J PROD INNOV MANAG
1987;4:185–198

OOOO

New Industrial Product Design and Evaluation Using Multiattribute Value Analysis

Ralph L. Keeney and Gary L. Lilien

Increasingly, the design of successful new industrial products is related to careful market assessment. Traditionally, managers and researchers have studied their markets by examining a small number of product attributes that are common across a range of informed respondents. In many ways, these techniques fail to meet the challenges posed by today's often heterogeneous, highly competitive, fast moving industrial markets. Ralph Keeney and Gary Lilien introduce us to a technique they call multiattribute value analysis, both describing the procedure and describing a comprehensive example. Their approach introduces considerable flexibility to the process of market assessment. Technically, it permits the evaluation of many more attributes, value tradeoffs, and synergies among attributes than do more traditional methods. In addition, it permits nonlinear evaluation functions that may be idiosyncratic to the individual. Practically, their approach, illustrated with a detailed case application, is shown to have significant potential for aiding product design decisions.

The New Industrial Product Development Process

The long-term health of industrial and consumer product companies is tied to their ability to innovate successfully—to provide existing and new customers with a continuing stream of attractive new products and services. The firm that does not maintain a program of managed innovation can quickly find itself behind competition. But the risks associated with innovation are significant; there are large investments involved and the likelihood of failure is high [14,17].

Hopkins [23] reports that well over half of all industrial firms find their success rates "disappointing" or "unacceptable." Cooper [13] reports a failure rate of 41% for fully developed new industrial products introduced into the market; i.e., for those that successfully passed the development process. Booz, Allen and Hamilton [4] report a failure rate of 35% for new products. In an earlier study, Booz, Allen and Hamilton [3] report that about 70% of the resources spent on new products are allocated to products that are not successful in the market.

There are many reasons to believe that successful new product development will be even harder in the future than it has been in the past. Those reasons include the fragmentation of markets, greater market competition, shorter time periods to adjust new products, increasing social and governmental constraints, capital shortages, and shorter product life cycles [31].

A number of studies on new product failures

Address correspondence to Gary L. Lilien, Research Professor of Management Science, Pennsylvania State University, 113 Business Administration Building II, University Park, PA 16802.

52 Vanderbilt Ave., New York, NY 10017

0737-6782/87/$3.50

BIOGRAPHICAL SKETCHES

Ralph L. Keeney is Professor of Systems Science at the University of Southern California. He has made contributions to both the theory and practice of decision analysis with a focus on problems involving multiple objectives. He was previously a professor in the OR Center at M.I.T. and was head of the decision analysis group at Woodward-Clyde Consultants. His published works include *Decision with Multiple Objectives* (Wiley, 1976) with H. Raiffa, *Siting Energy Facilities* (Academic Press, 1980), and *Decision Analysis Video Tapes and Study Guide* (M.I.T. Center for Advanced Engineering Study, 1978) with A.W. Drake. Dr. Keeney is associate editor for OR practice for the journal, *Operations Research,* and serves as editor-at-large for *Interfaces*. His current interests include applying multiattribute value analysis in the new product development process.

Gary L. Lilien is Research Professor of Management Science in the College of Business Administration at The Pennsylvania State University. He has published more than 50 articles and three books: *Market Planning For New Industrial Products,* (Wiley, 1980) with Jean-Marie Choffray; *Marketing Decision Making: A Model Building Approach* (Harper and Row, 1983) with Philip Kotler; and *Marketing Mix Analysis With Lotus 1-2-3* (Scientific Press, 1986). He is Editor-in-Chief of *Interfaces,* a journal focusing on practical applications of management science and is a member of the editorial board of the *Journal of Marketing*. Dr. Lilien is co-founder and Research Director of The Institute for the Study of Business Markets (ISBM) an organization aimed at fostering research on nonconsumer markets. Research in the new business products area is a major research interest of Dr. Lilien and is a major research priority for the ISBM.

In preparing this paper, Keeney was partially supported by the Office of Naval Research under Contract N00014-84-K-0332 titled, "Value-Focused Thinking and The Study of Values." Additional support was provided by The Pennsylvania State University's Institute for the Study of Business Markets.

[1,3–5,10–13,16,18,29,32] have found that although there are often many causes, a predominant reason that products fail is for lack of a clear understanding of market needs. For example, Calentone and Cooper [6], in a study of new-industrial-product failures, found that the largest category, 28%, included products that met a nonexistent need while only 15% of the product failures were "bad" products—that is, did not do what they were supposed to do. Cooper and Kleinschmidt [15] point out that a major reason for this lack of understanding of market needs results from inadequate market studies. They note that

> Preliminary Market Assessment [was] very weakly rated [as well done] overall, yet strongly correlated with all four measures of project performance . . . Detailed Market Study/Market Research [was] omitted altogether in 74.6 percent of the projects . . . yet [was] significantly correlated with all four measures of project performance. (p. 84)

Thus, there is strong reason to believe that an organized, analytic approach to new industrial product development has the potential to streamline the process and reduce new product failure rates.

In the next section we outline several popular analytic evaluation approaches for new products and suggest that there is an important class of problems for which these approaches are inappropriate. The next section introduces multiattribute value analysis for this class of problems. We develop that approach, illustrate its use in a real application, and evaluate its potential for new industrial product evaluation.

Analytic Approaches for the New Product Development Process

Urban and Hauser [34] discuss several methods available to aid in new product design and positioning. Most of those methods link product *attributes* (or dimensions, i.e., physical product features—speed, efficiency—as well as psychological aspects of the product—perceived quality, service, vendor reliability) to a *preference* measure through some functional form (usually linear). Three commonly used procedures are expectancy value methods, preference regression methods, and conjoint analysis. Each of these procedures relies on the product designers or analysts, rather than potential customers, to specify the attributes. While the designers are often guided by prior input from customers, their decisions may be somewhat arbitrary.

Expectancy Value Methods

These methods [37] ask respondents to score each of a set of products or product-concepts (on a 1–7 scale, say) on each product attribute. The respondent also provides an importance weight for each of these attributes. The value (or preference) the individual has for that product is then calculated as the sum of the attribute scores multiplied by the importance weight of the corresponding attribute.

