# Estimating Temporal Trends in Preferences Measured by Graded Paired Comparisons

Joel Huber, *Duke University*
Morris B. Holbrook, *Columbia University*

*A sizable marketing research literature on preference measurement, prediction, and explanation has thus far been more concerned with point estimates of preference than with consistent changes over time. Such temporal trends can be estimated using data collected from graded paired comparisons. An appropriate technique for deriving individual and group trend parameters is first described and then illustrated using data from an experiment dealing with aesthetic preferences. Suggestions are offered concerning its advantages and limitations for applications in marketing.*

Marketing researchers have used numerous techniques for measuring, predicting, and explaining brand preferences across a number of different offerings. For example, multiattribute attitude models predict affect as a linear combination of probabilistic attribute expectancies (or satisfaction ratings) weighed by attribute evaluations (or importance scores) [23]. Similarly, conjoint analysis uses an additive function to decompose brand evaluations into a set of part-worth utilities for separable product features [8,9]. Meanwhile, spatial models explain preference by the relative multidimensional distances of brands from idiosyncratic ideal locations [6,7]. Such approaches share a common focus on the microstructure of preferences: attempting to relate brand affects to their underlying attributes or dimensions. Generally, however, they have adopted a static viewpoint and have neglected the process wherein preferences consistently change over time. Yet, where affective change is likely to occur in response to exposure to new stimuli, such temporal phenomena may be crucial to understanding dynamics of the consumer's ultimate affective distinctions among brands.

This paper explores a technique for estimating systematic preference-development trends in data based on graded paired comparisons. One of the major weaknesses of graded paired comparisons—the necessity of numerous comparisons costing considerable respondent time and effort—becomes an advantage in that trends in preferences, reflecting

*Address correspondence to Joel Huber, Graduate School of Business, Duke University, Durham, NC 27706.*

Journal of Business Research **10**, 459–473 (1982)

52 Vanderbilt Ave., New York, NY 10017

0148-2963/82/040459-15$2.75

that time and effort, can be measured. Applications of this approach are described and illustrated in the present study.

## Graded Paired Comparisons

Analysis of data from graded paired comparisons has been discussed elsewhere [16] and is reviewed here only briefly. In a graded paired comparison, the respondent specifies which of two objects is preferred and indicates the strength of that preference. If a large number of these comparisons is made on pairs within a set of stimuli, a one-dimensional scale of value can be constructed such that the signed differences on that scale correspond maximally to the original paired preference comparisons. Moreover, when such judgments are collected on a set of $n$ objects, the large number of elicited judgments [$\frac{1}{2}n(n - 1)$] typically provides more than enough observations to estimate the $n$ object-specific preference parameters. Accordingly, some of the excess degrees of freedom may be used to estimate temporal trends or systematic changes in these preference values within the time span of the data collection task.

Scheffe [22] introduced an analysis of variance for graded paired comparisons that was modified by Bechtel [1–3], to apply to the case of individual respondents. Pessemier and Teach [21] developed a similar model using absolute rather than squared error as the criterion of fit. Finally, Hauser and Shugan [11] augmented an axiomatic tradition related to algebraic difference structures [18] to provide nonparametric tests for ascertaining the metric quality of preference measures.

The method given below differs in that additional parameters are added to calibrate the effect of consistent temporal trends and the analysis is framed in such a way to highlight the heterogeneity of responses found across subjects.

## Applicability

In business practice one is often concerned with temporal changes in preferences—especially among stimuli that are primarily aesthetic in nature. For example, producers of phonograph records need to identify recordings for which relative liking will increase over time. Conversely, advertisers are concerned with choosing artwork that has minimum wearout with repeat exposure. Accordingly, aesthetic stimuli—specifically, musical recordings—were chosen, for purposes of illustration, to represent the class of products for which preferential tastes would be expected to develop differentially during the first few exposures.

## Least-Squares Estimation of Order Bias, Preference Value Temporal Trend

The analysis proceeds from the assumption that each graded paired comparison represents a directional difference on a one-dimensional preference scale plus a within-pair order bias. Thus, for a given respondent or a group

$$P_{ijt} \quad B + V_{it} \quad V_{jt} + e_{ijt}$$

where

$P_{ijt}$ = the graded paired comparison between stimulus $i$ (first) and stimulus $j$ (second) at time $t$;
$B$ = the presentation order bias due to $i$ being presented before $j$;
$V_{it}$ = the preference value of stimulus $i$ at time $t$; and
$e_{ijt}$ = error, assumed to be independently, identically distributed.

