Conjoint Analysis in Consumer Research: Issues and Outlook

PAUL E. GREEN
V. SRINIVASAN*

Since 1971 conjoint analysis has been applied to a wide variety of problems in consumer research. This paper discusses various issues involved in implementing conjoint analysis and describes some new technical developments and application areas for the methodology.

The modeling of consumer preferences among multiattribute alternatives has been one of the major activities in consumer research for at least a decade. Undoubtedly, the expectancy-value class of attitude models (Fishbein 1967; Rosenberg 1956) has occupied more researchers' time and journal pages than any other approach. However, a more recent contender, conjoint analysis, shows indications of coming into its own as a practical set of methods for predicting consumer preferences for multiattribute options in a wide variety of product and service contexts.

The purpose of this paper is to trace the development of conjoint methodology and relate it to relevant topics in applied psychology, decision theory, and economics. We then discuss the merits and demerits of the alternatives that have been proposed for implementing the different steps in conjoint analysis. Next, we focus on reliability and validity testing of the methodology. The discussion then proceeds to applications of conjoint analysis to the evaluation of products and services in the public and private sectors. The paper concludes with a brief discussion of some new developments in methodology and application areas.

Readers who are unfamiliar with conjoint analysis may want to read Green and Wind (1975) before attempting a detailed study of this paper.[1]

CONJOINT ANALYSIS IN REVIEW

While the foundations of the field go back to at least the 1920s, it is generally agreed that 1964 marks the start of conjoint measurement, with the seminal paper by Luce, a mathematical psychologist, and Tukey, a statistician (Luce and Tukey 1964). Shortly thereafter, a number of theoretical contributions (Krantz 1964; Tversky 1967) and algorithmic developments (Kruskal 1965; Carroll 1969; Young 1969) appeared.

Conjoint measurement, as practiced by mathematical psychologists, has primarily been concerned with the conditions under which there exist measurement scales for both the dependent and independent variables, given the order of the joint effects of the independent variables and a prespecified composition rule. Computer programs have been developed and applied in examining whether a set of data meets the necessary conditions for applying various composition rules (e.g., additive) hypothesized by the researcher (Ullrich and Painter 1974; Barron 1977). However, applications by psychometricians and consumer researchers have emphasized the scaling aspects—finding specific numerical scale values, *assuming* that a particular composition rule applies, possibly with some error. Accordingly, it now seems useful to adopt the name, "conjoint analysis," to cover models and techniques that emphasize the transformation of subjective responses into estimated parameters.

While conjoint methodology was discussed briefly in the working paper by Green and Rao (1969) and the book by Green and Carmone (1970), the first detailed, consumer-oriented paper did not appear until 1971 (Green and Rao). Following this, a spate of papers dealing with either algorithms or applications (Green, Carmone, and Wind 1972; Srinivasan and Shocker 1973b; Johnson 1974; Westwood, Lunn, and Beazley 1974) appeared in a variety of journals. Theoretical justification for the multiattribute modeling of consumer preferences was provided in the growing literature on

* Paul E. Green is S.S. Kresge Professor of Marketing, Wharton School, University of Pennsylvania, Philadelphia, PA 19104, and V. Srinivasan is Professor of Marketing and Management Science, Graduate School of Business, Stanford University, Stanford, CA 94305. The authors wish to thank all those contributors to the field who were kind enough to send various bibliographic materials relating to conjoint analysis. The authors also extend their appreciation to John Hauser, Subrata Sen, and the referees for helpful comments on an earlier version of this paper.

[1] For a technical description of parameter estimation algorithms with illustrations, the reader is urged to see the excellent review of conjoint analysis by Rao (1977). A monograph-length treatment of conjoint analysis and related procedures for analyzing multiattribute data can be found in the text by Green and Wind (1973).

the Fishbein-Rosenberg class of expectancy-value models and the new economic theory of consumer choice (Lancaster 1971; Ratchford 1975). However, economists have generally been most interested in the aggregate implications of multiattribute utility structures and less concerned with estimation of individual utility functions per se.

As Wilkie and Pessemier (1973) have observed, expectancy-value models draw upon a *compositional* or build-up approach in which the total utility for some multiattribute object is found as a weighted sum of the object's perceived attribute levels and associated value ratings, as *separately* (and explicitly) judged by the respondent. In contrast, conjoint methodology is based on a *decompositional* approach, in which respondents react to a set of "total" profile descriptions. It is the job of the analyst to find a set of *part worths* for the individual attributes that, given some type of composition rule (e.g., an additive one), are most consistent with the respondent's overall preferences. Furthermore, a key distinction between these two approaches lies in the predominant purpose for which each approach is used. Users of conjoint analysis have generally emphasized predictive validity and regarded explanation largely as a desirable (but secondary) objective, while the converse has generally been true for the expectancy-value theorists.

Conjoint analysis, both in spirit and computational detail, is closely related to two other developments in applied psychology—the modeling of clinical judgments and functional measurement.

Since 1960, the modeling of clinical judgments has been the principal pursuit of a group of behavioral scientists initially identified with the Oregon Research Institute (Dawes and Corrigan 1974). This approach has involved decompositional modeling of subjects' responses to profile descriptions representing such diverse topics as gastric ulcer symptoms (Hoffman, Slovic, and Rorer 1968), psychological test scores (Goldberg 1968), and student applications for graduate study (Dawes 1971). Multiple regression has been the main technique for parameter estimation in these models, although analysis of variance (ANOVA) has been applied occasionally.

In clinical judgment modeling the dependent variable is, for instance, an admissions officer's overall evaluation of the likelihood of success in graduate study of an applicant described in terms of the predictors: individual Graduate Record Examination score, undergraduate grade-point average, and rating of the undergraduate institution's quality. The standardized partial regression coefficients (or beta weights) are often interpreted as measures of the relative importance of predictors in estimating the dependent variable.

Functional measurement, as originally proposed by Anderson (1970), also employs a decompositional approach utilizing, in this case, ANOVA and full factorial designs. Functional measurement has been used for both parameter estimation and model testing (e.g., to see if certain theoretically predicted interaction effects occur) in such content areas as information integration, attitude change theory, person perception, decision theory, and consumer behavior.

In terms of methodology, functional measurement proceeds similarly to clinical judgment modeling and conjoint analysis. For example, a researcher may be interested in how a subject judges the friendliness of a person described in terms of levels on two factors, say, boldness and laziness. Each of the two factors in the experiment would consist of a set of levels, and all levels of the first factor would be crossed with all levels of the second. Overall friendliness ratings are then decomposed by ANOVA procedures to yield a scale value for each level of each of the two factors. Moreover, assuming that replicate judgments are available, the researcher can check for the significance of interaction effects.

A very different set of procedures has been developed by Keeney and Raiffa (1976) for multiattribute utility estimation in normative contexts. The form of the utility function is derived deductively from a set of assumptions. The parameters of the utility function are obtained from tradeoff judgments and from preferences for alternative gambles (lotteries). The data collection is considerably more complex than conjoint analysis. Although the utility function approach employs decompositional modeling in several parts of its implementation (Hauser and Urban 1977), it is not an estimation procedure in that virtually no measurement error is assumed (in contrast to the just described statistical approaches).

In what follows, we use the term conjoint analysis broadly to refer to any decompositional method that estimates the *structure* of a consumer's preferences (e.g., part worths, importance weights, ideal points) given his/her overall evaluations of a set of alternatives that are prespecified in terms of levels of different attributes. This type of estimation has been referred to as "external analysis" in the psychometric literature (Carroll 1972).

ISSUES IN IMPLEMENTING CONJOINT ANALYSIS

Because of the substantial amount of among-person variation in consumer preferences, conjoint analysis is usually carried out at the individual level. The form of the preference model (composition rule) is generally assumed to be the same for all individuals, but the parameters of the model are permitted to vary across the sample of individuals from the relevant target population.

Several alternate means exist for identifying the attributes which are relevant to consumers in forming their preferences (Alpert 1971). A preliminary data collection effort, questioning consumers regarding attributes important to them, usually helps in identifying those attributes that are most frequently regarded as relevant (Braun and Srinivasan 1975). Kelly's (1955) repertory grid, focus group interviews, or judgments of product managers, retailers and others knowledgeable about the product/service and its uses can be used for this purpose. The more difficult and often subjective task is to reduce the number of attributes to a manageable size so that the estimation procedures are reliable while at the same time accounting for consumer preferences sufficiently well.

The various steps in conjoint analysis and the alternative methods of implementing each of the steps are summarized in the table. We discuss each of these steps in turn and indicate the empirical results, if any, that are available for comparing the alternatives. There is considerable scope for empirical research in this area to determine the methods that are most appropriate for each of the steps.

Among the numerous combinations of methods that can be chosen for the different steps, some of the combinations are not feasible and these will be pointed out in the ensuing discussion. To date, most applications of conjoint analysis have utilized only a few of the many possible combinations. By focusing attention on the steps themselves, better overall combinations may emerge. A worthwhile goal for empirical research is to identify the combination of methods that provides the maximum predictive validity for a given amount of respondent time (or research budget). Of course, the best combination will probably depend on factors such as the type of product/market, the number of relevant attributes, the type of respondent, and so on. Future studies might entail several combinations, each dealing with a separate part of the data collection and analysis.

TABLE

STEPS INVOLVED IN CONJOINT ANALYSIS

Step	Alternative methods
1. Selection of a model of preference	Vector model, ideal-point model, part-worth function model, mixed model
2. Data collection method	Two-factor-at-a-time (trade-off analysis), full-profile (concept evaluation)
3. Stimulus set construction for the full-profile method	Fractional factorial design, random sampling from multivariate distribution
4. Stimulus presentation	Verbal description (multiple cue, stimulus card), paragraph description, pictorial or three-dimensional model representation
5. Measurement scale for the dependent variable	Paired comparisons, rank order, rating scales, constant-sum paired comparisons, category assignment (Carroll, 1969)
6. Estimation method	MONANOVA, PREFMAP, LINMAP, Johnson's nonmetric tradeoff algorithm, multiple regression, LOGIT, PROBIT

Preference Models

First, let

$$p = 1, 2, \ldots, t \tag{1}$$

denote the set of t attributes or factors that have been chosen. Next, let y_{jp} denote the level of the pth attribute for the jth stimulus. We first consider the case where y_{jp} is inherently a continuous variable (e.g., travel time or price). The case of categorical (or polytomous) attributes will be considered later. The vector model of preference, referred to as the Composite Criterion Model by Srinivasan and Shocker (1973b) and Parker and Srinivasan (1976), posits that the preference s_j for the jth stimulus is given by

$$s_j = \sum_{p=1}^{t} w_p y_{jp}, \tag{2}$$

where the $\{w_p\}$ are the individual's weights for the t attributes. Thus, the vector model is identical in mathematical form to the Fishbein-Rosenberg class of multiattribute models. As remarked earlier, the weights $\{w_p\}$ will, in general, be different for different individuals in the sample. Geometrically, the preference s_j can be represented as the projection of the stimulus point $\{y_{jp}\}$ on the vector $\{w_p\}$ in the t-dimensional attribute space.

The *ideal-point model* posits that the preference s_j is negatively related to the squared (weighted) distance d_j^2 of the location $\{y_{jp}\}$ of the jth stimulus from the individual's ideal point $\{x_p\}$, where d_j^2 is defined as

$$d_j^2 = \sum_{p=1}^{t} w_p(y_{jp} - x_p)^2. \tag{3}$$

Thus, stimuli which are closer to the ideal point (smaller d_j^2) will be the more preferred ones (larger s_j). It turns out that the simultaneous estimation of $\{w_p\}$ and $\{x_p\}$ is feasible for the weighted Euclidean measure of distance as specified in equation (3). If, however, the exponent 2 in equation (3) is replaced by a general Minkowski metric r, the estimation of $\{x_p\}$ becomes very difficult. Fortunately, however, the Euclidean metric is often a close enough approximation to the general Minkowski metric (Green 1975).

