Separating Perceptual Dimensions from Affective Overtones: An Application to Consumer Aesthetics

Morris B. Holbrook; Joel Huber

Separating Perceptual Dimensions from Affective Overtones: An Application to Consumer Aesthetics

MORRIS B. HOLBROOK
JOEL HUBER*

Different types of perceptual distortion due to affective overtones are identified. A procedure, which deals with problems of distortion more comprehensively than competing methods, is proposed and illustrated in the case of consumer aesthetics. The resulting bias-free joint space is clearly interpretable and performs relatively well in predicting affect.

E ach approach to building a model of the consumer's attitude structure has suffered from its own shortcomings. A *decompositional* approach like multidimensional scaling, based on pairwise similarity judgments, may generate a perceptual space with dimensions that are virtually impossible to name (Silk 1969; Smith and Lusch 1976). When ideal points are projected into such spaces by means of unfolding analysis, their positions may be extremely unstable and/or difficult to interpret (Holbrook and Williams 1978; Kruskal, Young, and Seery 1973). Worse, when unfolding analysis is used to develop maps based on preference data alone, the result may be an affective space with dimensions that cannot be regarded as purely perceptual (Green and Carmone 1969). But such intrusions of affect into supposedly perceptual dimensions may be even more serious in *compositional* approaches based on the ratings of objects on a number of attributes where affective overtones—often called the "halo" effect—may cause attribute ratings to be colored by overall preferences. Indeed, such distortions of perceptual judgments by feedback from affect may render the predictive results obtained by expectancy-value or adequacy-importance models largely spurious (Beckwith and Lehmann 1975), thus invalidating their use as bases for product-positioning decisions.

Decompositional approaches to developing ideal-point attitude models may be beset by instability and interpretive difficulties while compositional techniques may be plagued by perceptual distortions due to affective overtones. In addition, both types of problem may be greatly intensified by the typical practice in consumer research of taking some set of products or brands with which the respondent may have only a passing acquaintance and eliciting a host of judgments that he cannot possibly be equipped or motivated to make intelligently. The prototypical application of multiattribute attitude models to consumer behavior, for example, might list eight brands of toothpaste and ask for ratings of each brand on five attributes, e.g., decay prevention, whitening power, taste/flavor, breath-freshening, value for the money (Bass and Talarzyk 1972; Bass and Wilkie 1973; Blackwell, Engel, and Talarzyk 1977; Mazis, Ahtola, and Klippel 1975; Sheth and Talarzyk 1972; Wilkie, McCann, and Reibstein 1974). Clearly, such judgments vastly exceed the likely range of experience and cognitive capabilities of the typical consumer.

Not surprisingly, then, the respondent tends to rely on overall preference to infer individual attributes. Crest is well-liked and, therefore, assumed to be high on all the favorable attributes (Blackwell et al. 1977). The halo effect, or "cognitive consistency," reigns supreme and the researcher learns little about true perceptual judgments.

Common Versus Idiosyncratic Perceptual Distortion

Although perceptual distortion due to affective overtones has been a widely recognized phenomenon

since Thorndike (1920) first described the "halo" in the evaluation of personnel, only rarely has it been pointed out that there are at least two separate processes at work (Blumberg, De Soto, and Kuethe 1966; Burnaska and Hollmann 1974; Stanley 1961; Willingham and Jones 1958). *Common perceptual distortion* arises to the extent that (a) preferences within a group are relatively homogeneous and (b) consensus exists concerning the favorability of certain attributes, so that the average perceptions on those attributes for the group as a whole are raised (lowered) for those objects generally viewed as good (bad).

By contrast, *idiosyncratic perceptual distortion* occurs when a particular individual's global evaluation of an object pulls his or her own rating of the object on some attribute away from the group's mean rating. Such idiosyncratic distortion does not depend on shared preferences or on consensus concerning attribute favorability. Indeed, either heterogeneity of preferences across subjects or disagreement concerning the favorability of attributes is required in order for idiosyncratic distortion to occur. Its effects may therefore be superimposed upon those of common perceptual distortion.

Measuring Common and Idiosyncratic Perceptual Distortion of Unequivocal Attributes

As recent but selective reviews in the consumer research literature have implied (Beckwith, Kassarjian, and Lehmann 1978; Huber and James 1978; James and Carter 1978), psychologists have focused primarily on attempting to assess the degree of perceptual distortion introduced by various rating formats, with relatively little attention to the problem of how to remove this bias once it exists (Guilford 1954; Stanley 1961). Although the present paper deals primarily with this latter issue, it is useful to review briefly the various ways psychologists and consumer researchers have measured both idiosyncratic and common distortion.

In psychology, the best developed methods for assessing the degree of idiosyncratic halo effect are Guilford's (1954) analysis-of-variance measure of rater–object interaction (Blumberg et al. 1966; Burnaska and Hollmann 1974; Friedman and Cornelius 1976; Johnson 1963; Kavanagh, MacKinney, and Wolins 1971; Stanley 1961; Willingham and Jones 1958) and Brown's (1968) measure of the variance in an object's ratings across attributes (Bernardin 1977; Bernardin and Walter 1977; Borman 1975, 1977; Borman and Dunnette 1975; James and Carter 1978; Rizzo and Frank 1977; Wilkie, McCann, and Reibstein 1974).

By contrast, common perceptual distortion has usually been assessed by the degree of intercorrelation between attributes defined across raters and objects (Borman and Vallon 1974; Gilinsky 1947;

Keaveny and McGann 1975; Landy, Farr, Saal, and Freytag 1976; Landy and Guion 1970; Taylor and Hastman 1956; Zedeck, Kafry, and Jacobs 1976)—a measure that Willingham and Jones (1958) have shown to be equivalent to the complement of the attribute–object interaction in the analysis-of-variance approach (Burnaska and Hollmann 1974). The magnitude of the attribute–object interaction has also been likened to the assessment of discriminant validity defined through Campbell and Fiske's (1959) multitrait-multimethod matrix with attributes as traits and raters as methods (Kavanagh et al. 1971).

Other ways of assessing common perceptual distortion have included Symonds' (1925) focus on an attribute's correlations between judges across objects and Latham, Wexley, and Pursell's (1975) and Nisbett and Wilson's (1977) tests of the effects of experimentally manipulated global evaluation on separate attribute ratings.

