11. J. N. Driscoll and A. W. Berger, *Improved Chemical Methods for Sampling and Analysis of Gaseous Pollutants from the Combustion of Fossil Fuels*, vol. 1, *Sulfur Oxides* (Report PF-209267, U.S. Environmental Protection Agency, Washington, D.C., 1971).
12. Committee on Medical and Biological Effects of Environmental Pollutants, *Nitrogen Oxides* (National Academy of Sciences, Washington, D.C., 1977).
13. W. T. Donaldson, *Environ. Sci. Technol.* 11, 348 (1977).
14. Committee on Biological Effects of Atmospheric Pollutants, *Particulate Polycyclic Organic Matter* (National Academy of Sciences, Washington, D.C., 1972).
15. L. L. Lamparski, T. J. Nestrick, R. H. Stehl,

Anal. Chem. 51, 2273 (1979); T. J. Nestrick and L. L. Lamparski, in preparation.
16. Analytical method ML-AM-73-97 consists of a base digestion, an acid wash of the hexane extract, and cleanup with alumina and silica gel columns. Measurements are made by low-resolution GC-MS.
17. Analytical method ML-AM-78-63 includes Soxhlet extraction through a silica gel bed followed by cleanup through beds of silica, silver nitrate-modified silica, and basic alumina. Various chlorinated dioxins are separated by high-pressure liquid chromatography and measured by low-resolution GC-MS.
18. Environmental Protection Agency methods 1, 2, 3, and 5, *Fed. Regist.* 42, 41754 (18 August 1977).

19. H. Freeman, *Environ. Sci. Technol.* 12, 1253 (1978).
20. H. H. Gill, Dow Chemical Co., unpublished report (1965).
21. R. H. Stehl, R. R. Papenfuss, R. A. Bredeweg, R. W. Roberts, in *Dioxins—Origin and Fate*, E. Blair, Ed. (American Chemical Society, Washington, D.C., 1973), p. 121; H. Prophet, Dow Chemical Company, unpublished report (1970).
22. "Disposition of Orange herbicide by incineration," Department of the Air Force, Final Environmental Statement, November 1974.
23. "At-sea incineration of herbicide Orange onboard the M/T *Vulcanus*" (Report EPA-600/2-7-086, Environmental Protection Agency, Washington, D.C., April 1978).

Multidimensional Scaling, Tree-Fitting, and Clustering

Roger N. Shepard

Without any quantitative information about the physical properties of colors, tones, speech sounds, or words, we can learn something about how humans process such stimuli from an analysis of ratings of perceived similarity, frequencies with which the stimuli are actually confused with each other, latencies of discriminative responses, or, in the case of infants and other animals, magnitudes of the "orienting reflex" when one stimulus is substituted for the other (*1, 2*). This

physical dimensions are as yet poorly characterized, and it is essential in the case of symbolic stimuli such as words, for which the relevant semantic dimensions are not even present in the physical stimuli.

Early History

Proposals that stimuli be modeled by points in a space in such a way that per-

Summary. American mathematical psychologists have developed computer-based methods for constructing representations of the psychological structure of a set of stimuli on the basis of pairwise measures of similarity or confusability. Applications to perceptual and semantic data illustrate how complementary aspects of the underlying psychological structure are revealed by different types of representations, including multidimensional spatial configurations and nondimensional tree-structures or clusterings.

purely psychological approach has advantages over a psychophysical one in the case of complex, naturalistic stimuli such as faces, for which the relevant

ceived similarity is represented by spatial proximity go back to the suggestions of Isaac Newton (*3*) that spectral hues be represented on a circle, of Helmholtz and Schrödinger (*4*) that colors in general be represented in a curved Riemannian manifold, of Drobisch (*5*) that pure tones be represented on a helix, and of Henning (*6*) that odors and tastes be represented within a prism and a tetrahedron, respectively. However, little prog-

The author is professor of psychology at Stanford University, Stanford, California 94305. This article was presented as an invited lecture on 2 April 1979 at the U.S.-U.S.S.R. symposium on "Normative and Descriptive Models of Decision-Making," jointly sponsored by the National Academy of Sciences of the United States and the Soviet Academy of Sciences, held at Tbilisi in the Soviet Republic of Georgia.

ress was made toward the development of data-analytic methods for the construction of such spatial representations on the basis of psychological data until the efforts of a group of psychometricians, beginning in the late 1930's at Chicago and subsequently moving to Princeton, culminated in the 1952 development by Torgerson of the first fully workable method of metric multidimensional scaling (*7, 8*).

This method is called "metric" because it requires psychological estimates of metric distances between the stimuli. Either one had to assume that the data (for example, subjective ratings of dissimilarity) increased linearly with such distances (*9*), or one had to use some preliminary (for example, "Thurstonian") scaling procedure to convert the data into numbers that could then be assumed to increase linearly with distance (*8*). Even after such numbers had been obtained, the computation required several more stages in which one successively (i) estimated the "additive constant" and thus obtained a matrix of estimated distances between the points; and then, on the basis of a theorem of Young and Householder (*10*), (ii) computed a matrix of scalar products between the points (interpreted as vectors issuing from their common centroid), and (iii) factored this matrix into its eigenvalues and vectors to obtain explicit coordinates for the stimuli in a Euclidean space of a number of dimensions corresponding to the number of large eigenvalues.

Meanwhile, I had been approaching the problem of analyzing such measures of similarity by estimating the nonlinear form of the monotonic function, f_{mon},

$$s_{ij} = f_{mon}(d_{ij}) \qquad (1)$$

where s_{ij} is the obtained measure of similarity between stimuli i and j and where d_{ij} is a distance between corresponding points i and j that satisfies quite general metric conditions. The conditions that I proposed were (i) the distance axioms of

positivity, symmetry, and the triangle inequality, required for any metric space (11),

$$d_{ij} > d_{ii} = 0 \quad \text{(for } i \neq j) \qquad (2a)$$
$$d_{ij} = d_{ji} \qquad (2b)$$
$$d_{ij} \leq d_{ih} + d_{hj} \qquad (2c)$$

and (ii) the limiting, additivity condition of the triangle inequality,

$$d_{ij} = d_{ih} + d_{hj} \qquad (3)$$

for those triples in which there is reason to believe that stimulus h falls between i and j on a shortest connecting path (12, 13). In this way I established that the relation of confusion frequency to distance generally approximates a negative exponential function (a function that I had also derived from a "diffusion" theory of the internal process). I was then able to use the inverse, logarithmic function to convert confusion frequencies into distance-like numbers and, by means of metric multidimensional scaling, to obtain a Euclidean spatial representation (13). The photograph on the cover shows a 1955 solution that I obtained in this way for nine colors (14).

On the basis of still weaker assumptions, Coombs and Hays [see (15)] devised procedures that required only the rank order of the similarity data. But their method correspondingly yielded only the "nonmetric" rank orders of the projections of the points on orthogonal axes of the space.

Various limitations of these early methods of multidimensional scaling discouraged their wide adoption. In addition to its strong assumption of linearity, Torgerson's original metric method was not readily extendable to other cases in which, for example, the underlying metric is non-Euclidean, or the matrix of data is incomplete. My own early method of dealing with nonlinearities was cumbersome and did not handle error variability in an optimum way. And the method of Coombs and Hays was impractical except for small matrices of data, and its nonmetric solutions failed to preserve what I subsequently found to be the essentially metric constraints of the ordinal data.