Expectancy value methods are perhaps the most widely used of all methods for new product screening. The method is easy to understand, simple to apply, and inexpensive as well. However, if several of the attributes tap the same underlying dimension, the method will over-weight (double count) the importance of those attributes. In addition, Beckwith and Lehman [2] show that the method can lead to the halo effect in which a favorable product is inappropriately rated favorably along all scales.

Preference Regression

The preference regression model looks very much like the expectancy value model but is developed differently. In the expectancy value method, respondents provide attribute-weights and the model is used to predict (infer) overall product preference. In the preference regression method, preference judgments obtained from individuals are used as dependent variables in a regression equation with attribute ratings as independent variables. Importance weights are then inferred as regression coefficients. Other differences between preference regression and expectancy value methods are (a) in preference regression, importance weights are usually assumed homogeneous across a group of respondents and (b) groups of attributes (combined through a factor analysis, perhaps) are often used as independent variables to reduce problems of multicollinearity.

Some advantages of the preference regression approach are that it is easy to use (requiring only a standard regression package), it is frequently more accurate in predicting preferences than the expectancy value approach, and the inferred importance weights can be used to guide product-design decisions.

Its limitations include the fact that a linear model-form is most frequently used. Although implementations such as PREFMAP [7] have options to deal with thresholds, nonlinear effects, and the like, Urban and Hauser [34] contend that such effects are not handled in an entirely satisfactory manner. In addition, importance weights are normally population-averages and provide little information about individual-level differences in attribute importance.

Conjoint Analysis

Conjoint analysis is an approach for predicting respondent preferences for a "product" defined by a set of attributes at specific levels (price = \$1400, horsepower rating = 28, etc.). The respondent usually rank-orders total product profiles, from most preferred to least preferred. These product profiles are combinations of product attributes, set at discrete levels. Four attributes set at four levels would lead to 4^4 (i.e., 256) combinations for rank ordering, so fractional factorial procedures [21] and trade-off analysis [24] are used to keep the respondent-task not too unwieldy. The conjoint analysis procedure then determines importance weights (known as "part-worths") for each of the attributes.

Conjoint analysis is most useful in evaluating design tradeoffs when a small number of important, discrete, alternatives are being considered. Analysis is normally done at the individual level, and market response is estimated by aggregating individual responses. Cattin and Wittink [8] review commercial uses of conjoint analysis. Green, Carroll, and Goldberg [20] discuss the POSSEE system, a decision support system for conjoint analysis studies.

Conjoint analysis has several limitations. First, no statistical inference procedures exist, a serious drawback for fitting a model form. Second, the procedure assumes that the appropriate experimental factors (the product attributes) are known in advance, are small in number, and are constant across respondents. Finally, the approach assumes that either the rank-ordered or paired-comparison data about individual preferences provide reliable information about likely purchase actions (a limitation of all the procedures, in fact).

The three procedures outlined here are the ones most commonly used in practice. But they are most frequently used for consumer products or for industrial products with a few important design dimensions that are identical across respondents.

There is an important class of products for which these approaches are not well suited. Many industrial product markets for high technology, and in particular, capital equipment, are characterized by heterogenous users (users with different needs), a large number of product attributes, and a small number of high-volume, buying firms. The multivariate value function approach, developed and illustrated below, is appropriate for markets of this type. Indeed, in many such markets, there will be a small number of highly influential buying firms who are consistent early adopters of new technologies. Von Hippel [35] refers to these customers as "lead users." He points out that in such markets, "most potential users will not have the real-world experience to problem solve and provide accurate data to inquiring marketing researchers" (p. 79). He suggests that a key to new product success in such markets is to aim in-depth market research at the (usually) small number of lead users.

The Multiattribute Value Analysis Approach

The multiattribute value analysis approach has three major steps. The first is specification of the product attributes for a given customer and prospective need. The second is identification of an appropriate evaluation model (i.e., the form of a multiattribute value function). And the third is the assessment of value judgments to calibrate the value function for that respondent.

Specification of Attributes

The specification of attributes provides a comprehensive list of criteria for evaluating prospective products. The discussion between the analyst and the prospective customer is used to develop the criteria. Particular emphasis must be placed on clarifying "softer," hard-to-measure criteria such as supplier problems, serviceability, or upgradeability, as these important criteria are often neglected with other approaches. The intent is to elicit a complete set of significant criteria.

Next, each criterion must be individually appraised to determine if it is fundamental to the product or a means to something fundamental (an end). For example, a criterion concerning redundancy may be a means to ensure reliability and low cost (ends). As such, it is inappropriate to include both redundancy (means) and reliability (ends) criteria. The process of identifying fundamental criteria should indicate which fundamental criteria are parts of which others and result in a hierarchical structure of the fundamental criteria. (The next section provides an example.)

For each criterion at the lowest level of the hierarchy, a measure (continuous or discrete scale) must be identified or constructed to indicate the degree to which products meet the corresponding criterion. Each measure has implicit value judgments, so it is desirable to make the measures customer-specific. Procedures to identify measures are discussed in Keeney [26].

Identification of a General Evaluation Model

The appropriate functional form of an evaluation model depends on how individuals "value" both relative levels of the product attributes and interactions between levels of these attributes. Different independence concepts are used to describe these value relationships. Various rather powerful theoretical results imply a specific functional form given sets of these independence properties. A major result concerning the independence of attributes is summarized in the Appendix. Such results are most appropriate for "fundamental" criteria (i.e., criteria that exclude means criteria as discussed above). If the tests described in the Appendix indicate significant dependencies, then it is usually appropriate to return to the first step—specification of attributes—and attempt to better define or restructure the attributes.

Assessment of Value Judgements

Assessment procedures first identify the independence conditions that are appropriate for a prospective customer, which will indicate the form of the value function. Then they estimate component value functions and scaling factors. The procedure outlined here is illustrated in the following case study.

Preferential independence (see Appendix) is verified by finding pairs of products that the customer finds indifferent that differ in terms of two criteria only. By varying the other criteria, we investigate whether this indifference is affected by the levels of the other criteria. If indifference does hold, preferential independence is confirmed. For weak-difference independence, we ask the customer for a level of one criterion that is half way in value between the value of two different levels of that criterion. If this "midvalue" level remains the same regardless of what levels are chosen for the other criteria, weak-difference independence holds.

