Chapter 13

Decision Analysis for Design Trades

13.1 INTRODUCTION

Decision making is a process undertaken by an individual or organization. The intent of this process is to improve the future position of the individual or organization in terms of one or more criteria. Most scholars [Howard, 1968] of decision making define this process as one that culminates in an irrevocable allocation of resources to affect some chosen change or the continuance of the status quo. The most commonly allocated resource is money, but other scarce resources are goods and services and the time and energy of talented people.

Watson and Buede [1987] have identified three primary decision modes: *choosing* one alternative from a list, *allocating* a scarce resource(s) among competing projects, and *negotiating* an agreement with one or more adversaries. Decision analysis is the common analytical approach for the first mode, optimization for the second, and a host of techniques have been applied to negotiation decisions [Jelassi and Foroughi, 1989]. Concepts of decision analysis are relevant to the second and third of these modes.

Section 13.2 provides a philosophical discussion of decision making and the elements of decision making: values, alternatives, and facts. Section 13.3 explains the rational basis of decision analysis in terms of a set of axioms that provide a compelling structure for some decision makers. Section 13.4 provides an analytical basis for modeling stakeholder values in the face of conflicting objectives, a critical element in design decisions when faster, better,

The Engineering Design of Systems: Models and Methods, Second Edition. By Dennis M. Buede

and cheaper are all desired but not mutually compatible. Section 13.5 discusses the modeling of uncertainty and risk preference for design decisions; decision trees, relevance diagrams, and influence diagrams are introduced as modeling tools. A sample application focused on the development of trade-off requirements consistent with an objectives hierarchy and performance requirements is presented in Section 13.6; this sample application is based upon a real application of decision analysis to requirements development.

This chapter describes a model of uncertainty (probability theory), a model of value (multiattribute value theory), a model of risk preference (utility theory), and a normative model for incorporating uncertainty, value, risk preference, and complexity for aiding the thought and conversation process needed to make explicit, rational decisions.

13.2 ELEMENTS OF DECISION PROBLEMS

Decision analysis is a normative theory for making a *decision* (an irrevocable allocation of scarce resources). The three major elements of a decision that make its resolution troublesome are the creative generation of alternatives, the identification and quantification of multiple conflicting criteria, and the assessment and analysis of uncertainty associated with the what is known and not known about the decision situation. Howard [1993] has drawn an analogy between the model building and analysis processes inherent in decision analysis and a conversation with a decision maker. The conversation (or modeling) needs to address what the decision maker (stakeholders in systems engineering) cares about (values), what the decision maker can do (alternatives), and what the decision maker knows (facts or absence thereof).

Many stakeholders and systems engineers claim to be troubled by the feeling that there is an, as yet unidentified, alternative that must surely be better than those so far considered. The development of techniques for identifying such alternatives is receiving considerable attention [Elam and Mead, 1990; Friend and Hickling, 1987; Keller and Ho, 1988; Keeney, 1992; West, 2007].

Ample research [von Winterfeldt and Edwards, 1986] has been undertaken to identify the pitfalls in assessing probability distributions that represent the uncertainty of a stakeholder. Research has also focused on the identification of the most appropriate assessment techniques. Similar research [von Winterfeldt and Edwards, 1986] has focused on assessing value and utility functions. Keeney [1992] has recently advanced concepts for the development and structuring of a value hierarchy for key decisions. While it will never be possible to turn decision support via decision analysis over to a computer, the vast number of real-world applications of decision analysis [Kirkwood and Corner, 1993] demonstrate that this analytic modeling support is well worth the time and effort.

13.3 AXIOMS OF DECISION ANALYSIS

There are five basic rules of thought [von Neumann and Morgenstern, 1947; Howard, 1992] that establish decision analysis: probability, order, equivalence, substitution, and choice. Probability is adopted as the representation of uncertainty. This is a well-founded discipline for addressing uncertainty and is the common approach within engineering.

The order rule states that our preferences are sufficiently well defined that any possible list of outcomes associated with the design alternatives can be ordered from least preferred to most preferred on each objective in the fundamental objectives hierarchy. In addition, once our preferences are aggregated across all objectives there is a single list of outcomes ordered by our preferences. Naturally, it is possible to be indifferent between two outcomes on a specific objective or on the aggregate. Our preference order does not need to be the same from one objective to the next; in fact, there would be no need to have multiple objectives if this were the case. The ordered list must be transitive, which is to say that any outcome can only appear once on any ordered list. If this is not the case, we become subject to the "money pump" argument; a disinterested party could entice us to put up an infinite amount of money by offering us a sequence of trades among three alternatives. For example, I would be intransitive if I stated that I preferred a Lexus to a Cadillac, a Cadillac to a BMW, and a BMW to a Lexus. With these preferences and ownership of a Lexus, I would pay to swap for your BMW, pay again to swap the BMW for your Cadillac, and then pay a third time to swap the Cadillac for the Lexus I originally owned. By this time I should realize there was something wrong with my preference structure.

The equivalence rule sets up a situation with three outcomes, A, *B,* and C, where A is preferred to ($>$) B, and $B > C$. This rule states that there is some lottery containing a probability, p, of obtaining outcome A and a probability of $(1 - p)$ of obtaining C that will make us indifferent to obtaining outcome B for sure.

The substitution rule states that we are willing to substitute any combination of outcomes in a decision-making situation if we are indifferent between them. This is just the operational definition of equivalence.

Finally, suppose we have two alternatives, each with exactly the same outcomes, and the probabilities of the outcomes are the same for all but two. If one of the alternatives has a higher probability associated with the outcome that is most preferred, then we should be happy to choose this alternative. This is the choice rule.

Given these four rules plus the axioms of probability theory, a normative theory of decision making results that dictates the maximization of expected utility. Utility in this case needs to be measured on an interval scale; an interval scale preserves equal intervals of measure and can be multiplied or divided by a constant and can have a constant added or subtracted from it. A ratio scale of measurement for utility could be used but is not necessary. Note that probabilities are constructed on a ratio scale.

13.4 MULTIATTRIBUTE VALUE ANALYSIS

Multiattribute value analysis is a quantitative method for aggregating a stakeholder's preferences over conflicting objectives to find the alternative with the highest value when all objectives are considered. (Note the phrases "multiattribute utility analysis" and "multiple objectives decision analysis" are also often used. In this book the word utility is reserved for situations in which uncertainty has been explicitly modeled and the stakeholder's risk preference is being included in the analysis.) Multiattribute value analysis can be addressed simply as is done in this chapter or with a great deal more sophistication [see French, 1986; Keeney and Raiffa, 1976]. Additional insights can be found in Kwinn and Parnell [2007]. Other approaches to value computations are also available: analytical hierarchy process (AHP) [Saaty, 1980, 1986], percentaging [Nagel, 1989], the technique for order preference by similarity to ideal solution (TOPSIS) [Yoon, 1980], a fuzzy algorithm [Yager, 1978], quality function deployment (QFD) [Akao, 1990], and Pugh matrix [Pugh, 1991]. None of these other approaches are based on an underlying set of axioms that provide a foundation for justifying an analytical process except the AHP. However, there are a number of analytical concerns that have been raised about AHP, percentaging, TOPSIS, and similar approaches [Buede and Maxwell, 1995; Dyer, 1990; Harker and Vargas, 1990].

The process for defining the objectives of interest for a system has been defined in Chapter 6. For the systems engineering application addressed in this book, the objectives are the performance requirements that have been defined as described in Chapter 6, as well as derived performance requirements that have been defined as part of the development of the allocated architecture.

Following the definition of the objectives, a value scale must be defined for each objective at the bottom of the objectives hierarchy. This value scale definition begins by defining the minimum acceptable value of performance for a given objective (constraining requirement) and the most desired value of performance for the objective (the design goal). Then the relative value of improving from the minimum acceptable threshold to the design goal is quantified in the form of a value curve. Objectives that are a combination of bottom-level objectives are in the hierarchy for ease of aggregation and communication; as a result these intermediate and the top-level (or fundamental) objectives are computed from lower level objectives.

After value scales are defined for each bottom-level objective, value weights that address the relative value associated with improving from the bottom (minimum acceptable threshold) of the value scale to the top (design goal) must be assessed from the stakeholder for all bottom-level objectives as well as the intermediate objectives. The discussion in this chapter is going to address the common, but not universal, case in which the values can be aggregated across objectives by using a weighted-average formula. The books by French [1986] and by Keeney and Raiffa [1976] address the general aggregation process and the assumptions required for various aggregation formulas.

The assumption that the general value function over the vector x of n bottom-level objectives can be written as a weighted additive function of value functions on the individual objectives:

$$v(\mathbf{x}) = \sum_{i=1}^{n} w_i v_i(x_i) \tag{13.1}$$

will be adopted from here on out. Note the weights are commonly normalized to sum to 1.0, and the value functions are normalized to range from either 0 to 1, or 0 to 10, or 0 to 100.

13.4.1 Eliciting Value Functions

The axioms of decision analysis produce the result that the value function over the vector x of bottom-level objectives must only be an interval function when the decision maker is risk neutral (the assumption made here). As a result, the individual value function v_i over bottom-level objective x_i must also be an interval-scaled function of x. This interval property is the key to eliciting value functions from stakeholders about the relative value they assign to improving from the threshold of acceptable performance of x_i, x_i^0, to the most desired value of x_i^*. Watson and Buede [1987] present the bisection and the equal differences methods for eliciting these functions.

These value functions take four general forms (see Fig. 13.1): decreasing returns to scale (RTS), linear RTS, increasing RTS, and an S-curve. The decreasing RTS signifies a satiation of preference near the most desired value. Decreasing RTS is commonly encountered when the threshold of acceptable performance is within the key performance range of interest to the stakeholders and the most desired value is outside this key performance range where satiation takes over. The linear RTS is commonly found when both the threshold of acceptable performance and the most desired value are within the key performance range of interest, or when there is no possible satiation of preference. The increasing RTS occurs when (1) the threshold of acceptable performance has been pushed below (in a value sense) the key performance range and (2) there is a technological or other cap on the most desired value so satiation of preference has not begun. Pushing the threshold of acceptable performance below the key performance range in a value sense means limited value is obtained by small increases in the performance parameter until some significant change is achieved. The S-curve reflects a joining of decreasing and increasing RTS and reflects the case in which the key performance range lies between the threshold of acceptable performance and the goal. The S-curve indicates that the range of possible designs has been maximized.

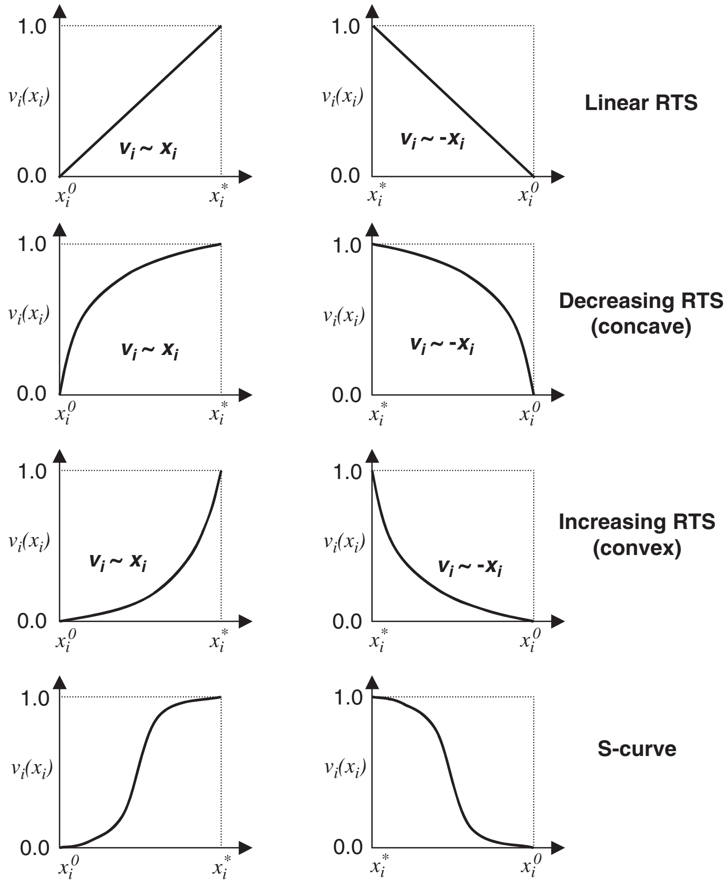


FIGURE 13.1 Common types of value curves.

Note no value curves that increase and then decrease, or decrease and then increase, have been shown. When value functions that are not monotonic (always increasing or always decreasing) are elicited, it is highly likely that there are two underlying objectives that have been combined. These two objectives should be separated so that the stakeholders are only considering one objective at a time when being asked to specify their preferences.

Exponential functions are most commonly used to approximate the value functions of stakeholders [Kirkwood, 1997]. Equation (13.2) shows a standard form for variables on which more is better and that is normalized to be 0 when the minimum acceptable threshold is met and 1.0 when the design goal is met. When a is greater than 1.0, this equation demonstrates decreasing RTS. When a is equal to 1.0, this equation becomes a straight line. When a is less than 1.0, this equation demonstrates increasing RTS.

$$v_i(x_i) = \frac{1 - e^{-\alpha(x_i - x_i^0)}}{1 - e^{-\alpha(x_i^* - x_i^0)}} \tag{13.2}$$

Wymore [1993] has suggested a value function (or figure of merit) family that can accommodate all of the above value curves to some degree.