In consumer research, a method to measure the extent of idiosyncratic perceptual bias was pioneered by Beckwith and Lehmann (1975) and has been applied in several subsequent studies (Bemmaor and Huber 1978; James and Carter 1978; Moore and James 1978; Johansson, MacLachlan, and Yalch 1976; Beckwith and Lehmann 1976). Briefly, this approach regresses each respondent's belief on a given attribute toward each object against both the group's mean belief on that attribute for each object and the respondent's own global evaluation of each object. The relative size and significance of the resulting beta coefficients are taken as measures of the degree of halo effect: the larger the relative role of an individual's global evaluation in predicting his belief, the greater the assumed perceptual distortion. Use of this test has suggested the presence of idiosyncratic perceptual distortion ranging from extreme (Beckwith and Lehmann 1975) to moderate (James and Carter 1978) to negligible (Bemmaor and Huber 1978; Moore and James 1978).

A key problem with the psychological procedures mentioned earlier and shared by the Beckwith-Lehmann method is that all such approaches require the assumption that each attribute is viewed as unequivocally favorable or unfavorable. But, if an individual's ideal point were located somewhere near the middle of the attribute scale, attribute ratings might be biased upwards for favorably evaluated low-rated objects and biased downwards for favorably evaluated high-rated objects. These contrary movements along the attribute scale might cancel and lead to the erroneous conclusion that idiosyncratic perceptual distortion was quite weak whereas, in fact, it had been quite strong.

Potentially, the key advantage of the Beckwith-Lehmann procedure over most of the previously mentioned approaches is that it does provide an estimate, in the form of group mean belief scores, of an ob-

ject's "true" perceived position. However, as pointed out by Huber and James (1978) and acknowledged by Beckwith and Lehmann (1976), these mean belief ratings may themselves be biased by common affective overtones due to shared preferences. In other words, the Beckwith-Lehmann procedure, as typically applied by consumer researchers, focuses exclusively on the individual level and takes no account of common perceptual distortion. Beckwith and Kubilius (1978) recently proposed a method of circumventing this problem by using an iterative computational procedure for estimating "true" belief scores. A more general approach is described and illustrated in this paper.

A GENERAL APPROACH FOR SEPARATING PERCEPTUAL DIMENSIONS FROM AFFECTIVE OVERTONES

Given the prevalence of consumer research on product classes for which underlying perceptual dimensions are not yet well understood, the preceding discussion indicates the need for an approach to constructing perceptual spaces that avoids the problems of interpretation encountered in decompositional methods, while circumventing the common and idiosyncratic perceptual distortions that plague compositional techniques. The multivariate statistical methods needed for such an approach are readily available. Indeed, the present study introduces no new statistical methodology. Rather its purpose is to illustrate a novel approach to chaining together regression, factor, and discriminant analyses in the construction of spatial representations of attitude structure.

General Procedure

The general approach may be broken down into six steps. Each step is described, then justified briefly before providing an illustrative application of the procedure.

Step 1. A set of stimulus objects is exposed, one at a time, to each subject and rated on a large number of semantic differential scales (Osgood, Suci, and Tannenbaum 1957). Several semantic differential scales are included that are intended to be highly evaluative in nature (e.g., good/bad, pleasant/unpleasant); their purpose is to provide a reliable estimate of affective overtones, so that they may later be removed.

Step 2. The normalized semantic differential ratings are correlated across subjects and objects to identify those that are strongly related to the good/bad scale. The emerging set of primarily affective scales is used to form an additive index of global evaluation for each object rated by each subject.

Step 3. Following a procedure proposed by Myers (1965), each of the remaining semantic differential scales is regressed (across subjects and objects) on this global evaluation index. Residuals from these regressions, which are equivalent to the semantic differential ratings with common affect partialled out, form the basis for subsequent analysis.

Step 4. The residuals from Step 3 are submitted to principal components analysis to obtain a factor structure based on intercorrelations between scales with the effect of global evaluation removed.

Step 5. In accord with Johnson's (1971) method for constructing perceptual maps from multidimensional data, the factor scores are then used as independent variables in multiple discriminant analysis with the stimulus objects themselves as the dependent category variables to be predicted. The resulting discriminant functions compose a multidimensional space with points representing the rating of each object by each subject. The group consensus concerning the position of each object is given by its mean position on each dimension. These centroids are, by definition, free from idiosyncratic perceptual distortion. Further, since common affective overtones have previously been partialled out of the factor structure, the positions of the centroids for each object should tend to be perceptual (rather than affective) in nature.

Step 6. The validity and usefulness of the perceptual space developed in the first five steps are gauged by its ability to predict the global evaluation of each object by each individual. For this purpose, semantic differential scores of ideal objects may be collected along with the ratings of real objects in Step 1 (Johnson 1971). Using the discriminant functions, ideal points may be positioned in the perceptual space along with the group centroids for the real objects. Then, in accord with the frequently proposed distance model of preference (Green and Carmone 1969; Holbrook and Howard 1977; Huber 1975), each individual's affect toward an object may be assumed to depend on the object's distance from his/her ideal point. This hypothesis may be tested with the intra-individual regression procedure suggested by Beckwith and Lehmann (1973), preferably using adjusted R^2's to account for degrees of freedom lost in the regression analysis (Montgomery and Morrison 1971).

Justification For the Procedure

Step 1. As research intended to build perceptual spaces is generally exploratory in nature, it is often necessary to include a large number of semantic differential scales, which may later be reduced by factor analysis (Osgood et al. 1957). Since the objective is to remove affective overtones, it may seem paradoxi-

cal that several of these scales are purposely evaluative in nature. Their role, however, is to reflect affect so that, in subsequent steps, its biasing effect can be taken into account explicitly and separated from primarily perceptual phenomena.

One advantage of the proposed method is that, in contrast to many consumer studies, subjects are exposed to each stimulus object being evaluated. The familiarity gained from recent firsthand experience with each object may contribute to the reduction of perceptual distortion (Brown 1968; Burnaska and Hollmann 1974; Huber and James 1978; Wilkie et al. 1974), particularly on those perceptual attributes that are relatively objective in nature (James and Carter 1978).

Step 2. The use of several evaluative scales to construct an additive index of global evaluation increases its reliability as measured, for example, by the coefficient alpha. Though alternative procedures for creating such an index are available (e.g., principal components analysis of the evaluative scales), these are probably unnecessarily complex for the present purposes.

Step 3. The computation of residuals left after removing each semantic differential rating's linear association with this index of global evaluation provides perceptual rating scores that are free from common affective overtones as defined earlier.

