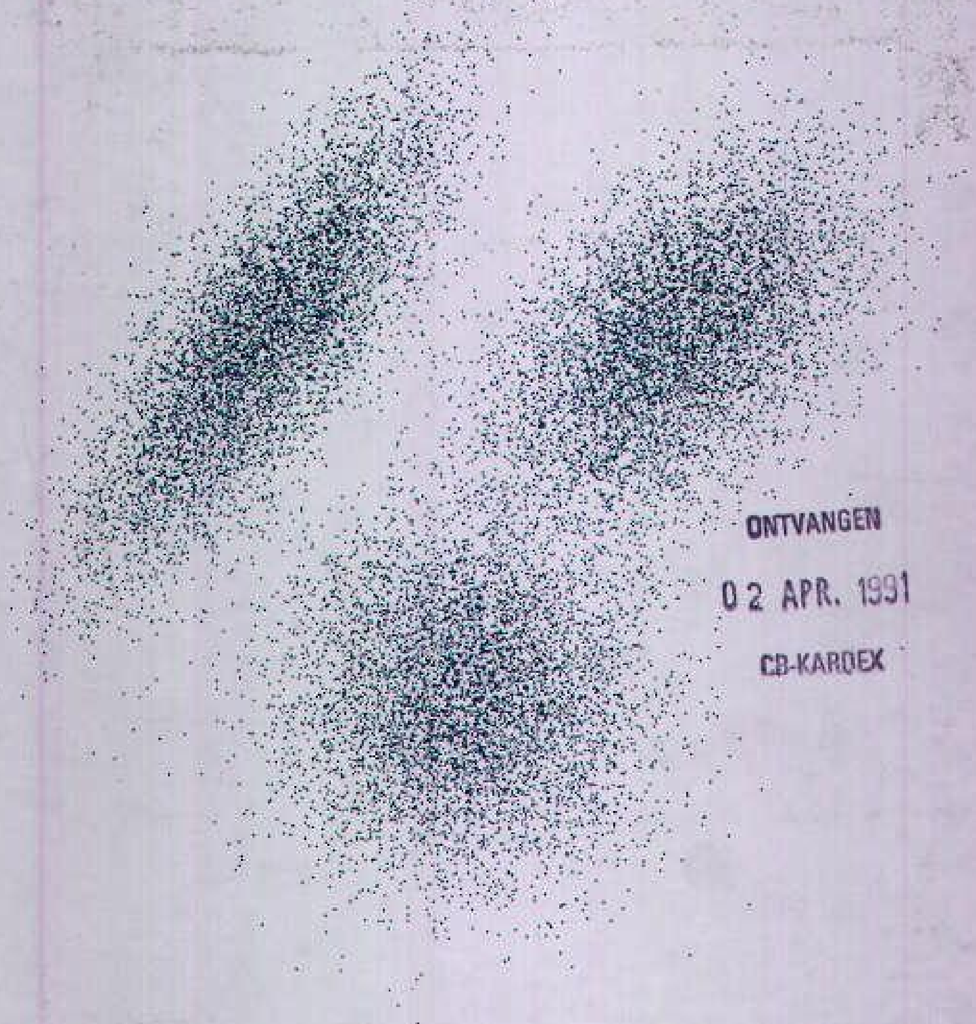
1406

Probabilistic Models of Perception



Daniel M. Ennis

NN08201, 1406

POSTULATES

1. A probabilistic model involving an exponential decay similarity function is consistent with an observed Gaussian relationship between similarity and distance involving confusable stimuli (Nosofsky, 1986, 1988; Shepard, 1988).

   Nosofsky, R.M. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.
   Nosofsky, R.M. (1988). On exemplar-based exemplar representations: Comment on Ennis (1988). *Journal of Experimental Psychology, General*, **117**, 412-414.
   Shepard, R.N. (1988). Time and distance in generalization and discrimination: Comment on Ennis (1988). *Journal of Experimental Psychology, General*, **117**, 415-416.
   This thesis.

2. The expected value of an exponential decay similarity function of city-block interstimulus distance will be experimentally insensitive to perceptual dependence (correlated psychological dimensions).

   This thesis.

3. Any psychological task which can be modeled by assuming that subjects choose one of two representations based on the smallest Euclidean distance to a third representation can be expressed in terms of the distribution of an indefinite quadratic form.

   This thesis.

4. Models based on the entire sequence from stimulus to percept will have greater validity, often with fewer parameters, than models which ignore the physicochemical parameters of stimuli.

   This thesis.

5. The power of the triangular and duo-trio discrimination methods is so low, relative to 2-alternative and 3-alternative forced choice procedures, that their extensive use in testing for small differences between stimuli is not to be recommended.

   Ennis, D.M. (1990). The relative power of difference testing methods in sensory evaluation. *Food Technology*, **44**, 114, 116 & 117.

6. Molecular models of chemical sensing show that molecular parameters, such as binding constants, can be estimated directly from perceptual information without using biochemical assays.

Ennis, D.M. (1989). A binary mixture model applied to the sweetness of fructose and glucose mixtures: De Graaf and Frijters revisited. *Chemical Senses*, **14**, 597-604.
Ennis, D.M. (1991). Molecular mixture models based on competitive and noncompetitive agonism. *Chemical Senses, in press.*

7. Biodegradable polymers can be developed by incorporating naturally occurring amino acids into synthetic polyamides such as the alternating copolymer of ε-amino caproic acid and glycine.

Ennis, D.M. and Kramer, A. (1974). Bacteria capable of degrading polymeric and low molecular weight amides. *Lebensm. -Wiss. u. Technol.*, **7**, 214-216.
Ennis, D.M. and Kramer, A. (1975). A rapid microtechnique for detecting the biodegradability of nylons and related polyamides. *J. Food Sc.*, **40**, 181-185.

8. All synthetic fertilizers applied to soils with structural problems, caused by high exchange capacity clays, should contain up to 10% dicarboxylic acid salts (such as magnesium, potassium, ammonium and hexanediamine salts of adipic, glutaric and succinic acids) to improve soil structure.

Ennis, D.M., Kramer, A., Mazzocchi, P.H., Jameson, C.W. and Bailey, W.J. (1975). Synthetic N-releasing biodegradable soil conditioners. *HortScience*, **10**, 505-506.
Ennis, D.M. 1978. Synthetic N-releasing soil conditioner. US Patent 4066431.
Ennis, D.M. 1981. Process for improving soil structure. Canadian Patent 1099942.

9. An outlier detection and elimination procedure should be implemented in international judging competitions involving sports such as gymnastics and diving to help to avoid the unfair effects of catastrophic errors.

10. Children should not be introduced by schools to computers until they have completed their secondary education.

Postulates accompanying the thesis:
Probabilistic Models of Perception
Daniel M. Ennis, 27 March 1991
Agricultural University Wageningen.

PROBABILISTIC MODELS OF PERCEPTION

Promotor: dr. J.E.R. Frijters, bijzonder hoogleraar in de psychologische en sensorische aspecten van voeding en voedsel

DANIEL M. ENNIS

PROBABILISTIC MODELS OF PERCEPTION

Proefschrift
ter verkrijging van de graad van doctor
in de landbouw- en milieuwetenschappen,
op gezag van de rector magnificus,
dr. H.C. van der Plas,
in het openbaar te verdedigen
op woensdag 27 maart 1991
des namiddags te vier uur in de aula
van de Landbouwuniversiteit te Wageningen

ACKNOWLEDGMENTS

The research reported in this thesis was conducted at the Philip Morris Research Center, Richmond, Virginia, U.S.A. I would like to thank the present and past managerial staff of the Research Center who made it possible for the author to conduct this work. A very special thanks is due to Drs. Max Hausermann and Carolyn Levy who provided support at the early stages of the research. Miles Waugh made extensive computer resources available and gave support and advice concerning computational methods. Dr. Ed Gee was an active supporter of the research and helped by facilitating extensive contact with scientists outside the Research Center. This work would not have been possible without the support of the present managerial staff, notably Dr. Ken Houghton, Jim Myracle and Sam Nelson. The author is grateful to all of these individuals.

The technical staff of the Research Center played a critical role in the development of the ideas in this thesis. Dr. Cliff Lilly recognized the value of this research from its inception and his generous support through the Philip Morris Visiting Scientist Program has been of critical value. Through this program, it was possible for me to work with, learn from and collaborate with a large number of scientists outside the Research Center. John Tindall read and commented on all of the papers in the thesis, always in a very helpful and constructive manner.

Throughout the course of this research, I have worked with a number of excellent scientists. My promotor, Dr. J.E.R. Frijters, has provided excellent leadership in the advice that he has given, in the quality of his own past and current work and in the collaboration that we have been able to develop together. My colleague and collaborator, Dr. Ken Mullen, deserves special recognition not only for his contributions to the solution of mathematical problems, but also because he has been an invaluable mentor for me in mathematics and statistics.

Important contributions to mathematical and computing problems have been made by Drs. Elise de Doncker-Kapenga and John Kapenga and by Joseph Palen. Thanks are due to Dr. Greg Ashby with whom I have had extensive discussion and contact on the subject of similarity and identification performance. I also thank Dr. Rob Nosofsky who very generously provided advice and comment on my early attempts to model similarity and identification. Drs. Jean-Claude Falmagne and Norman Johnson provided important advice and insights into the formulation of some of the problems in the thesis as quadratic forms in normal variables. These ideas are important to the future development of these probabilistic models.

Finally, I would like to thank my wife, Tricia, who has been enthusiastically supportive of this work.

Probabilistic Models of Perception

Thesis, Wageningen Agricultural University, The Netherlands

Daniel M. Ennis

ABSTRACT

Mental representations of objects may fluctuate or change from moment to moment. Many models of similarity, identification, classification, and preferential choice are deterministic. These models cannot formally account for perceptual fluctuations. In this thesis, it is assumed that there exists a probability density function for psychological magnitudes (usually assumed to be multivariate normal) and a judgment function which defines how these magnitudes are used to make a particular decision. Based on these ideas, probabilistic models of triad discrimination, similarity, identification and preferential choice are derived and evaluated. Several of these models can account for differences in self-similarity, asymmetric similarities and violations of the triangle inequality because the metric axioms are not assumed to apply to proximity measures among stimulus means. A paradox, created when deterministic models of identification are compared, concerning the universal form of the similarity function and the distance metric, is resolved using a probabilistic model. The use of nonlinear least squares to estimate parameters is illustrated in the case of several of the models. *Fechner-Thurstone* models, in which stimulus variability, a psychophysical transformation, and psychological variability are formally included, are discussed.

CONTENTS

Abstract

Chapter 1 **General Introduction** 1

Chapter 2 **Overview** 5

Modelling similarity and identification when there are momentary fluctuations in psychological magnitudes. *In* F.G. Ashby (Ed.) *Probabilistic Multidimensional Models of Perception and Cognition.* 1991. Hillside, N.J.: Lawrence Erlbaum Associates, Inc, *in press.* 6

A general probabilistic model for triad discrimination, preferential choice and two-alternative identification. *In* F.G. Ashby (Ed.) *Probabilistic Multidimensional Models of Perception and Cognition.* 1991. Hillside, N.J.: Lawrence Erlbaum Associates, Inc, *in press.* 50

Chapter 3 **Unidimensional Probabilistic Models** 68

Variants of the method of triads: Unidimensional Thurstonian models. 1988. *British Journal of Mathematical and Statistical Psychology* **41**, 25-36. 69

Decision conflicts: Within-trial resampling in Richardson's method of triads. 1989. *British Journal of Mathematical and Statistical Psychology*, **42**, 265-269. 81

Fechner-Thurstone models. 1991. *Mathematical Social Sciences, submitted.* 86

Chapter 4 **Multidimensional Probabilistic Models** 102

The effect of dimensionality on results from the triangular method. 1985. *Chemical Senses* **10**, 605-608. 103

A multivariate model for discrimination methods. 1986. *Journal of Mathematical Psychology* **30**, 206-219. 107

Theoretical aspects of sensory discrimination. 1986. *Chemical Senses* **11**, 513-522. 121

A multidimensional stochastic theory of similarity. 1988. *Journal of Mathematical Psychology*, **32**, 449-465. 131

Confusable and discriminable stimuli: Comment on Nosofsky (1986) and Shepard (1986). 1988. *Journal of Experimental Psychology: General* **117**, 408-411. 148

A comparison of selected probabilistic multidimensional models of identification with respect to perceptual dependence. 1991. *Journal of Mathematical Psychology, submitted.* 152

Chapter 5 **General Discussion and Conclusions** 197

Chapter 6 **Summary** 202

Chapter 7 **Samenvatting** (Summary in Dutch) 207

Curriculum Vitae 212

CHAPTER 1

GENERAL INTRODUCTION

Probabilistic Models

A wide range of decision processes, such as those involved in identification, forced choice, preferential choice, grouping and categorization, can be mathematically modelled as stochastic or probabilistic processes. This means that the momentary psychological magnitudes or percepts are treated as if they were drawn from a particular probability distribution. In this thesis it will be assumed that this distribution is unidimensional or multidimensional normal. The primary goal of the thesis is the development and evaluation of models of several different types of subject tasks all of which share a common framework concerning the probability distribution of the percepts used in each task.

The work to be reported in this thesis started with an interest in extending the unidimensional Thurstone-Ura model for the triangular method (Frijters J.E.R., 1979, *British Journal of Mathematical and Statistical Psychology*, **32**, 229-241) to the multidimensional case. The triangular method is a tri-stimulus discrimination method commonly used in research on the chemical senses and in applications to food and beverage sensory testing. Frijters very clearly set out the framework for modelling a task such as the triangular method under the assumption that the momentary psychological magnitudes or percepts are univariate normally distributed. Once this research was initiated, several new research areas opened up and appeared worthy of pursuit. These included the development of Thurstonian models for other tristimulus methods, models which formally connect physicochemical measures, including variances, to their corresponding psychological magnitudes, and multidimensional probabilistic models of other perceptual processes such as similarity, identification and preferential choice.

Given a number of alternative probabilistic models for a particular task, such as identification, it has been of interest to compare the models with regard to how they respond to changes in the models' parameters (such as the degree of perceptual dependence or correlation between dimensions). This has been a fruitful area to pursue. There are large differences between the predictions made by several of the newly developed probabilistic models that may be related to easily manipulated behavioral variables.

Other results from the multidimensional model evaluations suggested new interpretations for existing identification data sets which may be more plausible than current deterministic identification models. A paradox concerning the metric which defines distances, and the judgment function used to determine similarity for highly confusable objects is resolved. These evaluations have led to a justification for the implementation of probabilistic models instead of deterministic ones in cases where perceptual noise is large.

A recurrent theme in many of the papers in this thesis is to extend existing models, find underlying connections between models and attempt to reach increasingly higher levels of consolidation and simplification. Some areas of perceptual measurement may appear unrelated, but actually share identical models. An example is the multidimensional probabilistic model for Torgerson's method of triads and a general model for preferential choice. This relationship is discussed along with other special cases of this model.

Organization of the Thesis

The thesis is divided into three main sections covered in Chapters 2, 3 and 4. Chapters 5, 6 and 7 contain the conclusions and summaries (in English and Dutch).

Chapter 2 contains an overview of the content of the thesis (in two papers) and will give the reader a background on the deterministic precursors of the probabilistic models discussed in the thesis.

There are three papers in Chapter 3 on unidimensional models. In the first of these papers, unidimensional models for Torgerson's and Richardson's methods of triads are derived and it is shown how the parameters of the models may be estimated using the method of nonlinear least squares. Two commonly used methods, the duo-trio and the triangular method are shown to be special cases of Torgerson's and Richardson's methods, respectively. In the second paper, the consequences of relaxing one of the assumptions concerning resampling within a trial in Richardson's method is explored. It is shown how decision conflicts may arise in which the same two stimuli may appear to be most alike *and* most different. A model for the probability of occurrence of this event is derived. The final paper in this section deals with models (called *Fechner-Thurstone* models) in which physicochemical parameters (means and variances), a psychophysical transformation and perceptual variance are included in a single model. Parameter estimates for a sample problem are provided along with a discussion of the parameter efficiency of models of this kind.