13.4.2 Eliciting Value Weights

Before discussing how to elicit the weights that are used in the additive value function of Eq. (13.1), the meaning of these weights must be made clear. In words, *the weights must reflect the relative value associated with increasing from the bottom to the top of each value scale*. Note in Figure 13.1 each of the value functions has been normalized to range from 0 to 1. Other normalizations, for example, 0 to 10, 0 to 100, 14 to 85, are all acceptable, but it is usually most meaningful to stakeholders and everyone else to have every value function normalized from the same bottom value to the same top value. Value weights that reflect the relative value in increasing from the bottom to the top of each value scale are called *swing* weights because they represent the value attached to the swing from bottom to top.

Why must the weights reflect this change in value from the bottom to the top of the value scale? Consider the most general assumption that we can make about the value function, namely that the value across all objectives is the sum of individual value functions, $v_i'(x_i)$, functions that have not yet been normalized *in* any way; see Eq. (13.3):

$$v(\mathbf{x}) = \sum_{i=1}^{n} v_i'(x_i) \tag{13.3}$$

Equation 13.4 normalizes $v_i'(x_i)$ to range from 0 to 1. Recall that the axioms of decision analysis implied that an interval-scaled value function was sufficient, meaning that we can add or subtract constants from an interval scale, as well as multiply or divide by constants and still have an interval scale. The normalized value function, $v_i(x_i)$, is computed by subtracting a constant from the unnormalized value function; this constant is the unnormalized value associated with the worst value (x_i^0) of x_i. This result is then multiplied by a constant, namely the range in unnormalized value from worst to best (x_i^*) levels of x_i. Note that when $x_i = x_i^*$, the numerator and denominator are equal. When $x_i = x_i^0$, the numerator equals 0.

$$v_i(x_i) = \frac{1}{v_i'(x_i^*) - v_i'(x_i^0)} [v_i'(x_i) - v_i'(x_i^0)] \tag{13.4}$$

Now solving for the unnormalized value function:

$$v_i'(x_i) = [v_i'(x_i^*) - v_i'(x_i^0)]^* v_i(x_i) + v_i'(x_i^0) \tag{13.5}$$

Substituting (13.5) into (13.3) we get

$$\begin{aligned} v(\mathbf{x}) &= \sum_{i=1}^{n} (v_i'(x_i^*) - v_i'(x_i^0))^* v_i(x_i) + v_i'(x_i^0) \\ &= \sum_{i=1}^{n} (v_i'(x_i^*) - v_i'(x_i^0))^* v_i(x_i) + \sum_{i=1}^{n} v_i'(x_i^0) \end{aligned} \tag{13.6}$$

The last summation is a constant that has no relevance to distinguishing among alternatives, so it can be subtracted from both sides of the equation.

Now divide both sides by the constant

$$\sum_{i=1}^{n} [(v_i'(x_i^*) - v_i'(x_i^0)]$$

and distribute this term throughout the summation on the right side of the equals sign. The weights for each objective are defined to be

$$w_i = \frac{v_i'(x_i^*) - v_i'(x_i^0)}{\sum_{i=1}^{n} (v_i'(x_i^*) - v_i'(x_i^0))}, \tag{13.7}$$

Substituting Eq. (13.7) into (13.6),

$$\frac{v(\mathbf{x}) - \sum_{i=1}^{n} v_i'(x_i^0)}{\sum_{i=1}^{n} [v_i'(x_i^*) - v_i'(x_i^0)]} = \sum_{i=1}^{n} w_i v_i(x_i), \tag{13.8}$$

which is a linear transformation of the original value function and therefore equivalent to Eq. (13.1). So the value weights in Eq. (13.1) must be defined to be the relative swing in value from the worst point x_i^0 to the best point x_i^* across all objectives.

Any mathematical approach employing interval scales that uses Eq. (13.1) to compute value but does not explicitly call for the use of swing weights is doing the equivalent of changing money from one currency to another by picking a random set of exchange rates rather than using the current market-derived exchange rates. The use of weights that are not swing weights may well suggest an alternative as best that is not consistent with the stakeholders' preferences. Some methods such as the Pugh methodology [Pugh, 1991] hope the objectives can be developed so that they are nearly equal in relative weight, without even defining what the weight means. No application of the authors (out of over a hundred) has generated a set of objectives that were nearly equal in importance.

While value functions only need to be interval scales, weights must be defined on a ratio scale. A *ratio* scale is one on which zero means zero value. In this case the value at the design goal must be equal to the value at the minimum threshold: $v_i(x_i^*) = v_i(x_i^0)$. A weight of zero means that the objective can be ignored.

Weight elicitation techniques can be divided into two categories: those that ask directly for numbers and *those* that ask for indirect ordinal or interval judgments that are used to derive a ratio scale.

13.4.2.1 Direct Weight Elicitation Techniques. The most common direct elicitation technique for ratio scale numbers is to ask people to *spread 100 points* among the objectives at any given level of the objectives hierarchy. This is a typical technique for eliciting weights in any multiattribute value application. The research literature [Watson and Buede, 1987] is not kind to this technique, and our experience confirms the literature findings. While it is relatively easy to do, people assign numbers that are far too close together to meet any ratio scale requirements; this is true no matter how many caveats the assessor presents to the participants to remember the ratio scale requirements [Stillwell et al., 1981].

Two other common direct assessment techniques involve *anchoring* on either the most important or least important objective. The stakeholder is then asked to *assign the most* (*least*) *important a score of 100* (*1*) *and scale the remaining down* (*up*) based upon ratio scale requirements. The research literature has not really examined this method. In practice, it has not worked well for making the initial assessment queries, but has worked reasonably well when it is introduced later in the assessment process. By this point, the stakeholders have become accustomed to thinking about ratio scale properties based upon a more detailed assessment process. The advantage of starting with the most important objective is that the stakeholders are probably most familiar with it and therefore, it is a useful anchor. The least important objective may not be that familiar to the stakeholders. In either case, the weights are normalized to sum to 1.0 at the end.

Edwards [1977] introduced a multi-attribute utility technique called SMART that was based upon importance weights. (Edwards describes this as a self-recognized intellectual error [Edwards and Barron, 1994].) Edwards and Barron [1994] introduced SMARTS and SMARTER. SMARTS is simply SMART recast with the intellectually proper swing weights. SMARTS employs anchoring on the best objective at 100 points and scaling the rest down, then normalizing the weights to sum to 1.0.

SMARTER involves using the *rank-order centroid* technique of transforming the swing ranks of criteria into swing weights. Stillwell et al. [1981] offered several ad hoc ways to translate rank orders into weights. In the following

equations, r_i is the rank of the ith objective, K is the total number of objectives, and w_i is the normalized approximate ratio scale weight of the ith objective.
Rank sum:

$$w_i = \frac{K - r_i + 1}{\sum_{j=1}^{K} K - r_j + 1}$$

Rank exponent:

$$w_i = \frac{(K - r_i + 1)^z}{\sum_{j=1}^{K} (K - r_j + 1)^z}$$

where z is an undefined measure of the dispersion in the weights. The larger z is the larger is the ratio of the most important objective to the least important objective.
Rank reciprocal:

$$w_i = \frac{1/r_i}{\sum_{j=1}^{K} \left(1/r_j\right)}$$

Rank-order centroid (ROC):

$$w_i = \left(1/K\right) \sum_{j=i}^{K} \left(1/r_j\right)$$

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

Barron and Barrett [1996] show that ROC weights accurately define the best alternative 75 to 90% of the time based upon a set of true swing weights elicited some other way. When the incorrect alternative was identified, the loss of utility averaged 3 to 7%. The ROC results were at the worst ends of these ranges when the attribute values of the alternatives were negatively correlated, which unfortunately is the most common situation in practice. Barron and Barrett [1996] show that the rank-reciprocal and rank-sum weights were nearly always worse than the ROC weights. Kirkwood and Corner [1993] use an actual application by Ulvila and Snider [1980] on oil tanker standards to provide some results that contradict claims concerning the effectiveness of rank-sum, rank-reciprocal, and rank-exponent weights.

SIDEBAR 13.1: ILLUSTRATION OF WEIGHTING TECHNIQUES

To illustrate the weight elicitation techniques, consider the following engineering design sample problem. Suppose a communication system to be deployed as part of a data collection system is being designed. As part of our requirements analysis the following five major performance parameters that determine successful and profitable data collection operations (our measure of effectiveness) have been identified and ranked based upon the importance of the swing from minimum acceptable to ideal performance:

Performance Parameter	Minimum Acceptable Performance	Design Goal	Rank Order
Throughput, mbits/sec	100	120	1
Availability	0.85	0.95	2
Operating life, yrs	5	7	3
Procurement cost, $	100	85	4
Operating cost, $/mo	1.00	0.70	5

For the rank-based techniques the results in the table below are obtained. (Note that a 0.4 was used for the parameter in the rank exponent method.)

Rank Method	Throughput	Availability	Operating Life	Procurement Cost	Operating Cost
Rank sum	0.33	0.27	0.20	0.18	0.07
Rank exponent	0.25	0.23	0.21	0.18	0.13
Rank reciprocal	0.44	0.22	0.14	0.11	0.09
ROC	0.45	0.26	0.16	0.09	0.04

13.4.2.2 Indirect Weight Elicitation Techniques. Indirect assessment of weights can be obtained via one of several paired comparison techniques and the use of graphical adjustments on a computer. These techniques are generally far superior to any of the direct techniques in their ability to capture the decision maker's trade offs across objectives.

The *paired* comparison *techniques* are the most common and include the analytical hierarchy process (AHP) [Saaty, 1980], trade offs [Watson and Buede, 1987], balance beam [Watson and Buede, 1987] judgments, and lottery questions [Keeney and Raiffa, 1976].

AHP (see Sidebar 13.2) can be used to assess the weights of the objectives. In the full implementation of AHP, it is not easy to elicit swing weights because the AHP does not use the full value scale from 0 to 1. In AHP the stakeholders are asked to compare each objective with every other objective; note it is possible to skip some comparisons, but the accuracy of the results decreases rapidly as the number of skipped comparisons grows. The AHP commonly does not ask the stakeholders to rank order the objectives in terms of overall benefit but begins by asking the stakeholders to compare objectives two at a time in whatever order they appear. The stakeholders are given the option of using a verbal scale, a numerical scale, or adjustable bar graphs. The numerical scale ranges from 9 times more valuable to one ninth as valuable. The verbal choices have numerical equivalents that also vary from 9 to one ninth. If there are K objectives, AHP would pose $K(K-1)/2$ questions of this sort. These responses are used as an input to form a matrix upon which an eigenvector calculation is performed; these mathematical operations are justified by a set of axioms that Saaty [1980, 1986] has developed. It is possible that the stakeholders' judgments have inconsistencies embedded in them. Saaty [1980] has developed an inconsistency index based upon the mathematical operations he developed. Typically, the stakeholders are asked to rethink selected judgments if the inconsistency index is greater than 0.1. This approach seems to work well when the number of objectives is greater than 3 and less than 7 or 8. Naturally, it is possible to break a large number of objectives into subsets too/de this approach more efficient.

Trade offs are used for swing weights and involve using the scores to help elicit the weights of the objectives. First, the objectives are ranked in order of their overall swing in value. Next, the stakeholders are asked if the overall swing weight of the second objective is as great as the swing from the lowest to some intermediate point of the value scale of the first ranked objective. For example, the stakeholders are asked whether the overall swing in value of the second ranked objective was closer to 80 or 60% of the swing in value of the first ranked objective. Suppose after some discussion the stakeholders agreed that the swing in value on the second objective was roughly equivalent to a swing from 0 to 0.7 on the value scale (normalized to a high of 1.0) of the first objective. This establishes that the weight of the second objective is 70% that of the first objective. The third ranked objective could now be compared to intermediate points on either the first or second ranked objectives. This method works very well when the value curves are firmly established and the value curves are continuous. If the value curves change significantly after trade offs have been used, the weights have to be reassessed.

SIDEBAR 13.2 AHP EXAMPLE

Returning to the example of design trade offs for a communication system, suppose the stakeholders provide the judgments shown in the following table into the AHP verbal mode.

	Throughput	Availability	Operating Life	Procurement Cost	Operating Cost
Throughput	(Equal) 1	2	4	6	(Absolutely) 9
Availability	1/2	(Equal) 1	4	(Strongly) 5	(Absolutely) 9
Operating Life	1/4	1/4	(Equal) 1	(Weakly) 3	(Strongly) 5
Procurement Cost	1/6	1/5	1/3	(Equal) 1	2
Operating Cost	1/9	1/9	1/5	1/2	(Equal) 1

The normalized eigenvector of the largest eigenvalue for the numerical version of the above matrix is 0.43, 0.39, 0.11, 0.03, and 0.01. (Note that the AHP process associates a 9 with absolutely, 7 with very strongly, 5 with strongly, 3 with weakly, and 1 with equal.)

The *balance beam approach* is another approach for assessing the weights of the objectives (see Sidebar 13.3). The stakeholders are initially asked to establish a rank order of the overall swing weights of the objectives. Next, a series of questions is posed to the stakeholders that begins with "Is the overall swing in value of the first objective (a) greater, (b) less than, or (c) equal to the combined overall swing in values of the second and third most important objectives?" To illustrate this question a balance beam analogy (see Fig. 13.2) is used. If the stakeholders respond that the first ranked objective has the highest overall swing weight, the attractiveness of the other choice is increased by adding the fourth ranked objective to the package of the second and third ranked objectives. If the stakeholders say the package of second and third ranked objectives has a higher swing value than the first ranked objective, the attractiveness of the combination package is decreased by dropping the third ranked objective and adding the fourth ranked objective. This process is continued until the stakeholders have found a package of objectives with an overall swing in value that is comparable to the first ranked objective. Next, the second ranked objective is compared with the third and fourth ranked objectives. This continues until only the last two objectives remain. The process creates a set of inequality and equality equations that relate the swing weights of the objectives. Typically, a weight of 1 is assigned to the least weighted objective, the stakeholders are asked to assign a swing weight to the second least weighted objective, and then the equations are used to bound the swing weights of the remaining objectives. It is possible that there will be an inconsistency in a subset of the equations. If such an inconsistency exists, the balance beam questions posed by this subset of equations are reexamined until the stakeholders identify their inconsistency and make an adjustment. This approach generally produces a wide spread in the swing weights for the objectives.

