PSYCHOMETRIKA—VOL. 39, NO. 4
DECEMBER, 1974

REPRESENTATION OF STRUCTURE IN SIMILARITY DATA: PROBLEMS AND PROSPECTS[1]

ROGER N. SHEPARD

STANFORD UNIVERSITY

I think I am safe in assuming that I am addressing you today as a consequence of my role in the development of methods for the discovery and representation of structures underlying matrices of similarity data and, specifically, the methods now most widely referred to as nonmetric multidimensional scaling. In the dozen years that have elapsed since I originally reported this development in the journal of this Society [Shepard, 1962a, b], we have gained an enormous amount of experience in using a number of methods of this general type in the analysis of diverse sets of similarity data. Accordingly, I thought this might be a suitable occasion, to set forth what appear to me to be the six major problems that still confront attempts to use nonmetric multidimensional scaling to represent structures underlying similarity data, and to indicate what appear to me to be the prospects for overcoming each of these six problems. First, though, I need to make a few remarks concerning the background and scope of what I am proposing to do.

Type of Data and Types of Models Considered

I shall limit myself entirely to the very simplest case in which we have a single, complete set of $n(n - 1)/2$ measures, of whatever sort, of the pairwise similarities among n objects of interest—whether stimuli, concepts, persons, traits, symptoms, cultures, or species. As usual, I shall suppose these measures to be arrayed in the below-diagonal triangular half of an $n \times n$ similarity matrix in which the n rows and n columns correspond to

[1] Presidential address delivered before the annual meeting of the Psychometric Society at Stanford University on March 28, 1974 (in a somewhat revised form necessarily, for purposes of publication, with more words and fewer pictures). Much of the work surveyed in this paper was supported by the Bell Telephone Laboratories, during my eight years as a member of the technical staff there, and by grants (principally, GS-1302, GS-2283, and GB-31971X) from the National Science Foundation, during my subsequent eight years at Harvard and at Stanford. I wish to acknowledge, too, the important contributions to this work made by my many students and associates during these 16 years including, particularly, Doug Carroll, Jih-Jie Chang, Steve Johnson, and Joe Kruskal (at the Bell Laboratories) and Phipps Arabie, Glen Crawford, and Jim Cunningham (at Stanford). Finally, I am indebted to Sherry Huntsberger for suggestions and extensive help in connection with the preparation of the paper itself, and to Bert Green for useful editorial comments.

the same n objects. Thus, I shall not explicitly consider a number of more complex cases of considerable inherent interest and practical importance, including cases in which the rows and columns do not correspond to the same set of objects and in which the matrix may in fact be rectangular rather than square (in accordance with the "ideal-points" or "unfolding" model of Coombs [1964]), and cases in which similarity data in several different $n \times n$ matrices, or, equivalently, in one three-way $n \times n \times m$ matrix are to be analyzed simultaneously (as in the powerful metric methods recently developed by Carroll and Chang [1970], by Harshman [1970], and by Tucker [1972]).

Most of my remarks will presuppose the basic type of model upon which all widely used methods of nonmetric multidimensional scaling are currently based; namely, that for monotone mapping of similarity data into distances in a coordinate space. However, as alternative ways of overcoming some of the six problems, I shall also present methods, together with illustrative applications, based upon some fundamentally different models, in which either or both the notions of distance or an underlying coordinate space are abandoned. It is in order to facilitate the consideration of models in which the notions of distance and, hence, proximity no longer play a central role that, throughout this paper, I am using the theoretically more neutral term "similarity data" rather than the term "proximity data," which was implicit in the name that I gave to my original method and which I, along with Coombs, have previously favored [Coombs, 1964; Shepard, 1962a, b; 1972b].

Purposes in Analyzing Similarity Data

I shall take the primary purpose of the analysis of such a triangular matrix of similarity measures to be the achievement of a concise, invariant, and assimilable representation of the essential pattern of structure that lies more or less hidden in the given array of numerical data. By the achievement of an "invariant" representation of the "essential" structure, here, I mean to exclude representations that are heavily influenced by arbitrary features of the data: representations, for example, that change appreciably when the data are subjected to seemingly permissible transformations. Within this primary, general purpose, one can further discern rather different subpurposes.

Use for Discovery of Hidden Structure

One of these is the use of such an analysis for the discovery of previously unknown structure and, hence, for the achievement of new scientific insight. In the past I have even suggested that this is the single most important purpose. I still regard it as of possibly the greatest *potential* importance and have myself already discovered some seemingly significant patterns in this way that were previously unknown, at least to me. One example is the discovery of the different dimensions that govern subjects' confusions among dot-and-dash signals of the Morse Code in two types of tasks that differed in the extents to which they depended upon perception and upon memory [Shepard, 1963a].

My analysis of Miller and Nicely's [1955] extensive data on acoustic confusions between consonant phonemes embedded in noise, also, disclosed some orderly patterns of substantive significance [Shepard, 1972c]. I found, first, that despite the 50-fold variation in overall mean level of confusion that resulted for different signal-to-noise ratios, the *pattern* of those confusions was virtually invariant. Second, this invariant pattern exhibited a consistent and striking departure from the parallelism that, on accepted distinctive-feature schemes, was expected to hold between the voiced and the corresponding unvoiced consonants (see Fig. 8A, below). As a further surprise, an analysis of data by Peters [1963] on judged similarities among those same consonants (but as articulated by the subject rather than heard), led to the finding of an entirely different pattern. The corresponding voiced and unvoiced consonants, which had almost never been *confused* with each other and had therefore emerged at opposite ends of what was interpreted as the dimension of voicing, were in fact *judged* to be maximally similar to each other and therefore tended to map almost directly on top of each other. This led me to the conclusion that, in the judgmental task, the obvious parallelism in the articulatory structure of the set of voiced and unvoiced consonants induced subjects to treat the task more as an analogy task than as a simple similarity task. They seemed to be using their similarity judgments to tell us that, despite the great dissimilarity in voicing, each voiced consonant was most *analogous* to its similarly articulated unvoiced consonant.

Use for Convenient, Parametric Representation

Nevertheless, I have come to realize that most of the discoveries that have so far been made in this way could in principle have been made without the aid of multidimensional scaling as such. I now believe that it probably is not the discovery of genuinely new phenomena that has been the most demonstrable, practical contribution of multidimensional scaling—whether of the recent nonmetric and metric varieties considered here or of the classical metric variety perfected earlier by Torgerson [1952, 1958]. Rather, it has been the facilitation of scientific advance in two other important ways.

First, multidimensional scaling has provided a convenient, objective, and uniform way of representing the essential pattern underlying experimental results for purposes of communication and comparison between studies or researchers. Thus virtually the same, readily appreciated spatial picture may be obtainable despite the wide variations in nature or absolute numerical magnitudes of similarity data that arise from judgments, confusions, or reaction times [Shepard, 1958b, 1960, 1962a; Shepard, Kilpatric, & Cunningham, in press] or, as just noted, under conditions of low, medium, or high absolute levels of confusion [Shepard, 1972c]. Moreover, owing to the great reduction and hence averaging of data that the picture represents, the relations embodied in the picture may have greater statistical reliability than the relations to which they most directly correspond in the original data.

Second, multidimensional scaling has furnished a quantitative method of psychologically calibrating or describing in a reduced, parametric way a set of objects to be used in further experimental research or for the test of theoretical models. Examples include attempts to refine the spatial representation of the perceived relations among colors, word meanings, or other stimuli for such purposes as the investigation of visual anomalies [Helm, 1964], the structure and function of semantic memory [Fillenbaum & Rapoport, 1971; Rips, Shoben, & Smith, 1973; Rumelhart & Abrahamson, 1973; Shepard, *et al.*, in press], the rules by which differences along individual underlying dimensions combine to produce an overall difference or similarity [Arnold, 1971; Attneave, 1950; Arabie & Boorman, 1973; Shepard, 1964a], and the functional form of the gradient of stimulus generalization [Shepard, 1958a, b; 1965] or of discriminative reaction-time [Shepard, *et al.*, in press].

Present Status of the Original Method of "Analysis of Proximities"

A number of the suggestions that I shall make for overcoming the major problems that I see as still confronting nonmetric multidimensional scaling derive directly from ideas incorporated in my original method of "analysis of proximities" [Shepard, 1962a, b]—particularly ideas used in the generation of unbiased initial configurations and in the reduction of dimensionality. Accordingly, a brief mention of a recent development concerning the status of that method may be in order.

Although my original method generally yielded spatial configurations that appeared indistinguishable from those furnished by subsequent methods, my mathematical colleague Joseph Kruskal [1964a] soon noted that the precise measure of departure from monotonicity that was being minimized by the method was neither explicitly defined nor even known to exist in an explicitly definable form. Thus, despite the intuitive plausibility and practical success of the method, it lacked the conceptual advantage of a strict mathematical specification of exactly what problem was being solved. Moreover, in the absence of an explicitly defined loss function, general techniques for the minimization of such functions (notably, gradient methods) were not apparently applicable. The iterative method that was used consequently appeared somewhat *ad hoc*.

Now however, some twelve years following my initial report of that method, my *two* former associates at the Bell Laboratories, Joe Kruskal (who succeeds me as President of this Society) and Doug Carroll (who has just been elected to succeed *him*), have jointly discovered that, unbeknown to all of us, my original method was in fact equivalent to a gradient method. Moreover they have even determined that the explicit form of the measure of departure from monotonicity that was (implicitly) minimized is one that appears to be quite reasonable (Kruskal and Carroll, personal communication; see Kruskal [in press]). Indeed, the method of optimization that was used apparently turns out, with a minor alteration in a normalizing factor, to belong to a class of "interchange methods" that Jan de Leeuw (personal communication) has recently shown to offer some advantages in the avoidance of merely local minima. However, these recently discovered aspects of my original method are mentioned, here, for their inherent or historical interest only; they do not form the basis for any of the recommendations that I shall offer in what follows.

SIX MAJOR PROBLEMS AND PROSPECTS FOR OVERCOMING EACH

On the basis of extensive first-hand experience with non-metric multidimensional scaling and examination of a large number of reports by other investigators using this type of method, I believe the following six problems to be most in need of further attention.

1. The problem of attaining the globally minimum departure from monotonicity (primarily the problem of avoiding merely local minima).
2. The problem of achieving a meaningful substantive interpretation of the spatial configuration.
3. The problem of determining the proper number of dimensions for the coordinate embedding space.
4. The problem of avoiding undesirable loss or extraneous imposition of structural information (especially the so-called problem of "degeneracy").
5. The problem of determining the form of the underlying metric (particularly the form of the rule governing how differences along two or more underlying dimensions combine to yield the overall similarity between two objects).
6. The problem of representing discrete or categorical structure (in view of the fact that the scaling model assumes a continuous underlying coordinate space).

I proceed, now, to consider each of these six problems in turn. For each I shall attempt, under the subheading "*Problem,*" to define and to illustrate the nature of the problem and some of its aspects and then, under the subheading "*Prospects,*" to suggest some directions in which I think a way of overcoming that problem might profitably be sought.

1. *Attaining the Globally Minimum Departure from Monotonicity*

Problem

All existing methods for nonmetric multidimensional scaling use some variant of the steepest descent or gradient method to minimize the chosen measure of departure from monotonicity. This is not because gradient methods guarantee either quick or certain achievement of the desired minimum; they do not. The choice is dictated, rather, by the present lack of a viable alternative for this particular class of large and nonlinear numerical problems. In fact the gradient method has two rather frustrating properties. First, convergence tends to become quite slow as a minimum point is approached—unless, perhaps, recourse is taken to second-order methods. But, owing to the enormously greater demands that second-order methods place on computer memory, they are apt to be prohibitive, except in the case of a relatively small matrix of data or of an exceptionally large computer. Second, entrapment in an undesired, merely local minimum is not unlikely—unless the initial configuration is constructed so as to fall (with sufficiently high probability) within the vicinity of the desired global minimum. But to ensure this is to solve, by some other means, a very large part of the problem of optimization itself.

The problem of entrapment in merely local minima is the more troublesome of the two because, rather than merely resulting in a delayed attainment of the absolute minimum, it can result in a failure to attain that minimum at all. I must confess to having underestimated the seriousness of the problem when, in previous papers, I have glibly remarked that, if a local minimum is suspected, one can always try a number of random starting configurations and simply take, as the absolute minimum solution, the one that (repeatedly) turns up with the same smallest value of departure from monotonicity ("stress"). Unfortunately, according to recent much more extensive experience (particularly by my former student and colleague Phipps Arabie), although the absolute minimum *is* often attained within the first four or five random starts, to be quite safe at least 20 different starts should be used if a Euclidean solution is sought. (Indeed, in the case of one set of data, the globally optimum Euclidean solution failed to emerge at all until the 26th random try!) And, as we shall see, in the case of severely non-Euclidean metrics such as the so-called city-block or dominance metrics, the situation is very much worse.

Ironically, local minima pose especially prevalent and therefore irksome obstacles to the attainment of the optimum configuration in what might otherwise seem to be the simplest case; viz., that of a one-dimensional space. Theoretical analysis confirms what has become clear from practical experience. Evidently, a point that is initially situated on the wrong side of some other points can gradually work its way around those other points in a space of two or more dimensions but, when confined to a single line, is unable to move through those points owing to forces of mutual repulsion. Even in published studies, one-dimensional solutions have been presented that I can show to correspond to such merely local minima.

A widely advocated remedy for the problem of local minima is, of course, the use of rationally constructed initial configurations instead of configurations generated either arbitrarily or purely at random. Generally, such rational starting configurations have been obtained by procedures derived from the classical metric approach to multidimensional scaling perfected by Torgerson [1952]. The hope is that such metric solutions will be sufficiently close to the global minimum for the nonmetric method to avoid entrapment in other merely local minima (cf., Young and Torgerson [1967]). Possibly such procedures will eventually be modified to the point where they can be demonstrated to be uniformly satisfactory. However, in the experience of my associates and myself (*e.g.*, [Arabie, 1973] and [Arabie & Boorman, 1973]), the ability of available procedures of this type to circumvent local minima has so far been disappointing—particularly so in cases of non-Euclidean metrics, one-dimensional spaces, and the highly nonlinear relations between similarity and distance that are characteristic of the types of behavioral data with which I have often been concerned (*viz.*, those of confusion [Shepard, 1957b, 1972c], association [Shepard, 1957a], or reaction time [Shepard, *et al.*, in press]).

Incidentally, the prevalence of the local-minimum problem implies that conclusions drawn from "Monte Carlo" investigations (*e.g.*, of stress, dimensionality, or alternative methods and programs) should be received with caution unless adequate assurances are provided that the reported statistics are not contaminated by the undetected occurrence of suboptimal solutions. I believe that Arabie [1973] is justified in suggesting (a) that several studies purportedly comparing different *methods* (*viz.*, M-D-SCAL, TORSCA, and SSA) succeeded only in demonstrating the undesirability of a certain type of initial configuration (*viz.*, the arbitrary L-shaped configuration used in early versions of M-D-SCAL), and (b) that the stress values reported even in some of the most useful Monte Carlo studies [Klahr, 1969; Stenson and Knoll, 1969] are inflated to an unknown extent.

Prospects

In the absence of a promising alternative to the gradient method for minimizing the chosen measure of departure from good fit, the judicious selection or construction of the initial configuration does seem to offer the best hope for ensuring convergence to the desired global minimum. However, since it is not yet clear which of a number of possible ways of doing this will eventually prove most uniformly efficient and successful, several quite different possibilities should be pursued. Currently, these appear to divide into two general classes. In one, the construction of an initial configuration that is sufficiently close to the globally minimum configuration is attempted within the space of specified, low dimensionality in which a solution is being sought. In the other, a completely unbiased configuration is first constructed in a space that is either sufficiently high-dimensional [Shepard, 1962a] or non-dimensional [Cunningham & Shepard, 1974] to make this possible and, then, a smooth and presumably trap-free mapping of this unbiased configuration is attempted down into the specified low-dimensional target space.

