

MARLEY, A. A. J. (1981a). Multivariate stochastic processes compatible with "aspect" models of similarity and choice. *Psychometrika* 46, 421-428.

MARLEY, A. A. J. (1981b). Joint independent random utility models where one of the choice structures satisfies the strict utility model. *Journal of Mathematical Psychology*, 23, 257-272.

MARSCHAK, J. (1960). Binary choice constraints and random utility indicators. In K. J. Arrow, S. Karlin, and P. Suppes (Eds.), *Mathematical methods in the social sciences* (pp. 312-329). Stanford: Stanford University Press.

TVERSKY, A. (1972a). Elimination by aspects: A theory of choice. *Psychological Review*, 79, 281-299.

TVERSKY, A. (1972b). Choice by elimination. *Journal of Mathematical Psychology*, 9, 341-367.

RECEIVED: May 28, 1987

## A Multidimensional Stochastic Theory of Similarity

DANIEL M. ENNIS

Philip Morris Research Center, Richmond, Virginia

JOSEPH J. PALEN

Department of Mathematics, University of Michigan

AND

KENNETH MULLEN

Department of Mathematics and Statistics, University of Guelph

A multidimensional theory of similarity in which the mental representations of stimulus objects are assumed to be drawn from multivariate normal distributions is described. A distance-based similarity function is defined and the expected value of similarity is derived. This theory is the basis for a possible explanation of paradoxical results with highly similar stimuli regarding the form of the similarity function and the distance metric. A stochastic approach to multidimensional scaling based on same-different judgments is demonstrated using artificial and real data sets. The theory of similarity presented is used as a basis for a Thurstonian extension of Shepard's model of identification performance. © 1988 Academic Press, Inc.

## INTRODUCTION

The goal of this paper is to describe a multidimensional theory of similarity and to show how estimates of the model parameters assumed to be involved in making similarity judgments can be obtained. From the viewpoint of a mathematical model, mental representations of physical objects (or their analogous in lower organisms) can be treated as  $n$ -dimensional vectors with particular distributional properties and multidimensional parameters. Thurstone (1927) provided a framework for thinking about scaling relative psychological magnitudes by

We thank M. Waugh and E. Gee for providing computer and noncomputer resources, respectively. We thank W. Kuhn and C. Lilly who helped to facilitate our collaboration notably through the Philip Morris Visiting Scientist program. We thank R. Shepard, R. Nosofsky, G. Ashby, P. Arabie, J. Kapenga, and J. Frijters for discussions and comments that helped to improve our understanding of important issues related to this paper. Requests for reprints should be sent to Daniel M. Ennis, Philip Morris Research Center, Commerce Road, Richmond, VA 23261.

specifying the statistical parameters of an internal unidimensional continuum. Hefner (1958) extended Thurstone's ideas to the multivariate case in which the psychological magnitudes are represented as  $n$ -dimensional random vectors, where the values on each dimension have been drawn at random from independent normal distributions of equal variance. This means that the variances across dimensions are equal, but that the variances for different stimuli may not be. Techniques to obtain maximum likelihood estimates of the location and variability parameters of Hefner's model have been developed by Zinnes and MacKay (1983, 1987). Choice probabilities are not monotonically related to the distances between the means of the distributions under the assumptions of the Hefner model, when variances between stimulus points are unequal. MacKay (1987) has extended the model to cases in which the psychological variance on each dimension may be unequal for each stimulus.

Ashby and Perrin (1988) proposed a multidimensional version of signal detection theory in an attempt to find a common theoretical basis for similarity and recognition (identification). In this approach, the probability of confusing one stimulus object with another depends on the degree of overlap of the representational distributions. For a given momentary value there are particular probabilities that the variate was drawn at random from either of the two distributions of interest and the subject's identification decision will depend on the ratio of these two probabilities. This model does not involve a distance-based similarity function.

De Soete, Carroll, and DeSarbo (1986) described an unfolding model, the wandering ideal point (WIP) model, for paired comparisons data. Their model differs from Hefner's in that the values corresponding to the stimuli are fixed, only the ideal points have multivariate normal distributions, and the variance-covariance matrix of the ideal point distribution need not be an identity matrix. In the WIP model it is assumed that a subject will prefer one stimulus object over another whenever the momentary Euclidean distance between the preferred stimulus and the (wandering) ideal point is smaller than the equivalent distance for the non-preferred stimulus. This model appears to have much in common with a Thurstonian variant of Torgerson's method of triads (Ennis, Mullen, & Frijters, 1988). In Torgerson's method of triads, the subject's task is to decide which of two stimuli is most like a third preselected stimulus. This third stimulus could be replaced by the ideal point from the WIP model. The stimuli evoke psychological magnitudes which are assumed to be modelled as if they were drawn from independent normal distributions. In the Thurstonian variant of Torgerson's method of triads,  $P_{jk}$  represents the probability that stimulus  $S_i$  will be perceived to be more similar to  $S_j$  than  $S_k$ . If  $S_i$  is replaced by the subject's ideal point, then  $P_{jk}$  is the probability that  $S_j$  will be preferred to  $S_k$ . This preference model is more general than the WIP model because the momentary psychological magnitudes evoked by the stimuli are not fixed. However, Ennis, Mullen, and Frijters (1988) only presented the unidimensional model for Torgerson's method of triads.

An attempt to find a multidimensional extension of Torgerson's method of triads might usefully begin with a multidimensional model for the duo-trio method (Ennis

& Mullen, 1986b; Mullen and Ennis, 1987; Mullen, Ennis, de Doncker, & Kapenga, 1988), which is itself a special case of the Thurstonian variant of Torgerson's method of triads. The duo-trio method involves three stimuli, two of which are physicochemically identical. The subject's task is to decide which of two (possibly) different stimuli is most like a third preselected stimulus. It is assumed that the momentary psychological magnitudes corresponding to the three stimuli have been drawn from multivariate normal distributions (two independently drawn from one distribution, the third from a possibly different distribution). Another tri-stimulus grouping technique, the triangular method (in which the subject's task is to select the most different stimulus), has also been modelled under distributional assumptions similar to the duo-trio method (Ennis & Mullen, 1986b; Mullen & Ennis, 1987; Kapenga, de Doncker, Mullen, & Ennis, 1987).