Modern "Nonmetric" Approach

At the Bell Telephone Laboratories, I began in 1960 to explore a new approach
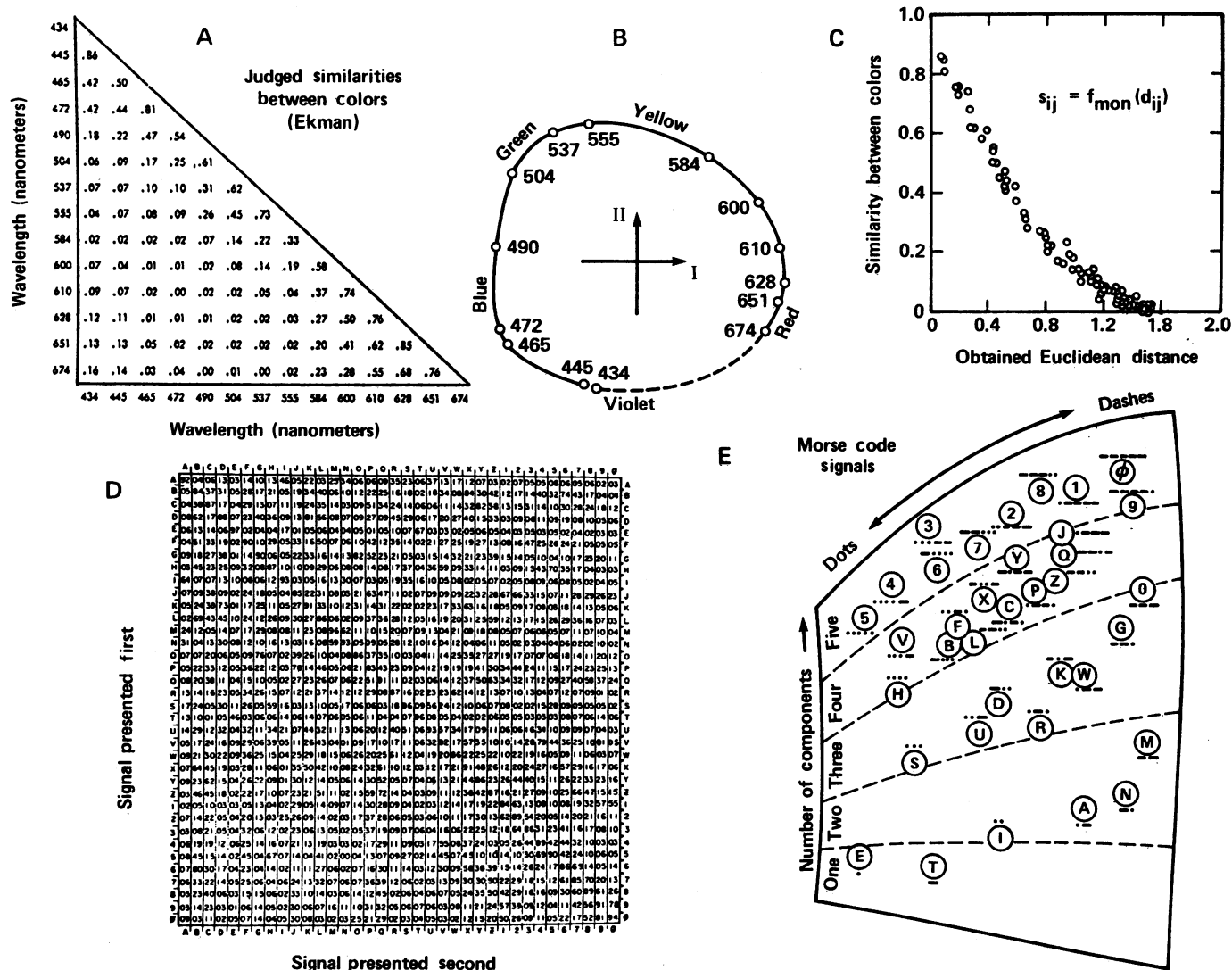

Fig. 1. Illustrations of nonmetric multidimensional scaling. (A) Judged similarities between 14 spectral colors. [From Ekman (22)] (B) Two-dimensional configuration obtained by my analysis of Ekman's data. (C) Obtained relation, for all pairs of the 14 colors, between judged similarities and corresponding Euclidean distances between points in my obtained configuration. [From Shepard (16)] (D) Percentage of "same" responses for all pairs of successively presented aural signals of the international Morse code. Entries in the principal diagonal correspond to correct responses. [From Rothkopf (23, 24)] (E) Two-dimensional configuration obtained by analysis of the Morse code data. The actual dot-and-dash patterns are indicated beside the points. [From Shepard (24)]

to multidimensional scaling, called "analysis of proximities," that proved capable of overcoming the limitations of the earlier approaches. I used a one-stage iterative method (i) to adjust the positions of points in a space until the rank order of the interpoint distances was as nearly as possible the inverse of the rank order of the corresponding similarities, and (ii) to find the space of the smallest number of dimensions for which the residual departure from a perfect inverse ranking was acceptably small (16).

Following a few adjustments, on 17 March 1961 the iterative process with which I had been experimenting finally converged to its first stationary configuration (at just 2:33 p.m. EST, according to the computer log). From then on, results of surprising precision were regularly obtained. Provided that the number of points was not too small relative to the number of dimensions, the merely qualitative, ordinal relations in the similarity data generally turned out to be sufficient to determine the quantitative, metric structure of the spatial representation. In two dimensions, a test configuration of as many as 15 random points could be essentially reconstructed on the basis merely of the rank order of the interpoint distances (16); and with as many as 45

random points, I later found that product-moment correlations between true and recovered distances averaged over 0.9999997 (17).

Such nonmetric multidimensional scaling soon reached essentially its present state of development when my associate J. B. Kruskal employed standard gradient methods to minimize an *explicitly defined* sum-of-squares measure of departure from the monotonic relation that I had posited between similarity and distance; namely, the "stress" measure (S) given by

$$ S = \left\{ \frac{\sum_{i,j} (d_{ij} - \hat{d}_{ij})^2}{\sum_{i,j} d_{ij}^2} \right\}^{1/2} \tag{4} $$

Here, the d_{ij} are the distances between the points at any particular iteration given, in terms of the $N \times K$ coordinates x_{ik} of the N points in the K-dimensional Euclidean space, by the usual distance formula

$$ d_{ij} = \left\{ \sum_{k=1}^{K} (x_{ik} - x_{jk})^2 \right\}^{1/2} \tag{5} $$

and the $\hat{d}_{ij}$ are numbers that (i) are monotonic with the similarity data $s_{i,j}$ and (ii) minimize stress relative to the spa-

tial distances d_{ij} at each iteration (18).

In some trial number of dimensions, a starting configuration of points is first constructed, either at random or by a metric method (19). On each ensuing iteration, then, (i) the best-fitting monotonic sequence $\hat{d}_{ij}$ is determined anew by an algorithm for least-squares monotone regression; (ii) the $N \times K$ partial derivatives of stress with respect to the coordinates, x_{ik}, are evaluated; and (iii) the coordinates are adjusted in the direction of the negative gradient or steepest descent by $x'_{ik} = x_{ik} - \alpha \cdot \partial S / \partial x_{ik}$, where α includes an adaptively modified step-size factor (18). The process is terminated when the components of the gradient have become small enough to indicate a close approach to a stationary configuration. (In order to exclude entrapment in a merely local minimum, different starting configurations should be tried.) The entire process can be repeated in spaces of higher or lower dimensionality, with the final solution chosen to achieve the best balance between parsimony, goodness of fit and, especially, substantive interpretability (20, 21).

Applications to Perception

Figure 1A displays the first significant set of empirical data to which I applied my original program for "analysis of proximities" (16); namely, Ekman's data on the perceived similarities between 14 spectral colors (22). The stationary two-dimensional configuration to which the iterative process converged is shown in Fig. 1B, together with a smooth curve subsequently drawn through the 14 points. Figure 1C shows the nonlinear relation between Ekman's similarity data and Euclidean distances in the obtained configuration. In addition to its good fit to the data, the two-dimensional solution is both more similar to Newton's (3) color circle and more parsimonious than the five-dimensional representation that Ekman (22) himself obtained by applying factor analysis to the data.