Chapter 4 contains six papers. In the first three papers, multivariate models for the duo-trio and triangular methods are described. These papers provide both Monte Carlo results as well as formal models for these methods which have been solved using numerical integration techniques. A fundamental assumption in multidimensional scaling, that there is a monotonic relationship between proximity measures and the perceptual distances between objects, is shown to be false when the objects are confusable.

Many commonly used models of similarity and choice are deterministic. The fourth paper in Chapter 4 addresses the development of a probabilistic model of similarity applied to *same-different* judgments. This paper contains a probabilistic approach to multidimensional scaling using both real and artificial data. In the paper, and the one following it, it is pointed out that what may appear to be a Gaussian judgment function and the Euclidean metric for distances between stimuli using a deterministic model may also be thought of as an exponential decay function and the city-block metric when perceptual variability is formally included in the model. The latter model is also more consistent with a large body of literature concerning the judgment function and metric in generalization experiments. Consequently, hidden structure in a data set may be revealed using a probabilistic model, where appropriate.

The sixth paper in Chapter 4 provides a comparison of three types of stochastic or probabilistic models of identification with regard to perceptual dependence or the degree to which perceptual dimensions are correlated. It is shown that some of these models are highly sensitive to perceptual dependence and others are not. The relevance of this result to experiments involving trained and untrained subjects is discussed.

CHAPTER 2

OVERVIEW

Modelling similarity and identification when there are momentary fluctuations in psychological magnitudes. *In* F.G. Ashby (Ed.) *Probabilistic Multidimensional Models of Perception and Cognition.* 1991. Hillside, N.J.: Lawrence Erlbaum Associates, Inc, *in press.*

A general probabilistic model for triad discrimination, preferential choice and two-alternative identification. *In* F.G. Ashby (Ed.) *Probabilistic Multidimensional Models of Perception and Cognition.* 1991. Hillside, N.J.: Lawrence Erlbaum Associates, Inc, *in press.*

Modelling Similarity and Identification when there are Momentary Fluctuations in Psychological Magnitudes

Daniel M. Ennis

Philip Morris Research Center

To discover the hidden structure in what we observe is a source of great enjoyment and a worthy goal for scientists to achieve. This structure is often revealed by successively employing models of increasing elegance and generality. The physical attributes of any set of objects will never occur at exactly the same value. Similarly, the chemical fluctuations in time and space around each cell, the cacophony in the living world, ensure that mental representations for the same and different objects will not be identical. Explicit models for these fluctuations might be based on known molecular/cellular processes and principles. In the absence of this knowledge, it is often useful to employ models that can be justified on the basis of experience in model fitting. These models may be shown later to have a basis in more fundamental processes, but initially must be viewed as operational. The models to be discussed in this chapter fall into this latter class.

The modest goal of this paper is to describe models which can be used to explore the consequences of momentary fluctuations in psychological values, irrespective of the processes responsible for these fluctuations. This exploration does indeed lead to the exposition of a hidden structure that cannot be seen from the perspective of models that ignore the existence of momentary fluctuations.

General Principles

An organism's behavior in responding to a stimulus can be modelled as if the organism transduced physical and/or chemical information into mental representations and employed some decision process. Many aspects of this sequence can be expressed in alternative mathematical forms, from which a particular model can be chosen. This selection would be based on fitting the models to experimental findings.

In general, psychological magnitudes can be treated as vectors in which each element of a vector corresponds to a value on a particular psychological continuum. Fluctuations in the vector magnitudes for a particular stimulus may occur because the physical stimulus may not be constant and/or because the information transduced to a mental representation (percept) may change from moment to moment. Fluctuations in the object or in the mental representation of it can be modelled using particular probability density functions (pdfs). In many probabilistic models, attention is paid only to fluctuations at the psychological level by assuming that physicochemical stimulus variance is zero and it has been common to assume that the psychological pdf is normal. Later in this chapter, the issue of physicochemical fluctuation will be discussed. If f is any psychological pdf, g is a judgment function for a particular task, and $\mathbf{z}$ is a vector which is a function of the momentary psychological magnitudes, then a very large number of tasks in psychology can be modelled based on the following simple equation:

$$P = \int_{D} f(\mathbf{z})g(\mathbf{z})d\mathbf{z}, \tag{1}$$

where P is the probability that a particular decision will be made and D is the joint domain of f and g. If f is a multivariate normal density function, then

$$f(\mathbf{z}) = \frac{\exp\{-0.5(\mathbf{z}-\underset{\sim}{\mu})' \; \Sigma^{-1}(\mathbf{z}-\underset{\sim}{\mu})\}}{(2\pi)^{n/2}|\Sigma|^{1/2}}. \tag{2}$$

The superscript, ′, denotes a row vector and $|\Sigma|$ is the determinant of Σ. Σ is the variance-covariance matrix for the **z** values and $\underset{\sim}{\mu}$ is their mean.

Similarity Models

Let $\mathbf{x}_i$ and $\mathbf{x}_j$ be vectors (n elements in each vector) of psychological magnitudes corresponding to two objects presented to a subject on a single trial. If the subject were to be presented with exactly the same objects an instant later, these psychological magnitudes might be different. Assume that the momentary psychological values are mutually independently distributed with $\mathbf{x}_i$ having density function f_i and $\mathbf{x}_j$ having density function f_j. The probability densities f_i and f_j are multivariate normal distributions with means $\underset{\sim}{\mu}_i$ and $\underset{\sim}{\mu}_j$ and variance-covariance matrices Σ_i and Σ_j. Based on the momentary psychological values, $\mathbf{x}_i$ and $\mathbf{x}_j$, the subject decides whether the stimuli are the same or different. Let $\mathbf{z} = \mathbf{x}_i - \mathbf{x}_j$. Σ is the variance-covariance matrix of **z**. When $n = 2$,

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

where σ_1^2 and σ_2^2 are the variances of the distributions from which x_{i1} and x_{i2} were drawn respectively; σ_3^2 and σ_4^2 are the variances of the distributions from which x_{j1} and x_{j2} were drawn respectively; ρ_1 is the correlation coefficient between the dimensions of $\mathbf{x}_i$; and ρ_2 is the correlation coefficient between the dimensions of $\mathbf{x}_j$. $\underset{\sim}{\mu}$ is a vector of differences between the means of the momentary psychological values, $\underset{\sim}{\mu}_i$ and $\underset{\sim}{\mu}_j$.

A general formula for the distance between the vectors $\mathbf{x}_i$ and $\mathbf{x}_j$ is the γ-Minkowski distance, d, where

$$d_{ij} = [\sum_{k=1}^{n} |z_k|^{\gamma}]^{1/\gamma} \qquad \gamma \geq 1. \tag{3}$$

If γ is 1, the distance, d, is referred to as the *city block* distance, and if γ is 2, d is referred to as the *Euclidean* distance.

One can similarly define the distance between population means as

$$\delta_{ij} = [\sum_{k=1}^{n} |\mu_{ik} - \mu_{jk}|^{\beta}]^{1/\beta} \qquad \beta \geq 1 \ . \tag{4}$$

It is extremely important to distinguish between d_{ij} and δ_{ij}. The distance, d_{ij}, is only defined for a particular trial. Once that trial is over, d_{ij} has no further meaning as far as the subject is concerned. The distance, δ_{ij}, is the distance between the means of the distributions of psychological magnitudes that give rise to d_{ij}. The means and the variance-covariance matrices of these distributions determine the likelihoods of occurrence of particular values of d_{ij} within a particular trial. The probability that a subject will ever directly experience psychological magnitudes equal to μ_i or μ_j is zero. In many traditional multidimensional scaling models that are not probabilistic (deterministic models), it is assumed that a subject will experience psychological magnitudes exactly equal to μ_i and μ_j whenever the two stimuli are presented. The difference between d_{ij} and δ_{ij} is central to differentiating between probabilistic models, which allow for fluctuations in psychological magnitudes from trial to trial, and deterministic models which make no probabilistic assumptions.

The Similarity Function

In Equation 1 it can be seen that the probability of making a particular decision depends on f and g. The function, g, is the judgment function. Suppose that g was concerned with perceived similarity between the momentary values, $\mathbf{x}_i$ and $\mathbf{x}_j$. Then g could be called the *similarity function*. *Similarity* can be defined in terms of d (which is a function of z). There are many different forms which could be proposed for the function, g. An obvious requirement would be that g decreased as d increased. Shepard (1987) proposed an exponential decay similarity function as a universal principle. A flexible function which includes the exponential decay function is

$$g(d) = \exp(-d^{\alpha}),\ \ \alpha \geq 0. \tag{5}$$

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

In order to use Equation 1, it is necessary to specify the probability density function for the momentary psychological magnitudes (Equation 2) and the judgment function to be used in making a particular decision (Equation 3). Both of these have now been defined and it follows from Equation 1 that the similarity of two objects over all possible trials is

$$P = \int_{R^n} \frac{\exp\{-0.5(z-\underset{\sim}{\mu})^t \Sigma^{-1}(z-\underset{\sim}{\mu})\}}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp(-d^{\alpha})\, dz. \tag{6}$$

Equation 6 does not include response bias, which can be accounted for by multiplying the right had side by a bias parameter. P is the *expected value* of g. Since Equation 6 postulates a normal distribution for the psychological magnitudes, originally proposed by Thurstone (1927) and then uses a general form to define similarity, which was motivated by Shepard's recent (1987) and earlier work, this model of similarity could be called a *Thurstone-Shepard* model.

Equation 6 can be evaluated numerically for a broad range of distance metrics (by varying γ) and similarity functions (by varying α). Computations of this type will be discussed a little later. There is a case, however, that deserves special mention because it leads to a closed form for Equation 6. If one assumes that the metric of d is Euclidean (γ is 2) and that α in the similarity function is 2, then

$$P = (|\Sigma||J|)^{-1/2} \exp[\underset{\sim}{\mu}'(2J^{-1} - I)\underset{\sim}{\mu}], \tag{7}$$

where $\mathbf{J} = \Sigma^{-1} + 2\mathbf{I}$ and $\mathbf{I}$ is the identity matrix.

The derivation of Equation 7 is given in Ennis, Palen and Mullen (1988). Naturally, Equation 7 is much faster to compute than Equation 6 for the special case of $\alpha = 2$ and $\gamma = 2$. For 5 decimal place accuracy, computational experience suggests a speed improvement of about 3 to 4 orders of magnitude. Of course, Equation 7 is only one special case, and not necessarily the most important one. A more interesting case may be when $\alpha = 1$ (g is exponential decay) and $\gamma = 1$ (city block metric). A simpler form for this case has not so far been derived.

Equation 5 gives g as a continuous function of d. Suppose that g was, instead, a step function of d. This would mean that G would be 0 or 1 depending on the value of d relative to some threshold or criterion, τ. In order to meet these objectives, let

$$g(d) = 0.5\{\mathrm{sgn}(\tau - d) + 1\}, \tag{8}$$

where sgn is the signum function. The signum function takes on the values 1, 1 and -1 whenever τ-d is greater than, equal to or less than zero, respectively. For instance, if $\tau - d > 0$, then $d < \tau$, sgn(τ-d) is 1 and $g(d)$ is 1. Similarly if $\tau = d$, $g(d)$ is 1. However, if $\tau - d < 0$, then $d > \tau$ and sgn(τ-d) is -1, leading to $g(d) = 0$. Using Equation 1, with this new definition of g results in

$$P = \int_{R^n} \frac{\exp\{-0.5(z-\mu)' \Sigma^{-1}(z-\mu)\}}{(2\pi)^{n/2}|\Sigma|^{1/2}} \, 0.5\{\mathrm{sgn}(\tau-d)+1\} \, dz. \quad (9)$$

τ may be a fixed value or may be drawn from a particular probability density function and vary from trial to trial.

Equations 6 and 9 have been discussed in terms of the expected value of the similarity function, but also correspond to the probability of giving a *same* response in a same-different task. Data of this kind will be analyzed later using Equations 6 and 7.

Perceptual Dependence and the Form of the Similarity Function

Ennis and Ashby (1990) recently showed that different probabilistic identification models differ greatly with regard to their sensitivity to perceptual dependence. It was shown that probabilistic identification models based on the idea of response regions (multidimensional signal detection theory) were more sensitive to perceptual dependence than models based on distance comparisons or the Shepard-Luce choice rule. See Ashby (1988), Ashby and Gott (1988), Ashby and Perrin (1988) and Ashby and Townsend (1986) for a discussion of recent developments in multidimensional signal detection theory. An identification model based on the exponential decay similarity function was particularly insensitive to perceptual dependence. This sensitivity was measured by comparing the difference between identification performance predictions when the variance-covariance matrices and mean difference vectors varied. Cases in which the correlation coefficients of the distributions being considered were either both positive or both negative were of special interest.

Some previous experience in evaluating Equation 9 (the step function) with different variance-covariance matrices, and a comparison of these evaluations with Equation 6, suggested that the sensitivity observed is related to the degree to which the judgment function, $g(\mathbf{z})$, approaches a step function. At the opposite extreme to the step function is a linear function of $\mathbf{z}$, the expected value of which (from Equation 1) is a linear function of the mean, $\underset{\sim}{\mu}$. Multidimensional signal detection identification models can be formulated in terms of Equation 1 by integrating over R^n with $f(\mathbf{z})$ as the probe's probability density function and $g(\mathbf{z})$ as a step function of the likelihoods that $\mathbf{z}$ is a random deviate from the two memory distributions. In order to test the step function hypothesis, Equation 6 was evaluated for a series of forms of the function, g, and for two cases of perceptual dependence and two levels of distributional overlap. Taking the basic form, $g(\mathbf{z}) = \exp(-d^{\alpha})$, α was varied from 1 to 25 in unit increments. These functions include exponential decay, Gaussian, and a series which, at $\alpha = 25$, approaches a step function. Figure 1a shows the form of g when α is 1, 2, 6 and 25. Figure 1b gives the difference between the values of P (from Equation 6) in the two perceptual dependence cases for the two levels of distributional overlap. When α is 1, the exponential decay case, the effect of perceptual dependence is very small. As α increases, especially if the level of distributional overlap is not great, the difference between the perceptual dependence cases increases and ultimately saturates. The maximum difference is seen when a step judgment function is operative.

This result has important implications for a variety of models which use Equation 5 to define similarity. One example is the Shepard-Luce choice rule, to be discussed in the next section. If Shepard's suggestion that the exponential decay similarity function formalizes a universal principle, then perceptual dependence will have only a small effect on task performance when objects are perceptually confusable. This theoretical result might be used to test hypotheses concerning the form of g that would then support or refute Shepard's theory concerning the nature of the similarity function.

Identification Models

Thurstone-Shepard-Luce Models

One approach to modelling absolute identification is to assume that there are m memory exemplars corresponding to m stimuli, $S_1, S_2, \ldots, S_m$. A probe stimulus, S_k, which may or may not have been in the original ensemble, is presented. It is assumed that the subject compares this probe to stored exemplars by determining the distances, $d_{k1}, d_{k2},..., d_{km}$ and then uses them in a similarity function such as Equation 5 to determine the similarities. Nosofsky (1986) modelled identification performance on the basis of the *Shepard-Luce* choice rule which models the probability that a subject will respond by identifying stimulus S_k as stimulus S_i(i.e. respond R_i to S_k). As a deterministic model, the choice rule is

$$P(R_i|S_k) = \frac{b_i h(d_{ki})}{\sum_{j=1}^{m} b_j h(d_{kj})}, \tag{10}$$

where $P(R_i|S_k)$ is the probability of responding R_i when the stimulus is S_k, b_i is the response bias parameter for the response R_i, d_{ki} is the distance between stimulus S_k and the memory representation of S_i, and h is a similarity function such as that given in Equation 5,

$$h(d) = \exp(-d^{\alpha}), \ \alpha \geq 0.$$

In a probabilistic model of identification, one might assume that the Shepard-Luce choice rule was being used within a trial and interest might center then on the expected value of the right hand side of Equation 10. Hence, when fluctuations in psychological magnitudes are to be accounted for,

$$P(R_i|S_k) = E\left[\frac{b_i h(d_{ki})}{\sum_{j=1}^{m} b_j h(d_{kj})}\right] \tag{11}$$

where E(r) is the expected value of r.

