Constrained Multidimensional Scaling, Including Confirmation

Willem J. Heiser and Jacqueline Meulman
University of Leiden

Constrained and confirmatory multidimensional scaling (MDS) are not equivalent. Constraints refer to the translation of either theoretical or data analytical objectives into computational specifications. Confirmation refers to a study of the balance between systematic and random variation in the data for modeling of the systematic part. Among the topics discussed from this perspective are the role of substantive theory in MDS studies, the type of constraints currently envisaged, and the relationships with other data analysis methods. This paper points out the possibility of using either sampling models or resampling schemes to study the stability of MDS solutions. Parallel to Akaike's (1974) information criterion for choosing one out of many models for the same data, a general stability criterion is proposed and illustrated, based on the ratio of within to total spread of configurations issued from resampling.

1. Scaling Analysis and Optimization

The coupling of constraints and scaling, although not new, is still extremely vital. The first modern scaling model, developed by Thurstone (e.g., 1927), immediately entailed the necessity of a judicious choice of constraints, as classified in the five famous "cases." These were internal constraints in the sense that the modified model would still accommodate the same set of observations, possibly less accurately, but with more predictive value. Guttman (1946) took a number of important additional steps. He showed that Thurstone's method, relying on an underlying multivariate normal distribution, could be replaced by an approach in which the optimization of a goodness-of-fit function is the central concept. This was an extension of similar developments occurring at that time in test theory and in contingency table analysis (for current reviews, see Nishisato, 1980, and Gifi, 1981) to the scaling of paired comparison and rank order data. Moreover, Guttman showed that introducing an additional type of constraint was not a problem. The proper objective would be to find scale values for the stimulus objects that are "best able to reproduce the judgment of each person in the population on each comparison" (Guttman, 1946, p. 145) and that are also perfectly linearly predictable from a predetermined set of (design) variables.

Although through hindsight it seems clear that it would have been possible to formulate a constrained multidimensional scaling (MDS) method within Guttman's framework, the early paradigm for MDS was to split the problem into *two distinct* steps. From the very first published psychometric application by Klingberg (1941)—based on earlier work by Richardson (1938)—to the authoritative presentation in Torgerson's (1958) book, it is evident that the main burden and first step of the analysis was to find a one-dimensional scale of psychological distances between pairs of stimulus objects. This was usually done by any of the var-

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 7, No. 4, Fall 1983, pp. 381–404

0146-6216/83/040381-24$2.45

ious Thurstonean scaling methods and preferably included a statement on the reliability of those distance estimates, e.g., by reporting the correlation between distance values obtained for two randomly selected subgroups of judges. In modern terminology, most effort was directed towards obtaining a good two-way matrix of dissimilarities. Granted that they were reliable and truly distance-like numbers (which could to some extent be ensured by selecting a proper zero-point of the scale), the dissimilarities were next submitted to the procedure described by Young and Householder (1938) in order to construct a spatial model of the stimulus objects. It was not until the well-known Shepard-Kruskal breakthrough (Kruskal, 1964; Shepard, 1962) that this second step was seriously challenged. In fact, conceptually, the breakthrough can be characterized as rendering the first step quite irrelevant ("given a set of similarities [or dissimilarities] between pairs of objects") and transforming the second step into an interesting problem of independent interest and wide applicability ("optimize a goodness-of-fit function which measures how well any given configuration matches any given transformation of the data").

Formulating MDS primarily as an optimization task has definite advantages. It offers the possibility of developing and using related computational schemes for what seems at first to be unrelated problems emerging from different fields. It provides a language for communication and a set of quality criteria so that insights can more easily be shared and accumulated. Whereas in the classical approach there are only very limited ways to prescribe specifically how the spatial model should look, prescriptions of various sorts can rather easily be translated into constraints on the optimization task. The reverse of this is that the analysis of one single (dis)similarity matrix becomes the focal point, even if these "data" were obtained by preliminary computations on replicated observations. This, in turn, unduly limits the possibility of computing some sort of *secondary statistic* (i.e., a number indicating how variable, or perhaps how stable, the resulting configuration seems to be) or other assessments of "what might have been." Moreover, whenever a fair amount of data had been collected for each of a number of subjects or other data sources, large systematic individual differences frequently showed up, urging recourse to a three-way model, which describes the effects of both stimuli and subjects and again complicates the assessment of random variability.

Does this imply that MDS is bound to be exploratory? If MDS is defined as a class of optimization problems, this is the wrong question to ask, in much the same way it cannot be discerned whether or not the computation of a smooth curve through a number of points is merely exploratory. However, if MDS is used in the course of analyzing data, certainly *confirmatory actions* can be taken. Then, of course, what it is that has to be confirmed needs to be specified more precisely. Sometimes it is desirable to make sure that what is seen is not method specific but emerges invariably under different treatments of the data. Sometimes it is desirable to stay within the realm of a distance model but to require that other pieces of information should be coherent with it. Sometimes there may even be a specific theory which can be translated into geometric terms.

A major aim of this paper is to contribute to the clarification of the methodological status of multidimensional scaling. Section 2 will distinguish three broad approaches towards the justification of applying MDS to real, substantive problems. They are called model-centered, data-centered, and content-centered; their distinctness is somewhat exaggerated, and many of the arguments are certainly not MDS specific. The role of a substantive theory is appreciated quite differently in these three approaches, but there is general agreement on the idea that constraints should play an important part in the analysis. Then, Section 3 sketches a provisional classification of constraints, based on their form and on the place where they enter the model. Most, if not all, recent proposals can be fitted into this, admittedly rather pragmatic, scheme. Section 4 discusses some interesting relationships with well-known objectives in multivariate analysis, thus pushing MDS beyond its conventional territory. Section 5 is concerned with the question of how to evaluate the results obtained from a scaling program—either constrained or unconstrained—and argues that in the absence of genuine replications, it is especially useful to examine (1) the performance of the technique by extensive plotting and (2) a comparison of trends against standard cases. If there are replications, a statistical stability argument may be invoked: What seems to be the case can be confirmed by examining what would have been the case if the sample were slightly different. Facing a choice among two or more specifications of the model, the more stable one should be selected. This is illustrated in Section 6 with the aid of the so-called *Bootstrap* technique, which can be thought of as systematically and efficiently carrying through the split-half reliability idea mentioned before for the first stage only. Ramsay's (1977, 1978) maximum likelihood method will be characterized as a radically smoothed version of the Bootstrap, in which he uses a shortcut to inference by assuming exact log-normality.

Throughout this paper it will be assumed that the reader is familiar with the elementary concepts of MDS, at the level of Kruskal and Wish (1978), Schiffman, Reynolds, and Young (1981), or Davison (1983). Some insights from linear algebra (in Sections 3 and 4) and from elementary statistics (in Sections 5 and 6) are also employed. For more technical and specialized reviews, the reader is referred to De Leeuw and Heiser (1982) and to Ramsay (1982).

2. Substantive Theory: Necessary, Convenient or Immaterial?

The Illusion of Transparency

The primary aim of an MDS method is to give a *full account* (or "reproduction") of the data. It is this conceptual characteristic that in many ways has caused controversy about the methodological status of an MDS analysis. It would give, in principle at least, a completely *tautological mapping* of the data; with a nonmetric technique complete reproduction can always be achieved provided a large enough dimensionality is chosen. Of course, in practice there is a tendency to avoid high dimensionality by being satisfied with a good approximation rather than by trying to get perfect reproducibility. However, Guttman (1977) is quite correct in insisting that the incumbent hypothesis in the social sciences should be to the effect that the data are very complex. Therefore, it is very surprising, indeed, that in most published applications of MDS, two- or three-space solutions prove to be enough to account for the data. In some cases this may have been overoptimistic (cf. Chang & Carroll, 1980, who argued that color space is best described in seven dimensions), and it is still somewhat unclear why this parsimony of description is attained in such a wide range of circumstances (not necessarily by nonmetricity, cf. Weeks & Bentler, 1979).

As a consequence, however, MDS has attained the trademark of *apparent transparency*. Here is a sophisticated visualization technique that maps—essentially without throwing away important information—a chaos of numbers and relations into a form more readily accessible to the human mind: Euclidean two- or three-space. As soon as there is access to this mapping, there is a strong tendency to assume that what is seen must capture precisely what there is to capture, or maybe even that the picture has been "underlying" the data. Just as most psychologists are aware of the selectivity of the human perceptual machinery and know that it imposes internal rules upon whatever there is to perceive, most MDS specialists agree that a picture may be worth a thousand numbers but that the thousand numbers may not be worth the picture. Transparency may be an illusion. However, they seriously disagree in their recommendations on how to appreciate the situation and what to do about it. Consequently, there are a variety of answers to Schönemann and Borg's (1981) rhetorical question: "What possible justification can there be for seeking best-fitting numerical representations of the data, whether or not the assumed model is really appropriate?" These answers can be grouped into three approaches that tend to exclude each other, at least in primary emphasis.

The Model-Centered Approach

In the spirit of classical statistics it can be said that initially there must be the assumption that the *structural model* is appropriate and, in addition, a *sampling* or *error* model must be specified that encompasses intuitions and experience about the manner in which every single observed dissimilarity varies about its fitted value. Ramsay (1977) has argued that a good all-round candidate for the distribution of a dissimilarity observation, given its errorless value, is the lognormal, in which the standard deviation is proportional to the mean. If, in addition, independence is assumed, then a best-fitting numerical representation can be found by maximum likelihood methods.

Note that in this approach it is absolutely essential to have more than one dissimilarity matrix and that a purely nonmetric analysis is no longer possible. As Ramsay (1977, p. 242) has remarked, "the nonmetric multidimensional scaling procedures suffer in general from the degrees of freedom used up in fitting the very general monotone transformations, which they employ, to the dependent variable. Such transformations ignore such considerations as smoothness, and the fact that certain types of data will almost certainly require only very specialized transformations." In a similar vein detailed assumptions can be made about specific data collection methods in order to extend the reach of the model to the process that supposedly generated the dissimilarity information. This has been done for pairwise comparisons of pairs of stimuli (Takane, 1978a, 1978b), rating methods (Takane, 1981), directional ranking methods (Takane & Carroll, 1981), and the tree-construction method (Takane & Carroll, 1982). The work cited deviates from the usual Shephard-Kruskal framework by again integrating what was called the first and second step in Section 1.