Figure 1D displays the frequencies with which unskilled listeners in an experiment by Rothkopf (23) judged successive aural signals of the international Morse code to be the same. Figure 1E shows the two-dimensional solution that I obtained by applying Kruskal's improved program, MDSCAL, to these data, after averaging each entry s_{ij} with its symmetrical counterpart s_{ji} (24). The two-dimensional solution seems more informative than the original 36×36 matrix of data. As is indicated by the added lines, perceptions of these aural signals
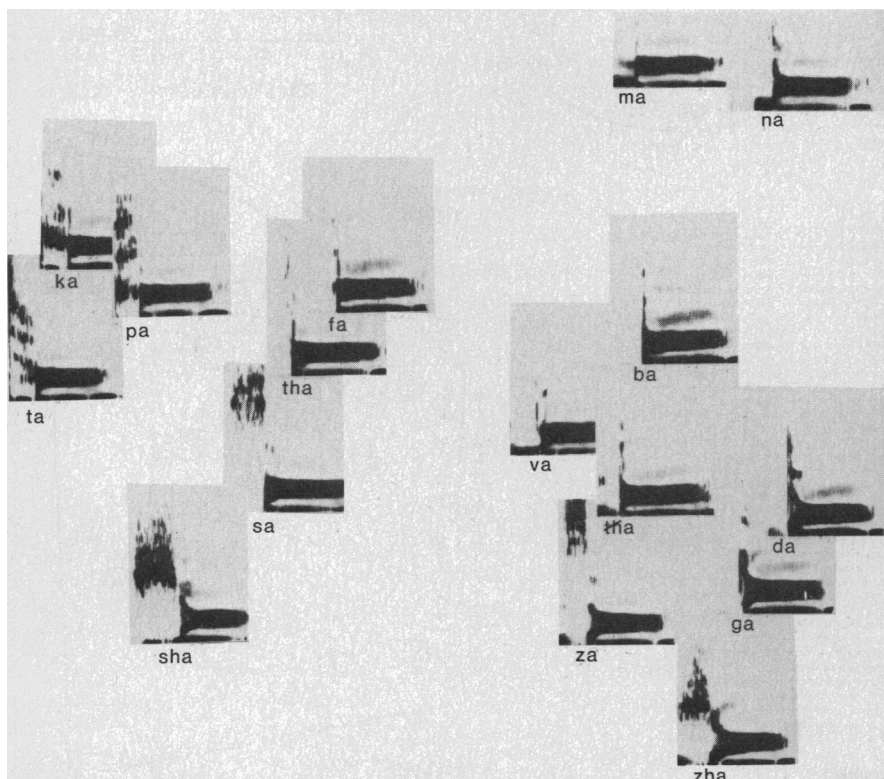

Fig. 2. Sound spectrograms for 16 syllables differing in the initial consonant, centered on the corresponding points of the two-dimensional solution obtained by applying Shepard and Chang's scaling program to the average confusion matrix for all of Miller and Nicely's conditions that imposed no differential filtering of high or low audio frequencies. In the spectrograms [rearranged from Carroll and Wish (30)] the distribution of audio energy is shown as a function of frequency (from 0 to 6 kHz) on the vertical axes, and time, on the horizontal axes.

differed primarily with respect to the number of components (dots or dashes) in each signal, and the relative preponderance of dots versus dashes among those components. Information about the sequential structure within each signal was largely lost on these unskilled listeners.

Figure 2 presents the solution that I obtained from an analysis of data collected by Miller and Nicely (25) on errors of identification of 16 consonant phonemes (all followed by the vowel /a/, "ah") in the presence of noise (26, 27). Centered over each of the 16 obtained points is a representative sound spectrogram for that syllable and, just below, its characteristic English spelling. The positions of the points are shown more precisely in Fig. 3A, by the phonetic symbols for the consonants. The added interpretive lines indicate how the speech sounds are organized on the basis of such phonetic features as voicing, nasality, affrication, and place of articulation. As can be seen in Fig. 3B, the resulting fit was very close, accounting for about 98 percent of the variance of the data. As before (13, 20, 24, 28), I found the confusion data to be well approximated by a negative exponential (the fitted curve).

Analysis of Multiple Matrices

We could also analyze separately the matrix of data for each subject, amount of training, or condition of stimulus presentation. But a more powerful analysis is achievable by simultaneously taking account of the entire set of matrices in each case. A metric method of individual difference scaling, INDSCAL, originated by another of my associates, J. D. Carroll, has proved particularly effective for this purpose (29). Carroll assumed that the same spatial configuration is appropriate for each individual subject or condition, but that the individuals differ with respect to the effective weights of the different dimensions. Accordingly, Carroll replaced the usual Euclidean distance formula (Eq. 5) by

$$d_{ij}{}^{(m)} = \left\{ \sum_{k=1}^{K} w_{mk}(x_{ik} - x_{jk})^2 \right\}^{1/2} \tag{6}$$

where w_{mk} is the weight of dimension k for individual m, and $d_{ij}{}^{(m)}$ is the resulting effective distance between stimuli i and j for that individual. Carroll and Chang adapted nonlinear iterative least squares to the canonical decomposition of the three-way $N \times N \times M$ matrix of stimuli by stimuli by individuals to obtain a metric configuration of the N stimuli in an orthogonal coordinate space and

also the unknown weights of the orthogonal dimensions of this space for each of the M individual subjects or conditions (29, 30). The larger, three-way matrix of data can support the extraction of a larger number of dimensions than is usually possible in the analysis of two-way matrices. Moreover, because the results are not, as in the Euclidean two-way case, arbitrary with respect to rotation, the unrotated axes of the solution should be immediately interpretable.

Carroll and Chang (31) applied IND-SCAL to the judged similarities of colors collected by Helm (32) from subjects with various degrees of color blindness. Over 90 percent of the total variance was accounted for by a two-dimensional solution in which the ten spectral hues formed a circular configuration very much like the one (Fig. 1B) that I had previously obtained. Now, however, the two orthogonal axes immediately corresponded to a red-green and to a blue-yellow dimension, in agreement with currently accepted "opponent-process" theories of color vision (33). Moreover, subjects were found to differ primarily
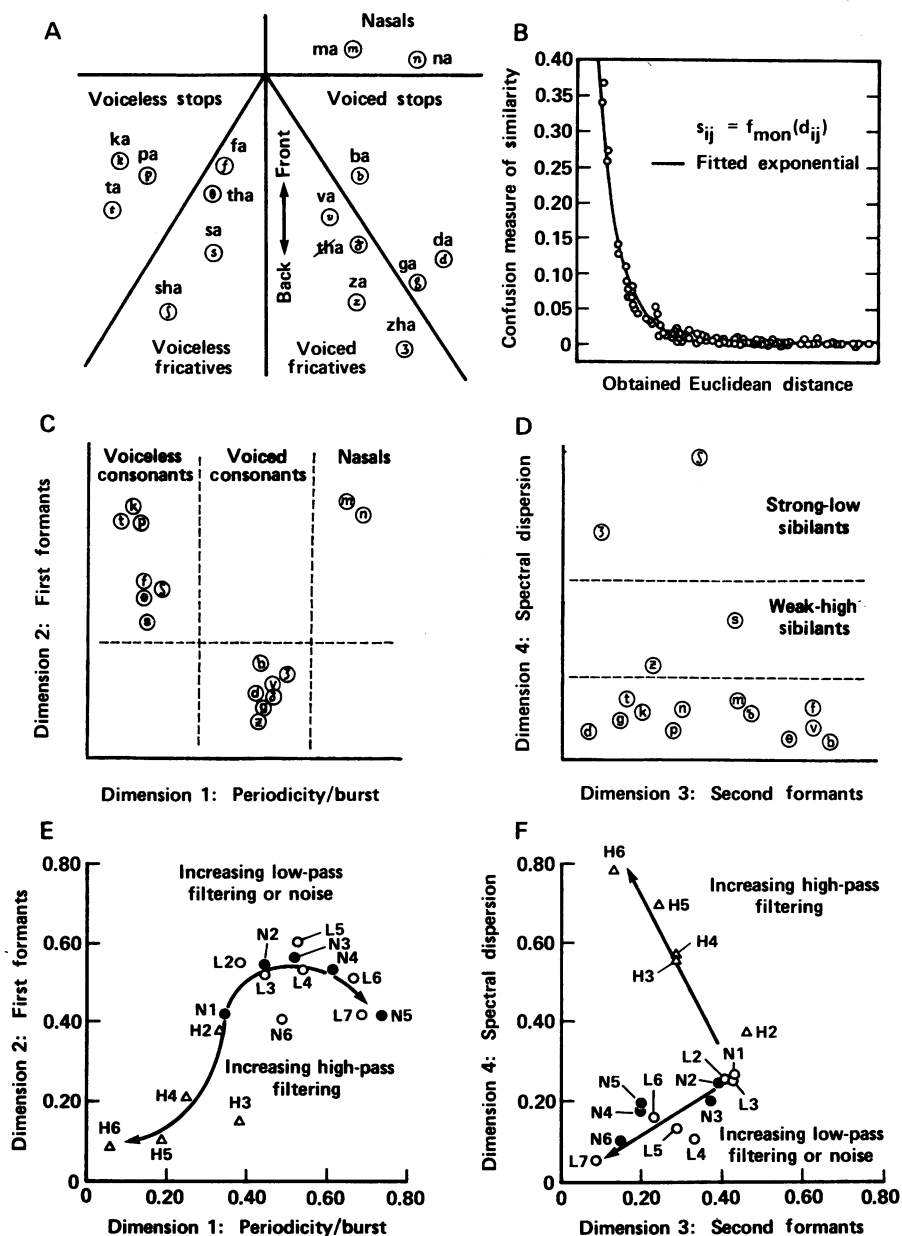