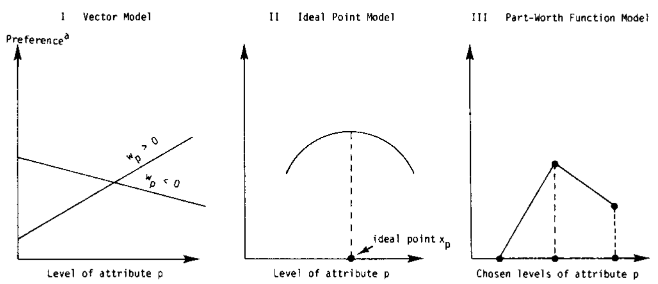
The *part-worth function* model posits that

$$s_j = \sum_{p=1}^{t} f_p(y_{jp}), \tag{4}$$

where f_p is the function denoting the part worth of different levels of y_{jp} for the pth attribute. In practice, $f_p(y_{jp})$ is estimated only for a selected set of levels for y_{jp} (usually three or four), with the part worth for intermediate y_{jp} obtained by linear interpolation. Thus the

FIGURE

ALTERNATIVE MODELS OF PREFERENCE



[a] Preference for different levels of attribute p while holding the values for the other attributes constant.

part-worth function is represented as a piecewise linear curve. To determine the part worth for a value of y_{jp} outside the range of estimation, extrapolation of the piecewise linear function would be needed and the validity of this procedure is questionable. (Hence, the researcher should try to employ the full range of the attribute, wherever practical.) Still, the part-worth function approach has received wide acceptance due, in part, to the ready interpretability of the graphicelly displayed attribute part-worth functions. The three models of preference are illustrated in the figure.

The part-worth function model provides the greatest flexibility in allowing different shapes for the preference function along each of the attributes. In particular, by defining $f_p(y_{jp}) = -w_p(y_{jp} - x_p)^2$ we get the ideal-point model and by setting $f_p(y_{jp}) = w_p y_{jp}$ we obtain the vector model. Similarly, the ideal-point model is more flexible than the vector model since it can be shown (Carroll 1972) that the vector model is a special case of the ideal-point model as $x_p \to \pm \infty$. Intuitively, as $x_p \to +\infty$, preference along the pth dimension increases as y_{jp} increases (since the ideal is at plus infinity) and this is essentially the same as the vector model with $w_p > 0$.

Although the part-worth function model seems to be the most attractive in terms of being compatible with any arbitrary shape for the preference function, this benefit comes at the cost of having to estimate additional parameters (thereby lowering their reliability) and the need to approximate intermediate values by linear interpolation. In particular, estimation of the vector model involves only the t parameters $\{w_p\}$. For the ideal-point model, $2t$ parameters have to be estimated, namely $\{w_p\}$ and $\{x_p\}$. If there are q levels, say, for each of the t attributes then $(q - 1)t$ parameters have to be estimated for the part-worth function model. Replacing $f_p(y_{jp})$ by $f_p(y_{jp}) + a_p$ does not alter the model in equation (4) in any essential way so that the part worth for level 1, say, can be taken to be zero without any loss of generality. Consequently, only $(q - 1)$ parameters need to be estimated for the pth attribute.

To summarize the discussion so far, the flexibility of the shape of the preference model is greater as we go from the vector to the ideal point to the part-worth function models; however, the reliability of the estimated parameters is likely to improve in the reverse order. Consequently, from the point of view of predictive validity, the relative desirability of the three models is not clear. Thus a priori notions of the shape of the part-worth function could help us in the choice of an appropriate model. One may always prefer greater durability (vector), smaller waiting time (vector), but may prefer moderate levels of sweetness or size of automobile (ideal point). However, one may prefer maximum temperature levels for both iced and hot tea (Carroll 1972) and have a lower preference for in-between temperature levels (part-worth function). If the attribute is categorical, (e.g., mode of travel—auto versus carpool versus public transit; type of educational institution—junior college, private university, state university), we are forced to use the part-worth function model. It is, of course, probable that for some attributes the vector model would be the best while for some others the ideal-point or part-worth models may be more appropriate.

We may combine the features of the three models to formulate a *mixed model*. It is well known (e.g., Green and Tull 1978, pp. 297–8) that a polytomous attribute with k levels can be converted into $(k - 1)$ dummy variables, where the ith dummy variable takes the value 1 for the ith level and 0 otherwise. (The kth level serves as the reference.) Thus the part-worth model can be converted to the vector model with the use of dummy variables. (The situation is analogous to the analysis of factorial experiments through multiple regression.) Similarly, if we consider the component of the ideal-point model along the pth attribute and relate the squared distance to preference by a negative sign, we obtain

$$f_p(y_{jp}) = -w_p(y_{jp}^2 - 2x_p y_{jp} + x_p^2). \qquad (5)$$

Equation (5) may be rewritten as,

$$f_p(y_{jp}) = a_p + b_p y_{jp}^2 + c_p y_{jp}, \qquad (6)$$

where $a_p = -w_p x_p^2$, $b_p = -w_p$ and $c_p = 2w_p x_p$. Thus, the essential nature of the ideal-point model can be captured in the vector model by considering the pseudo-attribute y_{jp}^2 in addition to y_{jp} (Carroll 1972). Furthermore, if $f_p(y_{jp})$ is substantially nonlinear (convex or concave), it may also be capable of being parsimoniously represented by equation (6). It may also be worthwhile to consider higher order polynomial terms, such as y_{jp}^3. Pekelman and Sen (1978) show that if the utility function has the form given in equation (6), then the estimation of that function would give better predictive results than a part-worth function approach combined with linear interpolation. However, if the functional form is likely to be very different from equation (6), the part-worth function approach may be appropriate.

A mixed model that may capture the advantages of all three models can be implemented by the notation,

$$s_j = \sum_{q=1}^{T} v_q z_{jq}, \qquad (7)$$

where T is the total number of pseudo-attributes (also the total number of estimated parameters) and the $\{z_{jq}\}$ are defined from the y_{jp}, as follows:

1. attributes where the preference is expected to be monotone and approximately linear: z_j is defined to be equal to y_j;
2. attributes for which the preference is expected to be either substantially nonlinear (convex or concave) or of the ideal-point type: for each attribute p, two z variables, one equal to y and the other equal to y^2, are defined;
3. attributes which are categorical, or the preference function is not well approximated by equation (6): for each attribute p with k levels, $(k - 1)$ dummy variable would be defined.

If the dependent variable (overall preference) is measured on an interval scale and multiple regression is used as the estimation procedure, then statistical testing of equation (7) could guide us in the choice of vector versus ideal point versus part-worth model for the pth attribute, e.g., if the coefficient of the y^2 term is not significant then the vector model could be chosen over the ideal-point model.

Much empirical research has examined whether consumers actually use the linear-compensatory model (i.e., the vector model) rather than the seemingly simpler evaluation models such as the lexicographic and conjunctive (cut-off) rules (Russ 1971; Wright 1975; Hansen 1976). (Actually, the lexicographic model is a special case of the vector model in equation (2) where the weight for the most important attribute is considerably larger than the second most important attribute whose weight, in turn, is considerably larger than the third most important attribute, etc.) This research has found that some consumers use each of the models, but generally prefer those requiring simpler processes. However, for predictive validity this problem is not as serious as it may initially seem. This is because the compensatory model of conjoint analysis can approximate the outcomes of other kinds of decision rules quite closely. In fact, a recent study by Berl, Lewis, and Morrison (1976) of high school seniors' choice of colleges showed that the linear compensatory model was more consistent with the respondents' actual behavior than the lexicographic and conjunctive rules.

Dawes and Corrigan (1974), Rorer (1971), and Green and Devita (1975b) discuss the general conditions under which linear-compensatory models perform well. Three of these conditions prevail in many situations in the context of modeling preferences for real brands/services: (i) the preference function is monotone (increasing or decreasing) over increasing levels of an attribute while holding other attributes constant, (ii) there are errors in the measurement of attribute levels (possibly because of perceptual differences across consumers), and (iii) the attributes tend to be correlated. Thus, even if the respondent's information processing strategy and decision model are complex, the compensatory model can usually produce good *predictions*, assuming that this is the main concern of the researcher. Furthermore, the ideal-point and the part-worth function models are more general than the linear-compensatory model, in that only additivity, not linearity, is assumed (i.e., only interaction terms are omitted). Consequently the predictive validity of these models can be expected to be very good.

In situations where specific interaction effects can be expected a priori, these could be captured in the mixed model by adding pseudo-attributes of the form $z = y_1 y_2$. In some other situations, interaction effects can be taken into account by combining two categorical attributes into one. For example, the researcher may believe that car roominess and gas mileage may interact in the following way:

- A respondent generally prefers a roomier car to a cramped one.
- A respondent generally prefers higher gas mileage to lower gas mileage.
- However, a roomy car *with* high gas mileage is evaluated much higher than the sum of the separate part worths.

If so, the researcher can construct a four-level factor from the two two-level factors and test his surmise using the *main-effect* estimates of the four-level superfactor.

Data Collection Alternatives

Data collection procedures in conjoint analysis have largely involved variations on two basic methods:

- the two-factor-at-a-time procedure, and
- the full-profile approach.

The two-factor-at-a-time procedure, also referred to as the "trade-off procedure" (Johnson 1974), considers factors (attributes) on a two-at-a-time basis. The respondent is asked to rank the various combinations of each pair of factor levels from most preferred to least preferred. Panel I of the exhibit shows an illustration of the approach, as applied to consumer evaluations of steel-belted radial replacement tires.

The two-factor-at-a-time procedure is simple to apply and reduces information overload on the part of the respondent. It also lends itself easily to mail questionnaire form, since no special props are needed. However, in actual problems, it displays a number of limitations:

EXHIBIT

ALTERNATIVE DATA COLLECTION METHODS

I. Two-Factor-at-a-Time Approach

Tire brand	Tread life: 30,000 miles	40,000 miles	50,000 miles
Goodyear	8	4	1[a]
Goodrich	12	9	5
Firestone	11	7	3
Sears	10	6	2

[a] 1 denotes the best-liked combination and 12 denotes the least-liked combination for a hypothetical respondent.

II. Full-Profile Approach (Sample stimulus card)

Brand
SEARS
Tread Life
50,000 MILES
Sidewall
WHITE
Price
$55

- By decomposing the overall set of factors to two-at-a-time combinations, there is some sacrifice in realism. Moreover, respondents are usually unclear as to what should be assumed about the t-2 factors that are *not* being considered in a specific evaluation task. For instance, in Panel I of the exhibit it is reasonable to expect that as the tread life increases the price of the tire is also likely to increase. Consequently, when the attributes of a product or service are correlated (e.g., for technological reasons) what the rank order in a particular table corresponds to is not clear.
- With, say, only six factors, each at four levels, the respondent could be asked to fill out 15 tables, each consisting of 16 cells. While partially balanced incomplete block designs (Green 1974) or related procedures (Johnson and VanDyk 1975) can be used to reduce the number of two-way tables, the total number of required judgments is still quite large.
- There is some tendency for respondents either to forget where they are in the table or to adopt patternized types of responses, such as always attending to variations in one factor before considering the other (Johnson 1976).
- The procedure appears to be most suited to *verbal* descriptions of factor combinations, rather than pictorial or other kinds of iconic representations. For example, a study of package designs in which color, logo, size, and shape can be simultaneously varied and portrayed graphically would not lend itself well to this approach.