To determine the value function, v, the responses from questions to appraise weak-difference independence are used. If for attribute, X, x′ is the midvalue level between the least desirable level x^0 and the most desirable level x^*, then we set $v(x') = 0.5$ (since $v(x^0) = 0$ and $v(x^*) = 1$ are assigned to scale v from 0 to 1). Other points between x^0 and x′ and between x′ and x^* can be similarly assessed to yield more points on v. Then a curve can be fit to the points. From the pairs of "indifferent products" identified in verifying preferential independence (see Appendix), at least n equations must be generated to estimate the n attribute weights.

Details on the independence concepts and assessment procedure are found in Keeney and Raiffa [28], Keeney [25], and von Winterfeldt and Edwards [36].

In essence, then, in a particular application (a) careful questioning of the prospective purchaser identifies a set of attributes and ascertains whether they are "ends" and not "means." (b) a series of indifference questions are used to determine and select an appropriate value function form and (c) multiple variations of those questions are then used to calibrate the value functions. A case illustration follows next.

Case Application: Capricorn Corporation and the OR9000

Background

Over the last decade the integrated circuit industry has gone through a cycle of birth, explosive growth, and, currently, has moved into a phase of severe competition. Few industries have evolved so quickly on the one hand and have found such severe competition on the other. The risks associated with customer and market misassessment are as significant here as in any other industry we have examined.

In early 1984 we were approached by the Capricorn Corporation (ficticious name), a well-known Silicon Valley firm, with an established reputation for producing high quality manufacturing, test, and control equipment. The firm had developed a technical breakthrough that, they felt, would give them a significant cost advantage in manufacturing test equipment for very large scale integrated circuits (VLSIC). The questions Capricorn asked were: how would prospective customers evaluate a Capricorn entry in this (highly competitive) market and how should Capricorn's product be designed?

The analysis described here was part of a larger study aimed at identifying likely customers for the product, the decision-process within firms for purchasing such test equipment, the likely future needs of such customers, and the likely competitors for the product within each customer-organization.

Identifying Attributes

Following a review of the technical literature and several meetings with technical and marketing staff at Capricorn, 17 decision criteria for VLSIC Tester evaluations were identified (Table 1, Column 1). Analyzing means-ends relationships to eliminated redundances, these 17 criteria were identified from a much larger list of 57 main criteria, many of which had subcriteria [22]. The decision criteria fell into four categories: technical, economic, software, and vendor support. Each criterion required an associated measure to describe characteristics of different products in terms of that criterion. These measures were either natural scales such as the number of picoseconds (psec) for timing accuracy or a constructed scale, such as yes/no for the availability of data analysis software (Table 1, Column 2).

Because the uses of the tester vary by the types of VSLIC devices tested by the customer, the range of desirable levels for performance criteria also vary by customer. For example, a customer may indicate that the "minimum acceptable level" for pin capacity is 64 while the "maximum level" (given current plans) is 256. Then the pin capacity dimension for this customer would only be evaluated between 64 and 256.

Technical criteria. There are six technical criteria for evaluating the testers, and each of these has a readily available natural measure. For example, pin capacity is measured by the number of pins and vector depth is measured by the memory size in megabits.

Economic criteria. A key criterion in evaluating integrated circuit testers is price. However,

Table 1. Ratings and Weights of Decision Criteria for Acorn Tester Selection

Decision criteria	Measure	Range of the measure			Rank-order of criteria within decision criterion category	Relative weights within decision criterion category
		Minimum acceptable level	Midvalue points	Maximum desirable level		
Technical						
X_1 = pin capacity	quantity	144	190	256	3	15
X_2 = vector depth	memory size (megabits)	1	3.9	4	2	20
X_3 = data rate	MHz	40	80	100	5	10
X_4 = timing accuracy	picoseconds	±500	±375	±250	1	35
X_5 = pin capacitance	picofarads	100	65	30	4	10
X_6 = programmable measurement units	number	4	8	16	6	10
Economic						
X_7 = price	total cost	2.5M	2M	1.5M	1	50
X_8 = uptime	percent	98	99	100	3	20
X_9 = delivery time	months	6	5	4	2	30
Software						
X_{10} = software translator	percent conversion	10	50	100	5	15
X_{11} = networking: communications	yes/no	no	–	yes	2	20
X_{12} = networking: open	yes/no	no	–	yes	3	20
X_{13} = development time	mean time (months)	4	3	2	1	30
X_{14} = data analysis software	yes/no	no	–	yes	4	15
Vendor Support						
X_{15} = vendor service	time until system works (hours)	4	2	1	2	30
X_{16} = vendor performance	time until response (hours)	4	2	1	2	30
X_{17} = customer applications	yes/no	no	–	yes	1	40

an important characteristic of the model is the ability of the potential purchaser to provide the appropriate price measure. Some prospective customers use the net purchase price. Others prefer to use the cost per unit tested or the total cost after tax implications. Two other related economic criteria are uptime and delivery time.

Software criteria. A key software criterion is whether a universal translator exists, measured by a yes/no scale. (A universal translator takes VLSIC testing software developed for another manufacturer's tester and translates it for use on Capricorn's tester.) Other software criteria include several forms of networking capabilities and software development time.

Vendor support criteria. Vendor support criteria include service, performance, and application support. Vendor service is measured by the time necessary to get the equipment running after it has gone down. Vendor performance is measured by the time until vendor personnel arrive at the customer's facility after such assistance has been requested. The criterion of vendor capability to assist in applications is measured by the simple yes/no scale.

An Evaluation Function for Mr. Smith of Acorn Industries

Acorn Industries is one of the most important potential customers for Capricorn, as it is one of the five U.S. industry leaders in the manufacturing of VLSICs. After careful preliminary evaluation of the tester acquisition process, Michael Smith, the Manager of Test Engineering was intensively interviewed to determine his value function. Not only was Mr. Smith identified by all other buying center members at Acorn as "most influential" in the buying process for test equipment, he was so identified in numerous interviews outside the firm as well. He seemed to serve as a "key referrent" for many firms in the region. In addition, he maintained a detailed matrix of test equipment and evaluation of that equipment along some 25 to 30 dimensions of his own. Thus, Acorn Industries in general and Mr. Smith in particular met the "lead user" criterion discussed by Von Hippel [35].

