7*, 326–340.

Edwards, W., & Newman, J. R. (1982). *Multiattribute evaluation*. Beverly Hills, CA: Sage.

Edwards, W., von Winterfeldt, D., & Moody, D. L. (1988). Simplicity in decision analysis: An example and a discussion. In D. Bell, H. Raiffa, and A. Tversky (Eds.) *Decision Making: Descriptive, normative and prescriptive interactions*. Cambridge, MA: Harvard Business School.

Einhorn, H., & Hogarth, R. M. (1975). Unit weighting schemes for decision making. *Organizational Behavior and Human Performance*, **13**, 171–192.

Fukuda, K., & Mizukoshi, I. (November 14, 1992) Vertex enumeration for convex polytopes and arrangements. (Version 0.14 beta). University of Tsukuba, Tokyo, Japan. (In Mathematica Software Package).

Keeney, R. L. (1992). *Value-focused thinking*. Cambridge, MA: Harvard University Press.

Keeney, R. L., and Raiffa, H. (1976). *Decisions with multiple objectives: Preferences and value tradeoffs*. New York: Wiley.

Kmietowicz, A. W., & Pearman, A. D. (1984). Decision theory, linear partial information and statistical dominance. *Omega*, **12**, 391–399.

Raiffa, H. (1968). *Decision analysis*. Reading, MA: Addison-Wesley.

Raiffa, H. (1969). *Preferences for multiattributed alternatives*. (Report No. RM-5868-DOT/RC). Santa Monica, CA: Rand Corporation.

Stillwell, W. G., Seaver, D. A., & Edwards, W. (1981). A comparison of weight approximation techniques in multiattribute utility decision making. *Organizational Behavior and Human Performance*, **28**, 62–77.

Srivastava, J., Beach, L. R., & Connolly, T. (in press). Do ranks suffice? A comparison of alternative weighting approaches in value elicitation. *Organizational Behavior and Human Decision Processes*.

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

Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin*, **83**, 312–317.

RECEIVED: July 18, 1994