With respect to the first of these two classes, several quite different approaches to the construction of a rational starting configuration appear worthy of further exploration. Within the spirit of those already in existence, is the approach—apparently first considered by Torgerson himself (personal communication)—in which a metric multidimensional scaling procedure based upon a parametrically specified, smooth functional relation between similarity and distance is alternated with a reestimation of the parameters of the functional relation. Then, when this preliminary iterative process becomes suitably stationary, the resulting configuration can be used to start the iterative process of the nonmetric method itself. Alternatively, some combinatorial method of arriving at an optimal assignment or permutation of the objects with respect to points in a pre-established configuration or ordering might prove desirable--particularly when the target space has only one dimension or, possibly, has a Minkowski power metric of one of the two limiting forms ($r = 1$ or $r = \infty$). Finally, there is the related possibility of building up the starting configuration by finding a near-optimum placement for each point individually as the points are added to the configuration, one at a time. In particular, if each successive point is allowed to migrate in from an extra dimension orthogonal to the space in which the rest of the configuration is confined, there should be no reason for any point to become trapped on the wrong side of any other points.

The technique of dimensional compression that I incorporated in my original method of "analysis of proximities" [Shepard, 1962a] belongs in the second of the two above-mentioned general classes of methods for generating initial configurations. It can however be regarded as carrying out the just-described point-by-point inward migration—but on all n points simultaneously. Such a way of looking at that method may help, moreover, to clarify why that method is not susceptible to entrapment in unwanted local minima. To start with, the n points are arranged as the vertices of a regular simplex in $n - 1$ dimensions. This is a completely unbiased configuration in which all $n(n - 1)/2$ interpoint distances are equal. As iteration proceeds, the variance of the distances is systematically increased by stretching distances that should be large and shrinking distances that should be small. The net effect is that each point migrates towards its optimum location within the hyperplane of the $(n - 2)$ dimensional simplex defined by the other $n - 1$ points in such a way as to instate and maintain the desired rank order of the distance from that point to each of the $n - 1$ other points. Clearly, there is again no reason for any point to become trapped in an inappropriate region. Experience reinforces our theoretical expectation that this process becomes stationary when all points have mutually gravitated into close proximity of the lowest dimensional hyperplane within which a good monotone fit to the similarity data is still possible. Rotation to principal axes then enables elimination of the superfluous dimensions, leaving us with the desired coordinates for the dimensionally-reduced and trap-free starting configuration.

Although this method requires that the n points be initially embedded in a space of $n - 1$ dimensions, the efficiency of the method and its invulnerability to entrapment may more than offset the disadvantage of having to carry along $n(n - 1)$ coordinates during these preliminary iterations. I know of no case in which this procedure has led to a merely local minimum. And, illustrative of its efficiency, for the first significant set of data analyzed by this method (*viz.*, Ekman's [1954] data on the judged similarities among 14 spectral colors), a highly satisfactory, virtually two-dimensional starting configuration was achieved in just two iterations [Shepard, 1962b].

A final alternative, which has close conceptual relations to the approach just considered, is offered by the method of "maximum variance nondimensional scaling" recently developed by my student, Jim Cunningham, and myself [Cunningham & Shepard, 1974]. This method finds that set of (generalized) distances among the n points such that (a) the distances satisfy only the metric axioms (positivity, symmetry, and triangle inequality) and such that (b) an appropriate balance is achieved between maximization of the variance of those distances and approximation to a monotone relation of the distances to the given similarity data. The distances are not required to satisfy the much stronger conditions entailed by embedding the points in a coordinate space of the Euclidean, Minkowskian, or any other variety. Owing to its coordinate-free and nondimensional nature, the obtained representation is, like the high-dimensional representation just considered, not subject to entrapment in merely local minima. But, owing to the device of maximizing variance, the obtained distances tend, where possible, toward consistency with a low-dimensional representation. Accordingly, purely linear metric multidimensional scaling based upon those distances should yield a near-optimum initial configuration for nonmetric multidimensional scaling. Unfortunately, convergence has so far proved to be rather slow with this method. So it remains to be seen whether the efficiency of this approach can be improved to the point where it becomes a practical competitor of my original method of dimensional compression.

2. Achieving a Meaningful Substantive Interpretation

Problem

The substantive interpretation of a spatial configuration obtained by multidimensional scaling is usually a matter of paramount importance. Typically, in fact, such an interpretation is *the* end result which the investigator is seeking and which, to the extent that it is meaningful and enlightening, justifies the often rather costly computations from which it derives. In addition to its importance for its own sake, moreover, the interpretability of the configuration often plays a crucial role in determining whether the obtained solution is valid and, particularly, whether it has been embedded in a space of the appropriate number of dimensions.

In view of its importance, it is distressing to see that this matter of interpretation is still sometimes neglected or mishandled in some way. In some cases, a spatial configuration of some number of dimensions has simply been presented without any compelling interpretation. For, even when an acceptably small measure of departure from monotonicity (stress) is achieved, in the absence of such an interpretation one can not determine (a) that the number of dimensions retained is appropriate or (b) that the configuration itself is valid (and not heavily determined by some mixture of random fluctuations in the data, degeneracy, and/or a merely local minimum).

In the worst cases, an overriding preoccupation with the reduction of stress to some desired level has led to solutions in four or more dimensions (a) where the configuration can not be visually apprehended and is probably not statistically reliable anyway, and even (b) where attempts at substantive interpretation, if any, have often been based solely on the coordinates for the points as they are printed out with respect to the unrotated axes of the solution. Such an approach to interpretation fails to appreciate two facts: First, except in certain special cases of non-Euclidean metrics, (or of methods of individual difference scaling [Carroll, 1972a] beyond the scope of this paper), the axes of the obtained configuration are entirely arbitrary. And second, even when appropriately rotated, axes do not necessarily offer the most interpretable features of a configuration.

Prospects

What seems to be most immediately needed, are efforts to educate potential users of multidimensional scaling concerning such matters as substantive interpretation, rotation of axes, and minimization of stress versus minimization of dimensionality. This paper is in part intended as one such effort. Some specific suggestions that may be helpful to some users are the following: First, always try for a solution in a space of three or, preferably, fewer dimensions where the spatial structure of the entire configuration can be seen and interpreted directly (rather than through the coordinates of the points on arbitrary axes). Second, when a representation in three or more dimensions can be shown to be both reliable and desirable, try objective methods for finding the most interpretable, rotated axes through the resulting high-dimensional space. (An excellent survey of such objective methods has been prepared by Carroll [1972b]. Less complete and up-to-date, though perhaps more readily available, is my own brief overview [Shepard, 1972b, pp. 39–43].) Third, search for *any* interpretable features of the spatial configuration—including clusters and circular orderings, as well as the linear ordering provided by (rotated) rectilinear axes.

Some concrete examples may help to illustrate the usefulness of searching for interpretable features other than axes. In fact the first significant application of nonmetric multidimensional scaling (again, my reanalysis of Ekman's [1954] data on 14 spectral colors, [Shepard, 1962b]) led to a quasi-circular configuration (similar to that presented in Fig. 7D) in which there was a perfect agreement between the ordering of the points around the circle and the ordering of the corresponding colors with respect to wave length. In this and some subsequent, quite different applications, interpretation in terms of a circular dimension (akin to Guttman's [1954] "circumplex") seems as inviting as interpretation in terms solely of rectilinear dimensions.

In Fig. 1 I exhibit the two-dimensional result of a nonmetric analysis of data (which I had originally collected in 1953) on the strengths of mental associations among 16 familiar kinds of animals. Subjects were asked simply to list as many kinds of animals as they could think of in ten minutes For the 16 most frequently listed kinds, I used, as a measure of the associative distance between the two items in each pair, a nonlinear average of the numbers of intervening items between those two items in the lists returned by the individual subjects.

In a first attempt to analyze association data by multidimensional scaling, I then applied the metric method described by Torgerson [1952, 1958] to these distance-like numbers in order to obtain a spatial representation for the 16 kinds of animals [see Shepard, 1957a]. At that time I obtained a three-dimensional representation in which [following Osgood, Suci, and Tannenbaum 1957] I tentatively interpreted the three axes, after appropriate rotation, as dimensions of size, potency, and activity. However, even though similar three-dimensional configurations were independently recovered from the data for two random subsets of the subjects, and even though subsequent studies have independently come up with a dimension of size and a dimension (of "predacity") that is obviously related to potency [see Rips, *et al.*, 1973], I did not at the time feel that the interpretations of the three axes sufficiently enlightening to warrant publication of the three-dimensional spatial configuration itself.

More recently, I have reanalyzed the very same set of measures of associative distance by nonmetric methods of multidimensional scaling and hierarchical clustering (using, specifically, M-D-SCAL [Kruskal, 1964a, b], HICLUS [Johnson, 1967], and embedding the clustering into the spatial representation as advocated in Shepard [1972c]). The two-dimensional solution (Fig. 1) provided a satisfactory monotone fit to the association

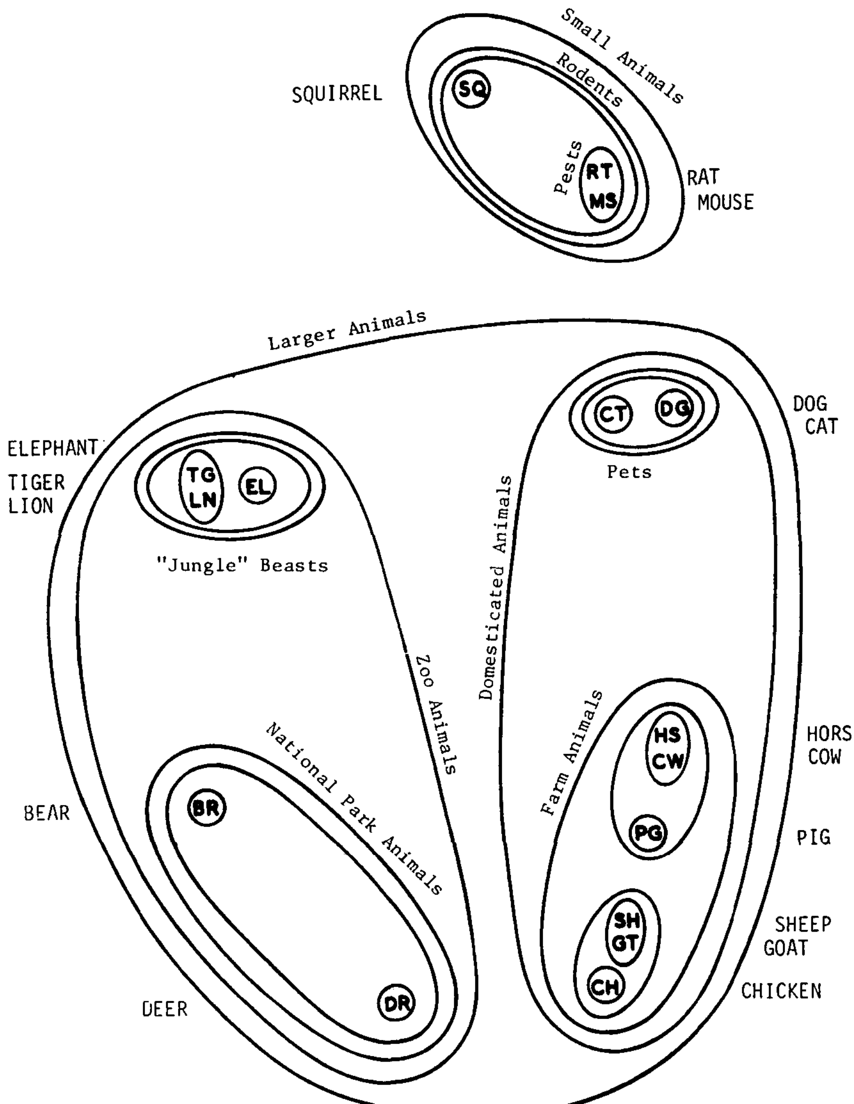

FIGURE 1

Two-dimensional spatial representation of the concepts of 16 animals, with embedded hierarchical clustering, based upon measures of association from a free-recall experiment by Shepard [1957a].

data and, also, was consistent with the (nondimensional) clustering result—as is indicated by the fact that the obtained clusters could be uniformly represented by smooth, convex, nonoverlapping contours. The principal point that I wish to make here, however, is that the way in which the points representing the kinds of animals cluster together in the spatial representation appears to be far more readily and compellingly interpreted than the order in which those points project onto any axes passing through this spatial representation. In some cases, the specific interpretative label printed in the figure may be open to dispute; *e.g.*, the label "Jungle Beasts" for lion, tiger, and elephant. But in every case the concept—as opposed to the verbal label—is, I hope, clear. Other examples in which clusters as well, possibly, as axes seem particularly susceptible to substantive interpretation will be presented in connection with later issues (see Figs. 8 and 13).

3. Determining the Proper Number of Dimensions

Problem

In my opinion, widespread practices and recommendations concerning the determination of dimensionality are tending to detract from the usefulness of multidimensional scaling. I believe that users, more often than not, are inclined to err in the direction of extracting too many dimensions. This inclination seems to be attributable to certain prevalent misconceptions about the nature of nonmetric multidimensional scaling and about the implications of Monte Carlo studies.

First, many users tend to place undue emphasis on the numerical value of the measure of departure from monotonicity (stress) to the virtual exclusion of much more important considerations of the statistical stability and substantive interpretability of the obtained configuration. In part, this may stem from an unfortunate tendency of users to accept, a bit too literally, the evaluative labels ("excellent," "good," "fair," and "poor") that Kruskal [1964a] once associated with particular numerical levels of stress. (Nobody likes to submit for publication a result that is only "fair" or "poor"!) Second, Monte Carlo studies, which generally recommend the extraction of more rather than fewer dimensions, are often limited in one or both of two respects: (a) they place excessive emphasis on approximating an underlying (artificially constructed) configuration (which the unprocessed similarity data themselves already do quite well!), while they disregard the more important considerations of stability and accessibility to substantive interpretation, and (b) they report mean stress values that in some cases may be inflated owing to the undetected occurrence of suboptimal solutions. Third, there has been a pervasive failure, even among otherwise sophisticated investigators, to recognize the guises under which a basically one-dimensional case can appear to the unwary to be two- or even three-dimensional.

As just one illustration of this last phenomenon, I present in Fig. 2A a two-dimensional representation that Levelt, Van de Geer, and Plomp [1966] obtained by a nonmetric analysis of the judged similarities among 15 aurally presented musical intervals. They attempted (without striking success, I feel) to give substantive interpretations to the two orthogonal dimensions of this representation and even to the third dimension of a three-dimensional solution as well. (Their interpretive effort included the fitting of a parabola—displayed, here, by the dashed curve—to the two-dimensional configuration.) To me, however, two aspects of this solution strongly suggest that the data should be represented in a one-dimensional space. First, the points fall essentially on a C-shaped curve that is very similar to the semi-circular configuration that, under the permissible monotone transformations of the interpoint distances is equivalent to a one-dimensional straight line [Shepard, 1962a, p. 130]. And second, the ordering of the points around this curve (as indicated by the solid curve terminating in an arrowhead) agrees nearly perfectly with an obvious physical property of the intervals—namely, their separation in terms of number of intervening half-tones on the musical scale.