The full-profile approach (also referred to as the concept evaluation task) utilizes the complete set of factors, as shown by the illustrative stimulus card for a four-factor design in Panel II of the exhibit. The major limitation of this approach is the possibility of information overload and the resulting temptation on the part of the respondent to simplify the experimental task by ignoring variations in the less important factors or by simplifying the factor levels themselves. The conjoint results obtained under such conditions may not be representative of the real life behavior of the individual where he/she may have more time and motivation to deliberate on the choice from among a small set of alternatives.

Because of the information overload problem, the full-profile procedure is generally confined to, at most, five or six factors in any specific sort. If a larger number of factors is entailed—some industrial studies have included 25 or more factors, each at from two to six levels—the analyst is more or less forced to incorporate "bridging-type" factors. The idea here is to prepare several card decks in which the full set of factors is first split into subsets of five or six factors each. Each card deck is then composed of factor combinations that involve, say, five factors, only. In each case one or two factors are common across decks so that they provide a basis for linking part-worth functions across the various subsets of factors. (For an illustration, see Hopkins, Larréché, and Massy 1977, p. 371.)

The full-profile approach—when implemented by various kinds of fractional factorial designs (to be discussed later)—entails fewer judgments to be made by the respondent, although each single judgment is more complex. As per later discussion on response (or dependent) variable data, the two-factor-at-a-time procedure provides only a set of rank orders while the full-profile approach can employ either a rank order or ratings (e.g., on a seven-point scale from "least liked" to "most liked"). Such flexibility in scaling seems desirable. In the full-profile approach, each stimulus card can potentially be assigned any level on a continuous attribute. However, in the two-factor-at-a-time approach, each continuous attribute is assigned only a few levels because of the need to construct manageable trade-off tables. (This is also true for the full-profile approach using fractional factorial designs.)

The main argument that seems to favor the full-profile approach is that it gives a more realistic description of stimuli by defining the levels of each of the factors and possibly taking into account the potential environmental correlations between factors in real stimuli. On the other hand, it has the disadvantage of making the task difficult for the respondent by having to consider several factors at one time. Based on these two considerations, we would speculate that in contexts where the environmental correlation between factors is large and the number of factors on the stimulus card is small (but greater than two), the full-profile approach is likely to be better in terms of predictive validity. However, if the environmental correlation between the factors is small and the number of factors on the stimulus card is large, the two-factor-at-a-time approach is likely to be better.

Four empirical studies have compared the two approaches to data collection. Montgomery, Wittink, and Glaze (1977), in a study of job choice by MBAs, found that the two-factor-at-a-time approach yielded higher predictive validity than the full-profile approach.

Similarly, Alpert, Betak, and Golden (1978) found that the goodness-of-fit to input data was better for the two-factor-at-a-time approach in their study of commuters' choice of transportation modes. However, Jain, Acito, Malhotra, and Mahajan (1978) found that the two methods yielded approximately the same level of cross-validity in the context of choosing checking accounts offered by various banks. Also, Oppedijk van Veen and Beazley (1977) found that the utilities determined by the two methods were roughly similar in the context of a durable good product class.

The first two studies used eight and nine attributes, respectively, and the resulting information overload may have biased the results against the full-profile approach. In comparison, the third and fourth studies used five and three attributes respectively. In all four studies, the problem contexts were such that there were no substantial environmental correlations across factors. (In any event, the full-profile approaches used orthogonal designs that exhibited no interattribute correlations.) Thus the results are not inconsistent with the conjecture advanced earlier comparing the effectiveness of the two alternative data collection procedures.

Stimulus Set Construction for the Full-Profile Method

The number of brands in a product class that a respondent may be familiar with is usually small. Furthermore, real brands and services are usually not distinctive enough to provide reliable estimates of parameters. For these reasons, conjoint analysis is usually done with hypothetical stimulus descriptions. This has the additional advantage of enabling us to compare predicted behavior with the actual behavior of the respondents towards real brands or services. In constructing the stimulus profiles for the full-profile approach several questions arise:

1. How many stimuli do we need to use?
2. What should be the range of attribute variation and interattribute correlation in constructing the stimuli?
3. How should the stimuli themselves be constructed?

The number of stimuli should obviously depend on the number of estimated parameters. From multiple regression theory (Darlington 1968), we know that the expected mean squared error of prediction is given by $(1 + T/n)\sigma^2$ where T is the number of estimated parameters, n is the number of stimuli to be evaluated, and σ^2 is the unexplained (error) variance in the model. Thus the ratio (n/T) should be as large as possible to minimize the increment to prediction error over and above the error (σ^2) that is unavoidable. For a given T, as n increases from $2T$ to $5T$ the prediction error decreases by 20 percent. (As the formula indicates, it is more appropriate to think in terms of the n/T ratio rather than $n - T$, the degrees of freedom.) However, a respondent typically takes about 20–30 minutes for $n = 25$ in a five attribute, full-profile rank ordering task. Given the usual considerations of maintaining the respondent's interest in the task, it is often difficult to increase n much above 30.

In deciding the range of variation of attribute levels and interattribute correlation (e.g., between auto versus bus and travel time to work, or between horsepower rating and gas mileage of cars), two conflicting considerations are relevant. The use of stimulus descriptions similar to those that currently exist (similar in terms of ranges of attribute levels and environmental correlations) will increase believability and hence validity of the preference judgments. On the other hand, if we make the ranges for attribute levels much larger than reality and/or decrease the magnitude of interattribute correlations to zero (as is implied by orthogonal designs), we may decrease believability and hence validity. But orthogonal designs and/or larger ranges for attribute values have the advantage of improving the accuracy of the parameter estimates for a given level of validity for the preference judgments. Thus, the extreme strategy of using descriptions similar to those that exist has the disadvantage of loss of accuracy in estimation while the alternate extreme strategy of using an orthogonal design and/or ranges for attribute values much larger than reality has the disadvantage of decreasing the validity of the respondent's preference judgments.

We would recommend, therefore, that the ranges be made larger than reality, but not so large as to be unbelievable. Further, we would recommend making interattribute correlations in the hypothetical descriptions smaller (in absolute value) than the environmental correlations that exist in real stimuli, but not to make them so small as to be unbelievable. However, if the environmental interattribute correlations are small to start with, there will be virtually no loss in believability by using orthogonal designs, and there is everything to gain in terms of accuracy of the estimated parameters.

The hypothetical stimulus descriptions can be constructed in either of two ways. The more popular method has been to define a number of levels (say, three or four) for each of the attributes over the range of attribute variation. The researcher should pretest the levels of the continuous factors to insure that they are far enough apart to be considered as realistically distinct. For categorical attributes the feasible levels are readily available but the selection of which levels are "representative" of the factor often requires careful study. If a full factorial design is used, the number of possible stimuli quickly becomes very large (e.g., with three attributes at three levels each and two attributes at two levels each, the total number of possible descriptions is $3^3 \times 2^2 = 108$). Green (1974) has suggested the use of various types of *fractional factorial designs* to reduce the number of combinations to a manageable size while at the same time maintaining orthogonality. These types of designs assume away most (sometimes all) interaction effects, which is realistic given the type of preference models discussed earlier. For instance, the preceding numerical example with five attributes can be reduced to an orthogonal main-effects plan with only 18 stimuli (Green and Wind 1975, p. 108).

A discussion of various kinds of fractional factorials, including both orthogonal and nonorthogonal designs, can be found in the paper by Green, Carroll, and Carmone (1978). This paper describes both main-effects plans, in which no interactions can be separately estimated, and less restrictive plans (such as Resolution IV and Compromise designs) that permit estimation of all main effects *and* selected two-factor interactions, without an inordinate increase in the number of stimuli. This paper also shows how various basic designs can be modified to handle different combinations of factor levels. In addition, a number of references to the specialized literature in this area are included.

If there is a substantial amount of environmental correlation between some of the attributes, an orthogonal design can produce some stimuli which may not be believable. If the environmental correlations are very high (e.g., 0–55 mph acceleration time, gas mileage, horsepower rating, and top speed), the researcher may wish to prepare a composite factor covering all four subfactors whose separate levels show various gradations of performance-mileage. Each level of the composite would reflect the subfactor correlations that exist technologically so that a 300 hp engine does not coexist with a 35 mpg fuel consumption. However, with this approach, it is no longer possible to separate the effects of the subfactors contained in the composite.

If one wishes to capitalize on the high efficiency of orthogonal arrays and other fractional factorials, factor independence is to be sought wherever possible. If some of the profiles turn out to be unbelievable, other orthogonal displays can be tried by permuting sets of factor levels. Finally, if worse comes to worst, a few profiles may have to be deleted or modified to incorporate correlated factor levels. However, methods such as the two-factor-at-a-time approach are less amenable to the correlated-factor approach, since all combinations of factor levels are typically displayed for evaluation.

An alternate procedure for creating the stimulus descriptions is that of random sampling from a multivariate distribution. Assuming for the moment that all the attributes are continuous, a multivariate distribution can be defined given the means, standard deviations (derived from the ranges), and interattribute correlations. The stimulus descriptions could then be randomly drawn from a multivariate normal distribution (Naylor, Balintfy, Burdick, and Chu 1966, pp. 97–9). Dichotomous attributes (e.g., auto versus bus as mode of travel) could also be handled in the above framework by defining a proxy continuous random variable and a cut-off value (e.g., if $x \geq 0$, then auto, if $x < 0$, then bus). For categorical variables with more than two categories, the values could be randomly drawn from a discrete distribution. If this categorical attribute is to be correlated with other continuous attributes, different sets of parameters for the multivariate distribution would have to be defined, depending on the value assigned to the categorical attribute.

Thus, if the attributes are to be correlated in the stimulus set, the random sampling procedure just described provides a systematic, although cumbersome procedure. By creating random descriptions in excess of the required number n, it is possible to delete descriptions which are dominated (i.e., have a less desirable attribute level on each of the attributes) by other stimuli. By trial and error, it is generally possible to construct the n descriptions so that none of the stimuli dominates any of the remaining $(n - 1)$ descriptions (Parker and Srinivasan 1976). Although this procedure is time consuming, it may still be desirable from the point of view of getting maximum potential information from the respondent's evaluative judgments.

To illustrate this point, consider a two-attribute product class with three levels for each attribute. Assume that the part-worth function is monotone (increasing or decreasing) over the three levels, e.g., as might be the case with attributes such as tread mileage, price, or waiting time. Without loss of generality, let us assume that greater levels are preferred to smaller levels on the attribute. Then in the (3×3) orthogonal design with nine descriptions it can be shown that only nine of the 36 potential paired comparisons have any information content. For example, the (1,1) and (3,3) descriptions provide no information at all; they are, by the monotonicity assumption, the worst and the best stimuli. Furthermore, the pairs comparing, say, the stimulus (3,2) to the stimuli (3,1), (2,2), (2,1), (1,2) and(1,1) also convey no information.

The random sampling procedure seems to be well suited to estimating ideal-point type models, as portrayed by equation (3). For example, in the fractional-factorial type approach, that uses only a few levels for each attribute, it is difficult to distinguish well between alternative ideal-point locations that lie between the adjacent levels used in the design. In the random sampling procedure, many levels of the attribute are likely to be used so that a finer discrimination regarding the ideal-point location can be made.