SIDEBAR 13.3: BALANCE BEAM EXAMPLE

Using the balance beam approach for the communication system design the stakeholders are asked to compare the swing in benefit of throughput (T) to that of the combined swings of availability (A) and operating life (OL). The stakeholders respond the combination is greater than that of throughput, or

$$T < A + \mathrm{OL}$$

However, throughput (T) is preferred to availability (A) and procurement cost (PC):

$$T > A + \mathrm{PC}$$

Availability is preferred to OL, PC, and operating cost (OC):

$$A > \mathrm{OL} - 1 - \mathrm{PC} + \mathrm{OC}$$

OL is preferred to PC and OC:

$$\mathrm{OL} > \mathrm{PC} + \mathrm{OC}$$

Next, the unnormalized weight of operating cost is fixed at 1 and the stakeholders are asked to provide a ratio weight for procurement cost; suppose they say 1.5. Now the weight for operating life is greater than 2.5, suppose the stakeholders say 3. The stakeholders now know that the weight for availability is greater than 4.5 (3 + 1.5) and agree to a weight of 6. Finally, the weight of throughput is between 7.5 (6 + 1.5) and 9 (6 + 3). The stakeholders choose 8. The normalized weights are 0.41, 0.31, 0.15, 0.08, and 0.05.

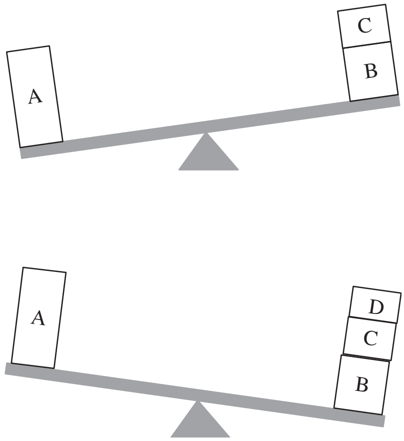


FIGURE 13.2 Balance beam analogy for paired comparisons.

Graphical elicitation procedures have been implemented in several software packages for the elicitation of scores and weights. Bar graph adjustment is most commonly used, but some software packages contain adjustable pie charts, where the wedges of the pie represent different objectives.

13.5 UNCERTAINTY IN DECISIONS

This section addresses the analysis of decisions when there is substantial uncertainty associated with outcomes impacting the relative value of the decision's alternatives. In systems engineering this uncertainty could be associated with the state of technology at some time in the future; the stakeholders' needs now and in the future; the ability to achieve cost, schedule, or performance goals; and environmental variables associated with the use or testing of the system.

Probability theory is discussed in Section 13.5.1 to refresh the reader's knowledge of this subject. Section 13.5.2 discusses the use of relevance diagrams to represent joint probability distributions. Influence diagrams are introduced in Section 13.5.3 as a way of representing a decision. The calculations of expected utility are described in terms of decision trees. Section 13.5.4 addresses risk preference.

13.5.1 Probability Theory

This section is not meant to be a detailed introduction to probability theory; for such an introduction see Roberts [1992] and Ghahramani [1996]. The reader is assumed to be familiar with the concepts of probability density functions for continuous random variables, probability mass functions for discrete random variables, the difference between marginal and conditional probability distributions, the notion of cumulative probability distributions, and joint probability distributions of two or more random variables. First, the concepts of probabilistic independence and dependence are discussed. Then two important equations, the law of total probability and Bayes rule, are provided. Finally, relevance diagrams are introduced as a way to describe the probabilistic dependencies among a set of random variables. This entire discussion will be conducted in terms of discrete random variables because the mathematics is easier to convey, and discrete random variables are more commonly encountered is systems engineering problems. In addition, decision analysis commonly discretizes continuous random variables for computational ease.

The *probabilistic independence* of two random variables, X and Y, is defined to occur when the conditional probability distribution on X given Y equals the marginal probability distribution on X. It can be shown that when the preceding is true for X, then the probability distribution on Y given X must also equal the probability distribution on Y. As a result, the joint probability distribution of instances of X, x_i, and Y, y_i, can be written as

$$p(x_i, y_j) = p(x_i|y_j)\,p(y_j) = p(y_j|x_i)\,p(x_i) = p(x_i)\,p(y_j) \tag{13.9}$$

when X and Y are probabilistically independent. Intuitively, probabilistic independence means that learning the value of X does not cause us to change our probability distribution about Y.

The *law of total probability* allows the computation of a marginal probability distribution of one random variable by summing over all possible values of a second random variable that is probabilistically dependent on the first. This law is used to compute $p(x_i)$ when the probabilities on the right-hand side of Eq. (13.10) are known better than $p(x_i)$ (shown in Fig. 13.3):

$$p(x_i) = \sum_{j=1}^{m} p(x_i|y_j)p(y_j) \tag{13.10}$$

Bayes rule is used to update our uncertainty on one random variable when information about another random variable becomes available, assuming the

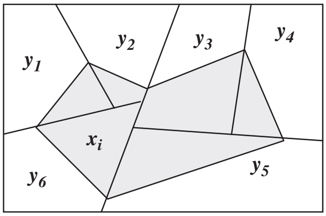


FIGURE 13.3 x_i as a subset of the universal event, which is partitioned by Y.

two random variables are probabilistically dependent on each other.

$$p(y_j|x_i) = \frac{p(x_i|y_j)p(y_j)}{\sum_{i=1}^{n} p(x_i|y_j)p(y_j)} = \frac{p(x_i|y_j)p(y_j)}{p(x_i)} \tag{13.11}$$

In the case of Eq. (13.11) information about the value of random variable X is obtained and is used to update our uncertainty about Y. The left-hand side of Eq. (13.11) is called the posterior probability distribution of Y when all values of $j = 1, 2, \ldots$ m are considered. The $p(y_j)$ in the numerator on the right-hand side of (13.10) is called the prior probability, the probability of Y before information on X became available. The values of $p(x_i|y_j)$ in the numerator and denominator are called the likelihood values of getting information on X given values of Y. Finally, the denominator of Eq. (13.10) is called the preposterior and is in fact equal to $p(x_i)$, as computed by the law of total probability [Eq. (13.10)]. The contrast between the law of total probability and Bayes rule can be seen by revisiting Figure 13.3. With the law of total probability the task is to compute the probability of a subset of the universal event using conditional probabilities that partition the universal event. With Bayes rule the universal event has been redefined based upon a new state of information, namely x_i is known to be true. Bayes rule provides the process for updating the probability of any variable based upon this new information.

Adoption of Bayes rule in practice requires a philosophical shift in the meaning of probabilities for most people. The most common philosophical interpretation of probability among engineers and statisticians is that of a long-run frequency associated with a set of events that have been or could be repeated many times, for example, flipping coins, removing production samples from a production line. However, in systems engineering the engineer of a system is typically involved in very early design decisions regarding the operational system, the test system for the operational system, the manufacturing system of the operational system, the test system for the manufacturing system of the operational system, and so forth. In these early design decisions there is typically a great deal of uncertainty about specific outcomes related to these decisions and very little data. In fact, it is often not possible to contemplate repeating experiments to develop long-run frequencies within a reasonable amount of time and money. Bayesian, or subjective, probability interprets a probability as a state of information about the uncertainty regarding a variable. Powerful mathematical and logical arguments have been put forward by Savage [1954], De Finetti [1974], Lindley [1994], and others for this interpretation of probability. Now that the computational power that we have on our desks is quite sizable, many theoreticians are becoming Bayesians due to the theoretical justification of the Bayesian argument. Yet many of these Bayesian converts still prefer to put uniform priors on the random variables and let the data shape the posterior distributions. This is fine when there is a lot of data, as there is late in the systems engineering development process. Early in the development process there is precious little data and uniform priors are not consistent with engineering judgment and likely to lead to poor design decisions. There is a vast amount of research available on the ability of humans to provide probability judgments [Hogarth, 1980; Kleindorfer et al., 1993; Wright and Ayton, 1994]. Serious probability elicitation processes have been developed and used extensively with successful results [Spetzler and Stael von Holstein, 1975; Merkhofer, 1987].

Bayes rule is useful during the design phase in systems engineering when there is little hard data available. During this phase there are often significant results available from analyses and simulations; these results are appropriately considered as data, making Bayes rule an appropriate tool.

Bayes rule has wide applicability in the world of testing. Before the test we have some uncertainty about the ultimate value of certain performance, cost, or schedule parameters. Data is collected during the test regarding the values of certain system or project characteristics that relate to the parameters of interest. These data should then be used to update our uncertainty about the parameters of interest. Test data should always be viewed as likelihood measures. All too often, the test result is viewed to be the answer, and only the data parameter associated with the largest likelihood value is reported.

13.5.2 Relevance Diagrams

A *relevance diagram* is a directed graph, or digraph, that is a statement of the joint probability distribution among a set of random variables as a factorization of conditional and marginal probability distributions. For example, the three possible factorizations of two random variables, X and Y, are shown in Figure 13.4. Each random variable is shown as a node with an oval encapsulation. The top case shows two probabilistically independent random variables; the absence of an arc indicates this independence. The next two cases show dependence or relevance in a Bayesian sense of probabilistic updating; the arc can go in either direction, with the direction reflecting a different conditional and marginal distribution that define the joint distribution. It is obvious from this simple graph that the arc in the bottom two graphs can be flipped (have its direction changed) without any repercussions. However, this is not true in general. A relevance diagram cannot have a cycle (see Chapter 5 for a definition), so flipping an arc that causes a cycle to form is never possible. In addition, when flipping an arc does not cause a cycle to be formed, it is possible that arcs will have to be added to the digraph [see Shachter, 1986].

As an example of relevance diagrams for systems engineering, consider an elevator design in which the state of technology related to control systems and power systems is highly uncertain in the time frame of the development effort (Fig. 13.5). The key performance requirements (design objectives) are elevator performance in terms of mean wait times; the operational cost of the system; and the availability of the elevator system. A relevance diagram depicting the probabilistic dependencies is shown in Figure 13.5. Note that there is no

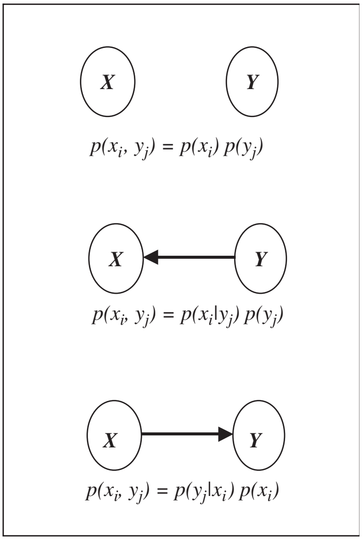


FIGURE 13.4 Relevance diagrams for two variables.

dependence between the three key performance requirements; these three variables are probabilistically independent of each other given the states of control technology and power technology. This is called conditional independence; if the variables for the control and power technologies were not present, there would be edges between the three requirements nodes (performance, availability, and cost). As discussed in previous chapters, there is great power to be gained in communicating the structure of reasoning (modeling) about design issues by using a graphical representation such as relevance diagrams.

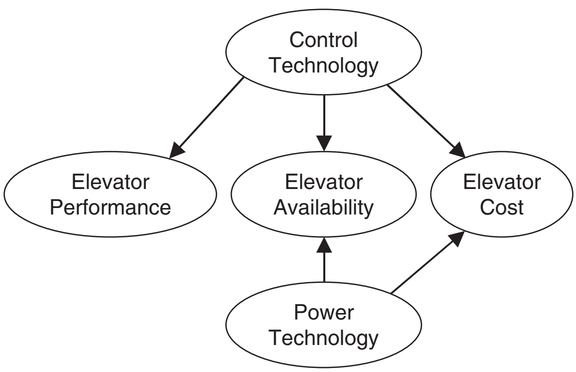


FIGURE 13.5 Notional relevance diagram for elevator design.

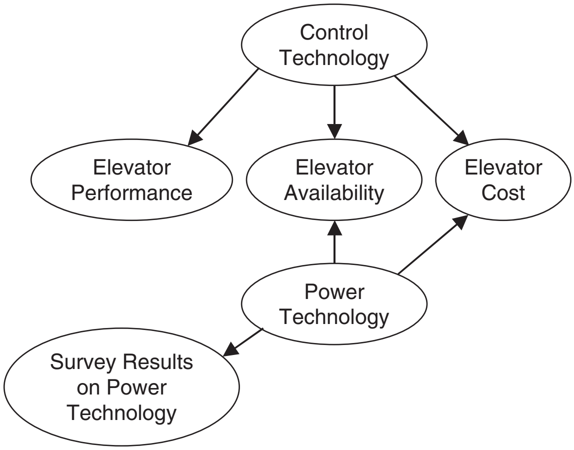


FIGURE 13.6 Relevance diagram with survey data on power technology.

As mentioned above, test results always provide likelihood information for Bayes rule. As a result, a relevance diagram that includes test results will have arcs going to the test result from the variable relevant to the test. A survey of power technology to assess the possible state of power technology in two years is an example of test data for the elevator design problem. This test data would be shown as a node with an arc coming to it from the Power Technology node in Figure 13.6. Bayes rule would then be used to flip this arc so that the survey results could be incorporated in the decision being made.