Step 4. When affect-free residuals are submitted to principal components analysis, the resulting factors are based on intercorrelations between scales with the effect of global evaluation partialled out. Factor scores should therefore tend to be primarily perceptual in nature with no common distortion due to shared preferences between objects and agreement concerning scale favorability.

A key reason for this step is the need to form a reduced space for the subsequent discriminant analysis. Multiple discriminant analysis (especially if conducted in a stepwise fashion) can be unstable due to overfitting when a large number of attributes is used to form a space on a relatively small number of respondents and objects. Indeed, in the study reported later, when discriminant analysis was attempted using raw data on 93 attributes for 15 objects rated by 16 respondents, the result was so unstable that a preliminary factor analysis was required to produce a meaningful discriminant solution. In other contexts, where the number of objects or respondents is large relative to the number of attributes, this step may be skipped.

Step 5. Besides generating group centroids that are free of idiosyncratic perceptual distortion, the discriminant analysis may serve further to reduce any remaining common affective distortion in the object space. The analysis produces a space wherein the rela-

tive distance between objects is maximized while minimizing the distances of the various individuals' ratings from each centroid. It therefore tends to emphasize those attributes about which respondents agree with respect to their ratings. To the extent that responses to relatively affective or affect-laden attributes are more heterogeneous than those to primarily perceptual attributes, the resulting discriminant space tends to be affect free. This purging of affective overtones is added onto that already accomplished by the partialling out process in Step 3. As discussed at length by Huber and Holbrook (1978), such an effect does not occur when only principal components analysis is used; it therefore constitutes a key distinction between the two multivariate techniques.

Step 6. In perceptual spaces built upon semantic differential ratings with affect partialled out, the use of ideal points is more appropriate conceptually than vector models of preference. This is because strictly perceptual dimensions do not necessarily possess "more is better" properties—activity and potency being typical examples (Osgood et al. 1957). Indeed, where the "more is better" phenomenon does occur, the subject may so indicate by the relative scale positions he assigns to real and ideal objects. The extra trouble required to obtain ideal points does, therefore, seem especially well justified in the case of affect-free perceptual spaces.

AN ILLUSTRATION

The study reported here applied the general approach just described to the case of consumer aesthetics. The objects used in the study were a type not generally investigated by consumer researchers—namely, recorded performances by jazz saxophonists. The traditional neglect of such "aesthetic" products by consumer researchers should not blind us, however, to the facts that artistic offerings are highly marketable commodities (often commanding staggering prices) and that concert promoters, art dealers, and record manufacturers are becoming increasingly interested in the application of consumer research techniques to marketing the arts.

Stimulus Objects

The stimulus objects in the present study were selected carefully to represent an a priori perceptual structure. Fourteen jazz recordings of 12-bar blues by alto and tenor saxophonists accompanied only by bass, drums, and piano or guitar were chosen to control for differences between musical form, instrument, and accompaniment—none of which is directly related to the aspects of musical style of major interest. Within the alto and tenor groupings an effort was made to include three players influenced primarily by

TABLE 1

DESCRIPTION OF JAZZ SAXOPHONE RECORDINGS USED AS STIMULUS OBJECTS

Style	Type of saxophone	Performer	Key	Tempo (beats/min.)	Title of tune	Record label and number
West Coast	Tenor	Lester Young	D-flat	222	Ad Lib Blues	Verve VE-2-2502
		Stan Getz	B-flat	238	Down Beat	Verve MG V-8321
		Zoot Sims	G	262	Zoot Swings the Blues	Prestige P-24061
		Al Cohn	B-flat	203	John's Bunch	Famous Door HL-107
	Alto	Paul Desmond	E-flat	210	St. Louis Blues	Columbia C2S-826
		Lee Konitz	B-flat	108	Cork 'N' Bib	Atlantic 1258
		Art Pepper	B-flat	200	Marty's Blues	Archives of Jazz AJ510
East Coast	Alto	Charlie Parker	F	218	Jam Blues	Verve VE-2-2508
		Sonny Stitt	F	259	Au Privave	Verve MG VS-6108
		Sonny Criss	F	262	California Screamin'	Prestige 7628
		Phil Woods	E-flat	145	The Stanley Stomper	Prestige P-24065
	Tenor	Sonny Rollins	B-flat	154	Bluesnote	Blue Note BN-LA401-H2
		Dexter Gordon	B-flat	232	Wee-Dot	Steeplechase SCS-1025
		John Coltrane	F	196	Some Other Blues	Atlantic 1354

Lester Young (usually regarded as the father of the "cool" school of saxophone playing) and three influenced most conspicuously by Charlie Parker (the acknowledged originator of the "bebop" saxophone style). The first style is sometimes called "West Coast" and is characterized by coolness, intellectualism, lyricism, and restraint. The second, often referred to as "hard bop" or "East Coast," is more rugged, emotional, harmonically adventurous, and uninhibited. Two factors for which it proved impossible to control were key and tempo. These aspects of each performance, together with each recording's saxophonist, type of saxophone, stylistic point of departure, tune title, and record label, are listed in Table 1.

Three choruses from each saxophonist's solo, with pauses at the beginning and end and with fairly characteristic playing in between, were recorded on separate tape cassettes and randomly labelled from "A" to "N." These excerpts served as the stimulus objects and were played back by each subject for himself on a Panasonic Model RQ-3245 tape recorder through Superex Model ST-PRO-B headphones.

Semantic Differential Scales

Because this was an exploratory study, a large number of scales were formed—based on informal interviews with jazz listeners, a perusal of the popular jazz periodicals, and introspection on the part of the experimenters. In all, 93 bipolar adjective pairs were selected to serve as the end points of 6-position semantic differential scales of the following form:

new __:__:__:__:__:__ old.

A sample of these scales appears in Table 2. The direction and order of the 93 scales were randomized, and a separate rating sheet was created for each saxo-

phonist and labelled from "A" to "N." These rating sheets were then randomized, numbered consecutively, and paper-clipped together with a sheet representing the "ideal" saxophonist at the end.

Subjects

Sixteen subjects were recruited by placing advertisements on Columbia University bulletin boards. Because of this recruiting procedure, all subjects were students or near-students (e.g., spouses, friends, or recent graduates). However, all those included in the sample expressed a strong interest in jazz (Ferber 1977). Each was paid $5.00 for participating in the study.