The fractional-factorial type design, on the other hand, is considerably easier to develop. If no attribute correlations are desired, it is likely to produce more accurate parameter estimates. It involves no limiting assumption such as multivariate normality which needs to be assumed in the random sampling approach. Finally, if the relative importances of the attributes are to be obtained, the orthogonal design produces less ambiguous answers than the correlated factors approach.

Based on these considerations, we would speculate that the fractional-factorial designs are better (in terms of predictive power) than the random sampling approach if the environmental interattribute correlations are not high. The random sampling approach is likely to be better if the attribute correlations are high or if most of the attribute part-worth functions are monotone with changes in attribute levels and ordinal overall preference judgments are obtained.

Stimulus Presentation

To date, the presentation of the hypothetical stimuli in the full profile approach has involved variations and combinations of three basic approaches:

- verbal description (multiple cue stimulus card),
- paragraph description, and
- pictorial representation.[2]

The two-factor-at-a-time approach has primarily used the verbal description approach. However, pictorial representation has been used in a few cases. For instance, Alpert, Betak, and Golden (1978) use pictures with different numbers of drops of gasoline to indicate the different levels of gasoline consumption of alternative modes of transportation.

Panel II of the exhibit illustrates the verbal description approach. In a typical task, the respondent is given n stimulus cards, each card defining the levels of each of the t attributes. The respondent is asked to either rank order them or rate them on a scale. The main advantage with this procedure is its simplicity and the efficiency with which the data can be collected.

Acito (1977) has found that the measured importance of an attribute is to some extent affected by the order or position of the attribute on the stimulus card. To reduce this potential bias, the order of the attributes is usually randomized over respondents. To reduce confusion, the order of attributes is kept the same for all the stimulus cards given to any one respondent. Of course, the stimulus cards themselves are shuffled thoroughly before giving them to a respondent.

Some researchers, such as Hauser and Urban (1977), have adopted the paragraph description approach. The advantage of this approach is that it provides a more realistic and complete description of the stimulus (similar to concept testing in new product development) and has the advantage of simultaneously testing advertising claims. A significant drawback of this procedure is that it limits the total number of descriptions to a small number, so that parameter estimates are likely to be very inaccurate when estimated at the individual level.

Pictorial representation using various kinds of visual props or three dimensional models provides several important advantages over verbal profiles:

- Information overload is reduced since the respondent is not required to read and then visualize large quantities of information.
- Higher homogeneity of perceptions of such things as car roominess or trunk capacity is obtained across respondents.
- The task itself is more interesting and less fatiguing.
- The stimuli are more realistic.

Alpert, Betak, and Golden (1978) report that a combination of pictures and words produced roughly the same results as the purely verbal approach, but the respondents took less time to complete the pictorial task. The primary disadvantage of the pictorial approach is the increased cost and time on the part of the researcher in preparing the stimulus descriptions. Furthermore, there is a danger in the picture displaying information different than the researcher intended (e.g., styling of the car may be conveyed in addition to roominess). The pictorial approach is usually administered by mail or personal interview. Data collection by time-shared computer terminals cannot, in general, employ this approach except through the use of sketches in cathode ray terminals or by providing the pictures to the respondents as props to be used in conjunction with a time-shared terminal. However, the pictorial approach has been successfully used on several occasions involving mailed props, followed by telephone interviews.

Based on these considerations, we feel that the verbal and, particularly, the pictorial approaches are likely to be the best methods of presenting stimulus descriptions, assuming individual-level parameter estimates are to be obtained. Choice of the pictorial approach depends upon the nature of the product class (the importance of imagery) and cost considerations.

Measurement Scale for the Dependent Variable

The various alternatives for defining a measurement scale for the dependent variable can be roughly classified as nonmetric (paired comparisons, rank order) or metric (rating scales assuming approximately interval scale properties, or ratio scales obtained by constant-sum paired comparisons [Torgerson 1958, pp. 105–12]). Depending on the purpose of the study, the measurement can be either in terms of overall preference or intention to buy (likelihood of purchase). The latter criterion is particularly suited to studies of new product classes and services that consumers do not purchase currently.

[2] We expect, however, that future applications of conjoint analysis, particularly in packaged goods (such as foods and beverages), will increasingly employ *actual* products, factorially composed using fractional designs from a set of physical/chemical attributes.

In comparing the metric with the nonmetric measurement scales, it should be noted that even though the dependent variable is nonmetric, the estimated parameters tend to satisfy close to interval-scaled properties, for typical values of n and T, the number of estimated parameters (Colberg 1978). The main advantage of the metric methods is the increased information content potentially present in these scales. However, based on results in nonmetric multidimensional scaling (Green and Carmone 1970, p. 36) we would expect the differential advantage of the metric scale to diminish as the n/T ratio becomes large. The nonmetric methods, on the other hand, have the following advantages:

- Ranked data are likely to be more reliable, since it is easier for a respondent to say which he/she prefers more as compared to expressing the magnitude of his/her preference.
- Data analysis based on a nonmetric dependent variable allows the part-worth functions to be combined in either an additive or multiplicative manner. This is because estimation of an additive model with a nonmetric dependent variable is also consistent with a multiplicative model, since the logarithmic transformation is just one of the permissible monotone transformations of the dependent variable.
- With the two-factor-at-a-time approach the nonmetric method is more appropriate than the metric method. The metric scale value for the dependent variable will necessarily depend on the levels of the $(t - 2)$ missing factors, whereas the rank order of the cells in a tradeoff table need not depend on the levels of the missing factors, except if the attributes are correlated.

The paired-comparison approach is the least efficient, in terms of information obtained per unit time, of all the methods. Its only advantages seem to be the increased reliability of the averaged rank order and the ability to test for intransitivities in the respondent's expressed preferences. All things considered, however, we conjecture that the rank-order approach will fare better in terms of predictive validity than the direct paired-comparisons approach for a given amount of the respondent's time.

In collecting data by means of paired comparisons, one generally collects more data than needed. Recognizing the potential redundancy of paired-comparisons data, a number of researchers (Johnson 1976; Shugan and Hauser 1977) have proposed interactive computer approaches for obtaining preference data. Johnson's procedure, in particular, has been designed to obtain conjoint analysis data from computer terminals placed in shopping malls and other high density locations. This proprietary procedure, the technical details of which are not available, is a sequential method in that it uses the results of earlier evaluations to select subsequent stimulus pairs, so that redundancy is kept to a minimum. Based on still early results, it appears that the number of paired comparisons may be reduced by as much as 25–30 percent without appreciable loss in accuracy. While practical experience with the method is still meager, it looks potentially promising for applications amenable to this type of data collection procedure. However, the relative efficiency of the interactive paired-comparison methods vis-a-vis the rank-order procedure is still not known.

One of the advantages of the rating-scale approach is that it can potentially be administered by mail. On the other hand, the rank-ordering task usually entails a personal interview since the procedure requires a considerable amount of explanation (such as first sorting the cards into two or more piles, corresponding to the more preferred versus the less preferred ones; then sorting the cards in each pile from most preferred to least preferred; and finally merging the different piles of cards and checking the final rank order).

Another way of obtaining interval-scale judgments is the so-called "dollar metric" approach (Pessemier, Burger, Teach, and Tigert 1971). In this method, the respondent compares stimuli A and B and if he/she prefers A to B, states how much the price of A has to increase until he/she will be indifferent between A and B. The results of such paired comparisons are aggregated to obtain an intervally scaled dollar metric of preference. This is a very slow procedure compared to the rating method; furthermore, the results may be influenced by the social biases involved in using dollar differences as a response measure.

The constant-sum method of obtaining ratio-scaled judgments in an interactive mode with a time-shared terminal has been compared (Hauser and Shugan 1977) to simply using the ordinal component of data obtained from paired comparisons. As might be expected, the ratio-scaled data provided better predictive results than the paired comparisons per se. However, this comparison is misleading. For a given amount of respondent's time, a very large number of paired comparisons can be obtained (e.g., by the use of rank order) as compared to the small number of constant-sum paired comparisons. Consequently, for a given amount of respondent's time, the relative desirability of the two methods is still unclear.

Given the conflicting considerations involved in the choice of a metric versus nonmetric dependent variable, additional empirical studies are needed to compare these alternative methods.

Estimation Methods

Parameter estimation methods in conjoint analysis can be roughly classified into three categories:

1. Methods which assume that the dependent variable is, at most, ordinally scaled. Methods in this class are MONANOVA (Kruskal 1965), PREFMAP (Carroll 1972), Johnson's nonmetric tradeoff procedure (Johnson 1973; Nehls, Seaman, and Montgomery 1976), and LINMAP (Srinivasan and

Shocker 1973a, 1973b; Pekelman and Sen 1974). For a more exhaustive list of this class of algorithms, see Rao (1977).

2. Methods which assume that the dependent variable is intervally scaled. Methods in this class are ordinary least squares (OLS) regression (Johnston 1972, Chap. 5), and minimizing sum of absolute errors (MSAE) regression (Srinivasan and Shocker 1973a, pp. 358–60).
3. Methods which relate paired-comparison data to a choice probability model. Methods in this class are LOGIT (McFadden 1976; Ben-Akiva 1973; Gensch, Golob, and Recker 1976; Green and Carmone 1977; Punj and Staelin 1978) and PROBIT (Goldberger 1964, pp. 250–1; Rao and Winter 1977).

In the class of algorithms designed for an ordinal-scaled dependent variable, MONANOVA is restricted to the part-worth function model (equation 4). The remaining approaches can be used for the vector or part-worth function models. For the ideal-point model, LINMAP is best suited since the use of other approaches may lead to negative weights and resulting interpretation difficulties (Srinivasan and Shocker 1973a, p. 338). The algorithms differ from each other in their operational definitions of the poorness-of-fit index, analogous to $(1 - R^2)$ in multiple regression studies, and the optimization method used to determine parameter estimates to achieve the minimum poorness-of-fit. While there is no a priori way to choose between the different poorness-of-fit definitions, empirical predictive validity tests could guide us in this direction.

LINMAP differs from the others in that it uses linear programming as compared to classical calculus methods employed by the other approaches. The use of linear programming enables LINMAP to obtain global optimum parameter estimates, while the other approaches cannot be guaranteed to achieve global optimums. In LINMAP, attribute weights can be constrained to be nonnegative, part-worth functions can be constrained to be monotone or of the ideal-point type, while such constraints cannot be imposed for the other approaches. Such constraints, as imposed on the basis of prior knowledge, can be useful in improving the accuracy of the parameter estimates when the (n/T) ratio is small—assuming, of course, that the prior knowledge is correct. Finally, the usage of sum-of-absolute-errors as the poorness-of-fit measure, rather than the sum-of-squared-errors, tends to produce more robust estimates, i.e., the estimated parameters are not as much affected by outliers or large errors in the input data (Blattberg and Sargent 1971). On the other hand, when the n/T ratio is small, the use of linear programming sometimes produces alternate optimums (i.e., different sets of estimated parameters have the same minimal poorness-of-fit), which is intuitively unappealing.

Among the metric methods, the OLS procedure has the important advantage of providing standard errors for the estimated parameters. None of the procedures in (1) has this advantage. However, the overall statistical significance of MONANOVA can be roughly ascertained using the computer simulation results obtained under the assumption of a random rank order for the dependent variable (Acito 1978). The MSAE procedure, on the other hand, is more robust than the OLS method (Blattberg and Sargent 1971) and, furthermore, permits us to impose a priori constraints on the estimated parameters.