13.5.3 Influence Diagrams and Decision Trees

Consider a standard design decision faced by systems engineers: Should a component for the system be bought from an existing supply source or be developed from more basic components? The uncertainty that may be most troublesome in this decision is how long it will to take to develop the major component and how much will it cost. The schedule and cost results could be better than, equal to, or worse than the result associated with purchasing the component. For this simple example assume the performance of both alternatives is equal. A *decision tree* depicting this decision is shown in Figure 13.7. The value computation at the end of each branch of the tree addresses the cost and schedule issues via a multiattribute value formulation. The decision node at the beginning of the tree depicts the two alternatives as branches emanating from a small square. After the Build alternative there are chance nodes (little circles) that represent the uncertainties concerning cost and schedule. The tree is ‘‘rolled back’’ by multiplying the value at the end of each branch times the probability value on the branch just before it. These probability-weighted

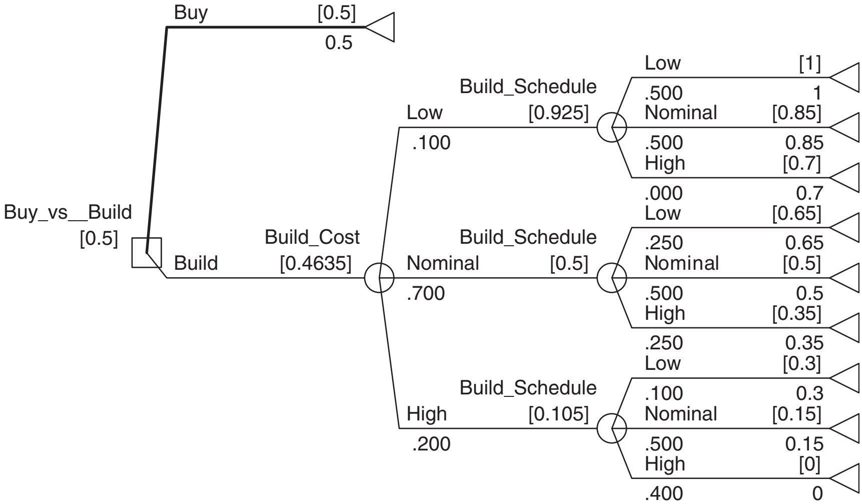


FIGURE 13.7 Decision tree for buy vs. build decision.

values are summed at each chance node to get an expected value at that node. These expected values are then multiplied by the probabilities on the branches before them and summed again. This process continues until the expected value of each alternative is available at the decision node. The preferred alternative should be the one with the highest expected value.

Influence diagrams are a graph-theoretic representation of a decision. Shachter [1986, 1990] presented the requirements and algorithms needed to transform an influence diagram from solely a communication tool into a computation and analysis tool capable of replacing the standard decision analysis tree. Significant additional research continues into influence diagrams for structuring decision problems, defining the underlying mathematics and graph theory of influence diagrams, and analyzing decision problems. When properly implemented, decision trees and influence diagrams provide identical solutions to the same problem. They are referred to as isomorphic since the decision tree can be converted to an influence diagram, and vice versa.

An influence diagram may include four types of nodes (decision, chance, value, and deterministic), directed arcs between the nodes, a marginal or conditional probability distribution defined at each chance node, and a mathematical function associated with each decision, value, and deterministic node. Each decision node, represented by a box, has a discrete number of states (or decision alternatives) associated with it; chance nodes, represented by an oval, must be discrete random variables. Deterministic nodes are represented by a double oval. A value node may be represented by a roundtangle, diamond, hexagon, or octagon.

An arc between two nodes (shown by an arrow) identifies a dependency between the two nodes. An arc between two chance nodes expresses relevance and indicates the need for a conditional probability distribution. An arc from a decision node into a chance or deterministic node expresses influence and indicates probabilistic or functional dependence, respectively. An arc from a chance node into a deterministic or value node expresses relevance; that is to say, the function in either the deterministic or value node must include the variables on the other ends of the arcs. An arc from any node into a decision node indicates information availability; that is, the states of these nodes are known with certainty when the decision is to be made.

Figure 13.8 shows an influence diagram for the buy versus develop decision described in the decision tree of Figure 13.7. The decision is represented in the box, the value node in the box with rounded corners, and the two chance nodes in ovals. Note that the alternatives and chance outcomes that were shown in the decision tree are not visible in the influence diagram. However, the edges in the influence diagram provide new information that was not readily available in the decision tree, namely the probabilistic and value dependencies inherent in the decision. Both cost and schedule are dependent on which alternative is selected. Cost and schedule are also probabilistically dependent on each other, with the influence diagram showing an arc from Build Cost to Build Schedule. Value only depends on cost and schedule.

The decision node represents a logical maximum (minimum) operation, that is, choose the alternative with the maximum (minimum) expected value or utility (cost). A deterministic node can contain any relevant mathematical function of the variables associated with nodes having arcs into the deterministic node. A value node also can contain any mathematical function of the variables with arcs entering the value node. In addition, the mathematical function in the value node defines the risk preference of the stakeholder.

A well-formed influence diagram meets the following conditions: (1) the influence diagram is an acyclic directed graph, that is, it is not possible to start at any node and travel in the direction of the arcs in such a way that one returns to the initial node; (2) each decision or chance node is defined in terms of mutually exclusive and collectively exhaustive states; (3) there is a joint probability distribution that is defined over the chance nodes in the diagram that is consistent with the probabilistic dependence defined by the arcs; (4) there is at least one directed path that begins at the originating or initial decision node, passes through all the other decision nodes, and ends at the value node;

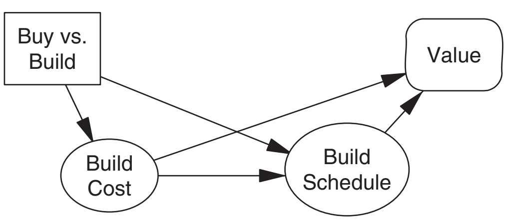


FIGURE 13.8 Influence diagram for build-buy decision.

(5) there is a proper value function defined at the value node (i.e., one that is defined over all the nodes with arcs into the value node); and (6) there are proper functions defined for each deterministic node. An influence diagram that is well formed can be evaluated analytically to determine the optimal decision strategy implied by the structural, functional, and numerical definition of the influence diagram. The analytic operations needed to evaluate an influence diagram numerically are evidence absorption, deterministic absorption, null reversal, arc reversal, and deterministic propagation [Shachter, 1986].

The influence diagram in Figure 13.9 shows an example of an influence diagram for a requirements allocation decision for the design of a new elevator system. The systems engineer is considering the use of one of two new technologies (power or controller); the large decision node (center left of Figure 13.9) defines the three alternatives. The requirements allocation (shown as three separate decision nodes) of costs, performance, and availability will be different if one or neither of these technologies is included in the design. Since this initial decision will be known when the three requirements allocation decisions are made, there are arcs from the initial decision node to the three requirements allocation decision nodes. The other arcs between the three requirements allocation decision nodes indicate the order in which the decisions will be made: performance, availability, and cost. (The decision maker is free to select any order among these three nodes.) These allocations and the prior uncertainty of the systems engineering team about the power and controller technologies will affect the uncertainty about the elevator's cost, performance, and availability. The arcs between the chance nodes are identical to those shown in Figure 13.5. Note, this diagram shows the uncertainty of elevator

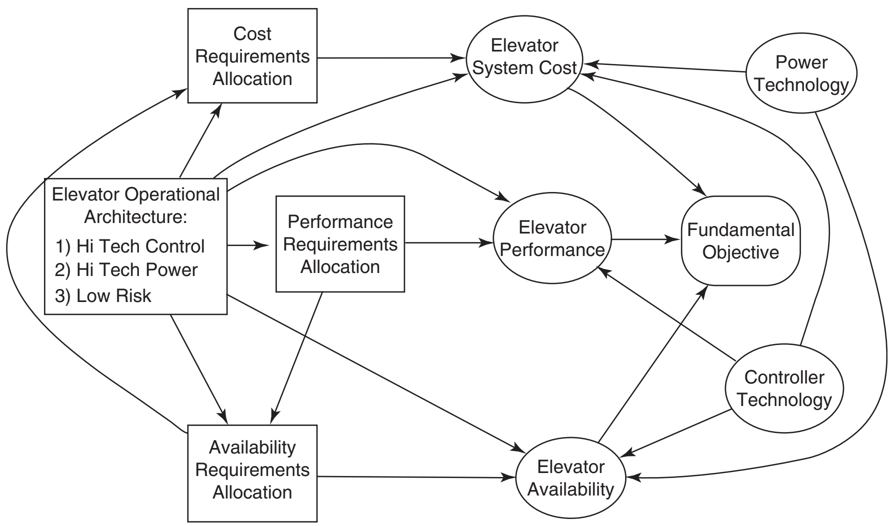


FIGURE 13.9 Sample influence diagram for requirements allocation.

performance to be independent of power technology. In this simplified example, the fundamental objective is comprised of three elements: cost, performance, and availability.

The results of a case study analysis of the above elevator architecture and requirements allocation decision are shown in Figure 13.10. First, the value functions for elevator performance (an index of various passenger waiting times), life-cycle cost, and availability and their weights are shown. Note that marginal decreasing returns to scale is shown in each curve as capability moves from the minimum acceptable threshold to the technological maximum. Next, the uncertainties associated with the two technologies in question are shown. The other uncertainties encoded as part of the analysis are not shown here. The analytical results show that the allocated architecture and the requirements allocation associated with the advanced power technology should be chosen to be consistent with the requirements (the value structure captured by the trade-off

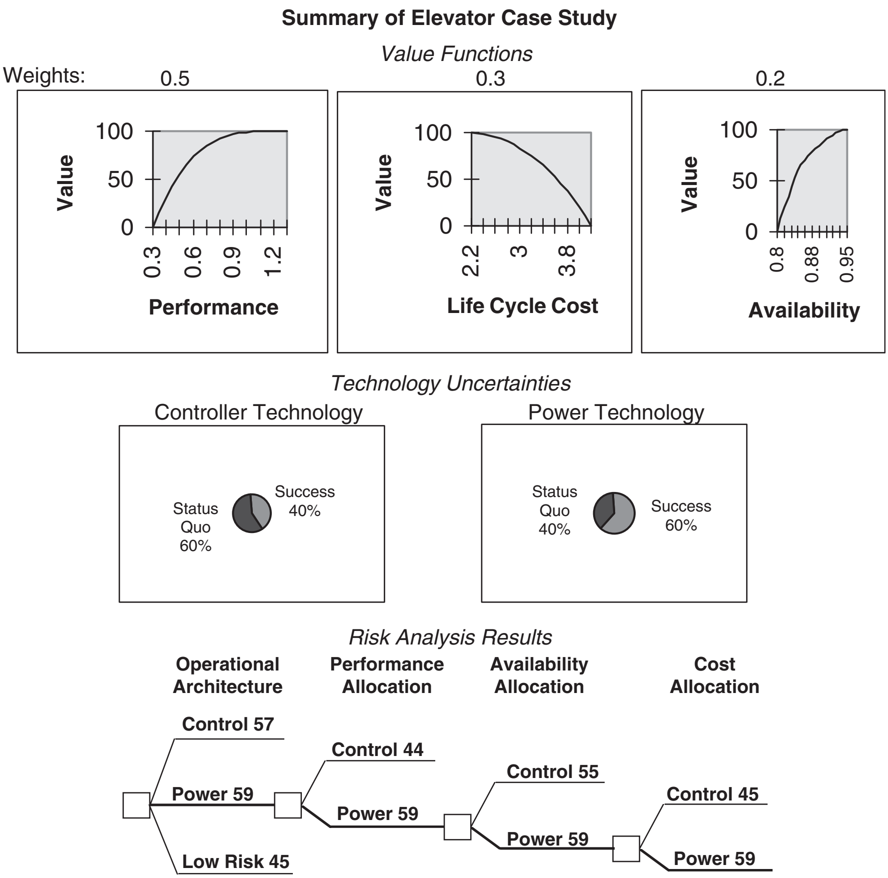


FIGURE 13.10 Summary of requirements allocation case study.

requirements) and the uncertainty about the technologies. The alternative associated with the control technology is very close; in fact, too close to be confident that the power technology is preferred given the limitations of value and probabilistic assessments. The low-risk alternative is clearly inferior; the design team could feel comfortable choosing either of the new technologies. The choice of technology would significantly change the requirements allocation decisions made in the three subsequent decision nodes.

13.5.4 Risk Preference and Expected Utility

Webster's dictionary defines risk simply as the "exposure to the chance of loss," and most people have at least an intuitive sense of what risk means to them. But from a decision-making perspective, it is essential to provide a more formal definition. The Defense Systems Management College (DSMC), [1989] in their *Risk Management Handbook*, defines *risk* as "the combination of the probability of an event occurring and the significance of the consequence of the event occurring" and defines risk management as "the various processes used to manage risk."

There are several strategies used for dealing with risk: avoidance, transference, management, and analysis. *Risk avoidance* is the selection of the low-risk alternative; unfortunately, what seems to be low risk intuitively is high risk in some cases. For example, consider a situation in which you have a sizable portfolio of U.S.-based stocks and are considering purchasing either another U.S. stock or what is considered a high-risk international stock. The international stock is often the lower risk alternative because its performance is either negatively correlated or uncorrelated with the performance of your portfolio while the performance of the low-risk U.S. stock is highly correlated with your current portfolio.

Risk transference involves options that transfer risk to others, an example being the purchase of insurance. The insurance purchaser is willing to pay a fixed price and have the insurance company take the risk of a major loss.

Risk management involves the use of hedging strategies; a hedging strategy is the maintenance of fallback options in case a riskier option fails. The failure is not catastrophic because the fall back option can be used. This is common in systems engineering when multiple contractors are asked to develop the same component; one contractor is pursuing the high-risk and high-performance approach that will be used if successful, while another contractor is pursuing a more conservative approach.