A completely independent solution, shown in Fig. 2B, is based upon similarity data that a former student of mine, Christopher Wickens, collected for his 1967 undergraduate honors thesis at Harvard before either of us knew of the study by Levelt *et al.* Although Wickens' experiment included only 12 of the 15 intervals investigated by Levelt *et al.*, the overall C-shaped nature of the two configurations in two-dimensional space is quite striking.

In analyses of many different sets of data that were known to be basically one-dimensional, I have found that two-dimensional solutions, when attempted, characteristically can assume either the simple C-shape (illustrated in Fig. 2) or the inflected S-shape, and that solutions in higher-dimensional spaces are even more various. (Note for example that, just as a semicircle in two dimensions is monotonely equivalent to a straight line in one, a helix in three dimensions is, in the same sense, monotonely equivalent to both.) Evidently, by bending away from a one-dimensional straight line, the configuration is able to take advantage of the extra degrees of freedom provided by additional dimensions to achieve a better fit to the random fluctuations in the similarity data. In some published applications, moreover, the possibility of the more desirable one-dimensional result was mistakenly dismissed because the undetected occurrence of a merely local minimum (which is especially likely in one-dimension) made the one-dimensional solution appear to yield an unacceptibly poor monotone fit and/or substantive interpretation.

Prospects

With the exception of my own original method, all methods of nonmetric multidimensional scaling require the user to specify, in advance, the number of dimensions of the space in which the solution is to be sought. When in doubt, the user must obtain solutions, separately, in spaces of different numbers of dimensions (perhaps 3, 2, and 1) and then use criteria such as goodness of fit, statistical stability, and substantive interpretability to choose among the resulting solutions [Shepard, 1972a, pp. 9–10]. In order to take maximum advantage of this approach and to avoid the specific pitfalls mentioned above, I believe it to be highly desirable (a) to recognize the importance of the criteria of stability and interpretability (including the special advantages of visually accessible two-dimensional representations and the embedded clusterings to which they particularly lend themselves), (b) to strive toward more careful Monte Carlo studies and, hopefully, more illuminating mathematical analyses, and (c) to be vigilant for configurations (especially *C*-shaped or *S*-shaped ones in two dimensions) which strongly indicate the attainability of an acceptable one-dimensional solution.

The possibility of a one-dimensional solution can also be determined by an examination of the matrix of similarity data itself. If and only if the underlying structure is truly one-dimensional, a permutation of the rows and columns of the matrix can be found such that, except for random fluctuations, the entries decrease monotonically with distance from the principal diagonal (as in the generalized "simplex" of Guttman [1955]). Fig. 3 displays such a premuted version of the matrix of data reported by Levelt, *et al.* [1966] and upon which the spatial representation in Fig. 2A was based. To assist visualization, the heaviness of the cell entries has been

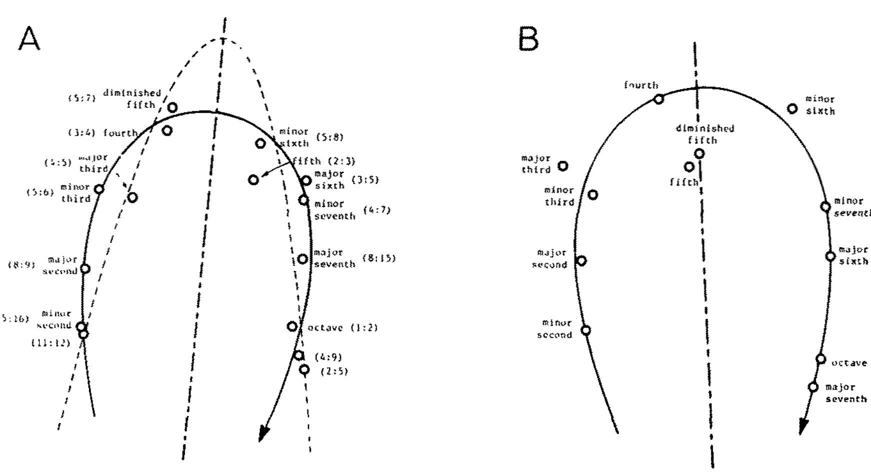

FIGURE 2

Two-dimensional spatial representations of musical intervals, obtained by Levelt, Van de Geer, and Plomp [1966] (A) and by Wickens (B), on the basis of independent sets of judged similarities.

chosen, here, in accordance with their numerical magnitudes. Notice that, except for minor fluctuations in seemingly isolated cells, these cell entries do tend to shade off quite uniformly as we move from the diagonal to the lower left corner. This is in contrast to the matrix of similarities among spectral colors reported by Ekman [1954], in which the similarities systematically increased again toward the lower left corner and in which monotonicity could only be maintained by a two-dimensional configuration in which the two ends of the C-shape (corresponding to red and violet) necessarily approached each other to form the familiar "color circle." (See Shepard [1962b] and the present Fig. 7D.)

	(2:5)	(4:9)	octave	maj 7th	min 7th	maj 6th	5th	min 6th	dim 5th	4th	maj 3rd	min 3rd	maj 2nd	(11:12)	min 2nd
(2:5)															
(4:9)	**30**														
octave	**25**	**29**													
maj 7th	13	13	**26**												
min 7th	14	**20**	**22**	**26**											
maj 6th	18	17	**73**	18	**30**										
5th	16	10	14	18	13	**25**									
min 6th	18	10	12	18	**27**	**24**	**27**								
dim 5th	10	9	9	19	16	13	**22**	**22**							
4th	12	12	15	10	14	**24**	**21**	17	**27**						
maj 3rd	7	8	14	17	18	**20**	18	**21**	**24**	**25**					
min 3rd	10	8	9	10	14	15	13	**20**	**25**	**22**	**28**				
maj 2nd	6	6	9	7	15	12	14	13	14	13	**23**	**28**			
(11:12)	14	14	3	10	11	11	10	7	8	10	17	**22**	**32**		
min 2nd	9	9	8	7	7	12	9	6	8	15	14	19	**29**	**32**	

FIGURE 3

Rearranged version of the matrix of similarities among musical intervals obtained by Levelt *et al.* [1966].

All of the preceding recommendations are with respect to the (almost universally adopted) approach of obtaining solutions in spaces of specified dimensionalities and then using various external criteria to choose among the resulting configurations. An entirely different approach, attempted in my original method [Shepard, 1962a], is to use constraints internal to the data themselves to determine the proper number of dimensions objectively. I still believe that such an approach is feasible, particularly in cases of relatively noise-free data, and that it avoids a number of problems, particularly the just-considered problems of inherently low-dimensional configurations appearing as curved structures in higher-dimensional spaces and of entrapment in merely local minima.

The first compelling demonstration of the possibility of determining the true number of underlying dimensions objectively was that illustrated in Fig. 2 of my original report [Shepard, 1962b, p. 223]. That figure showed that, by means of the above-mentioned device of stretching large distances relative to small, an initially regular simplex in 14 dimensions flattened down into a stationary configuration of 15 points that was virtually two-dimensional and essentially identical to the true underlying structure from which the input data were monotonically derived. And, again, this process of dimensional compression was sufficiently effective that the true number of underlying dimensions could be estimated after only three iterations.

Even more striking demonstrations of the potential power of this approach have arisen from subsequent attempts to extend the approach to permit the determination of the "intrinsic dimensionality" of curved structures in general [Bennett, 1969; Shepard & Carroll, 1966]. Fig. 4 presents the results of two tests of a method of this type, for "conformal reduction of nonlinear data structure," that I developed at the Bell Laboratories with the collaboration of Jih-Jie Chang. The method uses essentially the same sort of differential stretching of distances as my original 1962 method but differs from that method in that the requirement of monotonicity between the original data and the final distances is enforced only locally rather than globally. The method is quite similar to that developed by Bennett [1965, 1969] except for the definition of the local neighborhood around each point. In our method each neighborhood is defined (a) with respect to distances between points in the (evolving) configuration rather than with respect to the (fixed) set of data, and (b) according to an exponential-decay weighting function of these distances rather than according to an arbitrary discontinuous cutoff.

As can be seen from the nine successive stages of the iterative process shown in Fig. 4A, the abandonment of global monotonicity permitted a configuration of 19 points in the form of a circle with a gap (similar to the configurations in Figs. 2 and 7D and to Guttman's [1954] "circumplex") to open out into a straight line. The iterative process achieved stationarity with this perfectly one-dimensional configuration; there were no further changes with continued iteration. A similar evolution is shown in Fig. 4B for an intrinsically two-dimensional configuration of 49 points on a sphere. As can be seen from the portrayed two-dimensional projections, the initially

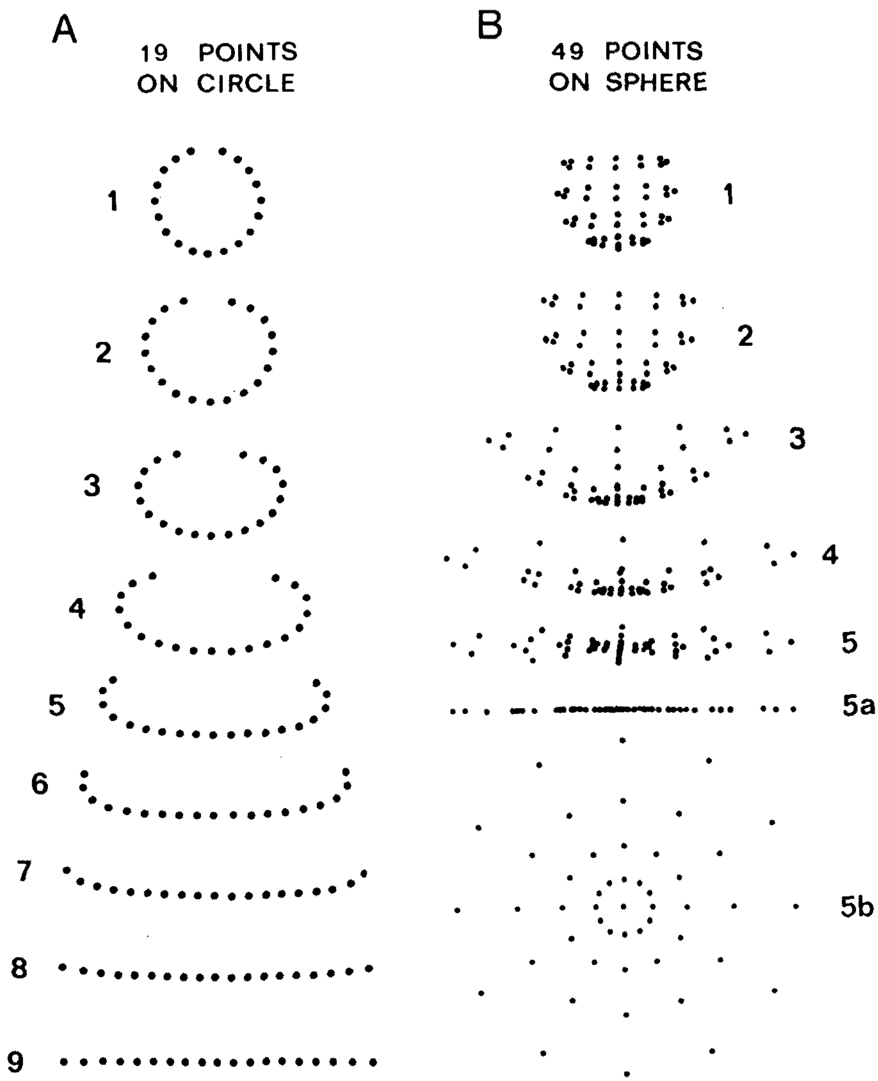

FIGURE 4

Successive stages in Shepard-Chang conformal reduction of 19 points on the perimeter of a circle (A) and of 49 points on the surface of a sphere (B) into flat spaces of the appropriate intrinsic dimensionalities.

spherical configuration (1), flattened out into shapes successively resembling a deep bowl (2), a parabolic microwave antenna (3), a shallow platter (4), and a perfectly flat disk (5). The striking degree of flatness of this final, stationary configuration is exhibited more clearly in view 5a, which shows the final configuration (of step 5) rotated into an edge-on orientation.

The maintenance of monotonicity, locally, ensured that the flattening in both cases, despite the global distortion, preserved local structure and hence was essentially "conformal." This is attested to by the evenness of the spacing of the points in the final configuration (9) that evolved from the 19 points on the circle, and by the systematically expanding regularity of the final configuration (5) that evolved from the 49 points on the sphere. This last regularity is most evident in view 5b, which shows the same configuration (of views 5 and 5a) rotated into a flat-on orientation. (The views in this figure are reproduced from the Shepard–Chang computer-generated movie "Illustrations of Conformal Mapping" which was first shown during an invited address before the 1966 meeting of the American Psychological Association [Shepard, 1966a].)

In the two examples presented in Fig. 4, the data were error free and the usual requirement of global monotonicity was relaxed. Nevertheless, the success of this objective method for dimensional reduction, together with the success of my original method from which this method derives (and in which global monotonicity was required with real and therefore fallible data) encourages me to believe that true underlying dimensionality is in principle determinable by automatic, objective methods.

4. Avoiding Loss or Imposition of Structure

Problem

Ironically, although we always seek to minimize the chosen measure of departure from monotonicity, special difficulties are apt to arise whenever the stress is zero or close to zero. Zero-stress solutions, in particular, are generally either *nonunique*—and to that extent introduce some degree of extraneous structure that is not contained in the data, or *degenerate*—and therefore fail to preserve some structure that *is* contained in the data. The problem of nonuniqueness is the less troublesome. One can always evaluate the degree of uniqueness of a solution by using several different starting configurations. Moreover, what the finding that substantially different solutions are obtainable with the same zero stress really indicates is that either the number of objects being scaled should be increased or the number of dimensions of the embedding space should be decreased. The problem of degeneracy, however, is sometimes bothersome enough to motivate a search for methodological innovations.

As a rough index of the *non*degeneracy of a solution, I would propose the ratio of the number of distinct values of distances among the n points to the total number of distances (*viz.*, $n(n - 1)/2$). Thus, if no two distances are tied we have a totally nondegenerate solution, while if all $n(n - 1)/2$ distances are tied we have a totally degenerate solution (namely, the regular simplex in $n - 1$ dimensions). Typically, because we require a solution in fewer than $n - 1$ dimensions, a zero-stress solution is not totally degenerate by this index, and the distances assume one of two or three different values.