Fig. 3. Various multidimensional scalings of 16 English consonants based on the confusion data of Miller and Nicely (25). (A) the two-dimensional configuration of Fig. 2 with interpretive lines added. (B) Obtained (exponential) relation, for all pairs of the 16 consonants, between the confusion data and corresponding Euclidean distances in the obtained configuration in (A). [From Shepard (26)] (C and D) Four-dimensional INDSCAL solution obtained by a simultaneous analysis of the individual matrices for all 17 of Miller and Nicely's conditions, projected onto the planes of dimensions 1 and 2 and dimensions 3 and 4. (E and F) The estimated weights of the various conditions of filtering and noise, projected onto these same two planes. [From Soli and Arabie (35)]

with respect to their weights for the red-green dimension, in agreement with the relatively greater prevalence of deficiency in the red-green system.

Taking advantage of my demonstration that the generally exponential relation of confusion frequency to distance holds, in particular, for Miller and Nicely's data on confusions between consonants (Fig. 3B), Arabie and Soli (34, 35) logarithmically transformed Miller and Nicely's data into distance-like numbers. They then applied Carroll and Chang's INDSCAL to the entire set of 17 resulting matrices—one matrix for each of the 17 different conditions of stimulus presentation; namely, six (N1 to N6) for increasingly added masking noise, six (L2 to L7) for increasingly restrictive low-pass filtering, and five (H2 to H6) for increasingly restrictive high-pass filtering. Panels C and D of Fig. 3 display the resulting four-dimensional solution, which accounts for 69 percent of the variance in the total set of 17 transformed matrices (as opposed to 61 percent without the prior log transformation).

Dimensions 1 and 2 are determined by low-frequency energy associated with voicing and nasality (compare Fig. 3A) or, in terms of acoustic events identifiable in the sound spectrograms (35), associated with the temporal relations between initial noise burst and onset of periodic pulsing (dimension 1) and with transition in the lowest resonance or "first formant" (dimension 2). Dimensions 3 and 4, by contrast, are determined by higher frequency energy associated with transition in the middle resonance or second formant (dimension 3) and extent of high-frequency "sh" noise characteristic of the sibilants /ʃ,ʒ/ and to a lesser extent /s,z/ (dimension 4). The weights of the 17 conditions, plotted in the planes of the same pairs of dimensions (Fig. 3, E and F) corroborate this interpretation: The low-pass conditions are more heavily weighted on the first two dimensions, while the high-pass conditions are more heavily weighted on the last dimension (34).

Non-Euclidean Representations

The Shepard-Kruskal approach to multidimensional scaling (16, 18) made feasible, for the first time, the search for solutions in non-Euclidean spaces. In his improved program MDSCAL, for example, Kruskal (18) replaced the Euclidean distance formula with the more general Minkowski r-metric formula

$$d_{ij} = \left\{ \sum_{k=1}^{K} |x_{ik} - x_{jk}|^r \right\}^{1/r} \tag{7}$$

The family of r-metrics had been of interest to students of perception because of indications that whereas perceived similarities conform to the locally Euclidean metric (r = 2) for perceptually "unitary" stimuli such as homogeneous colors, perceived similarities tend to conform to something closer to the "city-block" metric (r = 1) for "analyzable" stimuli such as geometrical shapes differing in perceptually distinct dimensions of size, orientation, and brightness (8, 36). For Ekman's color data, Kruskal in fact obtained lowest values of stress when r was close to 2 (18), while for more analyzable stimuli better fits tend to be obtained with r close to 1 (37). (An example of an r = 1 solution for the semantic domain will be given later in this article.)

Spaces that are non-Euclidean in the sense of being globally curved though locally Euclidean also have been proposed—particularly for perceived colors (4), positions of luminous points (38), and orientations in three-dimensional space (39). One step in the extension of multidimensional scaling to such situations has been based on an assumption of constant curvature (40). Another approach has sought to simplify the problem of interpreting globally curved structures by mapping them down into a flat Euclidean space of the same "intrinsic dimensionality" (20, 39, 41). Figure 4 shows results that Carroll, Chang, and I obtained in tests with artificial similarity data derived from distances between points on the surface of a sphere. We obtained the flat two-dimensional solutions
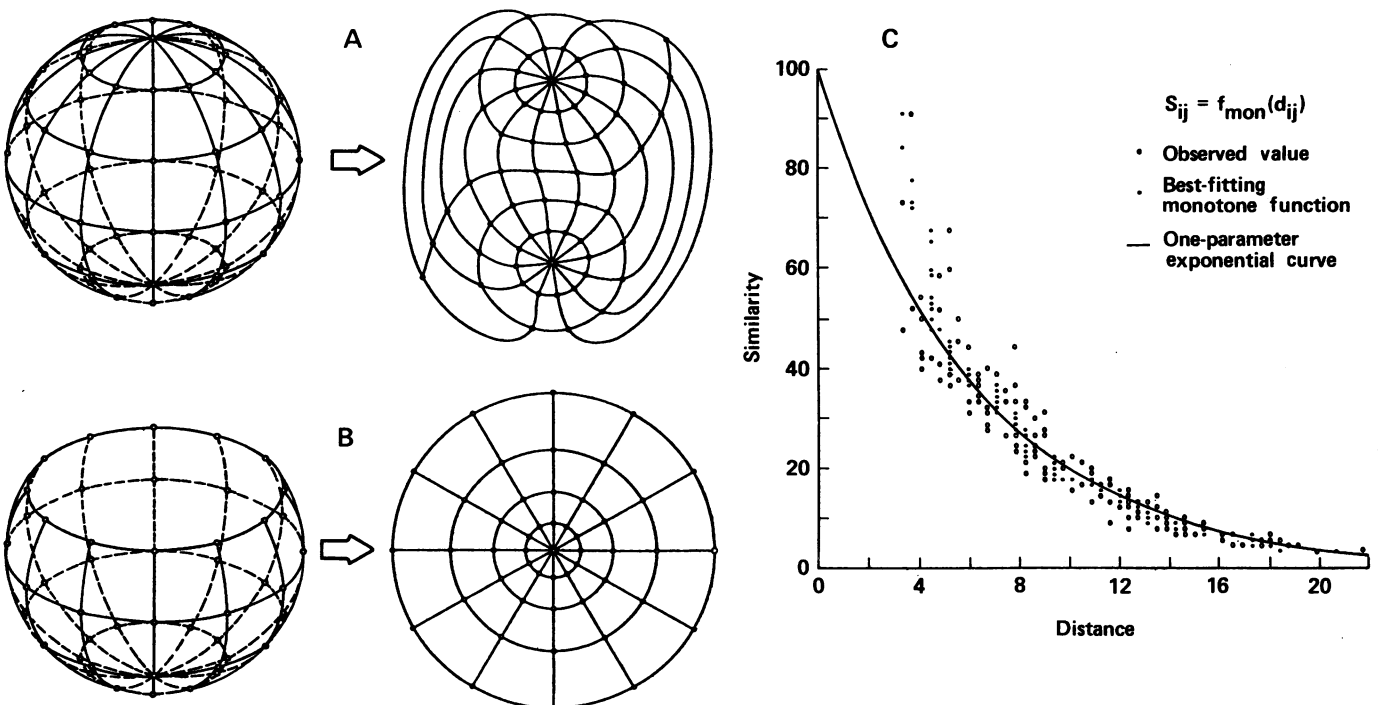

Fig. 4. Non-Euclidean analyses. (A and B) Parametric mapping of a curved configuration down into a flat space of the same intrinsic dimensionality by optimizing an index of "continuity" (A) and by "conformal reduction" (B). [From Shepard (20, 39), and Shepard and Carroll (41)] (C) Obtained relation for pairs of Morse code signals, between Rothkopf's similarity data (Fig. 1D) and corresponding distances in a general, nondimensional metric space. [From Cunningham and Shepard (43)]

by iterative procedures that (in Fig. 4A) minimized Carroll's index of departure from smoothness or continuity of the mapping, and (in Fig. 4B) minimized the Shepard-Chang measure of departure from local monotonicity. Both procedures yield a parametric representation in an appropriately reduced space, and the second tends to be conformal.