The probabilistic approaches, (3), explicitly model the errors that may be present in the preference functions, thereby leading to a deductive development of the choice model. The LOGIT approach has the advantage that the estimation procedure produces global maximum likelihood estimates (McFadden 1976). However, the use of the LOGIT model involves the "independence of irrelevant alternatives" assumption, which may not be a realistic assumption in many consumer behavior contexts. Briefly, if two substantially different services, say automobiles (A) and public transit (B), have choice probabilities of 0.75 and 0.25, respectively, and another mode of public transit (C), very similar to B, is introduced, then the LOGIT model leads to choice probabilities of 0.6, 0.2, and 0.2, respectively, for A, B, and C. Thus the *total* probability for public transit has risen which is intuitively unappealing. Since the two transit modes, B and C, are very similar, the introduction of C should draw its choice probability largely from B rather than A.

The PROBIT procedure is particularly suited to cases where a dichotomous intention-to-buy scale is used as the dependent variable (Rao and Winter 1977). The estimation procedure, however, cannot guarantee global maximum likelihood estimates.

The probabilistic choice models such as LOGIT and PROBIT assume that the paired comparisons are probabilistically independent. If the dependent variable data are obtained directly as paired comparisons, this may be a realistic assumption. (However, we have argued earlier that this is an inefficient way of collecting data.) If the data are obtained as a rank order and then converted to the equivalent $n(n - 1)/2$ paired comparisons, this assumption is not realistic.[3] The asymptotic estimates of standard errors are not likely to be valid either. (In any event, the limited number of stimuli usually used in consumer research studies casts doubt on the appropriateness of such asymptotic statistics.) Despite these limitations, however, the probabilistic choice models appear to have substantial predictive validity.

The choice between the nonmetric, (1) and (3), and metric methods, (2), should logically depend on the

[3] However, Punj and Staelin (1978) utilize a procedure based only on the $n - 1$ *independent* choices that mitigates this problem.

scale properties of the dependent variable. The simulation studies by Cattin and Wittink (1976) and Carmone, Green, and Jain (1978) have found that the OLS regression applied to integer ranks (the rank ordered dependent variable is redefined as a pseudo-interval-scaled variable, taking values 1, 2, . . . , n depending on the rank given that stimulus) produces solutions that are very close, in terms of predictive validity, to those obtained by the more expensive nonmetric algorithms. However, by employing integer ranks as the dependent variable, the usual standard errors and statistical tests are not strictly valid.

Cattin and Wittink (1976) report that the results from OLS and MONANOVA are virtually indistinguishable: the difference in predictive validity is, at most, 0.003 in terms of Pearson's rho. In other words, because of local optimum problems with MONANOVA, it may produce solutions which are only marginally different from OLS. The difference between OLS on the one hand and LOGIT and LINMAP on the other is about 0.03 units in terms of Pearson's rho, with OLS being the better approach when the attribute weights are normally distributed (compensatory model) and LINMAP and LOGIT being the preferred methods when the attribute weights exhibit a lexicographic structure. In general, this simulation study found that the methods differed by only *very small amounts* in their predictive validities.

A few empirical (as opposed to simulation) studies have compared alternative estimation procedures. Hauser and Urban (1977, p. 600) report that the monotonic methods have a slightly better fit to saved data than OLS. (The dependent variable was a rank order.) Hauser and Shugan (1977) state that the MSAE regression performs better than OLS because of its ability to incorporate constraints on the estimated parameters. (The dependent variable in this case was intervally scaled.) Montgomery, Wittink, and Glaze (1977) report that the predictive validity of OLS was better than that of MONANOVA. The dependent variable was a rank order but the small $n/T = 32/19$ ratio is likely to have biased this result in favor of OLS. Fidler and Thompson (1977) report that LINMAP and a PROBIT-like procedure yielded roughly the same validity and substantive conclusions. McCullough (1978) reports that Johnson's nonmetric procedure and MONANOVA predict at roughly the same level of validity and better than a Fishbein-Rosenberg type multiattribute model. Rao and Solgaard (1977) report that MONANOVA, Johnson's nonmetric procedure, LINMAP, and CCM (Carroll 1969) yield roughly the same level of cross validity. Jain et al. (1978) report that MONANOVA, Johnson's nonmetric procedure, LINMAP, LOGIT, and OLS yield roughly the same level of cross-validity, with LOGIT and LINMAP being the slightly preferred procedures.

Overall, the estimation methods do not seem to differ very much in their predictive validities. The metric procedures (2) seem to perform slightly better if the preference model is approximately compensatory. If the model is approximately lexicographic, the nonmetric methods seem to do slightly better. Based on these results, we would recommend that researchers estimate parameters using both a metric and a nonmetric method as a rough check on the robustness of their results, at least until more empirical evidence on their correspondence has been assembled.

RELIABILITY AND VALIDITY TESTS

While some researchers have been careful to test the reliability and/or validity of conjoint analysis, a large number of applications of conjoint analysis have ignored these issues. We will now discuss methods for assessing reliability and validity with the hope that future applications will incorporate at least some of these tests.[4]

Tests of Reliability

Tests of reliability can be carried out either at the level of input judgments of the respondent or at the level of estimated parameters. To obtain the reliability of a respondent's input judgments, the researcher can ask for preference judgments on a second set of stimulus cards which contain a subset of the original set of stimulus cards. This needs to be done only for a subsample of the respondents after the respondent has completed some intervening task, such as supplying a set of demographic data. The repeated evaluations can be used in determining the *test-retest reliability of the input preference judgments.*

If we view conjoint analysis as an instrument to measure the parameters of the preference model, then the *alternate forms method with spaced testing* may be more appropriate (Parker and Srinivasan 1976). After completing a conjoint analysis task, a subset of the respondents can be approached after a period of time and asked to perform the rank ordering or rating task on a second set of stimuli. The second stimulus set ("alternate form") would also have n descriptions from the same factor levels as the first set but would avoid duplication of stimuli from the initial set. This could be done by using a second fractional-factorial design from the same factor levels or by randomly drawing a second set of n descriptions from the multivariate distribution. Product moment correlations of the estimated parameters from the two tasks provides a measure of reliability, sometimes referred to as the *coefficient of equivalence.*

The second method of reliability testing is more rigorous than the first one, in that it takes into account

[4] Our discussion emphasizes predictive validity rather than validation of the actual decision process which, in large measure, is not known.

four sources of error: inaccuracies in the input data, variability in the set of constructed stimuli, errors in the estimation procedure, and lack of stability (variations from one time period to another). By contrast, the first method focuses only on the first source of error.

Tests of Validity

The *internal validity* of conjoint analysis can be reported in terms of the correlation (Pearson's or Spearman's rho depending on the scale properties of the dependent variable) between the input versus estimated values of the dependent variable. If Pearson's rho is used, it is desirable to adjust the correlation for the number of estimated parameters, as in the coefficient of multiple determination (R^2) adjusted for degrees of freedom.

Data used for the reliability tests also provide a method of cross-validation. The parameters of the preference function estimated from the first set of preference data can be used to predict the preferences for the second set. The predicted preferences can be correlated to the actual to obtain a measure of cross-validity. The procedure can then be reversed by predicting from set 2 to set 1, thus completing a double cross-validation.

While internal validity tests the goodness of the model, cross-validation also takes into account the predictive ability of the model. But neither method tests the *external validity* of the model. Since conjoint analysis studies are usually carried out using hypothetical stimulus descriptions, we can test for external validity by comparing predictions against a respondent's actual behavior with respect to real stimuli. This method, referred to as *predictive validity* by Parker and Srinivasan (1976, p. 1,017), involved predicting the respondent's rank order of the n real stimuli (choice set) using the estimated preference function. The respondent's actual choices will fall somewhere in this rank order (1 to n). The closer the number is to 1, the better is the external validity for that respondent. By repeating the predictions for each of the respondents in the sample, we obtain a frequency distribution of the number of respondents choosing the first, second, . . . , nth most preferred stimuli. The median of the distribution can be compared to 1 (best) and $(n + 1)/2$ (random model) to get a measure of the external validity. The obtained frequency distribution (which should be skewed positively) can be tested for statistical significance against a uniform distribution (random model) by the Kolmogorov-Smirnov statistic.

One of the important uses of conjoint analysis is to predict the behavior of the respondents toward *new* stimuli. The most rigorous tests for conjoint analysis predict the reaction of a respondent toward a new stimulus and compare it with actual behavior. However, since conjoint analysis is, at best, a static analysis, it can only be expected to predict responses in "steady state," assuming that consumers become knowledgeable about the new stimulus.

Additional Considerations in Reliability/ Validity Testing

If paired comparison preference judgments are obtained directly, one can also test for intransitivities. Hauser and Shugan (1977) have developed measures of intransitivity for interval and ratio-scaled preference data obtained through constant-sum paired comparisons. In addition, consistency can be tested at the aggregate level in terms of market shares. For example, in a conjoint analysis study designed by one of the authors for a large transportation company, data were available on market share, by supplier, for various pairs of cities served by the company. Shares of choices, estimated at the individual level and then aggregated, agreed closely enough with the actual market share data to provide reasonable assurance of the model's validity in the aggregate. Similar tests have been carried out by Fiedler (1972), Davidson (1973), and Punj and Staelin (1978).

Scott and Wright (1976) suggest some additional consistency checks to test whether the estimated parameters make sense. First, the signs of the estimated parameters should agree with a priori expectations, based on prior theory or reasoning. Second, the parameters derived for different subpopulations should differ in the direction that would be expected from prior theory or reasoning. In addition, the "face validity" of the results can be checked by comparing a respondent's subjective estimates (self-reports) with estimated parameters. A somewhat unobtrusive method of testing for face validity is to aggregate respondents' spontaneous verbalizations during the interview regarding the factors or attributes important to them and comparing them against aggregated results from estimated model parameters (Parker and Srinivasan 1976, p. 1,013).

To get some feel for the cost/benefit of conjoint analysis, its predictions can be compared against "naive" models which do not involve any data collection. For instance, if the vector model of preference is a reasonable approximation and the attributes are reoriented, if necessary, so that higher attribute levels are more preferred, and the attributes are standardized to unit standard deviation, then a *unit weighting model* (i.e., each weight w_p in equation (2) is set equal to unity) is often a good contender for a naive model (Dawes and Corrigan 1974). Alternately, the most important attribute from a priori considerations (e.g., price) can be used to define a naive preference model in which the weight for the most important attribute is unity, while all other attributes are assigned zero weights. An extension of this would be a lexicographic naive model with the ordering of attributes obtained from prior judgments.[5] For an illustration of these tests against naive models, see Parker and Srinivasan (1976, pp. 1,015–17).

Published empirical results on the predictive validity of conjoint analysis are encouraging (e.g., Fiedler 1972, condominiums; Davidson 1973, air transportation; Parker and Srinivasan 1976, primary health care facilities; Montgomery, Wittink, and Glaze 1977, job choice by MBAs; Scott 1977, solar heating for homes). Furthermore, the reliability and cross validity reported by Parker and Srinivasan (1976) are high. The only reported study we know of which showed extremely poor predictions is that by Bither and Wright (1977) in the context of selection of a set of movies dealing with golfing demonstrations.

APPLICATIONS OF CONJOINT ANALYSIS

The typical output from conjoint analysis is a set of estimated parameters of the preference model for each individual in the sample. A direct use of these results is to describe the degree to which each of the attributes is considered important by the respondent sample. If the vector model of equation (2) or ideal-point model of equation (3) is used, the weights $\{w_{ip}\}$ for individual i are first standardized by multiplying w_{ip} by the standard deviation s_p for attribute p—similar to the determination of standardized regression weights (beta weights) in multiple regression—and then normalized so that the attribute importances sum to unity. If the part-worth function model of equation (4) is used, the range of the values for the part-worth function over the levels of attribute p serves as a measure of importance for attribute p. Again, the ranges are normalized so that the t attribute importances sum to unity. The resulting distributions (over the respondent sample) of importances may be summarized in terms of the means and standard deviations for each of the t attributes. If the ideal-point model of equation (3) is used, the distribution of ideal points can be plotted for each of the attributes.