The trend $T_i$ in preference for stimulus $i$ is assumed to be a linear function of time $t$ so that

$$V_{it} = V_i + T_i \quad t, \qquad (2)$$

where $V_i$ is the preference value at time $t = 0$. In the present case, the linear trend is used as a sensitive measure of preference change during the time span of the experiment (approximately four hours). Other forms of temporal change could be assumed by defining levels of time as dummy variables or by taking a quadratic or exponential transformation of the time variable.

Directional dummy variables allow estimation of the above model. Specifically, Eq. (1) and (2) may be combined and rewritten as follows:

$$P_{ijt} = B + \sum_k V_k \quad d_k + \sum_k T_k \quad t \cdot d_k + e_{ijt} \qquad (3)$$

where

$$d_k = \begin{cases} & \text{if } k = i \\ & \text{if } k = j \\ 0 & \text{otherwise} \end{cases}$$

Using this formulation, least-squares approximations of the parameters can be estimated with common regression techniques by defining a dummy variable $d_k$ for each stimulus and a time-by-stimulus variable $(t \cdot d_k)$ for each growth trend. The constant term $B$ serves to estimate within-pair order bias, whereas the regression coefficients $V_k$ and $T_k$ estimate the preference value and temporal trend parameters for stimulus $k$.

In practice, since the complete set of dummy terms is linearly dependent, the full matrix of predictor variables is singular and cannot be inverted. Accordingly, one stimulus is dropped from the estimation procedure. Its value and trend coefficients are set arbitrarily at zero to serve as reference points for scaling the corresponding parameters of the other stimuli. It should be noted that this analytic step places all comparisons on a relativistic basis. That is, one is not able to assess absolute preference levels or growth rates, but only preferences and trends relative to other stimuli. This restriction is, of course, common to most forms of preference analysis in marketing and seems appropriate in all cases where ratio-scaled attitudinal data are not available. In some cases, it may be possible to convert these scales to have absolute meaning by suitable anchoring questions. For example, a behavioral anchor could be formed by labeling one of the stimuli "the product you most use now." Alternatively, a probabilistic anchor could be formed by asking respondents to estimate the likelihood of purchasing the most and the least liked items in the set. Such anchors would allow a researcher to track better the consumer's response to the entire set of stimuli as well as their relative positions.

**Decomposition into Group and Individual Components**

A remaining issue concerns the estimation of parameters at the group as opposed to the individual level. The sum of squares (SSR) shown in Table 1 for the various groups are estimated by the following regressions to predict the paired comparisons as a function of various combinations of order bias $B$, preference value $V$, and temporal trend $T$:

Step 1. *A pooled regression* (stacking all responses into one long matrix) to predict graded pairs as a function of order bias and preference values and produces $SSR^G(B)$ from the sum of squares due to the mean and $SSR^G(B,V)$ from the sum of squares due to the regression.

Step 2. *The above run including the temporal trend terms*. This run results in $SSR^G(B,V,T)$, from which $SSR^G(T)$ can be computed.

Step 3. *Separate runs using order bias and preference value* for each of the $m$ subjects in the segment. $SSR^P(B)$ and $SSR^P(B,V)$ result from aggregating the individual sums of squares.

Step 4. *Individual runs*, as in Step 3, *with the full model*, including temporal trend terms. These when aggregated, result in $SSR^P(B,V,T)$.

As displayed in Table 1, the analysis is strongly hierarchical. The order bias (constant) is removed first, followed by average preference values (main effects) and then by any remaining temporal trends (interactions). With respect to the level of analysis, group effects are removed prior to their individual-level counterparts.