Form of Smith's Evaluation Function

The independence assumptions necessary to use the results in the Appendix were verified with Mr. Smith. Preliminary questioning suggested that the additive form (Eq. (A.2)) would be appropriate, and, as discussed in the next section, checks of this assumption confirmed that impression. Hence, the particular evaluation function chosen was

$$\begin{aligned} v(x_1, \ldots, x_{17}) = {} & k_T v_T(x_1, \ldots, x_6) \\ & + k_E v_E(x_7, x_8, x_9) \\ & + k_S v_S(x_{10}, \ldots, x_{14}) \\ & + k_V v_V(x_{15}, x_{16}, x_{17}), \end{aligned} \tag{1}$$

where

$$v_T(x_1, \ldots, x_6) = \sum_{i=1}^{6} k_i v_i(x_i), \tag{2a}$$

$$v_E(x_7, x_8, x_9) = \sum_{i=7}^{9} k_i v_i(x_i), \tag{2b}$$

$$v_S(x_{10}, \ldots, x_{14}) = \sum_{i=10}^{14} k_i v_i(x_i), \tag{2c}$$

$$v_V(x_{15}, x_{16}, x_{17}) = \sum_{i=15}^{17} k_i v_i(x_i), \tag{2d}$$

where v is a measurable value function scaled from 0 to 1 for evaluating testers represented by $(x_1, \ldots, x_{17})$; T, E, S, and V stand for technical, economic, software, and vendor support criteria, respectively; v_T, v_E, v_S, and v_V are component measurable value functions for the four respective decision criteria categories; k_T, k_E, k_S, and k_V are the relative importance weights of the four decision criteria categories given the ranges indicated by the prospective customer; v_i is the component measurable value function for decision criterion X_i scaled from zero to one; x_i is a specific level of X_i; and k_i is the relative importance weight of decision criterion X_i within its decision criterion category. To use this model, we need to assess 17 component value functions (the v_i), 17 importance weights for the criteria (the k_i), and four importance weights on decision criterion categories (k_j, j = T, E, S, V).

Calibration of the Evaluation Function

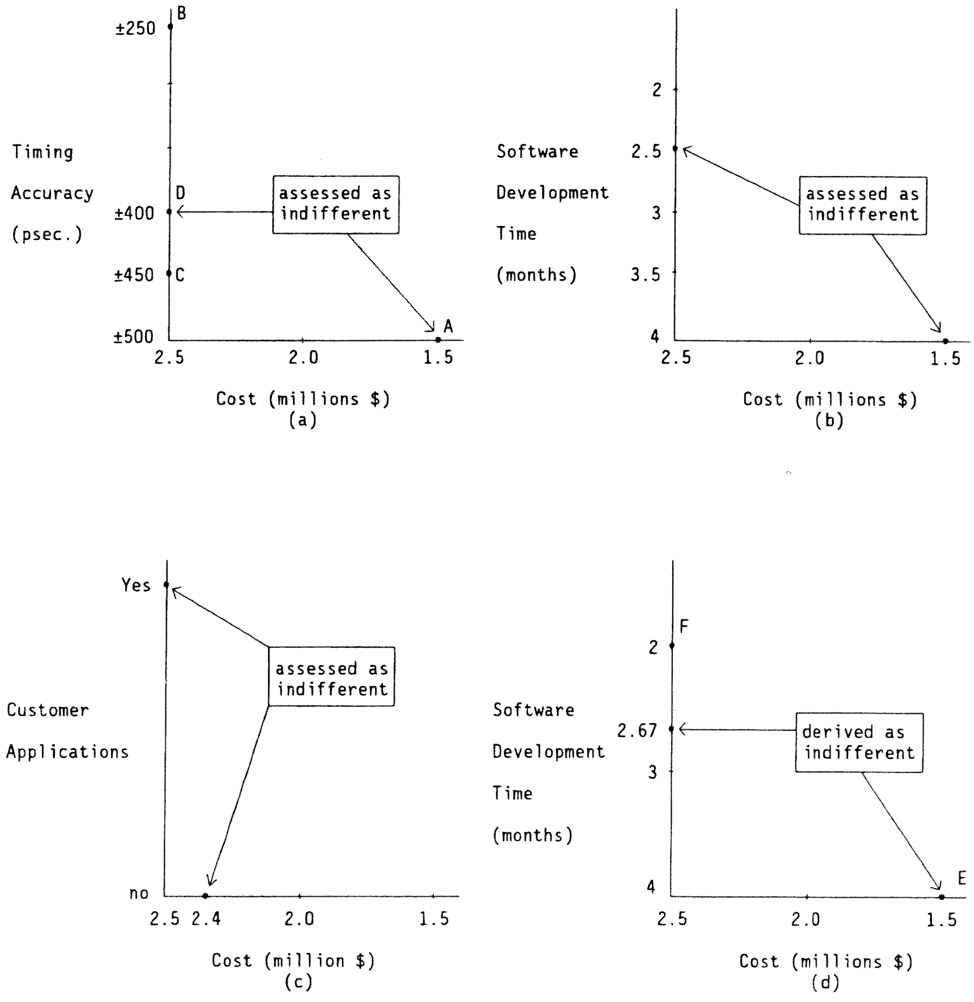
The 17 decision criteria were reviewed with Mr. Smith. The measure he chose for criterion X_7 was total cost, including tester cost plus initial training and spares. He also changed decision criterion X_{10} to the "percent conversion of the software translator." To determine the range over which criteria could vary, Mr. Smith was asked to specify a "minimum acceptable level" and a "maximum desirable level" for each criterion. These results are displayed in Table 1, Columns 3 and 5.

Importance weights. Before assessing information to directly determine the importance weights, we asked easier questions to rank-order these weights. For instance, with respect to the technical decision criteria, we begin by asking the following. "Suppose all the technical criteria were set at their respective minimum acceptable levels. If you could raise only one of these criteria from that level to its 'maximum desirable level,' which criterion would you move?" His response was that timing accuracy would be moved from ±500 picoseconds to ±250 picoseconds. This implies that k_4 is Eq. (2a) is the largest of the importance weights for the technical decision criteria evaluation function, v_T.

The next question concerned which of the technical criteria would be the second most important to move from its minimum acceptable level to its maximum desirable level. His response this time was that vector depth would move from 1 to 4 megabits. This process continued and was repeated for economic, software, and vendor support decision criteria (Table 1, Column 6).

