Proceedings of the 2006 Winter Simulation Conference
L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds.

RANKING AND SELECTION WITH MULTIPLE "TARGETS"

Douglas J. Morrice

Department of Information, Risk and
Operations Management
The University of Texas at Austin
Austin, TX 78712-1175, U.S.A.

John C. Butler

Department of Accounting and MIS
The Ohio State University
Columbus, OH 43210, U.S.A.

ABSTRACT

Managers of large industrial projects often measure performance by multiple attributes. In previous work, we developed a multiattribute ranking and selection procedure to allow tradeoffs between conflicting objectives. More recent developments in ranking and selection incorporate the notion of "constraints", or "targets", that must be satisfied. In this paper we demonstrate how some forms of single attribute utility functions can be used to create a target or constraint. We re-analyze our original problem under the assumption that there are unacceptable levels for some attributes.

1 INTRODUCTION

The evaluation of projects typically involves the use of multiple performance measures, e.g. cost vs. quality vs. time. In recent work, Butler et al. (2001) extended traditional single-attribute ranking and selection procedures to multiple attributes by using multiattribute utility theory (MAU) to convert multiple performance measures to a single scalar performance measure. The technique was applied to a real project evaluating configurations for a land seismic survey in geophysical exploration for oil and gas and was viewed as a success by the client company.

Recently, there has been increasing interest in the incorporation of constraints in ranking and selection procedures (e.g. Andradottir et al. 2005; Batur and Kim 2005). The purpose of this paper is to illustrate that the Butler et al. (2001) procedure can be extended to handle constraints in a natural way, including multiple constraints, *when it is appropriate to do so*. While it may be difficult to assess a multiattribute utility function, it is done on a regular basis in both the private and public sector. In fact, enforcing the notion of constraints can simplify assessment compared to a traditional MAU analysis.

The rest of this paper is organized as follows. In Section 2, we briefly summarize multiattribute utility theory, ranking and re-scaling. In Section 3, we address various issues regarding the use of constraints in a multiattribute analysis and apply this to the example from Butler et al. (2001) in Section 4. In Section 5, we conclude and offer some areas for additional research.

2 MULTIATTRIBUTE UTILITY THEORY

2.1 Multiattribute Utility Theory

MAU theory (Keeney and Raiffa 1976) is one of the major analytical tools associated with the field of decision analysis (see, for example, Clemen 1991). The first step of the procedure is to identify the fundamental objectives that determine the decision maker's preferences and structure these objectives into a means-ends hierarchy (Keeney 1992). Then representative measures for each lowest level sub-objective are identified to evaluate the alternatives.

The second step in an MAU analysis is to identify the alternatives and to estimate the performance of each alternative on each performance measure. When these estimates are uncertain, it is often appropriate to quantify them with ranges or with probability distributions determined using risk analysis methods (e.g., Clemen 1991; Keeney and von Winterfeldt 1991), i.e., Monte Carlo simulation.

Next, a single attribute von Neumann and Morgenstern (1947) utility function is assessed for each performance measure that scales performance between 0 and 1, inclusive. Finally, a multiple attribute utility function determines how the performance on each measure affects overall performance vis-à-vis a set of assessed weights, or measures of relative importance. The notion of relative – as opposed to absolute – importance is driven by the range of performance of each attribute. One reason why the utility functions should be assessed prior to the weights is to ensure that the decision maker is aware of these ranges.

There are many varieties of multiattribute utility functions, but the most commonly used forms rely on the notion of utility independence. If we assume that an alternative's performance is represented by two performance measures X_1 and X_2 then we say that attribute X_1 is utility independent of X_2 if preference for lotteries on X_1 given $X_2 = x_2$ (note: we will use the standard convention in the probability and statistics literature of representing a realization of the random variable X_i by lower case notation, i.e., by x_i) does not depend on the level of x_2 (Keeney and Raiffa 1976). The concept of utility independence allows us to consider the utility function for consequences of attribute X_1 independent of X_2. Mutual utility independence holds when X_1 is utility independent of X_2 and X_2 is utility independent of X_1. For the general case when there are n performance measures, i.e. $\boldsymbol{X} = (X_1, X_2, \ldots X_n)$, mutual utility independence holds if for $X_1, X_2, \ldots X_n$, every subset of X_is is utility independent of its complement.

If there are more than two attributes and X_i is utility independent of X_j for all $j \neq i$, then it is appropriate to model the utility of a realization of $\boldsymbol{X}$, $u(\boldsymbol{x})$, using the multilinear model:

$$u(\boldsymbol{x}) = \sum_{i=1}^{n} w_i u_i(x_i) + \sum_{i=1}^{n}\sum_{j>i} w_{ij} u_i(x_i) u_j(x_j) + \sum_{i=1}^{n}\sum_{j>i}\sum_{m>j>i} w_{ijm} u_i(x_i) u_j(x_j) u_m(x_m) + \ldots w_{123\ldots n} u_1(x_1) u_2(x_2) \ldots u_n(x_n) \tag{1}$$

where $u_i(\cdot)$ is a single attribute utility function over measure i that is scaled from 0 to 1, w_i is the weight for measure i where $0 \leq w_i \leq 1$ for all i, and w_{ijm} are scaling constants that represent the impact of the interaction between attributes i, j and m on preferences (see, for example, Keeney and Raiffa 1976, page 293).

