CHAPTER 3

Multidimensional Scaling

J. Douglas Carroll
Phipps Arabie

I. INTRODUCTION

This technique comprises a family of geometric models for representation of data in one or, more frequently, two or more dimensions and a corresponding set of methods for fitting such models to actual data. A much narrower definition would limit the term to spatial distance models for similarities, dissimilarities, or other *proximity data*. The usage we espouse includes nonspatial (e.g., such discrete geometric models as tree structures) and nondistance (e.g., scalar product or projection) models that apply to nonproximity (e.g., preference or other dominance) data as well as to proximities. As this chapter demonstrates, a large class of these nonspatial models can still be characterized as dimensional models—but with discrete rather than continuously valued dimensions.

The successful development of any multivariate technique and its incorporation in widely available statistical software inevitably lead to substantive applications over an increasingly wide range both within and among disciplines. Multidimensional scaling (MDS) is no exception, and within psychology and closely related areas we could catalog an immense variety of different applications (not all of them cause for celebration, however); several thousand are given in the annual bibliographic survey *SERVICE* (Murtagh, 1997) published by the Classification Society of North America.

Measurement, Judgment, and Decision Making

Further evidence of the vitality of developments in MDS can be found in the numbers of recent (1) books and edited volumes and (2) review chapters and articles on the topic. In the former category, we note Arce (1993); Ashby (1992); Cox and Cox (1994); de Leeuw, Heiser, Meulman, and Critchley (1986); De Soete, Feger, and Klauer (1989); Gower and Hand (1996); Green, Carmone, and Smith (1989); Okada and Imaizumi (1994); and Van Cutsem (1994). The conference proceedings volumes are too numerous even to cite, and the monograph series of DSWO Press at the University of Leiden has many noteworthy contributions. Concerning review chapters and articles, the subareas of psychology recently targeted include counseling (Fitzgerald & Hubert, 1987), developmental (Miller, 1987), educational (Weinberg & Carroll, 1992), experimental (L. E. Jones & Koehly, 1993; Luce & Krumhansl, 1988), and cognitive (Nosofsky, 1992; Shoben & Ross, 1987). Multivariate statistical textbooks also continue to pay due attention to MDS (e.g., Krzanowski & Marriott, 1994, chap. 5). Iverson and Luce's chapter in this volume focuses on a complementary aspect of measurement in psychology and the behavioral sciences, measurement (primarily, but not exclusively, unidimensional) based on subjects' orderings of stimuli, whereas we are concerned with measurement (primarily, but not exclusively, multidimensional, or multiattribute) based on proximity data on pairs of stimuli or other entities.

In this chapter we focus almost exclusively on that substantive area where we see the strongest bonds to MDS and its underpinnings and that seems most likely to spur new methodological developments in MDS, namely that answering fundamental questions about the psychological representation of structure underlying perception and judgment, especially in terms of similarities and dissimilarities. From its inception (Shepard, 1962a, 1962b), nonmetric MDS has been used to provide visualizable depictions of such structure, but current research focuses on much more incisive queries. Question 1 is whether any particular stimulus domain is better fitted by a discrete than by a continuous (usually) spatial model. The latter possibility gives rise to Question 2, which concerns the nature of the metric of the multidimensional stimulus space (often assumed to be either Euclidean or city-block, as defined later).

Question 1, of course, is at the heart of such controversies in experimental psychology as categorical perception (Tartter, in press, chap. 7) and neural quantum theory (Stevens, 1972). With the advent of increasingly general models (discussed later) for discrete structure and associated algorithms for fitting them, it has become possible in some cases to run empirical comparisons of selected discrete versus spatial models for given data sets (cf. Carroll, 1976; De Soete & Carroll, 1996). Pruzansky, Tversky, and Carroll (1982) compared data from several stimulus domains and concluded that: "In general, colors, sounds and factorial structures were better represented by a plane [i.e., a two-dimensional MDS solution], whereas conceptual stimuli from various semantic fields were better modelled by a[n] additive] tree" (p. 17).

Within the literature of experimental psychology, Question 2 effectively begins with Attneave's (1950, p. 521) reflections on "the exceedingly precarious assumption that psychological space is Euclidean" (1950, p. 521). He instead argued: "The psychological implication is that there is a unique coordinate system in psychological space, upon which 'distances' between stimuli are strictly dependent [as opposed to rotation invariant]; and thus our choice of axes is to be dictated, not by linguistic expediency, but by psychological fact." Moreover, Attneave (1950, p. 555) began the tradition of distinguishing between integral and analyzable stimulus domains with his sharp contrast between Euclidean and city-block metrics: "Perhaps the most significant psychological difference between these two hypotheses is that the former assumes one frame of reference to be as good as any other, whereas the latter implies a unique set of psychological axes." For the development of theoretical positions on this distinction between integral and analyzable stimuli, see Shepard's (1991) and other chapters in Lockhead and Pomerantz's (1991) Festschrift for W. R. Garner. For a review of theoretical and algorithmic approaches to city-block spaces, see Arabie (1991).

Since the mid-1980s, the most innovative and significant results pertaining to Question 2 have come from Nosofsky (e.g., 1992) and from Shepard (1987, 1988). In the latter papers, Shepard returned to his earlier interest in stimulus generalization to formulate and derive a universal law of generalization based on the distinctions between analyzable and integral stimuli and between the Euclidean and city-block metrics.

Reviewing recent work on "models for predicting a variety of performances, including generalization, identification, categorization, recognition, same-different accuracy and reaction time, and similarity judgment," Nosofsky (1992, p. 40) noted that "The MDS-based similarity representation is a fundamental component of these models." Additionally (Nosofsky, 1992, p. 34), "*The role of MDS in developing these theoretical relations is critical* [italics added]." The literature on Question 2 has become quite extensive; for example, see chapters in Ashby (1992) and work by Ennis and his collaborators (e.g., Ennis, Palen, & Mullen, 1988).

To explain how MDS can be used to address Questions 1 and 2, we must immediately make some distinctions among types of data matrices, and we do so by summarizing a lengthier taxonomy found in Carroll and Arabie (1980, pp. 610–611). Consider two matrices: one with n rows and the same number of columns (with entries depicting direct judgments of pairwise similarities for all distinct pairs of the n stimuli) and the other matrix with n rows of stimuli and K columns of attributes of the stimuli. Although both matrices have two *ways* (namely, rows and columns), the former is said to have one mode because both its ways correspond to the same set of entities (i.e., the n stimuli). But the matrix of stimuli by their attributes has two disjoint sets (and thus two *modes;* Tucker, 1964) of entities corresponding to the ways. For a one-mode two-way matrix, an additional consideration is whether conjugate off-diagonal entries are always equal, in which case the matrix is symmetric; otherwise it is nonsymmetric.

Another important distinction concerns whether the data are conditional (i.e., noncomparable) between rows/columns or among matrices. Row conditional data arise most commonly when a subject is presented with each of n stimuli in turn and asked to rank the remaining $n - 1$ according to their similarity to the standard. If the ranks are entered as a row/column for each successive standard stimulus in a two-way one-mode matrix, the entries are comparable within but not between rows/columns, and such data are therefore called row/column conditional (Coombs, 1964). If the data are a collection of I one-mode, two-way matrices, all $n \times n$ for the same set of n stimuli, a more general question is whether numerical entries are comparable among the matrices. If not, such three-way data are said to be matrix conditional (Takane, Young, & de Leeuw, 1977).

It is not our intention to dwell on traditional methods of collecting data for multidimensional scaling, given the excellent summaries already available (e.g., Kruskal & Wish, 1978; Coxon, 1982, chap. 2; Rosenberg, 1982, for the method of sorting; L. E. Jones & Koehly, 1993, pp. 104–108). An important distinction offered by Shepard (1972) is whether the input data are the result of direct judgments (e.g., from subjects' judging all distinct pairs of stimuli—say, on a 9-point scale of similarity/dissimilarity—or confusions data) or of indirect or profile data, as result when the data at hand are two-mode, but the model to be fitted requires one-mode data. In such cases, the user typically preprocesses the data by computing an indirect measure of proximity (e.g., squared Euclidean distances) between all pairs of rows or columns to obtain a one-mode matrix of pairwise similarities/dissimilarities. Although Shepard's (1962a, 1962b) original development of nonmetric MDS greatly emphasized applications to one-mode two-way direct similarities, applications of various MDS models to indirect or profile data are quite common.

A noteworthy development of recent years is that of models and associated algorithms for the direct analysis of types of data not previously amenable to MDS without preprocessing: free recall sequences (Shiina, 1986); row conditional rank-order data (Takane & Carroll, 1982): similarity/dissimilarity judgments based on triples (Daws, 1993, 1996; Joly & Le Calvé, 1995; Pan & Harris, 1991) or even n-tuples of stimuli (T. F. Cox, M. A. A. Cox, & Branco, 1991); triadic comparisons (Takane, 1982); and sorting data (Takane, 1981, 1982).

Carroll and Arabie (1980) organized their *Annual Review* chapter on

MDS around the typology of ways and modes for data and for corresponding algorithms. Although these distinctions remain crucial in considering types of data, the typology is now less clear-cut for algorithms. As we predicted in that chapter (p. 638), there has been intensive development of three-way algorithms, and the two-way special cases are often by-products. Thus, in our present coverage, the two-way algorithms and models are mentioned only as they are subsumed in the more general three-way approaches.

II. ONE-MODE TWO-WAY DATA

The inventor of the modern approach to nonmetric MDS (Shepard, 1962a, 1962b) began by considering a single one-mode two-way matrix, typically some form of similarity, dissimilarity, or other proximity data (sometimes also referred to as "relational" data). Another type of ostensibly dyadic data is so-called paired comparisons data depicting preferences or other forms of dominance relations on members of pairs of stimuli. However, such data are seldom utilized in multidimensional (as opposed to unidimensional) scaling. We do not cover paired comparisons data in this chapter because we view such data *not* as dyadic but as replicated monadic data (having $n - 2$ missing data values within each replication); see Carroll (1980) for an overview.

III. SPATIAL DISTANCE MODELS (FOR ONE-MODE TWO-WAY DATA)

The most widely used MDS procedures are based on geometric *spatial distance* models in which the data are assumed to relate in a simple and well-defined manner to recovered *distances* in an underlying *spatial* representation. If the data are interval scale, the function relating the data to distances is generally assumed to be inhomogeneously linear—that is, linear with an additive constant as well as a slope coefficient. Data of interval or stronger (ratio, positive ratio, or absolute) scale are called *metric*, and the corresponding models and analyses are collectively called *metric MDS*. In the case of ordinal data, the functional relationship is generally assumed to be monotonic—either monotonic nonincreasing (in the case of similarities) or monotonic nondecreasing (for dissimilarities). Ordinal data are often called *nonmetric* data, and the corresponding MDS models and analyses are also referred to as *nonmetric MDS*. The distinction between metric and nonmetric approaches is based on the presence or absence of metric properties in the data (not in the solution, which almost always has metric properties; Holman, 1978, is an exception).

Following Kruskal's (1964b, 1965) innovative work in monotone regression (as the basic engine for fitting any of the ordinal models considered in this review), first devised by Ayer, Brunk, Ewing, Reid, and Silverman (1955), there has been much activity in this area of statistics. In addition to Shepard's (1962a, 1962b) early approach and Guttman's (1968) later approach based on the rank image principle, alternative and related methods have been proposed by R. M. Johnson (1975), Ramsay (1977a), Srinivasan (1975), de Leeuw (1977b), de Leeuw and Heiser (1977, 1980, in developing their SMACOF algorithm, considered later), and Heiser (1988, 1991). McDonald (1976) provided a provocative comparison between the approaches of Kruskal (1964b) and Guttman (1968), and the two methods are subsumed as special cases of Young's (1975) general formulation. More recently, Winsberg and Ramsay (1980, 1983), Ramsay (1988), Winsberg and Carroll (1989a, 1989b), and Carroll and Winsberg (1986, 1995) have introduced the use of monotone splines as an alternative to the totally general monotone functions introduced by Kruskal, while other authors (e.g., Heiser, 1989b) have proposed using other not completely general monotonic functions, which, like monotone splines, can be constrained to be continuous and to have continuous derivatives, if desired. (Carroll and Winsberg, 1986, 1995, and Winsberg and Carroll, 1989a, 1989b, have used monotone splines in a somewhat unique manner—predicting data as monotone function(s) of distances, rather than vice versa as is typically the case in fully nonmetric approaches. As discussed later, these authors argue that this quasi-nonmetric approach avoids degeneracies that occur with fully nonmetric approaches.)

A. Unconstrained Symmetric Distance Models (for One-Mode Two-Way Data)

Although one of the more intensely developed areas in recent years has been the treatment of nonsymmetric data (discussed in detail later), most of the extant data relevant to MDS are symmetric, owing in part to the previous lack of models allowing for nonsymmetric data and the ongoing absence of readily available software for fitting such models. Therefore, we first consider recent developments in the scaling of symmetric data, that is, where the proximity of j to k is assumed identical to that obtained when the stimuli are considered in the reverse order.

The most widely assumed metric in MDS is the Euclidean, in which the distance between two points j and k is defined as

$$d_{jk} = \left[\sum_{r=1}^{R} (x_{jr} - x_{kr})^2 \right]^{1/2},$$

where x_{jr} and x_{kr} are the rth coordinates of points j and k, respectively, in an R-dimensional spatial representation. Virtually all two-way MDS procedures use either the Euclidean metric or the Minkowski ρ (or L_ρ) metric, which defines distances as

$$d_{jk} = \left[\sum_{r=1}^{R} |x_{jr} - x_{kr}|^{\rho} \right]^{1/\rho} \qquad (\rho \geq 1) \tag{1}$$

and so includes Euclidean distance as a special case in which $\rho = 2$. (Because a variable that later will appear extensively in this chapter will be labeled "p," we are using the nontraditional ρ for Minkowski's exponent.)

B. Applications and Theoretical Investigations of the Euclidean and Minkowski ρ Metrics (for One-Mode Two-Way Symmetric Data)

1. Seriation

A psychologist who harbors proximity data suspected of being *uni*dimensional is caught between a Scylla of substantive tradition and a Charybdis of deficient software. Concretely, the custom in experimental psychology has been to discount unidimensionality and seek only higher-dimensional solutions. For example, Levelt, van de Geer, and Plomp (1966) developed an elaborate two-dimensional substantive interpretation of data later shown to be unidimensional by Shepard (1974) and Hubert and Arabie (1989, pp. 308–310). Similarly, Rodieck (1977) undermined a multidimensional theory of color vision proposed by Tansley and Boynton (1976, 1977).

But a data analyst willing to counter the tradition of overfitting immediately encountered a suspicion that gradient-based algorithms for nonmetric MDS could not reliably yield solutions faithful to an underlying unidimensional structure in a proximities matrix (cf. Shepard, 1974). De Leeuw and Heiser (1977) pointed out that this is in fact a *discrete* problem of analysis masquerading as a continuous one. Hubert and Arabie (1986, 1988) demonstrated analytically why gradient methods fail in the unidimensional case and then provided an alternative algorithm based on dynamic programming, guaranteed to find the globally optimal unidimensional solution. Pliner (1996) has provided a different algorithm that can handle much larger analyses. Also see related work by Hubert and Arabie (1994, 1995a); Hubert, Arabie, and Meulman (1997); Mirkin (1996); and Mirkin and Muchnik (1996, p. 319).

2. Algorithms

Kruskal's (1964a, 1964b) option to allow the user to specify $\rho \neq 2.0$ in Eq. (1) ostensibly made it much easier for experimenters to decide which Minkowski metric was most suitable for their data. But evidence (Arabie, 1973) and hearsay soon accumulated that, at least in the city-block case (where $\rho = 1$), the algorithm found suboptimal solutions, and there was a suspicion (e.g., Shepard, 1974) that the same conclusion was true for unidimensional solutions (no matter what value of ρ was used, because all are mathematically equivalent in the case of one dimension).

As noted earlier, de Leeuw and Heiser (1977) made the crucial observation that the unidimensional case of gradient-based two-way MDS is in fact a discrete problem, and Hubert and Arabie (1986) provided an appropriately discrete algorithm to solve it. Hubert and Arabie (1988) then analytically demonstrated that the same discreteness underlies the problem of city-block scaling in two dimensions and conjectured that the result is actually much more general. Hubert, Arabie, and Hesson-Mcinnis (1992) provided a combinatorial nonmetric algorithm for city-block scaling in two and three dimensions (for the two-way case) and demonstrated the highly inferior fits typically obtained when traditional gradient methods were used instead on the same data sets. Nonetheless, such misguided and clearly suboptimal analyses continue to appear in the experimental psychology literature (e.g, Ashby, Maddox, & Lee, 1994). Using a majorization technique, Heiser (1989a) provided a metric three-way city-block MDS algorithm. Neither the approach of Hubert and colleagues (1992) nor that of Heiser can guarantee a global optimum, but they generally do much better than their gradient counterparts.

In MDS the city-block metric has received more attention during the past two decades than any other non-Euclidean Minkowski metric (see Arabie, 1991, for a review), but more general algorithmic approaches are also available. For example, Okada and Imaizumi (1980b) provided a three-way nonmetric generalization of the INDSCAL model, as in Eq. (5) (where a monotone function is fitted to the right side of that equation). Groenen (1993; also see Groenen, Mathar, & Heiser, 1995) has extended the majorization approach for $1 \leq \rho \leq 2$ in Eq. (1). His impressive results have usually been limited to two-way metric MDS but appear to have considerably greater generality. There have been some attempts at fitting even more general non-Euclidean metrics such as Riemannian metrics (see the review in Carroll & Arabie, 1980, pp. 618–619), but none have demonstrated any lasting impact on the field. Although Indow (1983, pp. 234–235) demonstrated, with great difficulty, that a Riemannian metric with constant curvature fits certain visual data slightly better than a Euclidean metric, Indow concluded that the increase in goodness of fit was not sufficient to justify the effort involved and that, in practice, Euclidean representations accounted exceedingly well for the data he and his colleagues were considering. In later work, however, Indow (1995; see also Suppes, Krantz, Luce, & Tversky, 1989, pp. 131–153, for discussion) has shown that careful scrutiny of the geometric structure of these visual stimuli within different planes of a three-dimensional representation reveals that the curvature is dependent on the specific plane being considered. This discovery suggests that a more general Riemannian metric with nonconstant curvature may provide an even more appropriate representation of the geometry of visual space.

3. Algebraic and Geometric Foundations of MDS in Non-Euclidean Spaces

Confronted with the counterintuitive nature of non-Euclidean and/or high-dimensional spaces, psychologists have regularly culled (and occasionally contributed to) the vast mathematical literature on the topic, seeking results relevant to data analyses in such spaces (see, for example, Carroll & Wish, 1974a; Critchley & Fichet, 1994; de Leeuw & Heiser, 1982; Suppes et al., 1989, chaps. 12–14). Linkages to that literature are impeded by its typical and unrealistic assumptions of (1) a large or even infinite number of stimuli, (2) error-free data, and (3) indifference toward substantively insupportable high dimensionalities. The axiomatic literature in psychology does not always treat these problems satisfactorily, because it postulates systems requiring errorless measurement structures that in turn entail an infinite number of (actual or potential) stimuli. For example, testing of the axiom of segmental additivity for geometric representations of stimuli would be exceedingly difficult in a practical situation in which only a finite number of stimuli are available, and the proximity data are subject to measurement or other error of various types, because, in principle, one has to demonstrate that an intermediate stimulus exists precisely *between* each pair of stimuli so that the distances sum along the implicit line connecting the three. (As a further complication, these distances may be only monotonically related to the true proximities, whereas observed proximities are, at best, measured subject to measurement or experimental error.) Given a finite sample of "noisy" stimuli, it is highly unlikely that, even under the best of circumstances (e.g., errorless data entailing distances measured on a ratio scale), one would find a requisite third stimulus lying precisely, or even approximately, between each pair of stimuli. This instance is but one extreme case illustrating the general difficulty of testing scientific models, whether geometric or otherwise, with finite samples of data subject to measurement or experimental error. For example, it would be equally difficult, in principle, to test the hypothesis that noisy proximity data on a finite sample of stimuli are appropriately modeled via a Euclidean (or city-block, or other metric) spatial model in a specified number of dimensions. In practice, we are often forced to rely on the principle of parsimony (or "Occam's razor"), that is, to choose, among a large set of plausible models for such a set of data, the most parsimonious model, which appears to account adequately for a major portion of the variance (or other measure of variation) in the empirical proximity data. This approach hardly qualifies as a rigorous scientific test of such a geometric model; rather it is more appropriately characterized as a practical statistical rule of thumb for choosing the best among a large family of plausible models. The axiomatic approach, as exemplified by Suppes and colleagues (1989), focuses more on the precise testing of a very specific scientific model and constitutes an ideal toward which researchers in multidimensional scaling and other measurement models for the analysis of proximity data can, at the moment, only aspire. We hope that a stronger nexus can be formed between the axiomatic and the empirical camps in future work on such measurement models, effecting a compromise that allows development of practical measurement models for real-world data analysis in the psychological and other behavioral sciences while, at the same time, approaching more closely the ideal of testing such models with a sufficiently well-defined rigor.

IV. MODELS AND METHODS FOR PROXIMITY DATA: REPRESENTING INDIVIDUAL DIFFERENCES IN PERCEPTION AND COGNITION

The kind of differential attention to, or differential salience of, dimensions observed by Shepard (1964) illustrates a very important and pervasive source not only of intraindividual variation but also of interindividual differences in perception. Although people seem to perceive the world using nearly the same dimensions or perceptual variables, they evidently differ enormously with respect to the relative importance (perceptually, cognitively, or behaviorally) of these dimensions.

These differences in sensitivity or attention presumably result in part from genetic differences (for example, differences between color-blind and color-normal individuals) and in part from the individual's particular developmental history (witness the well-known but possibly exaggerated example of the Eskimos' presumably supersensitive perception of varieties, textures, and colors of snow and ice). Although some attentional shifts might result simply from instructional or contextual factors, studies by Cliff, Pennell, and Young (1966) have indicated that it is not so easy to manipulate saliences of dimensions. If a more behavioral measure of proximity were used, for example, one based on confusions in identification learning, the differential weighting could result at least in part from purely behavioral (as opposed to sensory or central) processes, such as differential gradients of response generalization. Nosofsky (1992) and Shepard (1987) have posited mechanisms underlying such individual differences.

A. Differential Attention or Salience of Dimensions: The INDSCAL Model

The INDSCAL (for *IN*dividual *D*ifferences *SCAL*ing) model (Carroll & Chang, 1970; Carroll, 1972; Carroll & Wish, 1974a, 1974b; Wish & Carroll, 1974; Arabie, Carroll, & DeSarbo, 1987) explicitly incorporates this notion of individual differences in the weights, or perceptual importances, of dimensions. The central assumption of the model is the definition of distances for different individuals. As with ordinary, or *two*-way, scaling, these recovered distances are assumed to relate in some simple way—for example, linearly or monotonically—to the input similarities or other proximities. INDSCAL, however, assumes a different set of distances for each subject. The distance between stimuli j and k for subject i, $d_{jk}^{(i)}$, is related to the dimensions of a group (or common) *stimulus* space by the equation

$$d_{jk}^{(i)} = \left[\sum_{r=1}^{R} w_{ir}(x_{jr} - x_{kr})^2 \right]^{1/2}, \tag{2}$$

where R is the dimensionality of the stimulus space, x_{jr} is the coordinate of stimulus j on the rth dimension of the group stimulus space, and w_{ir} is the weight (indicating salience or perceptual importance) of the rth dimension for the ith subject. This equation is simply a weighted generalization of the Euclidean distance formula.