To specify the importance weights numerically, value tradeoffs among criteria must be considered. We chose the criterion ranked first within its decision criterion category and developed indifference pairs of criteria as illustrated in Figure 1a. This exhibit shows that, with all other criteria held fixed, the respondent is indifferent between a cost of $1.5 million for a tester with a timing accuracy of ±500 picoseconds and a tester that costs $2.5 million and has a timing accuracy of ±400 picoseconds. First, we asked which of points A or B in Figure 1a was preferred. Mr. Smith stated that B was preferred. Next, we asked which consequence was preferred between A and C in Figure 1a. Here, his response was A. This indicated that an increase in timing accuracy

Figure 1. Assessed Value Tradeoffs



of 50 picoseconds from 500 and 450 was not worth to him the additional cost of one million dollars per tester. We then found out that Mr. Smith was indifferent between consequences A and D. The other pairs of indifference points indicated in Figure 1 were assessed in a similar manner. These assessments also directly verified several of the preferential independence assumptions necessary to use the evaluation model in Eq. (1).

Since each of the value tradeoffs in Figure 1 show how much the respondent would pay for more performance along a given criterion, they provide a measure of decision criteria importance. An advantage of this procedure is that it directly addresses value tradeoffs. The procedure, however, is somewhat complex. A simpler procedure that does not directly address value tradeoffs is to allocate 100 points among the decision criteria of concern.

Within each decision criterion category, Mr. Smith was asked to allocate 100 points to represent the relative values of moving the criteria from their minimum acceptable levels to their maximum desirable levels. Among technical decision criteria, of the total of 100 points, he assigned 35 to timing accuracy, 20 points to vector depth, etc. The other associated weights indicated in Table 1, Column 7, were assigned to sum to 100. For ease of implementation, relative weights were first assigned and then normalized to sum to 100. Also, as a consistency check and for guidance in assigning weights, the ranking of importance weights within decision categories was available.

Component value functions. To evaluate alternative testers, we needed the relative desirability of any given level of any criterion within the range specified in Table 1. To do this, we asked questions such as (for pin capacity, which ranged from 144 to 256 pins), "What number of pins, call this y, is such that the increased desirability of going from 144 pins to y is equal to the increase in desirability of going from y pins to 256?" Mr. Smith's response in this case was 190, indicating that the relative desirability assigned to 190 must be midway between that assigned to 144 and to 256 pins. Since this midvalue point did not depend on the level of the other criteria, pin capacity was weak-difference independent of the other criteria. This is another required assumption for evaluation model (1).

We must scale the component value functions, v_i, from 0 to 1. Thus, for pin capacity, $v_1(144) = 0$, $v_1(190) = 0.5$, and $v_1(256) = 1$. Fitting this information to an exponential curve yields

$$v_1(x_1) = 1.929 \times [1 - \exp(0.0065(144 - x_1))].$$

An exponential, linear, or other functional form can be used to summarize these judgements. By using the midvalue assessments for additional pairs of criterion levels, additional data points are generated to help select and fit an appropriate curve.

The other assessed midvalue points are presented in Table 1, Column 4. In the case where the measure is a yes/no scale, we simply assign 0 to the no level and 1 to the yes level of the value function in each case. The resulting component value functions are listed in Table 2.

Synergies. To determine whether there were any synergistic effects in combinations of different criteria, we posed questions illustrated by Figure 1d referring to pairs of costs and mean development times for software. We assigned a relative desirability of zero to the least desirable combination (i.e., a cost of $2.5 million and four months development time) and a relative desirability of 100 to the best pair (a cost of $1.5 million and 2 months development time). We then determined the relative value assigned to the other two corners indicated by E and F in the figure. The relative weight Smith assigned to E was 40 and that assigned to F was 60. This result was an indication that the additive value function was appropriate since these summed to 100, and were not otherwise constrained to do so. Thus, the appropriate specific form of the evaluation model (A.1) was the additive case (A.2). Similar comparisons were performed with other dimension pairs with similar results.

The overall evaluation function. Using the information above, we calculated the parameters for the value function (1) and (2). The decision criteria group weights are

$$\begin{aligned} k_T &= 0.52,\ k_E = 0.14, \\ k_S &= 0.32,\ k_V = 0.02. \end{aligned} \tag{3}$$

These weights were assessed from the type of value tradeoffs presented in Figure 1 and normalized to sum to 1.0. Thus, the overall value function ranges from 0 to 100 since we assigned the

Table 2. Tester Evaluations and Evaluation Function

	Tester			Component value function
Decision criteria	OR 9000	J941	Sentry 50	
Technical				
X_1 = pin capacity	160	96	256	$v_1(x_1) = 1.929[1 - \exp(0.0065(144 - x_1))]$
X_2 = vector depth	.128	.256	.064	$v_2(x_2) = -0.9736E - 09[1 - \exp(6.917(x_2 - 1)]$
X_3 = data rate	50	20	50	$v_3(x_3) = -0.3091[1 - \exp(0.02406(x_3 - 40)]$
X_4 = timing accuracy	±1000	±1000	±600	$v_4(x_4) = (500 - x_4)/250$
X_5 = pin capacitance	55	50	40	$v_5(x_5 = (100 - x_5)/70$
X_6 = programmable measurement units	8	2	4	$v_6(x_6) = 1.309[1 - \exp(0.1203(4 - x_6))]$
Technical Evaluation	−54.9	−78.2	10.4	
Economic				
X_7 = price	1.4M	1.0M	2.8M	$v_7(x_7) = 2.5 - x_7$
X_8 = uptime	98	95	95	$v_8(x_8) = (x_8 - 98)/2$
X_9 = delivery time	3	6	6	$v_9(x_9) = (6 - x_9)/2$
Economic Evaluation	100	45	−45	
Software				
X_{10} = software translator	90	90	90	$v_{10}(x_{10}) = 2.768[1 - \exp(0.00498(10 - x_{10}))]$
X_{11} = networking: communications	yes	yes	yes	$v_{11}(x_{11}) = v_{11}(no) = 0.0; v_{11}(yes) = 1.0$
X_{12} = networking: open	yes	no	no	$v_{12}(x_{12}) = v_{12}(no) = 0.0; v_{12}(yes) = 1.0$
X_{13} = development time	3	4	4	$v_{13}(x_{13}) = (4 - x_{13})/2$
X_{14} = data analysis software	yes	yes	yes	$v_{14}(x_{14}) = v_{14}(no) = 0.0; v_{14}(yes) = 1.0$
Software Evaluation	83.7	48.7	48.7	
Vendor Support				
X_{15} = vendor service	2	4.75	6	$v_{15}(x_{15}) = -0.3091[1 - \exp(0.4811(4 - x_{15}))]$
X_{16} = vendor performance	4	4	4	$v_{16}(x_{16}) = -0.3091[1 - \exp(0.4811(4 - x_{16}))]$
X_{17} = customer applications	yes	yes	yes	$v_{17}(x_{17}) = v_{17}(no) = 0.0; v_{17}(yes) = 1.0$
Vendor Support Evaluation	55.0	37.2	34.3	
Overall Evaluation	13.3	−18.0	15.4	