Maximum likelihood methods derive their strength from an a priori probabilistic theory about *response behavior;* they do not capitalize on substantive knowledge about the stimulus objects. Nevertheless, specific constraints on the position of the stimulus points fit naturally into this framework and are indeed welcomed, for they make the associated likelihood ratio tests potentially more powerful. In fact, a whole family of models might be considered, specified a priori, with that particular member selected that is most likely in the light of the data. There is never any certainty whether or not the appropriate family is considered, and, of course, some of the many possible members may be overlooked.

The Data-Centered Approach

The second approach is affiliated with the "data-analytical" outlook in statistics. Here, the answer to Schönemann and Borg's (1981) question would be something like, "If we want to find out what is going on, we have to look at the data, and at as many transformations of them as possible." *Both the expected and the unexpected* must be looked for. Tukey has emphasized repeatedly that methodologies have no assumptions and can deliver no certainties. Justification can be obtained only after a cyclic process of skillful trying and educated guessing. "Quantitative methodologies are often, perhaps usually, developed with a particular 'leading situation' in mind. . . . Their performance in such a situation—or class of situations—may have little to do with their practical usefulness, since the differences between leading situations and practical arenas are often large, if not catastrophic. If all we know about our methodologies is their leading situation behavior, we are truly ill informed" (Tukey, 1980, p. 493).

Thus there must be restraint from tuning the technique to specific situations. This does not at all imply that complete trust should be placed in general-purpose number crunching; it does imply that there must be a willingness to try out different techniques (different "families of models") on the same data, that techniques be kept as *flexible* as possible, and that as much attention should be paid to the unexpected patterns as to the expected ones by looking hard at a variety of pictures. Although whatever is known about the context in which the data were collected ("substantive theory") is welcome to *tune the analysis,* it is not to be trusted either. From this perspective, MDS is only one of the available tools in need of better diagnostic devices, and certainly not to be confused with a particular geometric model of psychological similarity (which was merely one of the leading situations that stimulated its development).

The data-centered approach tends to reject the use of sampling models as a general strategy because these are seen as approximations of unknown quality. For the study of random variability, it is advisable to use *resampling* techniques, such as the Jackknife and the Bootstrap (see Section 6).

The Content-Centered Approach

Rather than directing the majority of attention to a family of models or to a variety of tactics and techniques, the third, *content-centered,* approach would advise focusing fully on the design of the observational framework, whether it be experimental or, say, correlational. If it is ensured that the variability of the observed dissimilarities across pairs of objects is genuine (by carefully controlled experiments or large sample sizes in a correlational study), there need not be so much concern about *sampling error* (exact replication will give the same gross, irreducible variability). Instead, the problem of *approximation error* (i.e., error due to misspecification of the structural model) must be addressed. Best-fitting numerical representations are "blind" to substantive theory and therefore will almost surely *not* give an appropriate model.

The major contention of the content-centered approach is that an *explicit rationale* is needed for the structure found, not so much precisely for the MDS structure—it should be recognizable by other methods too, although perhaps in a more roundabout way—but for the systematic variability itself. Such a rationale would guarantee that upon replication with another stratified *sample of stimuli or items* from the same "universe of content," essentially the same regularity shows up. The loss function may be modified to incorporate these a priori ideas "to guide the computations, and thus obtain results more generalizable than those from a purely blind analysis. In any event, the configuration must be basically the same whether viewed in the smallest space for it, or in a larger space, since . . . [monotonicity] is required to be satisfied by metric as well as nonmetric solutions for the same data, or by content oriented calculations" (Guttman, 1968, p. 471–472). If the unrestricted and the restricted spaces are basically different, presumably puzzlement must be admitted. (For a quasi-statistical decision strategy towards model selection from this point of view, see Lingoes & Borg, 1983).

The set of principles and rules that Guttman and his coworkers have formulated in order to generate useful designs and hypotheses is called *facet theory*. This framework was primarily invoked to convince social scientists to use principles of experimental design for their correlational studies. (Of course, it is not necessary to convince experimental psychologists to do this: if anything, they must be convinced to use similarity as a dependent instead of an independent variable.) Tversky has stressed that regularities found in human judgments of similarity sometimes cannot be described *completely* by a simple geometric model because there can be a familiarity effect, making the similarities asymmetric, a coincidence effect, making them intransitive, or a variety of other effects (Gati & Tversky, 1982; Tversky & Gati, 1982). Thus he addresses the issue of approximation error from the other side: Tversky's (1977) *feature theory,* inspired by earlier work of Attneave (1950, 1962) and Restle (1959, 1961), provides an explicit rationale for constructing stimuli in such a way that their judged similarity will display certain patterns that are incompatible with distances in any space.

From Ideas to Constraints

Overlooking the three approaches, it seems that despite considerable differences in emphasis and spirit, it is agreed that substantive considerations can be used profitably to modify the loss function and that the best fitting representations arising from different loss functions should be somehow compared. What form should these modifications have?

It is important to realize that substantive theory can enter the problem in a variety of ways:

1. As a *specific structural theory,* translating psychological invariances and transformations into geometric terms. A good recent example of this is Shepard's (1982) derivation of the doubly helical structure of musical pitch.
2. As a *theory of response behavior,* indicating how subjects process their internal impression

of overall similarity into the choices and judgments that are observed. Earlier work in this area, as reviewed and elaborated in Krantz (1967), ended up somewhat in the background due to the upsurge of nonmetric MDS. The recent series of papers by Ramsay (e.g., 1977, 1978, 1982) and Takane (e.g., 1978a, 1978b, 1981; Takane & Carroll, 1981) may cause renewed interest in these matters.

3. As a *theory of overlap and interaction*. Typically, this takes the form of first giving a characterization of the objects under study in terms of physical or semantic aspects, facets, features, or attributes and then posing a hypothesis about the organization, integration, or redundancy of these characteristics explaining the pattern of similarity or association among the objects.
4. As a *theory of structural coherence*. Here are considered two or more sets of proximity relations arising from different homogeneous groups of subjects or occurring in different points in time, under different external conditions, and so forth. Alternatively, there can be consideration of two or more different empirical relations on the same set of objects, e.g., a simple ordering and an ordering on pairs. Then a hypothesis may be posed that specifies a correspondence between the various relations—not directly, but mediated by a psychological space that by itself need not yet be fully understood.

These different types of theory require different types of constraints on the optimization task. In the next section a classification of constraints that takes account of these various objectives will be given. In addition, it will turn out that even for any given idea about geometrical structure, there are often several ways for translation into constraints. This will be illustrated for the general notions of a *circular order* or circumplex and of *in-betweenness* of (groups of) points with respect to each other.

The former case was first considered as a formal structure in psychometrics by Thurstone (1947, p. 183–185) under the name "cone configuration," but its special status was articulated by Guttman (1954) and his coworkers in subsequent work. Circumplexes have been suggested in such diverse fields as mental ability testing (Guttman, 1957), perception (Shepard, 1978), and interpersonal behavior (Benjamin, 1974; Wiggins, Steiger, & Gaelick, 1981). The latter is somewhat more general and thus weaker; if the circumplex is characterized by saying that *every* object is somehow in between its closest neighbors (insistence upon "every" forces the structure to become closed and hollow), then it will be clear that in-betweenness alone, being a local condition, allows the specification of a great variety of special structures. It is a primitive concept in information integration theory (Anderson, 1981), and it is bound to be important in developmental psychology; but it could also be useful in more exotic applications of MDS, such as in aiding the design of chips. In this last application, the objects are the components of the chip, the proximities specify how close they should preferably be due to functional interdependence, and there are a number of special components that should *not* be in between the others because they serve to communicate with the outside world.

3. A Provisional Classification of Constraints

At Least Two Independent Aspects

The core of the modeling process in any MDS method can be displayed in the following fashion:

$$\delta_{ij} \xrightarrow{\text{transformation}} \hat{d}_{ij} \overset{\text{fit}}{—} d_{ij} \xrightarrow{\text{distance function}} X$$

where for the moment the fact is ignored that frequently the dissimilarities δ_{ij} are obtained by some aggregation process on the raw data. For any choice of distance function, fit function, and class of transformations a different MDS method is obtained. These are the global choices, and the modeling process can be adjusted by placing constraints on the *coordinates* X of the points, on the *distances* d_{ij} among the points, or on the *transformation* that maps the dissimilarities into the disparities $\hat{d}_{ij}$. This constitutes the first, and most important, aspect of the classification: the *place in the model*.

There are a number of ways to refine the classification. A very fruitful one, which refers to both substantive considerations and mathematical form, is to distinguish between *equality* constraints (i.e., things should be identified with each other) and *inequality* constraints (i.e., one thing should dominate another). Calling this a distinction in *form*, and crossing form and place, gives the classes of constraints exemplified in the cells of Table 1. It should be pointed out that these are logical classes; most of them have not yet been implemented in generally available programs. In the remaining parts of this section, all six cells will be considered in turn, where the upper right one will get most attention.

Constraints on the Coordinates

Probably the oldest type of constraints in the field are linear equalities on the coordinates. In programs like POLYCON (Young, 1972) it was already possible to fix coordinates to some known value; and in the INDSCAL model the individual coordinates are constrained to be dimensionwise equal up to a scale parameter. Bloxom (1978) has given a generalized framework. Focus in this section will first be on his formulation. (In the discussion section of his paper Bloxom has carefully pinpointed the partial overlap with other classes of equality constraints proposed by Bentler and Weeks, 1978; Green, Carroll, and Carmone, 1976; and Carroll, Pruzansky, and Kruskal, 1980; so these intricacies need not be repeated here.)

The constraints considered by Bloxom have the form

$$\mathbb{X}_s = \mathbb{U}_s\mathbb{V}_s \qquad [1]$$

where $\mathbb{X}_s$ is the $I \times K$ matrix of stimulus coordinates for the sth replication; the parameters in $\mathbb{U}_s$ and $\mathbb{V}_s$ can be either fixed or free, and some free parameters can be restricted to be equal to others. If $\mathbb{U}_1 = \ldots = \mathbb{U}_s = \mathbb{U}$, and all $\mathbb{V}_s$ are constrained to be diagonal, this results in the INDSCAL model; if $\mathbb{V}_1 = \ldots = \mathbb{V}_s = \mathbb{V}$ and all $\mathbb{U}_S$ are constrained to be diagonal, this results in the "vector weighting" model included in the Lingoes and Borg (1978) PINDIS family of models. Other individual differences models can be specified similarly.