Another way of expressing the same model is provided by the following equations. We first define coordinates of what might be called a "private perceptual space" for subject i by the equation

$$y_{jr}^{(i)} \equiv (w_{ir}^{1/2}) x_{jr} \tag{3}$$

and then calculate ordinary Euclidean distances according to these idiosyncratic or private spaces, as defined in

$$d_{jk}^{(i)} \equiv \left[\sum_{r=1}^{R} (y_{jr}^{(i)} - y_{kr}^{(i)})^2 \right]^{1/2} = \left[\sum_{r=1}^{R} w_{ir}(x_{jr} - x_{kr})^2 \right]^{1/2}. \tag{4}$$

[The expression on the right was derived by substituting the definition of $y_{jr}^{(i)}$ in Eq. (3) into the middle expression in Eq. (4), defining $d_{jk}^{(i)}$.] Thus the weighted distance formulation is equivalent to one in which each dimension is simply *rescaled* by the square root of the corresponding weight. This rescaling can be regarded as equivalent to turning the "gain" up or down, thus relatively increasing or decreasing the sensitivity of the total system to changes along the various dimensions.[1]

[1] Tucker and Messick's (1963) "points of view" model, which assumes that subjects form several subgroups, each of which has its own private space, or point of view, can be incorporated within the scope of INDSCAL. At the extreme, the group stimulus space includes the union of all dimensions represented in any of the points of view, and an individual would have positive weights for all dimensions corresponding to the point of view with which he or she is identified and zero weights on all dimensions from each of the other points of view. For an updated treatment of points of view, see Meulman and Verboon (1993).

The input data for INDSCAL, as with other methods of three-way MDS, constitute a matrix of proximity (or antiproximity) data, the general entry of which is $\delta_{jk}^{(i)}$, the dissimilarity (antiproximity) of stimuli j and k for subject i. If there are n stimuli and I subjects, this three-way matrix will be $n \times n \times I$. The ith two-way "slice" through the third way of the matrix results in an ordinary two-way $n \times n$ matrix of dissimilarities for the ith subject. The output in the case of INDSCAL (although not necessarily for other three-way scaling methods) consists of two matrices. The first is an $n \times R$ matrix, $\mathbf{X} \equiv (x_{jr})$ of stimulus coordinates, the second an $I \times R$ matrix $\mathbf{W} \equiv (w_{ir})$ of subject weights. The input and output arrays for INDSCAL are illustrated in Figure 1. The coordinates described in the two matrices **X** and **W** can be plotted to produce two disjoint spaces, both with dimensionality R, and which we have called, respectively, the group stimulus space and the subject space. These are illustrated in Figure 2 for a purely hypothetical data set, as are two of these subjects' idiosyncratic or private perceptual spaces. Geometrically they are derived by stretching or shrinking each dimension by applying a rescaling factor to the rth dimension, proportional to $(w_{ir})^{1/2}$. The rth weight, w_{ir}, for subject i can be derived from the subject space by simply projecting subject i's point onto the rth coordinate axis.

Quite different patterns of similarity/dissimilarity judgments are predicted in Figure 2 for Subjects 2, 3, and 4. Subject 3 (who weights the dimensions equally and so would have a private space that looks just like the group stimulus space) presumably judges Stimulus A to be equally similar to B and D, because these two distances are equal in that subject's private space. In contrast, Subject 2 would judge Stimulus A to be more similar to D than to B (because A is closer to D), and Subject 4 would judge Stimulus A to be more similar to B than to D. There would, of course, be many other differences in the judgments of these three subjects, even though all three are basing their judgments on exactly the same dimensions.

Subjects 1 and 5, who are both one-dimensional, represent two extreme cases in the sense that each gives nonzero weight to only one of the two dimensions. Geometrically it is as though (if these were the only dimensions and the model fitted the data perfectly) Subject 1 has simply projected the stimulus points onto the Dimension 1 axis so that Stimuli A, D, and G, for example, project into the same point and so are seen by this subject as identical. Subject 5 exhibits the opposite pattern and presumably attends only to Dimension 2; this subject would see Stimuli A, B, and C as identical. Thus, as a special case, some subjects can have private perceptual spaces of lower dimensionality than that of the group stimulus space.

Distance from the origin is also meaningful in this subject space. Subjects who are on the same ray issuing from the origin but at different distances from it would have the same pattern of distances and therefore of *predicted* similarities/dissimilarities. They would have the same private space, in fact,

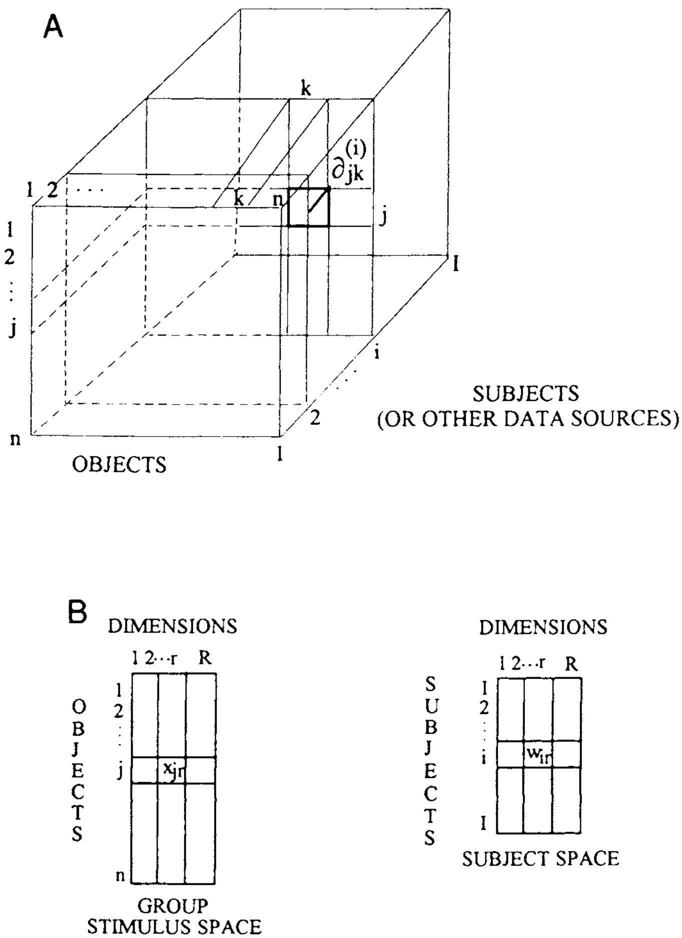

FIGURE 1 A schematic representation of input for (A) and output from (B) INDSCAL. Input consists of $I(\geq 2)$ $n \times n$ square symmetric data matrices (or half-matrices) one for each of I subjects (or other data sources), $d_{jk}^{(i)}$ is the dissimilarity of stimuli (or other objects) j and k for subject (or other data source) i. This set of I square matrices can be thought of as defining the rectangular solid, or three-way array, of data depicted at top in the figure. (This is the form of the input for other three-way scaling methods also.) The output from INDSCAL consists of two matrices, an $n \times R$ matrix of coordinates of the n stimuli (objects) on R coordinate axes (or dimensions) and an $I \times R$ matrix of weights of I subjects for the R dimensions. These matrices define coordinates of the *group stimulus space* and the *subject space*, respectively. Both of them can be plotted graphically, as in Figure 2, and a private space for each subject can be constructed, as shown there, by applying the square roots of the subject weights to the stimulus dimensions, as in Equation 3. Note: "Objects" need not be "stimuli." "Subjects" may come from other data sources.

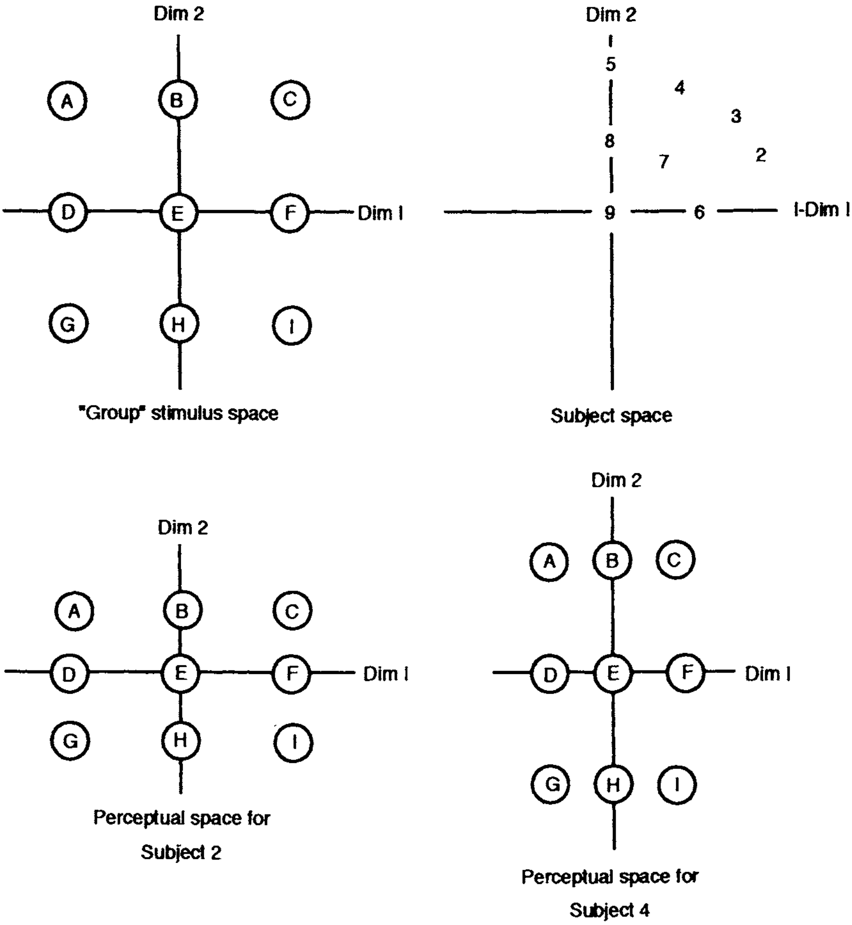

FIGURE 2 Illustration of the Carroll-Chang INDSCAL model for individual differences in multidimensional scaling. Weights (plotted in subject space) are applied to group stimulus space to produce individual perceptual spaces for Subjects 2 and 4, shown at the bottom of the figure. (For purposes of illustration, the dimensions are multiplied by the weights themselves, rather than by their square roots as is more technically correct.)

except for an overall scale factor. The main difference between such subjects is that this same private space and pattern of predicted judgments account for less of the variance in the (scalar products computed from the) data for subjects who are closer to the origin. Thus, although Subjects 3 and 7 in Figure 2 would have the same private space (the one corresponding to the group stimulus space), these two dimensions would account for more variance in the (hypothetical) matrix of Subject 3 than of Subject 7. Subject 9, being precisely at the origin (indicating zero weight on both dimensions), would be completely out of this space; that is, none of that subject's data could be accounted for by these two dimensions. The residual variance may be accounted for by other dimensions not extracted in the present analysis or simply by unreliability, or error variance, in the particular subject's responses.

The square of the distance from the origin is closely analogous to the concept of communality in factor analysis. In fact, the square of that distance is approximately proportional to variance accounted for. Although only an approximation, it is generally a good one and is perfect if the coordinate values on dimensions are uncorrelated. The cosine of the angle between subject points (treated as vectors issuing from the origin) approximately equals the correlation between distances (or, more properly, between scalar products) in their private perceptual spaces. Distances between these points are also meaningful—they approximate profile distances between reconstructed distances (or, again more properly, scalar products) from the respective private perceptual spaces in which the overall scale *is* included. We therefore reject arguments made by Takane et al. (1977), MacCallum (1977), and others that lengths (or distances from the origin) of these subject weight vectors are not meaningful. We believe the lengths (as well as directions) of these subject vectors *are* meaningful and interpretable, *even when the data are matrix conditional rather than unconditional;* in the latter case, Takane et al. (1977), MacCallum (1977), and others have argued these lengths have *no* meaning; thus those authors normalize subject weight vectors to *unit* lengths, contrary to the practice in the INDSCAL/SINDSCAL method of fitting the INDSCAL model. The lengths, in fact, often contain information that is quite critical in distinguishing among well-defined groups of subjects. Wish and Carroll (1974) presented one very good example, entailing perception of the rhythm and accent of English words or phrases by various groups of subjects. Most compelling, in this respect, is the fact that native and nonnative speakers of English were distinguished most clearly by the subject vectors—those for the former group having systematically greater length (terminating farther from the origin) than those for the latter group—implying that *all* dimensions characterizing the rhythm and accent (or stress patterns) of English words were much more salient to native than to nonnative speakers of English.[2]

[2] In statistical terms, the small set of "common" dimensions in the group space accounted for more variance in scalar products computed from the data of the native English speakers—the square of the length of the subject vector approximating the proportion of variance accounted for—whereas the nonnative English speakers apparently were largely accounted for by other variables not emerging from this analysis, such as unique linguistic dimensions (which might emerge if higher dimensional solutions were sought) more appropriate to their individual native languages or greater systematic or random errors stemming from an imperfect assimilation of English stress patterns.

One of the more important aspects of INDSCAL is the fact that its dimensions are unique, that is, not subject to the rotational indeterminacy characteristic of most two-way MDS procedures involving the Euclidean metric. INDSCAL recovered dimensions are generally defined uniquely up to what is called an "extended permutation," defined later. In the psychological model, the dimensions are supposed to correspond to fundamental perceptual or other processes whose strengths, sensitivities, or importances vary among individuals. Mathematically, the family of transformations induced by allowing differential weighting (which corresponds geometrically to stretching or compressing the space in directions parallel to coordinate axes) will differ for the various orientations of coordinate axes—that is, the family of admissible transformations is *not* rotationally invariant, as can be seen graphically by considering what kinds of private spaces might be generated in the case illustrated in Figure 2 if one imagines that the coordinate system of the group stimulus space were rotated, say, 45°. Instead of the square lattice transforming into various rectangular lattices, it would transform into various rhombuses, or diamond-shaped lattices. Rotating the coordinate system by something other than 45° would generate other families of parallelograms, generally a unique family for each different angle of rotation. These families are genuinely different, because they allow different admissible *sets* of distances among the objects or stimuli. Statistically speaking, a rotation (not corresponding to a reflection, permutation, or extended permutation) of the axes generally degrades the solution in the sense that the variance accounted for in fitting the model decreases after such a rotation, even if optimal weights are recomputed for the rotated coordinate system.

This dimensional uniqueness property is important because it obviates the need, in most cases, to rotate the coordinate system to find an interpretable solution. If one adopts the psychological model underlying INDSCAL, then these statistically unique dimensions should be psychologically unique as well. Indeed, practical experience has shown that the dimensions obtained directly from INDSCAL are usually interpretable without rotation (even when there is little reason to believe the underlying model's assumptions). Kruskal (1976) has provided a formal proof of this uniqueness property of INDSCAL (and of a wider class of three-way models of which it is a special case). Technically, the INDSCAL stimulus space is identified, under very general conditions, up to a permutation and reflection of coordinate axes, followed by a rescaling of all dimensions via a diagonal scaling matrix (with scale factors that may be either positive or negative). The rescaling transformation is generally resolved via the usual INDSCAL normalization convention, in which stimulus dimensions are scaled so as to have unit sum of squared (and zero mean) coordinate values; this way only the *signs* of the scale factors are nonidentified. In practice INDSCAL dimensions are identified up to a permutation and possible reflection of axes—what we call an *extended* *permutation*. In fact, even the permutation indeterminacy generally is resolved by ordering axes based on a variance accounted for (averaged over all subjects) criterion.

Space limitations preclude us from giving substantive illustrations of fitting the INDSCAL model (or any of the others covered in this chapter). Two protracted analyses are given in Arabie et al. (1987, pp. 12–16, 25–33).

Because of the particular normalization conventions used in the "standard" formulation described earlier, distances in the group stimulus space are not immediately interpretable but must instead be compared to the interstimulus distances of a hypothetical (or real) subject who weights all dimensions equally.

As is so often the case, the (weighted) Euclidean metric in Eq. (4) was chosen for mathematical tractability, conceptual simplicity, and historical precedence. In many stimulus domains (typically with nonanalyzable or unitary perceptual stimuli, or even with more conceptual analyzable stimuli when dimensionality becomes large) the Euclidean metric seems to fit quite well (Shepard, 1964). Furthermore, there is considerable evidence that methods based on it are robust, so that even if the basic metric is non-Euclidean, multidimensional scaling in a Euclidean space may recover the configuration adequately. We regard this particular choice of basic metric, then, as primarily heuristic and pragmatic, although on many grounds it does seem to be the best single choice we could have made. It is, however, within the spirit of the INDSCAL model to assume a much wider class of weighted metrics, and Okada and Imaizumi (1980) have provided such a generalization, along with gradient-based software to fit the model. Also, as argued in the discussion of two-way MDS models, among certain non-Euclidean metrics, the L_1 or *city-block* metric in particular appears to be more appropriate for the more cognitive or conceptual stimulus domains involving analyzable stimuli in which the dimensions are psychologically separable. For this reason we consider an obvious generalization entailing a weighted Minkowski ρ or power metric of the form

$$d_{jk}^{(i)} = \left[\sum_{r=1}^{R} w_{ir} |x_{jr} - x_{kr}|^{\rho} \right]^{1/\rho} \qquad \rho \geq 1 . \tag{5}$$

According to the rescaling of dimensions, the private space for this generalized L_ρ model would be defined as

$$y_{jr}^{(i)} = w_{ir}^{1/\rho} x_{jr} . \tag{6}$$

It is evident that computing ordinary Minkowski ρ metric in this rescaled space, now involving the ρth root of the weights, is equivalent to the weighted Minkowski ρ metric in Eq. (5). See Carroll and Wish (1974b, pp. 412–428) for a technical discussion concerning metrics in MDS.

B. The IDIOSCAL Model and Some Special Cases

The most general in this Euclidean class of models for MDS is what has been called the IDIOSCAL model, standing for *I*ndividual *D*ifferences *i*n Orientation *SCAL*ing (Carroll & Chang, 1970, 1972; Carroll & Wish, 1974a). The intuitive appeal of the IDIOSCAL model is demonstrated by the number of times it, or special cases of it, have been invented or reinvented (e.g., "PARAFAC-2" by Harshman, 1972a, 1972b; "Three-Mode Scaling," by Tucker, 1972, and other procedures proposed by Bloxom, 1978, and by Ramsay, 1981, incorporating this general Euclidean metric or some variant of it); indeed, sometimes it has been simultaneously reinvented *and* renamed (e.g., "the General Euclidean Model" by Young, 1984a). In the IDIOSCAL model, the recovered distance $d_{jk}^{(i)}$ between objects j and k for the ith source of data is given by

$$d_{jk}^{(i)} = \sqrt{\sum_{r}^{R} \sum_{r'}^{R} (x_{jr} - x_{kr}) c_{rr'}^{(i)} (x_{jr'} - x_{kr'})}\,, \tag{7}$$

where r and r' are indices of the R dimensions in the object space and (separately) the source space. This model differs from the INDSCAL model in Eq. (2) by the inclusion of matrix $\mathbf{C}^{(i)} \equiv (c_{rr'}^{(i)})$, which is an $R \times R$ symmetric positive definite or semidefinite matrix, instead of matrix $\mathbf{W}_i$, which is diagonal, with the weights w_{ir} on the diagonals. If each $\mathbf{C}_i$ is constrained to be such a diagonal matrix[3] $\mathbf{W}_i$ with nonnegative entries, then the diagonal entries in the $\mathbf{C}_i$ matrices are interpretable as source weights in the INDSCAL formulation of distance, and the INDSCAL model follows as a special case. This result can be seen by noting that if in Eq. (7), $c_{rr'}^{(i)} = w_{ir}$ when $r = r'$, and 0 when $r \neq r'$, then the terms $(x_{jr} - x_{kr})c_{rr'}^{(i)}(x_{jr'} - x_{kr'})$ drop out if $r \neq r'$ and become $w_{ir}(x_{jr} - x_{kr})^2$ for $r = r'$, thus producing the INDSCAL model of Equation (2). In the general IDIOSCAL model, C_i provides a rotation of the object space to a new (or IDIOsyncratic) coordinate system for source i, followed by differential weighting of the dimensions of this rotated coordinate system. In the Carroll and Chang (1970, pp. 305–310; 1972) approach to interpreting the model, this rotation will be orthogonal. The alternative approach suggested independently by Tucker (1972) and by Harshman (1972a, 1972b) entails no such rotation but assumes differing correlations (or, more geometrically, cosines of angles) between the same dimensions of the object space over different sources. (Further details on the two interpretations of the C_i matrices are given by Carroll and Wish, 1974a, and in the source articles; also see de Leeuw and Heiser, 1982.)

[3] We note here that the matrix $\mathbf{W}_i$ is an $R \times R$ diagonal matrix for the ith subject whereas, previously, the symbol $\mathbf{W}$ has been used to demote the $I \times R$ matrix of weights for the I subjects on the R dimensions (so that the ith row of $\mathbf{W}$ contains the diagonal entries of $\mathbf{W}_i$).

In vector and matrix form, this model can be written as

$$d_{jk}^{(i)} \equiv [(\mathbf{x}_j - \mathbf{x}_k)\mathbf{C}_i(\mathbf{x}_j - \mathbf{x}_k)']^{\frac{1}{2}}, \tag{8}$$

where $\mathbf{C}_i \equiv (c_{rr'}^{(i)})$ is an $R \times R$ matrix. The matrix $\mathbf{C}_i$ is generally assumed to be *symmetric and positive definite or semidefinite.* This metric is exactly what we would obtain if we defined a private perceptual space for individual i by a general linear transformation defined as

$$y_{jr}^{(i)} = \sum_{s=1}^{R} x_{js} q_{sr}^{(i)}, \tag{9}$$

which in vector-matrix notation is

$$\mathbf{y}_j^{(i)} = \mathbf{x}_j \mathbf{Q}_i, \tag{10}$$

and we then computed ordinary Euclidean distances in these private spaces. Matrix $\mathbf{C}_i$ in Eq. (8) will, in this case, simply be

$$\mathbf{C}_i = \mathbf{Q}_i \mathbf{Q}_i', \tag{11}$$

because

$$\begin{aligned}[d_{jk}^{(i)}]^2 &\equiv (y_j^{(i)} - y_k^{(i)})(y_j^{(i)} - y_k^{(i)})' \\ &= (\mathbf{x}_j - \mathbf{x}_k)\mathbf{Q}_i\mathbf{Q}_i'(\mathbf{x}_j - \mathbf{x}_k)',\end{aligned} \tag{12}$$

which is equivalent to Eq. (8) with $\mathbf{C}_i$ as defined in Eq. (11).

Another closely related interpretation is provided by decomposing the (symmetric, positive definite) matrix $\mathbf{C}_i$ into a product of the form

$$\mathbf{C}_i = \mathbf{T}_i \beta_i \mathbf{T}_i', \tag{13}$$

with $\mathbf{T}_i$ orthogonal and β_i diagonal. (This decomposition, based on the singular value decomposition and closely related to principal components analysis, can always be effected. If the $\mathbf{C}_i$'s are positive definite or semidefinite, the diagonal entries of β_i will be nonnegative.)

Then we can define

$$\Phi_i = \mathbf{T}_i \beta_i^{\frac{1}{2}}, \tag{14}$$

and clearly

$$\mathbf{C}_i = \Phi_i \Phi_i'. \tag{15}$$

Actually, Φ_i provides just one possible definition of the matrix $\mathbf{Q}_i$ in Eq. (10). Given any *orthogonal* matrix Γ, we may define

$$\mathbf{Q}_i = \Phi_i \Gamma, \tag{16}$$

and it will turn out that

$$\mathbf{Q}_i\mathbf{Q}_i' = \Phi_i\Gamma\Gamma'\Phi_i' = \Phi_i\Phi_i' = \mathbf{C}_i. \tag{17}$$

Any $\mathbf{Q}_i$ satisfying Eq. (11) can be shown to be of the form stipulated in Eq. (16), but the decomposition of $\mathbf{C}_i$ defined in Eqs. (13)–(14) or (15) (with Γ as the identity matrix) leads to a particularly convenient geometric interpretation. $\mathbf{T}_i$ can be viewed as defining an orthogonal rotation of the reference frame, and thus of the Individual Differences In Orientation (of the reference system) referred to earlier. The diagonal entries of β_i can be interpreted as weights analogous to the w_{ir}'s in the INDSCAL model that are now applied to this IDIOsyncratic reference frame. The considerable intuitive appeal of the IDIOSCAL model notwithstanding, it has empirically yielded disappointing results in general. A major practical drawback of using the IDIOSCAL model is the potential need to provide a separate figure (or set of them) for the spatial representation of each source.