In this paper, we extend the mathematical models developed for grouping techniques to same-different judgments and identification performance. This is accomplished by defining an explicit distance-based similarity function from which the expected value of similarity for confusable stimuli can be computed. We then show how the multivariate psychological parameters corresponding to a selection of hypothetical and real objects can be obtained.

## A MULTIDIMENSIONAL THEORY OF SIMILARITY

### Assumptions

Consider the case of a single pair of stimulus objects,  $S_x$  and  $S_y$ , which give rise to momentary psychological values of the respective magnitudes  $x$  and  $y$  where  $x' = (x_1, x_2, \dots, x_n)$ ,  $y' = (y_1, y_2, \dots, y_n)$ ;  $x'$  indicates an  $n$ -dimensional row vector and  $n$  is the number of psychological dimensions. The momentary psychological values are mutually independently distributed with  $x$  having density function  $h(x)$  and  $y$  having density function  $h(y)$ . The probability densities  $h(x)$  and  $h(y)$  are multivariate normal distributions with means  $\mu_x$  and  $\mu_y$  and variance-covariance matrices  $V_x$  and  $V_y$ . On the basis of the momentary psychological values,  $x$  and  $y$ , the subject decides whether the stimuli are the same or different. Let  $z = x - y$ .

Let  $d$  represent the momentary distance between  $x$  and  $y$  perceived by the subject, where

$$d = \left[ \sum_{k=1}^n |z_k|^\gamma \right]^{1/\gamma}, \quad \gamma \geq 1.$$

The distance between population means is

$$\delta = \left[ \sum_{k=1}^n |\mu_{xk} - \mu_{yk}|^\beta \right]^{1/\beta}, \quad \beta \geq 1.$$

Let the similarity of two particular momentary psychological values be  $g(d)$ . The form of  $g$  specifies the similarity function, or the function relating similarity to distance. If the subject invokes a step function,  $g(d)$  will be 0 or 1 depending on the value of  $d$  relative to some threshold value. If the subject invokes a continuous function, then  $g(d)$  will be a value that may be different from 0 or 1. If  $g$  is continuous,  $g(d)$  should decrease as  $d$  increases. Continuous and step functions will be considered for  $g$ .

#### The Continuous Function

There are many different functional forms which could be proposed for the function,  $g$ . Shepard (1987) argued in favor of an exponential decay similarity function. A flexible function which includes the exponential decay function is

$$g(d) = \exp(-d^\alpha), \quad \alpha \geq 0.$$

In order to satisfy the earlier stated requirement that  $g(d)$  should decrease as  $d$  increases,  $\alpha$  must be  $\geq 0$  or  $g(d)$  would become larger as  $d$  became larger. The particular value for  $\alpha$  may be different for different subjects and experimental conditions, although it is conceivable that  $\alpha$  may be a constant.

$V$  is the variance-covariance matrix of the difference between psychological values,  $z$ .  $\mu$  is a vector of differences between the means of the momentary psychological values,  $\mu_x$  and  $\mu_y$ .

The probability of declaring two randomly sampled psychological values from  $h(x)$  and  $h(y)$  to be the "same" is the expected value of  $g$  (in the absence of response bias), or

$$f(\mu, V, \alpha, \gamma) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\exp\{-0.5(z-\mu)' V^{-1}(z-\mu)\}}{(2\pi)^{n/2} |V|^{1/2}} \times \exp(-d^\alpha) dz_1 dz_2 \cdots dz_n, \quad (1)$$

where  $f(\mu, V, \alpha, \gamma)$  represents the expected value of the similarity of the two objects.

Equation (1) can be evaluated numerically for any  $\alpha$  and  $\gamma$  (which defines the metric of  $d$ ), but can be simplified significantly for the case when  $\alpha = 2$  and  $\gamma = 2$ . For this case,

$$f(\mu, V) = (|V| |J|)^{-1/2} \exp[\mu'(2J^{-1} - I)\mu], \quad (2)$$

where

$$J = V^{-1} + 2I$$

and  $I$  is the identity matrix.

A proof of Eq. (2) is given in the Appendix.

#### The Step Function

Conceptually, the step function can be handled in a way similar to that of the continuous function, except that  $g(d)$  is either 0 or 1 depending on the value of  $d$  relative to a threshold value,  $\tau$ .

If  $g(d) = 0.5\{\text{sgn}(\tau - d) + 1\}$ , where  $\text{sgn}$  is the signum function, then  $g(d)$  will be 0 when  $d > \tau$  (stimuli are different) and 1 when  $d \leq \tau$  (stimuli are the same).

The formula for calculating  $f(\mu, V, \tau)$  is

$$f(\mu, V, \tau) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\exp\{-0.5(z-\mu)' V^{-1}(z-\mu)\}}{(2\pi)^{n/2} |V|^{1/2}} \times 0.5\{\text{sgn}(\tau - d) + 1\} dz_1 dz_2 \cdots dz_n. \quad (3)$$

$\tau$  may be a fixed value or may be drawn from a particular probability density function and vary from trial to trial. In the examples given later, however, we consider  $\tau$  to be fixed.