For convenience, other special cases will be discussed only in an unreplicated context, i.e.,

$$\mathbb{X} = \mathbb{U}\mathbb{V} \qquad [2]$$

because their implementation in a three-way context is straightforward. Suppose $\mathbb{U}$ is fixed to a given set of outside variables, then Equation 2 states that the configuration should be equal to a linear combination of the given set. The variables in $\mathbb{U}$ can be quantitative, as when there are independently obtained rating scales or physical measurements for the objects. Note that Equation 2 does not describe a one-to-one relationship between specific outside variables and dimensions; as long as $\mathbb{V}$ is not further restricted, it is simply necessary to find the K linear combinations of the columns of $\mathbb{U}$ that perfectly predict $\mathbb{X}$, and thus the distances between the points, which in turn describe the var-

Table 1
Types of Constraints, Cross-Classified by Form and Place in the Model

Type of Constraint	Transformations	Distances	Coordinates
Equality	Special types of metric MDS, e.g., negative exponential transformation.	Fitting one-dimensional distances or equal eccentricity.	Fixing or equalizing points on some axes or placing them on curves or surfaces.
Inequality	Special types of nonmetric MDS, e.g., smoothed monotone transformation.	Regular patterns, e.g., block structure or free tree distances.	Order constraints on the axes or limited eccentricity.

iation among the observed proximities optimally. If in addition $\mathbf{V}$ is restricted to be diagonal, then each dimension of $\mathbf{X}$ will correspond to exactly one outside variable.

$\mathbf{U}$ can also be chosen as a binary table specifying the design of the objects (rows) in terms of features (columns). As an illustration, consider Table 2a representing a 3×3 crossed factorial design. A factor in a design partitions the objects into mutually exclusive groups that share the same feature (level). Each level of a factor is represented by a vector (a row of $\mathbf{V}$), and the combined effect of two levels from different factors is obtained by adding the corresponding vectors. The geometrical result of these operations is shown in Figure 1a. Differences between the levels of a factor are represented by translations; distances between points that are equivalent in one level remain invariant when moving to another level of the same factor (compare A, B, C and D, E, F or A, D, G and B, E, H).

The structural relationships within levels are not necessarily simple; parallel but irregular regions are obtained. It might be preferable to obtain regions that are somehow more homogeneous, because there is a tendency to think about a factor as a unit, as a single direction in space. One way of accomplishing this is to restrict part of the matrix $\mathbf{V}$ to be zero, as shown in Table 2b. Now each factor corresponds to one dimension, and objects within the same level are homogeneous with respect to that dimension; the points must form a regular grid (see Figure 1b).

How can the idea that certain points should be in between others be translated into equality constraints? Suppose the set of objects can be partitioned into two subsets, S_1 and S_2, and it is desired to obtain points for the objects in S_1 that are in between selected points corresponding to the objects in S_2. (This approach was taken by Van der Kloot and Van den Boogaard, 1978, for single and combined traits in a person-perception experiment, and by Heiser, 1981, for the row and column objects of an unfolding analysis). $\mathbf{X}$ in Equation 2 can be identified with the coordinates of the S_1 points; $\mathbf{V}$, with the coordinates of the S_2 points; and $\mathbf{U}$, with a matrix of fixed nonnegative coefficients, whose rows sum to one. The coefficients may reflect any kind of weights attached to the S_2 points in the spatial model.

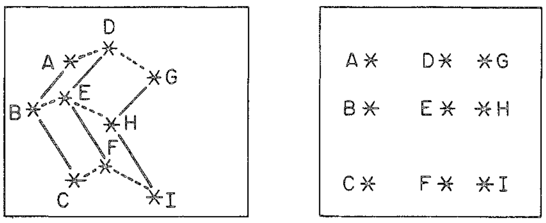
How about a circumplex within this framework? Again, a special fixed binary matrix $\mathbf{U}$ and a $\mathbf{V}$ that is restricted to be diagonal (De Leeuw & Heiser, 1980) must be used. Together, this forces the configuration matrix $\mathbf{X}$ to have the form given in Table 2c for the case of five stimulus objects. Note there are as many dimensions as there are objects, yet simple structure prevails. The squared Euclidean distances between the rows of Table 2c can be displayed as the *shortest path-length* distances in a so-called circular graph (see Figure 2). The distance between two nonadjacent nodes is simply the sum of the values of the links constituting the shortest path connecting them (e.g., for A to D this sum is $v_5^2 + v_3^2 + v_4^2 + v_2^2$, which equals the squared

Table 2
Special Equality Constraints: (a) U for a Cross-Classified Design, (b) Further Constraints on V to Obtain a Grid, (c) Combined Structure of the Free Parameters of a Circumplex

	(a)						(b)			(c)				
A	1			1			v_1	0	A	v_1	v_2	v_3	0	0
B	1				1		v_2	0	B	0	v_2	v_3	v_4	0
C	1					1	v_3	0	C	0	0	v_3	v_4	v_5
D		1		1			0	v_4	D	v_1	0	0	v_4	v_5
E		1			1		0	v_5	E	v_1	v_2	0	0	v_5
F		1				1	0	v_6						
G			1	1										
H			1		1									
I			1			1								

Figure 1
Spatial Representation of a Factorial Design: (a) Vector Additivity (Translations into Different Directions) and (b) Intra-Dimensional Homogeneity (Translations along the Axes Only)

(a) (b)



Euclidean distance between Rows 1 and 4 of Table 2c). Thus, a nonmetric MDS program (in full dimensionality but with appropriate constraints) can be used to fit the path-length distances in a circular graph in which the order of the stimuli is fixed in advance.

This principle carries over to other discrete structures specified in a binary matrix, but detailed discussion of these would be outside the scope of this paper. Table 1 also mentions placement of points on *curves or surfaces*. In the simplest case a decomposition of the form of Equation 2 is required, with $\mathbb{U}$ and $\mathbb{V}$ free but of column (row) order $R < K$. This places the points on an R-dimensional flat or subspace, and if $R = 1$ on a line. More general, nonlinear functional relations can also be specified (De Leeuw & Heiser, 1980; Lee & Bentler, 1980). The hypothesis of a circular order among the objects now will take the form of the requirement that the points should be on a circle. Notice the important difference with the implementation via Table 2c: It is no longer necessary to specify the order along the circumplex in advance. The line requirement and the circle requirement will be combined in an application in Section 6, where a capital Q has been fit to a set of color data.

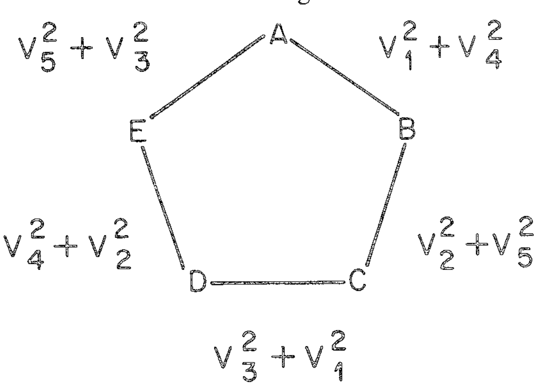
Inequality constraints on the axes can be motivated in many ways. More will be said about them in Section 4. For now, consider the following example of the in-betweenness notion. Glushko (1975) used the concept that the *redundancy of visual patterns* should be characterized by the number of distinct alternatives obtained under rigid rotations in 90-degree steps or reflections of the original patterns. According to this definition, a circle or a cross is highly redundant because both remain perceptually equivalent under the indicated set of transformations. Glushko partitioned his visual patterns into three groups: highly redundant, very nonredundant, and a group in between. It need not be assumed that such a partitioning will explain *all* the variation among the observed proximities—capitalization on such tentative and unrefined measures is not warranted. There is also no certainty that MDS will reveal the presupposed order *automatically*.

Thus, the coordinate values on the first axis can be constrained to be of the form

$$x_{\alpha 1} \leqslant x_{\beta 1} \leqslant x_{\gamma 1}\,, \qquad [3]$$

where α, β, and γ denote objects from the three different groups and it is understood that Equation 3 should be true for all different choices of such triples. Even if redundancy would not be a strong explanatory variable, Equation 3 guarantees that it will emerge as long as there is one corresponding direction in the space that is needed to describe the proximities. If it is insisted that the presupposed order should be dominant, without going so far as to require unidimensionality, Equation 3 can be repeated for all axes; if it turns out to be justified, this will give the configuration the form of a chain of rectangular regions. In fact, Glushko's (1975) unconstrained MDS results already satisfy such multiple constraints nearly perfectly, which is of

Figure 2
Circumplex as a Circular Graph with Path-Length Metric



course strong support for the primacy of his redundancy measure.

Finally, an example of more complex inequality constraints is worth mentioning. A subset of the points can be forced to be of limited eccentricity by restricting the sum of squares of their coordinates. This would be helpful in models that predict radial expansion of complexity (Guttman, 1954) or of saturation (as in the Munsell color system).

Constraints on the Distances

Borg and Lingoes (1980) have argued that instead of constraining coordinates, it may be useful to constrain the *distances* in an MDS problem. For example, separation between groups of varying redundancy can also be obtained by requiring that all within-group distances should be smaller than the distances between groups $\alpha - \beta$ and $\beta - \gamma$, which in turn should be smaller than the distances between the extreme groups. This application is discussed in Borg (1981), who also remarks that such a specification might be stronger than intended because it, in fact, requires the largest within-group distances to be smaller than the smallest between-group distances, causing a very compact clustering. It seems that this can be repaired on a very ad-hoc basis only, or by abandoning the separation concept altogether. Fitting a circular order in this approach can be done by introducing a dummy object that has missing proximities towards all other objects. Next, all distances from the dummy point towards the other points are required to be equal. (For an example using circular color data, see Borg & Lingoes, 1980, p. 30).

Similarly, points can be forced on a tree, a line, or a grid by selecting appropriate distance constraints. Frequently, a difficulty is encountered that is shared with the "binary coordinates" approach: The ordering of points in the structure must be known in advance. For instance, for three points on a line there must be a choice between

$$\begin{aligned} d_{ij} &= d_{ii'} + d_{i'j} \\ d_{ii'} &= d_{ij} + d_{ji'} \\ d_{ji'} &= d_{ji} + d_{ii'} \quad , \end{aligned} \qquad [4]$$

depending on which point is to be in between. In the case of a free tree, the so-called *four-point condition* (Buneman, 1971) can be used, which states that for all i, j, i', j' there must be

$$d_{ij} + d_{i'j'} \leqslant \max \{(d_{ii'} + d_{jj'}), (d_{ji'} + d_{ij'})\} \qquad [5]$$

but, again, it is not known beforehand which of the two sums on the right-hand side of Equation 5 will be "max." Once the correct selection of constraints is known, the remaining computational problem is not too difficult (Cunningham, 1978).