If mutual utility independence holds, then the correct choice is the multiplicative MAU model, $1 + wu(\boldsymbol{x}) = \prod_{i=1}^{n}\left[1 + ww_i u_i(x_i)\right]$. If we expand this compact form of the multiplicative MAU we get:

$$u(\boldsymbol{x}) = \sum_{i=1}^{n} w_i u_i(X_i) + \sum_{i=1}^{n}\sum_{j>i} ww_i w_j u_i(x_i) u_j(x_j) + \sum_{i=1}^{n}\sum_{j>i}\sum_{m>j>i} w^2 w_i w_i w_m u_i(x_i) u_j(x_j) u_m(x_m) + \ldots + w^{n-1}\prod_{i=1}^{n} w_i u_i(x_i) \tag{2}$$

where $0 \leq w_i \leq 1$ and $-1 < w < \infty$. Note there is no subscript on the common w, so this multiplicative form is a special case of the multilinear model (1) where the strength of all interactions among criteria is the same. Finally, if a more restrictive preference condition called additive independence is satisfied then it is the marginal, *not the joint*, distribution of each performance measure that determines preference. In this case we can represent the decision-maker's preferences with an additive MAU model:

$$u(\boldsymbol{x}) = \sum_{i=1}^{n} w_i u_i(x_i) \tag{3}$$

where $0 \leq w_i \leq 1$ and $\sum_{i=1}^{n} w_i = 1$. Inspection of (1), (2) and (3) reveals that the first term in both the multilinear and multiplicative models is the additive model. When the preferential interactions have no impact on preferences, i.e., when the interaction terms are all zero, the additive model is a special case of the more general aggregation schemes.

2.2 Re-scaling Multiattribute Utility Functions

The "even swaps" re-scaling procedure described in Hammond et al. (1998) and formalized in Butler et al. (2001) is based on the intuitive notion of exchanging performance in one attribute for another. For example, we might ask a decision maker how many dollars they would pay to change the horsepower of a car from 140 to 200 horsepower. We could continue through all of the K cars under consideration asking how much the decision maker would pay to change the horsepower of each car from its current level to 200 horsepower. Note that cars that have more than 200 horsepower would be have to be "paid", i.e. made cheaper, rather than pay to change to 200 horsepower. If we repeat this swapping for all of the measures used to evaluate the cars, we end up with a new set of hypothetical cars that share common levels on all attributes but cost. Further, these equivalent costs can be used to make the final decision as they reflect the performance of each car on all of the other attributes considered. See Hammond et al. (1998) and Butler et al. (2001) for more detailed examples of the even swaps procedure.

It is important to note that when walking the decision-maker through the swaps procedure that converts the original alternatives into equivalent hypothetical cars there is no need to make any assumptions about the form of the MAU aggregation function. In other words, the swaps procedure holds for all MAU, because the decision-maker uses his or her internal utility function to provide the numbers required. We have outlined the swaps procedure to provide the intuition of the utility re-scaling, but the decision-maker will never be required to make the swaps explicitly.

The first step in creating the hypothetical $K \geq 2$ equally preferred alternatives is to select a measure as the medium for exchange or standard measure (e.g., cost in the previous example). Without loss of generality, let the standard measure be performance measure 1. Next, select a common level of utility c_i for the other criteria i, $2 \leq i \leq n$ (e.g., u(200 HP) = c_i in the previous example). In other words specify x_{ki}' such that $u_i(x_{ki}') = c_i$ for all $i > 1$ and k, $1 \le k \le K$. The final step is to find the level of measure 1, x_{k1}', such that the two alternatives are equally preferred.

As shown in Proposition 1 of Butler et al. (2001), the equation of $u(x_{k1}')$ for the multilinear, multiplicative and additive MAU are all of the following general form:

$$u_1\left(x_{k1}'\right) = \frac{u\left(\boldsymbol{x_k}\right) - \theta_1}{\theta_2} \tag{4}$$

where θ_1 and θ_2 are constants that depend on the specific MAU form and the assessed utility functions and weights. As demonstrated in Butler et al. (2001), making decision based on x_{k1}' is strategically equivalent to making decisions based on $u_1\left(x_{k1}'\right)$.

3 CONSTRAINTS AND MULTIATTRIBUTE ANALYSIS

Butler et al. (2001) was motivated by Gupta and Panchapakesan (1979), page 141: "*... if we consider two bivariate normal populations with mean vectors* $\boldsymbol{\mu}_1 = (\mu_{11},\mu_{12})$ *and* $\boldsymbol{\mu}_2 = (\mu_{21},\mu_{22})$ *and a common covariance matrix* $\boldsymbol{\Sigma}$ *.... In this case we will naturally prefer the first population if* $\mu_{1j} \ge \mu_{2j}$, *j = 1,2. However if* $\mu_{11} > \mu_{21}$ *and* $\mu_{12} < \mu_{22}$, *the two are not comparable. There is practically no result available in this direction.*" One goal of Butler et al. (2001) was to provide a fully compensatory technique to address this issue.