The interpretation of the decomposition and the appropriate $F$-tests must take this hierarchy into account. In performing the $F$-tests, the mean square due to error is typically used as the denominator in the $F$-ratio. However, if a lower-order effect (such as individual preference values) is not significant, its sum-of-squares and degrees of freedom should be pooled with the error term in testing for higher-level effects. Conversely, the interpretation of a lower-order effect also follows the hierarchy. That is, the test on individual preference values examines the null hypothesis that all subjects have preferences equal to the average, whereas the test on group preference values determines whether the average preferences can themselves be interpreted as equal to zero. If both tests are significant, the group effect and the deviations of individuals from the group norm are both greater than could be expected by chance. The interesting case occurs, however, when the individual effect is significant and the *group effect* is not. In that case the individual preference values have effectively cancelled when aggregated so that one cannot reject the hypothesis of a negligible group component. The appropriate adjustment to the $F$-test is to pool the nonsignificant sum of squares from the group with that due to the individuals. The resultant $F$-ratio then tests the more general hypothesis that all individuals have zero parameter values. The decomposition given here does not include a term for deviation from subtractivity found in the work of Bechtel and O'Connor [3]. This term, representing consistent intersubject deviations from a one-dimensional preference scale, would be difficult to estimate using the dummy variable framework presented here because the deviation from subtractivity requires $\frac{1}{2}(n\text{-}2)\,(n\text{-}1)$ parameters, where $n$ is the number of stimuli. Fortunately, the deviation from subtractivity, while statistically significant, has been found to account for "a negligible proportion of the sum of squares" [3, p. 253]. Furthermore, the sum of squares due to segment order bias (included here but not in [3]) is a subset of the sum of squares due to deviation from subtractivity and may account in large part for its significance.

Table 1: Hierarchical Analysis of Variance for Order Bias $B$, Preference Value $V$, and Temporal Trend $T$

| Source | d.f. | Sum of Squares[a] |
|---|---|---|
| Group order bias | | $SSR^G (B)$ |
| Group preference values | $n - 1$ | $SSR^G (B, V) - SSR^G (B) = SSR^G (V)$ |
| Group temporal trends | $n - 1$ | $SSR^G (B, V, T) - SSR^G (B, V) = SSR^G (T)$ |
| Individual order biases | $m - 1$ | $SSR^P (B) - SSR^G (B) = SSR^I (B)$ |
| Individual preference values | $(m - 1)(n - 1)$ | $SSR^P (B, V) - SSR^P (B) - SSR^G (V) = SSR^I (V)$ |
| Individual temporal trends | $(m - 1)(n - 1)$ | $SSR^P (B, V, T) - SSR^P (B, V) - SSR^G (T) = SSR^I (T)$ |
| Error | #OBS $- (2mn - m)$ | $SSE^P (B, V, T)$ |

[a] $SSR^G$ = The sum of squares due to group analysis defined across all $n$ stimuli and $m$ individuals.
$SSR^P$ = The pooled sum of squares of $m$ individual analyses.
$SSR^I$ = The net individual sum of squares due to differences among individuals within the group.

**Strength of Association:** $\omega^2$ For many consumer applications, given that the statistical tests are satisfied, attention is likely to be less on levels of significance than on descriptive statistics that characterize subgroups of respondents. For this purpose, the $\omega^2$ measure is useful [5,12]. This statistic is estimated for factor $X$ as

$$\omega^2 = \frac{\mathrm{SSR}(X) - \mathrm{d.f}\,(X) \cdot \mathrm{MSerror}}{\mathrm{SStotal} + \mathrm{MSerror}} \tag{5}$$

and can be viewed as the proportion of variance due to a factor, adjusted for degrees of freedom.

The regression model, partititioning scheme, and $\omega^2$ statistic described above can be used flexibly to test a host of possible hypotheses specific to particular stimuli.

## An Illustrative Study

**Stimulus Objects** Stimulus objects consisted of 16 jazz recordings identical to those described in previous studies [13,15,16] except that two additional recordings were added in the present study. Since the focus of the present paper is methodological rather than substantive, the delineation of these recordings will not be repeated except to note that they represented sharply contrasting styles of saxophone playing, recorded on separate tape cassettes, and played back to subjects individually through headphones.

**Sample** Respondents with heterogeneous musical backgrounds were recruited with sign-up sheets on university bulletin boards that asked for estimates of (1) proficiency in playing a musical instrument and (2) the number of jazz recordings listened to in a typical month. On the basis of this information, four groups of eight respondents were established to reflect extreme levels on these two dimensions: i.e., High Jazz–High Music, High Jazz–Low Music, Low Jazz–High Music, and Low Jazz–Low Music.