Young's (1984a) approach to fitting what he called the "General Euclidean Model," specifically in the form of his "Principal Directions Scaling," can be viewed as a special case of IDIOSCAL in which the $\mathbf{C}_i$ matrix for each subject is positive semidefinite, with rank R_i less than R (generally R_i = 2). Young assumes each subject projects the IDIOSCAL-type stimulus space defined by $\mathbf{X}$ into an R_i-dimensional subspace so that in this model $\mathbf{Y}_i = \mathbf{X}\Phi_i$ where Φ_i is an $R \times R_i$ projection matrix (so $\Phi_i'\Phi_i = \mathbf{I}_R$; that is Φ_i is an *orthonormal section* projecting orthogonally from $\mathbf{X}$ into an R_i-dimensional subspace, $\mathbf{Y}_i$). In this case $\mathbf{C}_i = \Phi_i\Phi_i'$ will be positive semidefinite and is of rank R_i. The main advantage of this particular special case of IDIOSCAL appears to be that it enables the graphic representation of each subject's private perceptual space in (usually the same) smaller dimensionality—typically two. It is not clear, however, that this model has a convincing rationale *beyond* this practical graphical advantage (see Easterling, 1987, for a successful analysis).

Other models closely related to IDIOSCAL are discussed at length in Arabie et al. (1987, pp. 44–53), but one final three-way model for proximities that bears mentioning generalizes the IDIOSCAL model by adding additional parameters associated with the stimuli (or other objects): the PINDIS (*P*rocrustean *IN*dividual *DI*fferences *S*caling) model and method of Lingoes and Borg (1978). PINDIS adds to the parameters of the IDIOSCAL model a set of *weights* for stimuli, so the model for an individual, in the scalar product domain, is of the form

$$\mathbf{B}_i \cong \mathbf{A}_i\mathbf{X}\mathbf{C}_i\mathbf{X}'\mathbf{A}_i,$$

where $\mathbf{B}_i$ is an $n \times n$ matrix of *scalar products* among the n stimuli for subject/source i, whereas $\mathbf{A}_i$ is an $n \times n$ diagonal matrix of rescaling weights for stimuli. (Although we shall not demonstrate the result here, the IDIOSCAL model in the scalar product domain is of this form, but with $\mathbf{A}_i \equiv \mathbf{I}$ for all i, so that, in effect, the pre- and postmultiplication by $\mathbf{A}_i$ is omitted.) The interpretation of these additional parameters is difficult to justify on psychological grounds. Even more parameters defining different translations of the coordinates of each individual or other source of data, i, are allowed in the general formulation of PINDIS in its scalar product form. Geometrically, the rescaling parameters for stimuli have the effect of moving each stimulus closer to or farther from the centroid in the stimulus space; they do this by multiplying the coordinates by the weight associated with that object. It is hard to envision a psychological mechanism to account for such nonuniform dilations. Moreover, Commandeur (1991, p. 8–9) provides a trenchant and compelling algorithmic critique of PINDIS. Thus, we pursue this model and method no further.

C. Available Software for Two- and Three-Way MDS

1. The Two-Way Case

KYST2A (Kruskal, Young, & Seery, 1973) is the dominant software for two-way MDS. The acronym stands for "Kruskal, Young, Shepard, and Torgerson," and the software synthesizes some of the best parts of various approaches to nonmetric (two-way) MDS that these four contributors have proposed. These algorithms are described in great detail in the previously cited references, so they will not be further described here. The important distinctions are the following:

1. KYST2A minimizes a criterion Kruskal calls STRESS. The standard version of STRESS, often called STRESSFORM1, is defined as

$$\text{STRESSFORM1} = \left(\frac{\Sigma_{jk}(\hat{d}_{jk} - d_{jk})^2}{\Sigma_{jk} d_{jk}^2} \right)^{1/2},$$

$$\text{where } d_{jk} = \left[\sum_{r=1}^{R} (x_{jr} - x_{kr})^2 \right]^{1/2}$$

(i.e., the Euclidean distance in the recovered configuration has coordiantes x_{jr}, for $j = 1, 2 \ldots n$, $r = 1, 2 = R$) and $\hat{d}_{jk}$ is, depending on the user's specification, a linear, monotonic, or other function of the input similarity, s_{jk}, or dissimilarity, δ_{jk}, of j and k (a decreasing or nonincreasing function in the former case and an increasing or nondecreasing function in the latter).

STRESSFORM2 differs only in the normalization factor in the denominator, which is $\sum_{jk}(d_{jk} - \bar{d})^2$, where $\bar{d}$ is the mean of the d_{jk}'s. All sums (and the mean if STRESSFORM2 is used) are over only the values of j and k for which data are given. Generally, the diagonals (s_{jj} or δ_{jj}) or self-similarities/dissimilarities are undefined and therefore are treated as missing data (so that sums and means *exclude* those diagonal values as well).

2. KYST2A allows both metric and nonmetric fitting (and, in fact, includes options for other than either linear or general monotonic functions transforming data into estimated distances; the most important special case allows polynomial functions up to the fourth degree—but such generalized linear functionals are *not* necessarily monotonic). KYST2A allows still other options (see Kruskal et al., 1977, for details) for analyzing three-way data, but fitting only two-way or nonindividual differences models to all subjects or other sources, as well as for performing what Coombs (1964) and others call "multidimensional unfolding" (to be discussed later).

3. KYST2A allows fitting of metrics other than Euclidean—specifically the "Minkowski ρ," or L_ρ, metric of the form given in Eq. (1). In practice, the only two values of ρ that are used at all frequently are $\rho = 2$, the Euclidean case, and, quite inappropriately, $\rho = 1$, the city-block or Manhattan metric case (see Arabie, 1991, for a review). As noted earlier, however, Hubert and Arabie (1988; Hubert, Arabie, & Hesson-Mcinnis, 1992) demonstrated that the problem of fitting an L_1 or city-block metric is more appropriately approached via combinatorial optimization.[4]

Another available algorithm for two-way nonmetric MDS is Heiser and de Leeuw's (1979) SMACOF (*S*caling by *MA*jorizing a *CO*mplicated *F*unction) procedure, based on a *majorization* algorithm, (see de Leeuw and Heiser, 1980, for details), which we will not discuss here except to say that SMACOF optimizes a fit measure essentially equivalent to Kruskal's STRESS. Majorization is an important algorithmic approach deserving much more coverage than space allows. Important references include de Leeuw (1988), Groenen (1993), Groenen, Mathar, and Heiser (1995), Heiser (1991, 1995), Kiers (1990), Kiers and ten Berge (1992), and Meulman (1992).

Wilkinson's (1994) SYSTAT allows many options and considerable flexibility for two-way MDS.

Two other valuable algorithmic developments in two-way (*and* three-way) MDS are the ALSCAL (Takane et al., 1977) procedure and Ramsay's (1978) MULTISCALE. ALSCAL (for *A*lternating *L*east squares *SCAL*ing) differs from previous two-way MDS algorithms in such ways as (1) its loss function, (2) the numerical technique of alternating least squares (ALS) used earlier by Carroll and Chang (1970) and originally devised by Wold (1966; also see de Leeuw 1977a, and de Leeuw & Heiser 1977), and (3) its allowance

[4] For other combinatorial approaches to MDS, see Hubert and Schultz (1976), Poole (1990), and Waller, Lykken, and Tellegen (1995).

for nominal scale (or categorical) as well as interval and ordinal scale data. ALSCAL and MULTISCALE are also applicable to two-mode three-way data, and a three-way version of SMACOF is under development. All three programs will be considered again under spatial distance models for such data.

MULTISCALE (*MULTI*dimensional *SCAL[E]*ing), Ramsay's (1977b, 1978a, 1978b, 1980, 1981, 1982a, 1983) maximum-likelihood-based procedure, although strictly a metric approach, has statistical properties that make it potentially much more powerful as both an exploratory and (particularly) a confirmatory data analytic tool. MULTISCALE, as required by the maximum likelihood approach, makes very explicit assumptions regarding distribution of errors and the relationship of the parameters of this distribution to the parameters defining the underlying spatial representation. One such assumption is that the dissimilarity values δ_{jk} are log normally distributed over replications, but alternative distributional assumptions are also allowed.

The major dividend from Ramsay's (1978) strong assumptions is that the approach enables statistical tests of significance that include, for example, assessment of the correct dimensionality appropriate to the data (via an asymptotically valid chi square test of significance for three-way data treated as *replications*) while fitting a two-way *model*. Another advantage is the resulting confidence regions for gauging the relative precision of stimulus coordinates in the spatial representation. The chief disadvantage is the very strong assumptions entailed for the asymptotic chi squares or confidence regions to be valid. Not least of these is the frequent assumption of ratio scale dissimilarity judgments. In addition, there is the assumption of a specific distribution (log normal, normal, or others with specified parameters) and of statistical independence of the dissimilarity judgments.

2. The Three-Way Case

The most widely used approach to fitting the three-way INDSCAL model is the method implemented in the computer program SINDSCAL (for *S*ymmetric *INDSCAL*, written and documented by Pruzansky, 1975), which updated the older INDSCAL program of Chang and Carroll (1969a, 1989).

SINDSCAL begins with some simple preprocessing stages, initially derived by Torgerson and his colleagues (Torgerson, 1952, 1958) for the two-way case (also see Gower, 1966, and Keller, 1962). The first step, based on the assumption that the initial data are defined on at most an *interval* scale (so that the origin of the scale is arbitrary, leading to the similarities/dissimilarities being related to distances by an inhomogeneous linear function), involves solving the so-called additive constant problem. Then a further transformation of the resulting one-mode two-way matrix of estimated distances to one of estimated *scalar products* is effected. (See Torgerson, 1952, 1958, or Arabie et al., 1987, pp. 71–77, for further details on these preprocessing steps.)

In the two-way classical metric MDS, as described by Torgerson (1952, 1958) and others, the derived (estimated) scalar product matrix is thus simply subjected to a singular value decomposition (SVD), which is mathematically equivalent to a principal components analysis of a correlation or covariance matrix, to obtain an estimate of the $n \times R$ matrix $\mathbf{X}$ of coordinates of the n stimuli in R dimensions, $\hat{\mathbf{X}}$, by minimizing what has been called the STRAIN criterion:

$$\text{STRAIN} = \|\mathbf{B} - \hat{\mathbf{X}}\hat{\mathbf{X}}'\|^2 = \sum_j^n \sum_k^n (b_{jk} - \hat{b}_{jk})^2, \text{ where } \hat{b}_{jk} = \sum_{r=1}^R \hat{x}_{jr}\hat{x}_{kr}.$$

This approach yields a least-squares measure of fit between derived scalar products $\mathbf{B} = (b_{jk})$ and estimated scalar products $\hat{\mathbf{B}} = (\hat{b}_{jk})$. (In some cases, e.g., when fitting nonmetrically, it might be necessary to normalize STRAIN by, say, dividing by the sum of squared entries in the $\mathbf{B}$ matrix; but for the current metric case, and with the preprocessing described earlier, we may use this raw unnormalized form without loss of generality.)

In the three-way case, preprocessing entails these same steps for each similarity or dissimilarity matrix, $\mathbf{S}_i$ or Δ_i, respectively, converting an initial three-way array $\mathbf{S}$ (of similarities) or Δ (of dissimilarities) into a three-way array $\mathbf{B}$ of derived scalar products, where each two-way slice, $\mathbf{B}_i$, is a symmetric matrix of derived (estimated) scalar products for the ith subject or other source of data. CANDECOMP, as applied in this case, optimizes a three-way generalization of the STRAIN criterion discussed earlier, namely,

$$\text{STRAIN} = \sum_i \sum_j \sum_k (b_{jk}^{(i)} - \hat{b}_{jk}^{(i)})^2 = \sum_i \text{STRAIN}_i,$$

where STRAIN_i is STRAIN defined for the ith subject or source and where, if the usual matrix normalization option is used, the constraint is imposed that

$$\sum_i \sum_i (b_{jk}^{(i)})^2 = 1.0, \qquad \text{for all } i,$$

$$\hat{b}_{jk}^{(i)} = \sum_{r=1}^R \hat{w}_{ij}\hat{x}_{jr}\hat{x}_{kr} \tag{18}$$

is a generalized (weighted) scalar product, and parameters $\hat{w}_{ij}$ and $\hat{x}_{jr}$ are (estimates of) the same parameters (without the "hats") as those entering the weighted Euclidean distance defined for INDSCAL in Eq. (2), as demonstrated in Carroll and Chang (1970) and elsewhere (e.g., Appendix B, Arabie et al., 1987). The INDSCAL/SINDSCAL approach to metric three-way MDS then applies a three-way generalization of the SVD, called (three-way) CANDECOMP (for *CAN*onical *DECOMP*osition of *N*-way arrays) to array $\mathbf{B}$, to produce estimates (minimizing the least-squares STRAIN criterion) $\hat{\mathbf{X}}$ and $\hat{\mathbf{W}}$, respectively, of the group stimulus space and the subject weight space. For details of this CANDECOMP procedure and its application to the estimation of parameters of the INDSCAL model, see Carroll and Chang (1970) or Arabie et al. (1987).

Probably the most widely used approach for *nonmetric* fitting of the INDSCAL model is ALSCAL (Takane et al., 1977), which fits the model by optimizing a criterion called SSTRESS, analogous to Kruskal's STRESS, except that it is a normalized least-squares criterion of fit between *squared* distances (in the fitted configuration) and monotonically transformed data (called "disparities" by Takane et al.).

For each subject or data source, SSTRESS is defined analogously to Kruskal's STRESSFORM1, except that, again, *squared* Euclidean distances replace first-power distances. Another difference, irrelevant to the solutions obtained but definitely important vis-à-vis interpretation of values of SSTRESS, is that the square root of the normalized least-squares loss function defines STRESS, whereas SSTRESS is the *untransformed* normalized least-squares criterion of fit. Thus, to the extent that SSTRESS is comparable to STRESS(FORM1) at all, SSTRESS should be compared with squared STRESS. In the three-way case, overall SSTRESS is essentially a (possibly weighted) sum of SSTRESS_i, where SSTRESS_i is the contribution to the SSTRESS measure from subject/source i. As in the case of KYST2A, ALSCAL allows either monotonic or linear transformations of the data, in nonmetric or metric versions, respectively. See Young and Lewyckyj (1981) for a description of the most recent version of the ALSCAL program.

In a recently published Monte Carlo study, Weinberg and Menil (1993) compared recovery of structure of SINDSCAL to that by ALSCAL, under conditions in which both metric and nonmetric analyses were appropriate. Because SINDSCAL allows only metric analyses, even if only ordinal scale data are given, one would expect ALSCAL to be superior in recovering configurations under such ordinal scale conditions because ALSCAL allows a more appropriate nonmetric analysis whereas SINDSCAL necessarily treats the data (inappropriately) as interval scale. It is not clear which of the two should yield better recovery of configurations in the case of interval scale data, because both can allow (appropriate) metric analyses in this case.

Surprisingly, the Weinberg and Menil (1993) Monte Carlo study found that SINDSCAL was superior in recovery both of the stimulus configuration and of subject weights, in the case both of interval and of ordinal scale data (with some fairly severely nonlinear monotonic transformations of the data). The Weinberg and Menil findings may confirm some preliminary results reported by Hahn, Widaman, and MacCallum (1978), at least in the case of mildly nonlinear ordinal data. The explanation of this apparent anomaly appears to rest in the SSTRESS loss function optimized by ALSCAL, probably because SSTRESS measures the fit of transformed data to squared rather than first power distances; the squaring evidently tends to put too much weight on the large distances. A STRESS-based three-way approach might do better in this respect, but unfortunately no such methodology exists at present. Willem Heiser (personal communication) has indicated that he and his colleagues expect eventually to have a three-way version of SMACOF available, which should fill this void.

Version 6 of SYSTAT (Wilkinson, 1994) included, for the first time, software for nonmetric fitting of the INDSCAL model. It is too early to evaluate SYSTAT's performance in this particular domain, but we note that the example given in the documentation (Wilkinson, 1994, p. 140) erroneously suggests that both the subjects and the stimuli are positioned in the same space, rather than in disjoint spaces having a common dimensionality.

Another widely available program for both two- and three-way MDS is Ramsay's (1978, 1982a) MULTISCALE, briefly discussed earlier, which generally assumes *ratio* scale data, and fits via a maximum likelihood criterion, assuming either additive normal error or a lognormal error process. Although a power transformation is allowed, Ramsay's approach generally entails only metric options and in fact makes even stronger metric assumptions than other metric approaches in that it generally requires ratio scale, not the weaker form of interval scale proximity data generally assumed in metric MDS. The main advantage of Ramsay's approach is that it does utilize a maximum likelihood criterion of fit and thereby allows many of the inferential statistics associated with that approach, notably the asymptotic chi square tests that can be used to assess the statistical significance of various effects. (This advantage is undermined somewhat by the fact that the additional parameters associated with subjects or other sources of data in the three-way case can be regarded as *nuisance parameters,* whose number increases with the number of subjects/sources, thus violating one of the key assumptions on which the asymptotic behavior of the asymptotic chi square is based. Ramsay, 1980, however, provided some Monte Carlo results that led to adjustments in the degrees of freedom for the associated statistical tests that correct, at least in part, for this problem.)

Ramsay (1982b) and Winsberg and Ramsay (1984) also introduced a quasi-nonmetric option in MULTISCALE, in which the proximity data are transformed via a monotone spline function or functions in the case of matrix conditional three-way data, which, incidentally, can include an inhomogeneous *linear* function as a special case (thus allowing for more general metric fitting). But this option is not available in most versions of MULTISCALE. It is important, however, to note that this quasi-nonmetric option in MULTISCALE is quite different from the one introduced by Winsberg and Carroll (1989a, 1989b) and Carroll and Winsberg (1986, 1995) in their extended Euclidean two-way MDS. It also differs from the extended INDSCAL (or EXSCAL) approach, described next.

D. The Extended Euclidean Model and Extended INDSCAL

Winsberg and Carroll (1989a, 1989b) and Carroll and Winsberg (1986, 1995) proposed an extension of the simple Euclidean model for two-way proximities and of the INDSCAL model for three-way proximities for which the continuous dimensions of common space are supplemented by a set of specific dimensions, also continuous, but relevant only to individual stimuli or other objects. Here we state the extended model for distances for the three-way, extended INDSCAL case because the two-way extended model is a special case,

$$d_{jk}^{(i)} = \left[\sum_{r=1}^{R} w_{ir}(x_{jr} - x_{kr})^2 + \sigma_{ij} + \sigma_{ik} \right]^{1/2}, \tag{19}$$

where σ_{ij}, called the "specificity" of stimulus j for subject i, is the sum of squares of coordinates of specific dimensions for subject i on stimulus j. Note that we cannot tell in this model how many specific dimensions pertain to a given subject-stimulus combination, only their total effect on (squared) distances in the form of this specificity.

Winsberg and Carroll have adduced both theoretical and strong empirical evidence for the validity of this extended (ordinary or weighted) Euclidean model. They discussed the topic in a series of papers on maximum likelihood methods for either metric or quasi-nonmetric fitting of both the two- and three-way versions of this extended model. As noted in considerable detail in Carroll and Winsberg (1995), there are theoretical reasons why the now classical approach to nonmetric analysis pioneered by Kruskal (1964a, 1964b)—in which a totally general monotonic function (or functions in the three-way case) of the data is sought optimizing either of two forms of Kruskal's STRESS measure (or a large class of other STRESS-like fit measures)—cannot be used for nonmetric fitting of these extended models. The basis of this assertion is the existence of theoretical degeneracies or quasi-degeneracies (solutions yielding apparent perfect or near-perfect fit, but retaining essentially none or very little of the information in the original data) that can always be obtained via such a fully nonmetric fitting. Instead, Winsberg and Carroll (1989a, 1989b) and Carroll and Winsberg (1986, 1995) use a form of quasi-nonmetric fitting in which very (though not totally) general monotonic functions constrained to be continuous and to have continuous first and possibly second derivatives are applied to the distances derived from the model, rather than to the data. Winsberg and Carroll use monotone splines, which can be constrained to have any desired degree of continuity of function and derivatives—although other classes of functions possessing these desiderata could also be utilized. For complete details, see Carroll and Winsberg (1995). Also, see the discussion of the primordial model presented later in this chapter.

Carroll (1988, 1992) has also demonstrated that similar degeneracies would affect attempts at fully nonmetric fitting of discrete models (e.g., ADCLUS/INDCLUS, or tree structures), to be discussed later, and that such quasi-nonmetric fitting would be appropriate here as well. In fact, we argue that even in more well-behaved cases, such as fitting the ordinary two-way Euclidean or three-way INDSCAL model, quasi-degeneracies tend to occur in the case of fully nonmetric fitting, so that such quasi-nonmetric fitting may be more appropriate even in standard MDS. The essence of such quasi-nonmetric fitting is twofold: (1) the monotone function is applied on the model side (as seems more appropriate statistically, in any case), not to the data, and (2) a less than totally general class of monotone functions, such as monotone splines, is utilized so that continuity of the function and at least some of its derivatives can be guaranteed.

Concerning the extended simple and weighted Euclidean models assumed in this work, such extensions, entailing assumptions of dimensions specific to particular stimuli in addition to common dimensions, can be made for such other generalized Euclidean models as IDIOSCAL, three-mode scaling, and PARAFAC-2, or even to non-Euclidean models such as those based on city-block or other L_p metrics.

E. Discrete and Hybrid Models for Proximities

In addition to the continuous spatial models so closely associated with traditional MDS, nonspatial models (which are still geometric in the generic sense of being distance models) entailing discrete, rather than continuous, parameters can also be used profitably for representing proximity data (see, e.g., Gordon, 1996, and other chapters in Arabie, Hubert, and De Soete, 1996; S. C. Johnson, 1967, Hartigan, 1967; Kruskal and Carroll, 1969; Carroll and Chang, 1973; Carroll, 1976; Carroll and Pruzansky, 1975, 1980, 1983, 1986; De Soete & Carroll, 1996; Shepard and Arabie, 1979; Carroll and Arabie, 1980; Arabie et al., 1987). As already argued, such discrete (or "feature") representations may be more appropriate than continuous spatial models for conceptual or cognitive stimuli.

A large number of these discrete models are special cases of a model originally formulated by Shepard and Arabie (1979; see also Shepard, 1974) called ADCLUS, for *AD*ditive *CLUS*tering, and generalized to the three-way, individual differences case by Carroll and Arabie (1983) in the form of the INDCLUS (*IN*dividual *D*ifferences *CLUS*tering) model. We can state both models in the single equation

$$s_{jk}^{(i)} \cong \hat{s}_{jk}^{(i)} = \sum_{r=1}^{R} w_{ir} p_{jr} p_{kr} + g_i , \tag{20}$$

where $s_{jk}^{(i)} \equiv$ proximity (similarity or other measure of closeness) of stimuli (or other objects) j and k for subject (or other source of data) i ($j, k = 1, 2 \ldots n; i = 1, 2 \ldots I$). Note that $\hat{s}_{jk}^{(i)}$ is the model estimate of $s_{jk}^{(i)}$, and "$\cong$" means "approximately equals" except for error terms that will not be further specified here. In addition, $p_{jr} \equiv$ a binary (0, 1) valued variable defining membership ($p_{jr} = 1$) or nonmembership ($p_{jr} = 0$) of stimulus (or other object) j in class or cluster r ($j = 1, 2 \ldots n; r = 1, 2 \ldots R$); $w_{ir} \equiv$ (a continuous nonnegative) importance weight of class or cluster r for proximity judgments (or other measurements) for subject (or other source of data) i; $g_i \equiv$ additive constant for subject (source) i, or, alternatively, that subject's weight for the universal class or cluster of which all the stimuli are members; and $R \equiv$ number of classes or clusters (excluding the universal one).

Equation (20) gives the basic form of the three-way INDCLUS model; ADCLUS is simply the two-way special case in which $I = 1$, so if desired we may drop the "i" subscript.

It might be noted immediately that the ADCLUS/INDCLUS model as stated in Eq. (20) is algebraically of the same form as the scalar product form of the INDSCAL model given in Eq. (18). We simply substitute $b_{jk}^{(i)} = s_{jk}^{(i)}$ and $x_{jr}^{(i)} = p_{jr}^{(i)}$ while we set $g_i = 0$, for all i, and because we are concerned here with models themselves, we may, conceptually, remove the "hats" from Equation (18), of course! In the INDSCAL approach the b's, or (approximate) scalar products, can be interpreted as proximity measures derived from directly judged similarities or dissimilarities.