Task

Each subject self-administered the listening test at his own pace. Subjects were instructed to begin with the first randomly-ordered rating sheet, find the tape cassette corresponding to the label from "A" to "N" at the top, and play the tape at least once (but more often if desired) before completing the 93 rating scales. Subjects were advised to work as rapidly as possible, consistent with accuracy, as their first reactions were of primary interest. They were forewarned that they would later be asked to rank the recordings in order of preference and that they should therefore write themselves a message in the space provided at the bottom of each page to facilitate making these subsequent preference judgments. After rating each of the 14 randomly-ordered recordings, subjects rated the "ideal" saxophonist on the same set of 93 semantic differential scales, removed the paper clip, and sorted the rating sheets into their order of preference from the most to least preferred.

The overall task involved 1,395 (15 × 93) ratings

TABLE 2

SEMANTIC DIFFERENTIAL SCALES WITH LOADINGS GREATER THAN 0.50 FOR 20 VARIMAX-ROTATED FACTORS

Factor	Positively loaded scales	Loadings	Negatively loaded scales	Loadings
Activity	Hard/soft	.624	Quiet/loud	−.621
	Firm/limp	.522	Slow/fast	−.797
	Active/passive	.751	Listless/energetic	−.779
	Exciting/restful	.726	Gentle/tough	−.619
	Busy/lazy	.809	Dragging/rushing	−.678
	Wild/confined	.569	Smooth/rough	−.511
	Cluttered/sparse	.510	Mellow/harsh	−.594
	Extroverted/introverted	.574	Sad/funny	−.520
	Tense/relaxed	.624	Calm/lively	−.760
	Piercing/breathy	.549		
	Agitated/calm	.740		
	Driving/pedestrian	.742		
	Staccato/legato	.564		
Coolness	Funky/pristine	.595		
	Emotional/intellectual	.660		
	Warm/cool	.732		
Deepness	Shallow/deep	.817		
	Empty/full	.826		
Freedom			Free/structured	−.737
Age	New/old	.746	Traditional/contemporary	−.804
	Progressive/regressive	.735	Classic/faddish	−.710
	Bop/swing	.569		
	Modern/old-fashioned	.711		
Conciseness			Concise/verbose	−.790
Heaviness	Light/heavy	.589		
	Sharp/flat	.664		
Intonation	Out-of-tune/in-tune	.697		
	Unsophisticated/sophisticated	.578		
Fidelity	Poorly/well-recorded	.803		
	Low-/high-fidelity	.842		
Conventionality	Distinctive/faceless	.775		
	Unique/conventional	.542		
Solo			Solo/ensemble	−.777
Power	Weak/strong	.623		
	Weak/powerful	.551		
Lyricism	Epic/lyrical	.454[a]	Wistful/sardonic	−.657
Awkwardness	Beautiful/ugly	.683		
	Graceful/awkward	.716		
Manliness	Feminine/masculine	.771		
Tightness	Open/pinched	.735	Tight/loose	−.550
Breathiness	Piercing/breathy	.516		
Labor	Effortless/labored	.566	Legitimate/inflected	−.554
Repetition			Imitative/original	−.702
			Repetitive/shifting	−.657
Sadness			Sad/funny	−.491[a]

[a] Less than 0.50, but included to facilitate interpretation.

from each subject and therefore tended to take at least an hour and a half, even for those subjects who moved fairly swiftly. Some spent up to two-and-a-half hours. All subjects showed considerable involvement in the task, and only a handful of incomplete scales was found. These occasional missing scores were

coded as the means on the relevant scales, computed across saxophonists for a given subject.

Computational Procedure

Semantic differential scores (S_{ijk}) were normalized within each subject according to the procedure suggested by Bass and Wilkie (1973), where the normalized score (N_{ijk}) for subject k rating saxophonist i on scale j is:

$$N_{ijk} = S_{ijk}/\sum_i S_{ijk}. \qquad (1)$$

Those scores correlated at $|r| \geq 0.70$ with the bad/good semantic scale were rescored, if necessary, in the positive affective direction and summed to construct an index of global evaluation (Step 2).

The residuals from regressing the remaining semantic differential scales on this evaluative index were computed (Step 3) and then served as input into principal components analysis (Step 4). The resulting factors with eigenvalues greater than 1.0 were varimax rotated to facilitate interpretation. Factor scores on this rotated reduced space served as independent variables in multiple discriminant analysis to predict the saxophonists' identities as the dependent category variable (Step 5).

The discriminant functions that emerged from this analysis became the dimensions for a joint space indicating the perceived position of each saxophonist and each subject's ideal point. The position (P_{ikn}) of each saxophonist i, as perceived by each subject k on each dimension n, and each subject's ideal point (I_{kn}) were computed by the following formulas:

$$P_{ikn} = A_n + \sum_m d_{mn} \cdot F_{ikm}, \qquad (2)$$

$$I_{kn} = A_n + \sum_m d_{mn} \cdot F_{Ikm}, \qquad (3)$$

where F_{ikm} and F_{Ikm} are the real and ideal saxophonists' scores for subject k on factor m, d_{mn} is the unstandardized discriminant function coefficient for factor m on function n, and A_n is the constant in discriminant function n.

Centroids $(\bar{P}_{in})$ representing the group's mean perceived position of each saxophonist i on each dimension n were given by:

$$\bar{P}_{in} = \sum_{k=1}^{16} P_{ikn}/16. \qquad (4)$$

These were used to construct a spatial map of the average perceived positions of each saxophonist and the location of each subject's ideal point. Examination of this map facilitated interpretation of the discriminant functions.

For each subject k, the distance (D^*_{ikn}) of each saxophonist i from his ideal position along dimension n was computed as follows:

$$D^*_{ikn} = |\bar{P}_{in} - I_{kn}|. \qquad (5)$$

These D^* measures along each dimension were then used to predict global evaluations in optimally-weighted city-block distance models (Step 6). Specifically, for each subject separately, global evaluation was regressed on the distance scores along each dimension (Beckwith and Lehmann 1973). To account for degrees of freedom lost in these regressions, adjusted R^2s were taken as the appropriate measures of goodness-of-fit (Montgomery and Morrison 1971).

RESULTS

Index of Global Evaluation

Eight of the semantic differential scales were correlated at $|r| \geq 0.70$ with the bad/good scale: (1) distasteful/tasty (0.72), (2) untalented/talented (0.78), (3) tasteless/tasteful (0.81), (4) unimaginative/creative (0.74), (5) dull/exciting (0.75), (6) pleasant/unpleasant (-0.71), (7) memorable/forgettable (-0.71), (8) interesting/boring (-0.73). These nine scales, with the last three reversed in direction, were summed to create an index of global evaluation. The reliability of this index was suggested by its coefficient alpha of 0.95. Its validity as a measure of affect was supported by its high average intra-individual Spearman correlation with preference ranks: mean $r = -0.80$ ($t = -24.6$, $df = 15$, $p < 0.001$).