#### Identification and Categorization Models

Identification and categorization performance models, such as those discussed by Nosofsky (1986), based on Shepard's (1957) work, could be extended to deal with stimuli whose psychological magnitudes may vary from trial to trial by formulating the models in terms of expected values. For instance, in the case of identification performance,

$$P(R_j | S_i) = E \left[ \frac{b_j g(d_{ij})}{\sum_{k=1}^m b_k g(d_{ik})} \right],$$

where  $P(R_j | S_i)$  is the probability that stimulus  $S_i$  leads to response  $R_j$ ;  $b_j$  and  $b_k$  are response bias parameters,  $0 \leq b_j \leq 1$ ;  $m$  is the number of stimuli; and  $g(d_{ij})$  is the similarity function evaluated at  $d_{ij}$ . According to this formulation of identification decisions, the subject obtains a distance-based similarity value on each trial for the stimulus in question ( $S_i$ ) and each of the memory representations of the  $m$  stimuli. The terms in the denominator may not be independent if, for instance, the subject uses the same momentary psychological magnitude corresponding to  $S_i$  in determining each of the  $d_{ik}$  ( $k = 1, \dots, m$ ). On the other hand, before obtaining similarity values [ $g(d_{ik})$ ] for  $S_i$  and each of the  $m$  memory representations, the subject may obtain different psychological magnitudes corresponding to  $S_i$  (i.e., resampling the stimulus distribution before referring to each memory representation). The model given by Nosofsky (1986) for categorization can be similarly formulated. These stochastic extensions of identification and categorization models will require more study and elaboration and will not be pursued further here.

## EVALUATION OF CONTINUOUS AND STEP SIMILARITY FUNCTIONS

## Computing

Equations (1), (2), and (3) were evaluated on a Gould 32/97 computer. Equations (1) and (3) were handled numerically using an adaptive routine by Genz and Malik (1980) in the bivariate case. These results agreed to third decimal place accuracy with Monte Carlo simulations of 100,000 trials per estimate.

When  $n$ , the number of dimensions, is equal to 2,

$$V = \begin{bmatrix} \sigma_1^2 + \sigma_3^2 & \rho_1 \sigma_1 \sigma_2 + \rho_2 \sigma_3 \sigma_4 \\ \rho_1 \sigma_1 \sigma_2 + \rho_2 \sigma_3 \sigma_4 & \sigma_2^2 + \sigma_4^2 \end{bmatrix},$$

where  $\sigma_1^2$  and  $\sigma_2^2$  are the variances of the distributions from which  $x_1$  and  $x_2$  were drawn, respectively;  $\sigma_3^2$  and  $\sigma_4^2$  are the variances of the distributions from which  $y_1$  and  $y_2$  were drawn, respectively;  $\rho_1$  is the correlation coefficient between the dimensions of  $h(x)$  and  $\rho_2$  is the correlation coefficient between the dimensions of  $h(y)$ .

## Similarity Functions and Distance Metrics

Shepard (1987) proposed the basis for a law of generalization involving the following two ideas: first, that the probability that a response learned to stimulus  $S_i$  will be made to stimulus  $S_j$  is approximately an exponential decay function of the distance between the stimuli in a space of a certain dimensionality; second, that the metric used to define this distance will be Euclidean when the psychological dimensions are integral and city-block when they are separable. Shepard noted that the theory applied only to experiments in which generalization is tested immediately after a single learning trial with a novel stimulus. Shepard pointed out that with highly similar stimuli or with delayed test stimuli, the relationship between similarity and distance was of a Gaussian form and that the distance metric appeared to be Euclidean for cases in which the theory would predict city-block. The work of Nosofsky (1986) exemplifies this kind of result. Using highly similar stimuli, Nosofsky (1986) discussed identification and classification performance and used a "Gaussian" function in modelling the relationship between the Euclidean distance separating the stimulus points and similarity. With regard to Nosofsky's results, Shepard conjectured that internal noise may make "the otherwise sharply peaked gradient of generalization ... more nearly Gaussian."

There were two distances defined earlier under the assumptions for the similarity model. The distance between momentary trial psychological magnitudes was represented by  $d$ , while the distance between the means of the distributions of psychological magnitudes was  $\delta$ . Nosofsky and Shepard define the distance between the points representing the stimuli without psychological error and, consequently, treat distance in a deterministic manner. This concept of distance corresponds better to  $\delta$  than it does to  $d$ , since it is not expected to vary from trial to trial.

When modelling the relationship between  $\delta$  and  $f(\mu, V)$ , it is instructive to consider, for a particular similarity function and metric ( $\alpha$  and  $\gamma$ ), the effect of the

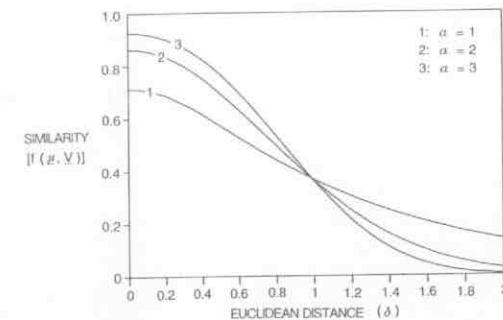


FIG. 1. Expected value of similarity as a function of the Euclidean distance between the means of the distributions of psychological magnitudes for values of  $\alpha$  of 1, 2, and 3 in the similarity function  $g(d) = \exp(-d^\alpha)$ .

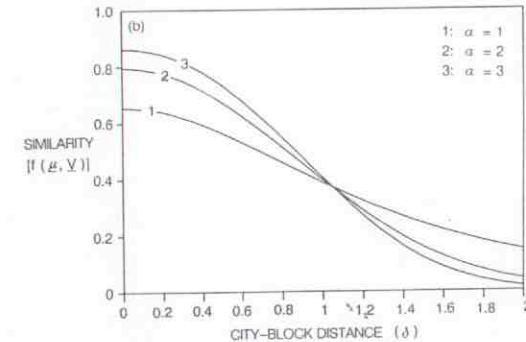
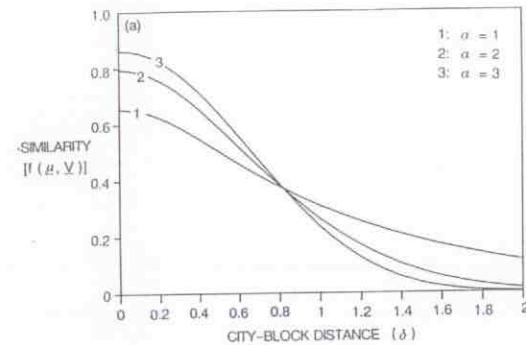