From the purely algebraic perspective, the ADCLUS/INDCLUS models can be viewed as scalar product models for proximities ($s_{jk}^{(i)}$), but with dimensions' coordinates ($x_{jr} \equiv p_{jr}$) constrained to be binary ($0 - 1$) rather than continuous. Thus, this particular discrete model for proximities can be viewed simply as a special case of the scalar product form of the continuous, spatial model discussed earlier, and most typically associated with MDS, albeit with the simple and straightforward constraint that the dimensions' coordinates must be discrete (specifically, binary).

An interpretation of the ADCLUS/INDCLUS model was provided by Shepard and Arabie (1979) using what is sometimes called a "common features" model, which can be viewed as a special case of Tversky's (1977) features of similarity model. Each of the R classes or clusters can potentially be identified with what Shepard (1974) called an attribute or what Tversky (1977) later dubbed a feature, which each stimulus or other object either has or does not have—a kind of all-or-none dimension, that is. The similarity (proximity) of two objects is incremented for subject (source) i by an amount defined by the weight (w_{ir}) associated with that particular subject/attribute combination if both objects have the attribute, but it is not incremented if either one fails to possess it. This model defines the similarity of a pair of objects as a *weighted* count of common attributes of those two objects—an intuitively quite compelling model. As with INDSCAL, in the three-way, individual differences case, the subjects or objects are differentiated by the profile of (cluster) weights characterizing the individual subjects.

Arabie and Carroll (1980) devised the MAPCLUS algorithm, the most widely used method for fitting the two-way ADCLUS special case of this model. Published data analyses using MAPCLUS include examples from psychoacoustics (Arabie & Carroll, 1980), marketing (Arabie, Carroll, DeSarbo, & Wind, 1981), and sociometry (Arabie & Carroll, 1989); other references are given in Arabie and Hubert (1996, p. 14).

A more widely used method for the discrete representation of similarity data is hierarchical clustering (Gordon, 1996; Hartigan, 1967; Johnson, 1967; Lance & Williams, 1967), which yields a family of clusters such that either two distinct clusters are disjoint or one includes the other as a proper subset. In the usual representation, the objects being clustered appear as terminal nodes of an inverted tree (known as a *dendrogram*), clusters correspond to internal nodes, and the reconstructed distance between two objects is the height of the internal node constituting their meeting point. The model implies that, given two disjoint clusters, all recovered distances between objects in the same cluster are smaller than distances between objects in the two different clusters, and that for any given pair of clusters these between-cluster distances are equal; all triangles are therefore acute isosceles (isosceles with the two larger distances equal). This property is equivalent to the ultrametric inequality, and the tree representation is called an *ultrametric tree*. The ultrametric inequality (u.i.) states that, for ultrametric distances h,

$$h_{jl} \leq \max(h_{jk}, h_{kl}) \qquad \text{for all } j, k, l. \tag{21}$$

Given a set of distances satisfying the u.i., the associated tree can easily be constructed and numerical height values defined (i.e., the numerical ultrametric values arising during the iterative clustering procedure and traditionally presented in the margins of the dendrogram beside their respective levels). An infinite family of ultrametric distance matrices is associated with the topology of a given rooted tree, but if the height values are specified, the particular ultrametric is correspondingly and uniquely specified. Those values must satisfy a partial order based on the hierarchy defined by the tree; namely, the height of an ancestral node corresponding to a superordinate class or cluster must be greater than or equal to the height of a descendant node representing a subordinate class/cluster (a proper subset of the former). This statement assumes the corresponding one-mode two-way proximity data are keyed as dissimilarities.

As an aside, we note that Holman's (1972) classic result relating (two-way one-mode) ultrametric and (two-way one-mode) Euclidean data engendered a highly productive tradition of formal investigations of the interconnections; see Arabie and Hubert (1996, p. 23–24) for an overview.

Although the two representations just discussed, ADCLUS (as fitted by the MAPCLUS algorithm in this case) and an ultrametric tree structure representation (as fitted by one of the standard hierarchical clustering approaches or other procedures to be discussed later for least-squares tree structure fitting), may at first blush seem to be quite distinct, it turns out that the latter is in fact a special case of the former. First, if we define a dissimilarity measure corresponding to a single linear transformation of $s_{jk}^{(i)}$—namely,

$$\delta_{jk}^{(i)} = t_i - s_{jk}^{(i)} \qquad (j \neq k), \tag{22}$$

where t_i is a large positive constant—the $\delta_{jk}^{(i)}$ can be so defined as to satisfy the triangle inequality, and thus the metric axioms, because the diagonal elements are undefined, and symmetry of $\mathbf{S}_i \equiv (s_{jk}^{(i)})$ is assumed by definition.

If we furthermore assume that the clusters for the ADCLUS representation are hierarchically nested, so that every pair of clusters is either disjoint or one is a proper subset of the other, and define $\hat{\delta}_{jk}^{(i)}$ as $t_i - \hat{s}_{jk}^{(i)}$, where $\hat{s}_{jk}^{(i)}$ is as defined in Eq. (20), then $\hat{\delta}_{jk}^{(i)}$ (for fixed i) will satisfy the ultrametric inequality and correspond to an ultrametric defined on a hierarchical tree. Thus, ultrametric trees, this very important class of discrete geometric models so closely associated with hierarchical clustering, can be viewed as a special case of the ADCLUS/INDCLUS model, after this simple linear transformation from the similarity to this dissimilarity form of the model. As will be evident shortly, a wider class of discrete models can also be viewed as special cases of ADCLUS/INDCLUS. For extended analyses using the INDCLUS model, see Arabie, Carroll, and DeSarbo (1987, chap. 6), as well as references cited there and in Arabie and Hubert (1996, p. 14).

F. Common versus Distinctive Feature Models

As already discussed, ADCLUS and INDCLUS are common feature models, in which similarity is defined by a (weighted) count of features shared by the two objects (a "feature" corresponding to a cluster of which an object is a member). (We can, as previously shown, translate this model into one for dissimilarities by simply subtracting the similarities for a given subject or source from a larger constant.) A different feature model for *dissimilarities* is a *distinctive features model,* which depicts dissimilarities using a weighted count of features *not* found in common; (i.e., distinctive features possessed by one or the other object but not by both). As Sattath and Tversky (1987) have shown, common and distinctive feature models are closely related. We discuss this relation in somewhat different terms from those offered by Sattath and Tversky.

One way to write a distinctive features model for three-way dissimilarities data, $\delta_{jk}^{(i)}$, is

$$\delta_{jk}^{(i)} \cong \hat{\delta}_{jk}^{(i)*} = \sum_{r=1}^{R} w_{ir}^{*} |p_{jr} - p_{kr}|, \tag{23}$$

where w_{ir}^{*} is a weight, and p_{jr} and R are as defined earlier. Therefore, Eq. (23) defines a weighted count of distinctive features. It is also a weighted city-block distance model with binary "dimensions" defined by the p's. Because the p's are binary (0, 1) variables, it might be noted that

$$|p_{jr} - p_{kr}| \equiv (p_{jr} - p_{kr})^2 \tag{24}$$

(in fact, $|p_{jr} - p_{kr}| = |p_{jr} - p_{kr}|^{\rho}$ for any $\rho > 0$). Therefore it is equally valid to write

$$\hat{\delta}_{jk}^{(i)*} = \sum_{r=1}^{R} w_{ir}^{*}(p_{jr} - p_{kr})^2 \tag{25}$$

[or $\hat{\delta}_{jk}^{(i)} = \Sigma_{r=1}^{R} w_{ir}^{*} |p_{jr} - p_{kr}|^{\rho}$, for $\rho > 0$], so that $\hat{\delta}^{*}$ can with equal validity be viewed as a weighted city-block, or L_1, metric defined on the (discrete) space whose coordinates are defined by $\mathbf{P} = (p_{jr})$ or as a weighted *squared* Euclidean metric defined on the same space (or, indeed, as any weighted L_{ρ} or Minkowski ρ metric, raised to the ρth power).

We now utilize the definition of $\hat{\delta}^{*}$ in Eq. (25), as squared Euclidean distances, for mathematical convenience. Expanding

$$\begin{aligned}\hat{\delta}_{jk}^{(i)*} &= \sum_{r} w_{ir}^{*}(p_{jr} - p_{kr})^2, \\ &= \sum_{r} w_{ir}^{*}(p_{jr}^2 - 2p_{jr}p_{kr} + p_{kr}^2),\end{aligned}$$

$$= \sum_r w^*_{ir} p^2_{jr} + \sum_r w^*_{ir} p^2_{kr} - 2 \sum_r w^*_{ir} p_{jr} p_{kr},$$

$$= \sum_r w^*_{ir} p_{jr} + \sum_r w^*_{ir} p_{kr} - 2 \sum_r w^*_{ir} p_{jr} p_{kr},$$

$$= u^*_{ij} + u^*_{ik} - 2 \sum_r w^*_{ir} p_{jr} p_{kr}, \tag{26}$$

(since $p^2 \equiv p$, given p binary), or

$$\hat{\delta}^{(i)*}_{jk} = \hat{\delta}^{(i)}_{jk} + u_{ij} + u_{ik}, \tag{27}$$

where $u_{ij} = u^*_{ij} - (t_i - g_i)/2 = \Sigma_r\, w^*_{ir} p_{jr} - t^*_i$, $w_{ir} = 2w^*_{ir}$, and $t^*_i = (t_i - g_i)/2$. As stated earlier, $\hat{\delta}^i_{jk} = t_i - \hat{s}^{(i)}_{jk}$, with t_i as defined in Eq. (22), whereas $\hat{s}^{(i)}_{jk}$ and g_i are as defined in Eq. (20).

Thus, the distinctive feature model can be viewed as a common features model supplemented by uniqueness u_{ij} and u_{ik} that have the same mathematical form as the specificities that transform the common space INDSCAL-model into the extended INDSCAL model discussed earlier. It should be stressed that the substantive interpretation of uniqueness in the present case differs greatly from the specificities in the extended Euclidean model/INDSCAL case. In the latter case, specificities are related to dimensions specific to the stimuli j and k, respectively, whereas in the distinctive features model, the uniqueness values pertain to a weighted count of *all* features that the stimulus possesses and can be viewed in the same spirit as Nosofsky's (1991, p. 98) stimulus bias.

As Sattath and Tversky (1987) have shown, the distinctive features model can always be formulated as a special case of the common features model, however, so that the uniquenesses are not (explicitly) necessary. This conversion is accomplished by supplementing the features in the common features model by a set of additional complementary common features—one complementary feature corresponding to each stimulus or other object. A complementary feature for a particular object is a feature possessed by all objects *except* that object, such as a class or cluster containing all $n - 1$ objects *excluding* that one. (Weights for the common features, including these complementary features, and the additive constants, must be adjusted appropriately.)

In the case of hierarchically nested features (classes or cultures), a distinctive features model will lead to the family of path length or additive trees, discussed later. Other discrete structures such as multiple trees (either ultrametric or path length/additive) are also special cases of either common or distinctive features models, whereas distinctive features models are special cases of common features models, as we noted earlier, so that all of a very large class of discrete models to be discussed later are special cases of the ADCLUS/INDCLUS form of common features model.

Although any distinctive features model can be formulated as a common features model with a large set of features (including complementary ones), the more parsimonious form (covering both common and distinctive features models) stated for similarity data is

$$s_{jk}^{(i)} \cong \hat{s}_{jk}^{(i)} = \sum_{r=1}^{R} w_{ir} p_{jr} p_{kr} - u_{ij} - u_{ik}, \tag{28}$$

where for a common features case $u_{ij} = u_{ik} = -g_i/2$ for all i, j, k.

G. The Primordial Model

We can now formulate a general model that includes the INDSCAL model, the two-way (Euclidean) MDS model, and this large class of discrete models, all as special cases. This *primordial model* will be the linchpin for much of the remaining discussion in Section IV and can be written as

$$s_{jk}^{(i)} \cong \hat{s}_{jk}^{(i)} = M_i\left(\sum_{r=1}^{R} w_{ir} x_{jr} x_{kr} - u_{ij} - u_{ik}\right), \tag{29}$$

where M_i is a monotone (nondecreasing) function. In the case of the INDSCAL model $u_{ij} = .5\Sigma_{r=1}^{R} w_{ir} x_{jr}^2$, so that the expression in parentheses on the right side of Eq. (29) equals $-.5(d_{jk}^{(i)})^2$, where, as before,

$$d_{jk}^{(i)} = \left[\sum_{r=1}^{R} w_{ir}(x_{jr} - x_{kr})^2\right]^{1/2}.$$

In the case of the extended INDSCAL (or EXSCAL) model,

$$u_{ij} = 1/2\left[\sum_{r}^{R} w_{ir} x_{jr}^2 + \sigma_{ij}\right] \tag{30}$$

where σ_{ij} denotes the (i, j)th specificity as defined in that model. Here M_i is a linear function only if similarities are assumed to be (inversely) linearly related to squared (weighted) Euclidean distances. (The two-way special cases of both of these should be obvious.) If the M_i's are assumed to be monotonic (but nonlinear), we recommend the quasi-nonmetric approach, for reasons discussed earlier in the case of fitting the extended INDSCAL (i.e., EXSCAL), or the extended Euclidean model in the two-way case.

As we have already shown, if $x_{jr} = p_{jr}$ (i.e., if the coordinates of the R dimensions are constrained to binary (0, 1) values), then the model becomes the common or distinctive features model and thus has all the other discrete models discussed earlier as special cases. Thus, Eq. (29) can be viewed as the primordial model, of which all others are descendants! It might be noted in passing that since, as Joly and Le Calvé (submitted) have shown, a city-block, or L_1, metric can always be written as the square of a Euclidean metric—although in a space, generally, of very high dimensionality—the L_1 metric models—including the three-way (weighted) version discussed earlier—can also, at least in principle, be included as special cases of this primordial scalar product model. In fact, although this primordial model has the form of a scalar product plus some additive constants, it is easy to show that it can in fact be formulated as an overall scalar product model that requires two additional dimensions to accommodate two additional scalar product terms with special constraints. Although most MDS models are based on distances between points, not scalar products among vectors, we have shown here that such distance models can easily be converted to this general scalar product form, at least in the case of Euclidean and city-block-based models. Some have argued, however, that the processes involved in computing, say, Euclidean distances are very "unnatural" (taking differences between coordinates of two stimuli in an internal spatial representation of the stimuli, squaring this difference, and then summing these squares of coordinate differences over all dimensions; this is possibly followed by a final step of taking the square root, at least in the case of ratio scale distance judgments). It is hard to imagine such operations being wired into the human neural apparatus. In contrast, calculating scalar products (simply multiplying the coordinates for the stimulus pair and summing these products) seems much more plausible as an innate neurological/psychological process. In fact, the general semantic model offered by Landauer and Dumais (1997) assumes the representation of words in a high-dimensional semantic space (about 300 dimensions for their data). Those authors argue that such scalar products can be computed by a very simple neural network. (The model assumes that the association of a given word with unordered strings of other words is based on finding the word in this semantic space closest to the centroid of the words in that string, in the sense of maximizing the cosine of the angle between the word and that centroid. The cosine of an angle in multidimensional space is, in turn, a simple function of the scalar products of vectors.)

Now if we just take the one additional evolutionary step of allowing some x's to be continuous and others discrete (binary, in particular), we immediately generate the hybrid models originally discussed by Carroll (1976; De Soete & Carroll, 1996; also see Hubert & Arabie, 1995a) as an even more general family of models in which continuous spatial structure is *combined* with discrete, nonspatial structure. We discuss some of the discrete and hybrid models that emerge as such special cases of the very broad, general model stated in Eq. (29). See Carroll and Chaturvedi (1995) for a general approach, called CANDCLUS, that allows fitting of a large class of discrete and hybrid models, including the (two- and three-way) common features models discussed previously, to many types of data that are two-way, three-way, and higher-way via either least-squares or a *least absolute deviations* (LAD) criterion. Chaturvedi and Carroll (1994) apply this approach to provide a more efficient algorithm, called SINDCLUS, for fitting the ADCLUS/INDCLUS models via an OLS criterion, whereas Chaturvedi and Carroll (1997) have extended this work to fit with a LAD criterion in a procedure called LADCLUS.

A tree with path-length metric (Carroll & Chang, 1973), or simply a path-length tree, is synonymous with what Sattath and Tversky (1977) called an "additive similarity tree." Unlike ultrametric trees, which have a natural root node, a path-length tree has no unique root. It is not necessary to think of it as being vertically organized into a hierarchy. (In fact, such a tree, for n objects, is consistent with $2n - 2$ *different* hierarchies, corresponding to rooting the tree along any one of its $2n - 2$ distinct branches.) Underlying the structure of a path-length tree is the *four-point condition* that must be satisfied by the estimated path-length distances. This condition, which is a relaxation of the ultrametric inequality, is satisfied by a set of distances (π_{jk}) if and only if, for all quadruples of points j, k, l, and m,

$$\pi_{jk} + \pi_{lm} \geq \pi_{jl} + \pi_{km} \geq \pi_{kl} + \pi_{jm} \text{ implies that } \pi_{jk} + \pi_{lm} = \pi_{jl} + \pi_{km}. \tag{31}$$

That is, the two largest sums of pairs of distances involving the subscripts j, k, l, and m must be equal. See Carroll (1976), Carroll and Pruzansky (1980), or De Soete and Carroll (1996) for a discussion of the rationale for this four-point condition and its relationship to the u.i.

H. Fitting Least-Squares Trees by Mathematical Programming

1. Fitting a Single Ultrametric Tree

Carroll and Pruzansky (1975, 1980) pioneered a mathematical programming approach to fitting ultrametric trees to proximity data via a least-squares criterion. This strategy basically attempts to find a least-squares fit of a distance matrix constrained to satisfy the u.i. by use of a penalty function, which measures the degree of violation of that inequality, as defined in Eq. (21), to a given matrix of dissimilarities. This approach can be extended easily but indirectly to the fitting of path-length trees satisfying the four-point condition, as described later. A more direct procedure entailing a generalization of the Carroll and Pruzansky penalty function approach was proposed and implemented by De Soete (1983) using a penalty function to enforce the four-point condition.

2. Fitting Multiple Tree Structures Using Mathematical Programming Combined with Alternating Least Squares

Many sets of proximity data are not well represented by either simple or hierarchical clusterings. A general model already discussed is the ADCLUS/INDCLUS model, in which proximity data are assumed to arise from discrete attributes that define overlapping but nonhierarchically organized sets. It may happen, however, that the attributes can be organized into two or more separate hierarchies, each of which could represent an organized family of subordinate and superordinate concepts. For example, in the case of animal names one might imagine one hierarchical conceptual scheme based on the phylogenetic scale and another based on function (or relationship to humankind) involving such categories as domesticated versus wild. The former could be classified as pets, work animals, and animals raised for food; pets could be further broken down into house versus outdoor pets, and so on.

This case requires a method to allow fitting *multiple* tree structures to data—a multidimensional generalization of the single tree structure, as it were. We now describe a procedure for fitting such multiple tree structures to a single two-way data matrix of *dis*similarities.

Consider fitting Δ, the two-way data matrix, with a mixture of hierarchical tree structures (HTSs), each satisfying the u.i. In particular, we want to approximate Δ as a sum

$$\Delta \cong \mathbf{H}_1 + \mathbf{H}_2 + \ldots \mathbf{H}_q, \tag{32}$$

where each $\mathbf{H}$ matrix satisfies the u.i. We use an overall *alternating least-squares* (ALS) strategy to fit the mixture of tree structures. In particular, given current fixed estimates of all $\mathbf{H}$ matrices except $\mathbf{H}_q$, we may define

$$\Delta_q^* = \Delta - \sum_{q' \neq q} \hat{\mathbf{H}}q' \tag{33}$$

and use the mathematical programming procedure discussed earlier to fit a least-squares estimate, $\hat{\mathbf{H}}_q$, of $\mathbf{H}_q$, to Δ_q^*.

3. Fitting a Single Path-Length Tree

J. S. Farris (personal communication), as Hartigan (1975, p. 162) noted, has shown that it is possible to convert a path-length tree into an ultrametric tree by a simple operation, given the distances from the root node to each of the nodes corresponding to objects. Letting π_{jO} represent the distance from the jth object to the root node O and π_{jk} represent the path-length distance from j to k, it can be shown that

$$h_{jk} = \pi_{jk} - \pi_{jO} - \pi_{kO}$$

(34)

satisfies the u.i. The h_{jk} will not, however, necessarily satisfy the positivity condition for distances. But both the u.i. *and* positivity will be satisfied by adding a sufficiently large constant Π by defining h_{jk} as

$$h_{jk} = \pi_{jk} - \pi_{jO} - \pi_{kO} + \Pi \equiv \pi_{jk} - u_j - u_k \ (j \neq k) \tag{35}$$

where $u_j = \pi_{jO} - \Pi/2$. An equivalent statement is that

$$\left.\begin{array}{ll} \pi_{jk} = h_{jk} + u_j + u_k & (j \neq k) \\ \text{or } \mathbf{\Pi} = \mathbf{H} + \mathbf{U} & \end{array}\right\}, \tag{36}$$

which states that the path-length distance matrix $\mathbf{\Pi}$ is decomposable into a distance matrix $\mathbf{H}$ that satisfies the u.i. plus an additive residual (which we shall simply call $\mathbf{U}$) where $u_{jk} = u_j + u_k$ for $j \neq k$, and the diagonals of $\mathbf{U}$ are undefined, or zero if defined. The decomposition can be defined so that the u_j's are nonnegative, in which case $\mathbf{U}$ is the distance matrix for a very special path-length tree, usually called a "bush" by numerical taxonomists or a "star" by graph theorists, and is a path-length tree with only one nonterminal (or internal) node. (We use the more standard graph-theoretic term *star* henceforth.) The nonnegative constant u_j is, then, just the length of the branch connecting terminal node j to that single internal node, and the distance between any two distinct terminal nodes, j and k, of the star tree equals $u_j + u_k$. Thus we may summarize Eq. (36) verbally as

A path-length tree = An ultrametric tree + A star tree.

It should be noted that this decomposition is not unique. Many different ways exist for decomposing a fixed path-length tree (PLT) into such a sum. In the case of multiple PLTs, because the sum of Q star trees is itself just a single star tree, we have the extended theorem that

$$\sum_{q}^{Q} \Pi_q = \sum_{q}^{Q} \mathbf{H}_q + \mathbf{U} \tag{37}$$

or, in words,

A sum of PLTs = A sum of ultrametric trees + One star tree.

It should also be noted that both single and multiple path-length or additive trees are also, by quite straightforward inference, special cases of the primordial model in Eq. (29).

We may thus fit mixtures of path-length trees by simply adding to the ALS strategy defined earlier an additional step in which the constants u_j, defining the single star component, are estimated via least-squares procedures. Details of this and of the procedure implementing estimation of the u_j's can be found in Carroll and Pruzansky (1975, 1980). A more computationally efficient, but heuristic (and therefore more likely to be suboptimal), approach to fitting multiple trees was also devised by Carroll and Pruzansky (1986).

I. Hybrid Models: Fitting Mixtures of Tree and Dimensional Structures

Degerman (1970) proposed the first formal hybrid model combining elements of continuous dimensional structure and of discrete class-like structure, using a rotational scheme for high-dimensional MDS solutions, and seeking subspaces with class-like rather than continuous variation. Since then, much has been said but little done about such mixed or hybrid models.

By further generalizing the multiple tree structure model that Carroll and Pruzansky proposed, it is possible to formulate a hybrid model that would include a continuous spatial component in addition to the tree structure components. To return to our hypothetical animal name example, we might postulate, in addition to the two hierarchical structures already mentioned, continuous dimensions of the type best captured in spatial models. In the case of animals, obvious dimensions might include size, ferocity, or color (which itself is multidimensional).