component weights to equal 100 within each of the decision criteria categories. The weights in Eq. (3) plus the weights in Table 1 and the component evaluation functions in Table 2 provide all the parts of the evaluation function.

Summary of judgements required. Let us consider the number of types and judgements that would be required from each respondent to identify and quantify an appropriate evaluation model. If there are N criteria identified for a particular problem, then approximately 3N judgments (depending on the evaluation form), are needed. Of these, N judgments are needed for the verification of N independence conditions to imply a functional form for the model. Another N judgments provide the importance weights, and a third N provide the component value functions.

Since the verification of independence conditions also provides information to specify either importance weights or component value functions, fewer than 3N judgments are absolutely necessary. However, a few additional judgements are useful as consistency checks. Also, if violations of independence conditions occur, either additional importance weights or component value functions that are dependent on criteria levels of other attributes are needed. The number of additional judgments increases roughly as the square of the number of criteria violations of independence conditions. If there are more than about three independence violations, the value model has been constructed with a poor set of overlapping criteria with complex value impacts. The criteria set should be reconsidered and modified in this case. With two or fewer independence violations, assessing value judgments to quantify synergies is both reasonable to do and provides insights.

Using the Model

The model developed above was used to determine the likely response Acorn Industries would have to the OR9000, a new tester Capricorn currently had in prototype form. Acorn Industries currently is interested in two preferred testers: the Sentry 50 and the J941. Table 2 presents a description of the three testers, in terms of the 17 decision criteria, and their evaluation using the value function developed above.

Base case evaluation. The decision criteria were scaled on a 0.0 to 1.0 scale with 1.0 corresponding to the most desirable level and 0.0 corresponding to the least desirable level of the scales listed in Table 1. Some evaluations in Table 2 are less than 0.0 since they are lower than the minimum acceptable level. This occurred because the range was set for the testers to be used in research as well as production engineering, while the testers evaluated were basically production models. Thus, none of these testers is truly acceptable to Acorn. This evaluation shows which of the three testers would be "least unacceptable," and indicates the characteristics of preferable testers.

Using base case evaluations, the Sentry 50 is slightly preferred to the OR9000 and both are much preferred to the J941. This follows from the overall evaluations of 15.4 for Sentry 50, 13.3 for OR9000, and −18.0 for J941.

In terms of decision categories, the OR9000 dominates J941, i.e., the OR9000 is better than the J941 in terms of technical, economic, software, and vendor support categories. The Sentry 50 does not dominate the J941.

The OR9000 is strongly preferred to the Sentry 50 in the economic, software, and vendor support categories. The Sentry 50 is strongly preferred to the OR9000 in the technical criteria category. This preference is basically due to pin capacity and timing accuracy, the latter being the major factor. We ran several sensitivity analysis to show the model can be used.

Sensitivity analysis 1: change in description. Consider the effect of a decrease in the data rate for the OR9000 from 50 to 20 MHz. This reduces the technical evaluation of OR9000 from −54.9 to −56.9 and reduces the overall evaluation from 13.3 to 12.3. How much is this difference worth? If the price of the OR9000 dropped from the $1.4 million with a 50 MHz data rate to $1.26 million with a 20 MHz data rate, the evaluations are equal at 13.3. This indicates that this change in data rate is worth $140,000. Similar assessments can be performed for any of the criteria.

Sensitivity analysis 2: indifferent equivalent costs. How much more is the Sentry worth than the OR9000? Note that the total difference in evaluation scores is 15.4 to 13.3 or 2.1 units. A $1 million price difference translates into $k_E k_7$ evaluation units or .14 × 50 = 7 units. The 2.1 unit difference is worth 2.1/7 × $1 million or $300,000. Thus Sentry is worth $300,000 more than OR9000.

Sensitivity analysis 3: weights and value tradeoffs. The evaluations of the OR9000 and the Sentry 50 are close, so customer choice is sensitive to weighting factors. The Sentry 50 is superior to the OR9000 only in the technical criteria category. Hence, we began lowering the weight on the technical criteria from 0.52 and proportionally increasing the weight on the other criteria categories to maintain a sum of 1.0. Once the technical weight drops to 0.505, the two testers are indifferent. If the weight on the technical criterion drops below 0.505, then the OR9000 is preferred to the Sentry 50. This indicates that the value tradeoffs are likely crucial to choice for this particular problem. Hence, additional market research should perhaps focus on their value tradeoff.

Sensitivity analysis 4: design upgrade analysis. Consider a Capricorn Machine that is the equal of the Sentry 50 in technical performance. Such a machine would have a value score of 47.3 vs. the OR9000's current score of 13.3 and Sentry's 15.4. This analysis points out that a substantial improvement in customer value will be seen by a technical upgrade of the OR9000.

Impact at Capricorn

Largely on the basis of this analysis, which showed the high importance of the technical criteria relative to the economic criteria, Capricorn abandoned market plans for the OR9000. This evaluation model, calibrated here and at other customer locations, helped focus attention on the market potential resulting from improvements in several technical dimensions. Research into producing a new machine with significant enhancements along key technical dimensions is currently in progress.