FIG. 2. Expected value of similarity as a function of the city-block distance between the means of the distributions of psychological magnitudes for values of  $\alpha$  of 1, 2, and 3 in the similarity function  $g(d) = \exp(-d^\alpha)$ . (a) Means differ on one axis only; (b) means differ equally on both axes.

multidimensional stochastic portion of the model on this relationship. Figures 1, 2a, and 2b show that for Euclidean and city-block metrics, the relationship between  $\delta$  and  $f(\mu, V)$  will have a modified Gaussian form for a range of similarity functions ( $\alpha = 1, 2$ , or 3). For all of the points in these figures, it was assumed that  $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 0.2$  (equal variance on all dimensions for all stimuli) and that  $\rho_1 = \rho_2 = 0.0$  (separable dimension stimuli). [Note that, for these parameters, the relative orientation of the stimulus means to each other will not affect  $f(\mu, V)$  when the similarity function involves a Euclidean distance metric; but when the city-block metric is assumed, it will.] These figures suggest, qualitatively consistent with Nosofsky's findings, that a modified Gaussian function relating  $f(\mu, V)$  and  $\delta$  should be expected, even if the similarity function is an exponential decay function and the metric defining  $d$  (within-trial distance) is city-block.

Assume that subjects employ an exponential decay similarity function ( $\alpha = 1$ ).

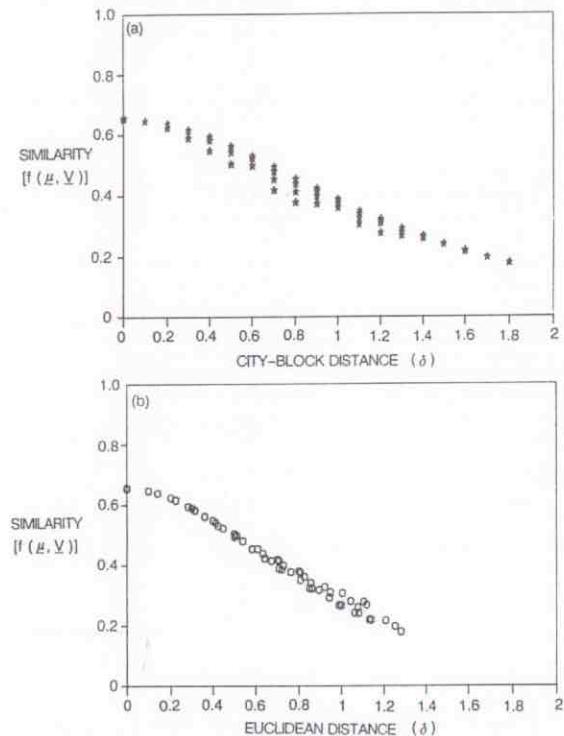


FIG. 3. Expected value of similarity between pairs of 16 stimuli plotted against the city-block and Euclidean distances ( $\delta$ ) between the means of the distributions of psychological magnitudes. An exponential decay function has been used to describe the relationship between city-block distance ( $d$ ) and within-trial similarity.

within each trial and that the city-block distance metric is also employed ( $\gamma = 1$ ). Consider 16 stimuli whose momentary psychological magnitudes can be represented mentally by independent multivariate normal distributions with means  $(0, 0)$ ,  $(0.8, 1.2)$ ,  $(0.3, 0.7)$ ,  $(0.9, 1.1)$ ,  $(1.2, 0.6)$ ,  $(0.8, 0.8)$ ,  $(0.1, 0.5)$ ,  $(0.3, 0.0)$ ,  $(0.7, 0.1)$ ,  $(1.1, 1.0)$ ,  $(0.9, 0.6)$ ,  $(0.6, 0.6)$ ,  $(0.4, 0.4)$ ,  $(0.2, 1.2)$ ,  $(0.9, 0.1)$ ,  $(0.7, 0.7)$ ; variances  $(0.2, 0.2)$  for all stimuli; and correlation coefficients of zero between dimensions for all stimuli. Imagine that the experimenter knows the location parameters (means) for each stimulus so that the distance between means (Euclidean and city-block metrics) can be computed. From Eq. (1) one can obtain the expected similarity value for each pair of stimuli assuming that  $\alpha = 1$  and  $\gamma = 1$  within each trial. It is interesting to inquire about the relationship between  $\delta$  (the distance between population means) and  $f(\mu, V)$  (the expected value of similarity). Figures 3a and 3b show this relationship for this set of 16 coordinates in two dimensions. Given the modified Gaussian form of these figures, it seems reasonable to attempt to fit a linear function relating  $\ln[f(\mu, V)]$  and  $\delta^2$  to the data for both metric forms of  $\delta$ . Such a linear regression analysis suggests that the Euclidean metric leads to a fit of the data ( $r^2 = .98$ ) which is at least as good as the city-block metric ( $r^2 = .95$ ). This conclusion might also be reached by simple inspection of the figures. Qualitatively, at least, one can conclude that the distance metric appropriate to the function relating distance to the expected value of the similarity of pairs of stimuli, evoking separable dimension representations, may be different from the metric employed by subjects within individual trials. It is possible, consequently, to reconcile Nosofsky's findings with those of Shepard's regarding the form of the metric provided that

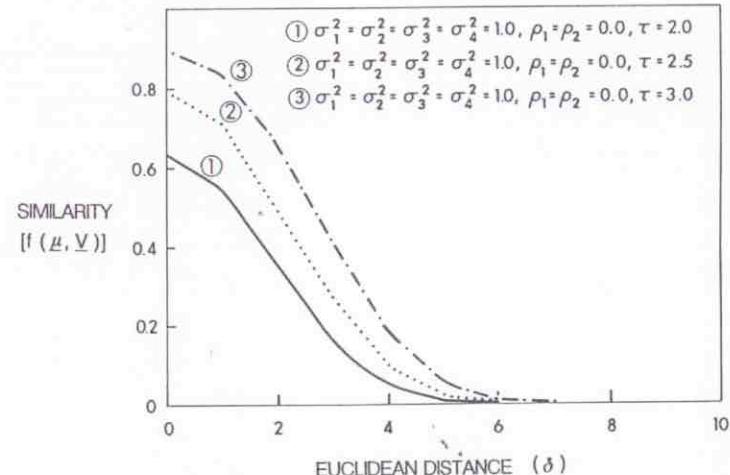


FIG. 4. Expected value of similarity as a function of the Euclidean distance between the means of the distributions of psychological magnitudes for different values of  $\tau$  when a step judgment function is assumed.