Consider the special case in which there are only two memory distributions with multivariate normal probability density functions, f_i and f_j and a probe with density function f_k. Let $\mathbf{u} = \mathbf{x}_k - \mathbf{x}_i$, $\mathbf{v} = \mathbf{x}_k - \mathbf{x}_j$, and $\mathbf{z} = (\mathbf{u},\mathbf{v})$. Σ is the variance-covariance matrix of the joint distribution of **u** and **v**, or **z**, and $\underset{\sim}{\mu}$ is the vector of mean differences previously defined. It can be shown that

$$\Sigma = \begin{bmatrix} \Sigma_1 & \Sigma_2 \\ \Sigma_2 & \Sigma_3 \end{bmatrix}, \text{ with}$$

$$\Sigma_1 = \Sigma_k + \Sigma_i, \ \Sigma_2 = \Sigma_k \text{ and } \Sigma_3 = \Sigma_k + \Sigma_j.$$

The judgment function introduced in Equation 1 can now be defined as a function of **z** using the Shepard-Luce choice rule by rewriting Equation 10 as

$$g(\mathbf{z}) = \frac{b_i h(d_{ki})}{\sum_{j=1}^{m} b_j h(d_{kj})}, \tag{12}$$

Applying Equation 1 to obtain the choice probability, it is necessary only to integrate over the vector space composed of all $2n$-tuples represented by **z**, weighting each element of the space by its probability of occurrence (given by the multivariate normal density function), or

$$P(R_i|S_k) = \int_{R^n}\int_{R^n} \frac{\exp\{-0.5(\mathbf{z}-\underline{\mu})^t\ \Sigma^{-1}(\mathbf{z}-\underline{\mu})\}}{(2\pi)^n\ |\Sigma|^{1/2}}\ \frac{b_i h(d_{ki})}{b_i h(d_{ki}) + b_j h(d_{kj})}\ d\mathbf{z}. \quad (13)$$

Since Equation 13 includes models which are probabilistic extensions of the Shepard-Luce choice rule, these models can be referred to as *Thurstone-Shepard-Luce* models.

If $\alpha = 1$ and $\gamma = 1$, h is an exponential decay function and the metric of d_{ki} is city-block. If $\alpha = 2$ and $\gamma = 2$, h is a Gaussian function and the metric of d_{ki} is Euclidean.

Identification Models Based on Ordinal Decision Rules

An alternative decision making process to the Shepard-Luce choice rule is to select a response based on the relative size of the momentary distances, d_{ki} and d_{kj}. If $d_{ki} < d_{kj}$ then the subject would give R_i as the response, for instance. The probabilistic model for this type of decision rule is identical (with two memory representations) to a Thurstonian variant of Torgerson's method of triads (Torgerson, 1958) and to a probabilistic generalization of Coombs' preference unfolding model (Coombs, 1964). Special cases of this model have been discussed by Ennis and Mullen (1986), Mullen and Ennis (1987), Mullen, Ennis, De Doncker and Kapenga (1988) and Ennis, Mullen and Frijters (1988). A major problem with this model had been the computing time needed to handle numerical integration of a $2n$-fold integral. Reduction of this integral to a single integral, for all values of n, is given in Mullen and Ennis (1990) and discussed in Chapter 5. This form of the model contributes importantly to solving the computational problem posed by the $2n$-fold integral. Reference to the use of the model for absolute identification is given in Chapter 5.

Identification Models Based on Category Distributions

Nosofsky (1986, 1990) discussed a model of categorization which was based on the context theory of classification proposed by Medin and Schaffer (1978). In this model, one assumes that a subject stores category exemplars in memory and that the probability of giving a category I response to stimulus S_k is

$$P(R_I|S_k) = \frac{b_I \sum_{i \varepsilon C_I} L_i(I) s_{ki}}{\sum_J b_J \sum_{j \varepsilon C_J} L_j(J) s_{kj}}, \tag{14}$$

where b_J is the response bias parameter for category C_J, $L_j(J)$ is the likelihood that exemplar$_j$ is presented during training as a member of category C_J, and s_{ki} is the similarity of exemplar$_k$ and exemplar$_i$.

In the previous section on Thurstone-Shepard-Luce models, it was assumed that each stimulus representation (either a probe or a memory representation) can be treated as a random value from a particular probability distribution. In this way, it was possible to capture the effect of momentary fluctuations in psychological magnitudes. It was also assumed that subjects make within-trial decisions based on single instances of stimulus and memory representations which were assumed to have been drawn from these distributions. The categorization model given in Equation 14 is a deterministic model with a finite set of training exemplars making up each category, and a finite set of likelihoods that each exemplar has been presented during training. Imagine, instead, that a category contains an infinite number of category exemplars with likelihoods of occurrence determined by a probability density function. The categories may now be viewed as the distributions corresponding to memory representations of individual stimuli and Equation 14 can be extended to yield identification probabilities for *particular* stimulus values. Hence,

$$\mathrm{P}(\mathrm{R}_i|\mathrm{S}_k) = \frac{b_i \int_{\mathrm{R}^n} f_i(\mathbf{x}) h(d_{ki})\, d\mathbf{x}}{\sum_{j=1}^{m} b_j \int_{\mathrm{R}^n} f(\mathbf{x})\, h(d_{kj}) d\mathbf{x}} . \tag{15}$$

All of the terms in Equation 15 have been defined in the previous section. Note that Equation 1 has been used in the numerator and denominator to obtain the expected values of the similarity of stimulus S_k to each of the m memory representations.

Equation 14 was a deterministic model for the categorization of a particular stimulus, and hence made no allowance for the possibility that the probe's mental representation may be a random variable. A more general model than Equation 15 is one in which presentation of the probe stimulus evokes a momentary value from a probability distribution itself. Thus, it is also necessary to integrate over the probe or stimulus pdf. Thinking again in terms of Equation 1 where,

$$g_k(x_k) = \frac{b_i \int_{R^n} f_i(\mathbf{x}) h(d_{ki}) d\mathbf{x}}{\sum_{j=1}^{m} b_j \int_{R^n} f_j(\mathbf{x}) h(d_{kj}) d\mathbf{x}},$$

this means that

$$P(R_i|S_k) = \int_{R^n} f_k(\mathbf{x}) g_k(\mathbf{x}) d\mathbf{x}. \tag{16}$$

Equation 16 gives the probability of identifying stimulus S_k as stimulus S_i based on the assumption that probes and memory representations of stimuli can be modelled as if they were drawn from particular probability distributions. A deterministic categorical model, such as that given in Equation 14 can be viewed as a special case of this model in which there are a finite number of elements in each category with a corresponding finite set of likelihoods of occurrence of each element.

A Resolution of the Shepard-Nosofsky Paradox

Shepard (1987) has proposed the exponential decay similarity function as a general form or universal principle of considerable importance when organisms make decisions about how to react to novel stimuli. Shepard has also given arguments in support of the city-block metric when psychological dimensions are separable (they can be attended to separately) and the Euclidean metric when they are integral (they cannot be attended to separately). Working with separable stimuli, Nosofsky (1986) provided very strong support for the Euclidean metric and a Gaussian form for the similarity function, which appeared to be incompatible with Shepard's theory. Both Nosofsky and Shepard used deterministic models and, therefore, did not take momentary fluctuations in the psychological magnitudes into account. Nosofsky's stimuli were highly similar, suggesting that perceptual variance should be included formally in models of his data. Equation 6 presents a form that should lead to a modified Gaussian relationship between P and δ. The extent of modification depends on the similarity function exponent, α, and the metric defining the within-trial distance, d. As can be seen from Figures 2, 3a and 3b, when the similarity function within a trial is exponential decay, the relationship between P and δ appear to be much more Gaussian in form than exponential decay. These results are discussed in Ennis, Palen and Mullen (1988) and Ennis (1988a,b). Based on these theoretical findings, it would be reasonable to infer a Gaussian similarity function if a deterministic model is used to uncover the underlying form in the presence of perceptual noise. Thus, although subjects may actually use an exponential decay similarity function in making decisions, this function may not be uncovered by using a deterministic modelling approach to the data.

It has also been shown (Ennis, 1988a) that if subjects actually use an exponential decay similarity function and the city-block metric *within a trial* to determine d, then the Euclidean metric distance between *stimulus means*, δ, is at least as satisfactory in relating similarity to distance as the city-block metric. Thus, the two hidden components - the form of the similarity function and the choice of metric - may not be revealed by employing a deterministic model in the presence of momentary fluctuations or noise. Further comments on this issue are in Nosofsky (1988) and Shepard (1988).

Nosofsky (1985) used the Shepard-Luce choice rule, Equation 10, to fit the absolute identification performance data of Kornbrot (1978) on tones. Both the exponential decay and the Gaussian similarity functions were used to model the data. The latter yielded a significantly better fit than the former for a neutral condition and a payoff biased condition. Equation 16, a probabilistic identification model, was used with a constant variance for stimuli and memory distributions of 1.0 to fit the same data. The exponential decay function fit the data at least as well the deterministic or probabilistic Gaussian models in the neutral condition (based on minimum chi-square and nonlinear least squares fits). In the payoff biased condition, the exponential decay probabilistic model fit significantly better than its deterministic counterpart, but not as well as the Gaussian probabilistic or deterministic models. However, all of the models were comparable when distance or squared distance in the similarity function was multiplied by a rate decreasing parameter [i.e. c in $g(d) = \exp(-cd^{\alpha})$]. This modelling provides further support for the idea that when the exponential decay similarity function is operative, it may require an appropriate probabilistic model to reveal it.

Multivariate Parameter Estimation

In order to make use of any of the models given earlier in fitting a particular data set, it is necessary to have an efficient, reliable procedure for parameter estimation. The parameters of interest in the probabilistic models are: γ, the metric-defining parameter; α, the exponent in the similarity function; $\underset{\sim}{\mu}_i$ and Σ_i, the vector of means and the variance-covariance matrix, respectively, of the stimulus or memory distributions. Hypotheses concerning γ and α can be tested by specifying particular values of these parameters (for example, 1 or 2) rather than allowing them to vary freely. This approach reduces the problem to one of estimating means and variance-covariance matrices under assumptions concerning the metric and form of the similarity function. There are a number of approaches to solving the estimation problem, one of which is to use the Levenberg-Marquardt algorithm for nonlinear least squares estimation (Dennis and Schnabel, 1983). Let a be a vector containing the parameters to be estimated.

Define

$$q_{ij}(\mathbf{a}) = p_{ij} - \mathrm{P}_{ij},$$

where p_{ij} is the observed proportion of judgments involving a comparison of stimulus S_i and stimulus S_j (this could be a "same-different" judgment or an identification error in which the response to presenting S_i is stimulus S_j). P_{ij} is the theoretical value obtained by solving the equation corresponding to the model being tested at the parameter values, **a**. Let **q(a)** be a vector with typical element q_{ij}(a). The value to be minimized is the residual sum of squares, $\mathbf{q}(\mathbf{a})^t\mathbf{q}(\mathbf{a})$. If $\mathbf{a}^0$ is an initial estimate of **a**, a series of approximations are computed as

$$\mathbf{a}^{n+1} = \mathbf{a}^{n} - [\alpha_n \mathbf{D}_n + \mathbf{J}_n^t \mathbf{J}_n]^{-1} \mathbf{J}_n^t \mathbf{q}(\mathbf{a}^n),$$

where $\mathbf{J}_n$ is the Jacobian matrix (matrix of partial derivatives) evaluated at $\mathbf{a}^n$, $\mathbf{D}_n$ is a diagonal matrix with entries equal to the diagonal of $\mathbf{J}_n^t\mathbf{J}_n$, and α_n is the Marquardt parameter, a positive constant. $\mathbf{J}_n$ is usually approximated using finite differences in double precision. The Marquardt parameter (α_n), initially 0.01, is quadrupled if the residual sum of squares increases from one iteration to the next and halved if it decreases.

Parameter Estimation using "Same-Different" Judgments

36 means and standard deviations were selected so that pairwise similarity values, P_{ij}, would be in the range 0.5 to 1.0. This selection ensured a high degree of distributional overlap. Figure 4 shows the 36 distributions with 2 standard deviation equal probability contours. One of the stimuli was assigned the mean vector $\mathbf{0}$. All correlation coefficients were assigned the value 0.0. Setting $\alpha = 2$ and $\gamma = 2$, the matrix of 666 similarity values (all stimulus pairs including self-comparisons) was obtained by solving Equation 7 for the selected means and standard errors. An iterative nonlinear least squares algorithm, as described above, was used to estimate the means and variance-covariance matrices from the matrix of paired simulated similarity values. A key to solving this problem and avoiding local minima is the generation of good initial estimates of the parameters.

An attempt to accurately recover all of the parameters for the 36 stimuli (all means and standard deviations) in one stage failed. Various strategies for estimating the parameters led to a successful analysis that was conducted in three stages. In the first stage, randomly generated values of the means were obtained and it was assumed that all standard deviations were 0.2 and that correlation coefficients were 0.0. The value of 0.2 for the standard deviations 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. Parameter estimates were obtained which minimized the residual sum of the squares,

$q(a)^t q(a)$. The parameter values at this minimum were then used as the starting configuration for a second stage in which all standard deviations were assumed to be equal across dimensions for a particular stimulus, but may vary across stimuli. The parameter estimates at the minimum from this stage were used as the starting configuration for the final stage in which the standard deviations were allowed to vary across both stimuli and dimensions. The residual sum of squares at this minimum was < 0.001.

The results of this analysis are given in Ennis, Palen and Mullen (1988) in which it is shown that estimates of the parameters differed from the original values only in the third decimal place. A plot of the recovered configuration was indistinguishable from a mirror image of Figure 4. It is interesting to note that the configuration obtained was a mirror image of the original configuration. Unlike traditional multidimensional scaling analysis based on deterministic models, the solution configuration from the type of probabilistic model used may not be invariant to rotation. This orientational uniqueness is a consequence of variance inequality.

An analysis of the Rothkopf (1957) Morse code data using Equation 7 to model the same-different judgments is given in Ennis, Palen and Mullen (1988). It was pointed out in that paper that 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 also provided this kind of interpretation of self-similarity. This means that P_{ii} will be closer to 1.0 when the variance is small than when it is large. In the Rothkopf Morse code data, many of the same-different judgment proportions for pairs of identical stimuli differ from each other and many are less than 1.0. In a deterministic model in which the proportions depend on δ, the probability for self-similarity must always be 1.0 when $\delta = 0$ and a judgment function of the form, $g(\delta) = \exp(-\delta^{\alpha})$, is used. This is not the case for a probabilistic model, such as that given in Equations 6 and 7. In these models, the within-trial distance, d, will almost always be different from zero when δ is zero because psychological magnitudes for a particular stimulus will vary from moment to moment.

Using the closed form, Equation 7, where $\alpha = 2$ and $\gamma = 2$, the Rothkopf Morse code data was modelled for the cases of equal and unequal variance within a stimulus distribution. 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 and the fit improvement was not significant at $p < 0.05$. Ennis, Palen and Mullen (1988) plotted the relationship between the size of the standard error for a stimulus distribution and 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). The size of the standard error was shown to decrease with increasing degree of isolation of a stimulus from the others in the ensemble. Stimuli which were located in close proximity to a large number of other stimuli had, therefore, distributions with the largest variance. Shepard's (1963) nonmetric multidimensional scaling analysis of the same data yielded a configuration of points which were quite similar to the location of the means of the distributions from probabilistic modelling. Certainly, the same interpretation of the dimensions (number of signal components and the dots/dashes ratio) would have resulted from both analyses. This result is consistent with the fact that 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.