Evaluation and Use

Our experience with this procedure for evaluating likely customer response to a new industrial product has been quite positive. Where the decision criteria—the key product attributes—are customer-specific, where there are many such attributes and where synergies, value tradeoffs, thresholds, and the like are important, no commonly used market evaluation procedure is satisfactory. Specifically, this approach has several benefits over existing methods. First, it permits the evaluation of many more dimensions than is possible with conjoint analysis, and permits interactions/synergies at any level as well. Second, it naturally allows for idiosyncratic, respondent-specified evaluation dimensions. Third, the model structures the evaluation process in much the same way that a careful industrial buyer does—and respondents take the evaluation task seriously because of that. Fourth, the procedure is hierarchical, bundling dimensions for value tradeoffs at higher levels and comparing individual criteria at lower levels. This eliminates much of the redundancy in lists of criteria by focusing on fundamental (i.e., ends) criteria and eliminating means.

On the other hand, this approach is not amenable for a large sample market research study. It requires heavy involvement of the respondent (most interviews take about half a day), and a well-trained, perceptive interviewer, since the interviewing procedure combines the specification and testing of functional forms with the calibration task. To perform this analysis in 10 to 20 firms would cost about the same as what pretest market models for consumer packaged goods costs ($25–75,000 [33]). This is somewhat more than most industrial marketers are used to paying for market analysis, but provides considerably more specific insight.

Many of the limitations of this approach derive from the more complex nature of this procedure relative to those procedures described in the second section of this article. The procedure requires skill to administer. The interviewer must be able to explain and administer trade-off questions and be sensitive to the level of knowledge (and the fatigue level) of the respondent. As a corollary, the respondent must have sufficient product-knowledge so that attribute trade-off questions are appropriate.

In addition, there are several possible biases involved in the approach. As most organizational buying involves multiple individuals [30] it is important that the key respondent's answers reflect the values of the firm. It is conceivable that the key respondents are "technological gatekeepers," wanting to encourage the supplier to undertake risky developments and produce products that may or may not be purchased when they ultimately reach the market. Although, ultimately, such a bias would lead to loss of the respondent's credibility with suppliers over time, it is a bias that should be considered.

To the extent that multiple individuals are involved in purchasing situations, it is important that the implications of those individuals' value functions be combined in a sensible way. One method is to use the value function outputs to predict "votes" for purchase decisions and then develop a (weighted) score from these votes. Choffray and Lilien [9] develop this combining rule and others; Keeney and Kirkwood [27] derive conditions under which a multilinear combining rule (similar to Eq. (A.1)) is appropriate. The combining rule problem is an issue of key importance and no existing procedure handles it adequately at the moment. Of comparable importance is the time-variance (stability) of the value assessment. Our assessment procedure is static; even if it is valid at the moment of assessment, the market (and customer values) may be different at the time of market entry. How different they may be is another unsolved problem.

The assessment procedure is clearly obtrusive. It forces the respondent to think in terms of hierarchies and levels of attributes, trade-offs, and the like. How much that assessment procedure affects (biases) the respondent is hard to say.

Many of the above problems would be alleviated if validation data were available. At this writing, we know of no other field application of the procedure in marketing. Indeed, assessments of multiple firms in multiple markets, using this and other methods of value assessment, will be required to address the validation issue. We hope to have the opportunity to address that question in the near future.

In summary, the use of multiattribute value analysis takes more time for both the interviewer and the respondent than other approaches addressing prospective customer purchase behavior. Clearly, multiattribute value analysis is not generally appropriate for most packaged goods. But for multimillion dollar purchases, it may be cost-effective to delve deeply into the value judgements of prospective purchasers.

It is reasonable to expect that purchasers will think hard about these values when "purchase time" comes. By forcing this hard thought early the seller may get biased responses and preferences may change over time. But with an explicit representation of these stated values, and having heard the reasoning and thoughts behind them, the seller should be in a better position to recognize misleading biases and to foresee causes for future changes in values. Since values are crucial in large industrial purchases, a focus on what buyers want balanced with what sellers can give them can provide valuable marketing intelligence.

On net, multiattribute value analysis is a procedure worth considering for important product design decisions in highly competitive, heterogeneous, fast moving markets. It avoids many of the disadvantages of the more common procedures and, though its cost is not trivial, it is an investment with the potential to pay handsome returns to leading firms.