Risk analysis addresses risk explicitly when decisions are made in uncertain situations. Addressing the uncertainty faced in a decision by assigning probabilities to the uncertain outcomes, producing a lottery, has been discussed above. If the outcomes are measured on a numerical scale (e.g., dollars) that captures the value associated with the outcome, the expected value of the lottery is used as a measure of the attractiveness of the lottery. However, if the outcomes of the lottery are substantial compared to the wealth or well being of the decision maker, the expected value may not be an appropriate measure of the value of the lottery, as judged by many decision makers. The value associated with a lottery is called the *certain equivalent,* the value the decision maker would be willing to accept in place of the lottery. Since this notion of certain equivalence is a subjective judgment that is special to the individual (or set of stakeholders) and the context at the time of the decision, a mathematical description of risk preference must be guided by the feelings of decision makers.

A utility or risk preference function, u, is introduced to be a function of the outcome values of the lottery. If such a function exists, the inverse function of the expected utility of the lottery is the value of the certain equivalent of the lottery that can then be used to compare the attractiveness of the lottery with other lotteries. For example, consider the two lotteries in Figure 13.11 in which the outcomes are measured in dollars. The expected values (EV) of these two lotteries are:

$$\mathrm{EV}(1) = 0.5\ \times \$1000 + 0.5 \times \$0 = \$500$$

$$\mathrm{EV}(2) = 0.1 \times \$100{,}000 + 0.9 \times -\$10{,}000 = \$1000$$

These expected values indicate that lottery 2 is preferred to lottery 1; $\mathrm{EV}(2) > \mathrm{EV}(1)$. Yet many people, who cannot afford a loss of \$10,000, would prefer the first lottery with the lower expected value. In other words, for those people, the *expected utility* of lottery $1 >$ the expected *utility* of lottery 2, or

$$0.5u(\$1000) + 0.5u(\$0) > 0.1u(\$100{,}000) + 0.9u(-\$10{,}000) \qquad (13.12)$$

Mathematically, if the inverse function of $u(.)$ exists, then Eq. (13.12) can be restated as

$$u^{-1}[0.5u(500) + 0.5u(0)] > u^{-1}[0.1u(100{,}000) + 0.9u(-10{,}000)] \qquad (13.13)$$

The question is: "Will such a function generally explain the decision maker's risk preference judgments over all possible lotteries?" The two expressions on

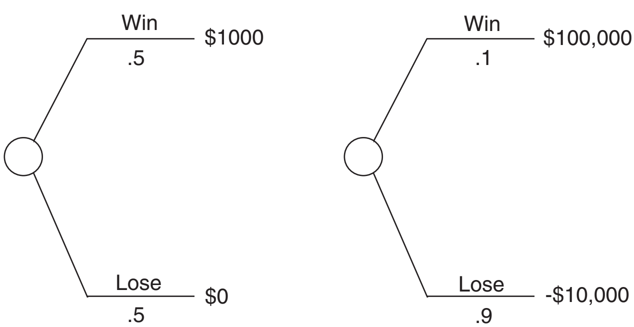


FIGURE 13.11 Comparison of two lotteries.

either side of the inequality in Eq. (13.13) are called the certain equivalents of the two lotteries.

The *risk premium*, x_p, of a lottery is defined to be the difference between the expected value of the lottery and the certain equivalent, $\tilde{x}$,

$$x_p = \bar{x} - \tilde{x} \tag{13.14}$$

For risk-averse decision makers the certain equivalent will always be less than the expected value and the risk premium will be positive.

13.5.4.1 Assessing A Risk Preference Function. Discussion of a risk preference function for a specific decision assumes that the outcomes of the decision have been characterized by a value function that collapses all dimensions of value onto one dimension, commonly called the *numeraire*. A money equivalent is the most common numeraire, but others are also possible. The risk preference function is then a function over the value numeraire.

There are two types of questions involving a certain equivalent and a two-outcome lottery that one can ask a decision maker during a risk assessment session. These two question types are shown in Figure 13.12. The first question type assumes the probabilities of the lottery are known and the decision maker

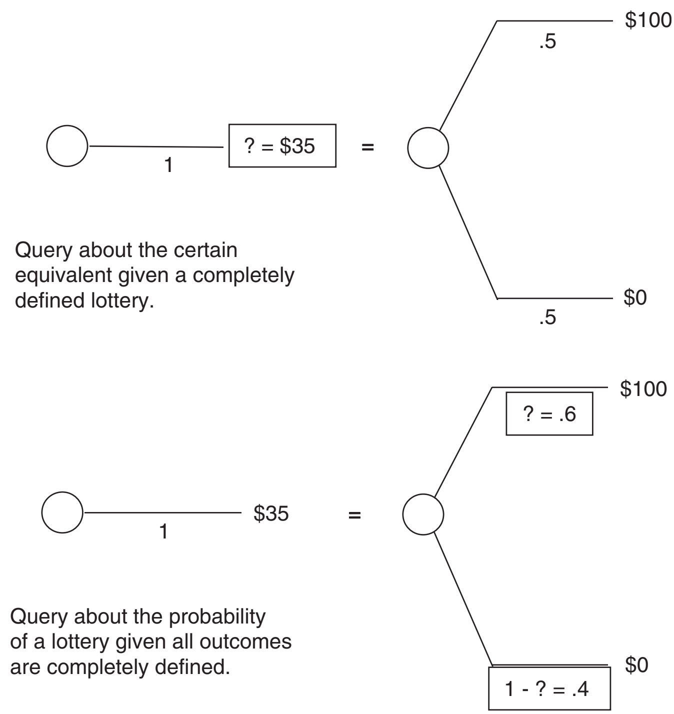


FIGURE 13.12 Simple risk preference assessment queries.

is asked to provide one of the outcome values, typically the value of the certain equivalent. However, one could fix the certain equivalent and ask for the value of either the best outcome or worst outcome. The second question assumes that all of the outcome values are known, including the certain equivalent, and the decision maker is asked to supply the probability value.

Unfortunately, research has shown that people do not provide coherent answers to these two types of queries. That is, in general the answers to the second question type are going to suggest much greater risk aversion than answers by the same individual to the first question type. A not uncommon response to the first query, which has an expected value of \$50, is \$35, yielding a risk premium of \$15. Now if \$35 is the certain equivalent in the second query, an individual might respond that the question mark for the probability of \$100 in the second lottery would be 0.6. The risk premium for this second lottery is \$25 (the expected value of \$60 minus the certainty equivalent of \$35).

The first question type is asking directly for the response that will be substituted into various analyses. Therefore, it is somewhat more appropriate to ask this question. However, very few decision makers have thought seriously about these issues in general, and even fewer have thought about them with respect to a specific decision situation. The assessment process is therefore a learning experience for the decision maker. The responses to the early questions should be treated as a warm-up process.

A second caution for the risk assessment process is that there is a very substantial zero effect. That is, people exhibit risk-averse behavior for gains but risk-seeking behavior for losses. Figure 13.13 shows responses for a certainty equivalent that demonstrates this behavior. The risk premium is \$15 for the top lottery and −\$15 for the bottom lottery. The risk-averse person in the top lottery would have a certain equivalent of less than −\$50 for the bottom lottery. Generally, people do not want to exhibit this "zero effect" once the seeming contradiction is pointed out to them and will switch to a consistent risk-averse (or risk-seeking) policy.

To investigate the decision maker's risk preference fully in the region of outcomes associated with the current decision, multiple lottery questions should be asked in this region. For illustrative purposes, suppose the decision involves gains of up to \$10,000 and losses as great as \$10,000. We arbitrarily set the end points of the utility scale as u(\$0) = 0 and u(\$10,000) = 1. Figure 13.14 provides six such lotteries and the responses of the decision maker shown in the boxes. Note that the utilities shown under each figure are calculated as in the following example:

$$\begin{aligned} u(\$2{,}500) &= .5\,u(\$10{,}000) + .5\,u(\$0) \\ &= .5\,(1) + .5\,(0) \\ &= .5 \end{aligned}$$

Figure 13.15 displays the resulting risk preference function. Note the decreasing rate of increase associated with this curve, mathematically known as a concave

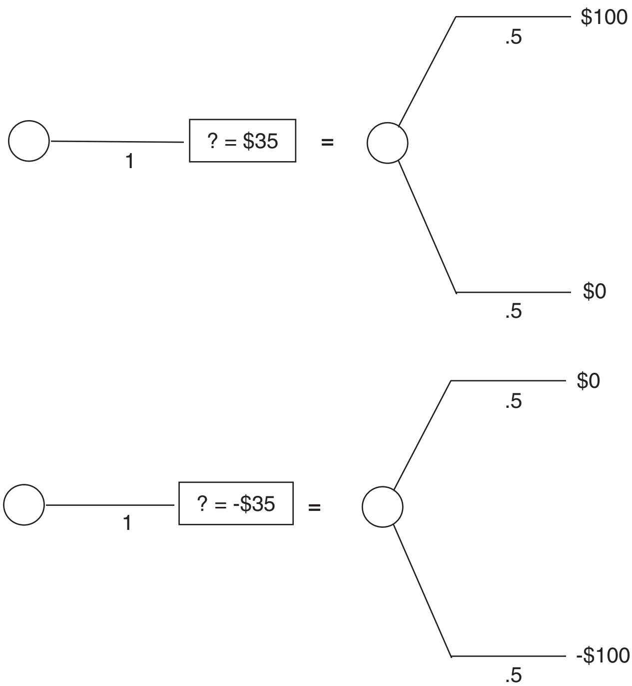


FIGURE 13.13 Illustration of the zero effect.

curve. A risk-neutral decision maker would have a straight line as a risk preference function; risk-seeking behavior is typified by a convex curve.

13.5.4.2 Exponential Risk Preference. Define the risk aversion coefficient $\gamma = -u''(x)/u'(x)$. If γ is a constant, it can be shown by simple integration that the risk preference function must take the form

$$u(x) = \begin{cases} k_1 x + k_2, & \text{if} \quad \gamma = 0 \\ k_1 e^{-\gamma x} + k_2, & \text{if} \quad \gamma \notin 0 \end{cases} \tag{13.15}$$

A common way to write such a risk preference function is

$$u(x) = \frac{1 - e^{-\gamma x}}{1 - e^{-\gamma x_{\max}}}, \tag{13.16}$$

where $x_{\max}$ is the largest value that x is expected to take. Thus, for any valued outcome x, the utility of x can be calculated using the exponential utility function. Note that this format produces

$$u(x_{\max}) = 1.0$$

$$u(0) = 0$$

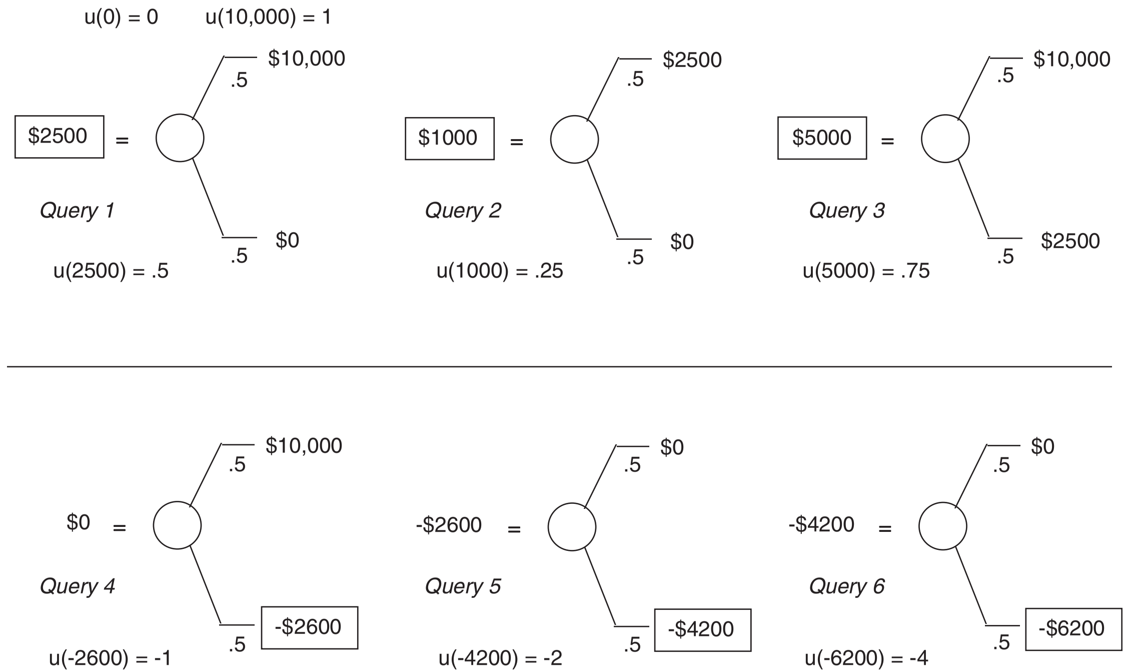


FIGURE 13.14 Assessment queries for risk preference function.

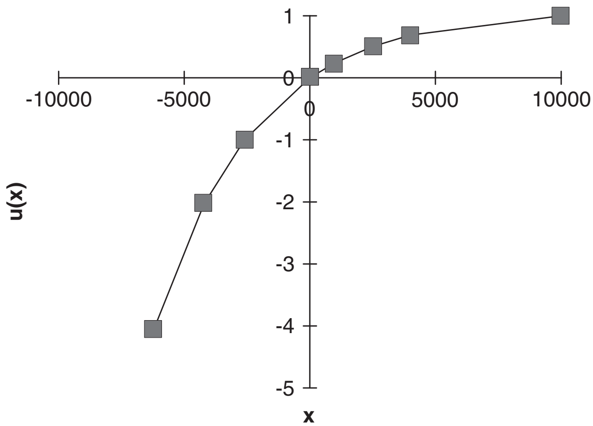


FIGURE 13.15 Assessed risk preference points.

The risk preference function plotted in Figure 13.15 is an exponential risk preference function with $\gamma = 0.00025$.