**The Design and Structure of the Graded Paired Comparison Task** The graded paired comparison task followed a rather complex development process. First, each of the 16 saxophone tapes was numbered randomly. Then, the 120 possible pairs were divided into 8 subsets, each containing one cyclic design as shown in Figure 1. Such cyclic designs have the advantage of automatically balancing stimulus and within-pair order effects [17]. That is, in a cycle, each stimulus

STIMULUS SHOWN SECOND

STIMULUS SHOWN FIRST

| | | 2 | | 4 | 5 | 6 | 7 | 8 | | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1** | | 2 | | 3 | | 4 | | 8* | 6 | | 7 | | 5 | |
| | | | 1 | | 2 | | 3 | | 4 | | 8* | 6 | | 7 | | 5 |
| | 5 | | | 1 | | 2 | | 3 | | 4 | | 8* | 6 | | 7 | |
| | | 5 | | | 1 | | 2 | | 3 | | 4 | | 8* | 6 | | 7 |
| 5 | 7 | | 5 | | | 1 | | 2 | | 3 | | 4 | | 8* | 6 | |
| 6 | | 7 | | 5 | | | 1 | | 2 | | 3 | | 4 | | 8* | 6 |
| 7 | 6 | | 7 | | 5 | | | 1 | | 2 | | 3 | | 4 | | |
| 8 | | 6 | | 7 | | 5 | | | 1 | | 2 | | 3 | | 4 | 8* |
| 9 | 8 | | 6 | | 7 | | 5 | | | 1 | | 2 | | 3 | | 4 |
| 10 | 4 | 8 | | 6 | | 7 | | 5 | | | 1 | | 2 | | 3 | |
| 11 | | 4 | 8 | | 6 | | 7 | | 5 | | | 1 | | 2 | | 3 |
| 12 | 3 | | 4 | 8 | | 6 | | 7 | | 5 | | | 1 | | 2 | |
| 13 | | 3 | | 4 | 8 | | 6 | | 7 | | 5 | | | 1 | | 2 |
| 14 | 2 | | 3 | | 4 | 8 | | 6 | | 7 | | 5 | | | 1 | |
| 15 | | 2 | | 3 | | 4 | 8 | | 6 | | 7 | | 5 | | | 1 |
| 16 | 1 | | 2 | | 3 | | 4 | | 8 | 6 | | 7 | | 5 | | |

**FIGURE 1.** Cyclic design. *Replication of pairs in the fourth set with reversed orders. **Read: In the first set stimulus 1 is shown first in a comparison with stimulus 2.

occurs exactly once in the first- and second-order position. As shown in Figure 1, this design provided 128 data points for each respondent tested.

The overall task exposed subjects to eight cyclic sets of stimulus pairs and lasted about four hours in all. The time parameter was defined as the temporal order of the set so that $t$ had 8 levels. This temporal order was rotated so that, for every group of eight respondents, each cycle occurred in each sequential position. Thus, within-pair presentation order was balanced within each cyclic set of stimulus objects; presentation time for each stimulus (defined as the temporal order of the set) was balanced within each subject; and finally, cycle order was balanced within each group of subjects. The somewhat elaborate balancing used in this study was not necessary for least-squares parameter estimation, but did render the design more efficient by minimizing multicollinearity among the variables.

Upon listening to a pair of tapes, a subject indicated which was preferred and the degree of preference on the following check-mark scale (illustrated for objects 2 and 5):

| Pair | Write Number Preferred | Degree of Preference (Small → Great) 0 1 2 3 4 5 |
|---|---|---|
| 5-2 | 2 | :__:__:✓:__:__:__ |

This task is one of a number of available graded paired comparison measures. For example, Bechtel and O'Connor [3] follow Cooper [4] in gauging "strength of preference" on a 10-point numerical scale running from "very weak" (0) to "very strong" (9). Pessemier and Teach [21] use the price gap that would make a person indifferent between two versions of a product. This "dollarmetric" analysis has also been employed by others [14, 20]. Finally, Hauser and Shugan [11] ask respondents to divide 100 chips between the members of a pair to indicate relative intensity of preference. The term "graded paired comparison" is intended as a general reference to all such measures that indicate both direction and strength of preference between objects. Indeed, the analytic procedure presented earlier can easily be adapted to any of the specific operationalizations mentioned.

## Results

The full set of individual order bias, preference value, and temporal trend parameters is too voluminous to report here. Accordingly, results are summarized in the form permitted by the partitioning into group and individual effects. Table 2 presents a decomposition of variance for the four segments defined earlier on the basis of jazz knowledge and ability to play a musical instrument. Results for the three factors in the general model—order bias, preference value, and temporal trend—can be summarized in terms of their strength and degree of homogeneity within segments.