An analysis of the Rothkopf data when $\alpha = 1$ and $\gamma = 1$ (the exponential decay, city-block metric model) using Equation 6, solved numerically, resulted in a configuration with a lower residual sum of squares than that obtained when $\alpha = 2$ and $\gamma = 2$ ($p < 0.01$). Some problems still remain to be solved with respect to local minima and computing efficiency for this case, and these will be taken up in a future paper. Based on the previous discussion concerning the judgment function and metric, the $\alpha = 1$ and $\gamma = 1$ case should fit the data best if the perceptual dimensions are separable. If the dimensions are integral, one would expect γ to be 2.

Fechner-Thurstone Models

In all of the models discussed so far, momentary fluctuations have been assumed to occur in the psychological magnitudes. Where specific probability density functions have been considered, attention has been restricted to multivariate normal functions, although Equation 1 is not necessarily restricted to this model. It is valuable to attempt to model performance in psychological tasks in terms of both psychological *and* physicochemical parameters because such models are likely to shed light on the material basis for perceptions. From the standpoint of modelling efficiency, it can be shown also that in many cases this type of model requires fewer parameters than models that ignore physicochemical parameters. Ennis and Mullen (1990) recently developed a general structure for connecting Fechnerian models with Thurstonian models. The purpose of this section is to introduce some of these ideas and to point the discussion in this chapter beyond probabilistic models which have all been based on Thurstonian concepts.

Assume that there are a number of objects with a common attribute that can be measured on a single physicochemical continuum. (A generalization to many continua will be mentioned later.) A stimulus magnitude, φ_i, is the value on this continuum for a particular stimulus object. Let $f_1(\varphi_i)$ be some probability density function of φ_i. Following a psychophysical transformation, the stimulus magnitude is represented mentally by a psychological magnitude, ψ_i. Let $g(\varphi_i)$ be any *one-to-one* function of φ_i which can operate on the entire domain of f_1. If any psychophysical transformation that connects physicochemical measures to mental representations is called a *Fechnerian* function, then g is such a function. Since g is one-to-one, it is invertible because for $g(\varphi_1) = g(\varphi_2)$, the only solution is $\varphi_1 = \varphi_2$. All monotonic functions are one-to-one. The pdf of ψ_i, $h_1(\psi_i)$, is

$$f_1 \circ g^{-1}(\psi_i)\, |dg^{-1}(\psi_i)/d\psi_i|,$$

where $f_1 \circ g^{-1}(\psi_i)$ is a composition function. If variation inherent in the biological system of the organism was zero, then each time an object with a particular physicochemical value was presented, it would be represented mentally by exactly the same psychological magnitude after the action of the transduction mechanism. However, it is more realistic to imagine that this value is a parameter, such as the mean, of a distribution of psychological magnitudes. The psychological magnitude, ψ_i, is then, a parameter in a probability density function of momentary psychological magnitudes, $x_i|\psi_i$, which may occur if central or peripheral noise is present, for instance. Let $f_2(x_i|\psi_i)$ be any pdf of $x_i|\psi_i$. Since the pdfs of ψ_i and $x_i|\psi_i$ are h_1 and f_2, the pdf of x_i is

$$\int_D f_2(x_i|\psi_i)h_1(\psi_i)d\psi_i \qquad (17)$$

where D is the domain of h_1.

The function f_1 may be a multivariate pdf of a vector of physicochemical measures instead of the single variable, φ_i. The function g , a one-to-one function, would map vectors from the domain space of f_1 to either single values (if g is a real-valued function) or vectors (if g is a vector-valued function) in the range space of g. Since g is a one-to-one function, g would have an inverse. Similar steps to those taken in deriving Equation 17 can be taken to derive the pdf of $\mathbf{x}_i$ (the vector equivalent of x_i, the momentary psychological magnitude). If z is some function of the momentary values in a trial, in terms of Equation 1, its pdf would be f. Returning to Equation 1, models for a great many psychological measures and tasks can be derived, such as similarity and identification performance. However, unlike many of the *specific* models discussed in this chapter, very special requirements for f_1, f_2, and g would be needed to ensure a multivariate normal pdf for the momentary psychological magnitudes, x_i. Ennis and Mullen (1990) discuss one such case leading to a unidimensional normal pdf.

In the previous models discussed in this chapter, no attention was given to stimulus parameters or to processes which might have led to the percepts. In this sense, the probabilistic models already discussed are much more restrictive than the Fechner-Thurstone models. These latter models, however, require a great deal of work before they can be put to use in modelling experimental results.

Some Computing Notes

Integral expressions such as Equations 6, 9, 13, and 16 were evaluated numerically on both Gould 32/97 and Trace Multiflow computers. An adaptive routine by Genz and Malik (1980) was found to be useful for multiple integration. A check for gross errors in the numerical computations was achieved by conducting large scale (100,000 trials per estimate) Monte Carlo evaluations.

In some cases, such as the evaluation of Equation 13, significant savings in computer time can be achieved by using Cholesky factorization to avoid the need to compute $f(\mathbf{z})$. Taking the standard multivariate normal pdf, a selection of values on each dimension (for example, the median of equal probability intervals) can be taken and converted to values in probabilistically equivalent intervals from the multivariate normal pdf of interest. This is achieved by computing $\mathbf{z} = \mathbf{L}\mathbf{y} + \underset{\sim}{\mu}$, where $\mathbf{L}$ is a lower triangular matrix such that $\mathbf{L}\mathbf{L}^t = \Sigma$, $\mathbf{y}$

is a vector with a standard multivariate normal pdf, and $\underset{\sim}{\mu}$ is the mean of the desired pdf. Since the interval bounding each **z** is probabilistically equivalent to each corresponding interval bounding **y**, it is only necessary to compute the probability contents of these intervals once. Numerical integration of $f(\mathbf{z})g(\mathbf{z})$ then becomes a dot product operation with a constant vector of probability weights [a vector computed from $f(\mathbf{z})$ at particular values of **z**] for all values of Σ and $\underset{\sim}{\mu}$. These values can be computed once, stored and reused as needed. It is still necessary, of course, to compute $g(\mathbf{z})$ for all values of **z**. In some cases, this approach to numerical integration can lead to significant improvements in computing speed compared with adaptive numerical integration of the entire function.

Concluding Remarks

Changes occur continuously in the physical and chemical properties of stimuli in the world in which we live. Our biological transduction and information processing systems do not remain static from moment to moment either. Although universal principles governing acquisition of information, judgment and behavior may exist, they may not be revealed by using deterministic or static models of these processes. One approach to modelling fluctuations or changes in psychological magnitudes from moment to moment is to treat the information-containing vectors probabilistically. A general approach to thinking about and formulating probabilistic models for many psychological tasks was presented in this chapter with specific applications to similarity and identification performance. It was shown that probabilistic models may be useful in revealing the nature of the judgment function and the distance metric when stimuli are perceptually similar and this led to a resolution of a paradox created when a deterministic model had been used to model identification performance. As a self-criticism, it must be pointed out that the probabilistic models chosen are somewhat arbitrary. A framework for building models with more concrete connections to physicochemical measures, however, has been sketched in the Fechner-Thurstone models.

REFERENCES

Ashby, F.G. (1988). Estimating the parameters of multivariate signal detection theory from simultaneous ratings on separate stimulus components. *Perception and Psychophysics*, **44**, 195-204.

Ashby, F.G., & Gott, R.E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **14**, 33-53.

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

Ashby, F.G., & Townsend, J.T. (1986). Varieties of perceptual independence. *Psychological Review*, **93**, 154-179.

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

Dennis Jr., J.E. and Schnabel, R.B. (1983) *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, New Jersey, pp. 221-228.

Ennis, D.M. (1988a). Confusable and discriminable stimuli: comments on Nosofsky (1986) and Shepard (1986). *Journal of Experimental Psychology, General*, **117**, 408-411.

Ennis, D.M. (1988b). Technical comment : Toward a universal law of generalization. *Science*, **242**, 944. [Comment on Shepard, R.N. (1987). Toward a universal law of generalization for psychological science. *Science*, **237**, 1317-1323.]

Ennis, D.M., & Ashby, F.G. (1990). A comparison of selected probabilistic models of identification with respect to perceptual dependence. Submitted.

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

Ennis, D.M., & Mullen, K. (1990). Fechner-Thurstone models. Submitted.

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

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

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.

Kornbrot, D.E. (1978). Theoretical and empirical comparison of Luce's choice model and logistic Thurstone model of categorical judgment. *Perception & Psychophysics,* **24**, *193-208.*

Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207-238.

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

Mullen, K. & Ennis, D.M. (1991). A simple multivariate probabilistic model for preferential and triadic choices. *Psychometrika, in press.*

Mullen, K., Ennis, D.M., de Doncker, E., & Kapenga, J.A. (1988). Multivariate models for the triangular and duo-trio methods. *Biometrics*, **44**, 1169-1175.

Nosofsky, R.M. (1985). Luce's choice model and Thurstone's categorical judgment model compared: Kornbrot's data revisited. *Perception & Psychophysics*, **37**, 89-91.

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

Nosofsky, R.M. (1988). On exemplar-based exemplar representations: Comment on Ennis (1988). *Journal of Experimental Psychology, General*, **117**, 412-414.

Nosofsky, R.M. (1990). Relations between exemplar-similarity and likelihood models of classification. *Journal of Mathematical Psychology, in press.*

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

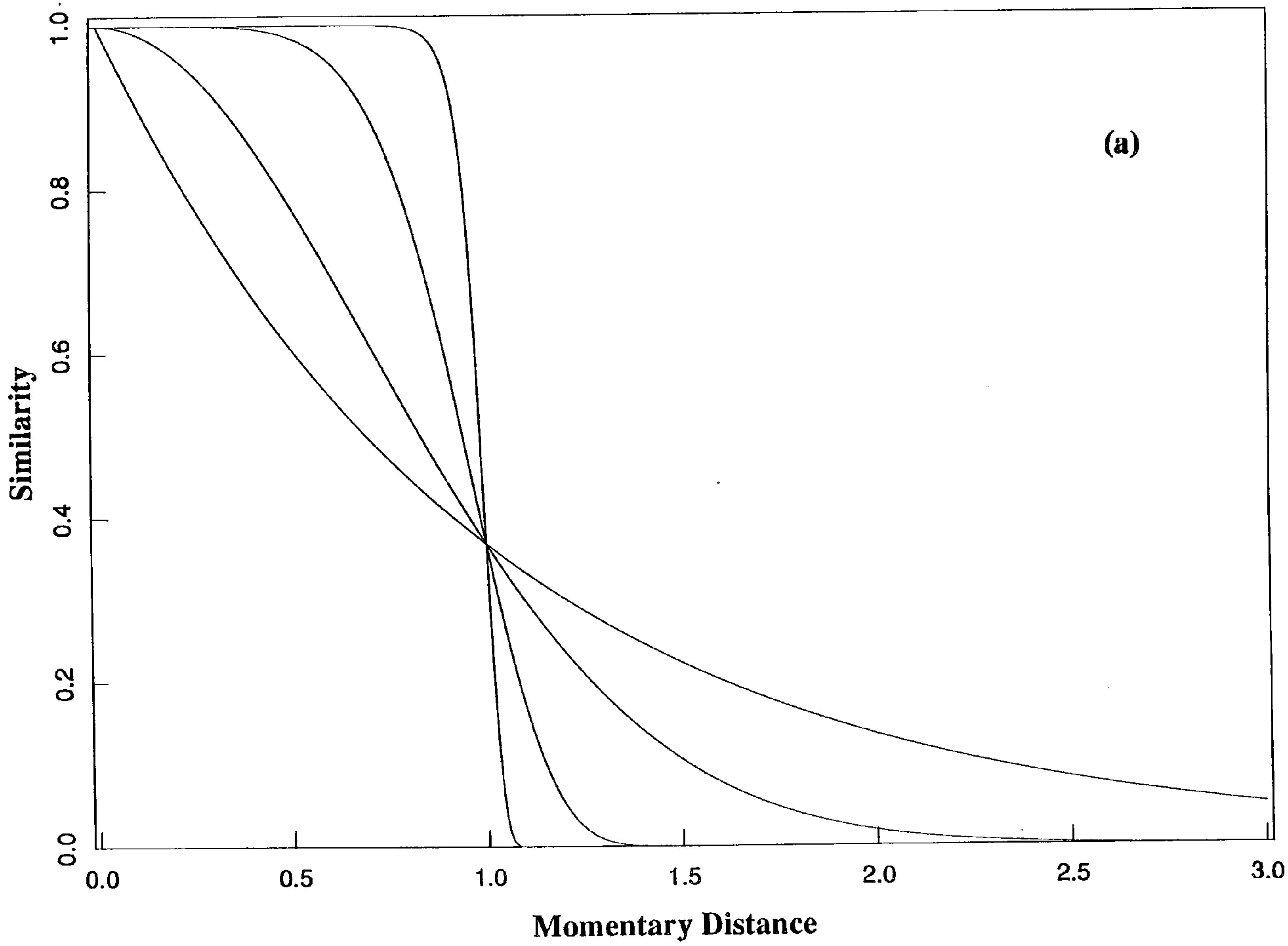
Shepard, R.N. (1987). Toward a universal law of generalization for psychological science. *Science*, **237**, 1317-1323.

Shepard, R.N. (1988). Time and distance in generalization and discrimination: Comment on Ennis (1988). *Journal of Experimental Psychology, General*, **117**, 415-416.

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

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

Figure 1. The effect of the form of g, the judgment function, on the sensitivity of its expected value, *similarity*, to perceptual dependence. (a) Four judgments of the form, $g(d) = exp(-d^{\alpha})$, where $\alpha = 1, 2, 6$ and 25. (b) The difference between the similarities of two pairs of distributions: In the first pair, correlation correlations between dimensions are both -0.8; in the second pair, correlation coefficients between dimensions are both +0.8. In both cases, $\delta = 1.0$. The curve with the higher asymptote corresponds to distributions with a standard deviation of 0.2, the lower curve corresponds to distributions with a standard deviation of 1.0.