As an illustration, Fig. 5A shows a degenerate solution that occurred when I analyzed data long ago sent to me by A. Howard, on the judged similarities among eight gustatory stimuli. The stimuli collapsed into the

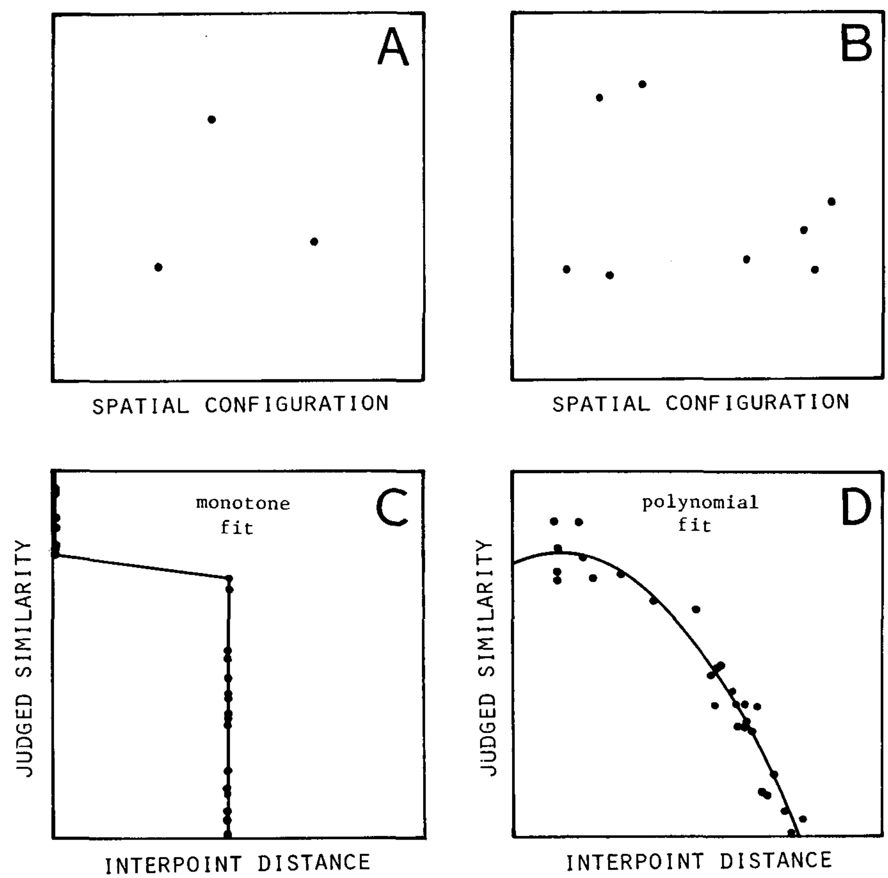

FIGURE 5

Degenerate (A) and nondegenerate (B) configurations for eight gustatory stimuli, based upon the "nonmetric" and "metric" assumptions that the relation of the similarity data to the interpoint distances has the form of a merely monotone function (C) or the form of a polynomial of low degree (D), respectively.

three vertices of an equilateral triangle and, so, there were just two values of distance: zero, for pairs of points located at the same vertex, and a fixed larger value, for pairs of points located at two different vertices. This degeneracy arose because the stimuli divided into three groups such that all of the similarities within any group were greater than any of the similarities between any two groups. The consequence is that much structural information is lost in the spatial solution. We learn only that the stimuli strongly cluster into the three groups; we learn nothing about either the relationships among these three groups or the relationships among the stimuli within any one of these groups. Correspondingly, the function relating the given similarities and the recovered distances assumes the implausible and uninformative shape of the single, discontinuous step shown in Fig. 5C. Although the fit is perfect (*i.e.*, the stress is zero), the representation is too degenerate to be of much use.

In Fig. 6 I display all maximally degenerate four-point configurations in two dimensions; *i.e.*, all two-dimensional configurations in which the distances among the four points take on only two distinct values. The smaller distances are represented by the heavier lines connecting pairs of points—unless those distances are zero, in which case the two or three coincident points are represented by concentric circles. (These configurations may be regarded as two-dimensional analogues of the one-dimensional "corner sequences" defined by Abelson and Tukey [1959].) Of the many strongly degenerate two-dimensional configurations that I have obtained in the analysis of real and artificial data, regardless of the number of points all

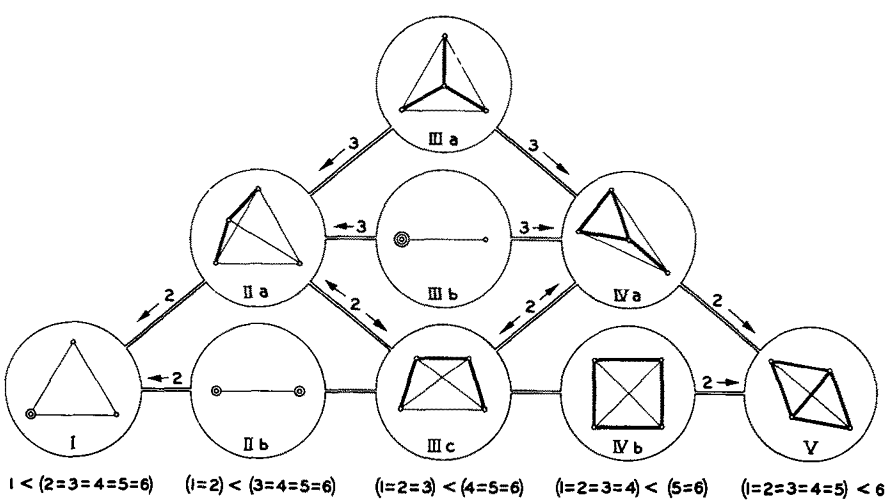

FIGURE 6
The nine degenerate four-point configurations in two dimensions, for which the six interpoint distances take on just two distinct values (as indicated along the bottom).

have taken one or another of the nine forms shown in Fig. 6. In cases of extreme degeneracy, apparently, the additional points usually collapse onto these same few vertices. Thus, the eight-point configuration shown in Fig. 5A corresponds to the equilateral triangular degeneracy labeled "I" at the left in Fig. 6. Therefore, whenever the points in an obtained solution cluster close to the vertices of one of the highly symmetric configurations exhibited in Fig. 6, one should suspect the occurrence of a rather marked degeneracy. In practice, the true extent of this degeneracy is often not immediately obvious because the criterion (*e.g.*, of low stress) for terminating the iterative process is attained before the clusters complete their collapse into the vertices. Even so, the tendency of the monotonicity diagram (see Fig. 5C) toward a discontinuous function of one or two steps will usually suffice as a signal of impending degeneracy.

Closely related to the problem of true degeneracy, in which a large proportion of the interpoint distances are tied, is the problem of quasidegeneracy, in which many of the monotone values being brought into a mutual best fit with the distances, if not the distances themselves, are tied. Actually, some degree of quasidegeneracy in this sense is always present in solutions obtained by nonmetric methods (although it may be negligible when the number of points is large and the data are sufficiently error-free [Shepard, 1966b]). It shows up in the zigzag or step-like shape that is characteristic of the best-fitting monotone functions obtained by nonmetric multidimensional scaling (see Fig. 7A). Roughly, the more pronounced the individual steps appear, the greater is the quasidegeneracy of the solution.

To the extent that we are interested in the functional form of the relation between the similarity data and metric distances (as in the study of stimulus generalization), such quasidegeneracy is undesirable for two related reasons; one substantive and one statistical. Substantively, the step-like function is unappealing if, as I assume, we generally believe that the true underlying relationship has some smooth functional form (such as the exponential decay form expected under some circumstances on theoretical grounds [Shepard, 1958a]). Statistically, the zigzag function is unreliable in the sense that, when we analyze a new set of similarity data for the same set of objects, we find that the individual zigs and zags of the function shift about in a quite unpredictable manner. The presumption, therefore, is that these individual zigs and zags do not represent any reliable or substantively meaningful phenomenon but, rather, reflect the attempt of the large number of degrees of freedom of a merely monotone function to fit the random fluctuations peculiar to each individual set of data.

Prospects

In order to minimize the likelihood of true degeneracy or of marked quasidegeneracy, researchers should try, whenever possible, to select objects for nonmetric scaling (a) that are not obviously grouped into a few psychologically compact clusters, and (b) that are not fewer than about ten in total number, for a two-dimensional solution, or more, for a higher-dimensional solution [Shepard, 1966b]. (A distressing number of two- and even three-dimensional solutions have been published in which, despite the inclusion of only six to eight objects, no evidence is provided that the configuration has a reasonable degree of metric determinacy and is not a prematurely arrested case of convergence toward a degeneracy.) Whether or not it is actually included in the published report, the monotonicity diagram (Figs. 5C and 7A) should be examined for step-like evidences of degeneracy and the results reported.

When true degeneracy does occur (as illustrated in Figs. 5A and 6), one currently has recourse to one or both of two further kinds of analysis in order to achieve a representation that preserves more of the structure in the similarity data. One can reapply the nonmetric analysis, separately, to the submatrix of similarities for the objects corresponding to each collapsed cluster containing enough points to make this worthwhile. If this does not lead to a further (hierarchically deeper) degeneracy, additional information can thereby be recovered about the internal structure of such a subset—though not about the relation of that subset to any other. Alternatively, if a representation of the overall structure of the entire set of objects is still desired, one must resort to metric methods, which depend upon stronger assumptions concerning the functional form of the monotone relation between similarity and distance. Sometimes, a combination of both approaches is quite successful. (See, for example, Fig. 18 and accompanying discussion in Shepard, *et al.* [in press].)

The present Figs. 5B and 5D illustrate the use of stronger, metric assumptions to overcome degeneracy. Here, the very same set of data already analyzed nonmetrically in Figs. 5A and 5C were reanalyzed using a program for "polynomial fitting in the analysis of proximities" that I developed in an early attempt to deal with the problem of degeneracy [Shepard, 1964b]. (Subsequently, of course, Kruskal and others have generalized their programs to provide for the fitting of polynomial or other parametric functions, also.) Notice that, although the fit is no longer perfect (as it should not be with fallible data), the spatial configuration preserves structural information about all eight of the stimuli (Fig. 5B), and (apart from a minor deviation from monotonicity, to be considered shortly) the relation between similarity and distance approximates a more plausible, smooth functional form—specifically, in this case, a quadratic (Fig. 5D).

The fitting of smooth, parametric functions rather than jagged, merely monotonic functions also permits us to circumvent the lesser problem of quasidegeneracy. This is illustrated in Fig. 7. The upper two panels (*A* and *B*) are both based upon measures of the tendency of pigeons trained to respond to each of a number of spectral colors to generalize that response to each of the other colors. These data, collected by Guttman and Kalish [1956], are of the sort for which the question of the functional form of the "gradient of stimulus generalization" is of central interest [Shepard, 1965]. From this standpoint, the function obtained by applying my polynomial-fitting program—in this case an exponential-like quartic curve (Fig. 7B) seems to be both substantively more plausible and statistically more reliable than the function obtained by nonmetric multidimensional scaling (Fig. 7A).

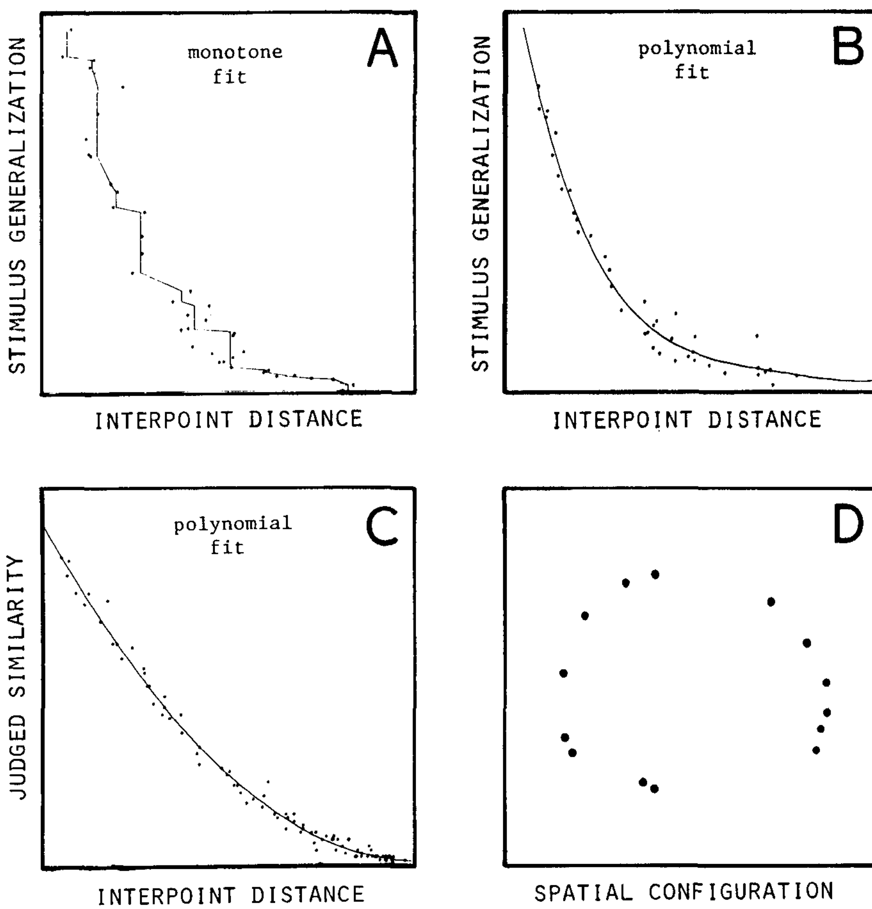

FIGURE 7

Measures of stimulus generalization between spectral colors, obtained for pigeons by Guttman and Kalish [1956], plotted against interpoint distances obtained by multidimensional scaling on the assumption of a merely monotone (A) or a polynomial (B) relation between similarity and distance; and a plot for a similar polynomial analysis of Ekman's [1954] judged similarities among 14 spectral colors (C) together with the quasicircular spatial configuration obtained by that analysis (D).

The lower panels (C and D) present a reanalysis of Ekman's [1954] data on the similarities among spectral colors as judged by human subjects. The quasicircular configuration obtained for the 14 colors by means of the polynomial-fitting program (Fig. 7D) is virtually identical to the nondegenerate one I originally obtained by a nonmetric analysis [Shepard, 1962b, p. 236]. Moreover, the obtained relationship between judged similarity and Euclidean distance between points in this configuration (Fig. 7C) exhibits a strikingly good fit to a smooth and plausible monotone decreasing function. Evidently, such a metric method of analysis can yield quite satisfactory results whether or not there is a problem of degeneracy or of quasidegeneracy.

One drawback of polynomial-fitting programs, of course, is that it is difficult to ensure that the resulting polynomial will be monotone over the range covered. Thus, a departure from monotonicity is evident in Fig. 5D and, although not included in the figure, a nonmonotonic upswing occurs just beyond the right-hand border of Fig. 7B. Kruskal (personal communication) has suggested the strategy of simultaneously minimizing departures from monotonicity and from a polynomial of specified degree (a possibility that is provided in recent versions of his program, *e.g.*, M-D-SCAL 5M). But, even with this strategy, I have found that the fitted polynomial can still become markedly nonmonotone in the range of the data.

In order to ensure monotonicity, one can of course iterate into a best fit with a parametric function of some other general type that can be more easily constrained to strict monotonicity. Since confusion and generalization data seem in general to decay exponentially with interstimulus distance [Shepard, 1958, a, b, 1965, 1972c], Jih-Jie Chang and I developed a gradient method for optimally adjusting, simultaneously, the coordinates of a spatial configuration and the parameters of an exponential decay function relating the similarity data to the interpoint distances [Chang & Shepard, 1966]. In Fig. 8 the results of applying this exponential-fitting method to Miller and Nicely's [1955] data on confusions among 16 consonants (A) is contrasted with the results of a nonmetric (M-D-SCAL) analysis of the same data (B). The closed curves show the embedded results of hierarchical clustering [Johnson, 1967] applied to the same data, as previously explained for Fig. 1. (See Shepard [1972c] for the full substantive interpretation of this configuration.)