Constraints on the Transformation

The previous examples already suggest that distances are a more elementary concept than coordinates (although the latter serve many useful purposes). Indeed, in some applications, obtaining a configuration of points is not even the prime objective or the most interesting part of the MDS analysis. Most interest might be in the way observable response measures are related to psychological distance. For instance, in learning experiments the gradient of generalization (the function relating response strength to interstimulus distance) may change radically under different conditions of reinforcement. This was, in fact, one of the major concerns of Shepard's (1958) earlier work on MDS. Thus it might be appropriate to restrict the transformation to a one-parameter class of functions (e.g., the negative exponential $y = \exp(-ax)$ with discrimination parameter a) and to check a hypothesis on the change of the parameter value as an effect of learning.

Another major reason why *constrained transformations* are interesting is the fact that the arbitrary monotone functions that are relied upon in ordinary nonmetric MDS methods include completely flat ones and step functions with one step only. The possibility of such pathological transformations causes the occurrence of degeneracies in the configuration: points collapse into clusters and almost all variability information in the data gets lost (cf. Shepard, 1974). The problem is particularly acute when the stimulus objects are radically distinct rather than gradually connected or when dissimilarities are missing in a systematic way (such as in unfolding). Heiser and Meulman (1983) developed a procedure for *smoothed* monotone regression, which prevents the function from accelerating too strongly. This is an effective device against degeneracies without imposing strong demands on the shape of the transformation. Also relevant in this context is the work by Winsberg and Ramsay (1980, 1981), who introduced so-called B-spline transformations as a very flexible way to get various types of smooth monotonicity.

Prospects

It is hoped that the discussion of constraints has made clear that there is plenty of scope for action in MDS. Many of the ideas mentioned in this section deserve closer scrutiny in empirical contexts. Some of them provide challenging optimization problems; all of them can be combined discreetly by model builders. It has been emphasized that often a single vague notion of structure can be implemented in a number of ways. This is no cause for alarm as long as the ultimate interest is in the data, not in the model used to display them. Moreover, it is true that some notions of structure and approaches have an independent history and existence outside the class of MDS models. The next section will point out a few of these similarities.

4. Similarities with Objectives in Multivariate Analysis

The introduction of constraints in the MDS framework enables the formulation of an MDS counterpart for many of the general objectives that have become standard in multivariate analysis (MVA). No attempt will be made to give a full account of the various interrelationships that exist between these two classes of techniques, but some of the more obvious correspondences will be pointed out.

That there is a very close connection between principal components analysis (PCA) and classical MDS has been aptly pointed out by Gower (1966), and the various optimality properties of PCA plots in terms of reconstructed distances have been emphasized by Gabriel (1971). Recently, much effort has been directed towards the analysis of special structure in covariance matrices (Bentler, 1980; Jöreskog, 1978). If the distributional assumptions (which are invoked in this approach to guide the selection of certain special structures over others) are separated from the structural ideas themselves, there is much overlap with constrained MDS. The imposition of a circumplex, for example, amounts to the specification that the factor pattern matrix should have the form of Table 2c. Just as in the MDS approach with linear equality constraints, the order of the points along the circle must be known in advance. Without such an order, the detection of a circumplex by means of ordinary PCA must be done by carefully looking at the complete eigenvector decomposition (Wiggins et al., 1981) because the circular components need not be the ones that account for most variance. Yet ordinary distance models are preeminently suited to capture the nonlinearity immediately (Lingoes & Borg, 1977).

Comparing More Than Two Configurations

Another rich source for making comparisons between MVA and MDS is the "S sets matching" or generalized canonical correlation problem (Carroll, 1968; Horst, 1961; Kettenring, 1971; McKeon, 1968; Van de Geer, 1980) against three-way scaling. In ordinary canonical correlation there are two sets of variables, say $\mathbf{X}_1$ and $\mathbf{X}_2$, and the idea is to find a linear transformation of each, say $\mathbf{X}_1\mathbf{L}_1$ and $\mathbf{X}_2\mathbf{L}_2$, which renders them *maximally similar*. Maximum similarity can be translated as follows: There must exist a *common space* $\mathbf{X}$ (the canonical variates) such that

$$\mathbf{X} = \mathbf{X}_s\mathbf{L}_s \qquad [6]$$

for $s = 1, 2$. This may not be the most usual formulation of the problem, but it is the most convenient one for the purposes of this section. In the first place, if there are $S > 2$ sets, Equation 6 is still valid when $s = 1, \ldots, S$ is specified. This generalized problem is intrinsically more difficult than the two-set case. There are many ways to put identification constraints on $\mathbf{X}$ and the $\mathbf{L}_s$, and a variety of optimality criteria can be used to fit Equation 6 in case no perfect common space exists. However, if $S = 2$, most of these criteria become essentially identical (Gifi, 1981, chaps. 6 & 7).

Now suppose the linear transformations are required to be *orthogonal,* which implies that they can be split up into a rotation and a rescaling of axes. Equation 6 can be rewritten into

$$\mathbb{X} = \mathbb{X}_s\mathbb{T}_s\Phi_s \quad [7]$$

for $s = 1, \ldots, S$ with $\mathbb{T}_s$ a rotation matrix and Φ_s a diagonal rescaling matrix. The orthogonality requirement is crucial, because otherwise $\mathbb{X}$ is left unidentified. For it is always possible to define $\mathbb{X}^* = \mathbb{X}\mathbb{A}$ and $\mathbb{L}_s^* = \mathbb{L}_s\mathbb{A}$ without invalidating Equation 6 whereas Equation 7 always holds. On the other hand, generality is lost because the analysis will no longer be invariant under preliminary arbitrary linear transformations (i.e., $\mathbb{X}_s^* = \mathbb{X}_s\mathbb{A}_s$), the classical requirement posed by Hotelling (1936), but only under preliminary rotations. Also note that imposing more stringent constraints on the transformations, such as that they should be rotations plus a *uniform* stretching (i.e., Φ_s is required to be a scalar matrix—a case usually called generalized Procrustes analysis; Gower, 1975; Ten Berge, 1977) reintroduces a rotational indeterminacy. For now any rotation matrix $\mathbb{T}$ can be chosen to postmultiply both sides of Equation 7 leaving its meaning unchanged (if $\mathbb{T}_s$ and $\mathbb{T}$ are both rotations, so is $\mathbb{T}_s\mathbb{T}$).

An optimality criterion for fitting the structural Equations 7 to a given set of X_s's still has not yet been selected. Suppose Equation 7 is put into the equivalent form

$$\mathbb{X}_s\mathbb{T}_s = \mathbb{X}\Phi_s^{-1} \quad [8]$$

for $s = 1, \ldots S$ and the criterion is chosen that the Euclidean distances between the rows of the S matrices on the left-hand side of Equation 8 should be approximately equal to the Euclidean distances between the rows of the S matrices on the right-hand side. Then the fact can be used that Euclidean distances between the rows of two matrices $\mathbb{U}$ and $\mathbb{V}$ are equal if and only if $\mathbb{U}\mathbb{U}' = \mathbb{V}\mathbb{V}'$. This leads to the following approximation problem (where $\approx$ denotes "approximately equal in the least squares sense"):

$$\mathbb{X}_s\mathbb{X}_s' \approx \mathbb{X}\Phi_s^{-2}\mathbb{X}', \quad [9]$$

which is precisely the problem solved by Carroll and Chang's (1970) CANDECOMP method to fit the INDSCAL model (with Φ_s^{-2} identified as the matrix of weights or saliences). It is obvious that this is a particularly clever way to solve the generalized canonical correlation problem because the individual rotations have disappeared in Equation 9! On the other hand, it should not be surprising that the INDSCAL common space has a unique orientation of axes since actually the matching problem in Equation 7 is being solved.

Some might remark that the above development is purely formal and not relevant methodologically; after all, INDSCAL is designed as a model to fit three-way dissimilarity data. This may be true, but is it not a good methodological idea to approach the three-way scaling problem by first doing separate MDS analyses on the individual dissimilarities and next attacking the problem of comparing configurations in its full generality? This position has been especially advocated by Borg and Lingoes (1978) and by Gower (1979). The weights of an INDSCAL analysis represent both the strength and the weakness of the method. The strength is that a simple display of individual differences is possible: the so-called subject space can be plotted independently from the stimulus configuration(s). Although the practical importance of such a display can hardly be overestimated, there are very reasonable alternatives to this INDSCAL feature (see, for example, Gower, 1971). The weakness is that the alleged simplicity is only apparent: The precise properties of the weights depend on the program used to fit them; but mostly poorness-of-fit, certain interaction effects, and genuine stretching are *confounded*. At the least, nonstandard so-called *directional statistics* are needed to further analyze the subject space (Schiffman, Reynolds, & Young, 1981, chap. 13).

Symmetry in Treatment of Sets

Thus various types of restricted three-way MDS can be regarded as fitting various special cases of the multivariate matching problem. Similarly, two-way MDS with constraints on the coordinate axes can be compared with two-sets canonical correlation analysis. In both of them there are two spaces characterizing the objects, and it is their communality that is of particular interest. The idea of a space in between the spaces given again follows naturally if it is appropriate to treat the sets *sym-*

metrically, i.e., to distribute the approximation error evenly over the sets.

For S sets, a symmetric treatment will almost always be reasonable; for two sets, however, it seems that a more asymmetric treatment is frequently called for. The most radical treatment is to project one set onto the other. In case one set contains only one variable, as for instance in multiple regression and discriminant analysis, the criterion is projected onto the set of predictor variables. This reflects the requirement that the best predictor should be contained in the predictor set at hand; a direction in between would serve no practical purpose. Recently, a more asymmetric treatment for the case of multiple variables in both sets has been recommended (Muller, 1981; Stewart & Love, 1968; Van den Wollenberg, 1977), the argument being that rather than mutual similarity, variance-accounted-for should be optimized in one of the sets.