In addition to this "descriptive" use of conjoint analysis, its "normative" uses are described next for both public and private sector applications.

Public Sector Applications

Conjoint analysis offers a tremendous potential for conducting cost/benefit analysis for many public policy decisions. We discuss this here in the contexts of planning rural primary health care delivery and evaluating the implications of energy policies for work-trip gasoline conservation. For additional applications in the public sector area, see McClain and Rao (1974), Whitmore and Cavadias (1974), Wind and Spitz (1976), and Hopkins, Larréché, and Massy (1977).

Parker and Srinivasan (1976) address the problem of determining the number, locations, and physical and operational characteristics of a set of health care facilities to be added to an existing health care delivery system so as to maximize the incremental benefit to the community, subject to a cost (budget) constraint. The conjoint analysis study used the mixed model of equation (7) with travel time (hours) to the facility as one of the attributes. Each respondent's estimated preference function was then transformed into a benefit function expressing the individual's benefit in dollars/year for an existing or potential health care facility. Briefly, this involved multiplying each of the estimated weights $\{v_q\}$ in equation (7) by a common positive number, chosen so that the transformed weight for the travel time attribute coincides with the annual dollar cost for the respondent's household to travel to a health facility that is one hour away.

For example, suppose there are currently only three facilities in the area with benefits for individual i of \$50/year, \$200/year, and \$150/year. Then assuming that individual i chooses the most preferred facility, his/her benefit from the existing system is \$200/year, corresponding to his/her choosing Facility 2. If we now add a fourth facility whose benefit for this individual is \$275/year, then his/her incremental benefit is \$75/year (i.e., he/she would, in "steady state," choose Facility 4 so that the new benefit is \$275/year). If, however, Facility 4's benefit was only \$175/year then the incremental benefit for this individual is zero—he/she would stay with Facility 2 so that his/her benefit has not changed. The total incremental benefit for adding Facility 4 to the community can be obtained by aggregating the individual incremental benefits, after differentially weighting different socioeconomic segments of the population, and then multiplying by the population/sample size ratio. Some optimization procedure can then be used to determine a set of additional facilities that would be feasible (i.e., total cost stays within the budget) and would maximize the total incremental benefit to the community.

In an ongoing study by one of the authors, potential energy policies (such as gasoline surcharges, transit subsidies) are being evaluated in terms of their impact on gasoline conservation. By developing a conjoint analysis model with mode of travel (car versus carpool versus public transit), time of travel, gasoline price, etc. as attributes, it is possible to predict whether changes in attributes, such as gasoline cost increased by a gas tax or waiting time decreased by transit improvements, would change the modal choice of the respondent, and if so, what the resulting gasoline conservation would be. The social costs of such policies can also be evaluated by a procedure similar to that in the health care study.

In short, conjoint analysis offers an excellent oppor-

[5] For an illustration of these tests against naive models, see Parker and Srinivasan (1976, pp. 1015–17).

tunity for consumer researchers to provide valuable cost/benefit analysis inputs for public policy decisions.

Private Sector Applications

While no precise figures are available, it is estimated that several hundred conjoint analysis studies have been conducted by corporate marketing research groups and consulting firms. The applications have spanned a wide variety of products/services: consumer nondurables and durables, industrial goods, financial and other services, transportation, etc.

In most private sector applications of conjoint analysis, some type of consumer choice simulation is carried out to see what share of choices would be generated by each of several product/service profiles if they were competing with each other in the market place. (The word "simulation" is used here to mean prediction of individual choices under hypothetical scenarios and subsequent aggregation of choices—it does not necessarily mean Monte-Carlo probabilistic simulation.)

The typical consumer choice simulator is quite simple in construction. Each individual's utility function and background description (current purchase behavior, socioeconomic, demographic, and psychographic characteristics) are entered into the simulator. Each of, say, six competitive products or services is entered as a full-profile description. Then, for each individual, in turn, a utility is computed for each of the competing items. The individual is assumed to choose that item displaying the highest utility to him. Elaborations on this theme incorporate various probability-of-choice rules, based on the utilities of all contending items. (See Shocker and Srinivasan 1977b for a detailed discussion.) The first-choice frequencies for each item are then simply added and expressed relatively.

Many variations on this simple procedure have been used in proprietary studies. For example, one may wish to employ the respondents' background data to obtain consumer segments; shares of choices are then cross-classified by these prior-defined segments. If profile data on current market brands are also available, it is a simple matter to examine brand switching behavior as new items are entered into the existing array, either as replacements for current items or as net additions to the product/service set.

More sophisticated choice simulators can also be constructed with additional features, such as:

- a provision for describing new item profiles in terms of probability distributions over factor levels, rather than as a deterministic level for each factor;
- a provision for simulating trial and repeat purchase, conditioned on assumptions about the relative amounts of sales effort and consumer satisfaction-in-use associated with each item.

Not surprisingly, one of the more important uses of conjoint analysis in the private sector has been in the evaluation of new product or service concepts. If augmented by cost data, this approach can also be extended to determine "optimal" new product/service ideas by searching the product space, taking into account existing brands/services in the market place and the possible effect of the cannibalization of one's own brands by the new entry. Shocker and Srinivasan (1977b) have reviewed such approaches and their applications to product concept evaluation and generation.

Another important use of conjoint analysis is in the context of guiding strategies for market communication. For instance, an analysis of the attribute importances and the relative position of the firm's brand vis-a-vis the firm's competitors can help in the development of suitable advertising strategies (Boyd, Ray, and Strong 1972).

Conjoint analysis is also useful in market segmentation. The idea is to divide a heterogenous population of consumers into more homogenous segments so that different marketing strategies can be tailored to different segments of consumers to achieve maximum marketing results. Cluster analysis can be employed to group respondents with similar "importances," for t attributes, into clusters. The clusters, in turn, can be cross-tabulated against various background variables, such as demographics, psychographics, product and media usage. Alternatively, the "importances" can be employed as a set of predictor variables in a multiple discriminant analysis in which subjects have been previously classified into groups on the basis of some other criterion, such as brand choice or product-class consumption levels.

For such market segmentation studies, one could also use the predicted, or input, preferences for the n stimuli as the segmentation variables instead of "importances." However, we have found that in many applications it is the importances rather than the preferences which are more discriminating. The reason for this is that *within* a specific attribute, e.g., tread life or price, there is often high agreement on at least the ordering of levels, e.g., more tread life is preferred to less. However, the importance of the factor, tread life, i.e., the relative part-worth range, is often more sensitive to individual variation.

NEW DEVELOPMENTS

In the relatively short time that conjoint analysis has been in use, a number of new developments have already appeared. Some of these are briefly described here.

Preference Models for Collections of Items

Most of the conjoint methodologies and their applications have been in the context of choosing a single item from a product (service) class. However, Green, Wind, and Jain (1972) have examined whether the total preference for, say, an entree-dessert combination can be decomposed, via an additive model, into the preference for an entree plus the preference for a dessert. Green and DeVita (1974, 1975a) discuss a method of taking into account two-factor interaction effects in the context of menu choice through MDPREF (Carroll 1972). Farquhar and Rao (1976)—see also Rao (1973)—consider the choice of subsets such as TV programs, liquor assortments, etc., by their "balance model." They model the utility for a set of n stimuli as composed of (i) preference components relating to *total* attribute level (over the n stimuli) on each of the t attributes and (ii) preference components relating to the standard deviation of attribute levels (within the n stimuli) for each of the t attributes. The latter terms correspond to the preference for "balance" or "counterbalance," depending on whether variability is desirable or undesirable.

McAlister (1978) considers the same problem of choice of subsets but models the overall preference for a set of magazines (say) using a part-worth function approach, where the argument in the part-worth function for attribute p is its *total* attribute level over the subset of magazines. By assuming that the part-worth function is approximated by linear and quadratic terms (to model attribute satiation), she is able to use LINMAP to estimate the parameters. McAlister (1978) also examines a second context of choice of subsets where the consumption is limited to only one of the elements in the subset (e.g., a student may apply to several universities but enrolls in only one of the schools which offer admission). She models this context by embedding preferences for individual items in a decision tree.

Incorporating Interaction Effects in Preference Models

Green, Carroll, and Carmone (1978) have developed an algorithm that incorporates selected interaction terms after main effect terms have been estimated. This algorithm appears to be most useful in cases involving a large number of factors, thus precluding the use of full factorial designs.

The authors use various kinds of fractional factorials that permit the researcher to estimate all main effects at the individual level and selected two-factor interactions at the group (or subgroup) level, without necessitating an unwieldy number of stimuli for respondent evaluation. The assumption is that the researcher will be primarily interested in interaction effects at the group (rather than individual) level. Under this assumption, the algorithm is able to take advantage of the additional degrees of freedom gained by pooling over groups of respondents.

Multi-Stage Consumer Decision Processes

Srinivasan (1978) observes that protocol analyses of consumer decision processes indicate that for many respondents the processing strategy seems to change during the task, starting from an initial strategy aimed at narrowing the choice to a small subset (through processes approximating conjunctive type rules) and ending in a detailed examination to choose the most preferred item through processes approximating compensatory rules. He models a multi-stage decision process as a sequence of linear models of the form of equation (7), where the weights change from one stage to the next (several of the weights in each stage may be zero). In an m-stage model, the ith stage preference function ($i = 1,2, \ldots, m$) discriminates only among stimuli which are tied by each of the previous stages $1,2, \ldots, i - 1$. Assuming that each attribute has only a small number of levels, this model includes the compensatory ($m = 1$), lexicographic (each stage has only one weight positive with all other weights equal to zero), conjunctive and disjunctive rules, and any combination thereof as special cases. Srinivasan provides a procedure to estimate the stage weights from rank-order preference judgments using LINMAP (Shocker and Srinivasan 1977a). The multi-stage model seems to offer the potential of taking us one step closer to actual consumer decision processes.

Componential Segmentation

As remarked earlier in the context of private sector applications, one of the uses of conjoint analysis has been in the development of market segments. A recent development, known as componential segmentation (Green 1977; Green, Carroll, and Carmone 1976, 1977; Green and DeSarbo 1977), takes a further step in this direction. Traditionally, market segments have been defined as groups of consumers whose responses to some market stimulus exhibit relatively little within-group variation but considerable among-groups variation. In contrast, componential segmentation places emphasis on the *interaction of a stimulus profile with a person profile*; that is, the concern here is less with market partitioning and more with predicting how a respondent, characterized by a particular set of attribute levels, will respond to a set of stimuli, each of which represents a particular profile of factor levels. It is the joint effect of the two sets of attribute and factor levels that results in response.