Carroll and Pruzansky (1975, 1980), in fact, generalized the multiple tree structure model just discussed in precisely this direction. The model can be formally expressed as

$$\Delta \cong D_1 + D_2 + \ldots + D_Q + D^2_{E_R}, \tag{38}$$

where D_1 through D_Q are distance matrices arising from tree structures based on either ultrametric or path-length trees, and $D^2_{E_R}$ is a matrix of squared distances arising from an R-dimensional Euclidean space. (The reason for adding squared rather than first-power Euclidean distances is a technical one largely having to do with mathematical tractability and consistency with the general primordial model in Eq. (29).) In effect, to estimate this additional continuous component, we simply add an extra phase to our alternating least-squares algorithm that derives conditional least-squares estimates of these components. Carroll and Pruzansky (1975, 1980) provided details of this additional step. The same reference also provides an illustrative data analysis with a protracted substantive interpretation. Hubert and Arabie (1995b) and Hubert, Arabie, and Meulman (1997) have provided yet another approach to fitting multiple tree structures.

J. Other Models for Two- and Three-Way Proximities

Another direction, already explored to some extent, involves generalization of the discrete models discussed to the case of nonsymmetric proximity data, such as two-mode matrices of proximities or nonsymmetric one-mode proximities (e.g., confusability measures) between pairs of objects from the same set. More extensive discussions of the analysis of nonsymmetric proximities are found in the next section, but we mention some particularly interesting discrete models and methods here. DeSarbo (1982) has devised a model/method called GENNCLUS, for example, which generalizes the ADCLUS/MAPCLUS approach to nonsymmetric proximity data. Furnas (1980) and De Soete, DeSarbo, Furnas, and Carroll (1984a, 1984b) have done the same for tree structures, in a general approach often called "tree 'unfolding."

Yet another fruitfully explored direction involves three-way extensions of a number of these models, which provide discrete analogues to the INDSCAL generalization (Carroll & Chang, 1970) of two-way multidimensional scaling. One such three-way generalization has already been discussed, namely the Carroll and Arabie (1983) INDCLUS generalization of ADCLUS/MAPCLUS to the three-way case—including an application of INDCLUS to some of the Rosenberg and Kim (1975) kinship data (where the third way was defined by those authors' various experimental conditions). In the case of tree structures and multiple tree structures, an obvious direction for individual differences generalization is one in which different individuals are assumed to base their judgments on the same family of trees, but are allowed to have different node heights (in the case of ultrametric trees) or branch lengths (for path-length or additive trees)—that is, single or multiple trees having identical topological structures, but different continuous parameters for each individual or other data source. Carroll, Clark, and DeSarbo (1984) implemented an approach called INDTREES, for fitting just such a model to three-way proximity data. In the hybrid case, a set of continuous stimulus dimensions defining a group stimulus space, together with individual subject weights similar to those assumed in INDSCAL, could also be introduced.

We emphasize that *all* the models discussed thus far for proximity data (even including IDIOSCAL, PARAFAC-2, Tucker's three-mode scaling model, and DeSarbo's GENNCLUS, if sufficiently high dimensionality is allowed) are special cases of the general primordial scalar products model in Eq. (29), some with continuous dimensions and others with discrete valued coordinates on dimensions constrained to binary values and often called "attributes" or "features." The only model discussed *not* in conformity with this generic framework is Lingoes and Borg's (1978) PINDIS—a model we have argued is substantively implausible and overparametrized, in any case. Thus, a very large class of continuous, discrete, and hybrid models can all be viewed as special cases of the primordial model—relatively simple in algebraic form, as well as in its theoretical assumptions concerning psychological processes underlying perception or cognition. Therefore, all can be viewed as special cases of this generic multidimensional model, with the different models varying only with respect to the class of continuous or discrete constraints imposed on the structure and interrelations of the dimensions assumed.

K. Models and Methods for Nonsymmetric Proximity Data

All the approaches to MDS discussed thus far have involved symmetric models for symmetric proximity data. Several types of proximity data are, however, inherently nonsymmetric; for example, the similarity/dissimilarity of j to k presented in that order is not necessarily equal to that of k to j when presented in the reverse order, so that theoretical problems may arise in modeling these data via distance models—which are inherently symmetric, because one of the metric axioms (which by definition is satisfied by all distance functions) demands that $d_{jk} = d_{kj}$ for all j and k. (We prefer the term *nonsymmetric* to *asymmetric*, which is often used as a synonym of the former, because some definitions of *asymmetric* imply *anti*symmetry—that is, that δ_{jk} is definitely *not* equal to δ_{kj}, or even that $\delta_{jk} = a(\delta_{kj})$, where a is a decreasing monotonic function [e.g., $a(\delta)$ = some constant $-\delta$]).

Examples of inherently nonsymmetric proximities include (1) confusions data, in which the probability of confusing k with j (i.e., responding j when stimulus k is presented) is not necessarily the same as that of confusing j with k; (2) direct judgments of similarity/dissimilarity in which systematic order effects may affect judgments, and the subject judges both (j, k) and (k, j) pairs (perhaps the best example of this involves auditory stimuli, where there may be systematic order effects, so that stimulus η followed by stimulus ζ may appear, and be judged, either more or less similar than ζ followed by η; visual and other psychophysical stimuli may be subject to analogous order and other effects; see Holman, 1979, and Nosofsky 1991, for impressive theoretical and substantive developments in this area); and (3) *brand-switching* data, in which the data comprise estimated probabilities (or observed relative frequencies) of consumers who choose brand η on a first occasion but select brand ζ at some later time (see Cooper & Nakanishi, 1988).

Tversky (1977) argued that even direct judgments of similarity/dissimilarity of conceptual/cognitive stimuli may be systematically nonsymmetric—largely depending (we would argue) on how the similarity or dissimilarity question is phrased—and he provided numerous empirical examples. For instance, if subjects are asked "How similar is Vietnam to China?" the response will be systematically different than if they are asked "How similar is China to Vietnam?" In this particular case Vietnam will generally be judged more similar to China than vice versa. Tversky (1977) argued that this occurs because China has more "features" for most subjects than Vietnam does, and that, in this wording of the similarity question, greater weight is given to "distinctive features" unique to the second stimulus than to those unique to the first. This example will be discussed in more detail later when we consider Tversky's (1977) "features of similarity" theoretical framework. We would argue that a slightly different wording of this question, namely "How similar are η and ζ?" would tend to produce symmetric responses (i.e., that any deviations from symmetry are not systematic but result only from random error). It is, in fact, this latter wording or some variation of it that is most often used when direct judgments of similarities/dissimilarities are elicited from human subjects.

1. The Two-Mode Approach to Modeling Nonsymmetric Proximities

The first of the two approaches to modeling nonsymmetric proximities is the *two-mode* approach, in which the stimuli or other objects being modeled are treated as two sets rather than one—in the two-way case, in effect, the proximity data are treated as two-mode two-way, rather than one-mode two-way, with one mode corresponding to rows of the proximity matrix and the other to columns. In the case of confusions data, for example, the rows correspond to the stimuli treated as *stimuli,* whereas the columns correspond to those same stimuli treated as *responses*. In the case of psychophysical stimuli for which there are or may be systematic order effects, the two modes correspond, respectively, to the first and second presented stimulus. More generally, we have the following important principle: any Θ-mode N-way data nonsymmetric in any *modes* corresponding to two *ways* (say, rows and columns) can be accommodated by a *symmetric* model designed for $(\Theta + 1)$-mode N-way data. The extra mode arises from considering the rows and columns as corresponding to distinct entities, so that each entity will be depicted twice in the representation from the symmetric model. (One could, of course, generalize this approach to data nonsymmetric in more than one mode—perhaps even to generalized nonsymmetries involving more than two-ways for a single mode—but we know of few, if any, actual examples of data of this more general type.)

The two-set distance model approach can be viewed very simply as a special case of Coombs's (1964) *unfolding model,* which is inherently designed for data having two or more modes. (In the two-mode case, with respective cardinalities of the stimulus sets being n_1 and n_2, the two-mode data can also be regarded as being in the "corner" of an augmented $(n_1 + n_2) \times (n_1 + n_2)$ matrix with missing entries for all but the $n_1 \times n_2$ submatrix of observed data—hence the traditional but unhelpful jargon of a "corner matrix.") Because most programs for two-way (and some for three-way) MDS allow for missing data. KYST2A allows the user to provide as input such an $n_1 \times n_2$ matrix. The case with which we are dealing, where $n_1 = n_2 = n$, leads directly to a representation in which the stimuli (or other objects) are modeled by $2n$ points—one set of points corresponding to each mode.

Coombs's (1964) distance-based unfolding model assumes preference is inversely monotonically related to distance between the subject's ideal point and a point representing the stimulus in a multidimensional space. Because of the historical association with Coombs's unfolding model, the general problem of analyzing two-mode proximity data (irrespective of whether they are row/column conditional or unconditional and whether ordinal, interval, or ratio scale) is often referred to as the *multidimensional unfolding* problem. From a methodological perspective, there are serious problems with the analysis of two-mode proximities, whether of the type discussed previously or of another type more normally associated with preferential choice (or other dominance) data—which in some cases can lead to data that, as defined earlier, are row or column conditional (e.g., an $I \times n$ matrix of preference ratings for I subjects on n stimuli).

Discussion of the problem of multidimensional unfolding as a special case of MDS, and the associated problems of theoretical degeneracies that make such analyses intractable if great care is not taken, can be found in Kruskal and Carroll (1969) or in Carroll (1972, 1980). To summarize the practical implications for the analyses of two-mode proximities: Either these analyses should be done metrically (i.e., under the assumption of ratio or interval scale data) while assuming row (or column) unconditional off-diagonal data, or they must be done using STRESSFORM2 (or its analogues, in case of other loss functions, such as SSTRESS[5]), whether doing a metric or nonmetric analysis, if row (column) conditional data are entailed. If a fully nonmetric analysis is attempted treating the data as unconditional (whether using STRESSFORM1 or 2), a theoretical degeneracy can be shown always to exist corresponding to perfect (zero) STRESS, although it will account for essentially none of the ordinal information in the data. On the other hand, either a metric or nonmetric analysis assuming (row or column) conditional data, but using STRESSFORM1 instead of STRESSFORM2, will always allow another, even more blatant theoretical degeneracy—as described in Kruskal and Carroll (1969) and Carroll (1972, 1980).

Discrete analogues of the two-set approach to the analysis of nonsymmetric data (or, more generally, rectangular or off-diagonal proximities) are also possible. The tree unfolding approach discussed briefly in the previous section is the most notable example. Note that this analysis was (necessarily) done metrically, assuming row/column unconditional data, for exactly the reasons cited earlier concerning possible degeneracies (which are even more serious in the case of such discrete models as tree structures, where, as discussed earlier, theoretical degeneracies arise in the case of nonmetric analyses—even in the case of symmetric proximities).

[5] It should be noted that ALSCAL should *not* be used for unfolding analyses, however, because the appropriate analogue to STRESSFORM2 is not available in any version of the ALSCAL software.

Tree unfolding has been generalized to the three-way case by De Soete and Carroll (1989). Various approaches to generalizing spatial unfolding to the three-way case have been pursued by DeSarbo and Carroll (1981, 1985); all are restricted to the metric case and to unconditional proximity data, for reasons discussed previously.

Although fully nonmetric analyses are inappropriate (except under the conditions mentioned in the case of spatial unfolding models and always in the case of discrete models), the type of quasi-nonmetric analyses described in the case of the extended Euclidean and INDSCAL models should be permissible, though to our knowledge no one has attempted this approach. Heiser (1989b), however, has pursued some different quasi-nonmetric methods as well as other approaches to unfolding by imposing various constraints on the configurations or by using *homogeneity analysis,* which is closely related to correspondence analysis; see Gifi (1990) for a fuller discussion of this approach to multivariate data analysis, or see Greenacre (1984), Greenacre and Blasius (1994), Lebart, Morineau, and Warwick (1984), and Nishisato (1980, 1993, 1996a, 1996b) for discussions of correspondence analysis. For reasons why correspondence analysis should not be considered a routine alternative to either metric or nonmetric MDS, see Carroll, Kumbasar, and Romney (1997) and Hubert and Arabie (1992).

We note tangentially that a large number of multidimensional models used for representing preferential choice data and methods for analyzing these data using these models have been proposed and can be included under the general rubric of multidimensional scaling (broadly defined). If one characterizes preferences, as does Coombs (1964), as measures of proximity between two sets (stimuli and subjects' ideal points), then the models can be classified as MDS models even if we restrict the domain to geometric models/methods for proximity data. In fact, as Carroll (1972, 1980) has pointed out, a large class of models called the *linear quadratic hierarchy* of models, including the so-called vector model for preferences (Tucker, 1960; Chang & Carroll, 1969a) can all be viewed as special cases or generalizations of the Coombsian unfolding or ideal point model.[6] The vector model, frequently fit by use of the popular MDPREF program (Chang & Carroll, 1969b, 1989), can be viewed as a special case of the unfolding model corresponding to ideal points at infinity (a subject vector then simply indicates the direction of that subject's infinitely distant ideal point). Overviews of these and other models/methods for deterministic (i.e., nonstochastic) analyses of preference data are provided by Carroll (1972, 1980), Weisberg (1974), Heiser (1981, 1987), and DeSarbo and Carroll (1985), whereas discussion of some stochastic models and related methods is found in Carroll and De Soete (1991), De Soete and Carroll (1992), and Marley (1992).

[6] In an important development, the ideal point model has been extended to the technique of discriminant analysis (Takane, Bozdogan, & Shibayama, 1987; Takane, 1989).

2. The One-Mode Approach to Modeling Nonsymmetric Proximities

The other general approach to analyzing nonsymmetric proximities entails a single set representation that assumes a nonsymmetric model. These models can be viewed as adaptations of either a spatial or a discrete (e.g., feature structure) model, modified to accommodate nonsymmetries.

Many of these models are subsumed as special cases of a *nonsymmetric* modification of what we called the primordial (symmetric) model for proximities in Eq. (29), which, in its most general (three-way) case, can be written for $\delta_{jk}^{(i)}$, the proximity between objects j and k for subject i, as

$$\delta_{jk}^{(i)} \cong M_i(b_{jk}^{(i)} + u_{ij} + v_{ik}), \tag{39}$$

where $b_{jk}^{(i)} = \Sigma_r^R w_{ir} x_{jr} x_{kr}$ (a weighted *symmetric* scalar product between j and k for subject/source i),

x_{jr} = continuous (discrete) value of jth object on rth dimension (feature),
w_{ir} = salience weight of rth dimension/feature for the ith subject,
u_{ij} = uniqueness of jth object for ith subject in the first (row) mode, and
v_{ik} = uniqueness of kth object for ith subject in the second (column) mode,

while M_i is a (nonincreasing or nondecreasing) monotonic function for subject i, depending on whether $\delta_{jk}^{(i)}$ is, respectively, a similarity or a dissimilarity measure.

For nonsymmetric proximities, among the special cases of this model are the following.

a. Tversky's Features of Similarity Model

A general statement of this model in set-theoretic terms is (Tversky, 1977)

$$S(j,k) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A) \quad \text{for } \theta, \alpha, \beta \geq 0, \tag{40}$$

where $S(j, k)$ is the similarity of stimuli j and k; A and B are corresponding sets of discrete dimensions/attributes/features (whichever term one prefers); $A \cap B$ is the intersection of sets A and B (or, the set of features common to j and k); $A - B$ is the set difference between A and B or, in words, the set of features possessed by j but not by k (whereas $B - A$ has the opposite meaning); θ, α, and β are numerical weights to be fitted; and f is a finitely additive function, that is,

$$f(\Omega) = \sum_{\Lambda_\Omega \in \Omega} g(\Lambda_\Omega)$$

(where Λ_Ω denotes a feature included in the feature set Ω). An MDS algorithm tailored to fit this model is described by DeSarbo, M. D. Johnson, A. K. Manrai, L. A. Manrai, and Edwards (1992).

When $a = \beta$, this model leads to symmetric proximities S; otherwise it leads to a nonsymmetric model. Tversky (1977) pointed out that the Shepard and Arabie (1979) ADCLUS model corresponds to the special case in which $\alpha = \beta$ (so that the model is symmetric) *and* $f(A) = f(B)$, for all A, B (that is, the weights of the feature sets for stimuli j, k, etc. are all equal). We now demonstrate that the more general model is a special case of the primordial nonsymmetric proximity model expressed in Eq. (39).

First, we rewrite Eq. (40) as

$$\begin{aligned} S(j, k) &= \theta f(A \cap B) + (\alpha + \beta) f(A \cap B) - \alpha f(A - B) \\ &\quad - \alpha f(A \cap B) - \beta f(B - A) - \beta f(A \cap B) \\ &= (\theta + \alpha + \beta) f(A \cap B) - \alpha f(A) - \beta f(B), \end{aligned} \tag{41}$$

with the last expression resulting from substitutions of the set identity $A \equiv (A \cap B) + (A - B)$.

Rewriting Eq. (41) with the same notation used in formulating the two-way ADCLUS model results in nonsymmetric (similarities) of the form

$$s_{jk} = \sum_r^R w_r p_{jr} p_{kr} - u_j - v_k, \tag{42}$$

where

$$u_j = -\alpha^* \sum_r w_r p_{ir}$$

and

$$v_k = -\beta^* \sum_r w_r p_{kr},$$

while

$$\alpha^* = \frac{\alpha}{\theta + \alpha + \beta} \quad \text{and} \quad \beta^* = \frac{\beta}{\theta + \alpha + \beta}.$$

Here $s_{jk} = \frac{1}{\theta + \alpha + \beta} S(j, k)$ (an unimportant scale transformation), and $w_r = \tau(\Lambda_r)$, where Λ_r is the rth "feature" and p_{jr} is a binary indicator variable; $p_{jr} = 1$ iff stimulus j has feature r, and τ is a nonnegative function.

This formulation, of course, is a two-way special case of Eq. (39). Extending this reinterpretation of the features of similarity model to the three-way case, we have

$$s_{jk}^{(i)} \cong \sum_{r=1}^{R} w_{ir} p_{ir} p_{kr} - u_{ij} - v_{ik}, \tag{43}$$

which is a special case of the three-way primordial nonsymmetric scalar product model of Eq. (39), with $x_{jr} \equiv p_{jr}$, that is, with discrete valued dimensions or features (and with $\delta_{jk}^{(i)} \equiv s_{jk}^{(i)}$, with M_i as the identity function). Because Eq. (43) is the three-way generalization of Eq. (39), the u and v terms now have an additional subscript for subject i. Thus, Tversky's (1977) features of similarity model leads to an extended (nonsymmetric) version of the ADCLUS/INDCLUS model—extended by adding the terms u_{ij} and v_{ik}.

Holman (1979) generalized Tversky's features of similarity model to include a monotone transformation of the expression on the right side of Eq. (40), making the model more nearly equivalent to Eq. (39), but only in the two-way case. Holman then formulated a general model for nonsymmetric proximities entailing response biases, a special case of which can be viewed as the two-way case of Eq. (39), with the terms u_j and v_k representing the response biases. Holman defined a general symmetric similarity function as part of his response bias model; our interpretation of Eq. (39) as a special two-way case is dependent on a particular definition of that general similarity function.

Krumhansl (1978) proposed a (continuous) model for nonsymmetric proximities based on what she called a distance-density hypothesis, which leads to an expression for modified distances $\bar{d}$ of the form

$$\bar{d}_{jk} = d_{jk} + \alpha\phi_j + \beta\phi_k, \tag{44}$$

where α, β, and $\bar{d}$, are unrelated to previous usage in this chapter.

The distance-density model has occasioned an impressive algorithmic tradition in two-way MDS. Okada and Imaizumi (1987; Okada, 1990) provide a nonmetric method in which a stimulus is represented as a point and an ellipse (or its generalization) whose center is at that very point in a Euclidean space. Although theirs is a two-way method, it could readily be extended to the three-way case. Distance between the points corresponds to symmetry, and between the radii to skew-symmetry. Bové and Critchley (1989, 1993) devised a metric method for fitting the same model and related their solution to work by Tobler (1979) and Weeks and Bentler (1982). Saito's approach (1991, 1993; Saito & Takeda, 1990) allows the useful option of including unequal diagonal values (i.e., disparate self-similarities) in the analysis. DeSarbo and A. K. Manrai (1992) devised an algorithm that, they maintain, links estimated parameters more closely to Krumhansl's original concept of density.

Krumhansl's original justification for her model, in which ϕ_j and ϕ_k are measures of the spatial density of stimuli in the neighborhoods of j and k, respectively, is actually equally consistent with a formulation using *squared* (Euclidean) distances, namely, modified squared distances $\tilde{d}^2$ defined as

$$\tilde{d}^2_{jk} = d^2_{jk} + \alpha\phi_j + \beta\phi_k, \tag{45}$$

which, in the three-way case, is a nonsymmetric generalization of the extended Euclidean model formulated in the symmetric case by Winsberg and Carroll (1989a, 1989b) and extended to the three-way (EXSCAL) case by Carroll and Winsberg (1986, 1995). It should be clear that this slight reinterpretation of Krumhansl's distance-density model also leads, in the most general three-way case, to a model with continuous spatial parameters of the same general form defined in Eq. (39).

b. Drift Models

As a final class of models leading to this same primordial generalized scalar product form, we now consider two frequently discussed models. One entails "drift" in a fixed *direction* (referred to as a slide-vector model in the implementation of Zielman & Heiser, 1993) and the second entails "drift" toward a fixed *point.* (The first can actually be viewed as a special case of the second, with the fixed point at infinity in some direction.)

Before stating the fixed directional form of the drift model in mathematical terms, we consider a stimulus identification task leading to confusions data, in which a stimulus is presented and the subject attempts to identify it by naming or otherwise giving a response associated with the stimulus presented. In the drift model, we assume the presented stimulus is mapped onto a point (in a continuous multidimensional spatial representation) corresponding to the "true" location of that stimulus *plus* a fixed vector entailing a drift in a fixed direction (and for a fixed distance).

Specifically, if $\mathbf{x}_j$ is the vector representing the true position of stimulus j, the effective position of the presented stimulus will be $\mathbf{x}_j + \psi$, where ψ is the fixed drift vector. If we then assume a Euclidean metric space, the perceived distance between j and another (nonpresented) stimulus k will be (in the two-way case)

$$\tilde{d}_{jk} = \left[\sum_{r=1}^{R} (x_{jr} + \psi_r - x_{kr})^2\right]^{1/2} \tag{46}$$

Now, if we assume that the probability of confusion is a decreasing monotonic function of $\tilde{d}$ then we have

$$\begin{aligned} s_{jk} \cong \text{Prob}\,(k|j) &= M^*(\tilde{d}_{jk}) \\ &= M^{**}\left[\sum_{r} (x_{jr} + \psi_r - x_{kr})^2\right] \end{aligned}$$

$$= M^{**}\left[\sum_r (x_{jr} - x_{kr})^2 + 2\sum_r \psi_r x_{jr}\right.$$

$$\left. - 2\sum_r \psi_r x_{kr} + \sum_r \psi_r^2\right]$$

$$= M^{**}\left[-2\sum_r x_{jr}x_{kr} + \sum_r x_{jr}^2 + \sum_r x_{kr}^2\right.$$

$$\left. + 2\sum_r \psi_r x_{jr} - 2\sum_r \psi_r x_{kr} + \sum_r \psi_r^2\right]$$

$$= M\left[\sum_r x_{jr}x_{kr} - u_j - v_k\right], \tag{47}$$

where M^* is (an arbitrary) monotonic function, and M^{**} and M are also monotonic functions (implied by absorbing first the square root transformation and then the multiplicative factor of -2). (If M^* is monotone decreasing, of course, M will be a monotone *increasing* function.) The important point is that Eq. (47) is of the same form as (the two-way case of) Eq. (39), with $u_j = -.5(\Sigma_r x_{jr}^2 + 2\Sigma_r \psi_r x_{jr} + \Sigma_r \psi_r^2)$ and $v_k = -.5(\Sigma_r x_{kr}^2 - 2\Sigma_r \psi_r x_{ir} + \Sigma_r \psi_r^2)$. Clearly, if we assume a separate drift vector for each subject/source in the three-way case, we get exactly the model form assumed in Eq. (39), with $u_{ij} = -.5(\Sigma_r w_{ir}x_{jr}^2 + 2\Sigma_r \psi_{ir}x_{jr} + \Sigma_r \psi_r^2)$ and $v_{ir} = -.5(\Sigma_r w_{ir}x_{kr}^2 - 2\Sigma_r \psi_{ir}x_{kr} + \Sigma_r \psi_r^2)$.