Shepard's theory concerning the similarity function and the metric is applied within trials for confusable stimuli. Specific comments on Nosofsky (1986) and Shepard (1986, 1987) have been made (Ennis, 1988a, 1988b).

### The Step Function

Figure 4 shows the relationship between  $\delta$  and  $f(\mu, V)$  for step functions where  $\tau$  is 2.0, 2.5, and 3.0 and where equal variances of 1.0 and correlation coefficients of 0.0 are assumed. In order to produce self-similarity values ( $\delta = 0.0$ ) in the 0.8–0.9 range,  $\tau$  should be between about 2.0 and 3.0 for this case. Although the step function model will make differential predictions for varying values of the stochastic parameters, it is quite limited compared to the continuous form of  $g$ . The only way to manipulate the rate of decrease of  $f(\mu, V)$  as a function of  $\delta$  for a given  $V$  matrix, for instance, is to change  $\tau$ . This will also have the effect of changing the predicted probability of declaring identical objects "same."

### STOCHASTIC MULTIDIMENSIONAL SCALING

Assuming that  $\alpha$  and  $\gamma$  are given, we have shown how the expected value of similarity is a function of the difference between the means of the distributions of psychological magnitudes ( $\mu$ ) and the variance-covariance matrix of the difference between psychological values ( $V$ ). It should, therefore, be possible to estimate the means and variance-covariance matrices of the psychological magnitudes corresponding to a selection of objects. For the case  $\alpha = 2$  and  $\gamma = 2$ , the means and standard errors for 36 stimuli in two dimensions were sampled at random from distributions that yielded values of  $f(\mu, V)$  in the range 0.5–1.0. One of the stimuli was assigned the mean (0, 0). All correlation coefficients were assigned the value 0.0. The matrix of 666 similarity values (all stimulus pairs including self-comparisons) was obtained by solving Eq. (2) for the selected means and standard errors. A modified Levenberg–Marquardt (steepest descent) algorithm was used to obtain multidimensional parameter values for which the difference between the similarities corresponding to the parameters obtained and the input similarities was minimum in a least-squares sense. Let  $a$  be a vector containing the parameters to be estimated. These are the estimates of the means and standard errors of the distributions of interest. From  $a$ , it is simple to compute  $\mu_y$  and  $V_y$  (the means of differences and variance-covariance matrix of differences for stimuli  $S_i$  and  $S_j$ ) and, consequently,  $f(\mu_y, V_y)$  can then be computed from Eq. (2). The function to be minimized is

$$q(a) = \sum_i \sum_j [P_{ij} - f(\mu_y, V_y)]^2, \quad j \leq i,$$

where  $P_{ij}$  is the observed probability of declaring  $S_i$  and  $S_j$  to be "same." A key to solving this problem and avoiding local minima is the generation of good initial starting values.

The analysis was initiated with randomly generated values of the means and assuming that all standard errors were 0.2 and that correlation coefficients were 0.0. The value of 0.2 for the standard errors was chosen because this value yields a self-similarity value of about 0.85, which roughly corresponded to the average diagonal value of the same-different matrix. The parameter values at this minimum were then used as the starting configuration for a second stage in which all standard errors were assumed to be equal across dimensions for a particular stimulus, but may vary across stimuli. The configuration at the minimum from stage 2 was used as the starting configuration for the final stage in which the standard errors may vary across both stimuli and dimensions.

The results of this analysis are given in Table 1. This table shows the means and standard errors of the original configuration of 36 points and their corresponding estimates. These estimates differ only slightly in the third decimal place from the actual values, supporting the validity of the strategy used to reach the minimum. The residual sum of squares at this minimum was  $< 0.001$ . An attempt to estimate all of the parameters in one stage failed to recover the original configuration. It is interesting to note that the results reported in Table 1 were obtained without rotation of the estimated configuration and are a mirror image of the original configuration. This orientational uniqueness is a consequence of variance inequality. The ability to directly interpret the results of a multidimensional scaling analysis without the arbitrariness introduced by rotation should prove useful in identifying the dimensions employed by subjects when comparing stimulus objects.

Differences between pairs of identical stimuli obtained from same-different judgments can be viewed as a consequence of differences in variances on one or more of the dimensions involved in the decision process. Ashby and Perrin (1988) have discussed this kind of interpretation of self-similarity. There may also be differences in self-similarity due to different numbers of psychological dimensions involved in the judgment. Krumhansl (1978) proposed a spatial density model to explain differences in self-similarity and asymmetrical similarities. Alternatively, it may be possible to formulate the effects of spatial density in terms of variance differences. Psychological magnitudes obtained from means located in a densely populated region of the space may have been sampled from distributions with higher variance than those located in a less densely populated area. Consequently, self-similarity measured in terms of  $f(\mu, V)$  would be lower in dense regions than in sparse regions. This hypothesis was supported by a reanalysis of the Rothkopf (1957) Morse code same-different matrix, as can be seen in Fig. 5.