Note that, in the configuration (B) obtained by the nonmetric analysis, only, there is an inconvenient partial degeneracy in which several points collapse together. Correspondingly, the fitted monotone function obtained by the nonmetric analysis manifested, again, a crude step-like shape (similar to that shown in Fig. 7A), whereas the fitted exponential function obtained by the metric analysis achieved an excellent fit (even better than that shown in Fig. 7B) and, in fact, accounted for some 99% of the variance of the confusion measures of similarity [Shepard, 1972c, p. 77]. Nevertheless, the

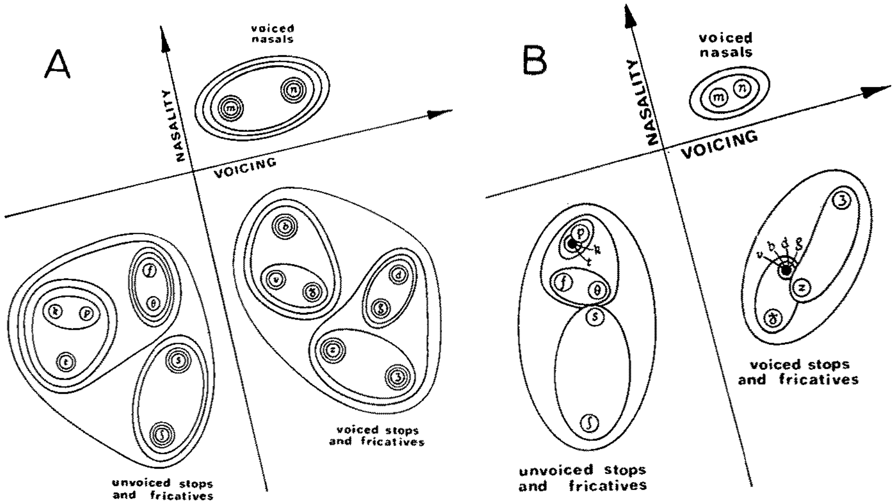

FIGURE 8

Two-dimensional spatial representations of 16 consonant phonemes obtained by multidimensional scaling on the assumptions that the relation of the confusion measures of similarity (obtained by Miller and Nicely [1955]) to the interpoint distances is exponential (A) or merely monotone (B), together with embedded hierarchical clusterings [see Shepard, 1972c].

exponential-fitting method is undesirably limited for general purposes, since many other types of similarity or dissimilarity data are not expected to bear an exponential relation to distance.

What may have seemed most remarkable in my original demonstrations of nonmetric scaling [Shepard, 1962b] was the extent to which a tightly constrained metric structure can be recovered from an analysis of merely ordinal relations in the data [cf., Shepard, 1966; Young, 1970]. It may seem odd, therefore, that I am now recommending consideration of methods that depend upon more than merely ordinal relations. Nevertheless, from the practical standpoint of trying to obtain results that are optimally meaningful and invariant under replication of the data as well as under reasonable (and therefore smooth) sorts of monotone transformations of the data, such a recommendation seems appropriate. The problem that remains, however, is to impose the desired condition of smoothness in some general and non-arbitrary way without having to specify a particular functional form—such as the polynomial (which can become nonmonotonic) or the exponential (which is often too restrictive).

Recently, with the collaboration of a student in computer science, Glen Crawford, I have been exploring a new approach to the problem of fitting functions which seems to provide a very natural and well-defined way of introducing general conditions such as convexity, concavity, or even "smoothness," as well as the condition of monotonicity. For each value x_i of an independent variable x (where the subscripts are assigned so that $x_1 < x_2 < \cdots < x_n$), we seek a theoretical estimate $\hat{y}_i$ that best fits the corresponding observed value y_i such that the theoretical values $\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_n$ satisfy certain explicitly prescribed constraints on their weighted first- and second-order differences as required to ensure that the sequence of values is monotone, is convex or concave, and/or is locally linear, as desired. The method that we are currently testing (which we call "least-squares regression to a constrained-difference function") uses a gradient method with penalty functions to obtain a least-squares solution subject to the prescribed constraints.

The results of one test application of this method are displayed in Fig. 9. In Fig. 9A the best-fitting monotone decreasing function is indicated for a set of artificial data (a quadratic with added random error). The given data are represented by the open circles and the fitted function is represented by the connected small solid circles. Note, as before, the characteristically step-like shape of the best-fitting monotone function. By contrast, Fig. 9B shows that, as soon as we use the method to impose just the condition of convexity in addition to the condition of monotonicity, the function fitted to the same data becomes very smooth. Moreover, the convex function also achieves a much closer fit to the true underlying (quadratic) function.

We plan to use this approach to curve fitting as the basis for a method of multidimensional scaling in which one will not have to assume a particular functional form for the relation between distance and similarity, but in which one can impose some further constraint beyond mere monotonicity (in order to obtain a functional relationship that is more plausible, more reliable, and less subject to degeneracy). In the most direct extension to multidimensional scaling, the fixed similarity values, s_{ij}, would be treated as the values of the independent variable x, while the to-be-varied distances, d_{ij}, would be treated as the values of the dependent variable y. However, other possibilities, e.g., turning the regression around, would also offer advantages, including that of being able to maximize the variance of the given similarity data accounted for by the spatial representation.

5. *Determining the Form of the Underlying Metric*

Problem

In addition to permitting the recovery of metric structure from merely ordinal data, the iterative procedures introduced in the original methods of nonmetric multidimensional scaling made possible the fitting of representations based upon non-Euclidean metrics. (See [Shepard, 1962b, p. 224] and, for the first actual implementation of this possibility, [Kruskal, 1964a, b].)

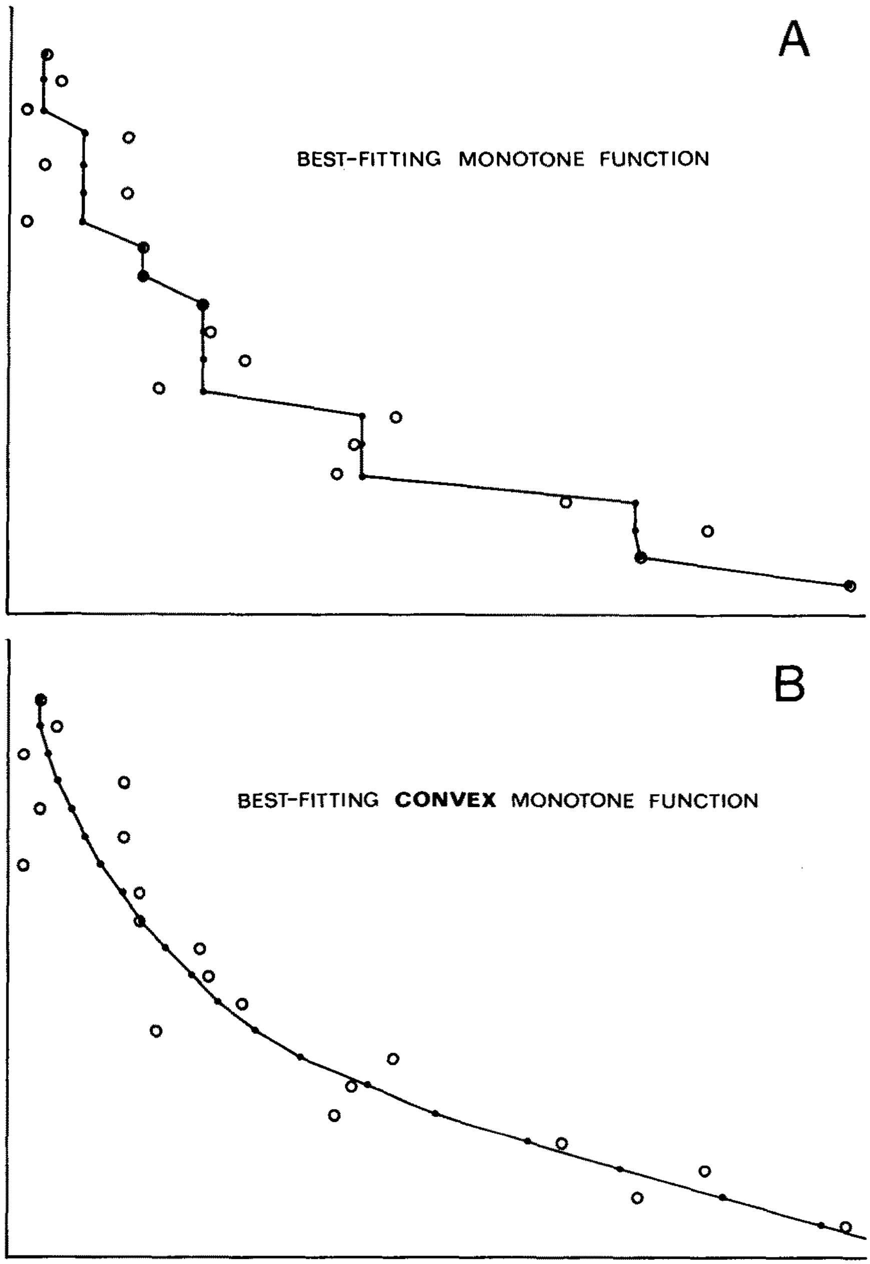

FIGURE 9

Identical artificially generated scatterplots with the best-fitting monotone decreasing function (A), and with the best-fitting monotone decreasing and convex function obtained by the Shepard-Crawford method of regression to a constrained difference function (B).

Methods of this type are therefore used in the continuing investigation into the conditions under which psychological similarity is determined by alternative Euclidean or non-Euclidean rules of combination of differences along underlying psychological dimensions [see, *e.g.*, Arnold, 1971; Attneave, 1950; Cross, 1965; Hyman & Well, 1967, 1968; Shepard, 1964a; Shepard & Cermak, 1973; Thomas, 1968; Torgerson, 1958]. Unfortunately, difficulties both of a persistent practical sort and of a recently discovered theoretical nature raise doubts about the ways in which these methods are usually used for this purpose.

Methods of nonmetric multidimensional scaling generally seek Minkowski power-metric representations by iterating to a best fit after specifying a value for the power, r (where the most widely discussed metrics—the so-called "city-block," "dominance" and, of course, Euclidean varieties—correspond to the limiting values of $r = 1$, $r = \infty$, and the intermediate value of $r = 2$, respectively). Thus, if the appropriate value of r is (as in the typical case) not known in advance, the user is faced with the tedious and costly procedure of obtaining solutions, separately, for a number of representative values of r spaced out between 1 and some very large value. The user must then decide among the resulting set of solutions in some way (presumably on the basis of which solution achieves the lowest residual stress [*e.g.*, Kruskal 1964a, p. 24]).

From the practical standpoint, the cost of obtaining solutions separately for many different values of r is apt to become prohibitive because problems of slow convergence and local minima are much more severe in the case of these non-Euclidean metrics. On the basis of his extensive attempts to obtain non-Euclidean solutions, Arabie (personal communication [see, also, Arabie & Boorman, 1973]) reports that, in three or more dimensions, random starting configurations are essentially useless. For this case, he recommends resorting to the incompletely validated strategy—apparently suggested, independently, by Arnold [1971] and Kruskal (personal communication)—of gradually working out from the Euclidean solution (for $r = 2$), by using the final configuration obtained for each value of r as the initial configuration for the next larger (or smaller) value of r. In the case of two dimensions, by contrast, Arabie reports that this (Arnold–Kruskal) strategy is generally unsuccessful and that one should therefore use random initial configurations (with perhaps as many as 100 different starts being required for each value of r in order to ensure attainment of the global minimum!).

Quite apart from this essentially practical problem, my own recent theoretical investigations have convinced me that the generally accepted practice of taking, as the correct metric, the one which yields the lowest residual departure from monotonicity is unfounded and probably leads to erroneous conclusions. Such a practice is based on the assumption, never explicitly justified, that values of stress are directly comparable across different values of r. I first came to question this assumption as a result of puzzling over the tendency (first noted by Arnold [1971] and, then, by Arabie and Boorman [1973]) of the points in spatial configurations conforming to the city-block or the dominance metric to be themselves disposed in a manner resembling the shape of the "unit sphere" for those particular metrics; that is, resembling the perimeter of a diamond or square, respectively, (in two dimensions) or the surface of an octahedron or cube, respectively, (in three). This led to the discovery of the purely geometrical fact that degeneracies and, hence, low values of stress are more prevalent for values of r close to 1 and, particularly so, for values of r approaching ∞.

The special nature of these extreme metrics (and a possible explanation for the puzzling phenomenon noted by Arnold and by Arabie and Boorman) is illustrated in Fig. 10, for the case of the city-block metric in two dimensions. In this case, the unit circle (*i.e.*, the set of all points equidistant from a

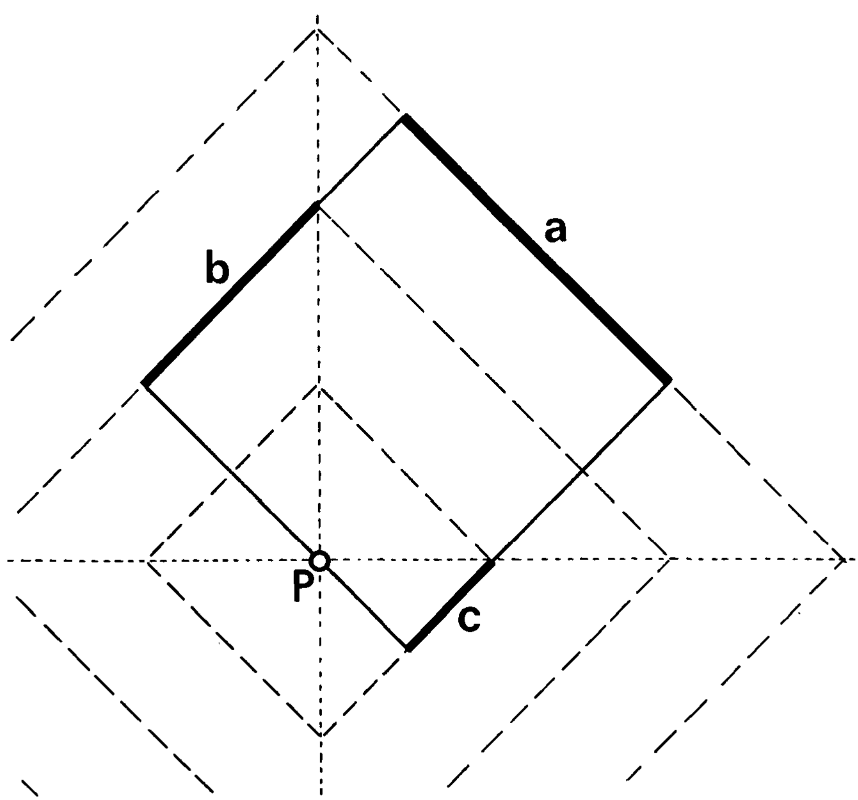

FIGURE 10
Demonstration of the prevalence of tied distances between points on an equidistance contour in the city-block metric.

given point) has the form of a diamond (or 45° square), as indicated by any one of the concentric dashed contours surrounding the point P. This implies that, for every point P, situated on a closed curve of the particular shape indicated by the solid curve, half of the remaining points on that same curve can be divided into three subsets (indicated by the heavier segments, a, b, and c) such that the distances from P to all points within any one of these subsets are tied. It follows that, if points (in finite number) are evenly distributed around the solid curve, no more than half of the interpoint distances will be distinct. According to the index of nondegeneracy proposed earlier, then, this situation is at least "half degenerate." But, as was illustrated in Fig. 5, whenever there is a choice between a more and a less degenerate configuration, the nonmetric method will move toward the more degenerate configuration, where the increased prevalence of ties permits the attainment of a lower value of stress.

When the point P coincides with any of the four corners of the solid curve, the three segments a, b, and c merge into one continuous region, which is coextensive with the two sides opposite P and within which *all* distances from P have the same value. Hence, a configuration of four points corresponding to these four corners has the completely degenerate property that all six interpoint distances have exactly the same value. This is in contrast to the Euclidean case, in which no more than three points (the vertices of an equilateral triangle) can be mutually equidistant, and in which a complete four-point degeneracy (the vertices of a regular tetrahedron) requires three dimensions.