Nondimensional Scaling

In some applications our primary interest is in determining the functional form of the relation between similarity and distance; for example, whether that form is exponential for confusion data (13) and hyperbolic for discrimination times (42). Accordingly, Cunningham and I developed a method of nondimensional scaling, which used a gradient-projection method to find distances, d_{ij}, in a completely general, coordinate-free metric space, that are as nearly as possible monotonically related to the given similarity data, s_{ij}. In order to obtain solutions that achieved an acceptable fit but that were, at the same time, as far as possible from the trivial and degenerate one with all points equally distant from each other, which in multidimensional scaling is ruled out by minimizing the number of dimensions, we had to maximize the variance of the distances. The only other condition on the distances (a condition that we imposed by a "penalty function") was that they satisfy the three distance axioms (Eqs. 2a, 2b, and 2c) [see (43)].

We found that we could accurately recover the nonlinear shape of the function used to generate artificial similarity data even though the program was provided no information about either the type of metric or the form of the function. By contrast, an MDSCAL solution failed to recover the form of the function when the data were derived from very non-Euclidean sum-over-path distances in a tree. Figure 4C shows the results obtained when we applied this program to the Morse code data of Fig. 1D. Without assuming anything about the nature of the underlying metric, we recovered a function essentially like the one I originally obtained on the assumption of a Euclidean metric (24). And, again, the relation is in rough agreement with a simple negative exponential function (the one-parameter fitted curve).

This method is nondimensional rather than multidimensional, because it does not furnish coordinates for a visualizable configuration of the stimuli. However, maximization of the variance of the dis-

Fig. 5. Additive tree obtained by analysis of Henley's data on the conceptual similarities between 30 species of animal, embedded in a two-dimensional space. [Rearranged from Sattath and Tversky (47)]
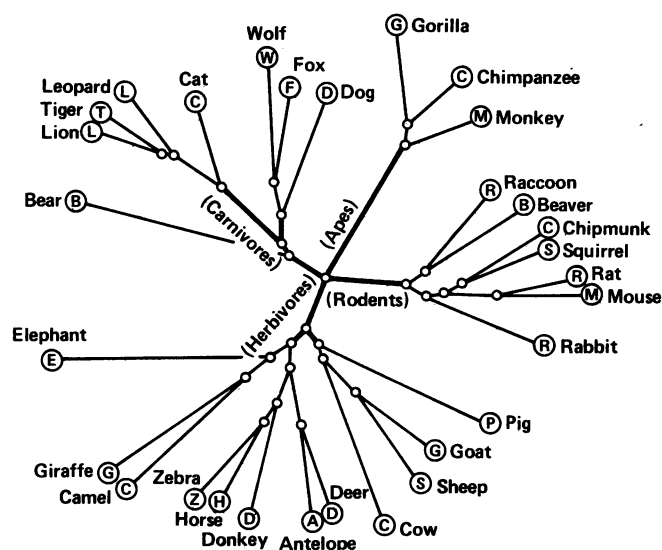
tances tends to drive the distances in every triangle toward the limiting additive case of the triangle inequality. (That is, Eq. 2c tends toward Eq. 3.) Hence, it is tempting to take the additional step of representing the distances as additive paths through a visualizable tree or graph (44).

Fitting Additive Trees

Procedures for fitting additives or path-length trees to similarity data were soon devised by Cunningham (45), Carroll and Chang (46), and Sattath and Tversky (47). An additive tree is a graph without any closed loops, in which the distance between any two nodes is given by the sum of the lengths of the links in the unique path between those nodes. The procedures for fitting such trees were based on the replacement of the triangle inequality by the stronger, four-point additivity condition

$$d_{hi} + d_{jk} \le \max \{(d_{hj} + d_{ik}), \\ (d_{hk} + d_{ij})\} \quad (8)$$

The resulting tree has N external nodes for the N stimuli, and the lengths of all connecting links are then estimated to minimize a least-squares measure of departure from good fit. Basically these methods are metric in that they treat the data as linearly related to underlying path-length distances. However, such distance-like data, if not initially available, could be obtained from similarity data by first applying Cunningham-Shepard maximum variance nondimensional scaling.

Figure 5 shows the additive tree that Sattath and Tversky (47) obtained by applying their program ADDTREE to Henley's (48) judged semantic dissimilarities

between 30 animal terms. I have arranged the branches in Fig. 5 so that the terminal nodes approximate the positions of the corresponding points in a multidimensional scaling solution for these same data. The major branches of the tree roughly correspond to apes, rodents, carnivores (both canines and felines), and large herbivores (hoofed animals and elephant). Perhaps trees or graphs are particularly well suited to the representation of semantic structures. Sattath and Tversky reported that the original dissimilarity data were more closely fitted by path-length distances in the tree (stress = .07, r = .91) than by distances in a two-dimensional Euclidean representation requiring about the same number of parameters (stress = .17, r = .86).

Hierarchical Clustering

The designation of one internal node as the "highest" node in an additive tree confers on every other node a derivative height by virtue of its distance from the designated node. The tree thereby becomes a hierarchical clustering. If all terminal nodes are equally distant from the highest node, the distance between terminal nodes is then given simply by the height of the highest node on the connecting path. Hence the hierarchical tree metric is a special case of the additive tree metric. In fact, as was independently shown by several workers in 1967 (49-51), the hierarchical tree metric is governed by the "ultrametric" inequality

$$d_{ij} \le \max (d_{ih}, d_{hj}) \quad (9)$$

which requires that all "triangles" be isosceles. Both nonmetric methods (49,

50) and least-squares metric methods (46, 51) have been devised for fitting such hierarchical tree structures to similarity data.