To apply these considerations in constrained MDS, consider the situation with one set of dissimilarities and one external variable, say an independent rating of psychological complexity. A symmetric treatment is exemplified by Ramsay's (1980) approach: he has set up a loss function in which the object points should simultaneously have distances approximating the dissimilarities and a direction among them that satisfies approximately the spacing given by the external variable. The object configuration is the common space that conveys information about both psychological complexity and dissimilarity of the objects. An asymmetric treatment is provided by Carroll's (1972) PREFMAP philosophy: first an unrestricted MDS is performed on the dissimilarities, possibly perfectly fitting them; next an optimal direction is found for the external variable, probably not accounting for all its variance. Alternatively, in the approach taken by Bentler and Weeks (1978), Bloxom (1978), De Leeuw and Heiser (1980), Heiser and Meulman (1983), and Noma and Johnson (1979), the external variable is *fitted perfectly*, e.g., along the first dimension, in the course of approximating the dissimilarities. Obviously, the first dimension need not be important in the sense of explaining much variance among the dissimilarities; just as in the case of the first canonical variate, it simply enforces a "correct" direction.

If there are several external variables, the situation does not change very much for Ramsay's (1980) and Carroll's (1972) approach: simply more directions are fit simultaneously or the PREFMAP procedure is repeated independently from earlier found directions. For the third approach the question arises, can several variables be fit by constraining successive orthogonal coordinate axes without acknowledging the correlation pattern that will generally exist among the constraints? The answer to this question is somewhat too involved to be fully explained here, but it can be said that if there are enough (say $M>K$) external variables spanning the same K-dimensional space of the dissimilarities, then an M-dimensional restricted MDS solution (where each axis corresponds with one external variable) will be of rank K and can be rotated to the correct common space. Thus, *the use of many constrained dimensions does not necessarily imply that the common space is high dimensional* (an analogous situation exists in confirmatory factor analysis). If the set of constraints is correlated, this will give the configuration a certain shape (see Kruskal, 1972, for a discussion of this concept), which can be studied, for instance, by computing its principal components.

This approach will now be illustrated with a classical example concerning the *perception of facial expressions*, using the dissimilarities collected by Abelson and Sermat (1962) and three median rating scale variables obtained for the same 13 objects by Engen, Levy, and Schlosberg (1958). The objects are selected photographs from the "Lightfoot-series" in which an actress expresses the emotions listed in Table 3. The columns of this table reflect judgments on the attributes "pleasant-unpleasant" (P-U), "attention-rejection" (A-R), and "tension-sleep" (T-S). The restricted MDS analysis was performed with the program SMACOF-II (Meulman, Heiser, & de Leeuw, 1983), which can treat both dissimilarity and rating scale values nonmetrically. Thus, a three-dimensional configuration was computed so that the order of the objects on the first coordinate axis corresponds with the order given by P-U; on the second one, with

A-R; and on the third one, with T-S. As a result, the three-dimensional solution is ordinally equivalent to the three-dimensional configuration given in Table 3. The monotone transformations of its columns are chosen in such a way that the distances among its rows approximate a monotone transformation of the dissimilarity data (stress = .098). The eigenvalues (measuring shape as the dispersion of the points along the principal axes) are .533, .374, and .083, indicating that the configuration is a reasonably flat disc; its first two principal components are displayed in Figure 3, in which the attribute vectors were obtained by projecting the coordinate axes of the three-dimensional solution onto the principal plane.

This principal plane is still a satisfactory representation in the following sense: The stress of the ordinal distance-dissimilarity relationship is .145; the attribute vectors fit ordinally with correlations of .980, .968, and .994 for P-U, A-R, and T-S, respectively; and the percentage of variance left unaccounted for with respect to the three-dimensional configuration is 8%. The conclusion must be that the attribute judgments and the dissimilarity judgments are *coherent;* they can be represented together as in Figure 3. This is not to say that the attributes describe everything the dissimilarities have to tell. The order of the stimulus projections on the direction perpendicular to the principal plane is given as the fourth column of Table 3: it presents the psychologist with the challenge to find an attribute that contrasts maternal love, anxiety, and light sleep with savoring a Coke, extreme pain, and physical strain. Or should this residual effect be regarded as random variation?

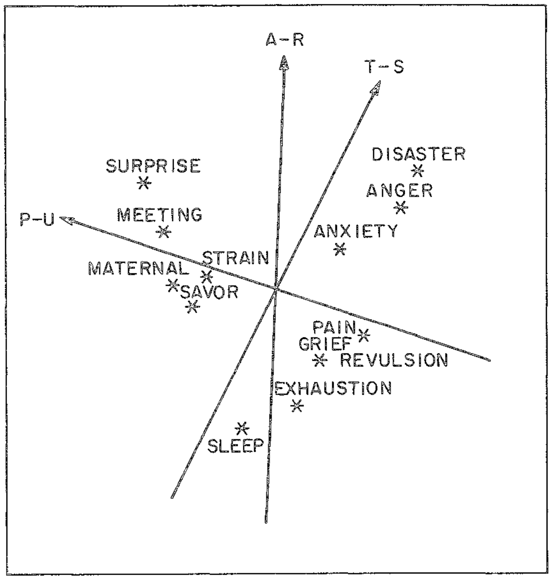
5. Confirmatory Scaling Analysis

Evidently, the possibility of constraining the MDS solution in various ways greatly enhances the options for analyzing data in a confirmatory fashion. Yet confirmation means trying to establish the balance between systematic and random effects. The question is how to decide in the MDS framework what is to be called systematic and what random variation. One strategy is to make this decision beforehand and to check its validity against a chance hypothesis. A case in point is that of Davison's (1980) proposal to analyze binary or ordinal response pattern data by Goodman's (1975) contingency table methods under an order hypothesis. It may be perfectly sensible to do so (provided the

Table 3
Median Rating Scale Values of 13 Facial Expressions on Three Schlosberg Scales (Engen et al., 1958); the Fourth Column is the Rank Order of Projections on the Third Dimension Omitted in Figure 3

Expression	Attributes P-U	A-R	T-S	Third Dimension Ranks
1. Grief at death of mother.	3.8	4.2	4.1	10
2. Savoring a Coke.	5.9	5.4	4.8	11
3. Very pleasant surprise.	8.8	7.8	7.1	9
4. Maternal love--baby in arms.	7.0	5.9	4.0	1
5. Physical exhaustion.	3.3	2.5	3.1	6
6. Anxiety--something is wrong with her plane.	3.5	6.1	6.8	2
7. Anger at seeing dog beaten.	2.1	8.0	8.2	5
8. Physical strain: pulling hard on seat of chair.	6.7	4.2	6.6	13
9. Unexpectedly meets old boy friend.	7.4	6.8	5.9	4
10. Revulsion.	2.9	3.0	5.1	8
11. Extreme pain.	2.2	2.2	6.4	12
12. Knows her plane will crash.	1.1	8.6	8.9	7
13. Light sleep.	4.1	1.3	1.0	3

Figure 3
Principal Plane of the Three-Dimensional Constrained Solution for the Combined Facial Expression Data



number of items is not too large); yet it would be a fallacy to regard this as confirmatory scaling.

Davison's (1980) method may be seen as an attempt to improve Goodman's (1975), but it does not scale anything. That is, it does not fit the parameters of a distance model, nor does it determine some optimal reordering of the rows/columns of a data matrix. Davison (1980) simply assumes that the scaling has already been done: "To be testable via the methods presented in this paper, an order hypothesis must satisfy one condition. It must divide the set of potential subject response patterns into two non-overlapping, complementary sets, a set of admissible response patterns and a set of inadmissible response patterns" (p. 124). The preferred division may turn out to be better than chance, but shouldn't it also be checked against others which might be even better? Indeed, Davison's procedure can be regarded as a particular loss function, evaluating for any single order of the items how well it fits the data. To return into the realm and spirit of scaling there should be a systematic search in the space of all possible orders for the one that minimizes the loss function under the hypothesis of unidimensionality.

Next consider the situation where there are *S* square symmetric tables of proximities available. Now a hypothesis on the order of the proximities may be confirmed by utilizing the statistical machinery of isotonic regression (Barlow, Bartholomew, Bremner, & Brunk, 1972). There would have to be found $I(I-1)/2$ mean values satisfying an increasing trend if arranged in the preconceived order; their variation would be the systematic part and could be compared with the residual variation in order to reject the chance expectation of equal rather than increasing mean proximity values. Again, this may be useful in some circumstances, but a typical MDS analysis will try to solve a completely different problem: Find one matrix of numbers representing the *S* original tables (its values may be means, but also medians, root mean squares, or still other representative numbers) and another one with numbers satisfying the metric axioms (and possibly much more than that) that are somehow maximally similar. Attempting an MDS analysis implies abandoning the idea of one a priori defined independent variable. It implies that the chance expectation of equal mean proximities can be rejected in favor of some MDS solution. In cases of doubt, any reasonable MDS analysis can be performed on one subset of the *S* tables and the resulting order of distances can be tested on the remaining data.

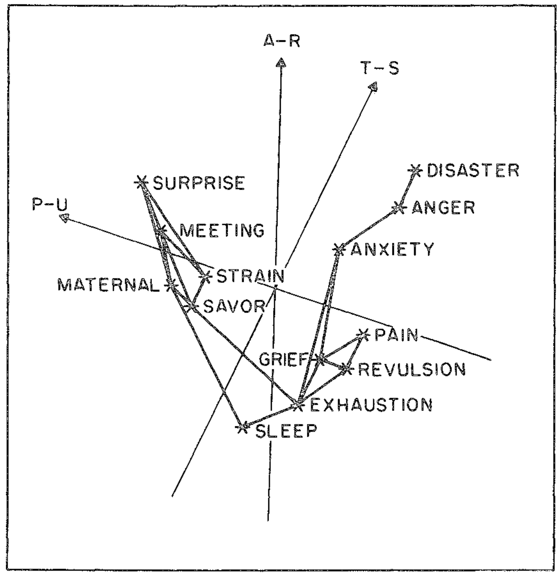
In conclusion, there is a tradeoff between confirmation in the narrow sense and doing MDS. Scaling models, constrained or not, reflect a posteriori systematic variability, and the real problem is not to decide between "effect" and "no effect" but to go beyond that and to evaluate several alternative specifications of the effect. For in most applications, the data would be asked very little indeed if the only interest were in finding out whether or not the expert's accumulated experience is better than chance. Before pursuing the issue of model selection, however, some remarks will be made about the assessment of the quality of any single solution.