(a)
Similarity
1.0
0.8
0.6
0.4
0.2
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Momentary Distance

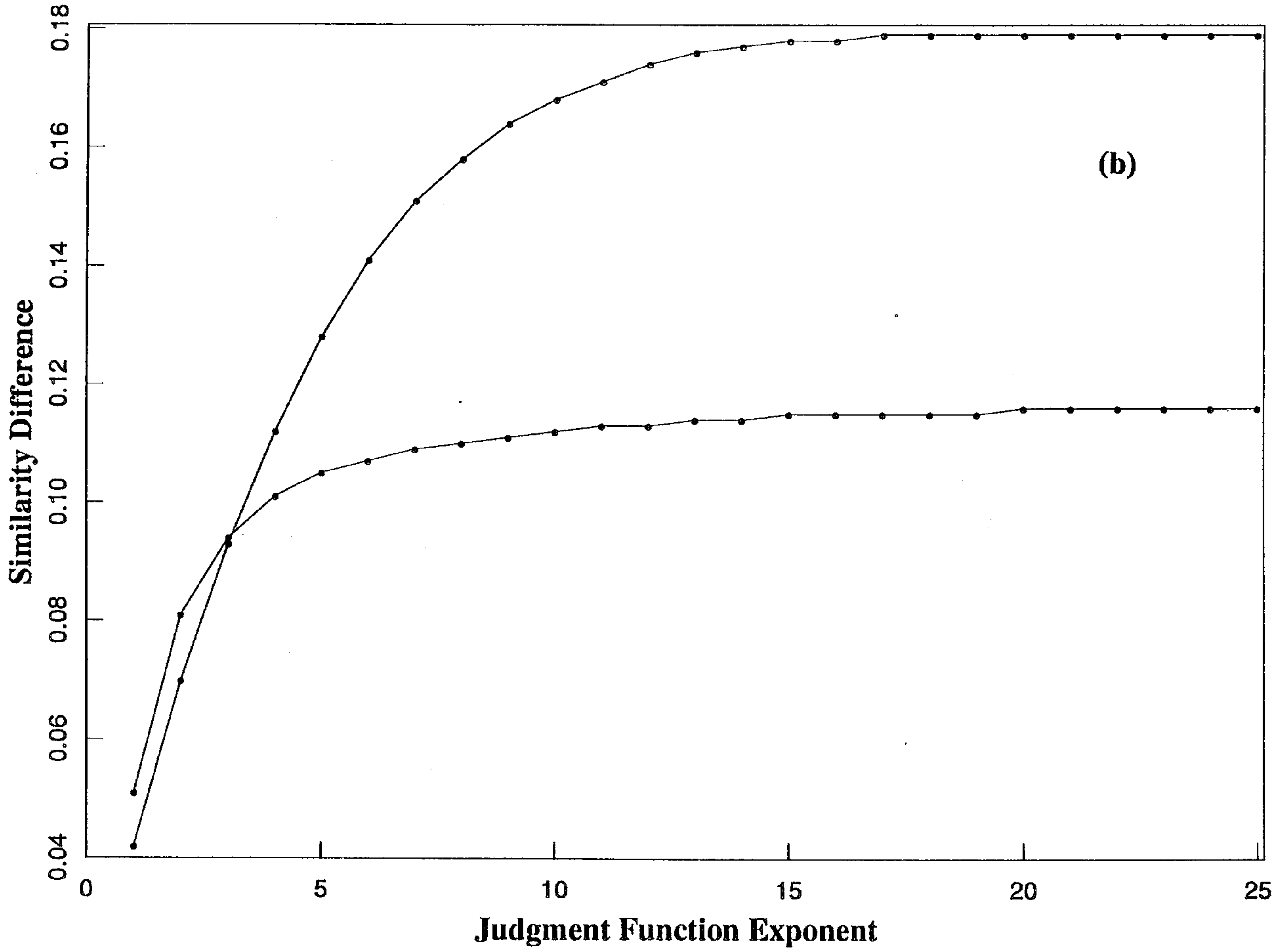

(b)
Similarity Difference
0.04
0.06
0.08
0.10
0.12
0.14
0.16
0.18
Judgment Function Exponent
0
5
10
15
20
25

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

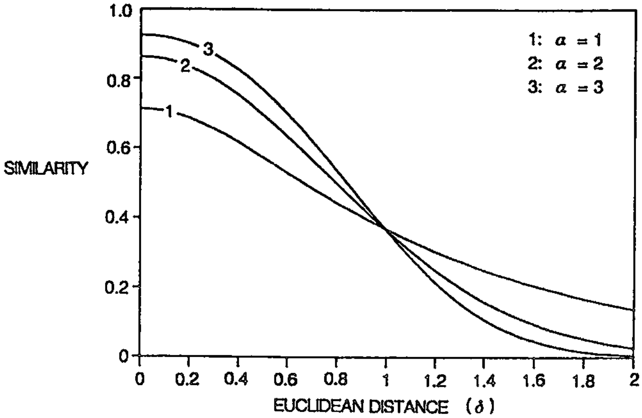
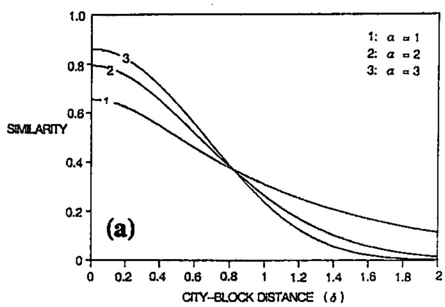

1: α = 1
2: α = 2
3: α = 3
1.0
0.8
0.6
0.4
0.2
0
SIMILARITY
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2
EUCLIDEAN DISTANCE (δ)
1
2
3

Figure 3. Expected value of similarity as a function of the city-block distance between the means of the distributions of psychological magnitudes for values of α 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.


1.0
0.8
0.6
0.4
0.2
0
SIMILARITY
1: α = 1
2: α = 2
3: α = 3
(a)
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2
CITY-BLOCK DISTANCE (δ)

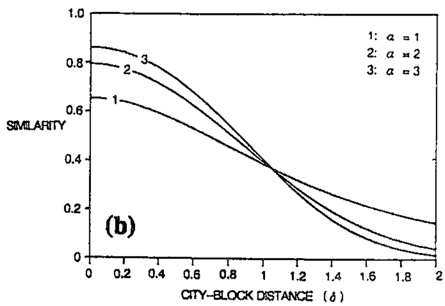
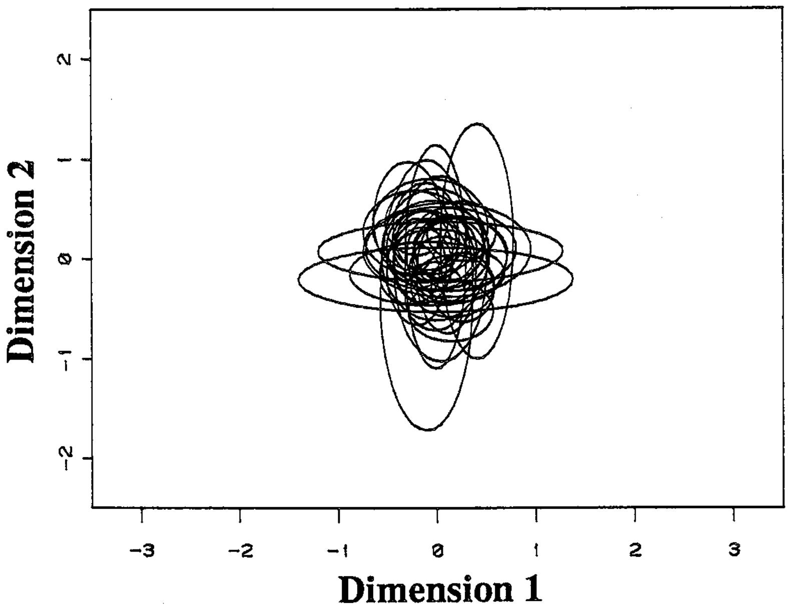

1.0
0.8
0.6
0.4
0.2
0
SIMILARITY
1: α = 1
2: α = 2
3: α = 3
(b)
0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6 1.8 2
CITY-BLOCK DISTANCE (δ)

Figure 4. Thirty six highly overlapping distributions with different variance-covariance matrices. Each distribution is represented by its 2 standard deviation equal probability contour.



Dimension 2
2
1
0
-1
-2
-3
-2
-1
0
1
2
3
Dimension 1

A General Probabilistic Model for Triad Discrimination, Preferential Choice and Two-Alternative Identification

Daniel M. Ennis

Philip Morris Research Center

Kenneth Mullen

Department of Mathematics and Statistics

University of Guelph

From a mathematical modelling viewpoint, there are very close parallels between triad discrimination, preferential choice and two-alternative identification under certain assumptions concerning the decision rules employed. The purpose of this chapter is to introduce a very general probabilistic model in a computationally simple form which can be used to model results obtained from several different types of psychological tasks. Chapter 16 contains a discussion of a number of probabilistic models of identification. One of these models, based on ordinal decision rules, will be covered in this chapter.

It might be useful to begin by providing a general overview of tasks involving three alternatives. First consider the situation in which all three alternatives are stimuli. Depending on the instructions, these tasks are variants of the method of triads. There are two methods which have been commonly discussed in the literature. Torgerson (1958) refers to one of them as the "complete method of triads", which we will call Torgerson's method of triads and the other as "Richardson's method of triadic combinations" (1938), which we will call Richardson's method of triads. In Torgerson's method of triads, the three objects are presented to the subject in three independent trials with the instruction to select one of the two objects most similar to the third. The third object is a different one of the three objects in the three trials. For instance, in the first trial, the subject's task might be to select which of S_j or S_k is most similar to S_i. ${}_iP_{jk}$ is the probability that S_i is more similar to S_j than S_k. The three trials are independent and may give rise to different psychological magnitudes from trial to trial for the same stimulus object. Richardson's method of triads involves a single presentation of the three objects and the subject's task is to judge which two objects are most alike perceptually and which two are most different. It is important to point out that the psychological magnitudes are assumed to remain at *fixed values* during a trial for both Torgerson's and Richardson's methods. If this is not the case (which is not unlikely for Richardson's method because two decisions are required per trial), the same two objects may appear to be most alike *and* most different. A model for this kind of result has been derived and is given in Ennis, Mullen, Frijters and Tindall (1989). A very common practice in the sensory evaluation of foods and beverages is to present three stimuli, two of which are (presumptively) physically identical. Three methods are commonly used: the duo-trio method, the ABX method and the triangular method. In the duo-trio method, one of the two "identical" stimuli is chosen as a standard and the subject's task is to decide which of the other two stimuli is most like the standard. In the ABX method, the two "different" stimuli are chosen as standards (A and B) and the task is to pick the standard which is most like the third stimulus (X). Both the duo-trio method and the ABX method are variants of Torgerson's method of triads. The third method, called the triangular method, involves the selection of "the most different" stimulus and is a special case of Richardson's method of triads.

The three alternatives just discussed do not have to be mental representations of physical stimuli presented during a trial. They could be memory representations of stimuli previously formed by subjects following presentation of the stimuli at a different time.

They could also be representations of objects or concepts that the subject never experienced materially. For example, one alternative might be an ideal point or ideal point distribution. Tasks which make use of the idea of memory representations occur in identification and categorization experiments. The concept of the ideal point is used in preference modelling. In a two-alternative identification experiment, a subject might be assumed to compare a stimulus representation with two memory representations and to "identify" the stimulus by naming the memory representation most similar to the stimulus. A model for paired preference might be based on the assumption that the preferred stimulus, of two, is the one which is most similar to an ideal point.

Due to the large number of applications for a viable model involving the comparison of three alternatives, a computationally tractable form would seem to be highly desirable. This model could then unify a large number of special cases dealing with a variety of experimental methods under one umbrella.

In a previous paper (Ennis, Palen & Mullen, 1988), it was mentioned that the wandering ideal point (WIP) model (De Soete, Carroll & DeSarbo, 1986) is closely related to a Thurstonian variant of Torgerson's method of triads (Ennis, Mullen & Frijters, 1988). The WIP model is a probabilistic interpretation of Coombs' (1964) preference unfolding model. In the WIP model one assumes that the stimulus points are fixed and that the ideal point "wanders" by supposing that momentary ideal point values within each trial in a preference task are drawn at random from a multivariate normal distribution. The preference decision depends on the Euclidean distances between the ideal point value and the stimulus values. The subject chooses the stimulus closest (smallest Euclidean distance) to the ideal point when making a preference judgment. In a different context, Hefner (1958) had postulated multivariate normal distributions (with identity variance-covariance matrices) for momentary psychological values for stimuli. If, unlike the WIP model, it is assumed that psychological values for both the stimuli and the ideal point are drawn from multivariate normal distributions (for which the variance-covariance matrix need not be the identity matrix), then the resulting choice model is more general than the WIP model and also more general with regard to the stimulus distributions than Hefner's model or the model of Zinnes and Griggs (1974), where it was assumed that the stimulus distribution variances are equal.

Consider the decision process specified in Torgerson's method of triads. The subject is presented with three stimulus objects, S_i, S_j and S_k and given S_i, for instance, is asked to decide which of S_j or S_k is most like S_i. If S_i is assumed to be the ideal point, then the probability of selecting S_j instead of S_k is the probability of preferring S_j over S_k. This decision is based on the Euclidean distances between the momentary values. An equation for computing the probability of this event, called $P_{i\,jk}$, is given in Ennis, Mullen and Frijters (1988) for the unidimensional case. In an identification experiment, S_i might represent a probe and S_j and S_k might represent the memory of stimuli S_j and S_k that have been established through training. An identification decision based on the distance between momentary values corresponding to S_i and S_j, and S_i and S_k could be modelled in exactly the same way as triad discrimination and preferential choice. Consequently, these three psychological tasks - triad discrimination, preferential choice and two-alternative identification - have exactly the same mathematical model under appropriate assumptions.

Model Assumptions

Multidimensional tri-stimulus models for the duo-trio and triangular methods have recently been developed (Ennis & Mullen, 1986a,b; Mullen & Ennis, 1987; Kapenga, de Doncker, Mullen & Ennis, 1987; and Mullen, Ennis, de Doncker & Kapenga, 1988). A useful starting point in developing a multidimensional extension of Torgerson's method of triads is to begin with a special case, the duo-trio method.

In the duo-trio method, one assumes that the distributions for two of the stimuli are identical. For instance, the distributions corresponding to S_i and S_j might be identical and one of these stimuli, say S_i, is designated as a standard. The subject's task is to decide which of S_j or S_k is most like S_i. The means of the distributions corresponding to the three stimuli are μ_i, μ_j and μ_k and $\mu_i = \mu_j$; their variance-covariance matrices are Σ_i, Σ_j and Σ_k and $\Sigma_i = \Sigma_j$. If $\mathbf{x}_i$ and $\mathbf{x}_j$ are the momentary psychological values corresponding to two stimuli, S_i and S_j, and $\mathbf{x}_k$ represents the corresponding magnitudes for a third stimulus, S_k, then in a particular trial, a correct overt response (with probability, ${}_iP_{jk}$) will be obtained if

(i) $|\mathbf{x}_i - \mathbf{x}_j| < |\mathbf{x}_i - \mathbf{x}_k|$ if S_i is the standard, or

(ii) $|\mathbf{x}_i - \mathbf{x}_j| < |\mathbf{x}_j - \mathbf{x}_k|$ if S_j is the standard.

Cases (i) and (ii) give identical results, so the following discussion will focus on case (i).

Mathematical Forms

Let $(\mathbf{x}_i - \mathbf{x}_j) = \mathbf{u}$ and $(\mathbf{x}_i - \mathbf{x}_k) = \mathbf{v}$. As shown in Mullen and Ennis (1987), ${}_iP_{jk}$ corresponds to the probability density content of the hypervolume inside the n-dimensional hypersphere $|\mathbf{u}| = R$ centered at 0, (where $R = |\mathbf{v}|$), or

$$ {}_i\mathrm{P}_{jk} = \int_C f(\mathbf{u},\mathbf{v})\, d(\mathbf{u})d(\mathbf{v}) \qquad (1) $$

where C is the region for which $|\mathbf{u}| < |\mathbf{v}|$

and

$$ f(\mathbf{u},\mathbf{v}) = \frac{\exp\{-0.5(\mathbf{z}-\boldsymbol{\mu})'\ \Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu})\}}{(2\pi)^n\ |\Sigma|^{1/2}}, $$

$$\mathbf{z} = (\mathbf{u},\mathbf{v}),$$

$$\boldsymbol{\mu} = [(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j), (\boldsymbol{\mu}_i - \boldsymbol{\mu}_k)],$$

$$\Sigma = \begin{bmatrix} \Sigma_1 & \Sigma_2 \\ \Sigma_2 & \Sigma_3 \end{bmatrix}, \text{ with}$$

$$\Sigma_1 = \Sigma_i + \Sigma_j,\ \Sigma_2 = \Sigma_i \text{ and } \Sigma_3 = \Sigma_i + \Sigma_k.$$

When Equation 1 is used to model the duo-trio method, $\boldsymbol{\mu}_i = \boldsymbol{\mu}_j$ and $\Sigma_i = \Sigma_j$. However, this does not have to be the case. If we assume that the distributions corresponding to S_i and S_j are *not* the same, then Equation 1 models the multidimensional Thurstonian variant of Torgerson's method of triads. Recalling that the probability of selecting S_j instead of S_k as the stimulus most similar to S_i is identical to the probability of preferring S_j to S_k if S_i represents the ideal point, it can be seen that Equation 1 also predicts preference probabilities. Note also that if $\boldsymbol{\mu}_i$ and Σ_i are the mean vector and variance-covariance matrix of the probe's distribution, then Equation 1 is also an identification model with $\boldsymbol{\mu}_j$, $\boldsymbol{\mu}_k$, Σ_j and Σ_k representing the memory distributions.