Figure 6A shows the hierarchical tree that I obtained by reanalyzing the average of Miller and Nicely's confusion matrices for unfiltered consonants, using the "diameter" (or "complete-link") variant of the nonmetric method of hierarchical clustering developed by my co-worker S. C. Johnson (26, 50). In Fig. 6B, I have embedded the clusters corresponding to cuts through this tree at representative levels as closed curves in the earlier spatial solution (Fig. 3A). The compatibility of the spatial and hierarchical representations is manifested in
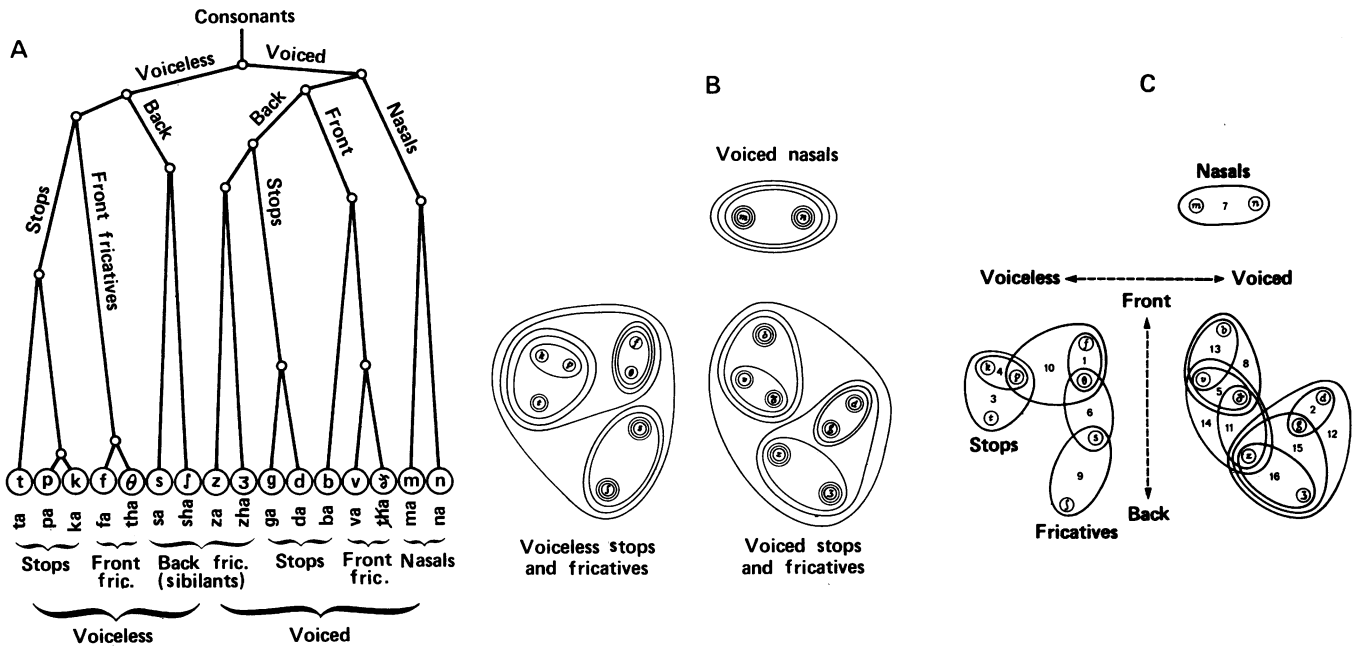

Fig. 6. Alternative clustering analyses of Miller and Nicely's (25) data on confusions of 16 consonants. (A) Hierarchical tree obtained by using Johnson's (50) nonmetric "diameter" (or "complete-link") method. (B) The same hierarchical clustering displayed as embedded in the two-dimensional scaling solution of Fig. 3A. [From Shepard (26)] (C) Nonhierarchical clustering obtained by ADCLUS analysis of the same data, embedded in the same two-dimensional scaling solution. The Arabic numerals indicate the ranks of the clusters by estimated weights. [From Shepard and Arabie (52)]
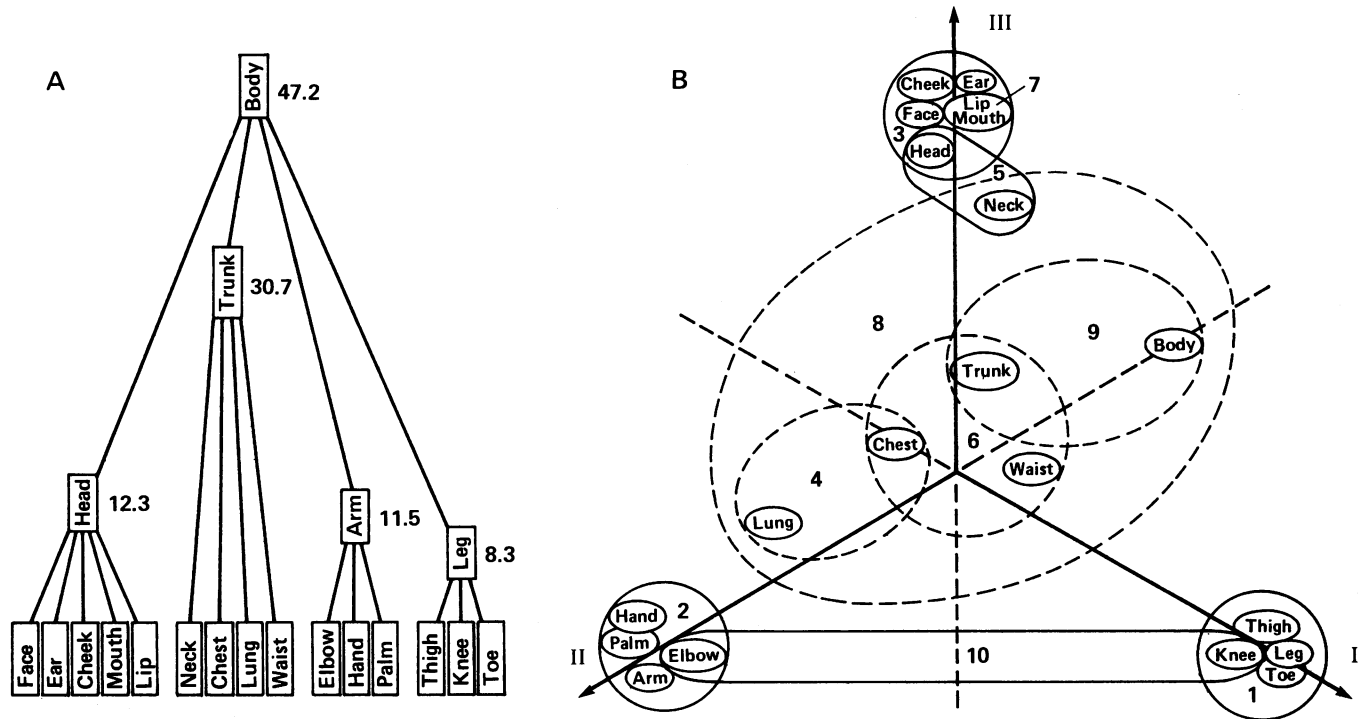

Fig. 7. Alternative clustering analyses of Miller's data on the semantic relationships between 20 names of parts of the body. (A) Hierarchical tree with stimuli assigned to internal as well as terminal nodes. The numbers attached to the internal nodes indicate their estimated heights in the hierarchy. [Rearranged from Carroll (46)] (B) Additive clustering obtained by ADCLUS analysis of the same data, embedded in a two-dimensional projection of the three-dimensional "city-block" solution (r = 1). The Arabic numerals indicate the ranks of the clusters by estimated weights. The "trunk" words, enclosed in dashed curves, fall in the back (-, -, -) orthant of the three-dimensional space. [From Shepard and Arabie (52)]

the compact and convex forms of the nested curves. However, each of the two types of representation brings out different aspects of the underlying structure. Only the continuous, spatial representation preserves the parallel orderings of the voiceless and the voiced fricatives /f,θ,s,ʃ/ and /v,ð,z,ʒ/ with respect to place of articulation and, hence, represents such facts as that the middle fricatives (for example, /θ/ and /s/) are more often confused than the extreme fricatives (for example, /f/ and /ʃ/). But only the discrete, clustering representation, separately obtained for each of Miller and Nicely's 17 conditions, reveals that whereas the front voiceless stop /p/ clusters with the other voiceless stops /k,t/, the corresponding front voiced stop /b/ uniformly groups with the front voiced fricatives /v,ð/ rather than with the other voiced stops /d,g/. Evidently, place of articulation was more salient than presence or absence of affrication for voiced consonants, while for the voiceless consonants absence of affrication became more salient owing to the correlated presence of an initial burst, which appears at the left in the spectrograms for /p,t,k/ in Fig. 2.

In semantic studies, superordinate words might reasonably be represented by internal nodes in the tree. Carroll and Chang developed a combinatorial optimization procedure for least-squares fitting of such a hierarchical model (46). Figure 7A snows the solution that they obtained from an analysis of semantic similarities that Miller had obtained for 20 names of body parts (46). This representation accounted for 92.7 percent of the variance with only five continuous parameters; namely, the height values for each of the five internal nodes of the tree. The semantic hierarchy seems to be well recovered, with "body" dominating "leg," "arm," "head," and "trunk," and with each of these dominating, in turn, the appropriate set of more subordinate terms.