Factors

Principal components analysis of the residuals on the remaining 84 semantic differential scales across all 16 subjects and all 15 saxophonists (including the ideal) produced 20 factors with eigenvalues greater than 1.0, accounting for 69 percent of the variance. Factor names were based on those semantic differential scales with varimax-rotated loadings greater than 0.50, as shown in Table 2.

As is often the case in the factor analysis of semantic differential scales (Osgood et al. 1957), a strong *activity* factor appeared (accounting for 22 percent of the variance), with high loadings on such scales as active/passive, busy/lazy, slow/fast, and calm/lively. Also typically, the deepness factor (accounting for five percent of the variance) seemed to represent *potency*, with high loadings on shallow/deep and empty/full. The remaining 18 factors account for 41 percent of the variance and represent a variety of perceived aspects of the recordings that might plausibly be used to describe the saxophonists' styles.

As anticipated, it appears that (with the possible exception of awkwardness, accounting for only 2 percent of the variance) the 20 factors, based on correlations with global evaluation partialled out, are *not* unequivocally favorable or unfavorable in nature. Rather some subjects might prefer players with more activity, coolness, deepness, structure, and so on,

whereas others might prefer just the reverse. Indeed, because of the manner in which they were derived, the correlation between each factor and global evaluation is, by definition, zero. Nor is any of the factors (including awkwardness) significantly correlated with preference rank. It therefore appears safe to conclude that partialling out global evaluation has effectively removed *common* affective overtones from the factor structure.[1]

Discriminant Space

Multiple discriminant analysis generated three key discriminant functions accounting for 71 percent of the variance (37 percent, 18 percent, and 16 percent respectively), with canonical correlations of 0.75, 0.63, and 0.60. The third discriminant function appeared to be a natural cutoff point as the next accounted for only eight percent of the variance, with a canonical correlation of 0.47. Examination of the standardized discriminant function coefficients shown in Table 3 suggests that the first is completely dominated by the activity factor (−0.905), with the next-highest loading only 0.191, and may therefore be regarded as an unambiguous activity dimension. Weights on the second discriminant function are much more evenly distributed, but interpretation of their signs suggests at least two key perceptual distinctions: (1) alto versus tenor (deepness, heaviness, manliness) and (2) West versus East Coast styles of saxophone playing (coolness, age, conciseness, good intonation, lyricism, tightness, and repetition as opposed to deepness, freedom, heaviness, power, awkwardness, manliness, labor, and sadness). The third function refers primarily to fidelity and age and may thus depend both on the recency of the recording and on the contemporaneity of the style of music being played.

Global evaluation is, by definition, uncorrelated with

---

[1] Another way to remove common affective overtones from the factor structure would be to perform the principal components analysis on the full matrix of 93 semantic differential scales and then to discard the factor that represents global assessment. But Myers (1965) has demonstrated that the remaining factors may be more difficult to interpret than those derived from residual scores with global evaluation partialled out. When the alternative full-matrix approach was applied to the present data, the results supported Myers' argument in that the varimax-rotated factors were weaker and somewhat less meaningful intuitively than those obtained from the residual scores. Only 16 nonevaluative factors appeared instead of 20. Seven intuitively meaningful factors—deepness, conventionality, power, lyricism, awkwardness, breathiness, and labor—disappeared from the factor structure and were either not replaced or replaced by factors that had to be interpreted on the basis of loadings substantially less than 0.50. In this instance, the distinction may be of only minor practical importance, however, since discriminant analyses of factors based on the two approaches produced perceptual maps that were virtually identical, thus suggesting that either procedure will lead to perceptual spaces that are relatively free from common distortion due to shared affective overtones.

TABLE 3

STANDARDIZED DISCRIMINANT FUNCTION COEFFICIENTS
FOR 20 FACTORS DISCRIMINATING BETWEEN
14 SAXOPHONE RECORDINGS

Factor	Function 1 Activity	Function 2 Alto/tenor West/East Coast	Function 3 Recency
Activity	−.905	.024	−.065
Coolness	.191	.197	−.235
Deepness	.076	−.251	.049
Freedom	−.008	−.144	−.259
Age	−.138	.216	−.514
Conciseness	−.132	.093	.187
Heaviness	−.094	−.421	.013
Intonation	−.134	.229	.093
Fidelity	.032	.309	.432
Conventionality	.119	−.071	−.170
Solo	−.114	−.025	.080
Power	−.027	−.201	.141
Lyricism	.135	.104	−.206
Awkwardness	−.020	−.239	.242
Manliness	−.015	−.342	.215
Tightness	.058	.301	.160
Breathiness	.042	−.024	−.315
Labor	.040	−.204	−.257
Repetition	.054	.254	.068
Sadness	.058	−.287	.023

any of these three discriminant scores. However, it is also important to note that, because of heterogeneity in preferences, global evaluation would have done a poor job of discriminating between saxophonists even if it had been included in the analysis. Indeed, whereas analysis of variance showed significant differences between the 14 real saxophonists on activity ($F = 15.498$; $df = 13$, 206; $p < 0.001$), age ($F = 3.751$, $p < 0.001$), heaviness ($F = 1.872$, $p < 0.05$), fidelity ($F = 2.832$, $p < 0.001$), and manliness ($F = 2.364$, $p < 0.01$), no such discriminating power was present in the index of global evaluation ($F = 1.246$, n.s.). This suggests that where, for some reason, the researcher has no choice but to work with attributes contaminated by affective overtones, some of this distortion may be removed by using objects toward which subjects are sufficiently heterogeneous in preference.[2]

---

[2] In the present case, a space built upon factors derived from the full set of semantic differential scales *without* partialling out the affective dimension was virtually identical to the perceptual map obtained with the recommended procedure. True, a conspicuously affective factor did emerge as the first rotated principal component. But, because of heterogeneity in preferences, scores on this factor showed no significant differences between saxophonists, and therefore failed to contribute noticeably to the discriminant analysis (see Step 5, "Justification" section). In this study, then, one could have omitted the partialling-out procedure. This is not a general result, however, and should not be counted on in other applications without first checking for preference heterogeneity (see Huber and Holbrook 1978).