As an illustration, consider the earlier example in which consumers are choosing among alternative steel-belted replacement tires, varying by brand, tread life, price, and sidewall color. Illustratively, the consumers are assumed to vary by sex, type of car ownership, and age of car owned. Traditional segmentation procedures might first group the consumers by type of tire owned and then see (e.g., by means of discriminant analysis) whether the a priori defined segments differ in terms of other consumer background variables. In contrast, componential segmentation decomposes total variability in preferences for alternative product descriptions (each involving a profile of brand, tread life, price, and sidewall color) by alternative respondents (each involving a profile of sex, type of car, and age of car) into three components: (i) variability due to product attributes, (ii) variability due to person attributes and (iii) variability due to interaction between person and product attributes. The first component, in a sense, explains the variation in preference for different levels of product attributes for the average respondent and should be of interest in product planning. The second component is usually not of interest, since it is simply due to different respondent profiles having a different mean scale value for preferences. (Thus, if the data were standardized so that each respondent gave the same average preference rating for the *n* profiles, this component would be zero.) The third component is due to interaction effects between stimulus and respondent profiles and, therefore, provides a direct measure of segmentability of the market. It may show, for instance, that compact car owners attach more importance to tire price than owners of medium and big cars. This could suggest that a firm has to be considerably more price competitive in the compact car segment, but could use a product differentiated, high-price strategy for the medium and big car segment.

Extensions of componential segmentation to three-way (respondents by scenarios by products) and higher-way matrices are also possible, thus providing a means for operationalizing the concept of situation dependence (Belk 1975) as it may interact with the stimulus attributes under evaluation by different types of respondents.

Preferences for Alternative Allocations of Scarce Resources

Recently conjoint analysis has been applied to study consumers' preferences for alternative allocations of some scarce resource, such as money or time. One pilot study (Carroll, Green, and DeSarbo 1978) considered preferences for alternative allocations of leisure time for different levels of (i) watching TV, (ii) recreational reading, (iii) sports activity, (iv) hobbies, and (v) socializing. The same type of approach can also be used to find consumer utilities for such things as alternative household budget allocations, or allocations of household savings across such investments as insurance, common stocks, municipal bonds, etc. If time (rather than money) is the scarce resource, possible applications can involve such things as preferences for TV news shows regarding the amount of time spent on national news, local news, weather, sports, and the like, or allocations of magazine space to various kinds of editorial matter and advertisements.

Conjoint Analysis in the Modeling of Perceptions Data

Until recently conjoint analysis has been restricted to the analysis of preference (and other kinds of dominance) data. If the analyst wished to explore consumers' *perceptions* of products or services, he/she usually fell back on the apparatus of multidimensional scaling, with its associated problems of dimension interpretation and the like.

More recently, conjoint analysis has been extended to the modeling of similarities data (Green and DeSarbo 1978). To illustrate the approach, assume that a researcher is interested in consumers' perceptions of various vacation sites, e.g., London, Bermuda, Las Vegas, Honolulu, etc. A set of six factors (say) are selected that are capable of describing vacation sites in general, such as (i) food quality, (ii) sightseeing opportunities, (iii) outdoor sports, (iv) night life/entertainment, (v) chance to meet new friends, and (vi) trip cost. Next assume that each factor is described in terms of three levels (e.g., superb, good, or fair food quality). An orthogonal main effects plan of 18 stimulus cards can be constructed from the 3^6 full factorial. The respondent is then shown a card on which each reference site, such as London, appears and is asked to sort the profiles along, say, a 9-point scale according to how similar each of the profiles is to each selected location (e.g., London). The similarities data are analyzed for each site separately via some type of nonmetric or metric conjoint algorithm. However, in this situation the analysis produces *part-similarity* functions (analogous to part-worth functions) and factor saliences (relative ranges of the part-similarity functions).

The results from such an analysis can be pooled over those who chose London (based on an earlier question) as the most preferred place to visit versus those who chose some other vacation site. Green and DeSarbo (1978) found, for instance, that London choosers attached higher salience to sightseeing, while others attached higher salience to outdoor sports and total trip cost.

Although not described here, the part-similarity functions can, in turn, be transformed into respondents' subjective probability distributions relating to the uncertainty of perceptions of London along each of the factors. This information together with the respondent's preference function (as developed from a conjoint analysis of the respondent's preference data) could be valuable in developing communication campaigns to correct misperceptions and attract more visitors to London. The implications of this extended conjoint methodology should be of considerable interest in positioning products and services.

CONCLUDING REMARKS

The wide support by academic and industry researchers over a relatively short time (since 1971) is an indication of the potential of conjoint analysis in providing a useful methodology for representing the structure of consumer preferences and some ability for predicting consumers' behavior towards new stimuli. Much empirical work needs to be done to identify the combination of alternatives in each of the various steps of conjoint analysis to achieve maximum predictive validity for a given problem definition and research budget. The answers could, of course, depend on situational factors such as type of product-market, number of relevant attributes, etc. Researchers are urged to pay considerably greater attention to testing reliability and validity in their applications. Although conjoint analysis has been extensively applied mainly in the private sector, it has a large potential for public sector applications as well. As the last section of the paper indicated, conjoint analysis is far from being a settled, cut-and-dried methodology. Many opportunities still exist for extending present techniques and applying them to a greater variety of substantive problems in consumer behavior.

[*Received August 1977. Revised March 1978.*]

REFERENCES

Acito, Franklin (1977), "An Investigation of Some Data Collection Issues in Conjoint Measurement," in *1977 Educators' Proceedings,* Chicago: American Marketing Association, 82–5.

——— (1978), "A Monte Carlo Investigation of Conjoint Measurement Under Random Data Conditions for Various Orthogonal Designs," Working Paper, School of Business, Indiana University.

Alpert, Mark I. (1971), "Definition of Determinant Attributes: A Comparison of Methods," *Journal of Marketing Research,* 8, 184–91.

———, Betak, John F., and Golden, Linda, L. (1978), "Data Gathering Issues in Conjoint Measurement," Working Paper, Graduate School of Business, The University of Texas at Austin.

Anderson, Norman H. (1970), "Functional Measurement and Psychophysical Judgment," *Psychological Review,* 77, 153–70.

Barron, F. Hutton (1977), "Axiomatic Conjoint Measurement," *Decision Sciences,* 8, 48–59.

Belk, Russell W. (1975), "Situational Variables and Consumer Behavior," *The Journal of Consumer Research,* 2, 157–64.

Ben-Akiva, Moshe (1973), "Program for Maximum Likelihood Estimation of the Multinomial Logit Model," Cambridge, MA: Massachusetts Institute of Technology, Department of Civil Engineering.

Berl, Janet, Lewis, Gordon, and Morrison, Rebecca Sue (1976), "Applying Models of Choice to the Problem of College Selection," in *Cognition and Social Behavior,* eds. J. S. Carroll and J. W. Payne, Hillsdale, NJ: Lawrence Elbaum Associates, 203–19.

Bither, Stewart W., and Wright, Peter (1977), "Preferences Between Product Consultants: Choices vs. Preference," *The Journal of Consumer Research,* 4, 39–47.

Blattberg, Robert, and Sargent, Thomas (1971), "Regression with Non-Gaussian Stable Disturbances: Some Sampling Results," *Econometrica,* 39, 501–10.

Boyd, Harper W., Jr., Ray, Michael L., and Strong, Edward C. (1972), "An Attitudinal Framework for Advertising Strategy," *Journal of Marketing,* 36, 27–33.

Braun, Michael A., and Srinivasan, V. (1975), "Amount of Information as a Determinant of Consumer Behavior Towards New Products," *1975 Combined Proceedings,* Chicago: American Marketing Association, 373–8.

Carmone, Frank J., Green, Paul E., and Jain, Arun K. (1978), "The Robustness of Conjoint Analysis: Some Monte Carlo Results," *Journal of Marketing Research,* 15, 300–3.

Carroll, J. Douglas (1969), "Categorical Conjoint Measurement," Meeting of Mathematical Psychology, Ann Arbor, MI.

——— (1972), "Individual Differences and Multidimensional Scaling," in *Multidimensional Scaling: Theory and Applications in Behavioral Sciences, Vol. I,* eds. R. N. Shepard, et al., New York: Seminar Press, 105–55.

———, Green, Paul E., and DeSarbo, Wayne S. (1978), "Optimizing the Allocation of a Fixed Resource: A Simple Model and Its Experimental Test," Working Paper, Bell Laboratories, Murray Hill, NJ.

Cattin, Philippe, and Wittink, Dick R. (1976), "A Monte-Carlo Study of Metric and Nonmetric Estimation Methods for Multiattribute Models," Research Paper No. 341, Graduate School of Business, Stanford University.

Colberg, Roger T. (1978), "A Monte Carlo Evaluation of Metric Recovery of Conjoint Measurement Algorithms," Research Paper, College of Business Administration, University of Nevada-Reno.

Darlington, Richard B. (1968), "Multiple Regression in Psychological Research and Practice," *Psychological Bulletin,* 69, 161–82.

Davidson, J. D. (1973), "Forecasting Traffic on STOL," *Operational Research Quarterly,* 24, 561–9.

Dawes, Robyn M. (1971), "A Case Study of Graduate Admissions: Application of Three Principles of Human Decision Making," *American Psychologist,* 26, 180–8.

———, and Corrigan, Bernard (1974), "Linear Models in Decision Making," *Psychological Bulletin,* 81, 95–106.

Farquhar, Peter H., and Rao, Vithala, R. (1976), "A Balance Model for Evaluating Subsets of Multiattributed Items," *Management Science,* 22, 528–39.

Fidler, Eduard, J., and Thompson, Gerald L. (1977), "An Experiment on Executive Decision Making," Management Science Research Report No. 407, Graduate School of Industrial Administration, Carnegie-Mellon University.

Fiedler, John A. (1972), "Condominium Design and Pricing: A Case Study in Consumer Trade-off Analysis," in *Proceedings (Third Annual Conference),* ed. M. Venkatesan, Chicago: Association for Consumer Research, 279–93.

Fishbein, Martin (1967), "A Behavior Theory Approach to the Relations between Beliefs about an Object and the Attitude towards the Object," in *Readings in Attitude and Theory Measurement,* ed. M. Fishbein, New York: John Wiley & Sons, Inc., 389–99.

Gensch, Dennis H., Golob, Thomas F., and Recker, Wilfred W. (1976), "Regression is Inappropriate for Analyzing Cross-Sectional Data," in *1976 Educators' Proceedings,* Chicago: American Marketing Association, 120–4.

Goldberg, Lewis R. (1968), "Simple Models or Simple Processes? Some Research on Clinical Judgments," *American Psychologist,* 23, 483–96.

Goldberger, Arthur S. (1964), *Econometric Theory,* New York: John Wiley & Sons, Inc.

Green, Paul E. (1974), "On the Design of Choice Experiments Involving Multifactor Alternatives," *The Journal of Consumer Research,* 1, 61–8.

——— (1975), "Marketing Applications of MDS: Assessment and Outlook," *Journal of Marketing,* 39, 24–31.

——— (1977), "A New Approach to Market Segmentation," *Business Horizons,* 20, 61–73.

———, and Carmone, Frank J. (1970), *Multidimensional Scaling and Related Techniques in Marketing Analysis,* Boston: Allyn and Bacon.

———, and Carmone, Frank J. (1977), "A BIB/Logit Approach to Conjoint Analysis," Working Paper, Wharton School, University of Pennsylvania.

———, Carmone, Frank J., and Wind, Yoram (1972), "Subjective Evaluation Models and Conjoint Measurement," *Behavioral Science,* 17, 288–99.

———, Carroll, J. Douglas, and Carmone, Frank J. (1976), "Superordinate Factorial Designs in the Analysis of Consumer Judgments," *Journal of Business Research,* 4, 281–95.

———, Carroll, J. Douglas, and Carmone, Frank J. (1977), "Design Considerations in Attitude Measurement," in *Moving A Head with Attitude Research,* eds. Y. Wind and M. G. Greenberg, Chicago: American Marketing Association, 9–18.

———, Carroll, J. Douglas, and Carmone, Frank J. (1978), "Some New Types of Fractional Factorial Designs for Marketing Experiments," in *Research in Marketing, Vol. I,* ed. J. N. Sheth, Greenwich, CT: JAI Press.