Additive Clustering

Sometimes we would not want the psychologically significant subsets of the stimuli to be nested in a strictly hierarchical fashion. While the kin terms "father" and "mother" should be classified together in contrast to "son" and "daughter" on the basis of generation, the terms "mother" and "daughter" should be classified together in contrast to "father" and "son" on the basis of sex. But these two cross-cutting classifications cannot be simultaneously accom-

modated within any one hierarchical representation. To deal with such cases, I developed with Arabie a nonhierarchical method of additive clustering ADCLUS, which differs from all of the more or less spatial representations in that it is not based on any notion of distance (20, 52, 53).

The basic idea behind additive clustering is that the perceived similarity between any two stimuli is simply the sum of positive psychological weights, w_k, of the discrete properties that both stimuli have in common. Formally,

$$s_{ij} = \sum_{k=1}^{K} w_k p_{ik} p_{jk} \qquad (10a)$$

where

$$p_{ik} = \begin{cases} 1 & \text{if object } i \text{ has property } k, \\ 0 & \text{otherwise.} \end{cases} \qquad (10b)$$

If it were not for the restriction that the p_{ik} be binary valued, the model would be essentially identical to that of factor analysis, which Ekman (22) had assumed to be suitable for the representation of continuous structures in similarity data. However, the imposition of this binary restriction converts the problem of computing the eigenvalues and vectors of the similarity matrix into a more difficult combinatorial problem of finding the smallest set of weighted subsets that will provide a satisfactory additive fit to the data (52).

Figure 6C displays the subsets obtained when we applied ADCLUS to the Miller-Nicely data (25) on confusions between 16 consonants, embedded in the previous two-dimensional scaling solution (see Figs. 3A and 6B). With these 16 subsets and the entire subset of all 16 consonants, the additive model was able to account for 94.5 percent of the variance. The Arabic number in each subset indicates the rank of that subset according to its estimated weight (ranging from .730 to .009). The subsets in the additive clustering are not nested as they were in the hierarchical clustering (Fig. 6B). For example, in accordance with the places of articulation, the sequence of voiceless fricatives /f,θ,s,ʃ/ now forms a chain of overlapping clusters, as does the parallel sequence of voiced fricatives /v,ð,z,ʒ/. And, for both the voiced and voiceless consonants, the front fricatives group with the front stops as well as with the relatively back fricatives.

Finally, Fig. 7B shows the ADCLUS solution that Arabie and I obtained (52) from a reanalysis of Miller's data on the relations among 20 names of body parts. Here, the ten obtained clusters are embedded in a two-dimensional projection

of the three-dimensional city-block solution that Arabie obtained for these same data using MDSCAL. (For these data, the residual stress was .019 for $r = 1$, as opposed to .046 for the Euclidean $r = 2$.) An advantage to be expected for such a non-Euclidean solution is that the orthogonal axes I, II, and III correspond, without rotation, to the immediately interpretable "leg," "arm," and "head" clusters. As before, the Arabic number within each subset indicates its rank with respect to weight (ranging from .820 to .119). When the entire set is included (with a weight of .048), 95.6 percent of the variance is accounted for by these subsets. In comparison with the hierarchical representation obtained by Carroll and Chang (Fig. 7A), the ADCLUS representation fails to represent the superordinate status of some of the terms explicitly. It does, however, provide an explicit representation of the nonhierarchical overlap of certain of the subsets. Most significant among the overlapping subsets are subset 10, which connects the functionally analogous "elbow" and "knee" parts of the "arm" and "leg" subsets, and subset 5, which, in the words of the Negro spiritual, confirms that, indeed, "the head bone is connected to the neck bone."

Concluding Remark

It would be a mistake to ask which of these various scaling, tree-fitting, or clustering methods is based on *the* correct model. As even my small sample of illustrative applications indicates, different models may be more appropriate for different sets of stimuli or types of data. Even for the same set of data, moreover, different methods of analysis may be better suited to bringing out different, but equally informative aspects of the underlying structure (54).

References and Notes

1. R. N. Shepard, A. K. Romney, S. Nerlove, Eds., *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences* (Seminar Press, New York, 1972), especially my "taxonomy" of types of data and methods, vol. 1, p. 21.
2. Concerning the orienting reflex, see E. N. Sokolov, *Annu. Rev. Physiol.* 25, 545 (1963).
3. I. Newton, *Opticks* (Smith & Walford, London, 1704), book 1, part 2, prop. 6.
4. H. von Helmholtz, *Handbuch der Physiologischen Optik* (Voss, Hamburg, ed. 2, 1896); E. Schrödinger, *Ann. Physik* (IV) 63, 481 (1920). See G. Wyszecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulas* (Wiley, New York, 1967). For recent Russian work based on a theory that color space is of constant positive curvature, see E. N. Sokolov, Ch. A. Izmailov, T. V. Izmailova, M. M. Zimachiov, *Vestn. Mosk. Univ. Psikhol.* No. 1, 4 (1977); Ch. A. Izmailov and E. N. Sokolov, *ibid.* No. 2, 47 (1978).
5. M. W. Drobisch, *Abh. Math. Phys. Kl. Königl. Sachs. Ges. Wiss.* 4, 1 (1855). See also R. N.

Shepard, *J. Acoust. Soc. Am.* **36**, 2346 (1964).
6. H. Henning, *Der Geruch* (Barth, Leipzig, 1916); *Z. Psychol.* **74**, 203 (1916). See F. A. Geldard, *The Human Senses* (Wiley, New York, 1962), pp. 283 and 313.
7. W. S. Torgerson, *Psychometrika* **17**, 401 (1952).
8. _____, *Theory and Methods of Scaling* (Wiley, New York, 1958).
9. T. Indow, in *Handbook of Perception*, R. C. Carterette and M. P. Friedman, Eds. (Academic Press, New York, 1974), vol. 2, p. 493.
10. G. Young and A. S. Householder, *Psychometrika* **3**, 19 (1938). Closely related theorems had previously been established by K. Menger; see L. M. Blumenthal, *Univ. Mo. Stud.* **13** (No. 2) (1938), especially pp. 55 ff.
11. The distance axioms were first abstractly formalized by M. Fréchet in 1906; see L. M. Blumenthal in *(10)*, p. 23. K. Menger has said the *"triangle inequality* seems to me to form, indeed, the central point of a very large part of all mathematics;'' [see L. M. Blumenthal in *(10)*, p. 7].
12. R. N. Shepard, *Psychometrika* **22**, 325 (1957). Also see the related "segmental additivity" condition of R. Beals, D. H. Krantz, and A. Tversky [*Psychol. Rev.* **75**, 127 (1968)].
13. R. N. Shepard, *J. Exp. Psychol.* **55**, 509 (1958); *Psychol. Rev.* **65**, 242 (1958).
14. The colors reproduced in the cover photograph may differ somewhat from the nine color samples that, following Torgerson *(8)*, I obtained from the Munsell Company for use in my 1955 dissertation experiment. The Munsell specifications of the original colors are listed in the first of my 1958 published reports *(13)*.
15. C. H. Coombs, *A Theory of Data* (Wiley, New York, 1964).
16. R. N. Shepard, *Psychometrika* **27**, 125 (1962); *ibid.*, p. 219.
17. _____, *J. Math. Psychol.* **3**, 287 (1966). The paradoxical possibility of recovering quantitative structure from merely qualitative data is connected with the circumstance that the number of pairs of points and hence the number of ordinal constraints on their distances increases approximately as the square of the number of to-be-estimated quantitative coordinates of the points. Methods taking advantage of this circumstance are called "nonmetric" because they use only ordinal properties of the input. However, the output can achieve great metric precision and is always metric in the sense that it satisfies the distance axioms.
18. J. B. Kruskal, *Psychometrika* **29**, 1 (1964); *ibid.*, p. 28.
19. F. W. Young and W. S. Torgerson, *Behav. Sci.* **12**, 498 (1967).
20. R. N. Shepard, *Psychometrika* **39**, 373 (1974). This paper includes a discussion of the common errors (i) of reporting a solution with too many dimensions, which gives a good fit but an uninterpretable, probably unreliable, and sometimes "degenerate" solution; and (ii) of failing to ensure that entrapment in a merely local minimum has not occurred, particularly in the case of one-dimensional solutions.
21. A variety of computer programs of this general type have subsequently been developed. However, if precautions are taken to exclude merely local minima, the generally high degree of overdetermination of the final configuration by the ordinal constraints in the data ensures that all such methods are capable of yielding essentially the same results—including, indeed, metric methods, if nonlinearity is not too extreme [see *(20)*; D. G. Weeks and P. M. Bentler, *Psychol. Bull.* **86**, 349 (1979)]. See *(1, 19, 20, 54)* for citations of some of the programs developed by