However, there *may* be situations where the decision maker's preferences are non-compensatory: poor performance on one criterion cannot be offset by good performance on another criterion. For example, a decision maker may refuse to spend more than \$20,000 on a car regardless of its performance of the other criteria. Setting such a minimum performance level for a criterion can be thought of as setting a target level, establishing a cutoff, using a screening criteria or setting a constraint.

Before proceeding it is important to note one caveat. As Keeney (2002) suggests, great care should be taken when using constraints to imply value judgments. Suppose that you told a car salesman that you would spend no more than \$20,000 on car. The salesman shows you several cars and you purchase one. The next day you encounter a friend who also has a new car that is much nicer than yours. When you inquire how much it costs you are bothered to learn that it cost \$20,100. When you return to the car lot, the dealer informs you that he didn't show you the car because it exceeded your maximum cost. If your response is "It was so close …" then \$20,000 is not really a constraint! It is the weights in a multiattribute analysis that should capture the relative importance of the criteria, not constraints on the acceptable attribute levels.

3.1 Using Constraints

Andradóttir, Goldsman, and Kim (2005), hereafter *AGK*, develop a procedure for selecting the best system when performance is measured by two criteria: a primary measure and a secondary, constrained, performance measure. We will use X_1 for the primary measure and X_2 for the constrained measure. The *AGK* setup is to find the largest of the K configurations:

$$\begin{aligned} &\text{Argmax}_{k=1,\ldots,K} \mathrm{E}\left[X_{k1}\right] \\ &\qquad \text{s.t. } \mathrm{E}\left[X_{k2}\right] \le Q \end{aligned} \tag{5}$$

where Q is a constraint. Or, using an indifference zone like approach the decision maker is asked to specify a range around the constraint $Q_L \le Q \le Q_U$ which specifies a desirable region ($\mathrm{E}[X_{k2}] \le Q_L$), an acceptable region ($Q_L \le \mathrm{E}[X_{k2}] \le Q_U$) and an unacceptable region ($\mathrm{E}[X_{k2}] > Q_U$). As *AGK* point out, obvious changes in sign can be made if Q is a minimum rather than a maximum and the same is true for our utility based approach.

Implicitly all single attribute utility functions contain a cutoff or minimum level of performance. However, a well specified utility function is defined on a range that spans the entire range of *possible* (i.e. allowable) outcomes. For example, Figure 1 shows an example utility function for costs specified on [\$200,000, \$60,000]. Normally, \$200,000 represents the most costly alternative being evaluated but could be interpreted as the maximum cost the decision maker was willing to pay; all more costly alternatives would have $u(x) = 0$.

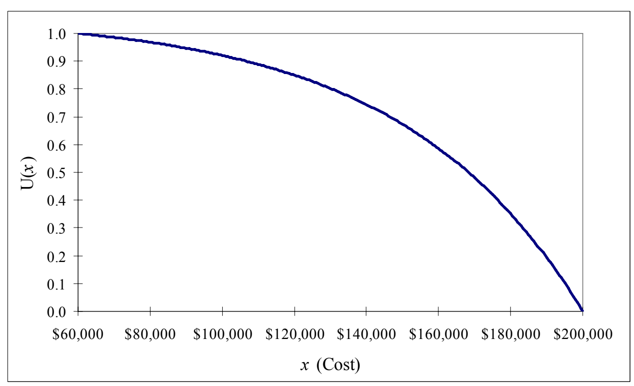
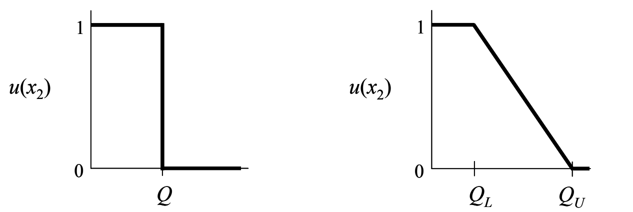


Figure 1: Example Utility Function for Cost

Figure 2 shows two single attribute utility functions that are consistent with the notion of a constraint or an acceptable range for the constrained attribute, X_2. The binary utility function represents an attribute that either does or does not satisfy a constraint. The range utility function is just a special case of a utility function that rewards attribute performance closer to Q_L and could be of any shape.