Using the parameter estimation procedure described earlier for the artificial data set, means and variances for the Rothkopf data were obtained assuming that  $\alpha = 2$  and  $\gamma = 2$ . Solutions in which it was assumed that the variances across dimensions for a particular stimulus were equal and unequal were obtained. The unequal variance model gave a slightly lower residual sum of squares than the equal variance model, but the configurations of means for the stimuli were almost identical. For convenience in comparing the relative variances of the stimuli, the equal variance model was used. Figure 5 shows that the size of the standard error for a stimulus

TABLE I  
Actual Means and Standard Errors for 36 Stimuli and  
Their Estimates Obtained Using Nonlinear Least-Squares Minimization

Means				Standard errors			
Dimension 1		Dimension 2		Dimension 1		Dimension 2	
(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
0.000	0.000	0.000	0.000	0.152	0.154	0.253	0.251
-0.046	-0.046	-0.235	0.234	0.182	0.184	0.240	0.238
0.059	0.059	0.149	-0.149	0.190	0.192	0.150	0.148
0.361	0.361	0.105	-0.105	0.679	0.680	0.236	0.235
-0.102	-0.103	-0.039	0.039	0.213	0.215	0.382	0.381
0.000	0.000	-0.224	0.224	0.216	0.218	0.240	0.238
-0.041	-0.041	-0.113	0.113	0.238	0.240	0.211	0.209
-0.093	-0.094	-0.100	0.100	0.237	0.238	0.425	0.424
-0.314	-0.314	0.117	-0.117	0.227	0.229	0.162	0.160
-0.405	-0.406	0.296	-0.295	0.286	0.288	0.173	0.170
0.144	0.144	0.101	-0.100	0.160	0.162	0.392	0.390
-0.023	-0.023	0.142	-0.142	0.231	0.233	0.152	0.149
-0.194	-0.194	0.111	-0.111	0.150	0.153	0.170	0.168
-0.106	-0.106	-0.119	0.119	0.158	0.160	0.179	0.177
-0.281	-0.282	-0.250	0.250	0.150	0.153	0.177	0.174
0.146	0.147	-0.096	0.096	0.298	0.300	0.151	0.149
0.140	0.141	0.278	-0.277	0.155	0.157	0.154	0.151
0.521	0.521	-0.138	0.138	0.150	0.153	0.217	0.215
0.095	0.096	-0.035	0.035	0.465	0.467	0.235	0.233
0.108	0.108	-0.047	0.047	0.168	0.170	0.164	0.162
-0.179	-0.179	-0.405	0.405	0.588	0.590	0.186	0.183
-0.047	-0.047	0.022	-0.022	0.330	0.332	0.272	0.271
-0.142	-0.143	0.032	-0.032	0.224	0.226	0.254	0.252
-0.099	-0.100	0.072	-0.072	0.155	0.158	0.181	0.179
-0.077	-0.077	-0.034	0.034	0.150	0.152	0.620	0.619
0.124	0.124	0.206	-0.206	0.267	0.268	0.151	0.149
0.270	0.271	-0.154	0.154	0.155	0.158	0.210	0.208
-0.268	-0.269	0.218	-0.218	0.210	0.212	0.266	0.262
0.205	0.206	0.012	-0.012	0.161	0.164	0.695	0.693
0.282	0.282	-0.261	0.261	0.172	0.175	0.160	0.158
0.369	0.369	0.023	-0.023	0.171	0.174	0.152	0.150
-0.025	-0.025	0.013	-0.013	0.559	0.561	0.156	0.154
-0.329	-0.329	0.068	-0.068	0.238	0.240	0.293	0.291
-0.127	-0.128	0.151	-0.150	0.179	0.182	0.181	0.179
-0.101	-0.101	-0.090	0.090	0.167	0.169	0.157	0.155
-0.368	-0.368	0.207	-0.207	0.162	0.165	0.160	0.158

Note. Actual values are designated (1) and estimates (2).

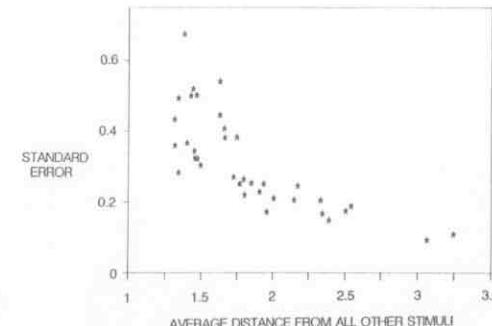


FIG. 5. The standard error on either dimension for 36 Morse code signals plotted in relation to the average Euclidean distance between each stimulus mean and the 35 other stimulus means.

distribution is related to the degree to which that stimulus is isolated from the other stimuli in the set under study (measured by the average Euclidean distance between a stimulus and all the other stimuli). Shepard (1963) had applied nonmetric multidimensional scaling to the same Rothkopf matrix and obtained a configuration of perceived Morse code signals. The configuration of means obtained using stochastic multidimensional scaling was quite similar to Shepard's configuration and would certainly have led to the same interpretation of the dimensions (number of signal components and the dots/dashes ratio). This result is not that surprising for this matrix because many of the same-different judgment probabilities were less than 0.5 suggesting, in the absence of response bias, that many pairs of signals were not highly confusable. Notwithstanding these comments, this data set proved useful in showing that some stimuli may be more precisely perceived than others and that this precision may depend on the location of a stimulus relative to the others in the data set. Variance differences between stimuli may explain differences in self-similarity. It would be interesting to analyse a matrix of same-different judgments obtained from more highly confusable objects than the Morse code signals.

When  $\alpha = 2$  and  $\gamma = 2$  it is possible to estimate the parameters of a sample problem rapidly without the need to use the much slower numerical methods to evaluate Eq. (1). To include  $\alpha$  and  $\gamma$  as parameters to be estimated is possible, but would require extensive numerical evaluation of Eq. (1), leading to a several hundred-fold increase in computation time, depending on the desired accuracy of the numerical analysis. A closed form for the special case  $\alpha = 2$ ,  $\gamma = 2$  is given in Eq. (2) and it would be very useful to have a similar algebraic form when  $\alpha = 1$ ,  $\gamma = 1$ . This is important because the city-block metric and the exponential decay function may be universally inherent in similarity judgments and identification decisions with separable dimension stimuli.

In considering the similarity function,  $g$ , we have restricted ourselves to a particular form,  $\exp(-d^\alpha)$ . If  $g$  were defined as any monotonically decreasing function of  $d$ , then it can be seen that nonmetric multidimensional scaling would be a special

case of the more general approach described here, but in which variances are assumed to be zero.