Guttman, Lingoes, McGee, Roskam, Torgerson, Young, as well as by Kruskal and Shepard. Perhaps the currently most versatile such program is the one, available from the Bell Telephone Laboratories, named KYST (after Kruskal, Young, Shepard, and Torgerson, from whose earlier programs it derives). KYST provides, also, for a very general analysis of rectangular matrices in which, in accordance with the "ideal-points" model of Coombs *(15)*, rows correspond to subjects, columns correspond to stimuli, and each entry is some measure of the preference of that subject for that stimulus and is interpreted as the spatial proximity between that stimulus and that subject's "ideal" stimulus *(1)*. The same kind of analysis, when applied to a square matrix of similarity data, can account for systematic asymmetries in the data. Other recent developments include (i) alternative approaches to the problem of asymmetry [for example, C. L. Krumhansl, *Psychol. Rev.* **85**, 445 (1978)], (ii) methods for imposing prespecified constraints on the spatial solution [for example, P. M. Bentler and D. G. Weeks, *J. Math. Psychol.* **17**, 138 (1978)], and (iii) maximum likelihood metric methods in which one can test the statistical significances of different numbers of dimensions and can establish confidence regions around the points [J. O. Ramsay, *Psychometrika* **42**, 241 (1977); *ibid.*, **43**, 145 (1978)].
22. G. Ekman, *J. Psychol.* **38**, 467 (1954).
23. E. Z. Rothkopf, *J. Exp. Psychol.* **53**, 94 (1957).
24. R. N. Shepard, *Hum. Factors* **5**, 33 (1963).
25. G. A. Miller and P. E. Nicely, *J. Acoust. Soc. Am.* **27**, 338 (1955).
26. R. N. Shepard, in *Human Communication: A Unified View*, E. E. David and P. B. Denes, Eds. (McGraw-Hill, New York, 1972), p. 67.
27. In order to avoid problems of partial "degeneracy," which sometimes arise in a purely nonmetric analysis *(20)*, I obtained the two-dimensional spatial solution by means of a more recent program that Chang and I developed for multidimensional scaling on the (metric) assumption that the similarity measures decay exponentially with distance.
28. R. N. Shepard, in *Stimulus Generalization*, D. I. Mostofsky, Ed. (Standford Univ. Press, Stanford, Calif., 1965), p. 94.
29. J. D. Carroll and J.-J. Chang, *Psychometrika* **35**, 283 (1970).
30. Related methods for the analysis of multiple or "three-way" matrices have also been developed by Harshman, Tucker, and others. See J. D. Carroll and M. Wish, in *Contemporary Developments in Mathematical Psychology*, R. C. Atkinson, D. H. Krantz, R. D. Luce, P. Suppes, Eds. (Freeman, San Francisco, 1974), vol. 2, p. 57; Y. Takane, F. W. Young, J. de Leeuw, *Psychometrika* **42**, 7 (1977); L. Tucker, *ibid.* **37**, 3 (1972).
31. See M. Wish and J. D. Carroll, in *Handbook of Perception*, E. C. Carterette and M. P. Friedman, Eds. (Academic Press, New York, 1974), vol. 2, p. 449.
32. C. Helm, *J. Opt. Soc. Am.* **54**, 256 (1964).
33. L. M. Hurvich and D. Jameson, *Psychol. Rev.* **64**, 384 (1957).
34. P. Arabie and S. D. Soli, in *Multidimensional Analysis of Large Data Sets*, R. Golledge and J. N. Raynes, Eds. (Univ. of Minnesota Press, Minneapolis, 1980).
35. S. D. Soli and P. Arabie, *J. Acoust. Soc. Am.* **66**, 46 (1979). Also see the six-dimensional solution obtained by Wish and Carroll *(31)*.
36. F. Attneave, *Am. J. Psychol.* **63**, 516 (1950); D. V. Cross, in *Stimulus Generalization*, D. I. Mostofsky, Ed. (Stanford Univ. Press, Stanford, Calif., 1965), p. 72; R. Hyman and A. Well, *Per-*

cept. Psychophys. **2**, 233 (1967); R. N. Shepard, *J. Math. Psychol.* **1**, 54 (1964).
37. Elsewhere, I have proposed an alternative, safer strategy for determining the form of non-Euclidean metrics by estimating the shape of the underlying "unit circle" *(20)*. A different and quite promising approach, which seeks a coordinate representation without assuming any particular form of the metric, has recently been proposed by E. W. Holman, *J. Math. Psychol.* **18**, 39 (1978).
38. R. K. Luneburg. *Mathematical Analysis of Binocular Vision* (Princeton Univ. Press, Princeton, N.J., 1947). Also see *(9)*.
39. R. N. Shepard, in *Theory Construction and Data Analysis in the Behavioral Sciences*, S. Shye, Ed. (Jossey-Bass, San Francisco, 1978), p. 29.
40. H. Lindman and T. Caelli, *J. Math. Psychol.* **17**, 89 (1978).
41. R. N. Shepard and J. D. Carroll, in *Multivariate Analysis*, P. R. Krishnaiah, Ed. (Academic Press, New York, 1966), p. 561. Also see the closely related approach described by R. S. Bennett, *IEEE Trans. Inf. Theory*, **IT-15**, 517 (1969).
42. R. N. Shepard, D. W. Kilpatric, J. P. Cunningham, *Cognitive Psychol.* **7**, 82 (1975).
43. J. P. Cunningham and R. N. Shepard, *J. Math. Psychol.* **11**, 335 (1974).
44. Like the semantic networks discussed in J. R. Anderson and G. H. Bower, *Human Associative Memory* (Winston, Washington, D.C., 1973); D. A. Norman and E. E. Rumelhart, *Explorations in Cognition* (Freeman, San Francisco, 1975).
45. See J. P. Cunningham, *J. Math. Psychol.* **17**, 165 (1978).
46. See J. D. Carroll, *Psychometrika* **41**, 439 (1976).
47. S. Sattath and A. Tversky, *ibid.* **42**, 319 (1977).
48. N. M. Henley, *J. Verb. Learn. Verb. Behav.* **8**, 176 (1969).
49. C. J. Jardine, N. Jardine, R. Sobson, *Math. Biosci.* **1**, 173 (1967).
50. S. C. Johnson, *Psychometrika* **32**, 241 (1967).
51. J. A. Hartigan, *J. Am. Stat. Ass.* **62**, 1140 (1967).
52. R. N. Shepard and P. Arabie, *Psychol. Rev.* **86**, 87 (1979).
53. What may be an improved method for fitting the ADCLUS model is described in P. Arabie and J. D. Carroll, *Psychometrika* **45**, 211 (1980). Also an individual-difference extension of this additive clustering model to the analysis of three-way matrices, INDCLUS, was recently reported by J. D. Carroll and P. Arabie (Annual Meeting of the Psychometric Society, Monterey, Calif., 9 June 1979).
54. For a short but authoritative introduction to multidimensional scaling, see J. B. Kruskal and M. Wish, *Multidimensional Scaling* (Sage, Beverly Hills, Calif., 1978). And, for the most comprehensive and up-to-date survey of developments in this explosively growing field, see J. D. Carroll and P. Arabie, *Annu. Rev. Psychol.* **31**, 607 (1980), which cites 334 references, mostly published in the last 7 years.
55. I thank the Bell Telephone Laboratories for support of my work on multidimensional scaling from 1960 to 1968 and the National Science Foundation for support of my work on multidimensional and nondimensional representation since 1968. A number of my associates and students are responsible for many of the developments reviewed here, including J. Douglas Carroll, Jih-Jie Chang, Stephen C. Johnson, Joseph B. Kruskal, and Myron Wish, at the Bell Telephone Laboratories, and Phipps Arabie, James P. Cunningham, and Amos Tversky, at Stanford.