Equation 1 can be transformed to constant limits of integration as shown in Mullen, Ennis, de Doncker and Kapenga (1988), but still requires the evaluation of a $2n$-fold integral. The $2n$-fold integral given in Equation 1 can be simplified significantly by defining the integral in terms of an indefinite quadratic form (Mullen & Ennis, 1991) which leads to a single integral, irrespective of the dimensionality of the vector space of psychological magnitudes. From Equation (1),

$$P_{ijk} = \int_{C<0} \frac{\exp\{-0.5(\mathbf{z}-\boldsymbol{\mu})^t \ \Sigma^{-1}(\mathbf{z}-\boldsymbol{\mu})\}}{(2\pi)^n \ |\Sigma|^{1/2}} d\mathbf{z}.$$

$$C = \mathbf{u}\bullet\mathbf{u} - \mathbf{v}\bullet\mathbf{v}$$

$$= \mathbf{z}^t\mathbf{J}\mathbf{z},$$

where, $\mathbf{J} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{-I} \end{bmatrix}.$

The goal of the following linear algebra steps is to reduce Equation (1) to a standard canonical form (i.e. with a diagonal matrix in the limit of integration) for which there is a known computationally simple solution. This reduction to a standard form is achieved in two stages: (a) Cholsky factorization to transform the density function to a standard multivariate normal and, (b) diagonalization of the matrix in the quadratic form which defines one of the limits of integration.

The first step involves Cholesky factorization.

Let $\mathbf{M}$ be a non-singular lower triangular matrix defined by $\mathbf{MM}' = \Sigma$.

Let $\mathbf{z}^* = \mathbf{M}^{-1}(\mathbf{z}-\boldsymbol{\mu})$. Then,

$$ {}_{i}P_{jk} = \int_{D<0} \frac{\exp(-0.5 z^{*} \bullet z^{*})}{(2\pi)^{n}} dz^{*} $$

where

$$ D = (z^{*}+\underset{\sim}{\xi})^{t} M^{t} J M (z^{*}+\underset{\sim}{\xi}) $$

with

$$ \underset{\sim}{\xi} = M^{-1}(\underset{\sim}{\mu}) $$

The next stage achieves the desired diagonalization.

$$ \underset{\sim}{\omega} = P^{t}(z^{*}) $$

$$ \mathbf{d} = -P^{t}(\underset{\sim}{\xi}), $$

where **P** is the matrix of normalized eigenvectors of $\mathbf{M}^{t}\mathbf{JM}$.

$$ D = (\underset{\sim}{\omega} - \mathbf{d})^{t} P^{t} M^{t} J M P (\underset{\sim}{\omega} - \mathbf{d}) $$

$$ = (\underset{\sim}{\omega} - \mathbf{d})^{t} \Delta (\underset{\sim}{\omega} - \mathbf{d}). $$

$$ {}_{i}P_{jk} = \int_{D<0} \frac{\exp(-0.5 \underset{\sim}{\omega} \bullet \underset{\sim}{\omega})}{(2\pi)^{n}} d\underset{\sim}{\omega} $$

$$ = \Pr[\sum_{i=1}^{2n} \delta_{i}(\omega_{i} - d_{i})^{2} < 0] $$

$$ = \Pr[\sum_{i=1}^{r} \delta_{i} \chi_{m_{i}, d_{i}} < 0], $$

where,

δ_{i} are the eigenvalues of $\Sigma \begin{bmatrix} \mathbf{I} & 0 \\ 0 & -\mathbf{I} \end{bmatrix}$,

m_{i} are the degrees of freedom of the ith non-central chi-square,

d_i is the non-centrality parameter of the ith chi-square and

r is the number of distinct eigenvalues.

In this form, from the work of Imhof (1961), the probability of choosing S_j or S_k is

$$P_{i\,jk} = 0.5 - \frac{1}{\pi}\int_0^{\infty} \frac{\sin\theta(t)}{t\ \rho(t)}\,dt, \tag{2}$$

where,

$$\theta(t) = 0.5 \sum_{i=1}^{r} [m_i \tan^{-1}(\delta_i t) + d_i^2\ \delta_i t(1 + \delta_i^2 t^2)^{-1}],$$

$$\rho(t) = \left[\prod_{i=1}^{r} (1 + \delta_i^2 t^2)^{m_i/4}\right] exp\{0.5 \sum_{j=1}^{r} (d_j \delta_j t)^2 /(1 + \delta_j^2 t^2)\},$$

Relationships Among Triad and Preference Models

Figure 1 summarizes the relationships between the model formulated in Equation 1 and other triad and preference models. It can be seen that the model presented in this paper is the general case for any model that specifies that a subject will choose the psychological magnitude from two alternatives that has the smaller Euclidean distance to a third psychological magnitude when the distributions of all magnitudes are assumed to be multivariate normal. The model includes the following special cases: The Thurstonian variant of Torgerson's method of triads, the duo-trio method model, the wandering ideal point model, the preference model of Zinnes and Griggs (1974) and Coombs' (1964) preference unfolding model. Also included is the two-alternative identification model based on a comparison of distances between a probe and the memory representations.

Computing

Numerical integration and parameter estimation techniques for equations such as Equations 1 and 2 have been discussed in Mullen and Ennis (1987), Mullen, Ennis, de Doncker and Kapenga (1988), and Ennis, Palen and Mullen (1988). The numerical evaluation of Equation 1 is fairly computationally intense, partly because it involves the numerical evaluation of a 2n-fold integral, where n is the number of dimensions. A comparison of the computational efficiency of Equations 1 and 2, on a Gould 32/97 computer when $n = 2$, revealed that equation 2 could be computed 10^3 to 10^4 times faster (to the same accuracy) than adaptive numerical integration of Equation 1, depending on the specific parameters chosen. When n is larger than 2, the relative efficiency of Equation 2 should become even greater because the computing time will be directly proportional to n rather than a function involving n as an exponent.

Parameters of the distributions corresponding to S_i, S_j and S_k can be estimated using nonlinear least squares as outlined in chapter 16. It would be interesting to simultaneously fit triad and preference matrices using Equation 2 for an ensemble of stimuli.

Concluding Remarks

Triad discrimination, preferential choice and two-alternative identification share a common mathematical model. This model is based on the assumption that subjects make choices between stimuli or memory representations based on Euclidean distances between stimuli, stimuli and ideal points, or probes and memory representations. These ideas were presented in the context of a probabilistic model that assumes multivariate normal probability density functions for stimuli, ideal points or memory representations. A computationally simple form to compute decision probabilities has been given.

References

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

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

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

Ennis, D.M., Mullen, K., Frijters, J.E.R. & Tindall, J. (1989). Decision conflicts: Within trial resampling in Richardson's method of triads. *British Journal of Mathematical and Statistical Psychology*, **42**, 265-269.

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

Imhof, J.P. (1961). Computing the distribution of quadratic forms in normal variables. *Biometrika*, **48**, 419-426.

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 Keast, P. and Fairweather, G. (eds.), *Numerical Integration*, 321-328.

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

Mullen, K. & Ennis, D.M. (1991). A simple multivariate probabilistic model for preferential and triadic choices. *Psychometrika, in press.*

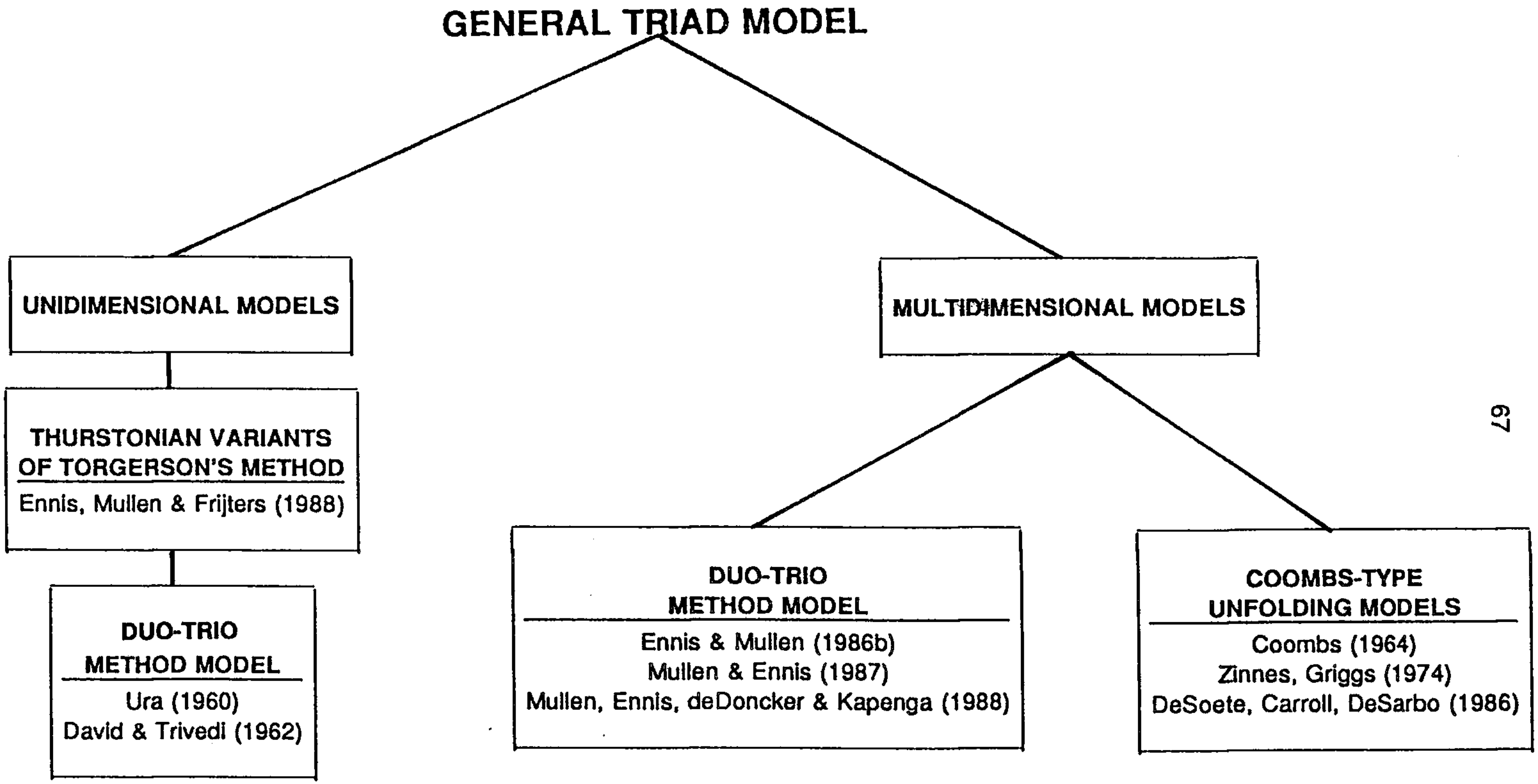
Mullen, K., Ennis, D.M., de Doncker, E., & Kapenga, J.A. (1988). Multivariate models for the triangular and duo-trio methods. *Biometrics*, **44**, 1169-1175.

Richardson, M.W. (1938). Multidimensional psychophysics. *Psychological Bulletin*, **35**, 659-660.

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

Zinnes, J.L. & Griggs, R.A. (1974). Probabilistic, multidimensional unfolding analysis. *Psychometrika*, **39**, 327-350.

Figure 1. The relationship among triad and preference choice models as special cases of the general triad model.


GENERAL TRIAD MODEL
UNIDIMENSIONAL MODELS
MULTIDIMENSIONAL MODELS
THURSTONIAN VARIANTS OF TORGERSON'S METHOD
Ennis, Mullen & Frijters (1988)
DUO-TRIO METHOD MODEL
Ura (1960)
David & Trivedi (1962)
DUO-TRIO METHOD MODEL
Ennis & Mullen (1986b)
Mullen & Ennis (1987)
Mullen, Ennis, deDoncker & Kapenga (1988)
COOMBS-TYPE UNFOLDING MODELS
Coombs (1964)
Zinnes, Griggs (1974)
DeSoete, Carroll, DeSarbo (1986)

CHAPTER 3

UNIDIMENSIONAL PROBABILISTIC MODELS

Variants of the method of triads: Unidimensional Thurstonian models. 1988. *British Journal of Mathematical and Statistical Psychology* **41**, 25-36.

Decision conflicts: Within-trial resampling in Richardson's method of triads. 1989. *British Journal of Mathematical and Statistical Psychology*, **42**, 265-269.

Fechner-Thurstone models. 1991. *Mathematical Social Sciences, submitted.*

British Journal of Mathematical and Statistical Psychology (1988), 41, 25–36 *Printed in Great Britain*

Variants of the method of triads: Unidimensional Thurstonian models

Daniel M. Ennis
Philip Morris Research Center, Commerce Road, Richmond, VA 23261, USA

Kenneth Mullen
Department of Mathematics and Statistics, University of Guelph, Canada

Jan E. R. Frijters
The Netherlands Agricultural University, Wageningen

This paper deals with a theory for two variants of the method of triads—Torgerson's and Richardson's methods. For both methods, the relationship between the decision probabilities and the parameters of the momentary psychological magnitudes is derived under the assumption that these magnitudes can be modelled as if they were drawn from normal distributions with particular means and variances. This approach differs from previous theoretical work on the methods of triads, where it has been assumed that distances between momentary psychological magnitudes are normally distributed. Cases where the variances of the distributions of psychological magnitudes are unequal can be handled by the models. It is shown that the duo–trio and triangular methods are special cases of Torgerson's and Richardson's methods of triads, respectively. Using non-linear least squares minimization, estimates of the means of the distributions of a sample problem are obtained for random samples of size 200. Decision conflicts, which may occur when using Richardson's method, are discussed.

1. Introduction

Since the work of Thurstone (1927), there has been a great deal of theoretical work dealing with unidimensional and multidimensional stochastic models for scaling psychological magnitudes. Generally, mental representations of physical or non-physical stimulus objects are modelled as if the momentary psychological magnitudes are random values from normal distributions. Having first derived the relationship between the behavioural response and the parameters of interest (means, variances, correlation coefficients), the scaling problem involves estimating these parameters for particular behavioural values obtained experimentally. There is a voluminous literature on this topic, particularly as it applies to two-stimulus methods such as the method of paired comparisons. With respect to three-stimulus methods, Torgerson

(1958) set out the unidimensional theory for two versions of the method of triads, a tri-stimulus procedure of some interest in the early development of multidimensional scaling. Unlike traditional assumptions formulated by Thurstone in this area, where it is assumed that the psychological magnitudes are randomly drawn from normal distributions. Torgerson assumed that the psychological distances between pairs of stimuli could be modelled as normal deviates. In this paper we deal with both versions of the method of triads from the more classical Thurstonian viewpoint and show how they are related to other tri-stimulus models. The models of interest focus on discrimination tasks involving decisions which depend only on the psychological magnitudes evoked in a particular triad, which would be the case for confusable stimuli, for instance.

2. Methods of triads—Methodology and assumptions

Let $\mathbf{S}_1, \mathbf{S}_2, \ldots, \mathbf{S}_n$ represent n sets of stimulus objects, and suppose that within each set the objects are physicochemically identical. Without using unique subscripts for each of the objects within each set, let $S_1, S_2, \ldots, S_n$ represent randomly selected stimulus objects from each of the n sets respectively. A triad is composed of three of the stimulus objects, S_i, S_j and S_k, $i \neq j \neq k \neq i$. Each object is assumed to be represented mentally by corresponding univariate psychological magnitudes, $x_1, x_2, \ldots, x_n$ which may vary from moment to moment but which are assumed to be drawn randomly from normal distributions with means $\mu_1, \mu_2, \ldots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$. The distributions of momentary psychological magnitudes are assumed to be independent.