For a problem with two attributes, the multilinear and multiplicative forms are identical and so either is the most general form of MAU consistent with utility independence

$$u(x_1,x_2) = w_1u_1(x_1) + w_2u_2(x_2) + wu_1(x_1)u_2(x_2). \quad (6)$$



Strict Target or Binary Utility   Acceptable Range Utility

Figure 2: Single Attribute Utility Functions Consistent With Constraint Interpretation

To determine the weights of (6) we first define x_i^o and x_i^* as the least and most preferred levels of attribute i, respectively, where $u_i(x_i^o)=0$ and $u_i(x_i^*) = 1$. We also assume that u is scaled such that $u(x_1^o, x_2^o) = 0$ and $u(x_1^*, x_2^*) = 1$. The weight for attribute 1 can be determined by assessing the probability p that makes the decision maker indifferent between receiving (x_1^*, x_2^o) and a lottery offering a p chance to receive (x_1^*, x_2^*) and a $1–p$ chance to receive (x_1^o, x_2^o). The expression of indifference implies that

$$u(x_1^*, x_2^o) = pu(x_1^*, x_2^*) + (1–p)u(x_1^o, x_2^o). \quad (7)$$

Applying (6) to the left hand side of (7) leads to

$$\begin{aligned} w_1u_1(x_1^*) + w_2u_2(x_2^o) + wu_1(x_1^*)u_2(x_2^o) \\ = pu(x_1^*, x_2^*) + (1–p)u(x_1^o, x_2^o) \end{aligned}$$

$$w_1(1) + w_2(0) + w(1)(0) = p(1) + (1–p)(0)$$

$$w_1 = p.$$

Alternatively we can assess $u(x_1^*, x_2^o) = w_1$ directly. If a decision maker believes that performance on attribute 1 is irrelevant if attribute 2 has not achieved its target level, x_2^o, then she would state that $p = 0$ which implies that $u(x_1^*, x_2^o) = u(x_1^o, x_2^o) = w_1 = 0$. In other words, strong performance on attribute 1 cannot compensate for (very) weak performance on attribute 2. There may be instances where this strong assumption is true, e.g., union rules, legal restrictions, but care should be taken in assuming that it holds in general.

If it is also true that $u(x_1^o, x_2^*) = u(x_1^o, x_2^o) = w_2 = 0$ then (6) reduces to

$$u(x_1,x_2) = u_1(x_1)u_2(x_2). \quad (8)$$

Abbas and Howard (2005) refer to (8) as satisfying attribute dominance utility. Not only can MAU handle "constraints", it also provides insight in to the conditions required for such a preference structure via the assessment of attribute weights.

The logic for the two attribute case can be extended so that multiple constraints can be considered as in Batur and Kim (2005). For three attributes the multilinear model offers some interesting preference models for constraint interpretation

$$u(\mathbf{x}) = u_1(x_1)u_2(x_2)u_3(x_3) \quad (9)$$

$$u(\mathbf{x}) = w_{13}u_1(x_1)u_3(x_3) + w_{23}u_2(x_2)u_3(x_3) \quad (10)$$

$$\begin{aligned} u(\mathbf{x}) = w_{13}u_1(x_1)u_3(x_3) + w_{23}u_2(x_2)u_3(x_3) \\ + w_{123}u_1(x_1)u_2(x_2)u_3(x_3). \end{aligned} \quad (11)$$

The first form (9) is analogous to a series model in electronics where all circuits must work, i.e., the targets must all be satisfied simultaneously, to provide any value to the decision maker. The second form (10) resembles a parallel circuit in that value can be provided through either of two paths as long as the target attribute X_3 is satisfied. Finally (11) is another parallel model but with a "bonus" when all three attributes score well.

4 AN EXAMPLE

4.1 Setup

We use the methodology developed in this paper to analyze the results generated by a simulator of the project described in Mullarkey et al. (2006). The simulator models a large outdoor operation called a land seismic survey. Land seismic surveys generate geophysical information used in oil and gas exploration (Dobrin and Savit 1988). They are conducted over large geographical areas (tens to hundreds of square kilometers). These projects take anywhere from a few days to a few years to complete, utilize from 20 to 1000 people, require capital equipment valued in the tens of millions of dollars, and generate survey revenues ranging from hundreds of thousands to hundreds of millions of dollars. The simulator was designed to support bidding, planning, and conducting these large, complicated, and expensive projects in a profitable manner.

The execution of a land seismic survey requires the coordination of five types of crews (see Figure 3). Briefly, the source crew sends signals (shock waves through the earth) from several geographic locations. The recording crew records reflections of these signals from the earth's subsurface layers. The layout crew places receiving (or monitoring) equipment at several geographic locations so that the recording crew can receive the reflected signals. The transport crew brings the layout crew receiving equipment. The packing crew prepares receiving equipment for the transport crew that is no longer required on a particular part of a survey for receiving signals sent by the source crew. Figure 3 summarizes the sequence above in an operations cycle diagram. The diagram shows that each crew cycles through its own local operations steps (e.g., the transport crew picks-up and drops off equipment) depicted by small loops on each crew. Additionally, each crew forms a step in the larger operations cycle that represents the progression of the land seismic survey. For more details on land seismic survey operations, see Mullarkey et al. (2006).