FIGURE

GROUP CENTROIDS FOR SAXOPHONISTS' AND INDIVIDUAL
SUBJECTS' IDEAL POINTS ON FIRST TWO
DISCRIMINANT FUNCTIONS
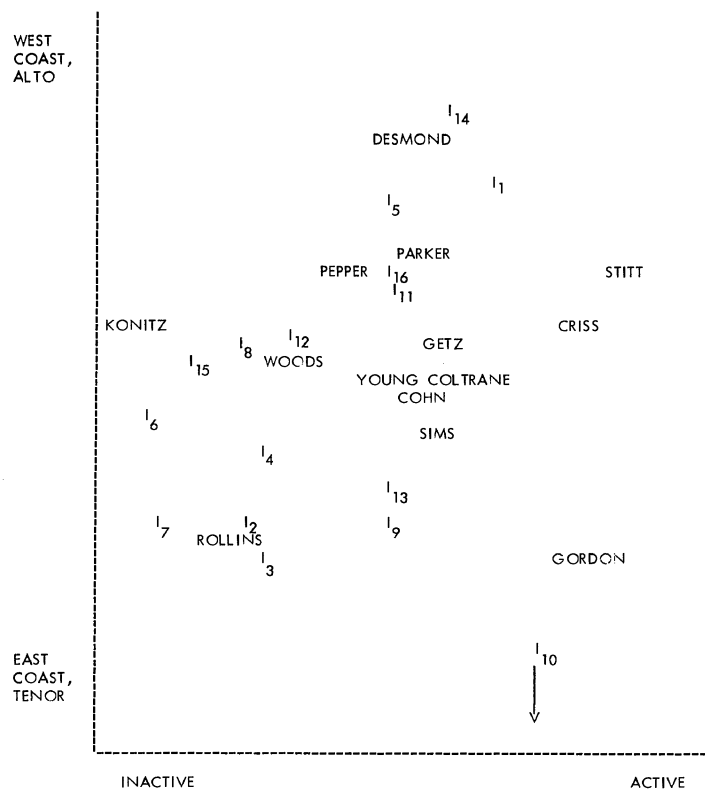
The centroids of each saxophonist on the first two discriminant functions appear in the Figure, which also shows the ideal point for each subject computed according to Equation 3. The positions of the saxophonists make good sense intuitively. Along the activity dimension, the comparatively fast performances by Stitt (259 beats per minute), Criss (262 beats), and Gordon (232 beats) are at one end of the continuum, whereas the slower tempos of Konitz (108 beats per minute), Rollins (154 beats), and Woods (145 beats) place them at the other end.

Vertically, there is an almost perfect discrimination between the altoists at the lighter end of the scale and the tenors at the heavier end. Within the regions defined by choice of instrument, there is fair, but not perfect discrimination between the West Coast players like Desmond and Pepper (altos) or Getz, Young, Cohn, and Sims (tenors) versus the East Coast hard-boppers like Parker, Stitt, Criss, and Woods (altos) or Gordon and Rollins (tenors). Among the 14 saxophonists, only Konitz (alto) and Coltrane (tenor) appear to be somewhat misplaced within their instrument categories.

On the recency dimension (not shown in the Figure), low scorers were Getz, Young, Konitz, and Sims—all relatively old, scratchy pre-1960 recordings of tradi-

tional, mainstream musical approaches. By contrast, the comparative modernists Coltrane, Rollins, and Woods—along with the especially well-recorded Cohn—all rated at the high end of the recency scale. The a priori reasonable positions of the players on the three dimensions lend considerable face validity to the interpretation of these dimensions as primarily perceptual in nature.

Distance Model of Attitude

Goodness-of-fit of the distance model of attitude was assessed by regressing global evaluation on distance scores (D^*_{ikn}) across saxophonists for each subject separately. The average fit of the optimally-weighted city-block distance model, as determined by this regression procedure, was highly significant though only moderately strong: mean adjusted $R^2 = 0.227$ ($t = 3.978, df = 15, p < 0.01$). This corresponds to an "adjusted" R of 0.47 and estimates the predictive performance of the optimal distance model after both common and idiosyncratic perceptual distortion have been removed.

One might wonder how good the fit would have been if distances (D_{ikn}) had been computed using each individual's perceived positions of the saxophonists (P_{ikn}) in Equation 5 instead of the group centroids $(\bar{P}_{ikn})$. Such D_{ikn} scores would, of course, reflect idiosyncratic perceptual distortions due to an individual's tendency to perceive more (less) preferred objects as lying closer to (farther from) his ideal point. Indeed, use of the individual's D_{ikn} scores in the optimal city-block distance model produced a mean adjusted R^2 of 0.37 ($t = 6.437, df = 15, p < 0.001$). This corresponds to an "adjusted" R of 0.61, and indicates that considerable improvement in goodness-of-fit may result from including perceptual positions that are biased by idiosyncratic perceptual distortion.

DISCUSSION

This study illustrates a general approach to modelling attitude that appears to create a clearly interpretable perceptual space relatively free from common and idiosyncratic distortion due to affective overtones. Thus, when the subject has recent firsthand experience with the objects being evaluated and when the objects are judged on a large number of semantic differential scales, including some that are intentionally affective in nature, global evaluation can be partialled out of the factor structure in a way that removes the feedback effects of preference on perceptual ratings, and results in factor scores that are relatively free of common distortion from shared affective overtones. Accordingly, when these scores are submitted to discriminant analysis, the position of group centroids in the resulting space may safely be regarded as primarily perceptual in nature.

This logic is supported in the present study, which generated a space corresponding closely to an a priori perceptual structure.[3] Moreover, there was evidence that, when individual ideal points are put into this perceptual space, idiosyncratic distances from these ideal points offer fairly good predictions of global evaluation (mean adjusted $R^2 = 0.37$). That they provide any predictive power is really rather remarkable. Both the factors and the discriminant functions themselves have been computed to be uncorrelated with global evaluation. Yet the model recognizes that different subjects have different ideal positions in the perceptual space. When these ideal points are taken into account in the determination of distance scores, significant and moderately strong predictions of evaluation result.