From a methodological standpoint, there are two versions of the method of triads. Torgerson (1958) refers to one of them as the 'complete method of triads', which we will call Torgerson's method of triads and the other as 'Richardson's method of triadic combinations' (Richardson, 1938), which we will call Richardson's method of triads. In Torgerson's method of triads, the three objects are presented to the subject in three independent trials with the instruction to select one of the two objects most similar to the third. The third object is a different one of the three objects in the three trials. For instance, in the first trial, the subject's task might be to select which of S_j or S_k is most similar to S_i. Let ${}_iP_{jk}$ denote the probability that S_i is more similar to S_j than S_k. The three trials are independent and may give rise to different psychological magnitudes from trial to trial for the same stimulus object. Richardson's method of triads involves a single presentation of the three objects and the subject's task is to judge which two objects are most alike perceptually and which two are most different. It is important to point out that the psychological magnitudes are assumed to remain at *fixed values* during a trial for both Torgerson's and Richardson's methods. If this is not the case (not unlikely for Richardson's method because two decisions are required per trial), the same two objects may appear to be most alike *and* most different. If the subject is unaware of this possibility, the subject may develop strategies to cope with what would appear to be illogical responses. This problem will be dealt with in a later section of the paper.

3. Torgerson's method of triads

Let ${}_iP_{jk}$ represent the probability that S_i will be perceived to be more similar to S_j than

S_k. $x_1, x_2, \ldots, x_n$ are the momentary psychological values, $\mu_1, \mu_2, \ldots, \mu_n$ are their means and $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ are their variances. Thus, x_i, μ_i and σ_i^2 are associated with S_i for $i = 1, \ldots, n$.

$$ {}_iP_{jk} = \int_0^\infty \int_0^\infty f(r,s)\,\mathrm{d}r\,\mathrm{d}s + \int_{-\infty}^0 \int_{-\infty}^0 f(r,s)\,\mathrm{d}r\,\mathrm{d}s, \tag{1} $$

where

$$ r = (x_j - x_k) \qquad \text{and} \qquad s = (2x_i - x_j - x_k), $$

$$ f(r,s) = \frac{\exp\{-0.5[(\mathbf{z}-\boldsymbol{\mu})' V^{-1}(\mathbf{z}-\boldsymbol{\mu})]\}}{(2\pi)|\mathbf{V}|^{1/2}}. $$

$$ \mathbf{z}' = (r, s), $$

$$ \boldsymbol{\mu}' = [(\mu_j - \mu_k), (2\mu_i - \mu_j - \mu_k)], $$

and

$$ \mathbf{V} = \begin{bmatrix} \sigma_j^2 + \sigma_k^2 & \sigma_k^2 - \sigma_j^2 \\ \sigma_k^2 - \sigma_j^2 & 4\sigma_i^2 + \sigma_j^2 + \sigma_k^2 \end{bmatrix}. $$

A proof of equation (1) is in Appendix 1.

The complement of ${}_iP_{jk}$ (that S_i is more similar to S_k than S_j) is

$$ {}_iP_{kj} = 1 - {}_iP_{jk}. $$

In the case of three particular objects, S_1, S_2 and S_3, the three probabilities, ${}_1P_{23}$, ${}_2P_{13}$ and ${}_3P_{12}$ can be obtained using equation (1).

When $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_n^2 = \sigma^2$,

$$ {}_iP_{jk} = 1 - \Phi[(\mu_k - \mu_j)\sigma\sqrt{2}] - \Phi[(\mu_k + \mu_j - 2\mu_i)/\sigma\sqrt{6}] $$
$$ + 2\Phi[(\mu_k - \mu_j)/\sigma\sqrt{2}]\Phi[(\mu_k + \mu_j - 2\mu_i)/\sigma\sqrt{6}], $$

where $\Phi(x)$ represents the area under the normal curve from $-\infty$ to x.

4. Richardson's method of triads

In Richardson's method, the subject's task is to judge which two stimulus objects are most alike perceptually and which two are most different. Once again, assume that the three objects are S_i, S_j and S_k. The corresponding momentary psychological magnitudes are x_i, x_j and x_k in a particular trial. These values are assumed to be randomly drawn from normal distributions of means μ_i, μ_j and μ_k respectively and variances σ_i^2, σ_j^2 and σ_k^2. It is assumed that only one sample of x_i, x_j and x_k occurs during a trial. In a given trial, if S_i and S_j are reported to be most alike, represent this decision by a_{ij} and if they are most different by d_{ij}. In the case of three particular objects, S_1, S_2 and S_3, six different decision pairs are possible:

$$ a_{12}\,d_{13};\ a_{12}\,d_{23};\ a_{13}\,d_{12};\ a_{13}\,d_{23};\ a_{23}\,d_{12};\ a_{23}\,d_{13}. $$

Since we assume that resampling does not occur, the probability that the same pair of objects are most alike *and* most different is zero. A particular decision pair, such as $(a_{12}\, d_{13})$ will occur with a certain probability, $P(a_{12}\, d_{13})$.

Let $u = (x_i - x_j)$ and $w = (x_j - x_k)$.

$$P(a_{ij}\, d_{ik}) = P(0 < u < w) + P(w < u < 0)$$

$$= \int_0^{\infty} \int_u^{\infty} f(u, w)\, \mathrm{d}w\, \mathrm{d}u + \int_{-\infty}^{0} \int_{-\infty}^{u} f(u, w)\, \mathrm{d}w\, \mathrm{d}u, \tag{2}$$

where $f(u, w)$ is the bivariate normal distribution of u and w or

$$f(u, w) = \frac{\exp\{-0.5[(\mathbf{z}-\boldsymbol{\mu})' V^{-1}(\mathbf{z}-\boldsymbol{\mu})]\}}{(2\pi)|\mathbf{V}|^{1/2}},$$

where

$$\mathbf{z}' = (u, w),$$

$$\boldsymbol{\mu}' = [(\mu_i - \mu_j), (\mu_j - \mu_k)],$$

and

$$\mathbf{V} = \begin{bmatrix} \sigma_i^2 + \sigma_j^2 & -\sigma_j^2 \\ -\sigma_j^2 & \sigma_j^2 + \sigma_k^2 \end{bmatrix}.$$

When $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_n^2$, then

$$\mathbf{V} = \sigma^2 \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}.$$

A proof of equation (2) is in Appendix 2.

5. Computing the decision probabilities

Equation (1) was evaluated numerically using an adaptive routine for multiple integrals by Genz & Malik (1980). Equation (2) was evaluated using the IMSL (1984) Fortran-callable subroutine, DBLIN. These results were checked by comparing them to Monte Carlo simulation values obtained from 100 000 trials per estimate. This comparison showed that numerical evaluation of equations (1) and (2) gave results which agreed to at least second decimal place accuracy with the Monte Carlo results for an extensive range of parameter values.

6. Parameter estimation for four stimulus objects

Consider a problem in which there are four sets of stimulus objects, $\mathbf{S}_1$, $\mathbf{S}_2$, $\mathbf{S}_3$ and $\mathbf{S}_4$ (the objects are physicochemically identical within each set) and for which estimates are known for the 12 probabilities obtainable using Torgerson's method of triads. Let these estimates be ${}_1\mathcal{Q}_{23}$, ${}_2\mathcal{Q}_{13}$, ${}_3\mathcal{Q}_{12}$, ${}_1\mathcal{Q}_{24}$, ${}_2\mathcal{Q}_{14}$, ${}_4\mathcal{Q}_{12}$, ${}_1\mathcal{Q}_{34}$, ${}_3\mathcal{Q}_{14}$, ${}_4\mathcal{Q}_{13}$, ${}_2\mathcal{Q}_{34}$, ${}_3\mathcal{Q}_{24}$ and ${}_4\mathcal{Q}_{23}$. Assuming that the variances for the populations of momentary psychological values are equal and unity, it is of interest to find estimates of the means of the distributions of psychological magnitudes. This problem can be handled as a non-linear least squares minimization problem in which there are 12 functions [all of the general form given in equation (1)] and 3 parameters (μ_2, μ_3 and μ_4; setting $\mu_1 = 0.0$). The objective is to find estimates of these parameters which minimize the residual sum of squares

$$\sum_i \sum_j \sum_k ({}_iQ_{jk} - {}_iP_{jk})^2 \qquad \text{for } j < k \quad \text{and} \quad i \neq j;\ i \neq k.$$

There are several approaches to locating the minimum for problems of this kind. A modified Levenberg–Marquardt algorithm, available as an option in the IMSL (1984) subroutine ZXSSQ, was used.

Suppose that $\mu_1 = 0.0$, $\mu_2 = 1.0$, $\mu_3 = 2.0$, $\mu_4 = 3.0$ and that all variances are equal and unity. Table 1 gives the 12 decision probabilities for five independent samples of size 200 drawn at random from normal distributions with the appropriate parameters and also the 12 theoretical probabilities. Table 2 gives the estimates of μ_2, μ_3 and μ_4 (assuming that $\mu_1 = 0.0$) for the five samples and the residual sums of squares.

Incidentally, any one of the four means could have been assigned a fixed value (which also does not have to be zero). The means of the five samples (0.979, 1.981 and 3.017) appeared to be good estimates of the means of the sampled distributions (1.0, 2.0 and 3.0), although it should be pointed out that a detailed study of the parameter estimation problem involved was not undertaken.

If n is the number of stimulus objects, there will be $n(n-1)(n-2)/2$ equations and probability estimates, and $n-1$ unknown means. For $n = 6$, for instance, there are 60 equations and five means to be estimated. Since, in a problem of this size, there are considerably more observations than unknowns, it should be possible to find reliable parameter estimates using only a fraction of the possible decision probabilities. These designs and their evaluation will not be pursued here although it is clear that this

Table 1. Theoretical and sampled values of decision probabilities for Torgerson's method of triads when $\mu_1 = 0.0$, $\mu_2 = 1.0$, $\mu_3 = 2.0$, $\mu_4 = 3.0$ and all variances are 1.0. Theoretical probabilities are designated as ${}_iP_{jk}$ and samples probabilities ($n = 200$) are designated as ${}_iQ_{jk}$.

Decision probabilities											
${}_1P_{23}$	${}_2P_{13}$	${}_3P_{12}$	${}_1P_{24}$	${}_2P_{14}$	${}_4P_{12}$	${}_1P_{34}$	${}_3P_{14}$	${}_4P_{13}$	${}_2P_{34}$	${}_3P_{24}$	${}_4P_{23}$
0.703	0.500	0.297	0.878	0.653	0.250	0.750	0.347	0.122	0.703	0.500	0.297
${}_1Q_{23}$	${}_2Q_{13}$	${}_3Q_{12}$	${}_1Q_{24}$	${}_2Q_{14}$	${}_4Q_{12}$	${}_1Q_{34}$	${}_3Q_{14}$	${}_4Q_{13}$	${}_2Q_{34}$	${}_3Q_{24}$	${}_4Q_{23}$
0.725	0.480	0.330	0.860	0.695	0.250	0.745	0.320	0.100	0.715	0.505	0.210
0.700	0.485	0.300	0.905	0.560	0.215	0.720	0.360	0.100	0.685	0.560	0.285
0.705	0.545	0.290	0.885	0.690	0.255	0.710	0.365	0.095	0.710	0.495	0.290
0.715	0.495	0.295	0.885	0.635	0.290	0.800	0.305	0.165	0.720	0.525	0.290
0.680	0.515	0.350	0.905	0.640	0.325	0.770	0.400	0.110	0.730	0.525	0.340

Table 2. Estimated population means using non-linear least squares minimization applied to the decision probabilities from Torgerson's method of triads

Case	Residual	μ_1	μ_2	μ_3	μ_4
Sample 1	0.009	0.000	0.953	2.090	3.118
Sample 2	0.010	0.000	1.167	2.044	3.050
Sample 3	0.004	0.000	0.923	2.016	3.005
Sample 4	0.009	0.000	0.995	1.991	3.031
Sample 5	0.007	0.000	0.855	1.763	2.883
Means			0.979	1.981	3.017
Standard error			0.117	0.127	0.086

work will need to be done if the method of triads is to be used experimentally for a large number of objects ($n > 5$).

7. Relationships to other tri-stimulus procedures

David & Trivedi (1962) and Ura (1960) provided the unidimensional (equal variance) theoretical basis for the duo–trio and triangular methods. Frijters (1979*a*) compared the triangular method model with the 3-alternative forced-choice model [Green & Swets (1966)] and based the resolution of the paradox of 'discriminatory non-discriminators' in taste psychophysics on this theory (1979*b*).

Both the duo–trio and triangular method models are concerned with discrimination tasks involving three stimulus objects, two drawn at random from two sets of physicochemically identical objects and the other from a third set of objects. In each trial of the duo–trio method, the subject's task is to decide which of two objects is most similar to a third object (drawn from a set which is physicochemically identical to one of the first two sets). In the triangular method, the subject's task is to identify the most different object. Under the assumption that the variances of the momentary psychological values corresponding to the stimulus objects are equal and that μ_j and μ_k are their means, the probability of a correct response for the duo–trio (P_d) and triangular method (P_t) are as follows (from David & Trivedi):

$$P_d = 1 - \Phi[(\mu_k - \mu_j)/\sigma\sqrt{2}] - \Phi[(\mu_k - \mu_j)/\sigma\sqrt{6}] \\ + 2\Phi[(\mu_k - \mu_j)/\sigma\sqrt{2}]\Phi[(\mu_k - \mu_j)/\sigma\sqrt{6})] \tag{3}$$

and

$$P_t = 2\int_0^{\infty} \Phi[-z\sqrt{3} + (\mu_k - \mu_j)\sqrt{(2/3)}] \\ + \Phi[-z\sqrt{3} - (\mu_k - \mu_j)\sqrt{(2/3)}]\, d\Phi(z). \tag{4}$$

The duo–trio method model is a special case of the model for one of the trials in Torgerson's method of triads where $\mu_i = \mu_j$. This is clear from the methodologies and instructions involved and also from a comparison of results from evaluating equation (1) and equation (3).

In Richardson's method of triads, there are six probabilities corresponding to six possible decision outcomes. Considered in pairs, these are the probability that S_i and S_j are most alike; that S_i and S_k are most alike; and that S_j and S_k are most alike. These are also the probabilities that S_k, S_j and S_i are the most different objects, respectively. Two of these six probabilities are $P(a_{ij}, d_{ik})$ and $P(a_{ij}, d_{jk})$ which, when summed, and assuming that $\mu_i = \mu_j$, give a value identical to P_t in equation (4). The triangular method model, therefore, predicts a behavioural response probability which can also be predicted from two of the six probabilities of Richardson's method of triads.

Multivariate models for discrimination methods, including the triangular and duo–trio methods, have been derived and evaluated using Monte Carlo and numerical methods (Ennis & Mullen, 1985, 1986*a*, *b*; Mullen & Ennis, 1987). Since these methods are special cases of variants of the method of triads, the same rationale can be used to derive multivariate stochastic models for both methods of triads. These models would prove useful in the development of stochastic multidimensional scaling.

8. Resampling within a trial

Earlier in this paper, it was pointed out that the momentary psychological values were assumed to remain at fixed values during a trial for both methods of triads. For instance, in deciding whether S_j or S_k is most similar to S_i in Torgerson's method, it is assumed that the values x_j, x_k and x_i, though randomly drawn from their respective distributions, will be used to select S_j or S_k and that resampling within a trial will not occur. One could argue, for example, that since S_i is compared to S_j and S_k that the subject might obtain two independent psychological magnitudes corresponding to S_i. In comparing the two methods of triads, Richardson's method would appear to offer the greatest potential for within-trial resampling because the subject is required to make two quite separate decisions — which two objects are most alike and which two objects are most different. In Torgerson's method, even if resampling occurs, the two distances obtained, e.g. $|x_i - x_j|$ and $|x_i^* - x_k|$ (where x_i^* is an independent psychological magnitude corresponding to S_i) will be different, even if only infinitesimally. Theoretically, then, the subject has a basis for making a decision. However, in Richardson's method, it is entirely possible for the subject to conclude that S_i and S_j are most alike and that the same objects are also most different. If x_i, x_j and x_k are the psychological magnitudes used to decide which two objects are most alike and x_i^*, x_j^* and x_k^* are the magnitudes used to decide which two objects are most different, there is some probability that

$$|x_i - x_j| < |x_i - x_k| \qquad \text{and} \qquad |x_i - x_j| < |x_j - x_k|$$

and that

$$|x_i^* - x_j^*| > |x_i^* - x_k^*| \qquad \text{and} \qquad |x_i^* - x_j^*| > |x_j^* - x_k^*|.$$

If the subject is not aware that this outcome is possible, it would lead to what may be called a decision conflict for the subject. Represent its probability by P_{con}. For convenience, consider only the equal variance case. If μ_i is the smallest mean, μ_k is the largest, and $\mu_j = (\mu_k + \mu_i)/2$, P_{con} will vary depending on the size of $(\mu_k - \mu_i)$.