In the case of the (two-way) model entailing drift toward a fixed point, we assume that the effective position of the presented stimulus, whose true location is $\mathbf{x}_j$, will be $\mathbf{x}_j + \omega(\mathbf{z} - \mathbf{x}_j)$, where $\mathbf{z}$ is the fixed point toward which stimuli drift, while ω is a parameter $(0 \leq \omega \leq 1)$ governing the degree to which $\mathbf{x}_j$ will drift toward $\mathbf{z}$. In this two-way case, the modified Euclidean distance will be

$$\tilde{d}_{jk} = \left[\sum^R (x_{jr} + \omega(z_r - x_{jr}) - x_{kr})^2\right]^{1/2}$$

$$= \left[\sum_r ([1 - \omega]x_{jr} + \omega z_r - x_{kr})^2\right]^{1/2}$$

$$= \left[-2(1 - \omega)\sum_r x_{jr}x_{kr} + (1 - \omega)^2\sum_r x_{jr}^2 + \sum_r x_{kr}^2\right.$$

$$+2\omega(1-\omega)\sum_r z_r x_{jr} - 2\omega \sum_r z_r x_{kr} + \omega^2 \sum_r z_r^2 \Bigg]^{1/2} .$$

Again, if we assume that the probability of confusion, as a measure of proximity, is a monotonic function of $\tilde{d}_{jk}$, we have after some simple algebraic manipulations that proximity is of the same form as in Eq. (47) (with M, u_j, and v_k defined appropriately), although, again the three-way generalization (assuming a possibly different fixed point $\mathbf{z}_i$ for each subject) will be of the same primordial form given in Eq. (39). It is important to note that, except for the additive constants u_{ij} and v_{ik}, this generalized (primordial) scalar product model is essentially *symmetric* (for each subject/source i).

To summarize this section, a large number of superficially disparate models for nonsymmetric proximities are of the same general form as the primordial modified three-way scalar product model stated in Eq. (39), although a very large class of discrete, continuous, and hybrid models for symmetric proximities are of that same general form but have the constraint that $u_{ij} = v_{ij}$, leading to the primordial symmetric model stated in Eq. (29).

It thus appears that a large class of seemingly unrelated models (both two- and three-way, symmetric and nonsymmetric) that have been proposed for proximity data of widely varying kinds are special cases of this generic three-way model that we call the primordial scalar product model, expressed in its most general form in Eq. (39).

3. Three-Way Approaches to Nonsymmetric Proximity Data

In a seminal two-way approach to representing structure underlying nonsymmetric one-mode data, Gower (1977) used areas of triangles and collinearities for the graphical representation of the skew-symmetric component of a nonsymmetric matrix. (Each stimulus was represented by two points, one for its row and another for its column.) The degree of nonsymmetry relates to the area (or sum of signed areas) of triangles, defined by pairs of points and the origin, in two-dimensional subspaces corresponding to matched pairs of eigenvalues in an SVD of the skew-symmetric component of the original matrix of proximity data (after a standard decomposition of the matrix into symmetric and skew-symmetric parts); the direction of the nonsymmetry depends on the sign of the area or of the summed signed areas. That approach forms the basis for numerous three-way models.

Bové and Rocci (1993) generalized Escoufier and Grorud's (1980) approach, in which nonsymmetries are represented by areas of triangles, to the three-way case. Kiers and Takane (1994) provided algorithmic advances on earlier work by Chino (1978, 1990). Similarly, Zielman (1993) provided a three-way approach emphasizing directional planes and collinearities for representing the skew-symmetric component of a nonsymmetric three-way matrix.

We have reviewed elsewhere (Arabie et al., 1987, pp. 50–53) other approaches to this problem (e.g., Kroonenberg & de Leeuw, 1980; also see Kroonenberg, 1983; for developments of Tucker's three-mode three-way principal component analysis, see Tucker, 1972) and will not repeat the discussion here. But Kroonenberg and de Leeuw's (1980, p. 83) empirical conclusion after a protracted analysis that "symmetrization does not really violate the structure of the data" they were analyzing is noteworthy. It is our impression that the extensive collective effort to provide MDS algorithms capable of faithfully representing the nonsymmetric psychological structure so emphasized by Tversky (1977) has borne little substantive fruit.[7] Two possible (and nonexclusive) explanations are (1) nonsymmetry is not very important psychologically or is a minor component of most proximity data, and (2) the extant models are failing to capture the implicit structure. Also see remarks by Nosofsky (1992, p. 38) on this topic.

Concerning the former explanation, Hubert and Baker's (1979) inferential test for detecting significant departures from symmetry has been greatly underemployed. Their examples suggest that presence of nonsymmetry in psychological data has been exaggerated. Similarly, Nosofsky's (1991) incisive treatment of the topic suggests that models incorporating terms like those for stimulus uniqueness in Eq. (39) may preclude the need to posit more fundamental nonsymmetries in similarity data. Concerning the appropriateness of extant models, integrative reviews (e.g., Zielman & Heiser, 1994) and comparative analyses (e.g., Takane & Shibayama, 1986; Molenaar, 1986) should afford a better understanding of exactly what is being captured by models for nonsymmetric data.

We now turn to a different class of such models.

4. Nonspatial Models and Methods for Nonsymmetric Proximity Data

The reader who expects to find nonspatial counterparts to the models just discussed will not be disappointed. For the case of one-mode two-way nonsymmetric data, Hutchinson (1981, 1989) provides a *network* model, NETSCAL (for *NET*work *SCAL*ing), in which a reconstructed distance, defined as the minimum path length between vertices corresponding to stimuli, is assumed to be a generalized power function of the input dissimilarities, and the topology of the network is based only on ordinal information in the data. Hutchinson's illustrative data analyses provide impressive support for the usefulness of his approach.

Klauer and Carroll used a mathematical programming approach to fit network models to one-mode two-way symmetric (1989) and nonsymmetric (1991) proximity data. Using a shortest path definition for the reconstructed distances, their metric algorithm, MAPNET (for *MA*thmetical *P*rogramming *NET*work fitting), seeks to provide the connected network

[7] Okada and Imaizumi (1997) have provided a noteworthy exception to this statement.

with a least-squares fit using a specified number of arcs. Klauer and Carroll (1991) compared their algorithm to Hutchinson's NETSCAL and found the two yielded comparable results, although MAPNET ran faster and provided better variance accounted for. (MAPNET has also been generalized to the three-way case called INDNET; see Klauer and Carroll, 1995.)

We note that neither Gower's (1977) approach nor these network models are subsumed in the primordial model.

V. CONSTRAINED AND CONFIRMATORY APPROACHES TO MDS

Substantive theory can provide a priori expectations concerning the configuration that MDS algorithms generate in the course of an analysis. Beyond being useful in interpreting the configuration, such expectations can actually be incorporated in the analysis in the form of constraints, if the algorithm and software at hand so allow.

Most of the literature on constrained MDS considers only two-way one-mode analyses, but the extension to the three-way case is usually fairly straightforward; thus, we invoke this distinction here much less than in some of the previous sections (also in contrast to our treatment of the topic in Carroll & Arabie, 1980, pp. 619, 628, 633).

A. Constraining the Coordinates

As Heiser and Meulman (1983a, 1983b) noted, most constrained approaches focus either on the coordinates of the configuration or on the function relating the input data to the corresponding recovered interpoint distances. We now consider the former case. Most of the discussion on this topic in our 1980 review centered on constraining the coordinates, and we will not repeat the coverage here. Important subsequent contributions include de Leeuw and Heiser (1980), Lee and Bentler (1980), Takane and Carroll (1981), Weeks and Bentler (1982), DeSarbo, Carroll, Lehmann, and O'Shaughnessy (1982), Heiser and Meulman (1983a, pp. 153–158; 1983b, pp. 387–390), Takane and Sergent (1983), Carroll, De Soete, and Pruzansky (1988, 1989), and Krijnen (1993).

1. Circular/Spherical Configurations

Shepard (1978) masterfully demonstrated the pervasive relevance of spherical configurations in the study of perception. In response, designers of MDS algorithms have made such configurations a popular form of constrained (two-way) MDS. T. F. Cox and M. A. A. Cox (1991) provided a nonmetric algorithm, and earlier metric approaches were devised by de Leeuw and

Heiser (1980) and Lee and Bentler (1980); also see Hubert and Arabie (1994, 1995a) and Hubert, Arabie, and Meulman (1997).

2. Hybrid Approaches Using Circular Configurations

It is too easy to think only of orthogonal dimensions in a metric space for representing the structure in proximities data via MDS, despite the emphasis earlier in this chapter on trees and related discrete structures. Yet other alternatives to dimensions are circles and the matrix form characterized by permuting input data according to a seriation analysis. That is, instead of a series of axes/dimensions or trees (as in Carroll & Pruzansky's hybrid approach, 1975, 1980, discussed earlier) accounting for implicit structure, a set of circles, for example, could be used to account for successively smaller proportions of variance (or components in some other decomposition of an overall goodness-of-fit measure). Taking this development a step further in the hybrid direction, one could also fit a circle as one component, the seriation form as another component, and yet another structure as a third, all in the same analysis of a one-mode symmetric proximities matrix, using the algorithms devised by Hubert and Arabie (1994) and Hubert, Arabie, and Meulman (1997). Those authors (1995a) subsequently generalized this approach to include two-way two-mode proximity matrices.

B. Constraining the Function Relating the Input Data to the Corresponding Recovered Interpoint Distances

In various programs for nonmetric two-way MDS, the plot of this function is appropriately known as the *Shepard diagram,* to give due credit to Shepard's emphasis on this function, which before the advent of nonmetric MDS was generally assumed be linear between derived measures. (Recall that the subtitle of his two 1962 articles is "Multidimensional scaling with an unknown distance function.") Shepard (1962a, 1962b) and Kruskal (1964a, 1964b) devised algorithms for identifying that function with assumptions no stronger than weak monotonicity. In later developments, Shepard (1972, 1974) pointed to the advantages of imposing such constraints as convexity on the monotone regression function. Heiser (1985, 1989b) extended this approach to multidimensional unfolding.

Work by Winsberg and Ramsay (1980, 1981, 1984) and Ramsay (1982a, 1988) using splines rather than Kruskal's (1964b) unconstrained monotone regression to approximate this function has afforded new approaches to imposing constraints on the monotonic function, such as continuity of the function and its first and possibly second derivatives. As already discussed extensively, these continuity constraints have allowed Winsberg and Carroll (1989a, 1989b) and Carroll and Winsberg (1986, 1995) to reverse the direction of the monotone function—treating the data as a (perturbed) monotone function of the distances in the underlying model rather than vice versa, as is done almost universally elsewhere in nonmetric (or even other approaches to quasi-nonmetric) MDS—in their quasi-nonmetric approach to fitting the Extended Euclidean model or its generalization, the Extended INDSCAL (or EXSCAL) model—which includes the ordinary two-way Euclidean MDS model or the three-way INDSCAL models as special cases. The statistical and other methodological advantages of this strategy have already been discussed. The imposition of some mild constraints on various aspects of MDS models often leads to considerable advantages of greater robustness; it also enables fitting, in many cases, of models that are essentially impossible to fit without such constraints.

C. Confirmatory MDS

As Heiser and Meulman (1983b, p. 394) note, "the possibility of constraining the MDS solution in various ways greatly enhances the options for analyzing data in a confirmatory fashion." Approaches to confirmatory MDS have taken several paths. For example, beginning with a traditional statistical emphasis of looking at the residuals, specifically of a nonmetric two-way analysis, Critchley (1986) proposed representing stimuli as small regions rather than points in the MDS solution. The advantage of this strategy is that the regions allow better goodness of fit to the ordinal proximity data. We noted earlier that Ramsay's maximum likelihood approach to two- and three-way MDS allows computing confidence regions for the stimulus mode.

An alternative strategy, used by Weinberg, Carroll, and Cohen (1984), employs resampling (namely, jackknifing and bootstrapping on the subjects' mode in INDSCAL analyses) to obtain such regions. The latter approach is more computationally laborious but less model-specific than Ramsay's, and the results suggest that Ramsay's estimates based on small samples provide an optimistic view of the actual reliability of MDS solutions. For resampling in the two-way case, de Leeuw and Meulman (1986) provide an approach for jackknifing by deleting one stimulus at a time. This approach also provides guidelines as to the appropriate dimensionality for a two-way solution. Heiser and Meulman (1983a) used bootstrapping to obtain confidence regions and assess the stability of multidimensional unfolding solutions.

Extending earlier results by Hubert (1978, 1979) to allow significance tests for the correspondence (independent of any model of MDS) between two or more input matrices, Hubert and Arabie (1989) provided a confirmatory approach to test a given MDS solution against an a priori, idealized structure codified in matrix form. Hubert's (1987) book is essential reading for this topic of research.

Vocational psychology has recently provided a setting for numerous developments related to confirmatory MDS (Hubert & Arabie, 1987; Rounds, Tracey, & Hubert, 1992; Tracey & Rounds, 1993), including a clever application of the INDSCAL model in such an analysis (Rounds & Tracey, 1993).

VI. VISUAL DISPLAYS AND MDS SOLUTIONS

A. Procrustes Rotations

It is often desirable to compare two or more MDS solutions based on the same set of stimuli. When the interpoint distances in the solution(s) to be rotated to maximal congruity with a target configuration are rotationally invariant (as in two-way MDS solutions in the Euclidean metric), the problem of finding the best-fitting orthogonal rotation and a dilation (or overall scale) factor (and even a possible translation of origin of one of the two to align the centroids of the two configurations, if not already done via normalization) has an analytic least-squares solution. But devising a canonical measure of goodness of fit between a pair of matched configurations has proven to be a more challenging problem (see Krzanowski and Marriott, 1994, pp. 134–141, for a concise history of developments).

Analogous to the shift in emphasis from two- to three-way MDS, advances in rotational strategies have progressed from an emphasis on comparing two MDS solutions to comparing more than two. This problem, one variant of which is known as *generalized Procrustes analysis* (Gower, 1975), has occasioned considerable algorithmic development (e.g., ten Berge, 1977; ten Berge & Knol, 1984; ten Berge, Kiers, & Commandeur, 1993; see Commandeur, 1991, and Gower, 1995a, for overviews) and can be cast in the framework of generalized canonical correlation analysis (Green & Carroll, 1988; ten Berge, 1988). As in the case of generalizing many two-way models and associated methods to the three-way (or higher) case, there are a plethora of different approaches to the multiset (e.g., MDS solutions) case, many (but not all) of which are equivalent in the two-set case. Also, in the case of Procrustes analyses, different techniques are appropriate, depending on the class of transformations to which the user believes, on theoretical or empirical grounds, the two (or more) configurations can justifiably be subjected. For example, Gower's generalized Procrustes analysis assumes that each configuration is defined up to an arbitrary similarity transformation (but that the translation component can generally be ignored because of appropriate normalization—e.g., translation of each so that the origin of the coordinate system is at the centroid of the points in that configuration). The canonical correlation-based approaches, on the other hand, allow more general *affine* transformations of the various configurations.

Yet another approach, first used by Green and Rao (1972, pp. 95–97) as a configuration matching approach (in the case of two as well as of three or more configurations) utilizes INDSCAL, applied to distances computed from each separate configuration, as a form of generalized configuration matching (or an alternative generalized Procrustes approach, implicitly assuming yet another class of permissible transformations too complex to be discussed in detail here). This INDSCAL-based approach to configuration matching has been quite useful in a wide variety of situations and has the advantage, associated with INDSCAL in other applications, of yielding a statistically unique orientation of common coordinates describing all the separate configurations. The general approach of configuration matching has long been used to assess mental maps in environmental psychology (e.g., Gordon, Jupp, & Byrne, 1989) and has also found many applications in food technology (see Dijksterhuis & Gower, 1991/1992) and morphometrics (Rohlf & Slice, 1990). In addition to the earlier applications in marketing by Green, cited earlier, a recent approach utilizing either (1) Gower's generalized Procrustes analyses, (2) INDSCAL-based rotation to congruence, or (3) a canonical correlation or generalized canonical correlation–based technique for configuration matching (Carroll, 1968; Green & Carroll, 1989)—or all three—has been quite successfully applied to provide a highly provocative and quite promising new paradigm for marketing analysis, synthesizing elements of a semantic differential approach in a neo-Kellyian framework with an MDS-type spatial representation (see Steenkamp, van Trijp, & ten Berge, 1994). Although devised in the context of a marketing problem, this novel methodological hybridization could very profitably be used in several areas of applied psychology. Other aspects of MDS that are applied to marketing and that could have useful analogues in psychology are discussed in Carroll and Green (1997).

B. Biplots

As Greenacre (1986) succinctly noted,

> "Biplot" is a generic term for a particular class of techniques which represent the rows and columns of a [two-way two-mode] data matrix $\mathbf{Y}$ as points in a low-dimensional Euclidean space. This class is characterized by the property that the display is based on a factorization of the form $\mathbf{AB}'$ [notation modified from the original] of a matrix approximation $\mathbf{Z}$ of $\mathbf{Y}$. The biplot recovers the approximate elements z_{ij} as scalar products $\mathbf{a}_i\mathbf{b}_j'$ of the respective i-th and j-th rows of $\mathbf{A}$ and $\mathbf{B}$, which represent row i and column j respectively in the display.

(Note: The names of these variables bear no necessary relation to usage elsewhere in this chapter.) Such representations have been available since the advent of MDPREF (Carroll & Chang, 1969), but by emphasizing the graphical presentation and by naming it a "biplot" (after its two modes), Gabriel (1971) contributed to the display's popularity. For advances in the underlying statistical techniques, see Gower (1990, 1992, 1995b), Gower and Harding (1988), Meulman and Heiser (1993), and Gower and Hand (1996).

C. Visualization

Young (1984b, p. 77) predicted that "methods for graphically displaying the results of scaling analyses rather than new scaling methods as such" were the new frontier of MDS developments and emphasized color and interactive graphic hardware. This prophecy has turned out to be highly myopic. Although the graphics capabilities of multivariate statistical packages like SYSTAT's SYSGRAPH (Wilkinson, 1994) are indeed impressive and will no doubt continue to improve, they are in no way specific to MDS analyses. The most dramatic graphics-based advances in our understanding of MDS techniques have come from black-and-white graphics portraying results of highly sophisticated investigations that rely on clever and insightful theoretical analyses and simulations (Furnas, 1989; W. P. Jones & Furnas, 1987; Littman, Swayne, Dean, & Buja, 1992).

VII. STATISTICAL FOUNDATIONS OF MDS

During the 1960s, MDS tended to be ignored in the statistical literature, but in the past 15 years, most comprehensive textbooks on multivariate data analysis have included at least one chapter on MDS (e.g., Krzanowski & Marriott, 1994, chap. 5). But relatively few papers (e.g., Cuadras, Fortiana, & Oliva, 1996; Groenen, de Leeuw, & Mathar, 1996) have looked intently at the problem of estimation in MDS. Focusing on the consistency of the Shepard-Kruskal estimator in two-way nonmetric MDS, Brady (1985) reached several interesting conclusions. For example, in aggregating over sources of data to go from a three-way two-mode matrix to a two-way one-mode matrix (as is typically done when two-way nonmetric MDS is applied), it is better to use medians than the traditional arithmetic mean when the data are continuous (e.g., collected using a rating scale). If the data are not continuous (e.g., aggregated over same-different judgments or overt confusions), then accurate recovery of the monotone function typically displayed as the Shepard diagram is unlikely. Brady also developed the beginnings of an hypothesis test for the appropriate dimensionality of MDS solutions.

Ramsay (1982b) provided a scholarly and comprehensive discussion of the underpinnings of his maximum likelihood–based MULTISCALE algorithms (described earlier).

Using matrix permutation/randomization techniques as the basic engine, Hubert and his collaborators (Hubert, 1985, 1987; Hubert & Arabie, 1989; Hubert & Golledge, 1981; Hubert & Subkoviak, 1979) have provided a variety of confirmatory tests applicable to MDS analyses. This general approach makes considerably weaker distributional assumptions than the other papers cited in this section.

Brady (1990) studied the statistical properties of ALS and maximum likelihood estimators when applied to two-way unfolding (e.g., Greenacre & Browne, 1986) and reached the unsettling conclusion that "even after making some strong stochastic assumptions, the ALS estimator is inconsistent (biased) for any squared Euclidean model with an error term." Further statistically based research that could lead to practical improvements in the everyday use of MDS is sorely needed.

Acknowledgments

We are indebted to Yuko Minowa and Zina Taran for bibliographic assistance and to Kathleen Power for editorial expertise.

References

Arabie, P. (1973). Concerning Monte Carlo evaluations of nonmetric scaling algorithms. *Psychometrika, 38,* 607–608.

Arabie, P. (1991). Was Euclid an unnecessarily sophisticated psychologist? *Psychometrika, 56,* 567–587.

Arabie, P., & Carroll, J. D. (1980). MAPCLUS: A mathematical programming approach to fitting the ADCLUS model. *Psychometrika, 45,* 211–235.

Arabie, P., & Carroll, J. D. (1989). Conceptions of overlap in social structure. In L. Freeman, D. R. White, & A. K. Romney (Eds.), *Research methods of social network analysis* (pp. 367–392). Fairfax, VA: George Mason University Press.

Arabie, P., Carroll, J. D., & DeSarbo, W. S. (1987). *Three-way scaling and clustering.* Newbury Park, CA: Sage. (Translated into Japanese by A. Okada & T. Imaizumi, 1990, Tokyo: Kyoritsu Shuppan)

Arabie, P., Carroll, J. D., DeSarbo, W., & Wind, J. (1981). Overlapping clustering: A new method for product positioning. *Journal of Marketing Research, 18,* 310–317. (Republished in 1989, *Multidimensional scaling,* pp. 235–246, by P. E. Green, F. J. Carmone, Jr., & S. M. Smith, Boston: Allyn and Bacon)

Arabie, P., & Hubert, L. (1996). An overview of combinatorial data analysis. In P. Arabie, L. J. Hubert, & G. De Soete (Eds.), *Clustering and classification* (pp. 5–63). River Edge, NJ: World Scientific.

Arabie, P., Hubert, L. J., & De Soete, G. (Eds.). (1996). *Clustering and classification.* River Edge, NJ: World Scientific.

Arce, C. (1993). *Escalamiento multidimensional* [Multidimensional scaling]. Barcelona: Promociones y Publicaciones Universitarias.

Ashby, F. G. (Ed.). (1992). *Multidimensional models of perception and cognition.* Mahwah, NJ: Erlbaum.

Ashby, F. G., Maddox, W. T., & Lee, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychological Science, 5,* 144–151.

Attneave, F. (1950). Dimensions of similarity. *American Journal of Psychology, 63,* 516–556.

Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T., & Silverman, E. (1955). An empirical distribution function for sampling with incomplete information. *Annals of Mathematical Statistics, 26,* 641–647.

Bloxom, B. (1978). Constrained multidimensional scaling in N spaces. *Psychometrika, 43,* 397–408.

Blumenthal, L. M., & Menger, K. (1970). *Studies in geometry.* New York: W. H. Freeman.

Bové, G., & Critchley, F. (1989). The representation of asymmetric proximities. *Proceedings of the First Meeting of the IFCS Italian Group of the Italian Statistical Society* (pp. 53–68). Palermo: Ila Palma.

Bové, G., & Critchley, F. (1993). Metric multidimensional scaling for asymmetric proximities when the asymmetry is one-dimensional. In R. Steyer, K. F. Wender, & K. F. Widaman (Eds.), *Psychometric methodology: Proceedings of the 7th European Meeting of the Psychometric Society in Trier* (pp. 55–60). Stuttgart: Gustav Fischer Verlag.

Bové, G., & Rocci, R. (1993). An alternating least squares method to analyse asymmetric two-mode three-way data. *Proceedings of the 1993 European Meeting of the Psychometric Society* (p. 58). Barcelona: Universidad Pompeu Fabra.

Brady, H. E. (1985). Statistical consistency and hypothesis testing for nonmetric multidimensional scaling. *Psychometrika, 50,* 509–537.

Brady, H. E. (1990). *Statistical properties of alternating least squares and maximum likelihood estimators for vector and squared Euclidean functional preference models.* Berkeley: University of California, Department of Political Science.