The apparent quality of such predictions may be elevated spuriously, however, by the presence of a second type of perceptual distortion due to idiosyncratic affective overtones. When this halo effect is removed by using centroid-based distance scores, mean adjusted R^2 falls to 0.22. The corresponding "adjusted" R of 0.47 might not seem too impressive when compared to the correlations of about 0.70 often obtained by multiattribute attitude models (Holbrook 1977). But when one considers the potentially large spurious component contributed to the latter correlations by both common and idiosyncratic perceptual distortions, the performance of the spatial model with $D_{ikn}{}^*$ scores seems almost surprisingly good. In short, it appears that, when carefully-collected belief scores are reduced to their undistorted perceptual underpinnings, those perceptions do play a measurable role in determining affect, accounting for almost half (0.22/0.49) of the variance normally explained. Thus, those who would argue that attitude research is nothing but one prolonged demonstration of the halo effect may need to revise and qualify their arguments since multiattribute models may apparently retain about half their usefulness when purged of their spurious affective overtones.

One limitation of the present application is that the extensive rating tasks, though they were cooperatively completed by the highly-involved students used in the study, might not be feasible for housewives evaluating toothpaste. One obvious solution is to remunerate such subjects in proportion to the drudgery of the task. As samples can be relatively small (assuming perceptual homogeneity), this should not overburden the resources of commercial researchers. Furthermore, it is important to keep in mind that the proposed approach is advocated only in cases where the perceptual foundations of attitude are poorly understood. Needless to say, once the number of meaningful

dimensions has been narrowed by the kind of exploratory investigation described here, subsequent research can eliminate the factor analytic stage and proceed directly with fewer semantic differential scales on larger samples.

A more fundamental limitation of the proposed approach to chaining regression, factor, and discriminant analyses is that some adjective may be a direct determinant of affect and may therefore be highly correlated with the bad/good scale or some other measure(s) of global evaluation. If such a variable were included in the index of evaluation and partialled out of the factor structure, its legitimate role as an underlying determinant of preference might be obscured.

For this reason, it can be claimed only that the map emerging from the chaining procedure is a *partial* representation of perception. There may be other aspects of perception that are so intimately linked with affect for *all* subjects that they are filtered out in the residual-computation step. Whether mushrooms are perceived as edible might be an example. It could be argued, however, that for an attribute to be thus perceived, valued, and sought by virtually everyone is the exception rather than the rule. Furthermore, where such cases exist, the researcher may rely on his own intuition to sort them out satisfactorily. By contrast, the procedure proposed here is far more useful when adjectival descriptors enter the twilight zone between perceptual reality and affective overtone, which is so important—but difficult—to clarify in most studies of consumer attitudes.

REFERENCES

Bass, Frank M., and Talarzyk, W. Wayne (1972), "An Attitude Model for the Study of Brand Preference," *Journal of Marketing Research*, 9, 93–6.

———, and Wilkie, William L. (1973), "A Comparative Analysis of Attitudinal Predictions of Brand Preference," *Journal of Marketing Research*, 10, 262–9.

Beckwith, Neil E., Kassarjian, Harold H., and Lehmann, Donald R. (1978), "Halo Effects in Marketing Research: Review and Prognosis," *Proceedings*, Association for Consumer Research, Eighth Annual Conference, 465–7.

———, and Kubilius, V. Victor (1978), "Empirical Evidence of Halo Effects in Store Image Research by Estimating True Locations," *Proceedings*, Association for Consumer Research, Eighth Annual Conference, 485–93.

———, and Lehmann, Donald R. (1973), "The Importance of Differential Weights in Multiple Attribute Models of Consumer Attitude," *Journal of Marketing Research*, 10, 141–5.

———, and Lehmann, Donald R. (1975), "The Importance of Halo Effects in Multi-Attribute Attitude Models," *Journal of Marketing Research*, 12, 265–75.

---

[3] This correspondence was investigated by Huber and Holbrook (1978), who concluded that various alternative approaches did not do as good a job as the proposed method in recovering the "objective" structure represented in Table 1.

——, and Lehmann, Donald R. (1976), "Halo Effects in Multiattribute Attitude Models: An Appraisal of Some Unresolved Issues," *Journal of Marketing Research*, 13, 418–21.

Bemmaor, Albert C., and Huber, Joel C. (1978), "Econometric Estimation of Halo Effect: Single Vs Simultaneous Equation Models," *Proceedings*, Association for Consumer Research, Eighth Annual Conference, 477–80.

Bernardin, H. John (1977), "Behavioral Expectation Scales Versus Summated Scales: A Fairer Comparison," *Journal of Applied Psychology*, 62, 422–7.

——, and Walter, C. S. (1977), "Effects of Rater Training and Diary-Keeping on Psychometric Error in Ratings," *Journal of Applied Psychology*, 62, 64–9.

Blackwell, Roger D., Engel, James F., and Talarzyk, W. Wayne (1977), *Contemporary Cases in Consumer Behavior*, Hinsdale, IL.: The Dryden Press.

Blumberg, Herbert H., De Soto, Clinton B., and Kuethe, James L. (1966), "Evaluation of Rating Scale Formats," *Personnel Psychology*, 19, 243–59.

Borman, Walter C. (1975), "Effects of Instructions to Avoid Halo Error on Reliability and Validity of Performance Evaluation Ratings," *Journal of Applied Psychology*, 60, 556–60.

—— (1977), "Consistency of Rating Accuracy and Rating Errors in the Judgment of Human Performance," *Organizational Behavior and Human Performance*, 20, 238–52.

——, and Dunnette, Marvin D. (1975), "Behavior-Based Versus Trait-Oriented Performance Ratings: An Empirical Study," *Journal of Applied Psychology*, 60, 561–5.

——, and Vallon, W. Robert (1974), "A View of What Can Happen When Behavioral Expectations Scales Are Developed in One Setting and Used in Another," *Journal of Applied Psychology*, 59, 197–201.

Brown, Eva M. (1968), "Influence of Training, Method, and Relationship on the Halo Effect," *Journal of Applied Psychology*, 52, 195–9.

Burnaska, Robert F., and Hollmann, Thomas D. (1974), "An Empirical Comparison of the Relative Effects of Rater Response Biases on Three Rating Scale Formats," *Journal of Applied Psychology*, 59, 307–12.

Campbell, D. T., and Fiske, D. W. (1959), "Convergent and Discriminant Validation by the Multi-Trait Multi-Method Matrix," *Psychological Bulletin*, 56, 81–105.

Ferber, Robert (1977), "What Is the JCR Editorial Policy on Samples and Sample Requirements?," *Newsletter*, Association for Consumer Research, 7, Nos. 4, 16.

Friedman, Barry A., and Cornelius, Edwin T. III (1976), "Effect of Rater Participation in Scale Construction on the Psychometric Characteristics of Two Rating Scale Formats," *Journal of Applied Psychology*, 61, 210–6.