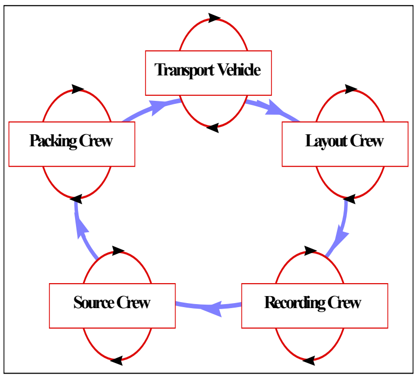


Figure 3: Crews in a Land Seismic Survey

During an actual land seismic survey, a project manager will monitor multiple performance measures. These include project cost, project duration, and utilization for all types of crews. Project cost represents the bottom line and is considered the most important performance measure. Project duration, which is positively correlated with project cost since variable costs such as labor increase with the duration of the job, is included because certain things such as reputation for finishing the job in a timely manner may be difficult to price.

The crew utilization is monitored to ensure that crews are not over worked. Crews work under very adverse conditions on many land seismic surveys. Overutilization of crews can lead to worker dissatisfaction, poor quality work, attrition and unsafe working conditions. Again, things such as worker dissatisfaction and work quality may be difficult to cost. Therefore, these measures are monitored in addition to cost.

To mimic reality, our simulator generates statistics on cost, duration, and all crew utilizations. We use the simulator to compare different project configurations (e.g., different survey designs or different levels of resources). Since the evaluation of different project configurations must be based on multiple, possibly correlated, measures we needed to develop a methodology to solve this problem. For more details on the simulator, see Mullarkey et al. (2006).

4.2 Multiattribute Utility Function

Butler et al. (2001) analyzed the output of the simulator using an additive model, and assessed utility functions and weights from company representatives. To demonstrate the use of targets, for this example we will assume that the crew unions re-negotiated their contracts and utilizations over 85% are now forbidden. Cost is still relatively more important than the project's duration, but these measures only factor in to preference if crew utilizations are all less than 85%.

As in Butler et al. (2001) Cost will be the measure of exchange and represented by x_1; project duration will be represented by x_2; and x_3, x_4, x_5, and x_6 will represent utilization percentages for source, layout, transport, and packing crews, respectively. Since source and recording crews work in such a synchronized fashion, we treat them as one entity in the simulation model. We will use the same assessed utility functions for Cost and Project Duration as Butler et al. (2001):

$$u_1(x_1) = 1.064 - (0.0195)\,e^{(x_1/50000)}, \text{ and}$$

$$u_2(x_2) = 1.021 - (0.00106)\,e^{(x_2/80)}$$

The crew utilization functions will be of the target form shown in the left panel of Figure 2 with $Q = 0.85$:

$$u_i(x_i) = \begin{cases} 1, x_i \le 0.85 \\ 0, \text{otherwise} \end{cases} \quad \text{for } i = 3, 4, 5, 6.$$

It would be straightforward to specify an indifferent zone and create an acceptable range as in the right panel of Figure 2, e.g., set $Q_L = 0.82$ and $Q_U = 0.88$ as we discuss in the conclusion.

The assumptions of the additive model are not satisfied because the joint distribution of attribute utilities factor into the decision maker's preferences. To see this we can assess the weights for the crew utilization measures. As discussed previously, the weight on x_3 can be assessed by asking the decision maker to specify $u(x_1^o, x_2^o, x_3^*, x_4^o, x_5^o, x_6^o)$. This quantity would have a utility of 0; all of the crews must be below 85% utilization, or the project configuration has no value to the decision maker. In fact, all combinations of utilization percentages other than $u(\cdot,\cdot, x_3^*, x_4^*, x_5^*, x_6^*)$ have utility of 0, which drastically simplifies the assessment of the multilinear model to the specification of three quantities: $u(x_1^o, x_2^o, x_3^*, x_4^*, x_5^*, x_6^*)$, $u(x_1^*, x_2^o, x_3^*, x_4^*, x_5^*, x_6^*)$ and $u(x_1^o, x_2^*, x_3^*, x_4^*, x_5^*, x_6^*)$; by assumption $u(x_1^*, x_2^*, x_3^*, x_4^*, x_5^*, x_6^*) = 1$.

Assuming that meeting the crew utilization targets is necessary but not sufficient to provide value, then $u(x_1^o,$

$x_2^o, x_3^*, x_4^*, x_5^*, x_6^*) = 0$. It may be easier to think about the utilities of the other configurations relative to the best possible case. For example, it may be that

$$u(x_1^*, x_2^o, x_3^*, x_4^*, x_5^*, x_6^*) = 2/3\, u(x_1^*, x_2^*, x_3^*, x_4^*, x_5^*, x_6^*) \text{ and}$$

$$u(x_1^o, x_2^*, x_3^*, x_4^*, x_5^*, x_6^*) = 1/3\, u(x_1^*, x_2^*, x_3^*, x_4^*, x_5^*, x_6^*),$$

which is consistent with the relative importance of Cost and Duration in Butler et al. (2001). Using the relations defined in Keeney and Raiffa (1976), page 293, the final form of the multilinear model is as follows:

$$u(\boldsymbol{x}) = 2/3u_1(x_1)u_3(x_3)u_4(x_4)u_5(x_5)u_6(x_6) + 1/3u_2(x_2)u_3(x_3)u_4(x_4)u_5(x_5)u_6(x_6)$$

$$= [2/3u_1(x_1) + 1/3u_2(x_2)]\, u_3(x_3)u_4(x_4)u_5(x_5)u_6(x_6).$$

It may be more intuitive to imagine using an additive utility function to evaluate the configurations that satisfy the union requirements:

$$u(\boldsymbol{x}) = \begin{cases} 2/3u_1(x_1) + 1/3u_2(x_2), \text{if } x_i \le 0.85, 3 \le i \le 6 \\ 0, \quad \text{otherwise} \end{cases}$$

The additive portion of the utility function can be re-scaled as in Butler et al. (2001). If we set $u_2(x'_2) = c_2$ then we seek x'_1 such that $2/3u_1(x'_1) + 1/3c_2 = 2/3u_1(x_1) + 1/3u_2(x_2)$. The rescaling is provided in (12)

$$u(\boldsymbol{x}) = u_1(x'_1) = \begin{cases} u_1(x_1) + 1/2(u_2(x_2) - c_2), \text{if } x_i \le 0.85, 3 \le i \le 6 \\ 0, \quad \text{otherwise} \end{cases} \tag{12}$$

4.3 Ranking and Selection Procedure

Assume that there are $K \ge 2$ alternatives and let $\boldsymbol{X}_k = (X_{k1}, X_{k2}, \ldots, X_{kn})$ denote a vector of random variables representing the performance measures for configuration k. Let $\mathrm{E}[u_1(X'_{k1})]$ denote the expected exchange utility for configuration k. From (4)

$$\mathrm{E}\left[u_1\left(X'_{k1}\right)\right] = \mathrm{E}\left[\frac{u\left(\boldsymbol{X}_k\right) - \theta_1}{\theta_2}\right].$$

Let

$$\mathrm{E}[u_1(X'_{[1]1})] \le \mathrm{E}[u_1(X'_{[2]1})] \le \ldots \le \mathrm{E}[u_1(X'_{[K]1})] \tag{13}$$

denote the ordered exchange values, i.e.,

$$[K] = \mathrm{Argmax}_{k=1,\ldots,K}\, \mathrm{E}\left[\frac{u\left(\boldsymbol{X}_k\right) - \theta_1}{\theta_2}\right].$$

The goal is to select the project configuration with the largest expected exchange utility $\mathrm{E}[u_1(X'_{[K]1})]$. If the R&S procedure accurately identifies the configuration with the-largest expected utility, we will say that a "correct selection" (CS) is made.

Because estimating the rank ordering in (13) reflects random fluctuation in configuration performance, it is impossible to guarantee a CS. Therefore, we ask the decision-maker to specify some level δ^* such that $\mathrm{E}[u_1(X'_{[K]1})] - \mathrm{E}[u_1(X'_{[K-1]1})] \ge \delta^*$ is practically significant. In general, the R&S procedure is designed to satisfy the following probability requirement:

$$P\{\mathrm{CS}\} \ge P^* \text{ whenever } \mathrm{E}[u_1(X'_{[K]1})] - \mathrm{E}[u_1(X'_{[K-1]1})] \ge \delta^*$$

where $(1/K) < P^* < 1$ and $0 < \delta^* < 1$. If $\mathrm{E}[u_1(X'_{[K]1})] - \mathrm{E}[u_1(X'_{[K-1]1})] < \delta^*$, then the procedure will select a configuration within δ^* of the best with probability at least P^*. See Butler et al. (2001) for procedures to aid in the selection of δ^*.

Butler et al. (2001) used an additive model (3) to evaluate four configurations for a seismic survey; the weights assessed were 0.4 for Cost (x_1), 0.2 for Duration (x_2) and 0.1 for each crew utilization measure (x_3 to x_6). To determine the best configuration we will calculate and present, the average re-scaled utility $\mathrm{E}[u_1(X'_{i1})]$, the re-scaled standard deviation, $\mathrm{StDev}(\mathrm{E}[u_1(X'_{i1})])$, the average cost $\overline{X}_{i1}$ from the simulator, and the average equivalent cost, $\overline{X}'_{i1}$, for each of the i configurations. The results are presented in Table 1 for comparison. Using the Rinott (1978) procedure the second stage results in Table 1 indicate that Configuration II is the superior approach.