———, and DeSarbo, Wayne S. (1977), "Demographic Stereotypes and Brand Preferences: An Application of Componential Segmentation," Working Paper, Wharton School, University of Pennsylvania.

———, and DeSarbo, Wayne S. (1978), "Additive Decomposition of Perceptions Data Via Conjoint Analysis," *The Journal of Consumer Research,* 5, 58–65.

———, and Devita, Michael T. (1974), "A Complementarity Model of Consumer Utility for Item Collections," *The Journal of Consumer Research,* 1, 56–67.

———, and Devita, Michael T. (1975a), "An Interaction Model of Consumer Utility," *The Journal of Consumer Research,* 2, 146–53.

———, and Devita, Michael T. (1975b), "The Robustness of Linear Models Under Correlated Attribute Conditions," *1975 Educators' Conference,* Chicago: American Marketing Association, 108–11.

———, and Rao, Vithala R. (1969), "Nonmetric Approaches to Multivariate Analysis in Marketing," Working Paper, Wharton School, University of Pennsylvania.

———, and Rao, Vithala R. (1971), "Conjoint Measurement for Quantifying Judgmental Data," *Journal of Marketing Research,* 8, 355–63.

———, and Tull, Donald S. (1978), *Research for Marketing Decisions (4th Edition),* Englewood Cliffs, NJ: Prentice-Hall, Inc.

———, and Wind, Yoram (1973), *Multiattribute Decisions in Marketing: A Measurement Approach,* Hinsdale, IL: The Dryden Press.

———, and Wind, Yoram (1975), "New Way to Measure Consumers' Judgments," *Harvard Business Review,* 53, 107–17.

———, Wind, Yoram, and Jain, Arun K. (1972), "Preference Measurement of Item Collections," *Journal of Marketing Research,* 9, 371–7.

Hansen, Fleming (1976), "Psychological Theories of Consumer Choice," *The Journal of Consumer Research,* 3, 117–42.

Hauser, John R., and Shugan, Steven M. (1977), "Efficient Measurement of Consumer Preference Functions: A General Theory for Intensity of Preference," Working Paper 602-001, Graduate School of Management, Northwestern University.

———, and Urban, Glen L. (1977), "A Normative Methodology for Modeling Consumer Response to Innovation," *Operations Research,* 25, 579–619.

Hoffman, Paul J., Slovic, Paul, and Rorer, L. G. (1968), "An Analysis of Variance Model for the Assessment of Configural Cue Utilization in Clinical Judgment," *Psychological Bulletin,* 69, 338–49.

Hopkins, David S. P., Larréché, Jean-Claude, and Massy, William F. (1977), "Constrained Optimization of a University Administrator's Preference Function," *Management Science,* 24, 365–77.

Jain, Arun K., Acito, Franklin, Malhotra, Naresh, and Mahajan, Vijay (1978), "A Comparison of Predictive Validity of Alternative Methods for Estimating Parameters in Preference Models," Working Paper, School of Management, State University of New York, at Buffalo.

Johnson, Richard M. (1973), "Varieties of Conjoint Measurement," Working Paper, Market Facts, Inc., Chicago.

——— (1974), "Trade-off Analysis of Consumer Values," *Journal of Marketing Research,* 11, 121–7.

——— (1976), "Beyond Conjoint Measurement: A Method of Pairwise Tradeoff Analysis" in *Advances in Consumer Research, Vol. III,* ed. B. B. Anderson, Proceedings of Association for Consumer Research, Sixth Annual Conference, 353–8.

———, and VanDyk, Gerald J. (1975), "A Resistance Analogy for Efficiency of Paired Comparison Designs," unpublished paper, Market Facts, Inc., Chicago.

Johnston, J. (1972), *Econometric Methods* (2nd Edition), New York: McGraw-Hill Book Co.

Keeney, Ralph L., and Raiffa, Howard (1976), *Decisions with Multiple Objectives: Preferences and Value Tradeoffs,* New York: John Wiley & Sons, Inc.

Kelly, George A. (1955), *The Psychology of Personal Constructs,* New York: W. W. Norton & Co., Inc.

Krantz, David H. (1964), "Conjoint Measurement: The Luce-Tukey Axiomatization and Some Extensions," *Journal of Mathematical Psychology,* 1, 248–77.

Kruskal, Joseph B. (1965), "Analysis of Factorial Experiments by Estimating Monotone Transformations of the Data," *Journal of the Royal Statistical Society, Series B*, 27, 251–63.

Lancaster, Kelvin (1971), *Consumer Demand: A New Approach*, New York: Columbia University Press.

Luce, R. Duncan, and Tukey, John W. (1964), "Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement," *Journal of Mathematical Psychology*, 1, 1–27.

McAlister, Leigh (1978), "Choosing Multiple Items from a Homogenous Product Class," unpublished doctoral dissertation, Graduate School of Business, Stanford University.

McClain, John O., and Rao, Vithala R. (1974), "Trade-offs and Conflicts in Evaluation of Health Systems Alternatives: A Methodology for Analysis," *Health Services Research*, 9, 35–52.

McCullough, James M. (1978), "Identification of Preference Through Conjoint Measurement: A Comparison of Data Collection and Analytical Procedures," Working Paper, College of Business Administration, The University of Arizona.

McFadden, Daniel (1976), "Quantal Choice Analysis: A Survey," *Annals of Economic and Social Measurement*, 5, 363–90.

Montgomery, David B., Wittink, Dick R., and Glaze, Thomas (1977), "A Predictive Test of Individual Level Concept Evaluation and Trade-off Analysis," Research Paper No. 415, Graduate School of Business, Stanford University.

Naylor, Thomas, H., Balintfy, Joseph L., Burdick, Donald S., and Chu, Kong (1966), *Computer Simulation Techniques*, New York: John Wiley & Sons, Inc.

Nehls, Lyle, Seaman, Bruce, and Montgomery, David B. (1976), "A PLI Program for Trade-Off Analysis," Cambridge, MA: Marketing Science Institute.

Oppedijk van Veen, Walle M., and Beazley, David (1977), "An Investigation of Alternative Methods of Applying the Trade-off Model," *Journal of the Market Research Society*, 19, 2–9.

Parker, Barnett R., and Srinivasan, V. (1976), "A Consumer Preference Approach to the Planning of Rural Primary Health Care Facilities," *Operations Research*, 24, 991–1025.

Pekelman, Dov, and Sen, Subrata (1974), "Mathematical Programming Models for the Determination of Attribute Weights, *Management Science*, 20, 1217–29.

———, and Sen, Subrata (1978), "Improving Prediction in Conjoint Measurement," Working Paper, Graduate School of Management, The University of Rochester.

Pessemier, Edgar A., Burger, Philip, Teach, Richard, and Tigert, Douglas (1971), "Using Laboratory Brand Preference Scales to Predict Consumer Brand Purchases," *Management Science*, 17, B371–85.

Punj, Girish N., and Staelin, Richard (1978), "The Choice Process for Graduate Business Schools," *Journal of Marketing Research*, 15 (forthcoming).

Rao, Vithala R. (1973), "On Models for Scaling Second Choice Data," Working Paper, Graduate School of Business and Public Administration, Cornell University.

——— (1977), "Conjoint Measurement in Marketing Analysis," in *Multivariate Methods for Market and Survey Research*, ed. J. N. Sheth, Chicago: American Marketing Association, 257–86.

———, and Solgaard, Hans S. (1977), "An Empirical Evaluation of Alternative Multiattribute Utility Models," Working Paper, Graduate School of Business and Public Administration, Cornell University.

———, and Winter, Frederick W. (1977), "Application of the Multivariate Probit Model for Market Segmentation and Product Design," Working Paper 388, School of Business Administration, University of Illinois at Urbana.

Ratchford, Brian T. (1975), "The New Economic Theory for Consumer Behavior: An Interpretive Essay," *The Journal of Consumer Research*, 2, 65–78.

Rorer, Leonard G. (1971), "A Circuitous Route to Bootstrapping," in *Conference on Personality Measurement in Medical Education*, eds. H. B. Haley, A. G. D'Costa, and A. M. Schafer, Washington, D.C.: Association of American Medical Colleges.

Rosenberg, Milton J. (1956), "Cognitive Structure and Attitudinal Affect," *Journal of Abnormal and Social Psychology*, 53, 367–72.

Russ, Frederick A. (1971), "Consumer Evaluation of Alternative Product Models," unpublished doctoral dissertation: Carnegie-Mellon University.

Scott, Jerome E. (1977), "Demand Curve and Utility Estimation in Industrial Markets," Paper MPB 18.4, Atlanta: Joint National ORSA/TIMS meeting.

———, and Wright, Peter (1976), "Modeling an Organizational Buyer's Product Evaluation Strategy: Validity and Procedural Considerations," *Journal of Marketing Research*, 13, 211–24.

Shocker, Allan D., and Srinivasan, V. (1977a), "LINMAP (Version II): A FORTRAN IV Computer Program for Analyzing Ordinal Preference (Dominance) Judgments via Linear Programming Techniques and for Conjoint Measurement," *Journal of Marketing Research*, 14, 101–3.

———, and Srinivasan, V. (1977b), "Multi-Attribute Approaches for Product Concept Evaluation and Generation: A Critical Review," Research Paper No. 326 (Rev.), Graduate School of Business, Stanford University.

Shugan, Steven M., and Hauser, John R. (1977), "P.A.R.I.S.: An Interactive Market Research Information System," Working Paper 602-002, Graduate School of Management, Northwestern University.

Srinivasan, V. (1978), "A Model and Estimation Procedure for Multi-Stage Decision Processes," Working Paper, Graduate School of Business, Stanford University.

———, and Shocker, Allan D. (1973a), "Linear Programming Techniques for Multidimensional Analysis of Preferences," *Psychometrika*, 38, 337–69.

———, and Shocker, Allan D. (1973b), "Estimating the Weights for Multiple Attributes in a Composite Criterion Using Pairwise Judgments," *Psychometrika*, 38, 473–93.

Torgerson, Warren S. (1958), *Theory and Methods of Scaling*, New York: John Wiley & Sons, Inc.

Tversky, Amos (1967), "A General Theory of Polynomial Conjoint Measurement," *Journal of Mathematical Psychology*, 4, 1–20.

Ullrich, James R., and Painter, John R. (1974), "A Conjoint Measurement Analysis of Human Judgment," *Organizational Behavior and Human Performance*, 12, 50–61.

Westwood, Dick, Lunn, Tony, and Beazley, David (1974), "The Trade-off Model and Its Extensions," *Journal of the Market Research Society*, 16, 227–41.

Whitmore, G. A., and Cavadias, G. S. (1974), "Experimental Determination of Community Preferences for Water Quality-Cost Alternatives," *Decision Sciences*, 5, 614–31.

Wilkie, William L., and Pessemier, Edgar A. (1973), "Issues in Marketing's Use of Multi-Attribute Attitude Models," *Journal of Marketing Research*, 10, 428–41.

Wind, Yoram, and Spitz, Lawrence, K. (1976), "An Analytical Approach to Marketing Decisions in Health Care Organizations," *Operations Research*, 24, 973–90.

Wright, Peter (1975), "Consumer Choice Strategies: Simplifying vs. Optimizing," *Journal of Marketing Research*, 12, 60–7.

Young, Forrest W. (1969), "Polynomial Conjoint Analysis of Similarities: Definitions for a Special Algorithm," Research Paper No. 76, Psychometric Laboratory, University of North Carolina.