In the case of higher-dimensional spaces, the contrast with the Euclidean metric becomes even more marked. Specifically, the maximum number of points, m, that can be arranged in k dimensions, so that the points in all $m(m - 1)/2$ pairs are separated by the same distance is given by $k + 1$, by $2k$, and by 2^k in the cases of the Euclidean, city-block, and dominance metrics, respectively (where the points then coincide with the vertices of the regular simplex, the regular cross-polytope, and the hypercube, respectively). Thus, the number of points that can be mutually equidistant tends, with increasing dimensionality, to be twice as great for the city-block metric as for the Euclidean and to increase exponentially faster than either for the dominance metric. Even in just three dimensions, the number of points for which complete degeneracy is possible is 4, 6, and 8, for the Euclidean, city-block, and dominance metrics. Thus, while it is not true in the Euclidean case, with a dominance metric we know in advance that in only three dimensions we can fit any similarity data for eight objects whatever, and that the resulting spatial configuration will preserve none of the structural information in those data.

The same phenomenon emerges in partial as well as complete degeneracy. Thus, for 16 points at the vertices of a regular 3×3 lattice in two dimensions, the number of distinct values of interpoint distance is 9, 6, and only 3 for the Euclidean, city-block, and dominance metrics. Likewise, for 27 points at the vertices of a regular $2 \times 2 \times 2$ lattice in three dimensions, the number of distinct values of distance is 9, 6, and only 2 for the same three metrics. And, for 16 points at the vertices of a regular $1 \times 1 \times 1 \times 1$ lattice (or hypercube) in four dimensions, the number of distinct values of distance is 4, 4, and only 1. Quite generally, then, tied distances, degeneracies and, hence, lower levels of stress are easier to come by with non-Euclidean and, particularly, with the dominance metric. Consequently, while the finding that the lowest stress is attainable for $r = 2$ may be evidence that the underlying metric is Euclidean, the finding that a lower stress is attainable for a value of r that is much smaller or larger may be artifactual.

These same investigations disclosed some other curious properties of these non-Euclidean metrics. One is that the completely degenerate configuration for the city-block and dominance metrics (unlike the regular simplex for the Euclidean) give the misleading appearance of containing structural information. Thus, in the cubical degeneracy for the three-dimensional dominance metric, points separated by one edge of the cube appear to be related to each other in a way that points separated by two edges (or a face diagonal) or by three edges (or a body diagonal) do not. But, in fact, all pairs of points are related in the same way, and the interpoint distances are unchanged by any permutation of the assignment of points to vertices. Another fact of perhaps greater practical relevance is that, in the cases of the two-, three-, and four-dimensional lattices considered above, the rank order of the interpoint distances for the Euclidean metric (though divided into more levels) is entirely consistent with the rank order for the dominance metric (and differs, at most, by a reversal of one pair of adjacent values from that for the city-block metric). Thus, although it might seem quite natural to generate a set of stimuli by using every combination of values from among two, three, or four levels on each of four, three, or two physical dimensions, respectively, nonmetric methods of analysis would be totally incapable of determining whether similarity data collected for such a set conformed with the Euclidean or dominance metrics (and would be extremely inefficient, at best, in revealing a better or worse correspondence to the city-block metric).

Finally, the limitations of the currently accepted practice of using nonmetric multidimensional scaling to test the appropriateness of the various Minkowski power metrics do not end with these practical and theoretical difficulties. There is the further limitation that this one-parameter class of so-called r-metrics is itself quite restricted. How much so may be seen from Fig. 11, which presents a hierarchy of some of the most thoroughly studied metric spaces, ranging from the most general, represented at the top, down to the most specific; represented at the bottom. Note that the class of Minkowski r-metric spaces occupies a relatively low position in this hierarchy.

This is because the distance formula (displayed in the rectangle for that class of spaces) is extremely restrictive and, in fact, entails that the unit sphere in k dimensions has exactly $2k$ prominences for $r < 2$ and exactly 2^k prominences for $r > 2$.

Much more general, is the class of general Minkowski spaces in which the unit spheres may have any convex, centrally symmetric shape. The conditions of symmetry and convexity of the unit spheres are connected with the distance axioms of symmetry and the triangle inequality, displayed in the top-most and next lower rectangles, respectively. An additional requirement, that these unit spheres be constant in size and shape throughout the space, entails that the general Minkowski space is isotropic and intrinsically flat. In this last respect, general Minkowski is, in turn, less general than Finsler space, in which these unit spheres can change continuously with location in the space. Thus, Finsler space permits both the locally non-Euclidean properties of flat Minkowski spaces and the globally non-Euclidean properties of intrinsically curved Riemann spaces (indicated on the left in Fig. 11). Finally, even Finsler space (along with all the more special cases arrayed below it) presupposes a continuous underlying co-

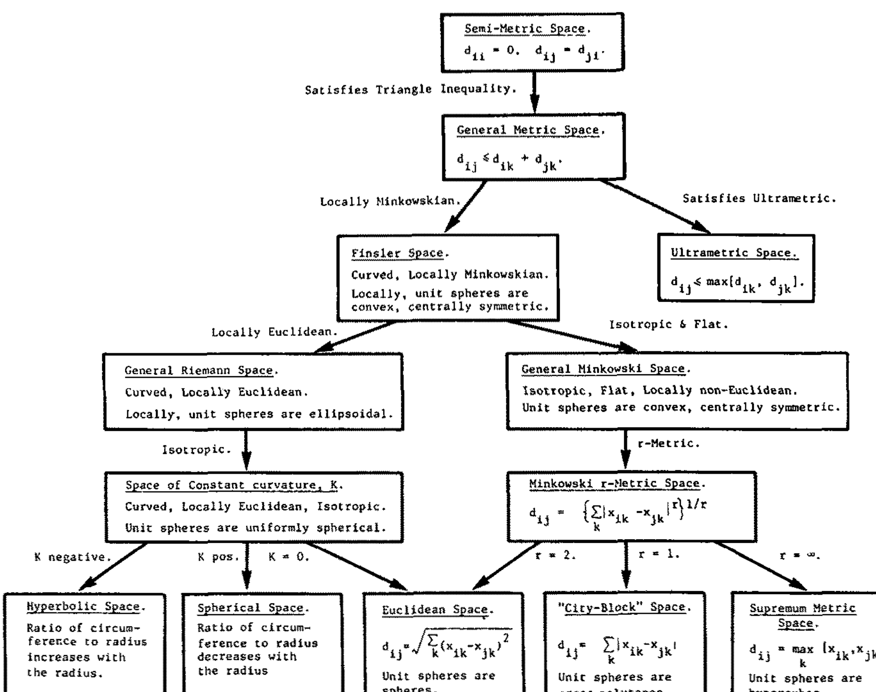

FIGURE 11
Hierarchy of some commonly considered metric and semimetric spaces.

ordinate space with its own intrinsic dimensionality and, so, is less general than the merely metric or semi-metric spaces. For, the constraints in these most general cases are defined solely in terms of interpoint distances and entail neither a specific dimensionality nor the embedability of a coordinate system. (For a fuller mathematical treatment of these various types of spaces see, *e.g.*, Beals, Krantz, and Tversky [1968]; Blumenthal [1959]; Busemann [1955]; and Rund [1959].)

Generally available methods for the analysis of similarity data have been designed to yield representations with metrics corresponding to only five of the thirteen boxes included in Fig. 11; namely, that of the ultrametric (as exemplified by Johnson's [1967] formulation of hierarchical clustering) and those of the Minkowski r-metric and, hence, the Euclidean, city-block, and dominance metrics (as already discussed). It is true (a) that general Riemann spaces and spaces of constant curvature have long been considered for representing, respectively, the perceived similarities among colors [*e.g.*, Silberstein, 1938; Silberstein & MacAdam, 1945] and the perceived distances among luminous points in a dark three-dimensional field [*e.g.*, Blank, 1958; 1959; Indow, in press; Luneburg, 1947, 1950]; (b) that general Minkowski spaces have been recognized as providing for the representation of considerably more diverse rules of combination than are representable just by the class of r-metrics [Shepard, 1964a]; and (c) that completely general metric spaces avoid the implication, possibly objectionable for the representation of semantic structure, of an underlying continuum. However, methods of multidimensional scaling have, up to now, not been fully extended to these cases.

Prospects

The strategy of obtaining Minkowski r-metric solutions for the same set of data using different values of r may still be useful in some cases. It apparently can provide evidence that the underlying metric is Euclidean or near-Euclidean. For those who wish to use this strategy, some savings in computation is possible owing to a kind of conjugate relationship that holds in certain cases between values of r above and below the Euclidean value of 2. Certainly in two dimensions there is no reason to seek optimum solutions both for $r = 1$ and for $r = \infty$. These two limiting metrics are identical in two-dimensions, except for a 45° rotation and uniform dilation or contraction of the configuration as a whole. Moreover, for intermediate values of r, Koopman and Cooper [1974] have just reported numerical results suggesting that, if r and r^* stand in the relation $1/r + 1/r^* = 1$, an approximate conjugacy holds in two dimensions such that, for practical purposes, it may be unnecessary to obtain two-dimensional solutions for values of r both above and below 2. Additional support for this suggestion may be found in a result that Phipps Arabie (personal communication) has obtained in further analyses of Ekman's [1954] color data. Corresponding to the r-value of about 2.5 (or 5/2) for which Kruskal [1964a, p. 24] had found that stress was minimum for these data, Arabie found a corresponding, conjugate minimum stress at an r-value of about 5/3 (in accordance with $2/5 + 3/5 = 1$).

However, even in two dimensions, this conjugacy does not hold strictly except in the limiting case of $r = 1$ and $r^* = \infty$. Although the proof is too long to include here, I established some years ago that, for any value of r in the open interval between 1 and 2, there is no value, r^*, greater than 2 such that the r-metric and the r^*-metric are exactly equivalent except for a rotation and change of scale. And, from the already mentioned fact that the number of points that can be arranged to be mutually equidistant in k-dimensions is $2k$ for the city-block metric but 2^k for the dominance metric, it is clear that conjugacy for $k > 2$ does not even hold in the limiting cases in which r and r^* approach 1 and ∞.

In any case, the very serious limitations already noted in the use of Minkowski r-metric solutions to investigate the underlying rule of combination has led me to explore quite different approaches. One which appears to avoid the practical and theoretical difficulties discussed above and to offer considerable generality (at least for the two-dimensional case), is based upon results that I obtained with the collaborative help of Doug Carroll and Jih-Jie Chang [Shepard, 1966]. Basically, these results amount to a demonstration that purely Euclidean solutions can be surprisingly robust in the face of certain kinds of rather marked departures from the assumed Euclidean metric. As I shall argue, this means that one can determine the form of the underlying unit circle and, hence, the nature of an underlying (flat and isotropic) space even though the metric is of the general Minkowski type or of still more general merely semimetric types.

Our investigation was based upon a set of artificial data derived from a square array of 50 random points. The Euclidean distances among these points were converted into very non-Euclidean and, in fact, merely semimetric "distances" by differentially stretching or shrinking all distances depending upon the angular orientation of the line connecting each pair of points in accordance with the six-lobed unit circle shown in Fig. 12A. The resulting "metric" was essentially in the spirit of the general Minkowski metric. But, since the unit circle was nonconvex and had six rather than four prominences the metric could not be approximated by any two-dimensional r-metric and was strictly speaking only a semimetric.

Nevertheless, when nonmetric multidimensional scaling (M-D-SCAL) was applied to these rather bizarre semimetric distances on the assumption that they were ordinary Euclidean distances, the obtained two-dimensional configuration achieved a close approximation to congruence with the randomly generated original configuration. For most of the individual points, the

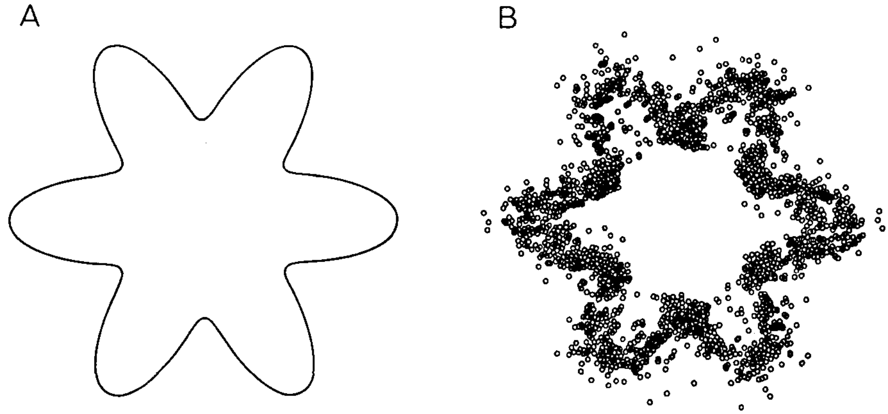

FIGURE 12
Six-lobed unit circle (A) and attempted reconstruction of that semimetric unit circle (B) on the basis of a Euclidean multidimensional scaling solution.

discrepancy in the recovered position of the point was small compared with the distance to even its nearest neighbor. When we then plotted, in polar coordinates, the ratio of the original (non-Euclidean) data to the recovered (Euclidean) distances as a function of the angle of the line between the two points in every pair, we obtained the scatterplot displayed in Fig. 12B. The six-lobed shape of the underlying unit circle reemerged quite clearly.

In the analysis of real data, of course, the similarity measures may be some unknown monotone function of the underlying (non-Euclidean) distances rather than those distances themselves. So, in general, it would not be appropriate merely to plot ratios as was done for Fig. 12B. Nevertheless, it should not be difficult to estimate, for each small angular interval, the (Euclidean) distance for which the function relating similarity to distance within that interval attains some specified level. I have outlined a procedure of this sort that, I am hopeful, will yield a good approximation to the desired unit circle. Depending upon whether the resulting plot is generally circular, square, or of some other (possibly even nonconvex) form, we could infer quite directly whether the metric is of the Euclidean, city-block (or dominance), or some other (perhaps merely semimetric) type.

The extension of existing methods of multidimensional scaling so as to yield solutions in spaces of constant curvature appears straightforward but, to my knowledge, has not been attempted. Perhaps the method that comes closest to being such an extension is the "nonmetric factor analysis" method, SSA-III, of Guttman and Lingoes [see Lingoes, 1972]. In this method scalar products of vectors, rather than distances, are brought into a best monotone fit with the similarity data. If the vectors were all constrained to be of the same length, the scalar products would become equivalent to distances on the surface of a hypersphere—a surface of constant positive curvature. Particularly since more recent findings regarding the binocular perception of luminous points in space have been in some ways inconsistent with Luneburg's [1947] model of visual space as a (hyperbolic) manifold of constant negative curvature [Foley, 1972], however, there has been no insistent demand for methods designed to yield representations in which the curvature is constrained to be constant, whether positive or negative.

Doug Carroll and I have pursued a quite different approach that appears to offer the appreciably greater generality of allowing for cases in which the intrinsic curvature of the underlying space may vary from region to region (as in general Riemann spaces) and, at the same time, the metric may behave locally as a general Minkowski metric or even as a still more general semimetric (such as that associated with the nonconvex unit circle illustrated above). Such an approach provides for the possibility that the underlying psychological space may be a kind of Finsler space or even some semimetric generalization of such a continuous space. This generality is achieved by abandoning the requirement that the triangle inequality be satisfied while retaining the requirement that the similarity between the objects represented by two points vary in an appropriately smooth and continuous manner with the positions of those points in the underlying space. The analysis itself is carried out by applying, to the given similarity data, a method of "parametric mapping" to optimize an index, developed by Carroll, for measuring departure from a smooth or "continuous" relation between the data and the coordinates for the points in the underlying parameter space [Shepard & Carroll, 1966].