### CONCLUSION

A multidimensional model of similarity has been described which involves a distance-based similarity function and an assumed distribution of momentary psychological magnitudes from which the distance is derived. Evaluation of the same-different judgment model shows that it is possible to produce a modified Gaussian function relating similarity to the distance between the means of the distributions of psychological magnitudes even if the within-trial similarity function is an exponential decay function. Nosofsky's findings regarding the form of the metric (Euclidean) for a particular set of confusable stimuli is consistent with Shepard's theory that the appropriate metric is city-block for separable stimuli, provided Shepard's theory is applied at the individual trial level.

Using a nonlinear least-squares procedure, it is shown how the parameters of a sample problem may be estimated from a matrix of hypothetical same-different judgments. Because of uniqueness introduced by unequal variances, where such variances exist, the multidimensional scaling analysis yields a solution configuration that does not require rotation to interpret the psychological dimensions used by the subject. Assuming that the judgment function is any monotonically decreasing function of the distance between the momentary within-trial psychological magnitudes, the stochastic multidimensional scaling procedure described in this paper is a general case which includes deterministic approaches, such as nonmetric multidimensional scaling, as special cases.

### APPENDIX

The momentary psychological values are  $\mathbf{x}$  and  $\mathbf{y}$  where  $\mathbf{x}' = (x_1, x_2, \dots, x_n)$ ,  $\mathbf{y}' = (y_1, y_2, \dots, y_n)$ ;  $\mathbf{x}'$  indicates an  $n$ -dimensional row vector and  $n$  is the number of sensory dimensions. The momentary psychological values are mutually independently distributed with  $\mathbf{x}$  having density function  $h(\mathbf{x})$  and  $\mathbf{y}$  having density function  $h(\mathbf{y})$ . The probability densities  $h(\mathbf{x})$  and  $h(\mathbf{y})$  are multivariate normal distributions with means  $\boldsymbol{\mu}_x$  and  $\boldsymbol{\mu}_y$ , where  $\boldsymbol{\mu}_x' = (\mu_{x1}, \mu_{x2}, \dots, \mu_{xn})$  and  $\boldsymbol{\mu}_y' = (\mu_{y1}, \mu_{y2}, \dots, \mu_{yn})$ , and variance-covariance matrices  $\mathbf{V}_x$  and  $\mathbf{V}_y$ .

On the basis of the momentary psychological values,  $\mathbf{x}$  and  $\mathbf{y}$ , the subject decides whether the stimuli are the same or different.

Let  $\mathbf{z} = \mathbf{x} - \mathbf{y}$  and  $d$  represent the momentary distance between  $\mathbf{x}$  and  $\mathbf{y}$  perceived by the subject, where

$$d = \left[ \sum_{k=1}^n |z_k|^\gamma \right]^{1/\gamma}.$$

$\mathbf{V}$  is the variance-covariance matrix of the difference between psychological values,  $\mathbf{z}$ .  $\boldsymbol{\mu}$  is a vector of differences between the means of the momentary psychological values,  $\boldsymbol{\mu}_x$  and  $\boldsymbol{\mu}_y$ .

The expected value of similarity, in the absence of response bias, is  $f(\boldsymbol{\mu}, \mathbf{V}, \alpha, \gamma)$ . In an individual trial, similarity is defined as  $g(d)$ , where

$$g(d) = \exp(-d^\alpha)$$

$$= \exp\left(-\left[\sum_{k=1}^n |z_k|^\gamma\right]^{\alpha/\gamma}\right)$$

$$= G(\mathbf{z}),$$

$$f(\boldsymbol{\mu}, \mathbf{V}, \alpha, \gamma) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\exp\{-0.5(\mathbf{z} - \boldsymbol{\mu})' \mathbf{V}^{-1}(\mathbf{z} - \boldsymbol{\mu})\}}{(2\pi)^{n/2} |\mathbf{V}|^{1/2}} \times \exp(-d^\alpha) dz_1 dz_2 \cdots dz_n$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{\exp\{-0.5(\mathbf{z} - \boldsymbol{\mu})' \mathbf{V}^{-1}(\mathbf{z} - \boldsymbol{\mu})\}}{(2\pi)^{n/2} |\mathbf{V}|^{1/2}} \times G(\mathbf{z}) dz_1 dz_2 \cdots dz_n.$$

Consider the case when  $\gamma = 2$ ,  $\alpha = 2$ ,

$$\gamma = 2, \quad \alpha = 2 \rightarrow G(\mathbf{z}) = \exp[-(\mathbf{z}' \mathbf{z})].$$

Since  $(\mathbf{z} - \boldsymbol{\mu})' \mathbf{V}^{-1}(\mathbf{z} - \boldsymbol{\mu}) = \mathbf{z}' \mathbf{V}^{-1} \mathbf{z} - \mathbf{z}' \mathbf{V}^{-1} \boldsymbol{\mu} - \boldsymbol{\mu}' \mathbf{V}^{-1} \mathbf{z} + \boldsymbol{\mu}' \mathbf{V}^{-1} \boldsymbol{\mu}$ , then

$$f(\boldsymbol{\mu}, \mathbf{V}) = (2\pi)^{-n/2} |\mathbf{V}|^{-1/2}$$

$$\times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp[-0.5(\mathbf{z}' \mathbf{J} \mathbf{z} - \mathbf{z}' \mathbf{b} - \mathbf{b}' \mathbf{z} + \boldsymbol{\mu}' \mathbf{V}^{-1} \boldsymbol{\mu})] dz_1 \cdots dz_n,$$

where  $\mathbf{J} = \mathbf{V}^{-1} + 2\mathbf{I}$  and  $\mathbf{b} = \mathbf{V}^{-1} \boldsymbol{\mu}$ .  $\mathbf{V}^{-1}$ ,  $\mathbf{I}$ , and  $\mathbf{ss}'$  are symmetric; thus  $\mathbf{J}$  is symmetric.