**Within-Pair Order Bias** Order bias reflects the tendency of subjects to prefer the item listened to first within a pair. In the present case, all the segments had *negative* order bias parameters, indicating a kind of "recency effect" in assigning higher scores to the recording that came second. The size of the group order bias ranged from $-0.40$ (on a 6-point scale) for the high–high group to $-0.16$ for the low-low group, and was statistically significant at the 0.05 level for all but the latter group. However, although group order biases were statistically significant, the

## Table 2: Decomposition of Variance for Four Groups

| | | High jazz–high music | | High jazz–low music | | Low jazz–high music | | Low jazz–low music | |
|---|---|---|---|---|---|---|---|---|---|
| Source | d.f. | *F*-test | $\omega^2$ | *F*-test | $\omega^2$ | *F*-test | $\omega^2$ | *F*-test | $\omega^2$ |
| Group order bias | 1 | 22.9[a] | 0.01 | 22.5[a] | 0.01 | 13.6[a] | 0.01 | 3.5 | 0.00 |
| Group preference values | 15 | 19.9[a] | 0.14 | 7.4[a] | 0.11 | 22.8[a] | 0.13 | 12.5[a] | 0.07 |
| Group temporal trends | 15 | 3.5[a] | 0.02 | 2.0[b] | 0.01 | 2.7[a] | 0.01 | 0.9 | 0.00 |
| Individual order bias | 7 | 1.7 | 0.00 | 8.4[a] | 0.02 | 2.0 | 0.00 | 13.4[a] | 0.04 |
| Individual preference values | 105 | 6.6[a] | 0.29 | 9.3[a] | 0.38 | 10.6[a] | 0.39 | 9.8[a] | 0.41 |
| Individual temporal trends | 105 | 1.4[b] | 0.02 | 1.4[b] | 0.02 | 2.6[a] | 0.06 | 1.4[b] | 0.02 |
| Error | 776 | | | | | | | | |

[a] F-test significant at 0.01 level.
[b] F-test significant at 0.05 level.

Table 3 Effects of Music and Jazz Knowledge on the Explained Variance and Homogeneity of Preference Values and Temporal Trends Within Segments

| Segments | Preference Values | | Temporal Trends | |
|---|---|---|---|---|
| | Total $\omega^2$ | Due to Group[a] | Total $\omega^2$ | Due to Group[a] |
| High jazz–high music | 0.43 | 0.33 | 0.04 | |
| High jazz–low music | 0.49 | 0.22 | 0.03 | |
| Low jazz–high music | 0.46 | 0.25 | 0.07 | 0.14 |
| Low jazz–low music | 0.49 | 0.15 | 0.02 | 0.00 |

[a] For each factor this is $\omega^2$ for the group divided by the total $\omega^2$ square.

$\omega^2$ measures shown in Table 2 indicate that the combined individual and segment effects accounted for only 2–4% of the total variance. Nevertheless, these effects were large enough that a careful analyst might want to use the present least-squares approach to remove them before estimating values or trends.

**Preference Value** The majority of the explained variance is accounted for by the preference values of the stimuli. Group and individual $\omega^2$ totaled approximately 45% in each segment (Table 3). There are, however, large differences in the distribution of explained variance between the individual and group components. As a relativistic measure of homogeneity, one can divide the $\omega^2$ due to preference values at the group level by the sum of $\omega^2$ due to preferences at the group and individual levels combined. As shown in Table 3, this measure of homogeneity ranges from 0.15 in the low–low group to 0.33 in the high–high group, indicating that individuals' preferences tend to be more like their groups' as knowledge of jazz and ability to play an instrument increase.

**Temporal Trends** An analogous measure of homogeneity in temporal trends produces essentially the same result (Table 3). That is, without specifying the magnitude of preference changes, the homogeneity of such trends increases (from 0.0 to 0.50) with knowledge of jazz and ability to play a musical instrument.

By contrast, the *amount* of such preference change—as measured by the combined group and individual $\omega^2$ shown in Table 3—is greatest for the segment with high musical training but low knowledge of jazz (0.01 + 0.06 = 0.07). This finding suggests that, while formal musical training may have provided evaluative criteria so as to facilitate change in preference values, knowledge of jazz might have acted conversely to inhibit preference development via the resistance of stereotyped or polarized tastes.