Carroll, J. D. (1968). Generalization of canonical correlation analysis to three or more sets of variables. *Proceedings of the 76th Annual Convention of the American Psychological Association, 3,* 227–228.

Carroll, J. D. (1972). Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences: Vol. 1. Theory* (pp. 105–155). New York: Seminar Press. (Reprinted in Key texts on multidimensional scaling; by P. Davies, & A. P. M. Coxon, Eds., 1984, Portsmouth, NH: Heinemann)

Carroll, J. D. (1976). Spatial, non-spatial and hybrid models for scaling. *Psychometrika, 41,* 439–463.

Carroll, J. D. (1980). Models and methods for multidimensional analysis of preferential choice (or other dominance) data. In E. D. Lantermann & H. Feger (Eds.), *Similarity and choice* (pp. 234–289). Bern: Hans Huber.

Carroll, J. D. (1988). *Degenerate solutions in the nonmetric fitting of a wide class of models for proximity data.* Unpublished manuscript, Rutgers University, Graduate School of Management, Newark, New Jersey.

Carroll, J. D. (1992). Metric, nonmetric, and quasi-nonmetric analysis of psychological data. Presidential Address for Division 5, 1992 American Psychological Association Meeting, Washington, DC. Abstract in October 1992 *Score* (Division 5 Newsletter).

Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. In M. R. Rosenzweig & L. W. Porter (Eds.), *Annual review of psychology* (Vol. 31, pp. 607–649). Palo Alto, CA: Annual Reviews. (Reprinted in *Multidimensional scaling: Concepts and applications,* pp. 168–204, by P. E. Green, F. J. Carmone, & S. M. Smith, 1989, Needham Heights, MA: Allyn and Bacon)

Carroll, J. D., & Arabie, P. (1983). INDCLUS: An individual differences generalization of the ADCLUS model and the MAPCLUS algorithm. *Psychometrika, 48,* 157–169. (Reprinted in *Research methods for multimode data analysis*, pp. 372–402, by H. G. Law, W. Snyder, J. Hattie, & R. P. McDonald, Eds., 1984, New York: Praeger)

Carroll, J. D., & Chang, J. J. (1969). A new method for dealing with individual differences in multidimensional scaling (Abstract). *Proceedings of the 19th International Congress of Psychology*. London, England.

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an *N*-way generalization of "Eckart-Young" decomposition. *Psychometrika, 35*, 283–319. (Reprinted in Key texts in multidimensional scaling, by P. Davies & A. P. M. Coxon, Eds., 1984, Portsmouth, NH: Heinemann)

Carroll, J. D., & Chang, J. J. (1972, March). *IDIOSCAL (Individual Differences in Orientation SCALing): A generalization of INDSCAL allowing IDIOsyncratic reference systems as well as an analytic approximation to INDSCAL*. Unpublished manuscript, AT&T Bell Laboratories, Murray Hill, NJ. Presented at a meeting of the Psychometric Society, Princeton, NJ.

Carroll, J. D., & Chang, J. J. (1973). A method for fitting a class of hierarchical tree structure models to dissimilarities data and its application to some "body parts" data of Miller's. *Proceedings of the 81st Annual Convention of the American Psychological Association, 8*, 1097–1098.

Carroll, J. D., & Chaturvedi, A. (1995). A general approach to clustering and multidimensional scaling of two-way, three-way, or higher-way data. In R. D. Luce, M. D'Zmura, D. D. Hoffman, G. Iverson, & A. K. Romney (Eds.), *Geometric representations of perceptual phenomena* (pp. 295–318). Mahwah, NJ: Erlbaum.

Carroll, J. D., Clark, L. A., & DeSarbo, W. S. (1984). The representation of three-way proximities data by single and multiple tree structure models. *Journal of Classification, 1*, 25–74.

Carroll, J. D., & Corter, J. E. (1995). A graph-theoretic method for organizing overlapping clusters into trees and extended trees. *Journal of Classification, 12*, 283–313.

Carroll, J. D., & De Soete, G. (1991). Toward a new paradigm for the study of multiattribute choice behavior. *American Psychologist, 46*, 342–351.

Carroll, J. D., De Soete, G., & Pruzansky, S. (1988). A comparison of three rational initialization methods for INDSCAL. In E. Diday (Ed.), *Data analysis and informatics V* (pp. 131–142). Amsterdam: North Holland.

Carroll, J. D., De Soete, G. & Pruzansky, S. (1989). Fitting of the latent class model via iteratively reweighted least squares CANDECOMP with nonnegativity constraints. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (pp. 463–472). Amsterdam: North Holland.

Carroll, J. D., & Green, P. E. (1997). Psychometric methods in marketing research: Part II, multidimensional scaling [Guest editorial]. *Journal of Marketing Research, 34*, 193–204.

Carroll, J. D., Kumbasar, E., & Romney, A. K. (1997). An equivalence relation between correspondence analysis and classical metric multidimensional scaling for the recovery of Euclidean distances. *British Journal of Mathematical and Statistical Psychology, 50*, 81–92.

Carroll, J. D., & Pruzansky, S. (1975). Fitting of hierarchical tree structure (HTS) models, mixtures of HTS models, and hybrid models, via mathematical programming and alternating least squares. *Proceedings of the U.S.–Japan Seminar on Multidimensional Scaling*, 9–19.

Carroll, J. D., & Pruzansky, S. (1980). Discrete and hybrid scaling models. In E. D. Lantermann & H. Feger (Eds.), *Similarity and choice* (pp. 108–139). Bern, Switzerland: Hans Huber.

Carroll, J. D., & Pruzansky, S. (1983). Representing proximities data by discrete, continuous or "hybrid" models. In J. Felsenstein (Ed.), *Numerical taxonomy* (pp. 229–248). New York: Springer-Verlag.

Carroll, J. D., & Pruzansky, S. (1986). Discrete and hybrid models for proximity data. In W. Gaul & M. Schader (Eds.), *Classification as a tool of research* (pp. 47–59). Amsterdam: North Holland.

Carroll, J. D., & Winsberg, S. (1986). Maximum likelihood procedures for metric and quasi-nonmetric fitting of an extended INDSCAL model assuming both common and specific dimensions. In J. de Leeuw, W. J. Heiser, J. Meulman, & F. Critchley (Eds.), *Multidimensional data analysis* (pp. 240–241). Leiden: DSWO Press.

Carroll, J. D., & Winsberg, S. (1995). Fitting an extended INDSCAL model to three-way proximity data. *Journal of Classification, 12,* 57–71.

Carroll, J. D., & Wish, M. (1974a). Models and methods for three-way multidimensional scaling. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2, pp. 57–105). San Francisco: W. H. Freeman.

Carroll, J. D., & Wish, M. (1974b). Multidimensional perceptual models and measurement methods. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 2, pp. 391–447). New York: Academic Press. (Reprinted in *Key texts in multidimensional scaling,* by P. Davies & A. P. M. Coxon, Eds., 1984, Portsmouth, NH: Heinemann)

Chandon, J. L., & De Soete, G. (1984). Fitting a least squares ultrametric to dissimilarity data: Approximation versus optimization. In E. Diday, M. Jambu, L. Lebart, J. Pagès, & R. Tomassone (Eds.), *Data analysis and informatics III* (pp. 213–221). Amsterdam: North-Holland.

Chang, J. J., & Carroll, J. D. (1969a). *How to use INDSCAL, a computer program for canonical decomposition of N-way tables and individual differences in multidimensional scaling.* Murray Hill, NJ: AT&T Bell Laboratories.

Chang, J. J., & Carroll, J. D. (1969b). *How to use MDPREF, a computer program for multidimensional analysis of preference data.* Murray Hill, NJ: AT&T Bell Laboratories.

Chang, J. J., & Carroll, J. D. (1989). A short-guide to MDPREF: Multidimensional analysis of preference data. In P. E. Green, F. J. Carmone, & S. M. Smith, *Multidimensional scaling: Concepts and applications* (pp. 279–286). Needham Heights, MA: Allyn and Bacon.

Chaturvedi, A., & Carroll, J. D. (1994). An alternating combinatorial optimization approach to fitting the INDCLUS and generalized INDCLUS models. *Journal of Classification, 11,* 155–170.

Chaturvedi, A., & Carroll, J. D. (1997). An L_1-norm procedure for fitting overlapping clustering models to proximity data. In Y. Dodge (Ed.), *Statistical data analysis based on the* L_1*-norm and related methods* (IMS Lecture Notes Monograph No. 30, pp. 443–456). Hayward, CA: Institute of Mathematical Statistics.

Chino, N. (1978). A graphical technique for representing the asymmetric relationships between N objects. *Behaviormetrika, 5,* 23–40.

Chino, N. (1990). A generalized inner product model for the analysis of asymmetry. *Behaviormetrika, 27,* 25–46.

Cliff, N., Pennell, R., & Young, F. W. (1966). Multidimensional scaling in the study of set. *American Psychologist, 21,* 707.

Commandeur, J. J. F. (1991). *Matching configurations.* Leiden: DSWO Press.

Coombs, C. H. (1964). *A theory of data.* New York: Wiley.

Cooper, L. G., & Nakanishi, M. (1988). *Market-share analysis.* Boston: Kluwer.

Corter, J., & Tversky, A. (1986). Extended similarity trees. *Psychometrika, 51,* 429–451.

Cox, T. F., & Cox, M. A. A. (1994). *Multidimensional scaling.* London: Chapman & Hall.

Cox, T. F., Cox, M. A. A., & Branco, J. A. (1991). Multidimensional scaling for n-tuples. *British Journal of Mathematical and Statistical Psychology, 44,* 195–206.

Critchley, F. (1986). Analysis of residuals and regional representation in nonmetric multidimensional scaling. In W. Gaul & M. Schader (Eds.), *Classification as a tool of research* (pp. 67–77). Amsterdam: North-Holland.

Critchley, F., & Fichet, B. (1994). The partial order by inclusion of the principal classes of dissimilarity on a finite set, and some of their basic properties. In B. Van Cutsem (Ed.), *Classification and dissimilarity analysis* (pp. 5–66). Heidelberg: Springer-Verlag.

Cuadras, C. M., Fortiana, J., & Oliva, F. (1996). Representation of statistical structures, classification and prediction using multidimensional scaling. In W. Gaul & D. Pfeifer (Eds.), *From data to knowledge* (pp. 20–31). Heidelberg: Springer-Verlag.

Daws, J. T. (1993). *The analysis of free-sorting data: Beyond pairwise cooccurrences.* (Doctoral dissertation, University of Illinois at Urbana-Champaign). (UMI Dissertation NO. 9411601).

Daws, J. T. (1996). The analysis of free-sorting data: Beyond pairwise cooccurrences. *Journal of Classification, 13,* 57–80.

Degerman, R. L. (1970). Multidimensional analysis of complex structure: Mixtures of class and quantitative variation. *Psychometrika, 35,* 475–491.

de Leeuw, J. (1977a). Applications of convex analysis to multidimensional scaling. In J. R. Barra, F. Brodeau, G. Romier, & B. van Cutsem (Eds.), *Recent developments in statistics* (pp. 133–145). Amsterdam: North-Holland.

de Leeuw, J. (1977b). Correctness of Kruskal's algorithms for monotone regression with ties. *Psychometrika, 42,* 141–144.

de Leeuw, J. (1988). Convergence of the majorization method for multidimensional scaling. *Journal of Classification, 5,* 163–180.

de Leeuw, J., & Heiser, W. (1977). Convergence of correction-matrix algorithms for multidimensional scaling. In J. C. Lingoes (Ed.), *Geometric representations of relational data: Readings in multidimensional scaling* (pp. 735–752). Ann Arbor, MI: Mathesis.

de Leeuw, J., & Heiser, W. (1980). Multidimensional scaling with restrictions on the configuration. In P. R. Krishnaiah (Ed.), *Multivariate analysis* (Vol. 5, pp. 501–522). New York: North Holland.

de Leeuw, J., & Heiser, W. (1982). Theory of multidimensional scaling. In P. R. Krishnaiah, & L. N. Kanal (Eds.), *Handbook of statistics Vol. 2: Classification, pattern recognition and reduction of dimensionality* (pp. 285–316). Amsterdam: North-Holland.

de Leeuw, J., Heiser, W., Meulman, J., & Critchley, F. (Eds.). (1986). *Multidimensional data analysis.* Leiden: DSWO Press.

de Leeuw, J., & Meulman, J. (1986). A special jackknife for multidimensional scaling. *Journal of Classification, 3,* 97–112.

DeSarbo, W. S. (1982). GENNCLUS: New models for general nonhierarchical clustering analysis. *Psychometrika, 47,* 446–449.

DeSarbo, W. S., & Carroll, J. D. (1981). Three-way metric unfolding. *Proceedings of the Third ORSA/TIMS Special Interest Conference on Market Measurement and Analysis,* 157–183.

DeSarbo, W. S., & Carroll, J. D. (1985). Three-way metric unfolding via weighted least squares. *Psychometrika, 50,* 275–300.

DeSarbo, W. S., Carroll, J. D., Lehman, D. R., & O'Shaughnessy, J. (1982). Three-way multivariate conjoint analysis. *Marketing Science, 1,* 323–350.

DeSarbo, W. S., Johnson, M. D., Manrai, A. K., Manrai, L. A., & Edwards, E. A. (1992). TSCALE: A new multidimensional scaling procedure based on Tversky's contrast model. *Psychometrika, 57,* 43–69.

DeSarbo, W. S., & Manrai, A. K., (1992). A new multidimensional scaling methodology for the analysis of asymmetric proximity data in marketing research. *Marketing Science,* 11, 1–20.

De Soete, G. (1983). A least squares algorithm for fitting additive trees to proximity data. *Psychometrika, 48,* 621–626.

De Soete, G., & Carroll, J. D. (1989). Ultrametric tree representations of three-way three-mode data. In R. Coppi & S. Bolasco (Eds.), *Analysis of multiway data matrices* (pp. 415–426). Amsterdam: North-Holland.

De Soete, G., & Carroll, J. D. (1992). Probabilistic multidimensional models of pairwise choice data. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 61–88). Mahwah, NJ: Erlbaum.

De Soete, G., & Carroll, J. D. (1996). Tree and other network models for representing proximity data. In P. Arabie, L. J. Hubert, & G. De Soete (Eds.), *Clustering and classification* (pp. 157–197). River Edge, NJ: World Scientific.

De Soete, G., DeSarbo, W. S., Furnas, G. W., & Carroll, J. D. (1984a). The estimation of ultrametric and path length trees from rectangular proximity data. *Psychometrika, 49,* 289–310.

De Soete, G., DeSarbo, W. S., Furnas, G. W., & Carroll, J. D. (1984b). Tree representations of rectangular proximity matrices. In E. Degreef & J. Van Buggenhaut (Eds.), *Trends in mathematical psychology* (pp. 377–392). Amsterdam: North-Holland.

De Soete, G., Feger, H., & Klauer, K. C. (Eds.) (1989). *New developments in psychological choice modeling.* Amsterdam: North-Holland.

Dijksterhuis, G. B., & Gower, J. C. (1991/1992). The interpretation of generalized Procrustes analysis and allied methods. *Food Quality and Preference, 3,* 67–87.

Easterling, D. V. (1987). Political science: Using the generalized Euclidean model to study ideological shifts in the U.S. Senate. In F. Young & R. M. Hamer (Eds.), *Multidimensional scaling: History, theory, and applications* (pp. 219–256). Mahwah, NJ: Erlbaum.

Ennis, D. M., Palen, J. J., & Mullen, K. (1988). A multidimensional stochastic theory of similarity. *Journal of Mathematical Psychology, 32,* 449–465.

Escoufier, Y., & Grorud, A. (1980). Analyse factorielle des matrices carrees non symetriques. [Factor analysis of square nonsymmetric matrices]. In E. Diday, L. Lebart, J. P. Pagès, & R. Tomassone (Eds.), *Data analysis and informatics* (pp. 263–276). Amsterdam: North-Holland.

Fichet, B. (1994). Dimensionality problems in L_1-norm representations. In B. Van Cutsem (Ed.), *Classification and dissimilarity analysis* (pp. 201–224). Heidelberg: Springer-Verlag.

Fitzgerald, L. F., & Hubert, L. J. (1987). Multidimensional scaling: Some possibilities for counseling psychology. *Journal of Counseling Psychology, 34,* 469–480.

Furnas, G. W. (1980). *Objects and their features: The metric analysis of two-class data.* Unpublished doctoral dissertation, Stanford University, Stanford, CA.

Furnas, G. W. (1989). Metric family portraits. *Journal of Classification, 6,* 7–52.

Gabriel, K. R. (1971). The biplot-graphic display of matrices with application to principal component analysis. *Biometrika, 58,* 453–467.

Gifi, A. (1990). *Nonlinear multivariate analysis.* New York: Wiley.

Glazer, R., & Nakamoto, K. (1991). Cognitive geometry: An analysis of structure underlying representations of similarity. *Marketing Science, 10,* 205–228.

Gordon, A. D. (1996). Hierarchical classification. In P. Arabie, L. J. Hubert, & G. De Soete (Eds.), *Clustering and classification* (pp. 65–121). River Edge, NJ: World Scientific.

Gordon, A. D., Jupp, P. E., & Byrne, R. W. (1989). The construction and assessment of mental maps. *British Journal of Mathematical and Statistical Psychology, 42,* 169–182.

Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika, 53,* 325–338.

Gower, J. C. (1975). Generalized procrustes analysis. *Psychometrika, 40,* 33–51.

Gower, J. C. (1977). The analysis of asymmetry and orthogonality. In J. Barra, F. Brodeau, G. Romier, & B. van Cutsem (Eds.), *Recent developments in statistics* (pp. 109–123). Amsterdam: North-Holland.

Gower, J. C. (1990). Three-dimensional biplots. *Biometrika, 77,* 773–785.

Gower, J. C. (1995a). Orthogonal and projection Procrustes analysis. In W. J. Krzanowski (Ed.), *Recent advances in descriptive multivariate analysis* (pp. 113–134). Oxford: Clarendon Press.

Gower, J. C. (1995b). A general theory of biplots. In W. J. Krzanowski (Ed.), *Recent advances in descriptive multivariate analysis* (pp. 283–303). Oxford: Clarendon Press.

Gower, J. C., & Greenacre, M. J. (1996). Unfolding a symmetric matrix. *Journal of Classification, 13,* 81–105.

Gower, J. C., & Hand, D. J. (1996). *Biplots.* New York: Chapman & Hall.

Gower, J. C., & Harding, S. A. (1988). Nonlinear biplots. *Biometrika, 75,* 445–455.

Green, P. E., Carmone, F. J., Jr., & Smith, S. M. (1989). *Multidimensional scaling: Concepts and applications.* Boston: Allyn and Bacon.

Green, P. E., & Rao, V. R. (1972). Configural synthesis in multidimensional scaling. *Journal of Marketing Research, 9,* 65–68.

Greenacre, M. J. (1984). Theory and applications of correspondence analysis. London: Academic Press.

Greenacre, M. J. (1986). Discussion on paper by Gabriel and Odoroff. In J. de Leeuw, W. Heiser, J. Meulman, & F. Critchley (Eds.), *Multidimensional data analysis* (pp. 113–114). Leiden: DSWO Press.

Greenacre, M. J., & Blasius, J. (Eds.). (1994). *Correspondence analysis: Recent developments and applications.* New York: Academic Press.

Greenacre, M. J., & Browne, M. W. (1986). An efficient alternating least-squares algorithm to perform multidimensional unfolding. *Psychometrika, 51,* 241–250.

Groenen, P. J. F. (1993). *The majorization approach to multidimensional scaling: Some problems and extensions.* Leiden: DSWO Press.

Groenen, P. J. F., de Leeuw, J., & Mathar, R. (1996). Least squares multidimensional scaling with transformed distances. In W. Gaul & D. Pfeifer (Eds.), *From data to knowledge* (pp. 177–185). Heidelberg: Springer-Verlag.

Groenen, P. J. F., Mathar, R., & Heiser, W. J. (1995). The majorization approach to multidimensional scaling for Minkowski distances. *Journal of Classification, 12,* 3–19.

Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika, 33,* 465–506.

Hahn, J., Widaman, K. F., & MacCallum, R. (1978). *Robustness of INDSCAL and ALSCAL with respect to violations of metric assumptions.* Paper presented at the Annual Meeting of the Psychometric Society, Hamilton, Ontario, Canada.

Harshman, R. A. (1972a). Determination and proof of minimum uniqueness conditions for PARAFAC1. University of California at Los Angeles, *Working Papers in Phonetics 22.*

Harshman, R. A. (1972b). PARAFAC2: Mathematical and technical notes. University of California at Los Angeles, *Working Papers in Phonetics 22.*

Hartigan, J. A. (1967). Representation of similarity matrices by trees. *Journal of the American Statistical Association, 62,* 1140–1158.

Hartigan, J. A. (1975). *Clustering algorithms.* New York: Wiley (Translated into Japanese by H. Nishida, M. Yoshida, H. Hiramatsu, & K. Tanaka, 1983, Tokyo: Micro Software).

Heiser, W. J. (1981). *Unfolding analysis of proximity data.* Unpublished doctoral dissertation, University of Leiden.

Heiser, W. J. (1985). *Multidimensional scaling by optimizing goodness-of-fit to a smooth hypothesis.* Internal Report RR-85-07. University of Leiden: Department of Data Theory.

Heiser, W. J. (1987). Joint ordination of species and sites: The unfolding technique. In P. Legendre & L. Legendre (Eds.), *Developments in numerical ecology* (pp. 189–221). Heidelberg: Springer-Verlag.

Heiser, W. J. (1988). Multidimensional scaling with least absolute residuals. In H.-H. Bock (Ed.), *Classification and related methods of data analysis* (pp. 455–462). Amsterdam: North-Holland.

Heiser, W. J. (1989a). The city-block model for three-way multidimensional scaling. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (pp. 395–404). Amsterdam: North-Holland.

Heiser, W. J. (1989b). Order invariant unfolding analysis under smoothness restrictions. In G. De Soete, H. Feger, & C. Klauer (Eds.), *New developments in psychological choice modeling* (pp. 3–31). Amsterdam: North-Holland.

Heiser, W. J. (1991). A generalized majorization method for least squares multidimensional scaling of pseudodistances that may be negative. *Psychometrika, 56,* 7–27.

Heiser, W. J. (1995). Convergent computation by iterative majorization: Theory and applications in multidimensional data analysis. In W. Krzanowski (Ed.), *Recent advances in descriptive multivariate analysis* (pp. 149–181). New York: Oxford University Press.

Heiser, W. J., & de Leeuw, J. (1979). *How to use SMACOF-III* (Research Report). Leiden: Department of Data Theory.

Heiser, W. J., & Meulman, J. (1983a). Analyzing rectangular tables by joint and constrained multidimensional scaling. *Journal of Econometrics, 22,* 139–167.

Heiser, W. J., & Meulman, J. (1983b). Constrained multidimensional scaling, including confirmation. *Applied Psychological Measurement, 7,* 381–404.

Holman, E. W. (1972). The relation between hierarchical and Euclidean models for psychological distances. *Psychometrika, 37,* 417–423.

Holman, E. W. (1978). Completely nonmetric multidimensional scaling. *Journal of Mathematical Psychology, 18,* 39–51.

Holman, E. W. (1979). Monotonic models for asymmetric proximities. *Journal of Mathematical Psychology, 20,* 1–15.

Hubert, L. J. (1978). Generalized proximity function comparisons. *British Journal of Mathematical and Statistical Psychology, 31,* 179–192.

Hubert, L. J. (1979). Generalized concordance. *Psychometrika, 44,* 135–142.

Hubert, L. J. (1985). Combinatorial data analysis: Association and partial association. *Psychometrika, 50,* 449–467.

Hubert, L. J. (1987). *Assignment methods in combinatorial data analysis.* New York: Marcel Dekker.

Hubert, L., & Arabie, P. (1986). Unidimensional scaling and combinatorial optimization. In J. de Leeuw, W. Heiser, J. Meulman, & F. Critchley (Eds.), *Multidimensional data analysis* (pp. 181–196). Leiden: DSWO Press.

Hubert, L., & Arabie, P. (1987). Evaluating order hypotheses within matrices. *Psychological Bulletin, 102,* 172–178.