Table 1: Results from Butler et al. (2001)

Configu-ration	Average Re-scaled Utility	Re-scaled Standard Deviation	Average Cost	Average Equiva-lent Cost
(I)	0.5759	0.0113	$112,354	$160,947
(II)	0.9595	0.0011	$68,551	$84,261
(III)	0.5403	0.0102	$151,197	$164,490
(IV)	0.8607	0.0036	$93,507	$117,275

To demonstrate that the results of Butler et al. (2001) are not dependent on the conservative Rinott procedure, we first replicate the results from Table 1 in Table 2 using the Kim and Nelson (2001) procedure as described by Branke, et al. (2005). Using batched means of size ten and δ^* =

0.00434 as assessed in Butler et al. (2001), the Kim-Nelson procedure stops after an initial examination of eight batched means and confirms that Configuration (II) is superior. For comparison, the results using the Rinott procedure in Table 1 were based on ten initial batches of size ten for all configurations in the first phase, and an additional ten and eighteen batches for Configurations I and III, respectively, in the second phase.

Table 2: Confirmation of Butler et al. (2001) Results using the Kim-Nelson (2001) Procedure

Configuration	Average Re-scaled Utility	Re-scaled Standard Deviation
(I)	0.5483	0.0116
(II)	0.9772	0.0009
(III)	0.5137	0.0502
(IV)	0.9297	0.0055

Both the Rinott and Kim-Nelson procedures indicate that Configurations II and IV are the top two performers. However, while these configurations perform well in terms of cost and duration they also have the highest crew utilizations. In fact, the average utilization of the layout crew exceeds the union negotiated maximum of 85% for these alternatives. Therefore, we should expect Configurations II and IV to decrease in terms of utility when we consider the constrained crew utilizations.

Table 3 presents the results when the Kim-Nelson (2001) procedure is applied to the target utility function in (12). Configuration IV never produced legal crew utilizations and is quickly eliminated after examining the eight batched means in round 1. In each subsequent round a single batched mean based on ten observations is added for each candidate configuration. After three additional rounds Configuration (I) is selected as the superior approach.

Table 3: Kim-Nelson Procedure for Target Rescaling

		Re-scaled Utility Configuration Mean (Standard Deviation)			
Round	Candidates	I	II	III	IV
1	I,II,III,IV	0.5333	0.0584	0.5117	0.0000
		(0.0108)	(0.0694)	(0.0106)	(0.000)
2	I,II,III	0.5323	0.0519	0.5132	
		(0.0105)	(0.0678)	(0.0109)	
3	I,II,III	0.5330	0.0467	0.5136	
		(0.0101)	(0.066)	(0.0104)	
4	I,II,III	0.5332	0.0424	0.5126	
		(0.0097)	(0.0642)	(0.0104)	
Final	**I**	**0.5332**			

One interesting feature of the analysis is the relatively large variance associated with Configuration (II) that leads to its continued consideration even though it has a relatively low average. This results from the fact that Configuration (II) meets the union crew utilization requirement in only 5% of the simulated cases, but when it does achieve this target it performs very well in terms of Cost and Duration resulting in high scores. It may be easier to see the similarity of the alternatives by inspecting the average and standard deviation of the re-scaled costs associated with each alternative as presented in Table 4.

Table 4: Average Equivalent Costs For Round 4 of Kim-Nelson Procedure

Configuration	Average Re-scaled Utility	Average Equivalent Cost	Std. Dev. Equivalent Cost
(I)	0.5332	$162,811	$3,675
(II)	0.0424	$193,787	$27,220
(III)	0.5126	$165,074	$3,227
(IV)	0.0000	$200,000	$0

5 CONCLUSIONS

This paper has demonstrated how the notion of constraints may be incorporated into a multiattribute utility analysis and applied to find the best simulated system, or subset of systems. The Butler et al. (2001) procedure was also applied to the more efficient Kim-Nelson (2001) ranking and selection procedure. The results from the example confirm the intuition associated with constraints and targets.

It is important to emphasize again that true constraints are likely to be rare in practice. Even our example with negotiated union crew utilizations might not be a true constraint. It is quite possible that in reality, the company would be fined or penalized in some other fashion if the negotiated maximums were exceeded. Alternatives that are close to the maximum should receive lower utility because they are more likely to penalized, but setting their utility to zero may be too extreme. Applying an acceptable range to these constraints may be a better approximation, effectively defining a utility function over performance in the acceptable range.

It might be more appropriate to use an additive model with binary utilities (e.g. the left panel of Figure 2). While not quite as strict as the constraint interpretation this formulation may be useful in some settings. For example, note that for three attributes with binary utilities $0.6u_1(x_1) + 0.3u_2(x_2) + 0.1u_3(u_3)$ implies non-compensatory preferences assuming the attributes are in decreasing order of preference. If an alternative satisfies the target for x_1, the only way it can be the worst performing alternative is if it does not satisfy any other targets and some other alternative satisfies at least one additional target. Similarly, if an alternative satisfies the target for x_3 it can only be the most preferred alternative if no other alternative satisfies any target, or the alternative satisfies at least one additional, relatively more important target. This preference model is consistent with the notion of elimination by aspects (EBA) (Tversky 1972) where a decision maker evaluates each alternative one attribute at a time, starting with the most important attribute, and proceeds until only one alternative has the highest score on a particular attribute or she runs out of attributes.