Gilinsky, Alberta S. (1947), "The Influence of the Procedure of Judging on the Halo Effect," *American Psychologist*, 2, 309–10.

Green, Paul E., and Carmone, Frank J. (1969), "Multidimensional Scaling: An Introduction and Comparison of Nonmetric Unfolding Techniques," *Journal of Marketing Research*, 6, 330–41.

Guilford, J. P. (1954), *Psychometric Methods*, New York: McGraw-Hill.

Holbrook, Morris B. (1977), "Comparing Multiattribute Attitude Models by Optimal Scaling," *Journal of Consumer Research*, 4, 165–71.

——, and Howard, John A. (1977), "Frequently Purchased Nondurable Goods and Services," in *Selected Aspects of Consumer Behavior*, ed. R. Ferber, Washington, D.C.: National Science Foundation.

——, and Williams, Rebecca S. (1978), "A Test of the Correspondence Between Perceptual Spaces Based on Pairwise Similarity Judgments Collected With and Without the Inclusion of Explicit Ideal Objects," *Journal of Applied Psychology*, 63, 373–6.

Huber, Joel (1975), "Predicting Preferences on Experimental Bundles of Attributes: A Comparison of Models," *Journal of Marketing Research*, 12, 290–7.

——, and Holbrook, Morris B. (1978), "Using Attribute Ratings for Product Positioning: Some Distinctions Between Reduced Spaces," working paper, Columbia University.

——, and James, William (1978), "A Measure of Halo," *Proceedings*, Association for Consumer Research, Eighth Annual Conference, 468–73.

James, William L., and Carter, Forrest S. (1978), "Halo Effects and Location Preferences," *Proceedings*, Association for Consumer Research, Eighth Annual Conference, 474–6.

Johansson, Johny K., MacLachlan, Douglas L., and Yalch, Richard F. (1976), "Halo Effects in Multiattribute Attitude Models: Some Unresolved Issues," *Journal of Marketing Research*, 13, 414–7.

Johnson, Donald M. (1963), "Reanalysis of Experimental Halo Effects," *Journal of Applied Psychology*, 47, 46–7.

Johnson, Richard M. (1971), "Market Segmentation: A Strategic Management Tool," *Journal of Marketing Research*, 8, 13–8.

Kavanagh, Michael J., MacKinney, Arthur C., and Wolins, Leroy (1971), "Issues in Managerial Performance: Multitrait-Multimethod Analyses of Ratings," *Psychological Bulletin*, 75, 34–49.

Keaveny, Timothy J., and McGann, Anthony F. (1975), "A Comparison of Behavioral Expectation Scales and Graphic Rating Scales," *Journal of Applied Psychology*, 60, 695–703.

Kruskal, J. B., Young, F. W., and Seery, J. B. (1973), "How to Use KYST, a Very Flexible Program to Do Multidimensional Scaling and Unfolding," unpublished paper, Bell Telephone Laboratories.

Landy, Frank J., Farr, James L., Saal, Frank E., and Freytag, Walter R. (1976), "Behaviorally Anchored Scales for Rating the Performance of Police Officers," *Journal of Applied Psychology*, 61, 750–8.

——, and Guion, Robert M. (1970), "Development of Scales for the Measurement of Work Motivation," *Organizational Behavior and Human Performance*, 5, 93–103.

Latham, Gary P., Wexley, Kenneth N., and Pursell, Elliott D. (1975), "Training Managers to Minimize Rating Errors in the Observation of Behavior," *Journal of Applied Psychology*, 60, 550–5.

Mazis, Michael B., Ahtola, Olli T., and Klippel, Robert E. (1975), "A Comparison of Four Multi-Attribute Models in the Prediction of Consumer Attitudes," *Journal of Consumer Research*, 2, 38–52.

Montgomery, David B., and Morrison, Donald G. (1971), "Adjusting R^2," Marketing Science Institute Technical Report, Cambridge, MA.

Moore, William L., and James, William L. (1978), "A Study of the Amount of Halo in the Perceptions of Automobiles," Proceedings, Association for Consumer Research, Eighth Annual Conference, 481–4.

Myers, James H. (1965), "Removing Halo from Job Evaluation Factor Structure," Journal of Applied Psychology, 49, 217–21.

Nisbett, Richard E., and Wilson, Timothy D. (1977), "The Halo Effect: Evidence for Unconscious Alteration of Judgments," Journal of Personality and Social Psychology, 35, 250–6.

Osgood, Charles E., Suci, George J., and Tannenbaum, Percy H. (1957), The Measurement of Meaning, Urbana: University of Illinois Press.

Rizzo, William A., and Frank, Fredric D. (1977), "Influence of Irrelevant Cues and Alternate Forms of Graphic Rating Scales on the Halo Effect," Personnel Psychology, 30, 405–17.

Sheth, Jagdish, N., and Talarzyk, W. Wayne (1972), "Perceived Instrumentality and Value Importance As Determinants of Attitudes," Journal of Marketing Research, 9, 6–9.

Silk, Alvin J. (1969), "Preference and Perception Measures in New Product Development: An Exposition and Review," Sloan Management Review, 11, 21–37.

Smith, Robert E., and Lusch, Robert F. (1976), "How Advertising Can Position a Brand," Journal of Advertising Research, 16, No. 1, 37–43.

Stanley, Julian C. (1961), "Analysis of Unreplicated Three-Way Classifications, With Applications to Rater Bias and Trait Independence," Psychometrika, 26, 205–19.

Symonds, Percival M. (1925), "Notes on Rating," Journal of Applied Psychology, 9, 188–95.

Taylor, Erwin K., and Hastman, Roy (1956), "Relation of Format and Administration to the Characteristics of Graphic Rating Scales," Personnel Psychology, 9, 181–206.

Thorndike, Edward L. (1920), "A Consistent Error in Psychological Ratings," Journal of Applied Psychology, 4, 25–9.

Wilkie, William L., McCann, John M., and Reibstein, David J. (1974), "Halo Effects in Brand Belief Measurement: Implications for Attitude Model Development," Proceedings, Association for Consumer Research, Fourth Annual Conference, 280–90.

Willingham, Warren W., and Jones, Marshall B. (1958), "On the Identification of Halo Through Analysis of Variance," Educational and Psychological Measurement, 18, 403–7.

Zedeck, Sheldon, Kafry, Ditsa, and Jacobs, Rick (1976), "Format and Scoring Variations in Behavioral Expectation Evaluations," Organizational Behavior and Human Performance, 17, 171–84.