Diagnostic Plotting

In order to check whether the loss function has been successfully minimized, it is imperative to examine auxiliary plots relating certain aspects of the model to the data. The scatterplot of distances versus disparities can reveal outliers, cluster effects, or misfit of small dissimilarities (in this last case it will be pear-shaped, a phenomenon almost always observed with ALSCAL). The transformation plot of disparities versus dissimilarities can give information about the way subjects utilized the dissimilarity scale (e.g., showing a ceiling effect), or about apparent inconsistency (e.g., exhibiting large plateaus in the step function). In case external variables were used to constrain the axes, it is also recommended that transformation plots be examined for the same types of reasons. A good review of diagnostic plotting is Chambers and Kleiner (1982).

Adding extra information into the standard plots of the results is almost always helpful. For example, Van der Kloot and Kroonenberg (1982) included a number of "pseudo-subjects" in their TUCKALS analysis to check the apparent individual differences against a priori defined pure types. A suggestion by Kruskal and Wish (1978) is to check apparent trends against the raw data. For example, Figure 3 could be characterized by saying that there are essentially three clusters: the joyful, the frightful, and the introverted expressions. Alternatively, it could be supposed—remembering the phenomenon that one-dimensional structures tend to become *V*-, *C*- or *S*-shaped if analyzed in two dimensions (Shepard, 1974)—that there is a bent chain here, going from SURPRISE through SLEEP to DISASTER. Either possibility could be substantiated in followup research with a larger number of stimuli from the same domain; but can one or the other be favored by relying on the data at hand? For the cluster conjecture, all small dissimilarities should correspond to within-cluster distances only; for the chain conjecture, all small dissimilarities will correspond to distances between groups of points consecutive along the chain; and if neither of them holds, then connecting highly similar points will produce a network extending in every direction. In Figure 4, all points have been connected whose dissimilarity value is smaller than 3.5 (which constitutes the lowest 27% of them). The chain conjecture is clearly supported.

Figure 4
Configuration for 13 Facial Expressions, with Each Dissimilarity Smaller Than 3.5 Indicated by a Line



This has another interesting consequence for understanding the relationships between the variables used as constraints. As is apparent from their almost orthogonal position in the plot and from their direct product-moment correlation of .18, the variables P-U and A-R are almost *uncorrelated*. However, because the stimuli tend to form a bent chain, the variables appear to be not unrelated. Upon inspection of their original scatterplot, it turns out that a quadratic curve fits with $r = .67$. Thus A-R can be predicted reasonably well from P-U, but not the reverse. This establishes that P-U is the more basic of the two, whereas A-R is merely a nonlinearly related effect. The constrained MDS analysis establishes the stronger result that the dissimilarities confirm the same tendency.

Goodness-of-Fit and Stability

Evaluating *goodness-of-fit measures* for the case $S = 1$ is still largely a matter of experience. Some guidance can be obtained from monte carlo work (Spence, 1983; Spence & Graeff, 1974), mainly for the determination of optimal dimensionality. In comparing a constrained solution with an unconstrained one, the former will generally have worse fit. In this situation, it can be useful to add extra parameters to the constrained model, as to make the degrees of freedom of the models comparable (cf. Bentler & Weeks, 1978). For instance, suppose the points are constrained to be on a circle, then I free parameters can be added to the model (a third dimension, or I unique dimensions) to bring the degrees of freedom on the same level as for an unconstrained two-dimensional solution. The enriched constrained model now has a fair chance to be better in the sense of goodness of fit.

For $S > 1$ this type of idea can be implemented more rigorously. Assuming that the S data matrices are indeed replications, goodness of fit can now be evaluated against the stability of the results under the supposition that the data might have been slightly different. That is, in selecting a model for the data there must be a willingness to exchange a certain decrease of fit for an increase in stability (cf. the well-known problem of selection of predictors in multiple regression). Conventional rules for doing this are primarily derived from the maximum likelihood principle. It was Ramsay (1977) who introduced these types of considerations in the MDS field. Suppose that Ramsay's approach is followed by adopting a log-normal distribution, independent sampling errors, and the log-likelihood function as a goodness-of-fit measure. Then a nested series of models can be tested by evaluating the ratio of their likelihood against the χ^2-distribution with the appropriate number of degrees of freedom.

There are definite advantages, however, to following a slightly different procedure, using the AIC statistic (Akaike, 1974, 1977) defined as

$$\mathrm{AIC}\ (\pi) = -2\ln L(\pi) + 2n_\pi \qquad [9]$$

where the log-likelihood $\ln L(\pi)$ is corrected for the number n_π of independently adjusted parameters in model π. The relative magnitudes of AIC may be directly compared across models that are not necessarily nested, and the one associated with the smallest value of AIC should be chosen as being statistically best. Here "best" means that the model is most stable in the sense that it will predict future observations (i.e., from the same distribution) most accurately (AIC may be considered as an estimate of the expected prediction error). This very useful secondary statistic was introduced in psychometrics by Takane (1978a, 1978b, 1981) and subsequently used by Winsberg and Ramsay (1981) and Stoop, Heiser, & de Leeuw (1981). Note that if two models have the same number of free parameters, the AIC criterion simply states that the one with best fit should be chosen.

With both the likelihood ratio test and the minimum AIC criterion, a very abstract and idealized reasoning is applied for choosing among alternative models. In the next section a class of procedures will be outlined that attempts to answer the same type of questions while remaining closer to the data. The stochastic elements are brought in by monte carlo methods, and actual computations will replace theoretical calculation.

6. Stability Analysis by Resampling

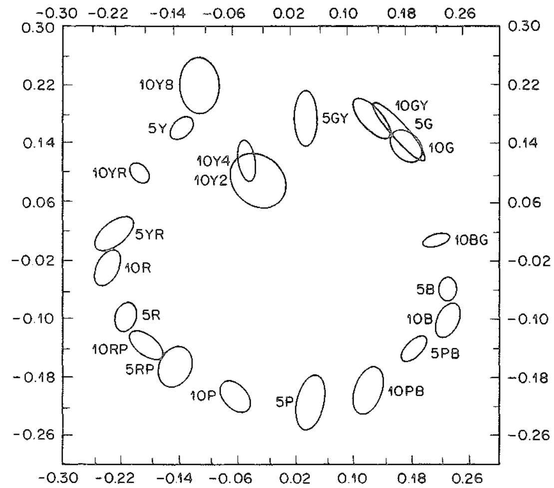
With the usual maximum likelihood methods, inference is obtained at the price of having to assume that the dissimilarities are sampled from independently distributed random variables with an a priori specified joint distribution F with unknown parameters. If such specific assumptions are unsuitable, the stability of the representation can still be investigated by resampling from the nonparametric maximum likelihood estimate of F, which is essentially the empirical distribution of the observations itself. Such resampling schemes are discussed in Efron (1979, 1982), Efron and Gong (1983), and Gifi (1981), are particularly suited to situations like those presented in this paper, where the combination of structural model, unknown distribution function, and complicated loss function renders a standard statistical analysis unwieldy.

One such resampling scheme, the so-called Bootstrap procedure introduced by Efron (1979), is illustrated. The example concerns data collected by Ramsay (1968) on the perceived differences between 21 color patches of constant value in terms of the Munsell color system. Ramsay obtained ratio judgments of dissimilarity with respect to one pair of standard stimuli for 20 subjects, with great reliability in terms of internal consistency. Thus, regarding each whole dissimilarity matrix as one single observation, there is a (multivariate) *dissimilarity distribution* characterized by 20 independent discrete values. Now, the idea of the Bootstrap is that an indication of the variability of the results may be obtained by resampling from this empirical dissimilarity distribution in order to obtain new data matrices of the same size that "just as well could have been observed." The resampling is done *with replacement* and repeated, say, B times. Then, an MDS analysis is performed on each of the B Bootstrap samples with identical options, and consequently B values are obtained instead of a single one for each numerical result of interest (coordinates, stress, and possibly further diagnostic numbers).

As a first analysis, a two-dimensional nonmetric scaling on the mean ratio judgments was performed, repeated $B = 20$ times. Because the optimal distances remain unchanged under arbitrary rotation of the coordinate axes, the 20 configurations were matched through separate orthogonal Procrustes rotations (Cliff, 1966) towards the solution for the original data. Now each stimulus is associated with a cloud of 20 points, and the question arises how to describe these succinctly. An irregular contour could be drawn so that each of the points lies precisely inside it, but the shapes of such contours would be heavily dependent on extreme points only, and the location of the bulk of the points would essentially be ignored. Instead a restricted class of contours has been chosen: the family of *ellipses* through which the major characteristics of the cloud can be grasped from their elongation and orientation. Precisely how this was done is described in Meulman and Heiser (1983). Here it suffices to say that the Bootstrap variances and covariances were used to estimate the shape and that the size of the ellipses to be considered below has been determined so as to include 95% of the Bootstrap points.

The final result of this series of operations is shown in Figure 5. Each ellipse is labeled with the Munsell notation for hue (e.g., 5YR denotes the fifth yellow-red), with a chroma designation for the 10Ys only because all the others are approximately of chroma 6. Except for the yellow-to-green corner

Figure 5
95% Confidence Ellipses Based on Bootstrapping Ramsay's Data without Constraints (20 Samples, Solutions Matched)

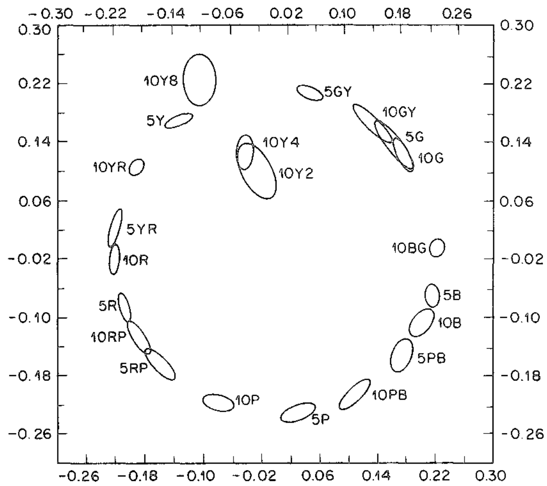


all ellipses are neatly nonoverlapping and rather small. According to the intention of the Munsell system the equal-chroma hues should be circularly arranged; however, there seems to be a slight tendency towards a pentagon. Also, there seems to be only weak evidence for equal spacing along the circular order.