As one test of this approach, again carried out with the collaborative assistance of Doug Carroll and Jih-Jie Chang [Shepard, 1968], the already-described square configuration of 50 random points was converted into a toroidal configuration by identifying opposite edges of the square and redefining distances so as to be non-Euclidean in two respects; local and global. The locally non-Euclidean structure was induced by stretching and shrinking each distance as a function of its angle in accordance with the six-lobed unit circle in Fig. 12A; and the globally non-Euclidean structure was secured by taking, as the distance between any two points, the shortest such distance within the surface of the torus (and so, where such a path is shorter, across what were previously bounding edges of the square). Finally, these doubly non-Euclidean distances were converted into artificial similarity measures by means of an exponential decay transformation.

Despite the intrinsic two-dimensionality of the underlying space, the best-fitting two-dimensional solution obtained by standard nonmetric multidimensional scaling (M-D-SCAL) failed (a) to achieve a good fit to the data, (b) to achieve an accurate representation of any (either opened out or projected down) version of the toroidal configuration, and (c) to permit a recognizable reconstruction of the six-lobed unit circle. It was only by going to a four-dimensional space, within which the torus itself can be isometrically embedded, that a satisfactory fit was achieved. But, because the four-dimensional solution did not achieve a reduction to the correct intrinsic dimensionality of the toroidal surface itself, it was still not possible to reconstruct the unit circle and, hence, to infer the locally non-Euclidean form of the underlying metric.

The best-fitting two-dimensional solution obtained by the method of parametric mapping did, after some difficulties with apparently merely local minima, achieve what corresponded to an opened-out version of the toroidal surface (akin to those presented in Shepard and Carroll, [1966] and in Shepard and Cermak [1973]). Moreover, the construction of a scatterplot in the manner described for Fig. 12B again revealed the six-lobed form of the unit circle. Although in our experience the method of parametric mapping, even more than standard methods of nonmetric multidimensional scaling, is thus beset with problems of slow convergence and local minima, I believe that some of the suggestions made earlier for the construction of good initial configurations may substantially increase the effectiveness of the method. If so, the representation of similarities in terms of quite general semimetric but continuous coordinate spaces becomes a viable alternative.

In this just-considered generalization, the triangle inequality was abandoned while the notion of a continuous underlying coordinate space was retained. The method of "maximum variance nondimensional scaling" that Jim Cunningham and I have recently been exploring represents a very different kind of generalization in which the triangle inequality is made central while the requirement of an underlying coordinate space is relinquished [Cunningham & Shepard, 1974]. In order to obtain unique nontrivial solutions, the requirement of minimum dimensionality (which has no meaning in the absence of a continuous space) was replaced by a requirement of maximum variance of the distances (which does and which, according to the above-reported results on reduction of dimensionality, seems to have a very similar effect).

In test analyses with both real and artificial data, we found that, if we thus maximize the variance of the distances subject only to (a) the strict satisfaction of just the metric axioms (particularly the triangle inequality) and (b) adequate maintenance of monotonicity, we can in fact recover underlying Euclidean or non-Euclidean distances with considerable accuracy without any assumption or use of a coordinate embedding space. Since we are able to recover the underlying distances, we are also able to recover the form of the unknown function relating the similarity data to the distances. In fact, in a test with a rather exotically non-Euclidean metric (sum-over-path distances in a graph-theoretic tree), the sigmoid form of this (artificially imposed) functional relation was recovered with great precision, while a straightforward application of standard nonmetric multidimensional scaling to the same artificial data gave a very poor fit and approximation to the sigmoid function. (Compare Figs. 3 and 4 in Cunningham & Shepard [1974].)

For many practical purposes, two drawbacks of this method may limit its general usefulness. First, although apparently not subject to local minima, so far it has tended to converge rather slowly and, so, to be relatively costly—particularly for large matrices. Second, it does not, of course, yield a spatial configuration in a two- or three-dimensional coordinate space of the sort that is usually sought for purposes of substantive interpretation. Nevertheless, it is of considerable theoretical interest as a demonstration of the possibility of recovering very general, merely metric representations—corresponding to the next to the top box in Fig. 11. Moreover, as already suggested, it should find some uses for the study of the form of the function relating distances, which are potentially very non-Euclidean, to similarity measures of various types (such as those of stimulus generalization or reaction time) and, possibly, for the generation of starting configurations for methods that are particularly susceptible to local minima. Finally, as I shall suggest in the next and final section, it may provide one basis for constructing more discrete, network-like or graph-theoretic representations such as have been proposed for certain, *e.g.*, semantic, domains.

6. *Representing Discrete or Categorical Structure*

Problem

Standard methods of multidimensional scaling and, in fact, nearly all of the methods discussed above have been based upon the assumption of an underlying space that is continuous and has a well-defined dimensionality. (The only exceptions discussed here were the two nondimensional methods of hierarchical clustering [Johnson, 1967] and maximum variance scaling [Cunningham & Shepard, in press].) Methods for mapping the data into a continuous coordinate space seem eminently appropriate for the investigation of domains of objects in which there is an underlying continuous physical variation—as there clearly is, for example, with colors which vary continuously in brightness, hue, and saturation; with tones which vary continuously in intensity, frequency, and duration; or even with facial expressions which also vary continuously (even though the physical dimensions of the variation are not yet completely specified). Experience indicates that, even when the stimuli vary only in discrete steps (as with the dot-and-dash signals of the Morse Code [Shepard, 1963a]), a representation within a continuous coordinate space is often quite satisfactory—particularly if the number of stimuli and number of steps of variation are not too small.

Other domains, particularly those of a more conceptual, linguistic, or semantic nature, appear to be inherently more discrete, categorical, or bipolar. With regard to the perception of speech, for instance, suggestive examples are provided by the empirical phenomenon of "categorical perception" [Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967] and by the related theoretical notion of "distinctive features" [Halle, 1964; Miller & Nicely, 1955]. And with regard to purely internal conceptual or semantic systems, inherently discrete structures seem to be the rule rather than the exception. Some especially clear-cut examples include the cognitive systems of kin terms [Haviland & Clark, 1974; Romney & D'Andrade, 1964], of linguistic marking [Clark, 1969, 1970; Greenberg, 1966], and of binary, componential, and algebraic structures in general [*e.g.*, Boyd, 1972; Lévi-Strauss, 1967]. I myself have several times argued that, although a frozen spatial configuration may well represent the perceived relations of pair-wise similarities among stimuli either on the average or else for any one individual at any one time, such a configuration does not exhaust the cognitive structure of a set of stimuli. In particular it may not adequately represent the ways in which different subjects or the same subject at different times may see subsets of the stimuli as grouped together or as having some property in common [Shepard, 1963b, 1964a; Shepard & Cermak, 1973; Shepard, Hovland & Jenkins, 1961].

It is of course true that subsets of scaled objects with strong mutual relations of similarity will tend to show up as visibly clustered groups of points even in the "continuous" spatial representations obtained by multidimensional scaling. Indeed, in their strongest form, such clusterings may take one of the more dramatic forms of complete degeneracy illustrated in Fig. 6. Moreover, objective methods of rotation or of affine transformation can bring out the discrete or discontinuous aspects of such a configuration more fully [Degerman, 1970; Kruskal, 1972; Torgerson, 1965]. Still, such basically continuous spatial representations, even when linearly transformed, do not (any more than transformations to simple structure in factor analysis) yield discrete subsets, as such, explicitly. Among other things, this makes it difficult to evaluate whether a seeming cluster is valid or reliable—in the sense that it would be if it repeatedly emerged, explicitly, in the analyses of independent sets of data.

It is also true that methods of hierarchical clustering [Jardine & Sibson, 1971; Johnson, 1967; Sokal & Sneath, 1963] do yield both explicit and categorical structures that, moreover, have been found to be quite useful and reliable in the representation of speech sounds (*e.g.*, see Fig. 8 and Shepard [1972c]) and of concepts (*e.g.*, see Fig. 1 and Shepard *et al.*, [in press]). However, the requirement that the clusters be hierarchically nested seems undesirably restrictive for many purposes. In Fig. 8A it means that, once just the back fricatives [z] and [ʒ] have been grouped with the unvoiced stops [d] and [g] (perhaps by virtue of their place of articulation), neither of these back fricatives can be grouped with either of the front fricatives [v] and [ð] (on the basis solely of their affrication). And in Fig. 1 it means that, once "cat" is grouped with the other household pet "dog," it can no longer be grouped with the other felines "lion" and "tiger." In short, although hierarchical systems can represent some of the discrete or categorical structure underlying a set of similarity data, it can not represent psychological properties, however salient, that correspond to overlapping subsets.

For the same reason, hierarchical systems can not represent parallel or analogical correspondences between the structures within two nonoverlapping subsets (such as the parallelism in articulation between the voiced and unvoiced consonants [Shepard, 1972c, p. 106]). The representation of such parallelisms requires the specification of connections (representable only as overlapping clusterings) between the subparts of disjoint clusters. And, finally, although a hierarchical clustering is equivalent to a graph-theoretic tree, arbitrary graph structures (containing closed paths or cycles) are precluded. So hierarchical representations can not, any more than continuous spatial representations, furnish the sorts of general graphs or networks currently being advocated for the representation of semantic structure [Anderson & Bower, 1973; Quillian, 1968; Rumelhart, Lindsay, & Norman, 1972].

Prospects

One already-noted limitation of maximum variance nondimensional scaling is that, because it drops the restrictive requirement that the distances be embedded in a coordinate space, it forfeits the pictorially presentable spatial configuration that has proved to be so useful in substantive applications of multidimensional scaling. Jim Cunningham and I are currently exploring a possible addition to this method of nondimensional scaling that we hope will be able to convert the obtained set of coordinate-free distances into a graph structure that will meet this need for a more picturable representation. If we succeed, the sort of general graph structure that is produced should also be much closer to the kind of discrete, network-like representations proposed for semantic memory. The distance estimates furnished by such maximum variance scaling appear ideally suited for the purposes of constructing a graph-theoretic representation because the maximization of variance together with the maintenance of the triangle inequality tends, wherever possible, to render distances additive—as they should be over a connected path through a graph. (Thus, as noted before, artificial data generated from sum-over-path distances were well fit by maximum variance nondimensional scaling but not by standard nonmetric multidimensional scaling.) As a first step toward the proposed addition, Cunningham [1974] has already reported encouraging results in the fitting of graphs of one particular type; namely, tree structures.

A last type of method for the representation of structure in similarity data to be considered here takes an entirely different approach. Although it is perhaps closest in spirit to the approach taken by Johnson [1967] in his formulation of hierarchical clustering, it departs from all methods for hierarchical clustering in abandoning the very strong restriction that the clusters never overlap. And, although it shares with both of the nondimensional methods mentioned (*viz.*, those of hierarchical clustering and maximum variance scaling) the abandonment of an underlying coordinate space, it goes beyond both of those methods in abandoning, also, the notion of distance.

As an initial basis for the exploratory development, with Phipps Arabie, of a method of this type, I chose a model according to which the perceived similarity between any two objects is a simple sum of the psychological weights associated with all and only those (discrete) properties that the two objects both share. Formally,

$$s_{ij} = \sum_{k=1}^{m} w_k p_{ik} p_{jk} ,$$

$$\text{where } p_{ik} = \begin{cases} 1 & \text{if object } i \text{ has property } k \\ 0 & \text{otherwise,} \end{cases}$$

and where w_k is a non-negative weight representing the psychological salience of property k.

The computational problem with which we are faced is that of finding a minimum set of m subsets of the objects and associated optimum weights such that the maximum possible variance of the similarities, s_{ij}, is accounted for. Although the model is closely related to the standard factor-analytic model, the restriction of the variables p_{ik} to binary values converts the computational problem into a much more difficult, combinatorial one. As we have currently developed it, the computer program, ADCLUS, for this type of additive cluster analysis [Arabie & Shepard, 1974] proceeds in two successive phases; a nonmetric and then a metric one. In the first phase, combinatorial methods are used to generate an ordered list of subsets with potentially positive weights that is invariant under monotone transformations of the data. In the second phase, a modified gradient method is then used to estimate optimum weights for these subsets and to eliminate all subsets for which the weights become sufficiently small.

As an illustration, Table 1 presents our nonhierarchical reanalysis of Miller and Nicely's [1955] data on the confusions among 16 consonant phonemes in the presence of white noise. These are the same data that have already been analyzed by a variety of more standard methods of multidimensional scaling and hierarchical clustering (see Fig. 8, Shepard [1972c] and, also, Wish and Carroll [in press]). In this case, approximately 99% of the variance was accounted for by the first 30 subsets (a representation that, in terms of number of parameters, is roughly comparable to the two-dimensional spatial solution which also accounted for about 99% of the variance). The first 16 subsets (roughly comparable to a one-dimensional solution) are listed in Table 1, in rank order according to their estimated weights. The obtained subsets appear to be readily interpretable except, possibly, for the four (ranked 3, 8, 10, and 13) for which the interpretations are given in parentheses. And the last three of these subsets, despite their uncertain interpretations, are very likely reliable in view of the facts (a) that the 8th (consisting of [b], [v], and [ð]) emerged repeatedly in hierarchical analyses of independent sets of data [Shepard, 1972c], (b) that the 13th (consisting of [p], [f], and [θ]) is, according to distinctive feature schemes, identical in structure to the 8th except for voicing, and (c) that the 10th (consisting just of [b] and [v]) is contained entirely within the 8th. What these results, along with those in Shepard [1972c], seem to indicate is the need for some revision of the distinctive feature schemes on the basis of which the interpretations were attempted.

The results show a marked departure from a strictly hierarchical structure. The overlapping nature of the obtained subsets is apparent in Fig. 13, where the first 16 are embedded as closed curves in the spatial configuration used earlier for Fig. 8A. Note, for example, that the relatively back fricatives, [z] and [ʒ], now cluster both with the relatively back stops, [g], and [d], and *also* (through [z]) with the relatively front fricatives, [v] and [ð]. Like-

TABLE 1
Nonhierarchical Additive Cluster Analysis of Confusion Measures of Similarity Among 16 Consonants
(Data from Miller and Nicely [1955])

Rank	Weight	Elements of Subset	Interpretation
1st	.282	f θ	front unvoiced fricatives
2nd	.214	d g	back voiced stops
3rd	.196	p k	(unvoiced stops omitting t)
4th	.157	p t k	unvoiced stops
5th	.155	v ð	front voiced fricatives
6th	.129	m n	nasals
7th	.090	θ s	middle unvoiced fricatives
8th	.074	b v ð	(front voiced consonants)
9th	.071	s ʃ	back unvoiced fricatives
10th	.061	b v	(front voiced consonants)
11th	.050	d g z ʒ	back voiced consonants
12th	.044	z ð	middle voiced fricatives
13th	.044	p f θ	(front unvoiced consonants)
14th	.035	z ʒ	back voiced fricatives
15th	.033	v z ð	front & middle voiced fricatives
16th	.030	g z ʒ	back voiced consonants

wise, the front unvoiced stop [p] clusters both with the other unvoiced stops, [t] and [k], and (as noted above) with the relatively front unvoiced fricatives, [f], [θ]. Finally, the overlapping clusters form chains connecting the four progressively further back unvoiced fricatives, [f], [θ], [s], and [ʃ], and, in parallel fashion, the four progressively further back voiced fricatives, [v], [ð], [z], and [ʒ], in a way that was not possible in the representation (Fig. 8) obtained by the hierarchical method.