To test the reliability of this joint effect of low jazz knowledge and high musical training, the percentage of each individual's trend coefficients significant at the 0.05 level in the high music–low jazz segment (15%) was compared with that in the other three (9%). This difference was significant at the 0.05 level, indicating an interaction between musical training and jazz knowledge in moderating the degree of affective change. Thus, although the percentage of variance due to temporal trend was small (on the order of 7%), the differences among groups were meaningful and significant.

## Discussion

In summary, this paper has employed data from a study of aesthetic preferences to illustrate the use of the graded paired comparison task together with a method for estimating within-pair order bias, preference values, and temporal trends through least-squares regression on signed dummy variables. Results can be categorized via a decomposition of variance that quantifies the homogeneity of the parameters across members of a group.

The methodology is quite flexible and not limited in ways that happened to be appropriate to the present data. The task, for example, can take many forms—from the division of 10 points between stimuli of a pair described by Haley and Case [10] to the intensity measures of Hauser and Shugan [11] and the dollarmetrics of Pessemier and Teach [21]. Furthermore, if one wishes to adopt the group level of analysis, a transformation of the proportion preferring one item over another (e.g., the logit, probit, or arcsin transformations) can be used as input.

In some studies the temporal trend and within-pair order bias might simply be viewed as "nuisance" parameters whose effects should be taken into account but not entertained (as in classical analysis of covariance). As such the regression analysis provides a way to account for these effects. In other applications, temporal trends might be of considerable substantive interest in their own right.

The jazz data appear to illustrate the former perspective. Thus, the order parameter indicated a recency bias that a careful analyst might wish to remove statistically. Similarly, the effect of temporal trend did not account for much of the variance in the current analysis and thus might be best thought of as a nuisance variable. In *other* applications, however, the temporal trends might be expected to be much stronger and to have direct managerial significance. Suppose consumers were asked to evaluate one wave (16 pairs) of competing brands each month. The temporal trend would then provide a rather sensitive measure of preference change that could be related to exposure to promotion and/or brand usage.

Apart from the measures on trends, the decomposition of preference components into group and individual effects permits a useful description of the *structure* of preferences both within and across customers. For example, a low within-subject $\omega^2$ means that individual tastes are not internally consistent, reflecting unreliable paired judgments. Such fluidity of individual tastes occurs when product differences are small (as can occur in a blind taste test) or when products tested are in the early stages of the product life cycle before consumers' preference orderings have been well formed. This fluidity for a product class may be a signal that promotional campaigns to stimulate trial and attitude changes are likely to be successful, particularly in contrast with product classes having high within-subject $\omega^2$, signifying a concomitant rigidity of tastes.

Also arising from the analysis, the $\omega^2$ for group preferences reflects the homogeneity of preferences *across* rather than within respondents. A low group $\omega^2$ (particularly relative to a high individual $\omega^2$ as occurred in the music data) implies that group members desire alternative product formulations. Such a result would indicate that the group could be further segmented in terms of benefits desired. This benefit segmentation could proceed by clustering the individual preference scores to arrive at subgroups with homogeneous product desires.

Thus, the proposed methodology can provide managerially useful input into the structure of the market. Although the current study attests to the feasibility of collecting numerous graded pairs in a rather difficult data collection context, it is reasonable to examine the context in which the extra data-collection work of graded pairs is justified. Generally, the extra effort of graded pairs is worthwhile if the research focus requires both stable estimates and statistical information at the individual level. For example, in a segmentation study it is important to be able to distinguish between real intersubject differences and intrasubject error—which is possible with graded pairs since, due to the extra degrees of freedom on which they are derived, the estimated parameters have confidence intervals at the individual level.

It should be acknowledged, however, that the degrees of freedom enabling the estimation of so many parameters are accompanied by a cost. First, in aggregating across subjects, the assumption of uncorrelated error terms is likely to be violated since a respondent with high error on one pair is also likely to have high errors on others. The effect of such heteroscedasticity is to render suspect the statistical test on the pooled regressions.

A second problem is behavioral and relates to the effects the paired task has on respondents. Simply having a pair, rather than monadic (one stimulus at a time) rating, focuses attention on the differences between alternatives. This comparative perspective may make the consumer more sensitive to product differences than when actually choosing a brand. Further, the large number of responses required may induce a response set in subjects to simplify the task. Thus the effect of having comparisons on pairs appears to lead to *more* processing whereas the large number of judgments leads to *less*, and it is unlikely that the effects would exactly cancel. Although these problems occur with any intense utility-estimation procedure, research is currently being conducted (and certainly more needs to be done) on the extent to which these attentional biases invalidate measures of preference.