Hubert, L. J., & Arabie, P. (1988). Relying on necessary conditions for optimization: Unidimensional scaling and some extensions. In H.-H. Bock (Ed.), *Classification and related methods of data analysis* (pp. 463–472). Amsterdam: North-Holland.

Hubert, L., & Arabie, P. (1989). Combinatorial data analysis: Confirmatory comparisons between sets of matrices. *Applied Stochastic Models and Data Analysis, 5,* 273–325.

Hubert, L., & Arabie, P. (1992). Correspondence analysis and optimal structural representations. *Psychometrika, 56,* 119–140.

Hubert, L., & Arabie, P. (1994). The analysis of proximity matrices through sums of matrices having (anti-)Robinson forms. *British Journal of Mathematical and Statistical Psychology, 47,* 1–40.

Hubert, L., & Arabie, P. (1995a). The approximation of two-mode proximity matrices by sums of order-constrained matrices. *Psychometrika, 60,* 573–605.

Hubert, L., & Arabie, P. (1995b). Iterative projection strategies for the least-squares fitting of tree structures to proximity data. *British Journal of Mathematical and Statistical Psychology, 48,* 281–317.

Hubert, L. J., Arabie, P., & Hesson-Mcinnis, M. (1992). Multidimensional scaling in the city-block metric: A combinatorial approach. *Journal of Classification, 9,* 211–236.

Hubert, L. J., Arabie, P., & Meulman, J. (1997). Linear and circular unidimensional scaling for symmetric proximity matrices. *British Journal of Mathematical and Statistical Psychology, 50.*

Hubert, L. J., & Baker, F. B. (1979). Evaluating the symmetry of a proximity matrix. *Quality and Quantity, 13,* 77–84.
Hubert, L. J., & Golledge, R. G. (1981). A heuristic method for the comparison of related structures. *Journal of Mathematical Psychology, 23,* 214–226.
Hubert, L. J., & Schultz, J. R. (1976). Quadratic assignment as a general data analysis strategy. *British Journal of Mathematical and Statistical Psychology, 29,* 190–241.
Hubert, L. J., & Subkoviak, M. J. (1979). Confirmatory inference and geometric models. *Psychological Bulletin, 86,* 361–370.
Hutchinson, J. W. (1981). *Network representations of psychological relations.* Unpublished doctoral dissertation, Stanford University.
Hutchinson, J. W. (1989). NETSCAL: A network scaling algorithm for nonsymmetric proximity data. *Psychometrika, 54,* 25–51.
Indow, T. (1983). An approach to geometry of visual space with no a priori mapping functions: Multidimensional mapping according to Riemannian metrics. *Journal of Mathematical Psychology, 26,* 204–236.
Indow, T. (1995). Psychophysical scaling: Scientific and practical applications. In R. D. Luce, M. D'Zmura, D. D. Hoffman, G. Iverson, & A. K. Romney (Eds.), *Geometric representations of perceptual phenomena* (pp. 1–28). Mahwah, NJ: Erlbaum.
Johnson, R. M. (1975). A simple method for pairwise monotone regression. *Psychometrika, 40,* 163–168.
Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika, 32,* 241–254.
Joly, S., & Le Calvé, G. (submitted). Realisable 0–1 matrices and city block distance.
Joly, S., & Le Calvé, G. (1995). Three-way distances. *Journal of Classification, 12,* 191–205.
Jones, L. E., & Koehly, L. M. (1993). Multidimensional scaling. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 95–163). Mahwah, NJ: Erlbaum.
Jones, W. P., & Furnas, G. W. (1987). Pictures of relevance: A geometric analysis of similarity measures. *Journal of the American Society for Information Science, 38,* 420–442.
Keller, J. B. (1962). Factorization of matrices by least-squares. *Biometrika, 49,* 239–242.
Kiers, H. A. L. (1990). Majorization as a tool for optimizing a class of matrix functions. *Psychometrika, 55,* 417–428.
Kiers, H. A. L., & Takane, Y. (1994). A generalization of GIPSCAL for the analysis of nonsymmetric data. *Journal of Classification, 11,* 79–99.
Kiers, H. A. L., & ten Berge, J. M. F. (1992). Minimization of a class of matrix trace functions by means of refined majorization. *Psychometrika, 57,* 371–382.
Klauer, K. C., & Carroll, J. D. (1989). A mathematical programming approach to fitting general graphs. *Journal of Classification, 6,* 247–270.
Klauer, K. C., & Carroll, J. D. (1991). A comparison of two approaches to fitting directed graphs to nonsymmetric proximity measures. *Journal of Classification, 8,* 251–268.
Klauer, K. C., & Carroll, J. D. (1995). Network models for scaling proximity data. In R. D. Luce, M. D'Zmura, D. Hoffman, G. J. Iverson & A. K. Romney (Eds.), *Geometric representations of perceptual phenomena* (pp. 319–342). Mahwah, NJ: Erlbaum.
Krijnen, W. P. (1993). *The analysis of three-way arrays by constrained PARAFAC methods.* Leiden: DSWO Press.
Kroonenberg, P. M. (1983). *Three-mode principal component analysis: Theory and applications.* Leiden: DSWO Press.
Kroonenberg, P. M., & de Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika, 45,* 69–97.
Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review, 85,* 445–463.

Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29,* 1–27.

Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika, 29,* 115–129.

Kruskal, J. B. (1965). Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society, Series B, 27,* 251–263.

Kruskal, J. B. (1976). More factors than subjects, tests and treatments: An interdeterminacy theorem for canonical decomposition and individual differences scaling. *Psychometrika, 41,* 281–293.

Kruskal, J. B., & Carroll, J. D. (1969). Geometrical models and badness-of-fit functions. In P. R. Krishnaiah (Ed.), *Multivariate analysis II* (pp. 639–671). New York: Academic Press.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling.* Newbury Park, CA: Sage.

Kruskal, J. B., Young, F. W., & Seery, J. B. (1973). *How to use KYST, a very flexible program to do multidimensional scaling and unfolding.* Murray Hill, NJ: AT&T Bell Laboratories.

Krzanowski, W. J., & Marriott, F. H. C. (1994). *Multivariate analysis. Part 1: Distributions, ordination and inference.* New York: Wiley.

Lance, G. N., & Williams, W. T. (1967). A general theory of classificatory sorting strategies. I. Hierarchical systems. *Computer Journal, 9,* 373–380.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104,* 211–240.

Lebart, L., Morineau, A., & Warwick, K. M. (1984). *Multivariate descriptive statistical analysis: Correspondence analysis and related techniques for large matrices* (E. M. Berry, Trans.). New York: Wiley. (Original work published 1977).

Lee, S.-K., & Bentler, P. M. (1980). Functional relations in multidimensional scaling. *British Journal of Mathematical and Statistical Psychology, 33,* 142–150.

Levelt, W. J. M., van de Geer, J. P., & Plomp, R. (1966). Triadic comparisons of musical intervals. *British Journal of Mathematical and Statistical Psychology, 19,* 163–179.

Lingoes, J. C., & Borg, I. (1978). A direct approach to individual differences scaling using increasingly complex transformations. *Psychometrika, 43,* 491–519.

Littman, L., Swayne, D. F., Dean, N., & Buja, A. (1992). Visualizing the embedding of objects in Euclidian space. In *Computing science and statistics: Proceedings of the 24th symposium on the interface* (pp. 208–217). Fairfax Station, VA: Interface Foundation of North America.

Lockhead, G. R., & Pomerantz, J. R. (Eds.) (1991). *The perception of structure.* Arlington, VA: American Psychological Association.

Luce, R. D., & Krumhansl, C. L. (1988). Measurement, scaling, and psychophysics. In R. C. Atkinson, R. J. Herrnstein, G. Lindzey, & R. D. Luce (Eds.), *Stevens' handbook of experimental psychology* (pp. 3–74). New York: Wiley.

MacCallum, R. C. (1977). Effects of conditionality on INDSCAL and ALSCAL weights. *Psychometrika, 42,* 297–305.

Marley, A. A. J. (1992). Developing and characterizing multidimensional Thurstone and Luce models for identification and preference. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 299–333). Mahwah, NJ: Erlbaum.

McDonald, R. P. (1976). A note on monotone polygons fitted to bivariate data. *Psychometrika, 41,* 543–546.

Meulman, J. J. (1992). The integration of multidimensional scaling and multivariate analysis with optimal transformations. *Psychometrika, 57,* 539–565.

Meulman, J. J., & Heiser, W. J. (1993). Nonlinear biplots for nonlinear mappings. In O. Opitz, B. Lausen, & R. Klar (Eds.), *Information and classification* (pp. 201–213). New York: Springer-Verlag.

Meulman, J. J., & Verboon, P. (1993). Points of view analysis revisited: Fitting multidimensional structures to optimal distance components with cluster restrictions on the variables. *Psychometrika, 58,* 7–35.

Miller, K. F. (1987). Geometric methods in developmental research. In J. Bisanz, C. J. Brainerd, & R. Kail (Eds.), *Formal methods in developmental psychology* (pp. 216–262). New York: Springer-Verlag.

Mirkin, B. (1996). *Mathematical classification and clustering.* Dordrecht: Kluwer.

Mirkin, B. G., & Muchnik, I. (1996). Clustering and multidimensional scaling in Russia (1960–1990): A review. In P. Arabie, L. J. Hubert, & G. De Soege (Eds.), *Clustering and classification* (pp. 295–339). River Edge, NJ: World Scientific.

Molenaar, I. W. (1986). Deconfusing confusion matrices. In J. de Leeuw, W. Heiser, J. Meulman, & F. Critchley (Eds.), *Multidimensional data analysis* (pp. 139–145). Leiden: DSWO Press.

Murtagh, F. (Ed.). (1997). *Classification Literature Automated Search Service, 26.*

Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications.* Toronto: University of Toronto Press.

Nishisato, S. (1993). *Elements of dual scaling: An introduction to practical data analysis.* Mahwah, NJ: Erlbaum.

Nishisato, S. (1996a). An overview and recent developments in dual scaling. In W. Gaul & D. Pfeifer (Eds.), *From data to knowledge* (pp. 73–85). Heidelberg: Springer-Verlag.

Nishisato, S. (1996b). Gleaning in the field of dual scaling. *Psychometrika, 61,* 559–599.

Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology, 23,* 94–140.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology, 43,* 25–53.

Okada, A. (1990). A generalization of asymmetric multidimensional scaling. In M. Schader & W. Gaul (Eds.), *Knowledge, data and computer-assisted decisions* (pp. 127–138). Heidelberg: Springer-Verlag.

Okada, A., & Imaizumi, T. (1980). Nonmetric method for extended INDSCAL model. *Behaviormetrika,* 7, 13–22.

Okada, A., & Imaizumi, T. (1987). Nonmetric multidimensional scaling of asymmetric proximities. *Behaviormetrika, 21,* 81–96.

Okada, A., & Imaizumi, T. (1994). *pasokon tajigen shakudo kouseihou* [*Multidimensional scaling using a personal computer*]. Tokyo: Kyoritsu Shuppan.

Okada, A., & Imaizumi, T. (1997). Asymmetric multidimensional scaling of two-mode three-way proximities. *Journal of Classification, 14,* 195–224.

Pan, G. C., & Harris, D. P. (1991). A new multidimensional scaling technique based upon association of triple objects-Pijk and its application to the analysis of geochemical data. *Journal of Mathematical Geology, 23,* 861–886.

Pliner, V. (1996). Metric unidimensional scaling and global optimization. *Journal of Classification, 13,* 3–18.

Poole, K. T. (1990). Least squares metric, unidimensional scaling of multivariate linear models. *Psychometrika, 55,* 123–149.

Pruzansky, S. (1975). *How to use SINDSCAL: A computer program for individual differences in multidimensional scaling.* Murray Hill, NJ: AT&T Bell Laboratories.

Pruzansky, S., Tversky, A., & Carroll, J. D. (1982). Spatial versus tree representations of proximity data. *Psychometrika, 47,* 3–24.

Ramsay, J. O. (1977a). Monotonic weighted power transformations to additivity. *Psychometrika, 42,* 83–109.

Ramsay, J. O. (1977b). Maximum likelihood estimation in multidimensional scaling. *Psychometrika, 42,* 241–266.

Ramsay, J. O. (1978a). Confidence regions for multidimensional scaling analysis. *Psychometrika, 43,* 145–160.

Ramsay, J. O. (1978b). *MULTISCALE: Four programs for multidimensional scaling by the method of maximum likelihood.* Chicago: National Educational Resources.

Ramsay, J. O. (1980). Some small sample results for maximum likelihood estimation in multidimensional scaling. *Psychometrika, 45,* 139–144.

Ramsay, J. O. (1981). MULTISCALE. In S. S. Schiffman, M. L. Reynolds, & F. W. Young (Eds.), *Introduction to multidimensional scaling: Theory, method and applications* (pp. 389–405). New York: Academic Press.

Ramsay, J. O. (1982a). *MULTISCALE II manual.* Mooresville, IN: International Educational Services.

Ramsay, J. O. (1982b). Some statistical approaches to multidimensional scaling data [with discussion]. *Journal of the Royal Statistical Society A, 145,* 285–312.

Ramsay, J. O. (1983). MULTISCALE: A multidimensional scaling program. *American Statistician, 37,* 326–327.

Ramsay, J. O. (1988). Monotone splines in action. *Statistical Science, 3,* 425–441.

Rodieck, R. W. (1977). Metric of color borders. *Science, 197,* 1195–1196.

Rohlf, F. J., & Slice, D. (1990). Extensions of the Procrustes method for the optimal superimposition of landmarks. *Systematic Zoology, 39,* 40–59.

Rosenberg, S. (1982). The method of sorting in multivariate research with applications selected from cognitive psychology and person perception. In N. Hirschberg & L. Humphreys (Eds.), *Multivariate applications in the social sciences* (pp. 117–142). Mahwah, NJ: Erlbaum.

Rosenberg, S., & Kim, M. P. (1975). The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research, 10,* 489–502.

Rounds, J., & Tracey, T. J. (1993). Prediger's dimensional representation of Holland's RIASEC circumplex. *Journal of Applied Psychology, 78,* 875–890.

Rounds, J., Tracey, T. J., & Hubert, L. (1992). Methods for evaluating vocational interest structural hypotheses. *Journal of Vocational Behavior, 40,* 239–259.

Saito, T. (1991). Analysis of asymmetric proximity matrix by a model of distance and additive terms. *Behaviormetrika, 29,* 45–60.

Saito, T. (1993). Multidimensional scaling for asymmetric proximity data. In R. Steyer, K. F. Wender, & K. F. Widaman (Eds.), *Psychometric methodology* (pp. 451–456). Stuttgart: Gustav Fischer.

Saito, T., & Takeda, S. (1990). Multidimensional scaling of asymmetric proximity: Model and method. *Behaviormetrika, 28,* 49–80.

Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika, 42,* 319–345.

Sattath, S., & Tversky, A. (1987). On the relation between common and distinctive feature models. *Psychological Review, 94,* 16–22.

Shepard, R. N. (1962a). The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika, 27,* 125–140.

Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function. II. *Psychometrika, 27,* 219–246.

Shepard, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology, 1,* 54–87.

Shepard, R. N. (1972). A taxonomy of some principal types of data and of multidimensional methods for their analysis. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences: Vol. I. Theory* (pp. 24–47). New York: Seminar Press.

Shepard, R. N. (1974). Representation of structure in similarity data: Problems and prospects. *Psychometrika, 39,* 373–421.

Shepard, R. N. (1978). The circumplex and related topological manifolds in the study of perception. In S. Shye (Ed.), *Theory construction and data analysis in the behavioral sciences* (pp. 29–80). San Francisco: Jossey-Bass.

Shepard, R. N. (1987). Toward a universal law of generalization. *Science, 237,* 1317–1323.
Shepard, R. N. (1988). Toward a universal law of generalization [Letter to editor]. *Science, 242,* 944.
Shepard, R. N., & Arabie, P. (1979). Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review, 86,* 87–123.
Shiina, K. (1986). A maximum likelihood nonmetric multidimensional scaling procedure for word sequences obtained in free-recall experiments. *Japanese Psychological Research, 28* (2), 53–63.
Shoben, E. J., & Ross, B. H. (1987). Structure and process in cognitive psychology using multidimensional scaling and related techniques. In R. R. Ronning, J. A. Glover, J. C. Conoley, & J. C. Witt (Eds.), *The influence of cognitive psychology on testing* (pp. 229–266). Mahwah, NJ: Erlbaum.
Srinivasan, V. (1975). Linear programming computational procedures for ordinal regression. *Journal of the Association for Computing Machinery, 23,* 475–487.
Steenkamp, J.-B. E. M., van Trijp, H. C. M., & ten Berge, J. M. F. (1994). Perceptual mapping based on idiosyncratic sets of attributes. *Journal of Marketing Research, 31,* 15–27.
Stevens, S. S. (1972). A neural quantum in sensory discrimination. *Science, 177,* 749–762.
Suppes, P., Krantz, D. M., Luce, R. D., & Tversky, A. (1989). *Foundations of measurement: Vol. II. Geometrical, threshold, and probabilistic representations.* New York: Academic Press.
Takane, Y. (1981). MDSORT: A special-purpose multidimensional scaling program for sorting data. *Behavior Research Methods & Instrumentation, 13,* 698.
Takane, Y. (1982). The method of triadic combinations: A new treatment and its application. *Behaviormetrika, 11,* 37–48.
Takane, Y. (1989). Ideal point discriminant analysis and ordered response categories. *Behaviormetrika, 26,* 31–46.
Takane, Y., Bozdogan, H., & Shibayama, T. (1987). Ideal point discriminant analysis. *Psychometrika, 52,* 371–392.
Takane, Y., & Carroll, J. D. (1981). Nonmetric maximum likelihood multidimensional scaling from directional rankings of similarities. *Psychometrika, 46,* 389–405.
Takane, Y., & Sergent, J. (1983). Multidimensional scaling models for reaction times and same-different judgments. *Psychometrika, 48,* 393–423.
Takane, Y., & Shibayama, T. (1986). Comparison of models for stimulus recognition data. In J. de Leeuw, W. Heiser, J. Meulman, & F. Critchley (Eds.), Multidimensional data analysis (pp. 119–138, 147–148). Leiden: DSWO-Press.
Takane, Y., Young, F. W., & de Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika, 42,* 7–67.
Tansley, B. W., & Boynton, R. M. (1976). A line, not a space, represents visual distinctness of borders formed by different colors. *Science, 191,* 954–957.
Tansley, B. W., & Boynton, R. M. (1977). Letter in reply to R. W. Rodieck. *Science, 197,* 1196.
Tartter, V. C. (in press). *Language processes* (2nd ed.). Newbury Park, CA: Sage.
ten Berge, J. M. F. (1977). Orthogonal procrustes rotation for two or more matrices. *Psychometrika, 42,* 267–276.
ten Berge, J. M. F. (1988). Generalized approaches to the maxbet problem and the maxdiff problem, with applications to canonical correlations. *Psychometrika, 53,* 487–494.
ten Berge, J. M. F., Kiers, H. A. L., & Commandeur, J. J. F. (1993). Orthogonal Procrustes rotation for matrices with missing values. *British Journal of Mathematical and Statistical Psychology, 46,* 119–134.
ten Berge, J. M. F., & Knol, D. L. (1984). Orthogonal rotations to maximal agreement for two or more matrices of different column order. *Psychometrika, 49,* 49–55.

Tobler, W. (1979). Estimation of attractivities from interactions. *Environment and Planning A, 11,* 121–127.

Torgerson, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika, 17,* 401–419.

Torgerson, W. S. (1958). *Theory and methods of scaling.* New York: Wiley.

Tracey, T. J., & Rounds, J. (1993). Evaluating Holland's and Gati's vocational-interest models: A structural meta-analysis. *Psychological Bulletin, 113,* 229–246.

Tucker, L. R (1960). Intra-individual and inter-individual multidimensionality. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: Theory and Applications* (pp. 155–167). New York: Wiley.

Tucker, L. R (1964). The extension of factor analysis to three-dimensional matrices. In N. Frederiksen & H. Gulliksen (Eds.), *Contributions to mathematical psychology* (pp. 109–127). New York: Holt, Rinehart, and Winston.

Tucker, L. R (1972). Relations between multidimensional scaling and three-mode factor analysis. *Psychometrika, 37,* 3–27.

Tucker, L. R, & Messick, S. J. (1963). An individual difference model for multi-dimensional scaling. *Psychometrika, 28,* 333–367.

Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327–352.

Van Cutsem, B. (Ed.). (1994). *Classification and dissimilarity analysis.* Heidelberg, Germany: Springer-Verlag.

Waller, N. G., Lykken, D. T., & Tellegen, A. (1995). Occupational interests, leisure time interests, and personality: Three domains or one? Findings from the Minnesota Twin Registry. In D. Lubinski & R. V. Dawis (Eds.), *Assessing individual differences in human behavior: New concepts, methods, and findings* (pp. 232–259). Palo Alto, CA: Consulting Psychologists Press.

Weeks, D. G., & Bentler, P. M. (1982). Restricted multidimensional scaling models for asymmetric proximities. *Psychometrika, 47,* 201–208.

Weinberg, S. L., & Carroll, J. D. (1992). Multidimensional scaling: An overview with applications in educational research. In B. Thompson (Ed.), *Advances in social science methodology* (pp. 99–135). Greenwich, CT: JAI Press.

Weinberg, S. L., Carroll, J. D., & Cohen, H. S. (1984). Confidence regions for INDSCAL using the jackknife and bootstrap techniques. *Psychometrika, 49,* 475–491.

Weinberg, S. L., & Menil, V. C. (1993). The recovery of structure in linear and ordinal data: INDSCAL versus ALSCAL. *Multivariate Behavioral Research, 28,* 215–233.

Weisberg, H. F. (1974). Dimensionland: An excursion into spaces. *American Journal of Political Science, 18,* 743–776.

Wilkinson, L. (1994). SYSTAT for DOS: *Advanced applications, Version 6 edition.* Evanston, IL: Systat.

Winsberg, S., & Carroll, J. D. (1989a). A quasi-nonmetric method for multidimensional scaling of multiway data via an extended INDSCAL model. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (pp. 405–414). Amsterdam: North-Holland.

Winsberg, S. & Carroll, J. D. (1989b). A quasi-nonmetric method of multidimensional scaling via an extended Euclidean model. *Psychometrika, 54,* 217–229.

Winsberg, S., & Ramsay, J. O. (1980). Monotonic transformations to additivity using splines. *Biometrika, 67,* 669–674.

Winsberg, S., & Ramsay, J. O. (1981). Analysis of pairwise preference data using B-splines. *Psychometrika, 46,* 171–186.

Winsberg, S., & Ramsay, J. O. (1983). Monotone spline transformations for dimension reduction. *Psychometrika, 48,* 575–595.

Wish, M., & Carroll, J. D. (1974). Applications of individual differences scaling to studies of human perception and judgment. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook* *of perception: Psychophysical judgment and measurement* (Vol. 2, pp. 449–491). New York: Academic Press.

Wold, H. (1966). Estimation of principal components and related models by iterative least squares. In P. R. Krishnaiah (Ed.), *Multivariate analysis* (pp. 391–420). New York: Academic Press.

Young, F. W. (1975). Methods for describing ordinal data with cardinal models. *Journal of Mathematical Psychology, 12,* 416–436.

Young, F. W. (1984a). The general Euclidean model. In H. G. Law, C. W. Snyder, Jr., J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 440–469). New York: Praeger.

Young, F. W. (1984b). Scaling. *Annual Review of Psychology, 35,* 55–81.

Young, F. W., & Lewyckyj, R. (1981). *ALSCAL.4 user's guide.* Unpublished manuscript, L. L. Thurstone Psychometric Laboratory, University of North Carolina, Chapel Hill.

Zielman, B. (1993). Directional analysis of three-way skew-symmetric matrices. In O. Opitz, B. Lausen, & R. Klar (Eds.), *Information and classification* (pp. 156–161). New York: Springer-Verlag.

Zielman, B., & Heiser, W. J. (1993). Analysis of asymmetry by a slide-vector. *Psychometrika, 58,* 101–114.

Zielman, B., & Heiser, W. J. (1994). *Models for asymmetric proximities.* Internal Report RR-94-04. Leiden: Department of Data Theory, University of Leiden.