Another important concept in risk preference is the *risk tolerance,* or the inverse of the risk aversion coefficient. For the exponential risk preference function and its constant risk aversion coefficient, the risk tolerance is constant. In Figure 13.15, the risk tolerance is \$4000. For an expected value decision maker the risk aversion coefficient is zero, making the risk tolerance infinity.

The exponential risk preference function has another very special property, called the *delta property* This property is stated as follows: An increase in all outcomes of the lottery by a constant amount, Δ, results in an increase of the certain equivalent by the same amount, A. So, for example, in the first example above suppose that the certain equivalent for fifty — fifty gamble of \$100 and \$0 was \$35. Now, if each prize is increased by \$100 and the certain equivalent of a fifty — fifty gamble on \$200 and \$100 becomes \$135, then the delta property is satisfied for at least this one case. The exponential risk preference function is the only function that can satisfy this property.

One very important implication of the delta property is that the buying and selling prices of a lottery are the same. For example, the maximum that a decision maker was willing to pay, *B*. for a lottery is the amount that when subtracted by every outcome made us indifferent to having the lottery and not having it, or a value of \$0. Similarly, the minimum that the decision maker would sell the lottery for, *S,* is its certain equivalent; also see Figure 13.16. If the risk preference function is exponential, it can be proven that $B = S$ through the use of the delta property. For other risk preference functions the buying and selling prices of a lottery are not necessarily equal.

There is a “quick and dirty” method for assessing a decision maker’s risk aversion coefficient for an exponential utility function. The value of *R* for which the decision maker is indifferent to accepting the lottery in Figure 13.17

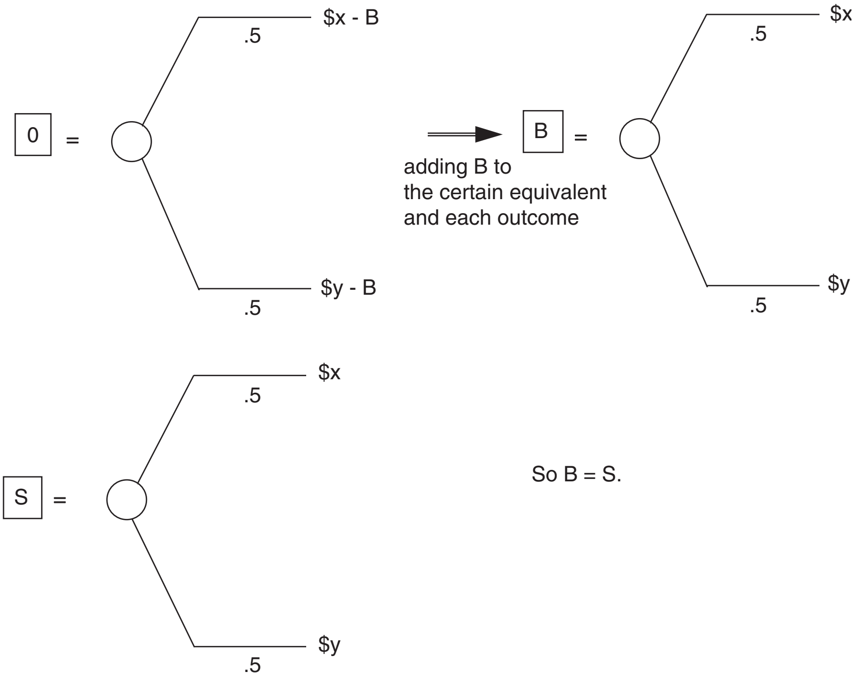


FIGURE 13.16 Buying and selling prices are equal for exponential risk preference.

is the risk tolerance. That is, the certainty equivalent of the lottery in Figure 13.17 is 0 when R is the risk tolerance of the decision maker. It can be shown that $\gamma = 1/R$.

The exponential risk preference function is used as an approximation early in risk analyses to determine the effect of risk preference on the choice of alternatives. If this choice is sensitive in the appropriate region of the decision maker's risk tolerance, then more detailed analysis of the decision maker's risk preference is appropriate.

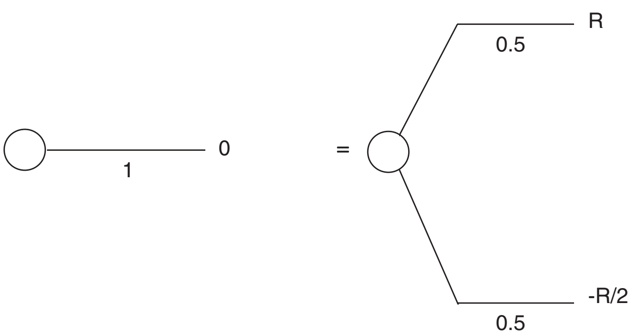


FIGURE 13.17 Risk aversion coefficient lottery.

13.6 SAMPLE APPLICATION

This application demonstrates how decision analysis can be used in the requirements development process of systems engineering. The requirements development process consists of the development of an operational concept, identification of the external systems that interact with the system and the context in which the system operates, an objectives hierarchy for the system's performance, and the requirements. These requirements are divided into requirements categories of input/output, system-wide and technology, trade-off, and test. The focus of this application is the use of multiattribute value analysis as the approach for defining the trade-off requirements that comprise the value model to be used by the stakeholder in evaluating the available alternatives. Implicit in this approach is an objectives hierarchy for defining the value space of the stakeholder (see Sidebar 13.4). Also included is the mathematical structure for the trade-off requirements.

SIDEBAR 13.4: ECONOMIC MODELS

Hazelrigg [1996] provides strong motivation to use decision analysis tools in systems engineering design decisions. In his treatment he addresses the results of Arrow's impossibility theorem [Sen, 1970] for achieving group consensus on preferences and recommends the use of the demand function from economics for defining consumer preferences for alternate design alternatives. The issue of gaining stakeholder consensus on trade offs needed during design is real; thus the systems engineering team must resort to accepting the position of one stakeholder (the bill payer) as king when these disagreements cannot be resolved. This was the method used in the application presented in this section.

The notion of a demand function for a military system is not helpful. However, for a commercial system the multiattribute value function can be considered to be a first-order, Taylor series approximation of the demand function. Hazelrigg [1996] does not go into detail about how to obtain the demand function; the suggestion made in this book is to elicit stakeholders' preferences and use the bill payer as king or queen to resolve disagreements.

Throughout this discussion a system called the Mobile Protected Weapons System (MPWS) is used to describe the development of the system engineering and decision analysis concepts. The MPWS was to be a helicopter-transportable, direct-fire support weapons system for the U.S. Marine Corps (USMC), with an initial operating capability of 1988. The basis of the example was a real application of decision analysis to the MPWS in 1980. After the evaluation structure embodied in the objectives hierarchy and trade-off requirements discussed below was used to evaluate proposed MPWS designs, the MPWS was stopped in favor of purchasing similar vehicles "off-the-shelf," as directed by Congress. The contractors who received the objectives hierarchy and trade-off requirements as part of the Request for Proposal were very complimentary of the USMC for providing this information to guide their design decisions.

13.6.1 MPWS Overview

An intuitive need for a highly mobile, helicopter-transportable weapons system that can provide the landing force assault fire support as well as an antiarmor capability first became apparent to the USMC in the early 1970s. There were several contributing factors:

- Naval gunfire support assets, so important during an amphibious assault, were steadily decreasing.
- Navy combatant ships with suitable guns for shore bombardment were being retired without replacements or being replaced with ships less capable of providing gunfire support to amphibious forces.
- The retirement from the Fleet Marine Force (FMF) of the ONTOS, a light, mobile, antitank weapon system carrying six 106-millimeter (mm) recoilless rifles.
- The retirement of the crew-served individual 106-mm recoilless rifle.
- The deletion of the 3.5-inch rocket launcher from the Marine Corps inventory.
- At a time when naval gunfire and direct-fire weapons were decreasing, the Soviet and Soviet aligned forces increased their capability with a wide array of armored weapons systems, including tanks, armored personnel carriers, and lightly armored weapons platforms.

In accordance with acquisition procedures contained in Circular A-109 of the U.S. Office of Management and Budget, Mission Area Analysis (MAA) was continuous, and a Mission Element Needs Statement was developed stating that:

- Amphibious forces possess capabilities that are uniquely featured by their responsiveness to the maritime aspects of the national strategy. Amphibious warfare requires the full spectrum of capabilities from naval combat effectiveness offshore and in the air to the close combat mission ashore. The close combat capability provides the mobility, shock action, and portions of the firepower necessary to enable landing forces to successfully attack and destroy enemy personnel and materiel, breach their defenses, link up surface-borne with helicopter-borne forces, defeat infantry and mechanized counterattacks, and exploit success in combat ashore.

- Capabilities currently possessed by the landing force provide limited mobility and direct fire combat power to enable assault units to rapidly close with and destroy enemy forces. Mobility and direct fire support capabilities required to enhance current capabilities are:

  a. Helicopter transportability of weapons systems by heavy-lift helicopter
  b. Vehicle and crew survivability through armor protection from nearby artillery airbursts and medium-caliber direct-fire weapons firing at medium range
  c. Rapid cross-country mobility, agility, and endurance without significant degradation of on-road capability and capable of competing with the expected mobility of the threat
  d. An on-board weapons suite with a long-range, high-kill probability capability against armored, light armored, materiel, and personnel targets characteristic of the threat
  e. The ability to engage and defeat the target spectrum in all weather conditions
  f. Nuclear, Biological and Chemical (NBC) detection and protection

The Marine Corps requirements defined an affordable weapons system that was to be highly mobile, helicopter-transportable, compatible with amphibious operations, and able to provide direct-fire support during landing force operations. The weapons system must provide protection from suppressive fires and be capable of engaging and defeating armored, personnel, and materiel targets.

13.6.2 Operational Concept for MPWS

In defining the mission needs for the MPWS, three employment scenarios were considered. These scenarios represent the spectrum of scenarios that drives the design of MPWS. The relative importance of each parameter in the design process changes as a function of scenario.

Scenario 1: *Offensive Role (assault support with the infantry)* MPWS would be used with the infantry in offensive operations. A red/blue force ratio of 1:4 and a northern NATO environment are established as the base for the determination of relative capability requirements in this scenario.

Scenario 2: *Defensive Role (blocking position)* MPWS would be employed with helicopter-borne forces to establish blocking positions. Friendly tanks are not available. The mission calls for delaying the enemy and channelizing his avenues of approach. It is assumed that enemy forces are mechanized to include T62, T64, and T72 tanks, BMP, BTR, assault guns, SP artillery, and attack helicopters. MPWS will be operating at altitudes higher than sea level. A red/blue force ratio of 4:1 in a Middle East environment is established as the base in this scenario.

Scenario 3: *Subsequent Operations* MPWS would be employed with a combined arms task force and would no longer be in an amphibious assault role. Blue forces are task organized, and there would most likely be low-mid-intensity nonnuclear conflict. Red/blue force ratio of 1:4 and a Middle East/Third World environment are the requirements determination base.

13.6.3 External Systems Diagram

The external systems of the MPWS during its operational and maintenance phase would be the operators (driver, gunner, and passengers), maintainers, targets (light armored vehicles, tanks, personnel, and helicopters), and a heavy lift helicopter that would have to transport the MPWS.

Figure 13.18 is an external systems diagram showing the inputs to and outputs from MPWS for the various external systems. This diagram was completed using the IDEFO Integrated Definition for Function Modeling process modeling (see Chapter 3). Four external systems are shown in Figure 13.18; the MPWS operators, the MPWS targets, the heavy-lift helicopter that will carry the MPWS from point to point, and the MPWS maintenance personnel. The interaction between the MPWS and its operators is shown by the three arrows; two leaving the operators' function and one leaving the MPWS function. Terrain forces are shown as part of the context, entering the MPWS function as input from outside the set of external systems. The primary benefits of this analytical construct are to bound the MPWS system very specifically by showing where MPWS ends and other systems begin, and to specify the inputs to and outputs of MPWS so that requirements can be defined to make these inputs and outputs possible.

Figure 13.19 portrays an objectives hierarchy similar to the one developed by a team of USMC experts and the decision analysts working the project. The three operational scenarios are the first decomposition of the hierarchy because the principal objectives of the USMC for the MPWS had different relative importance depending upon the scenario. The top-level objectives, or measures of effectiveness (MOEs), were firepower, mobility, availability, and survivability. Firepower was broken into measures of performance (MOPs): lethality, servicing rate, stowed kills (a combination of the number of stowed rounds and the lethality of those rounds), and target acquisition. Lethality is composed of the various types of targets, followed by the ranges at which those targets would be engaged. Target acquisition is composed of identification and recognition in good weather as well as the bad weather capability. Mobility is broken into capabilities related to cross-country, long-distance airlift, road, and water. Survivability is measured by means proxies for agility, protection, and signature.