In the meantime, however, the wealth of data collected in a graded pairs task that allows individual preferences and trends to be estimated is often worth the cost. Furthermore, the partitioning of variance into group and individual components permits a useful measure of group homogeneity. Thus the approach given here to collecting and analyzing graded paired comparisons offers potential applications, particularly where market researchers are concerned with the development of individual brand or product preferences over time.

*Morris B. Holbrook gratefully acknowledges the support of Columbia University's Faculty Research Fund.*

## References

Bechtel, Gordon G., The Analysis of Variance and Pairwise Scaling, *Psychometrika* 39 (March 1967): 47–65.

2. Bechtel, Gordon G., *Multidimensional Preference Scaling*. Mouton, The Hague-Paris, 1976.

3. Bechtel, Gordon G., and O'Connor, P. J., Testing Micropreference Structures, *J. Marketing Res.* 16 (May 1979): 247–257.

4. Cooper, Lee G., Multivariate Investigation of Preferences, *Multivar. Behav. Res.* 8 (April 1973): 253–272.

5. Green, Paul E., On The Analysis of Interactions in Marketing Research Data, *J. Marketing Res.* **10** (November 1973): 410–420.

6. Green, Paul E., and Carmone, Frank J., *Multidimensional Scaling and Related Techniques in Marketing Analysis.* Allyn and Bacon, Boston, MA, 1970.

Green, Paul E., and Wind, Yoram, *Multiattribute Decisions in Marketing: A Measurement Approach.* Dryden Press, Hinsdale, IL, 1973.

Green, Paul E., and Srinivasan, V., Conjoint Analysis in Consumer Behavior: Issues and Outlook, *J. Consumer Res.* **5** (September 1978): 103-123.

Green, Paul E., and Rao, Vithala, R., Conjoint Measurement for Quantifying Judgmental Data, *J. Marketing Res.* **8** (August 1971): 355-363.

10. Haley, Russell I., and Case, Peter B., Testing Thirteen Attitude Scales for Agreement and Brand Discrimination, *J. Marketing* **43**, (Fall 1979): 20-23.

11 Hauser, John, and Shugan, Steven M., Intensity Measures of Consumer Preference, *Oper. Res.* **28** (March-April 1980): 279-320.

12. Hays, William L., *Statistics for the Social Sciences,* 2nd ed. Holt Rinehart and Winston, New York, 1973.

13. Holbrook, Morris B., and Huber, Joel, Separating Perceptual Dimensions from Affective Overtones: An Application to Consumer Aesthetics, *J. Consumer Res.* **5** (March 1979): 272-283.

14. Huber, Joel, and James, William, The Marginal Value of Physical Attributes: A Dollarmetric Approach. Presented at the 7th American Marketing Association Attitude Conference, Hilton Head, SC, February, 1976.

15. Huber, Joel, and Holbrook, Morris B., Using Attribute Ratings for Product Positioning: Some Distinctions Among Compositional Approaches, *J. Marketing Res.* **16** (November 1979): 507-516.

16. Huber, Joel, and Holbrook, Morris B., The Determinants of Esthetic Value and Growth, *Adv. Consumer Res.* **8**, Association for Consumer Research, 1980.

17. John, J. A. Wolock, F. W., and David, H. A., *Cyclical Designs.* National Bureau of Standards, Applied Mathematics Series, No. 62, 1972.

18. Krantz, D. H., Luce, R. D., Suppes, P., and Tversky, A., *Foundations of Measurement,* Vol. 1. Academic Press, New York, 1971.

Montgomery, D. B., and Morrison, D. G., A Note on Adjusting $R^2$, *J. Fin.* **28** (1973): 1009-1013.

20. Nevin, John R., Laboratory Experiments for Estimating Consumer Demand: A Validation Study, *J. Marketing Res.* **10** (August 1974): 261-268.

21 Pessemier, E. A., and Teach, R. D., Dissaggregation of Analysis of Variance for Paired Comparisons: An Application to a Marketing Experiment, Krannert Paper #282, Krannert School of Management, Purdue University, Lafayette, IN, 1970.

Scheffe, H., An Analysis of Variance for Paired Comparisons, *J. Am. Stat. Assoc.* **47** (1952): 381–400.

Wilkie, William L., and Pessemier, Edgar A., Issues on Marketing's Use of Multiattribute Attitude Models, *J. Marketing Res.* **10** (November 1973): 428–441.