References

1. Angelus, T. L. Why do most new products fail?" *Advertising Age* 40:865–886 (March 24, 1969).
2. Beckwith, Neal E. and Lehmann, D. R. The importance of halo effects in multi-attribute attitude models. *Journal of Marketing Research* 12:265–275 (August 1975).
3. Booz, Allen & Hamilton. *Management of New Products*. New York: Booz, Allen & Hamilton, Inc., 1971.
4. Booz, Allen & Hamilton. *New Product Management for the 1980's*. New York: Booz, Allen & Hamilton, Inc., 1982.
5. Briscoe, C. Some observations on new industrial product failures. *Industrial Marketing Management* 2:151–162 (February 1973).
6. Caletone, Roger J. and Cooper, Robert C. A typology of industrial new product failure. In: *1977 Educator's Conference Proceedings*, ed. B. Greenberg and D. Bellenger (eds.). Chicago: American Marketing Association, 1977, pp. 492–497.
7. Carroll, J. D. Individual differences and multidimensional scaling. In: *Multidimensional Scaling: Theory and Application in the Behavioral Science*, R. N. Shepard, A. K. Rommey, and S. Nerlove, (eds.). Vol. 1. New York: New York Seminar Press, Inc., 1972.
8. Cattin, P. and Wittink, D. R. Commercial use of conjoint analysis: a survey. *Journal of Marketing* 4:44–53 (Summer 1982).
9. Choffray, Jean-Marie and Lilien, Gary L. *Market Planning for New Industrial Products*. New York: Wiley, 1980.
10. Cooper, R. G. Why new industrial products fail. *Industrial Marketing Management* 4:315–326 (December 1975).
11. Cooper, R. G. The dimensions of industrial new product success and failure. *Journal of Marketing* 43:93–103 (Summer 1979).
12. Cooper, R. G. Project New Prod: factors in new product success. *Journal of Marketing* 14:277–292 (1980).
13. Cooper, R. G. New product success in industrial firms. *Industrial Marketing Management* 11:215–223 (1982).
14. Cooper, R. G. The performance impact of product innovation strategies. *European Journal of Marketing* 18:43–54 (1984).
15. Cooper, Robert G. and Kleinschmidt, Elko. An investigation into the new product process: steps, deficiencies and impact. *Journal of Product Innovation Management* 3:71–85 (June 1986).
16. Crawford, C. M. Marketing research and the new product failure rate. *Journal of Marketing* 41:51–61 (April 1977).
17. Crawford, C. M. New product failure rates—facts and fallacies. *Research Management* (September 1979), pp. 9–13.
18. Davidson, J. H. Why most new consumer brands fail. *Harvard Business Review* 54:117–121 (March–April 1976).
19. Dyer, J. S. and Sarin, R. K. Measurable multiattribute value functions. *Operations Research* 27:810–822 (July–August 1979).
20. Green, P. E., Carroll, J. D., and Goldberg, S. M. A general approach to product design optimization via conjoint analysis. *Journal of Marketing* 45:17–37 (Summer 1981).
21. Green, Paul E. and Wind, Y. *Multi-Attribute Decisions in Marketing*. Hinsdale, IL: Dryden Press, 1973.
22. Healy, J. A VLSI-ATE Selection Matrix. *Solid State Technology* (November 1982), pp. 81–88.
23. Hopkins, D. S. *New Product Winners and Losers*. Conference Board Report No. 773, 1980.
24. Johnson, Richard M. Tradeoff analysis of consumer values. *Journal of Marketing Research* 11:121–127 (May 1974).
25. Keeney, R. L. *Siting Energy Facilities*. New York: Academic Press, 1980.
26. Keeney, R. L. Measurement scales for quantifying attributes. *Behavioral Sciences* 26:29–36 (January 1981).
27. Keeney, Ralph L. and Kirkwood, Craig W. Group decision making using cardinal social welfare functions. *Management Science* 22:430–437 (December 1975).
28. Keeney, R. L. and Raiffa, H. *Decisions with Multiple Objectives*. New York: Wiley, 1976.
29. Maidique, M. A. and Zirger, B. J. A study of success and failure in product innovation: the case of the U.S. electronics industry. *IEEE Transactions on Engineering Management* EM-31:192–203 (November 1984).
30. Moriarty, Rowland T. *Industrial Buying Behavior*. Lexington, MA: Lexington Books, 1983.
31. Qualls, William, Olshavsky, Richard W. and Michaels, Ronald E. Shortening of the PLC—an empirical test. *Journal of Marketing* 45:76–80 (1981).
32. Rothwell, R., Freeman, C., Horsley, A., Jervis, T. T. P., Robertson, A. B., and Townsend, J. SAPPHO updated—Project SAPPHO Phase II. *Research Policy* 3:258–291 (1974).

33. Shocker, Allan D. and Hall, William G. Pretest market models: a critical evaluation. *Journal of Product Innovation Management* 3:86–107 (June 1986).
34. Urban, Glen L. and Hauser, John R. *Design and Marketing of New Products*. Englewood Cliffs, NJ: Prentice-Hall, 1980.
35. Von Hippel, Eric. Lead users: a source of novel product concepts. *Management Science* 32:791–805 (July 1986).
36. von Winterfeldt, D. and Edwards, W. *Decision Analysis and Behavioral Research*. New York: Cambridge University Press, 1986.
37. Wilkie, William L. and Pessemier, Edgar A. Issues in marketing's use of multi-attribute models. *Journal of Marketing Research* 10:428–441 (November 1973).

Appendix: Independence Concepts and the Selection of a Value Function

To define the independence concepts, let us assume that we have n criteria denoted $X_1, \ldots, X_n$ with x_i being a level of criterion X_i, $i=1, \ldots n$. Thus, a prospective product can be described by the vector $x = (x_1, \ldots, x_n)$. A value model is then a function v that assigns higher numbers to preferred sets of product characteristics. That is, $v(x) > v(x')$ if and only if x is preferred to x′ by the customer whose evaluation is of interest.

Two independence concepts are most important in developing value models.

Preferential Independence. The pair of criteria $[X_1, X_2]$ is preferentially independent of the other criteria $X_3, \ldots, X_n$ if the preference order for products involving only changes in the levels of X_1 and X_2 does not depend on the levels at which $X_3, \ldots, X_n$ are fixed.

Weak Difference Independence. Criterion X_1 is weak-difference independent of criteria $X_2, \ldots, X_n$ if the order of preference differences between pairs of X_1 levels does not depend on the level at which criteria $X_2, \ldots, X_n$ are fixed.

Using these independence concepts, numerous value models can be developed [25]. They are all constructed hierarchically from the following result, proven in Dyer and Sarin [19].

Result

Given criteria $X_1, \ldots, X_n$, $n \geq 3$, a model of value differences v with the form

$$\begin{aligned} v(x_1, \ldots, x_n) &= \sum_{i=1}^{n} k_i v_i(x_i) \\ &+ k \sum_{i=1}^{n} \sum_{j>1} k_k k_j v_i(x_i) v_j(x_j) \\ &+ k \sum_{i=1}^{n} \sum_{j>1} \sum_{h>1} k_i k_j k_h v_i(x_i) v_j(x_j) v_h(x_h)] \qquad \text{(A.1)} \\ &+ \cdots + k^{n-1} k_1 \cdots k_n v_1(x_1) \cdots v_n(x_n) \end{aligned}$$

exists if and only if $[X_1, X_i]$, $i=2, \ldots, n$ is preferentially independent of the other criteria and if X_1 is weak-difference independent of the other criteria.

To determine v in Eq. (A.1), we need to assess v_i, $i=1, \ldots, n$ on a 0 to 1 scale and the scaling constants k_i $i=1, \ldots n$. The additional constant k concerning synergy among criteria is determined from the k_i.

If $\sum_{i=1}^{n} k_i = 1$, then $k = 0$, and Eq. (A.1) reduces to the additive form

$$v(x_1, \ldots, x_n) = \sum_{i=1}^{n} k_i v_i(x_i). \qquad \text{(A.2)}$$

When $\sum_{i=1}^{n} k_i \neq 1$, then $k \neq 0$ and there is a value synergy among the criteria. Then, multiplying each side of Eq. (A.1) by k, adding 1, and factoring yields

$$kv(x_1, \ldots, x_n) = \prod_{i=1}^{n} [1 + kk_i v_i(x_i)], \qquad \text{(A.3)}$$

which is referred to as the multiplicative form. Any v_i in either Eq. (A.2) or Eq. (A.3) can itself be a multiattribute model, so these results can be used in a nested fashion to create more complex models if appropriate.