Suppose the lack of equal spacing is considered to be of minor importance, and it is decided to study the restricted model enforcing the equal-chroma colors on a circle with both center and radius unknown and the yellows of varying chroma on some straight line: this is the Q configuration described in Section 3, and it appears to be so well established that it should not be discarded too quickly. The program SMACOF-II (Meulman et al., 1983) was used for the constrained scaling. Applying the same Bootstrapping scheme as before gives the constellation shown in Figure 6.

Most ellipses in Figure 6 are smaller (not all of them, however, cf. 5B), and again relatively large ones are found in the yellow-to-green region. Some diagnostic measures are summarized in Table 4. These include the stress values for the original unrestricted and Q-restricted analyses, which are not too different; using the idea that the number of model parameters should be approximately equal

Figure 6
95% Confidence Ellipses Based on Bootstrapping Ramsay's Data with Constraints (20 Samples, Solutions Matched)



for a fair comparison, one free dimension has been added to the Q-restricted model and the superior optimum stress value of .0616 was obtained. It should be noted here that the Q-restricted part of the configuration stayed primarily in the plane of the first two principal components, so that indeed it can be argued backwards that the two-dimensional Q is already sufficient. In what way do the Bootstrap results confirm the same conclusion?

A Criterion for Comparing Different Bootstrap Solutions

Not much work has yet been done on the possibility of defining decision criteria on the basis of the Bootstrap distribution of configurations. A good criterion to compare MDS models in terms of overall stability is to look for the smallest *Within/Total (W/T) ratio* of sums of squares, in which Within is the pooled sum of squared distances from points in a single Bootstrap cloud to its centroid and Total the sum of squared distances from all points to the origin. If instability prevails so that all Bootstrap clouds overlap almost completely, the W/T ratio will approach one; if there is a completely stable situation, it will become zero. As Table 4 indicates, the data are very homogeneous according to this definition: For both models the Within variance is less than .5% of the Total variance. Again, the Q-restricted model turns out to be the better one (its W/T ratio is smallest).

The W/T ratio can be interpreted as *stability corrected for fit*, because it can be shown that T equals $B - \Sigma\sigma_b^2$ where σ_b is the stress of Bootstrap b (at least for an algorithm properly minimizing Kruskal's stress). This normalizes the coefficient in a desirable fashion: If a restricted solution would be relatively stable but associated with high stress, the W/T criterion will tend to reject it. Research currently under way into the behavior of this ratio indicates that the correction for fit might not be too crucial in many cases; for it turns out that—contrary to what might be surmised on a priori grounds—restricted solutions are not necessarily more stable than unrestricted ones in terms of W (Meulman, 1983). This is especially true if constraints that are unreasonable are imposed: these already have a higher value of W and will certainly be rejected by the minimum W/T criterion.

A detailed comparison will not be attempted of the present authors' results with those obtained by Ramsay (1978) using his MULTISCALE method on the same data. Obviously, there is considerable conformity of objectives. Ramsay relies on the asymptotic normality of ML estimates and, as Efron (1982) has remarked, "Fisher's familiar theory for assigning a standard error to a maximum likelihood estimate is itself a 'bootstrap theory,' carried out in a parametric framework" (p. 29). Rather than deriving variability information directly from the empirical dissimilarity distribution, Ramsay (1978) is willing to accept a certain multivariate log-normal model and obtains a parametric or *smoothed* estimate of the unknown F, say F_{smooth}. He could then execute the Bootstrap algorithm exactly as is being done in this paper, starting with F_{smooth}. However, it is not necessary to actually carry this out because a shortcut is possible by means of theoretical calculations, due to the choice of a narrow parametric family of models. Evidently this saves considerable computing time but it could easily give somewhat overly optimistic results if the specific distributional assumptions do not provide a fair approximation. The correct comparison

Table 4
Diagnostic Measures for Ramsay's (1968) Color Data

Solution	Stress Original Solution	W/T Ratio Superimposed Bootstraps
Unrestricted 2-dimension	.0704	.0047
Q-restricted	.0791	.0032
Q-restricted + 1 dimension	.0616	—

would be to check the standard MULTISCALE results against a Bootstrapping scheme using the MULTISCALE loss function but ignoring the estimated standard errors.

The analysis presented here differs from Ramsay's in that Kruskal's loss function was used, the mean data were treated nonmetrically, and no power transforms were estimated for each replication. It has been illustrated that the employment of stability analysis as a diagnostic tool need not be confined to the framework of parametric maximum likelihood estimation. Given enough raw computing power and a criterion like the one proposed, the performance of both premeditated and ad hoc models on a given set of data can always be studied. To be sure, there must be replications. In case the proximities are collected in judgment experiments, this will frequently be no problem, except when there is particular interest in individual differences—for individuals cannot simultaneously be modeled and resampled. If the proximities were computed as measures of distance or association (correlation, confusion, joint occurrence), then resampling can be applied on the raw observations, and the slightly different versions of the proximity matrix thus obtained can be repeatedly scaled.

Conclusion

Choice of dimensionality is not the only feasible way to influence the arrangement of points in geometric models of data. Additional specifications can be introduced to mold the model into a particular shape. Such specifications will generally aim to enhance simplicity or order, and some of them have clear precedents in related fields of data analysis. They require an adaptation of the general MDS optimization problem that is definitely more involved than setting dimensionality equal to a fixed number. Their introduction does not automatically transform MDS from an exploratory into a confirmatory technique.

Exploration refers to a mode of analysis in which there is more concern about finding out what seems to be the case than about the correctness of expectations. In this mode, a constrained analysis may be useful because it can reveal patterns that would otherwise have been left unnoticed.

Confirmation involves the assessment of the influence and balance of systematic and random effects in the data. As things stand now, only partial knowledge exists on how particular systematic effects reveal themselves in the modeling process. It is also important to realize that sampling and specification errors are confounded if only one matrix of proximities is available for analysis. As always, replicated observations form the only sound basis for attempts to make general statements. They are a prerequisite for both sampling models and resampling techniques that can be used to reduce uncertainty. In this mode of analysis, the advantage of constrained MDS is in the possible increase in power and precision.

This paper has not touched upon algorithms and programs, in part because they are treated elsewhere, in part because the development of flexible and reliable software seriously lags behind the theoretical possibilities. This lag seems unavoidable in view of the enormous investments in time and effort that are required for proper implementation. It does not alter the fact, however, that the prospects for constrained MDS look bright.

References

Abelson, R. P., & Sermat, V. Multidimensional scaling of facial expressions. *Journal of Experimental Psychology,* 1962, *63,* 546–554.

Akaike, H. A new look at the statistical model identification. *IEEE Transactions on Automatic Control,* 1974, *AC-19,* 716–723.

Akaike, H. On entropy maximization principle. In P. R. Krishnaiah (Ed.), *Applications of statistics*. Amsterdam: North-Holland, 1977.

Anderson, N. H. *Foundations of information integration theory*. New York: Academic Press, 1981.

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

Attneave, F. Perceptions and related areas. In S. Koch (Ed.), *Psychology: A study of a science* (Vol.4). New York: McGraw-Hill, 1962.

Barlow, R. E., Bartholomew, D. J., Bremner, J. M., & Brunk, H. D. *Statistical inference under order restrictions*. New York: Wiley, 1972.

Benjamin, L. S. Structural analysis of social behavior. *Psychological Review,* 1974, *81,* 392–425.

Bentler, P. M. Multivariate analysis with latent variables: causal modeling. *Annual Review of Psychology,* 1980, *31,* 419–456.

Bentler, P. M., & Weeks, D. G. Restricted multidimensional scaling models. *Journal of Mathematical Psychology,* 1978, *17,* 138–151.

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

Borg, I. *Anwendungsorientierte Multidimensionale Skalierung*. Berlin: Springer Verlag, 1981.

Borg, I., & Lingoes, J. C. What weight should weights have in individual differences scaling? *Quality and Quantity,* 1978, *12,* 223–237.

Borg, I., & Lingoes, J. C. A model and algorithm for multidimensional scaling with external constraints on the distances. *Psychometrika,* 1980, *45,* 25–38.

Buneman, P. The recovery of trees from measures of dissimilarity, In F. R. Hodson et al. (Eds.), *Mathematics in the archeological and historical sciences*. Edinburgh: Edinburgh University Press, 1971.

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

Carroll, J. D. Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Vol. 1. Theory*. New York: Seminar Press, 1972.

Carroll, J. D., & Chang, J. J. Analysis of individual differences in multidimensional scaling via an *N*-way generalization of "Eckart-Young" decomposition. *Psychometrika,* 1970, *35,* 283–319.

Carroll, J. D., Pruzansky, S., & Kruskal, J. B. CANDELINC: A general approach to multidimensional analysis of many-way arrays with linear constraints on parameters. *Psychometrika,* 1980, *45,* 3–24.

Chambers, J. M., & Kleiner, B. Graphical techniques for multivariate data and for clustering. In P. R. Krishnaiah & L. N. Kanal (Eds.), *Handbook of statistics* (Vol. 2). Amsterdam: North Holland, 1982.

Chang, J. J., & Carroll, J. D. Three are not enough: An INDSCAL analysis suggesting that color space has seven (± 1) dimensions. *COLOR research and application,*1980, *5,* 193–206.

Cliff, N. Orthogonal rotation to congruence. *Psychometrika,* 1966, *31,* 33–42.

Cunningham, J. P. Free trees and bidirectional trees as representations of psychological distance. *Journal of Mathematical Psychology,* 1978, *17,* 165–188.

Davison, M. L. A psychological scaling model for testing order hypotheses. *British Journal of Mathematical and Statistical Psychology,* 1980, *33,* 123–141.

Davison, M. L. *Multidimensional scaling*. New York: Wiley, 1983.

De Leeuw, J., & Heiser, W. J. Multidimensional scaling with restrictions on the configuration. In P. R. Krishnaiah (Ed.), *Multivariate analysis* (Vol.5). Amsterdam: North-Holland, 1980.

De Leeuw, J., & Heiser, W. J. Theory of multidimensional scaling. In P. R. Krishnaiah & L. Kanal (Eds.), *Handbook of statistics* (Vol.2). Amsterdam: North-Holland, 1982.

Efron, B. Bootstrap methods: Another look at the Jackknife. *Annals of Statistics,* 1979, *7,* 1–26.

Efron, B. *The Jackknife, the Bootstrap and other resampling plans* (CBMS NSF Regional Conference Series in Applied Mathematics No. 38). Philadelphia: SIAM, 1982.