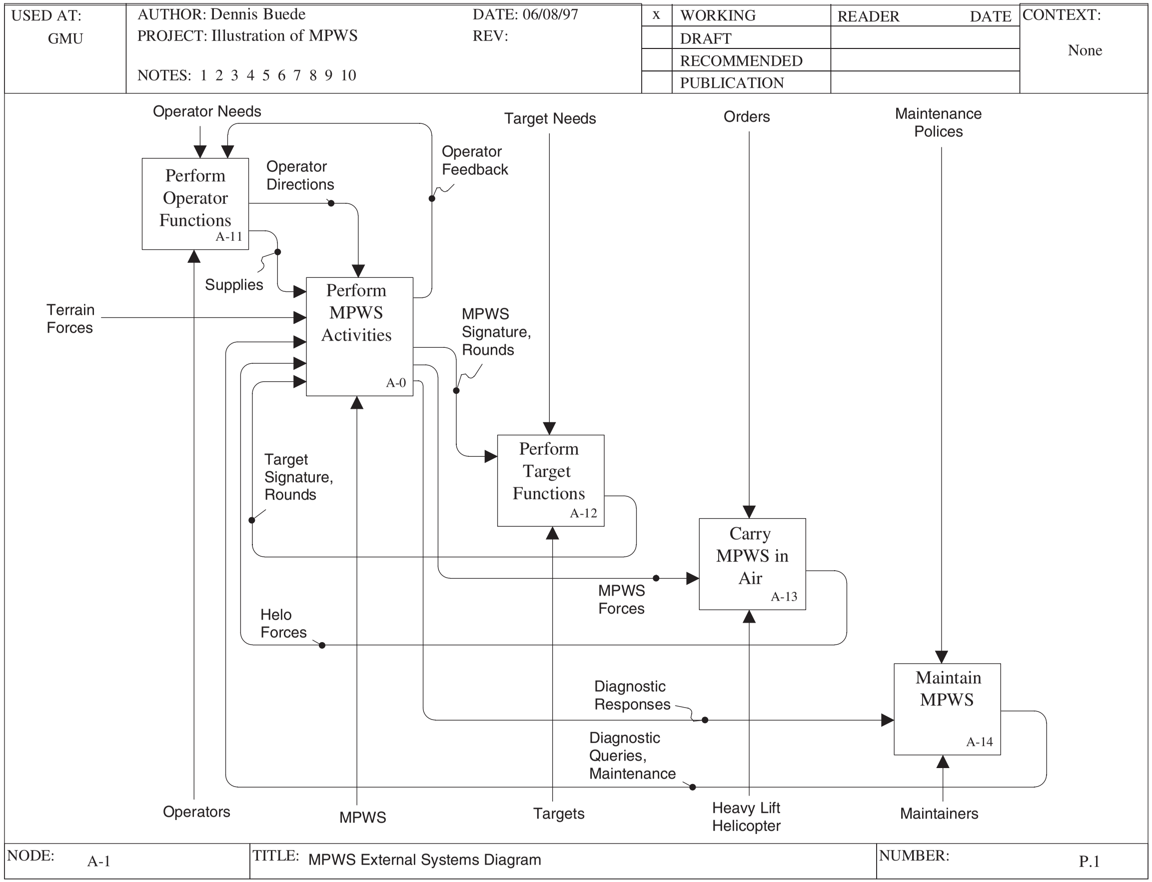


FIGURE 13.18 External systems diagram for MPWS.

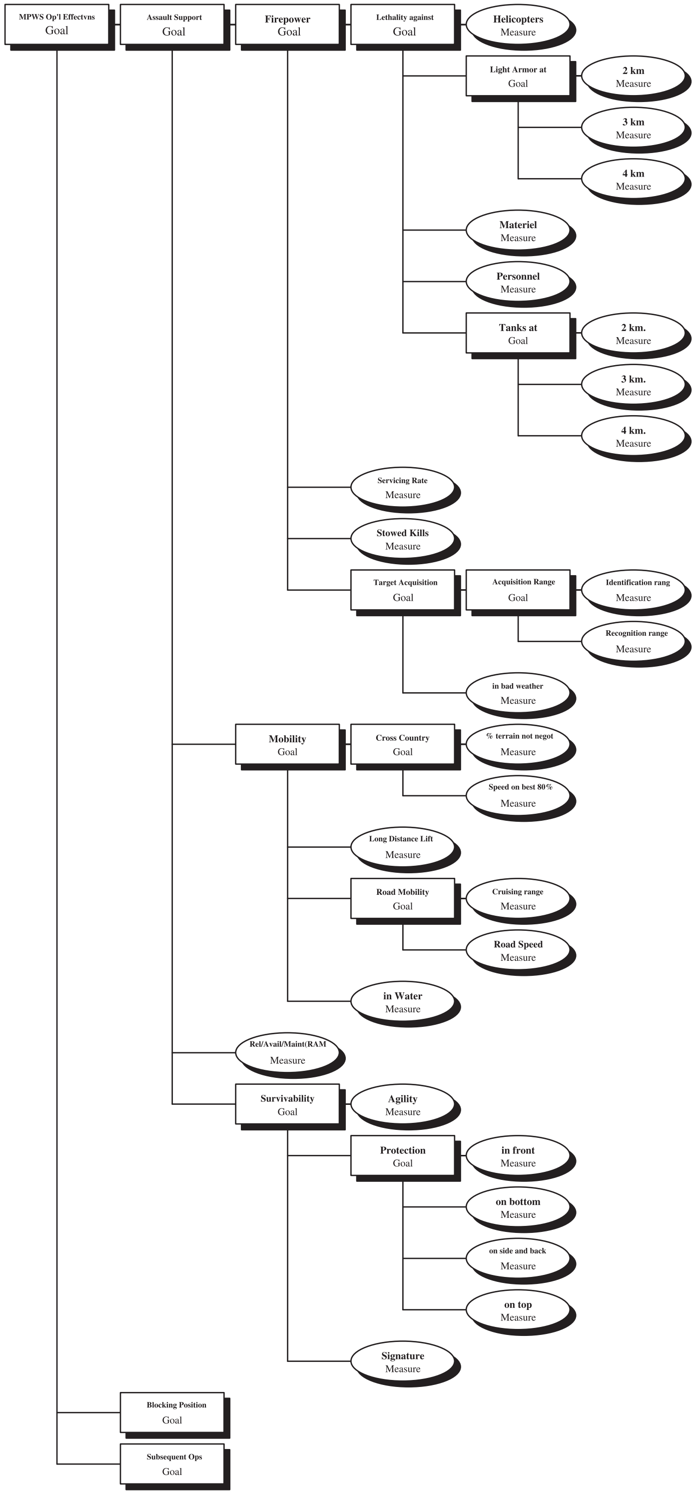


FIGURE 13.19 Operational effectiveness performance parameters.

13.6.4 Requirements

The focus of this application is the set of requirements called trade-off requirements, algorithms for comparing any two alternate designs on the aggregation of cost and performance objectives. As discussed in Chapter 6, these algorithms are divided into (a) performance trade offs, (b) cost trade offs, and (c) cost-performance trade offs.

In the development of requirements for MPWS substantial attention was devoted to the trade-off requirements for performance. The structure that describes the mission-related objectives on which these performance trade offs were defined is the objectives hierarchy shown in Figure 13.19. The trade-off requirements consist of a utility or value curve for each bottom-level objective and a set of weights at each branch in the tree.

13.6.4.1 Utility Curves. Figure 13.19 portrays the many operational effectiveness variable performance parameters whose utility for improvement were quantified for guidance by the USMC committee. Inherent in these value or utility curves for the many performance parameters is the notion that design trade offs are acceptable within the 0-to-100 range of utility; that is, MPWS performance in some area can be sacrificed to the point of zero marginal utility, but no further, in order to achieve performance gains in other areas. The zero utility point on each performance parameter does not mean that a system with this capability has no utility to the Marine Corps. Rather, it means that this level of performance is the minimum acceptable to the Marine Corps across its range of missions. So, for example, to be helicopter-transportable the MPWS must not weigh any more than 16 ton at 3000 feet on a 91.5°F day. The utility curve for helicopter transportability is shown in Figure 13.20. Increased performance for each parameter has value to the Marine Corps as shown by the shape of the utility curves.

The shapes of these utility curves are the same for all of the above scenarios. However, the relative values of improvements in one parameter compared to improvements in another parameter do not vary across the three scenarios. These relative values of performance parameter improvements are described in Section 13.6.4.2.

13.6.4.2 Weights. Improvements in performance determined from the curves for each parameter are not equally important in the overall analysis of an MPWS. Therefore, a weighting procedure is applied to define the relative value of improving from the 0 to the 100 level of utility on one performance objective compared to another. The meaning of the weights can be described as follows: the weight given to parameter A reflects how much more valuable it is to improve from a score of 0 to 100 in parameter A as compared to the improvement in parameter B from 0 to 100. Note weights are not a generic measure of value but are dependent upon the swings from 0 to 100 on the associated utility curves.

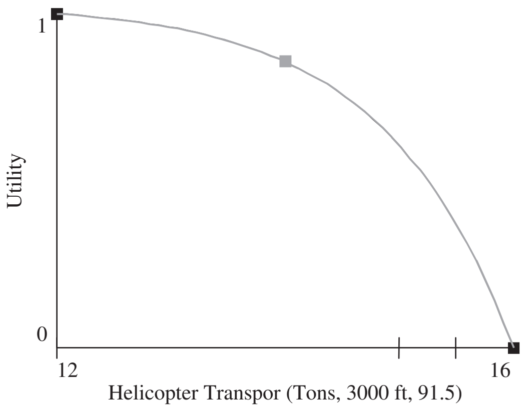


FIGURE 13.20 Utility curve for helicopter transportability, measured in tons.

For MPWS, weights played a large role in distinguishing among scenarios. While the shapes of utility curves remain constant across scenarios, their relative importance changed significantly. For example, an improvement in utility for helicopter transportability was very important in the blocking position role since the MPWS might have to be lifted into position. This same improvement was far less important in the subsequent operations role since the force would be traveling over land. Therefore, the weight that helicopter transportability has, relative to other operational effectiveness factors, was greater in the former role than in the latter.

13.6.5 Use of Utility Curves and Weights

Value (or utility) curves and weights can be used as follows: the abscissa (x axis) of each curve is a measurable attribute that provides input to the curve. The ordinate *(y* axis) is a measure of relative value or utility ranging from 0 to 100. As an example, value or utility curves for V80 and Percent No-Go are shown in Figures 13.21 and 13.22. Note that an improvement in V80 from 10 to 15 mph is valued as highly as a gain from 15 to 25 mph. Both improvements would net 50 utility points. Using these curves, a candidate propulsion system yielding a V80 speed of 15 mph would receive 50 utility points while one with a V80 speed of 20 mph would receive 80 points; a candidate with 6% No-Go scores 85 while one with 16% scores 35.

These value or utility scores would not be very meaningful for comparing systems without a relative measure of importance between attributes. Thus, a weighting procedure is applied to the scores to allow evaluation based upon a combination of parameters. Again, consider the value or utility curves illustrated in Figures 13.21 and 13.22: Suppose propulsion system 1 yields a V80 speed of 15 mph and Percent No-Go of 6%, while propulsion system 2 had values of 20 mph and 16%. System 1 scores would be 50 and 85, while system 2

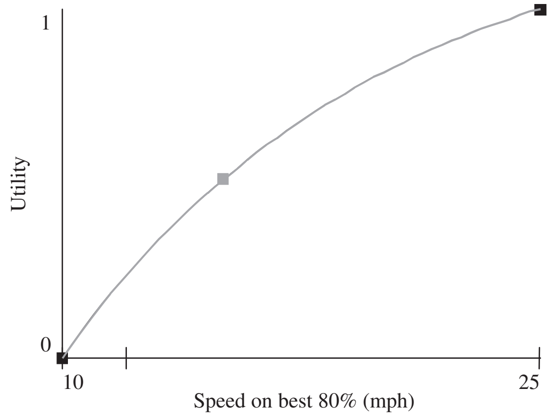


FIGURE 13.21 Utility curve for V80, speed on the best 80% of terrain.

scores would be 80 and 35. If both V80 and Percent No-Go were equally important, the weighted scores for both systems would be:

$$\text{System 1} : 1/2\,(50) + 1/2\,(85) = 67.5$$
$$\text{System 2} : 1/2\,(80) + 1/2\,(35) = 57.5$$

This would indicate that propulsion system 1 was superior on these factors. However, if V80 was considered to be two times as important as Percent No-Go, the weighted scores would be:

$$\text{System 1} : 2/3\,(50) + 1/3\,(85) = 61.7$$
$$\text{System 2} : 2/3\,(80) + 1/3\,(35) = 65$$

In this case, propulsion system 2 would be better.

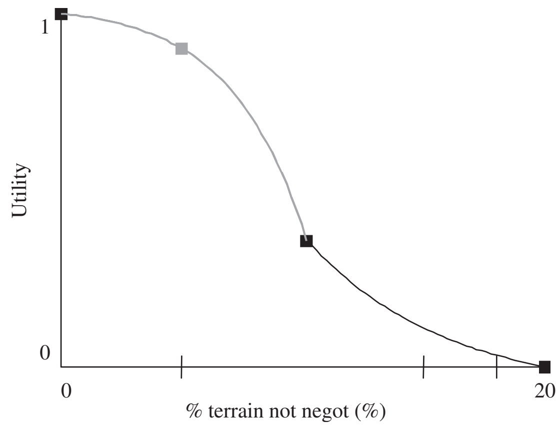


FIGURE 13.22 Utility curve for % No Go, % of terrain that is not negotiable.

It should be clear that the relative weights of the objectives play a major role in the design and evaluation processes.

13.6.6 Conclusions

As discussed in Chapter 6 the requirements development process is a systematic one that considers how the system is to be used, how the system is going to interact with other systems and the general environment, and the user objectives and priorities. Since user objectives and priorities are inherently subjective, the ultimate requirements for the system have to be subjective, reflecting trade offs of the users. This is not to say that substantial analysis is not critical to the development of good requirements. In the case of the MPWS the USMC used a great deal of analysis about alternate sites around the world in which it might be involved in conflict and the capabilities of the CH-53E helicopter to develop the utility curve for helicopter transportability and its relative weight to other performance objectives. By using many analysis techniques and a broad base of experts, logical and explicit statements of requirements were developed based upon informed consensus. The appropriate, detailed requirements inputs to the process can be obtained at lower organizational levels using appropriate experts and analyses, yet the more difficult, high-level requirement questions can be addressed at the highest levels of the organization.