Define the following:

$\lambda_i$  are the  $n$  distinct eigenvalues of  $\mathbf{J}$  (since  $\mathbf{J}$  is symmetric),

$\mathbf{V}_{\lambda i}$  are the eigenvectors,

$\mathbf{C} = (\mathbf{V}_{\lambda 1}, \dots, \mathbf{V}_{\lambda n})$ , and,

for any  $\mathbf{x}$ ,  $\mathbf{D}^x = (c_{ij})$ ;  $c_{ii} = \lambda_i^x$ ;  $c_{ij} = 0$ ,  $i \neq j$ .

$\mathbf{V}_{\lambda 1}, \dots, \mathbf{V}_{\lambda n}$  form an orthonormal basis; therefore

$$\mathbf{C}^{-1} \mathbf{J} \mathbf{C} = \mathbf{D} \text{ and } \mathbf{C}^{-1} = \mathbf{C}^T,$$

$$\mathbf{z}' \mathbf{J} \mathbf{z} = \mathbf{z}' \mathbf{C} \mathbf{C}^{-1} \mathbf{J} \mathbf{C} \mathbf{C}^{-1} \mathbf{z}$$

$$= \mathbf{z}' \mathbf{C} \mathbf{D} \mathbf{C}^{-1} \mathbf{z}$$

$$= (\mathbf{z}' \mathbf{C} \mathbf{D}^{1/2}) (\mathbf{D}^{1/2} \mathbf{C}^T \mathbf{z}).$$

Let

$$t = D^{1/2}C^T z - D^{-1/2}C^T b$$

$$t' = z' C D^{1/2} - b' C D^{-1/2}.$$

For each  $z_i$  there exists a  $t_j$  such that  $dt_j = \lambda_i dz_i$ ; thus

$$dt_1 \cdots dt_n = (\lambda_1 \lambda_2 \cdots \lambda_n)^{1/2} dz_1 \cdots dz_n$$

$$= |D|^{1/2} dz_1 \cdots dz_n$$

$$= |J|^{1/2} dz_1 \cdots dz_n.$$

$$f(\mu, V) = (2\pi)^{-n/2} |V|^{-1/2} \cdot |J|^{-1/2}$$

$$\times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp[-0.5(t't + \mu'V^{-1}\mu - b'Jb)] dt_1 \cdots dt_n$$

$$= (2\pi)^{-n/2} |V|^{-1/2} |J|^{-1/2} \exp[-0.5(\mu'V^{-1}\mu - b'Jb)]$$

$$\times \int_{-\infty}^{\infty} \exp(-t_1^2/2) dt_1 \int_{-\infty}^{\infty} \exp(-t_2^2/2) dt_2 \cdots$$

$$\times \int_{-\infty}^{\infty} \exp(-t_n^2/2) dt_n$$

$$= (|V| |J|)^{-1/2} \exp[\mu'(2J^{-1} - I)\mu],$$

where

$$J = V^{-1} + 2I.$$

#### REFERENCES

ASHBY, F. G., & PERRIN, N. (1988). Toward a unified theory of similarity and recognition. *Psychological Review*, **95**, 124-150.

DE SOETE, G., CARROLL, J. D., & DESARBO, W. S. (1986). The wandering ideal point model: A probabilistic multidimensional unfolding model for paired comparisons data. *Journal of Mathematical Psychology*, **30**, 28-41.

ENNIS, D. M. (1988a). Confusable and discriminable stimuli: Comments on Nosofsky and Shepard. *Journal of Experimental Psychology*, in press.

ENNIS, D. M. (1988b). Technical comment: Toward a universal law of generalization. *Science*, in press. [Comment on SHEPARD, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, **237**, 1317-1323.]

ENNIS, D. M., & MULLEN, K. (1986a). A multivariate model for discrimination methods. *Journal of Mathematical Psychology*, **30**, 206-219.

ENNIS, D. M., & MULLEN, K. (1986b). Theoretical aspects of sensory discrimination. *Chemical Senses*, **11**, 513-522.

ENNIS, D. M., MULLEN, K., & FRUTTERS, J. E. R. (1988). Variants of the method of triads: Unidimensional Thurstonian models. *British Journal of Mathematical and Statistical Psychology*, **41**, 25-36.

GENZ, A. C., & MALIK, A. A. (1980). Remarks on algorithm 006: An adaptive algorithm for numerical integration over an  $N$ -dimensional rectangular region. *Journal of Computing and Applied Mathematics*, **6**, 295-302.

HEFNER, R. A. (1958). Extensions of the law of comparative judgment to discriminable and multidimensional stimuli. Unpublished doctoral dissertation, University of Michigan.

KAPENGA, J. A., DE DONCKER, E., MULLEN, K., & ENNIS, D. M. (1987). The integration of the multivariate normal density function for the triangular method. In P. Keast and G. Fairweather (Eds.), *Numerical integration* (pp. 321-328).

KRUMHANSL, C. L. (1978). Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density. *Psychological Review*, **85**, 445-463.

MACKAY, D. B. (1987). Personal communication.

MULLEN, K., & ENNIS, D. M. (1987). Mathematical formulation of multivariate Euclidean models for discrimination methods. *Psychometrika*, **52**, 235-249.

MULLEN, K., ENNIS, D. M., DE DONCKER, E., & KAPENGA, J. A. (1988). Multivariate models for the triangular and duo-trio methods. *Biometrics*, in press.

NOSOFSKY, R. M. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.

ROTHKOPF, E. Z. (1957). A measure of stimulus similarity and errors in some paired-associate learning tasks. *Journal of Experimental Psychology*, **53**, 93-101.

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

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

SHEPARD, R. N. (1986). Discrimination and generalization in identification and classification: Comment on Nosofsky. *Journal of Experimental Psychology: General*, **115**, 58-61.

SHEPARD, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, **237**, 1317-1323.

THURSTONE, L. L. (1927). A law of comparative judgment. *Psychological Review*, **34**, 273-286.

ZINNES, J. L., & MACKAY, D. B. (1983). Probabilistic multidimensional scaling: Complete and incomplete data. *Psychometrika*, **48**, 27-48.

ZINNES, J. L., & MACKAY, D. B. (1987). Probabilistic multidimensional analysis of preference ratio judgments. *Communication and Cognition*, **20**, 17-44.

RECEIVED: November 16, 1987