The results of this early test application of this nonhierarchical method of additive cluster analysis encourage me to believe that methods based on models quite different from those underlying standard methods of multidimensional scaling can provide potentially useful, complementary information about the psychological structure underlying a set of objects. Our current efforts are being directed toward improving the efficiency of the iterative method used to adjust the weights and to eliminate unimportant

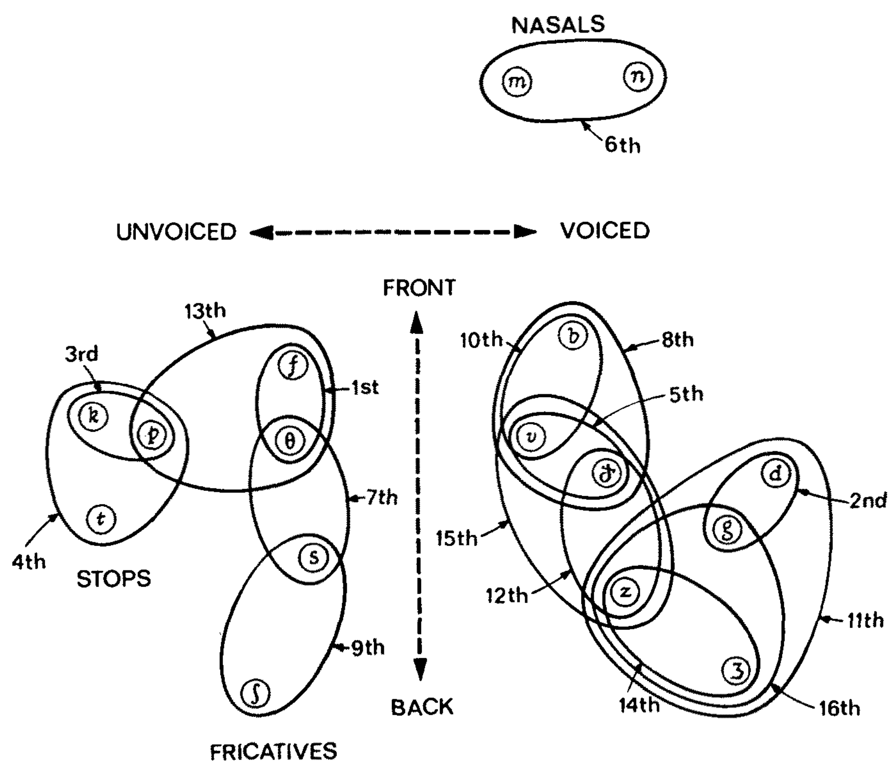

FIGURE 13

The first 16 clusters from Table 2, embedded in the two-dimensional spatial configuration (of Fig. 8A) representing the same confusion measures of similarity among the 16 consonant phonemes.

subsets, and toward investigating the stability of the weights for independent sets of data and as more or fewer subsets are eliminated.

CONCLUSION

After struggling with the problem of representing structure in similarity data for over 20 years, I find that a number of challenging problems still remain to be overcome—even in the simplest case of the analysis of a single symmetric matrix of similarity estimates. At the same time, I am more optimistic than ever that efforts directed toward surmounting the remaining difficulties will reap both methodological and substantive benefits. The methodological benefits that I forsee include both an improved efficiency and a deeper understanding of "discovery" methods of data analysis. And the substantive benefits should follow, through the greater leverage that such methods will provide for the study of complex empirical phenomena—perhaps particularly those characteristic of the human mind.

REFERENCES

Abelson, R. P., and Tukey, J. W. Efficient conversion of nonmetric information into metric information. *Proceedings of the American Statistical Association Meetings*, Social Statistics Section, 1959, 226–230.

Anderson, J. R., and Bower, G. H. Human associative memory. Washington, D. C.: V. H. Winston & Sons, 1973.

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

Arabie, P. and Boorman, S. A. Multidimensional scaling of measures of distance between partitions. *Journal of Mathematical Psychology*, 1973, **10**, 148–203.

Arabie, P. and Shepard, R. N. Representation of similarities as additive combinations of discrete overlapping properties. Presented at the Mathematical Psychology Meeting in Montreal, August, 1973.

Arnold, J. B. A multidimensional scaling study of semantic distance. *Journal of Experimental Psychology Monograph*, 1971, **90**, 349–372.

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

Beals, R., Krantz, D. H., and Tversky, A. Foundations of multidimensional scaling. *Psychological Review*, 1968, **75**, 127–142.

Bennett, R. S. The intrinsic dimensionality of signal collections. Doctoral dissertation, The Johns Hopkins University, 1965. University Microfilms, Inc., Ann Arbor, Michigan.

Bennett, R. S. The intrinsic dimensionality of signal collections. *IEEE Transactions on Information Theory*, 1969, Vol. IT-15, No. 5, 517–525.

Blank, A. A. Axiomatics of binocular vision: The foundations of metric geometry in relation to space perception. *Journal of the Optical Society of America*, 1958, **48**, 328–334.

Blank, A. A. The Luneburg theory of binocular space perception. In S. Koch (Ed.), *Psychology: A study of a science*. New York: McGraw-Hill, 1959. Pp. 395–426.

Blumenthal, L. M. *Theory and applications of distance geometry*. Oxford: Clarendon Press, 1953.

Boyd, J. P. Information distance for discrete structures. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences*, Vol. I. New York: Seminar Press, 1972. Pp. 213–223.

Busemann, H. *The geometry of geodesics.* New York: Academic Press, 1955.

Carroll, J. D. Individual differences and multidimensional scaling. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences*, Vol. I. New York: Seminar Press, 1972. (a) Pp. 105–155.

Carroll, J. D. Algorithms for rotation and interpretation of dimensions and for configuration matching. Bell Telephone Laboratories, 1972. (b) (Multilithed handout prepared for the workshop on multidimensional scaling, held at the University of Pennsylvania in June, 1972.)

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

Chang, J.-J., and Shepard, R. N. Exponential fitting in the proximity analysis of confusion matrices. Presented at the annual meeting of the Eastern Psychological Association, New York, April 14, 1966.

Clark, H. H. Linguistic processes in deductive reasoning. *Psychological Review*, 1969, **76**, 387–404.

Clark, H. H. Word associations and linguistic theory. In J. Lyons (Ed.), *New horizons in linguistics.* Baltimore, Maryland: Penguin Books, 1970. Pp. 271–286.

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

Cross, D. V. Metric properties of multidimensional stimulus generalization. In D. I. Mostofsky (Ed.), *Stimulus generalization.* Stanford, Calif.: Stanford University Press, 1965. Pp. 72–93.

Cunningham, J. P. On finding an optimal tree realization of a proximity matrix. Presented at the mathematical psychology meeting, Ann Arbor, Michigan, August 29, 1974.

Cunningham, J. P., and Shepard, R. N. Monotone mapping of similarities into a general metric space. *Journal of Mathematical Psychology*, 1974, **11**, (in press).

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

Ekman, G. Dimensions of color vision. *Journal of Psychology*, 1954, **38**, 467–474.

Fillenbaum, S., and Rapoport, A. *Structures in the subjective lexicon.* New York: Academic Press, 1971.

Foley, J. M. The size-distance relation and the intrinsic geometry of visual space. *Vision Research*, 1972, **12**, 323–332.

Greenberg, J. H. *Language universals.* The Hague: Mouton, 1966.

Guttman, L. A new approach to factor analysis: The radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences.* Glencoe, Illinois: Free Press, 1954. Pp. 258–348.

Guttman, L. A generalized simplex for factor analysis. *Psychometrika*, 1955, **20**, 173–192.

Guttman, N., and Kalish, H. I. Discriminability and stimulus generalization. *Journal of Experimental Psychology*, 1956, **51**, 79–88.

Halle, M. On the bases of phonology. In J. A. Fodor and J. J. Katz (Eds.), *The structure of language.* Englewood Cliffs, N. J.: Prentice-Hall, 1964. Pp. 324–333.

Harshman, R. A. Foundations of the parafac procedure: Models and conditions for an explanatory multi-modal factor analysis. Unpublished doctoral dissertation, University of California at Los Angeles, 1970.

Haviland, S. E., and Clark, E. V. 'This man's father is my father's son': A study of the acquisition of English kin terms. *Journal of Child Language*, 1974, **1**, 23–47.

Helm, C. E. Multidimensional ratio scaling analysis of perceived color relations. *Journal of the Optical Society of America*, 1964, **54**, 256–262.

Hyman, R., and Well, A. Judgments of similarity and spatial models. *Perception and Psychophysics*, 1967, **2**, 233–248.

Hyman, R., and Well, A. Perceptual separability and spatial models. *Perception and Psychophysics*, 1968, **3**, 161–165.

Indow, T. Applications of multidimensional scaling in perception. In E. C. Carterette and M. P. Friedman (Eds.), *Handbook of perception*, Vol. 2. New York: Academic Press, in press.

Jardine, N., and Sibson, R. *Mathematical taxonomy*. New York: Wiley, 1971.

Johnson, S. C. Hierarchical clustering schemes. *Psychometrika*, 1967, **32**, 241–254.

Klahr, D. A Monte Carlo investigation of the statistical significance of Kruskal's nonmetric scaling procedure. *Psychometrika*, 1969, **34**, 319–330.

Koopman, R. F., and Cooper, M. Some problems with Minkowski distance models in multidimensional scaling. Presented at the annual meeting of the Psychometric Society, Stanford University, March 28, 1974.

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

Kruskal, J. B. Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, 1964, **29**, 28–42. (b)

Kruskal, J. B. Linear transformation of multivariate data to reveal clustering. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences*. Vol I. New York: Seminar Press, 1972. Pp. 179–191.

Kruskal, J. B. Multidimensional scaling and other methods for discovering structure. In A. J. Ralston, H. S. Wilf, and K. Enslein, (Eds.), *Statistical methods for digital computers*. New York: Wiley, in press.

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

Lévi-Strauss, C. *The savage mind*. Chicago: University of Chicago Press, 1967.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., and Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, **74**, 431–461.

Lingoes, J. C. A general survey of the Guttman-Lingoes nonmetric program series. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences*. Vol. I. New York: Seminar Press, 1972.

Luneburg, R. K. *Mathematical analysis of binocular vision*. Princeton, N. J.: Princeton University Press, 1947.

Luneburg, R. K. The metric of binocular visual space. *Journal of Optical Society of America*, 1950, **40**, 637–642.

Miller, G., and Nicely, P. E. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 1955, **27**, 338–352.

Osgood, C. E., Suci, G. J., and Tannenbaum, P. H. *The measurement of meaning*. Urbana, Ill.: University of Illinois Press, 1957.

Peters, R. W. Dimensions of perception of consonants. *Journal of the Acoustical Society of America*, 1963, **35**, 1985–1989.

Quillian, M. R. Semantic memory. In M. Minsky (Ed.), *Semantic information processing*. Cambridge, Mass.: M.I.T. Press, 1968.

Rips, L. J., Shoben, E. J., and Smith, E. E. Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 1–20.

Romney, A. K., and D'Andrade, R. G. Cognitive aspects of English kinship terms. *American Anthropologist*, 1964, **66**, 146–170.

Rumelhart, D. E., and Abrahamson, A. A. A model for analogical reasoning. *Cognitive Psychology*, 1973, **5**, 1–28.

Rumelhart, D. E., Lindsay, P. H., and Norman, D. A. A process model for long-term memory. In E. Tulving and W. Donaldson (Eds.), *Organization of memory.* New York: Academic Press, 1972.

Rund, H. *The differential geometry of Finsler space.* Berlin: Springer-Verlag, 1959.

Shepard, R. N. Multidimensional scaling of concepts based upon sequences of restricted associative responses. *American Psychologist,* 1957, **12**, 440–441. (abstract) (a)

Shepard, R. N. Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika,* 1957, **22**, 325–345. (b)

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

Shepard, R. N. Stimulus and response generalization: Tests of a model relating generalization to distance in psychological space. *Journal of Experimental Psychology,* 1958, **55**, 509–523. (b)

Shepard, R. N. Similarity of stimuli and metric properties of behavioral data. In H. Gulliksen and S. Messick (Eds.), *Psychological scaling: Theory and applications.* New York: Wiley, 1960. Pp. 33–43.

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

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

Shepard, R. N. Analysis of proximities as a technique for the study of information processing in man. *Human Factors,* 1963, **5**, 33–48. (a)

Shepard, R. N. Comments on Professor Underwood's paper "Stimulus selection in verbal learning." In C. N. Cofer and B. S. Musgrave (Eds.), *Verbal behavior and learning: Problems and processes.* New York: McGraw-Hill, 1963. Pp. 48–70. (b)

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

Shepard, R. N. Polynomial fitting in the analysis of proximities. *Proceedings of the XVIIth international congress of psychology.* Amsterdam: North-Holland Publisher, 1964, 345–346. (abstract) (b)

Shepard, R. N. Approximation to uniform gradients of generalization by monotone transformations of scale. In D. I. Mostofsky (Ed.), *Stimulus Generalization.* Stanford, Calif.: Stanford University Press, 1965. Pp. 94–110.

Shepard, R. N. Computer explorations of psychological space. Invited research address presented at the annual meeting of the American Psychological Association, New York, September 3, 1966. (a)

Shepard, R. N. Metric structures in ordinal data. *Journal of Mathematical Psychology,* 1966, **3**, 287–315. (b)

Shepard, R. N. Continuity versus the triangle inequality as a central principle for the spatial analysis of similarity data. Presented in a symposium on alternative models for the geometric representation of psychological data, at the mathematical psychology meetings, Stanford University, August 28, 1968.

Shepard, R. N. Introduction to Volume I. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences,* Vol. I. New York: Seminar Press, 1972. Pp. 1-20. (a)

Shepard, R. N. A taxonomy of some principal types of data and of multidimensional methods for their analysis. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences,* Vol. I. New York: Seminar Press, 1972. Pp. 21–47. (b)

Shepard, R. N. Psychological representation of speech sounds. In E. E. David & P. B. Denes (Eds.), *Human communication: A unified view.* New York: McGraw-Hill, 1972. Pp. 67–113. (c)

Shepard R. N., and Carroll, J. D. Parametric representation of nonlinear data structures. In P. R. Krishnaiah (Ed.), *Multivariate analysis: Proceedings of an international symposium.* New York: Academic Press, 1966. Pp. 561–592.

Shepard, R. N., and Cermak, G. W. Perceptual-cognitive explorations of a toroidal set of free-form stimuli. *Cognitive Psychology*, 1973, **4**, 351–377.

Shepard, R. N., Hovland, C. I., and Jenkins, H. M. Learning and memorization of classifications. *Psychological Monographs*, 1961, **75**, (13, whole no. 517).

Shepard, R. N., Kilpatric, D. W., and Cunningham, J. P. The internal representation of numbers. *Cognitive Psychology*, in press.

Silberstein, L. Investigations of the intrinsic properties of the color domain. *Journal of the Optical Society of America*, 1938, **28**, 63–85.

Silberstein, L., and MacAdam, D. L. The distribution of color matchings around a color center. *Journal of the Optical Society of America*, 1945, **35**, 32–39.

Sokal, R. R., and Sneath, P. H. A. *Principles of numerical taxonomy.* San Francisco: W. H. Freeman, 1963.

Stenson, H. H., and Knoll, R. L. Goodness of fit for random rankings in Kruskal's nonmetric scaling procedure. *Psychological Bulletin*, 1969, **71**, 122–126.

Thomas, H. Spatial models and multidimensional scaling of random shapes. *American Journal of Psychology*, 1968, **81**, 551–558.

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

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

Torgerson, W. S. Multidimensional scaling of similarity. *Psychometrika*, 1965, **30**, 379–393.

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

Wish, M., and Carroll, J. D. Applications of "INDSCAL" to studies of human perception and judgment. In E. C. Carterette and M. P. Friedman (Eds.), *Handbook of perception*, Vol. 2. New York: Academic Press, in press.

Young, F. W. Nonmetric multidimensional scaling: Recovery of metric information. *Psychometrika*, 1970, **35**, 455–473.

Young, F. W., and Torgerson, W. S. TORSCA, A FORTRAN IV program for Shepard-Kruskal multidimensional scaling analysis. *Behavioral Science*, 1967, **12**, 498.