Efron, B., & Gong, G. A leisurely look at the Bootstrap, the Jackknife, and cross-validation. *The American Statistican,* 1983, *37,* 36–48.

Engen, T., Levy, N., & Schlosberg, H. The dimensional analysis of a new series of facial expressions. *Journal of Experimental Psychology,* 1958, *55,* 454–458.

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

Gati, I., & Tversky, A. Representations of qualitative and quantitative dimensions. *Journal of Experimental Psychology: Human Perception and Performance,* 1982, *8,* 325–340.

Gifi, A. *Non-linear multivariate analysis*. Leiden: University of Leiden, Department of Data Theory, 1981.

Glushko, R. J. Pattern goodness and redundancy revisited: Multidimensional scaling and hierarchical clustering analyses. *Perception and Psychophysics,* 1975, *17,* 158–162.

Goodman, L. A. A new model for scaling response patterns: An application of the quasi-independence concept. *Journal of the American Statistical Association*, 1975, *70*, 755–768.

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

Gower, J. C. Statistical methods of comparing different multivariate analyses of the same data. In F. R. Hodson et al. (Eds.), *Mathematics in the archaeological and historical sciences*. Edinburgh: Edinburgh University Press, 1971.

Gower, J. C. Generalized Procrustes analysis. *Psychometrika*, 1975, *40*, 33–51.

Gower, J. C. Comparing multidimensional scaling configurations. In A. P. M. Coxon & P. M. Davies (Eds.), *Working papers in multidimensional scaling*. Cardiff, Wales: MDS (X) Project, 1979.

Green, P. E., Carroll, J. D., & Carmone, F. J. Superordinate factorial designs in the analysis of consumer judgments. *Journal of Business Research*, 1976, *4*, 281–295.

Guttman, L. An approach for quantifying paired comparisons and rank order. *Annals of Mathematical Statistics*, 1946, *17*, 144–163.

Guttman, L. A new approach to factor analysis: The radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences*. Glencoe IL: The Free Press, 1954.

Guttman, L. Empirical verification of the radex structure of mental abilities and personality traits. *Educational and Psychological Measurement*, 1957, *17*, 391–407.

Guttman, L. A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, 1968, *33*, 469–506.

Guttman, L. What is not what in statistics. *The Statistician*, 1977, *26*, 81–107.

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

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

Heiser, W. J., & Meulman, J. *Smoothed monotone regression*. Paper presented at the Workshop on Nonmetric Data Analysis, Paris, 1983.

Horst, P. Relations among *m* sets of measures. *Psychometrika*, 1961, *26*, 129–149.

Hotelling, H. Relations between two sets of variates. *Biometrika*, 1936, *28*, 321–377.

Jöreskog, K. G. Structural analysis of covariance and correlation matrices. *Psychometrika*, 1978, *43*, 443–477.

Kettenring, J. R. Canonical analysis of several sets of matrices. *Biometrika*, 1971, *58*, 433–460.

Klingberg, F. L. Studies in measurement of the relations between sovereign states. *Psychometrika*, 1941, *6*, 335–352.

Krantz, D. H. Rational distance functions for multidimensional scaling. *Journal of Mathematical Psychology*, 1967, *4*, 226–245.

Kruskal, J. B. Multidimensional scaling by optimizing goodness-of-fit to a nonmetric hypothesis. *Psychometrika*, 1964, *29*, 1–28; 115–129.

Kruskal, J. B. Linear transformation of multivariate data to reveal clustering. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Vol.1. Theory*. New York: Seminar Press, 1972.

Kruskal, J. B., & Wish, M. *Multidimensional scaling*. Beverly Hills CA: Sage, 1978.

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

Lingoes, J. C., & Borg, I. Identifying spatial manifolds for interpretation. In J. C. Lingoes (Ed.), *Geometric representations of relational data*. Ann Arbor MI: Mathesis Press, 1977.

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

Lingoes, J. C., & Borg, I. A quasi-statistical model for choosing between alternative configurations derived from ordinally constrained data. *British Journal of Mathematical and Statistical Psychology*, 1983, *36*, 36–53.

McKeon, J. J. Canonical analysis: Some relations between canonical correlation, factor analysis, discriminant function analysis, and scaling theory. *Psychometric Monographs*, 1968, No.13.

Meulman, J. *Stability of constrained multidimensional scaling solutions*. Paper presented at the joint meeting of the Psychometric Society and the Classification Society, Jouy-en-Josas, France, 1983.

Meulman, J., & Heiser, W. J. *The display of Bootstrap solutions in MDS* (Technical Memorandum). Murray Hill NJ: Bell Laboratories, 1983.

Meulman, J., Heiser, W. J., & de Leeuw, J. *Progress notes on SMACOF-II*. Leiden: University of Leiden, Department of Data Theory, 1983.

Muller, K. E. Relationships between redundancy analysis, canonical correlation, and multivariate regression. *Psychometrika*, 1981, *46*, 139–142.

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

Noma, E., & Johnson, J. *Constrained nonmetric multidimensional scaling* (Technical Report MMPP 1979–4). Ann Arbor: University of Michigan, Mathematical Psychology Program, 1979.

Ramsay, J. O. Economical method of analyzing perceived color differences. *Journal of the Optical Society of America*, 1968, *58*, 19–22.

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

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

Ramsay, J. O. Joint analysis of direct ratings, pairwise preferences, and dissimilarities. *Psychometrika*, 1980, *45*, 149–165.

Ramsay, J. O. Some statistical approaches to multidimensional scaling data. *Journal of the Royal Statistical Society, Series A*, 1982, *145*, 285–312.

Restle, F. A metric and an ordering on sets. *Psychometrika*, 1959, *24*, 207–220.

Restle, F. *Psychology of judgment and choice*. New York: Wiley, 1961.

Richardson, M. W. Multidimensional psychophysics. *Psychological Bulletin*, 1938, *35*, 659–660.

Schiffman, S. S., Reynolds, M. L., & Young, F. W. *Introduction to multidimensional scaling*. New York: Academic Press, 1981.

Schönemann, P. H., & Borg, I. Measurement, scaling, and factor analysis. In I. Borg (Ed.), *Multidimensional data representations: When and why*. Ann Arbor MI: Mathesis Press, 1981.

Shepard, R. N. Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychological Review*, 1958, *65*, 242–256.

Shepard, R. N. Analysis of proximities: Multidimensional scaling with an unknown distance function. *Psychometrika*, 1962, *27*, 125–140; 219–246.

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

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

Shepard, R. N. Geometrical approximations to the structure of musical pitch. *Psychological Review*, 1982, *89*, 305–333.

Spence, I. Monte carlo simulation studies. *Applied Psychological Measurement*, 1983, *7*, 405–425.

Spence, I., & Graeff, J. The determination of the underlying dimensionality of an empirically obtained matrix of proximities. *Multivariate Behavioral Research*, 1974, *9*, 331–341.

Stewart, D., & Love, W. A general canonical correlation index. *Psychological Bulletin*, 1968, *70*, 160–163.

Stoop, I., Heiser, W. J., & De Leeuw, J. How to use SMACOF-IA. Leiden: University of Leiden, Department of Data Theory, 1981.

Takane, Y. A maximum likelihood method for nonmetric multidimensional scaling: I. The case in which all empirical pairwise orderings are independent—theory. *Japanese Psychological Research*, 1978, *20*, 7–17. (a)

Takane, Y. A maximum likelihood method for nonmetric multidimensional scaling: II. The case in which all empirical pairwise orderings are independent—evaluations. *Japanese Psychological Research*, 1978, *20*, 105–114. (b)

Takane, Y. Multidimensional successive categories scaling: A maximum likelihood method. *Psychometrika*, 1981, *46*, 9–28.

Takane, Y, & Carroll, J. D. Nonmetric maximum likelihood multidimensional scaling from directional rankings of similarities. *Psychometrika*, 1981, *46*, 389–405.

Takane, Y., & Carroll, J. D. *Multidimensional scaling from the tree construction method of similarity data*. Paper presented at the joint meeting of the Psychometric Society and the Classification Society, Montreal, 1982.

Ten Berge, J. M. F. Orthogonal Procrustes rotation for two or more matrices. *Psychometrika*, 1977, *42*, 267–276.

Thurstone, L. L. A law of comparative judgment. *Psychological Review*, 1927, *34*, 273–286.

Thurstone, L. L. *Multiple factor analysis*. Chicago: University of Chicago Press, 1947.

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

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

Tversky, A., & Gati, I. Similarity, separability, and the triangle inequality. *Psychological Review*, 1982, *89*, 123–154.

Tukey, J. W. Methodological comments focused on opportunities. In P. R. Monge (Ed.), *Multivariate techniques in human communication research*. New York: Academic Press, 1980.

Van de Geer, J. P. *Introduction to multivariate linear data analysis. Part V: Relations between K sets of data*. Leiden: University of Leiden, Department of Data Theory, 1980.

Van den Wollenberg, A. L. Redundancy analysis, an alternative for canonical correlation analysis. *Psychometrika*, 1977, *42*, 207–219.

Van der Kloot, W. A., & Kroonenberg, P.M. Group and individual implicit theories of personality: An application of three-mode principal component analysis. *Multivariate Behavioral Research*, 1982, *17*, 471–491.

Van der Kloot, W. A., & Van den Boogaard, T. *Weights of information in impression formation* (Report MT001-78). Leiden: University of Leiden, Department of Psychology, 1978.

Weeks, D. G., & Bentler, P. M. A comparison of linear and monotone multidimensional scaling models. *Psychological Bulletin*, 1979, *86*, 349–354.

Wiggins, J. S., Steiger, J. H., & Gaelick, L. Evaluating circumplexity in personality data. *Multivariate Behavioral Research*, 1981, *16*, 263–289.

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

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

Young, F. W. A model for polynomial conjoint analysis algorithms. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Vol.1. Theory*. New York: Seminar Press, 1972.

Young, G., & Householder, A. S. Discussion of a set of points in terms of their mutual distances. *Psychometrika,* 1938, *3,* 19–22.

Acknowledgments

This research was partially performed at Bell Telephone Laboratories, Murray Hill, NJ, U.S.A. Several useful comments on an earlier manuscript were made by Pieter M. Kroonenberg and by the editors of this issue. The authors are indebted to Peter M. Bentler for his encouragement in this project.

Author's Address

Send requests for reprints or further information to Willem J. Heiser, Middelstegracht 4, 2312 TW Leiden, The Netherlands.