13.7 SUMMARY

This chapter has introduced the complexities associated with decision making in general and addressed the difficulty of decision making in the engineering of a system. With respect to engineering a system, the definition of clear and meaningful alternatives for the design and integration of a system involves the use of sophisticated processes and modeling techniques as described in the first 12 chapters of this book. The development of the value structure for selecting design and integration alternatives was discussed in Chapter 6 and involves complex trade offs across stakeholders and stages of the system's life cycle. Finally, there is significant uncertainty regarding the relative effectiveness and cost of competing technologies as well as future needs of the stakeholders.

The axioms of decision analysis, as presented in this chapter, provide a sound basis for a coherent, rational decision-making process that incorporates meaningful approaches for addressing value trade offs and uncertainty. Multiattribute value analysis, a product of the axioms of decision analysis, uses value functions and weights to quantify the trade offs across objectives. These value functions and weights require that the stakeholders answer questions that have meaningful interpretations to them in terms of the decision being made; the quantification of values is not an ad hoc set of numbers producing an index of goodness.

Dealing with uncertainty is a difficult problem; decision analysis relies upon probability theory to capture the uncertainty faced by the decision maker. In the engineering of a system the uncertainty is not often described by existing data and interpretable as the long-run frequency of a set of known events. Instead the uncertainty deals with processes that change with time and for which no (or at most a few) known events have occurred. Instead of ignoring the uncertainty faced in the engineering of a system, decision analysis permits the engineers to capture the expert judgment of the engineers, stakeholders, and other experts and use this information to provide insights about the design choices with the best information available at the time. Recent advances in decision analysis provide graph-theoretic models for representing probabilistic dependence (relevance diagrams) and decisions problems (influence diagrams).

Once uncertainty is modeled explicitly, the risk preference of the decision maker has to be addressed as part of decision analysis. The concepts of risk aversion, neutrality, and preference are defined mathematically and illustrated as part of the decision analysis process. Using the decision maker's risk preference requires computing the certainty equivalent as the inverse of the utility function.

Clearly, it is inappropriate to use the sophisticated tools of decision analysis for every decision that is part of the engineering of a system. Many times engineers have described the benefit of thinking about the decision in the terms of decision analysis. At other times developing the value model and using a quick scoring and weighting evaluation provides insight into which alternatives are serious and which should be ignored. For really complex and contentious decisions, the full power of decision analysis can provide an explicit and rational process for defining and discussing the alternatives to reach a conclusion consistent with the values of the stakeholders and the uncertainty as defined by relevant experts.

PROBLEMS

13.1 In defining reliability of a system, we talk about the probability of a failure. Failure here is an event or distinction, but not one that passes the clarity test. As a result, systems engineers work very hard to focus on the distinction, mission failure, where a mission failure is a failure that precludes the user from completing her/his mission. This definition still does not pass the clarity test because we have not defined the mission, a definition that is system and context dependent.

For the elevator system where you work or go to school,

a. Define mission in a way that meets the clarity test.

b. Define as many failures as possible and show which would be classified a mission failure. Be sure to keep the clarity test in mind when defining these failures.

c. Discuss whether it is sufficient to discuss failures one at a time or whether it is necessary to examine possible combinations of failures to define fully all *possible* mission failures.

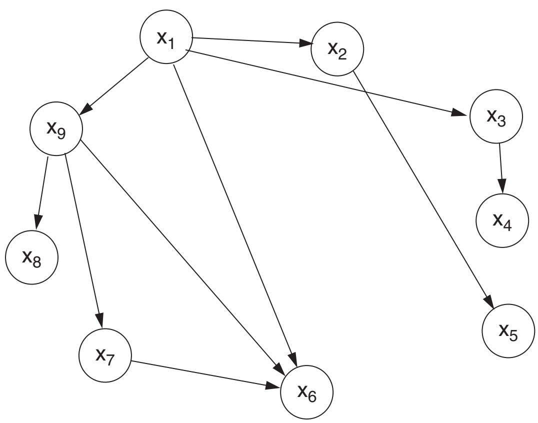
13.2 Garbled Communications, Ltd. is designing a new system for special-purpose use that only requires three signals to be sent and received. The derived requirements below list the probability that signal s_i is received given that signal s_i is sent:

p(s_j received | s_i sent, &)	Receive s_1	Receive s_2	Receive s_3
Send s_1	0.80	0.10	0.10
Send s_2	0.05	0.90	0.05
Send s_3	0.02	0.08	0.90

For the operational concept each signal is equally likely to be sent. The stakeholders' requirement for this scenario is that each signal should have a 0.85 probability of being sent given that it was received. Is this requirement met if these derived requirements can be satisfied? Note the symbol "&" on the right-hand side stands for all prior information.

13.3 Garbled Communications, Ltd. has begun producing its new communications system and has built three assembly lines: LI, L2, and L3. Ll is the most productive, accounting for 40% of the production; L2 is the least productive, accounting for 25%. L3 accounts for the rest. Test data show that L1 has a 2% chance of producing a lemon, L2 a 4% chance, and L3 a 3% chance. What is the probability that a lemon picked at random will come from each of the assembly lines?

13.4 Write the joint probability distribution that is consistent with the relevance diagram shown below.



13.5 Create a relevance diagram that is consistent with the following joint probability distribution:

$$
\begin{aligned}
&p(x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8|\&) \\
&\quad = p(x_8|x_7, x_5, \&)\, p(x_7|x_6, x_5, x_4, \&)\, p(x_6|x_3, \&) \\
&\qquad \times p(x_5|x_2, x_1, \&)\, p(x_4|x_3, x_1, \&)\, p(x_3|\&)\, p(x_2|\&)\, p(x_1|\&)
\end{aligned}
$$

13.6 You have been tasked with providing a recommendation for a test site at which an acceptance test will be conducted. There are three possible test sites (A, B, and C). Site A is the preferred site during good weather. Site C is the least preferred. Unfortunately, there is a long-range weather forecast for 3 months from now when the test needs to be conducted. The weather forecasters described the possibilities for weather as "good," "fair," and "poor." These possibilities have been defined very carefully and their forecast for the time period of the test is: 0.3 for good, 0.6 for fair and 0.1 for poor.

You have tried to find a way to reserve site A for a long enough period of time that the weather will certainly be good. However, site A is used by many people, and management has determined that the project cannot afford to rent site A for this extended time period. The cost at which the sites can be reserved for the time period in question is \$1000 for site A, \$700 for site B, and \$400 for site C.

Usage of each of these sites has varying positives and negatives for being able to analyze the results and recommend that the system be accepted. You have queried your colleagues to determine how much they would be willing to pay to change a specific site in the different weather conditions to the preferred site A and weather condition. These relative dollar values do not include the cost of renting the site for the needed time period. The relative dollar value equivalents for sites and weather conditions are shown below:

	Weather Is Good	Weather Is Fair	Weather Is Poor
Site A	\$1000	\$200	\$0
Site B	\$950	\$300	\$200
Site C	\$500	\$450	\$300

That is, site A in good weather is worth \$1000 more dollars in terms of test performance than it is in poor weather. Similarly, site A in good weather is worth \$500 more than site C in good weather.

a. Draw the influence (or decision) diagram for this problem.
b. Draw the decision tree for the problem.
c. Compute the expected values for the three sites to determine which site should be recommended.
d. What is the value of perfect information for weather? Show the influence diagram and decision tree for computing the value of perfect information.
e. Using the following u curve, what is the best expected utility decision?

$$u(x) = (1 - e^{-0.01x})/(1 - e^{-0.01})$$

where x is the total monetary value associated with using the site in question.
f. What is the value of perfect information using the above u curve.

13.7 As part of the management group of the systems engineering team, Bill D. Orby has been given the task of recommending whether to "build" or "buy" a particular component. Bill has called several manufacturers of this component and found the best "buy" alternative will cost \$200,000 for the quantity needed. The performance of this component that is available from outside is categorized as moderate; this categorization includes many performance parameters and is rather coarse, but Bill hopes sufficient for an initial analysis.

Next, Bill spent significant time talking to several design engineers within his company who would be given the task of building this component, and several others who have built similar components in the past. There is uncertainty concerning both the cost and ultimate performance of this component if it is built by Bill's organization. Bill has modeled the uncertainty about total cost for developing and building the total quantity of the component as follows:

Build Cost	Probability
\$100,000	0.2
\$200,000	0.6
\$400,000	0.2

The performance of the built component expected by the engineers with whom Bill spoke is substantially greater than the performance to be provided by the bought component. Bill has devised three performance categories to describe the uncertainty surrounding the built component: low, moderate, and high. The assessed probabilities of these performance outcomes, which are independent of the cost uncertainty, are

Build Performance	Probability
Low	0.2
Moderate	0.3
High	0.5

The last issue that must be addressed is the combination of costs and performance, including the difference between spending money outside the organization for the component versus spending the money inside the organization. You have found that management can think of an "equivalent purchase price" for the nine possible combinations of outcomes associated with building the component. The following table provides this equivalent purchase price. [Note that (1) negative numbers are equivalent to receiving money and (2) the cost of building the component has been included in the values in the table.]
Table of Equivalent Outside Purchase Price as a Function of "Built Performance" and "Built Cost"

	Built Cost inside the organization		
Built Performance	\$100,000	\$200,000	\$400,000
High	−300,000	−200,000	0
Moderate	0	100,000	300,000
Low	100,000	200,000	400,000

Note that management prefers to build the component inside because the \$200,000 build cost with moderate performance is equivalent to spending \$100,000 outside. Assume that management's value function on "Outside Purchase Price" is a linear function with coefficient of −1.

a. Draw an influence diagram for this problem.
b. What is the best expected value decision?
c. What is the expected value of perfect information for built performance? for built cost? and for the combination of built performance and built cost? Show the influence diagram for each of these perfect information calculations.

13.8 Consider Problem 13.6. The first paragraph holds except we will drop the fair weather condition. The probability of good weather is 0.3; the probability of poor weather is 0.7.

We are now going to enhance this model to address the need to test our system under a specified test condition. The weather affects the ability of each site to provide the necessary elements (e.g., terrain, visibility) that define the test condition. Our test experts visit each site and return with probabilities that each site can do a "good" versus "poor" job of reproducing the needed test condition. Assume that we have definitions of good and poor that meet the clairvoyant's test. (Note we could have defined more than two categories if we felt we needed to achieve more accuracy.)

Site	Weather	p(test condition good|site, Weather, &)	p(test condition poor|site, Weather, &)
A	Good	1.0	0.0
A	Poor	0.5	0.5
B	Good	0.9	0.1
B	Poor	0.5	0.5
C	Good	0.7	0.3
C	Poor	0.2	0.8

The test engineers have determined that they would be willing to pay \$10,000 to move from a test site providing a poor version of the test condition to a test site providing a good version of the test condition.

Which site should we choose? Remember the rental cost of each site. What is the value of perfect information on the weather?

13.9 Now we are going to take Problem 13.8 and increase the modeling complexity by defining three different test conditions that must be reproduced by the test site. We call these test conditions X, Y, and Z. We first generate descriptions of "good" and "poor" for each test condition. Then we ask the wizard to help us elicit the values for having good versus poor representations of the three test conditions. We respond that having a poor representation of each test condition is worth no money to us. Test condition X is the most important for obtaining a good representation and we would pay \$10,000. Similarly, we would pay \$5000 to obtain a good representation of Y and \$1667 to obtain a good representation of Z.

If we were using multiattribute value theory, what would our swing weights be for these three test conditions?

Site	Weather	p(test condition X is good|site, Weather, &)	p(test condition Y is good|site, Weather, &)	p(test condition Z is good|site, Weather, &)
A	Good	1.0	1.0	1.0
A	Poor	0.5	0.5	0.5
B	Good	0.9	0.9	0.6
B	Poor	0.5	0.5	0.5
C	Good	0.7	0.5	0.7
C	Poor	0.2	0.2	0.2

Which site should we choose? Remember the rental cost of each site.

What is the value of perfect information on the weather?

13.10 Returning to Problem 13.6, there is another way in which we could have expanded the analysis from this point. In fact, the systems engineers and stakeholders have to determine whether the system is acceptable after these tests are over and the test results are in; that is, they have to make a decision. In addition, going into the test, they are not sure whether the system has acceptable performance for the stakeholders. If the system does, and it is accepted, then there should be relatively few and inexpensive fixes needed relative to the case where the system's performance is unacceptable, but the decision is made to accept the system.

So we have two decision nodes: which test site to choose and whether to accept the system for use by the stakeholders.

The weather has two states and associated probabilities as in Problem 13.8.

The ability of the three sites to reproduce good versus poor test conditions in the weather conditions is as it was in Problem 13.8.

Now we must introduce our prior probabilities on the acceptability of the system's performance. Suppose we start with only two possibilities (acceptable and unacceptable) with probabilities of 0.8 and 0.2.

We must also introduce our uncertainty that the test will say the system is "acceptable." This uncertainty is dependent on the system's actual performance and our ability to reproduce the test condition. The table below describes this probability distribution.

Actual System Performance	Ability to Reproduce the Test Condition	p(test says accept|system is., test condition is, &)
Acceptable	Good	0.95
Acceptable	Poor	0.60
Unacceptable	Good	0.10
Unacceptable	Poor	0.25

The quality engineers are called in to help us determine what the relative value of accepting a system is given it is or is not acceptable, over the life time of the system. These engineers conduct an analysis over the 10-year life time of our system and present the net present value (NPV) to our organization for the following conditions:

Actual System Performance	Decision to Accept or Not	Justification for Last Column	NPV over System Life Time
Acceptable	Accept	Best profit	\$100,000
Acceptable	Do Not Accept	Make some unneeded fixes	\$80,000
Unacceptable	Accept	Have many repairs under warranty, damage reputation	−\$10,000
Unacceptable	Do Not Accept	Make needed fixes, delay hurts sales	\$20,000

Which site should we choose? Remember the rental cost of each site. What is the expected value of perfect information on the weather?