Also, recent work has demonstrated the equivalence of utility functions and preferences based on achieving a target (e.g. Bordley and Licazi 2000; Bordley and Kirkwood 2004; Abbas and Matheson 2005). These papers show that we could either assess a utility function for x_i, $u_i(x_i)$, or could assess a probability distribution that the target level is achieved given attribute performance x_i, $p(x_i)$. When the target is known with certainty, $p(x_i) = 0$ or $p(x_i) = 1$. In other words, the expected utility of an attribute with a known target Q as shown in the left panel of Figure 2 is the proportion of times that the target is satisfied. In some cases it may be more natural for a decision maker to think about achieving a target rather than a utility function but there is a mapping from one interpretation to the other, and each yield identical decisions. Thus Butler et al. (2001) can be reinterpreted using this target-based interpretation.

REFERENCES

Abbas, A.E., and R.A. Howard. 2005. Attribute dominance utility. *Decision Analysis*, 2, 185-206.

Abbas, A.E., and J.E. Matheson. 2005. Normative target-based decision making. *Managerial and Decision Economics*, 26, 373–385.

Andradóttir, S., D. Goldsman, and S. Kim. 2005. Finding the best in the presence of multiple constraint. In *Proceedings of the 2005 Winter Simulation Conference*. M.E. Kuhl, N.M. Steiger, F.B. Armstrong, and J.A. Jones, eds. The Institute of Electrical and Electronic Engineers, Piscataway, New Jersey, 732 – 738.

Batur, D., and S. Kim. 2005. Procedures for feasibility detection in the presence of multiple constraints. In *Proceedings of the 2005 Winter Simulation Conference*. M.E. Kuhl, N.M. Steiger, F.B. Armstrong, and J.A. Jones, eds. The Institute of Electrical and Electronic Engineers, Piscataway, New Jersey, 692 – 698.

Bordley, R. F., and C.W. Kirkwood. 2004. Multiattribute preference analysis with performance targets. *Operations Research* , 52, 823-0835.

Bordley, R., and M. LiCalzi. 2000. Decision analysis using targets instead of utility functions. *Decisions in Economics and Finance*, 23, 53-74.

Branke, J., S.E. Chick, C. Schmidt. 2005. Sensitivity analysis in ranking and selection for multiple performance measures. In *Proceedings of the 2005 Winter Simulation Conference*. M.E. Kuhl, N.M. Steiger, F.B. Armstrong, and J.A. Jones, eds. The Institute of Electrical and Electronic Engineers, Piscataway, New Jersey, 708-717.

Clemen, R. T. 1991. *Making Hard Decisions*. PWS Kent Publishing, Boston.

Dobrin, M. D., C.H. Savit. 1988. *Introduction to Geophysical Prospecting*. McGraw-Hill, New York.

Gupta, S.S., S. Panchapakesan. 1979. *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations*. Wiley: New York.

Hammond, J. S., R. L. Keeney, and H. Raiffa. 1998. Even swaps: A rational method for making trade-offs. *Harvard Business Review* (March-April 1998), 137-149.

Keeney, R. L. 1992. *Value-Focused Thinking: A Path to Creative Decisionmaking*, Harvard University Press, Cambridge, Massachusetts.

Keeney, R. L. 2002. Common mistakes in making value trade-offs. *Operations Research*, 50, 935-945.

Keeney, R. L., H. Raiffa. 1976. *Decisions with Multiple Objectives*. Wiley, New York.

Keeney, R. L., and D. von Winterfeldt. 1991. Eliciting probabilities from experts in complex technical problems. *IEEE Transactions on Engineering Management* 38, 191-201.

Kim, S.-H., and B. L. Nelson. 2001. A fully sequential procedure for indifference-zone selection in simulation. *ACM TOMACS* 11 251–273.

Mullarkey, P., G. Butler, S. Gavirneni, D. Morrice. 2006. Schlumberger uses simulation in bidding and executing land seismic surveys. To appear in *Interfaces*.

Rinott, Y. 1978. On two-stage procedures and related probability-inequalities. *Communications of Statistics: Theory and Methods* A8, 799-811.

Tversky, A. 1972. Elimination by aspects: A theory of choice. *Psych. Rev.*, 79, 281–299.

von Neumann, J. and O. Morgenstern. 1947. *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ.

AUTHOR BIOGRAPHIES

DOUGLAS J. MORRICE is a Professor in Operations Management at The University of Texas at Austin. He has ORIE Ph.D. from Cornell University. His research interests include simulation design, modeling, and analysis. Dr. Morrice was Co-Editor of the *Proceedings of the 1996 Winter Simulation Conference*, and 2003 Winter Simulation Conference Program Chair. He is currently serving as a representative for the INFORMS Simulation Society on the Winter Simulation Conference Board of Directors. His email address is <morrice@mail.utexas.edu>.

JOHN C. BUTLER is an assistant professor of Accounting and MIS at the Ohio State University. He has a Ph.D. in Management Science and Information Systems from the University of Texas. His research interests are in decision analysis and decision support systems <butler_267@cob.osu.edu>.