Figure 1 is a plot of the relationship between P_{con} and $(\mu_k - \mu_i)$, when $\mu_j = (\mu_k + \mu_i)/2$. P_{con} was obtained by Monte Carlo simulation using 20 000 trials per estimate. From Fig. 1 it can be seen that the probability of a decision conflict will be greatest in this case when the means of the psychological magnitudes are similar and is 1/3 when $(\mu_k - \mu_i) = 0$. When this difference is between 0 and 3, this problem will occur from 17 to 33 per cent of the time.

Although Richardson's method appears to provide some efficiency from an experimental point of view (one trial instead of three in Torgerson's method), this benefit must be offset by the possibility of creating a situation where the subject may be forced to deal with what would appear to be an illogical decision. Presumably, the subject would then invoke a strategy to deal with this problem which would lead to results that cannot be modelled by equation (2).

9. Conclusion

In discussing methods of triads, Torgerson (1958) assumed that the distances between momentary psychological magnitudes are normally distributed and then applied Thurstonian ideas to distances instead of psychological magnitudes (discriminal processes). This consequently led to a procedure for estimating the coordinates of points corresponding to the stimuli in a multidimensional space. In this paper, we

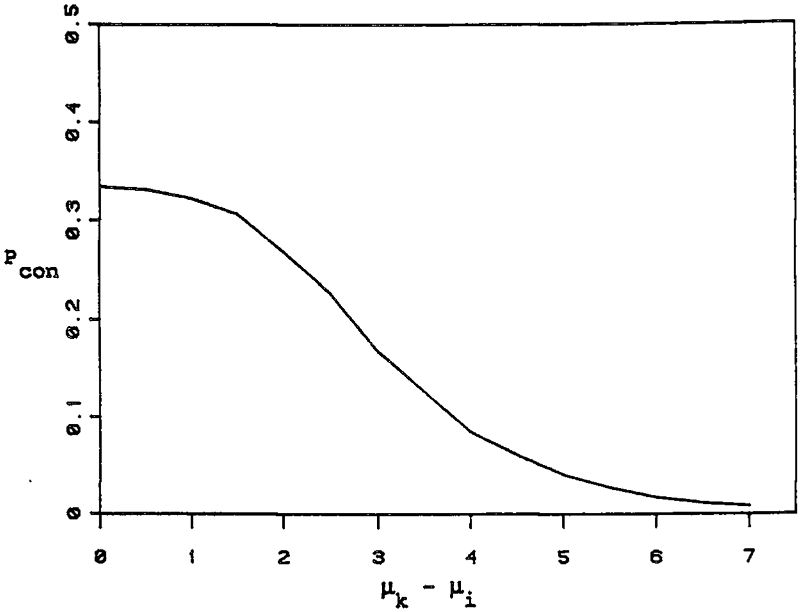


Figure 1. The probability of a decision conflict (P_{con}) in Richardson's method of triads as a function of the difference between the largest and smallest mean ($\mu_k - \mu_i$), when the intermediate mean, μ_j, is equal to $(\mu_k + \mu_i)/2$.

considered the methods of triads from a more classical Thurstonian viewpoint in the sense that we modelled the relationship between decision probabilities from these methods and the unidimensional parameters of the psychological magnitudes (means and variances). By assuming that the momentary psychological magnitudes can be modelled as if they were drawn at random from normal distributions with particular means and variances, we are simultaneously assuming that the distances between pairs of these magnitudes cannot be modelled as normal deviates.

Equations relating the probabilities associated with particular decision outcomes and the means and variances of the psychological magnitudes were derived and evaluated numerically for both Torgerson's and Richardson's methods of triads. Using non-linear least squares minimization, the means of the distribution of psychological magnitudes were estimated. In order to employ either of the methods of triads for estimation work with large numbers of means (>5), it will be necessary to work out optimum designs aimed at reducing the number of decision probabilities needed. For large problems, it should be necessary only to estimate a fraction of the probabilities to obtain reliable parameter estimates.

It was shown how the triangular method is a special case of Richardson's method and that the duo–trio method is a special case of Torgerson's method. The problem of producing decision conflicts in Richardson's method was discussed. This problem suggests that Richardson's method should be avoided unless the experimenter is certain that psychological magnitudes remain at fixed values during a trial.

The theory derived and discussed in this paper is unidimensional. In previous papers on multivariate models (Ennis & Mullen, 1986*a*, *b*; Mullen & Ennis, 1987) it was shown how decision probabilities are related to several parameters other than the distance between the means of the psychological magnitudes and that a multivariate model for discrimination methods should be a multiparameter model incorporating the effects of correlation structure, relative orientation of the means to each other, variances and distances. Unless simplifying assumptions are made about the variance–covariance matrix, decision probabilities in discrimination tasks will not be monotonically related to distances. In order to use either of the methods of triads for multidimensional scaling, it will be necessary to formulate a multivariate extension of the theory given in this paper. As noted above, since this has already been done for the triangular and duo–trio methods, the next task will be to generalize these ideas.

Acknowledgements

We greatly appreciate the help provided by C. Lilly through the Philip Morris Visiting Scientist Program which facilitated our collaboration. J. Tindall, P. Martin and two anonymous reviewers provided helpful comments on an earlier version of the manuscript. Management support and library and computer resources were provided by L. Meyer, P. Martin, G. McDonald and M. Waugh. Graphics support was given by C. Nilles.

References

David, H. A. & Trivedi, M. C. (1962). Pair, triangle and duo–trio tests. Technical Report 55, Department of Statistics, Virginia Polytechnic Institute, Blacksburg, Virginia.

Ennis, D. M. & Mullen, K. (1985). The effect of dimensionality on results from the triangular method. *Chemical Senses*, **10**, 605–608.

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

Ennis, D. M. & Mullen, K. (1986*b*). Theoretical aspects of sensory discrimination. *Chemical Senses*, **11**, 513–522.

IMSL (1984). IMSL Library (Version 9.2), developed by IMSL, Inc., 7500 Bellaire Blvd, Houston, Texas 77036.

Fritjers, J. E. R. (1979*a*). Variations of the triangular method and the relationship of its unidimensional probabilistic models to three-alternative forced choice signal detection theory models. *British Journal of Mathematical and Statistical Psychology*, **32**, 229–241.

Fritjers, J. E. R. (1979*b*). The paradox of the discriminatory nondiscriminators resolved. *Chemical Senses and Flavour*, **4**, 335–358.

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.

Green, D. M. & Swets, J. A. (1966). *Signal Detection Theory and Psychophysics*. New York: Wiley.

Mullen, K. & Ennis, D. M. (1987). Mathematical formulation of multivariate Euclidean models for discrimination methods. *Psychometrika*, **52**, 235–249.

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

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

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

Ura, S. (1960). Pair, triangle and duo–trio tests. *Japanese Union of Scientists and Engineers*, **7**, 107–119.

Received 13 May 1987; revised version received 10 December 1987

Appendix 1

${}_iP_{jk}$ is the probability that S_i will be perceived to be more similar to S_j than S_k. $x_1, x_2, \ldots, x_n$ are the momentary psychological values, $\mu_1, \mu_2, \ldots, \mu_n$ are their means and $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$ are their variances. Thus, x_i, μ_i and σ_i^2 are associated with S_i for $i = 1, \ldots, n$.

$$\begin{aligned}
{}_iP_{jk} &= P(|x_i - x_j| < |x_i - x_k|) \\
&= P(|x_i - x_j| < x_i - x_k) + P(|x_i - x_j| < -x_i + x_k) \\
&= P(x_i - x_j < x_i - x_k \text{ and } -x_i + x_j < x_i - x_k) \\
&\quad + P(x_i - x_j < -x_i + x_k \text{ and } -x_i + x_j < -x_i + x_k) \\
&= P(x_k - x_j < 0 \text{ and } -2x_i + x_j + x_k < 0) + P(x_j - x_k < 0 \text{ and } 2x_i - x_j - x_k < 0).
\end{aligned}$$

Let $x_j - x_k = r$ and $2x_i - x_j - x_k = s$, then r and s have a bivariate normal distribution with means $\mu_j - \mu_k$ and $2\mu_i - \mu_j - \mu_k$ and variance–covariance matrix $\mathbf{V}$.

Hence

$${}_iP_{jk} = \int_0^\infty \int_0^\infty f(r,s)\,\mathrm{d}r\,\mathrm{d}s + \int_{-\infty}^0 \int_{-\infty}^0 f(r,s)\,\mathrm{d}r\,\mathrm{d}s,$$

where

$$r = (x_j - x_k) \qquad \text{and} \qquad s = (2x_i - x_j - x_k),$$

$$f(r,s) = \frac{\exp\{-0.5[(\mathbf{z} - \boldsymbol{\mu})' V^{-1}(\mathbf{z} - \boldsymbol{\mu})]\}}{(2\pi)|\mathbf{V}|^{1/2}},$$

$$\mathbf{z}' = (r, s),$$

$$\boldsymbol{\mu}' = [(\mu_j - \mu_k), (2\mu_i - \mu_j - \mu_k)],$$

and

$$\mathbf{V} = \begin{bmatrix} \sigma_j^2+\sigma_k^2 & \sigma_k^2-\sigma_j^2 \\ \sigma_k^2-\sigma_j^2 & 4\sigma_i^2+\sigma_j^2+\sigma_k^2 \end{bmatrix}.$$

When $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_n^2$, $x_k - x_j$ and $-2x_i + x_j + x_k$, being orthogonal, are independently distributed, so that

$${}_iP_{jk} = P(x_k - x_j < 0)P(-2x_i + x_j + x_k < 0) + P(x_j - x_k < 0)P(2x_i - x_j - x_k < 0).$$

The notation $N(\mu, \sigma^2)$ will be used to indicate the normal function with mean and variance parameters μ and σ^2 respectively, and $\Phi(x)$ to indicate the area under the standard normal curve from $-\infty$ to x.

Let

$$p = P(x_j - x_k) < 0) = \int_{-\infty}^{0} N(\mu_j - \mu_k, 2\sigma^2) = \int_{-\infty}^{(\mu_k-\mu_j)/(2\sigma^2)^{0.5}} N(0,1)$$
$$= \Phi[(\mu_k - \mu_j)/(2\sigma^2)^{0.5}]$$

and

$$q = P(2x_i - x_j - x_k < 0) = \int_{-\infty}^{0} N(2\mu_i - \mu_j - \mu_k, 6\sigma^2)$$
$$= \int_{-\infty}^{(\mu_k+\mu_j-2\mu_i)/(6\sigma^2)^{0.5}} N(0,1)$$
$$= \Phi[(\mu_k + \mu_j - 2\mu_i)/(6\sigma^2)^{0.5}].$$

Therefore,

$${}_iP_{jk} = (1-p)(1-q) + pq = 1 - p - q + 2pq$$
$$= 1 - \Phi[(\mu_k - \mu_j)/\sigma\sqrt{2}] - \Phi[(\mu_k + \mu_j - 2\mu_i)/\sigma\sqrt{6}]$$
$$+ 2\Phi[(\mu_k - \mu_j)/\sigma\sqrt{2}]\Phi[(\mu_k + \mu_j - 2\mu_i)/\sigma\sqrt{6}].$$

Appendix 2

$$P(a_{ij}\, d_{ik}) = P(|x_i - x_j| < |x_i - x_k| \text{ and } |x_i - x_j| < |x_j - x_k| \text{ and}$$
$$|x_i - x_k| > |x_i - x_j| \text{ and } |x_i - x_k| > |x_j - x_k|).$$

The first and third terms are the same, so that

$$P(a_{ij}\, d_{ik}) = P(|x_i - x_j| < |x_i - x_k| \text{ and } |x_i - x_j| < |x_j - x_k| \text{ and}$$
$$|x_i - x_k| > |x_j - x_k|).$$

Let $(x_i - x_j) = u$; $(x_i - x_k) = v$; and $(x_j - x_k) = w$.

It can be seen that $u - v + w = 0$.

$$P(a_{ij}\, d_{ik}) = P(|u| < |v| \text{ and } |u| < |w| \text{ and } |w| < |v|).$$

$$P(|u| < |v| \text{ and } |u| < |w|) = P(|u| < |v| \text{ and } |v| < |w|) + P(|u| < |w| \text{ and } |w| < |v|).$$

Therefore

$$P(a_{ij}\, d_{ik}) = P(|u| < |v| \text{ and } |v| < |w| \text{ and } |w| < |v|)$$
$$+ P(|u| < |w| \text{ and } |w| < |v| \text{ and } |w| < |v|).$$

In this sum, the first probability contains a contradiction and therefore has a probability of zero, while the second probability contains a redundancy, so that

$$\begin{aligned}P(a_{ij}\, d_{ik}) &= P(|u| < |w| \text{ and } |w| < |v|)\\ &= P(|u| < |w| \text{ and } |w| < |w+u|),\end{aligned}$$

since $u - v + w = 0$. For both of these inequalities to be simultaneously true, u and w must be either both positive or both negative. Therefore,

$$\begin{aligned}P(a_{ij}\, d_{ik}) &= P(|u| < |w| \text{ and } |w| < |w+u| \text{ and } u > 0, w > 0)\\ &\quad + P(|u| < |w| \text{ and } |w| < |w+u| \text{ and } u < 0, w < 0)\\ &= P(0 < u < w) + P(w < u < 0).\\ &= \int_0^\infty \int_u^\infty f(u, w)\, dw\, du + \int_{-\infty}^0 \int_{-\infty}^u f(u, w)\, dw\, du\end{aligned}$$

where $f(u, w)$ is the bivariate normal distribution of u and w or

$$f(u, w) = \frac{\exp\{-0.5[(\mathbf{z}-\boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{z}-\boldsymbol{\mu})]\}}{(2\pi)|\mathbf{V}|^{1/2}},$$

where

$$\mathbf{z}' = (u, w),$$

$$\boldsymbol{\mu}' = [(\mu_i - \mu_j), (\mu_j - \mu_k],$$

and

$$\mathbf{V} = \begin{bmatrix} \sigma_i^2 + \sigma_j^2 & -\sigma_j^2 \\ -\sigma_j^2 & \sigma_j^2 + \sigma_k^2 \end{bmatrix}.$$

British Journal of Mathematical and Statistical Psychology (1989) 42, 265–269 *Printed in Great Britain*

Decision conflicts: Within-trial resampling in Richardson's method of triads

Daniel M. Ennis†
Philip Morris Research Center, Commerce Road, Richmond, VA 23261, USA

Kenneth Mullen
Department of Mathematics and Statistics, University of Guelph, Ontario, Canada

Jan E. R. Frijters
The Netherlands Agricultural University, Wageningen

J. Tindall
Philip Morris Research Center, VA

A unidimensional model is derived which can provide the probability that the same two stimuli will be perceived to be most alike *and* most different in Richardson's method of triads under the assumption that resampling occurs within a trial. This probability is shown to depend on the extent to which the stimulus distributions overlap and their relative locations on a unidimensional continuum. Recommendations on how to estimate this probability experimentally are given.

1. Introduction

In a previous paper (Ennis, Mullen & Frijters, 1988), Thurstonian models for Richardson's and Torgerson's methods of triads were derived and discussed. Both of these methods involve decisions related to perceived differences between three stimulus objects. In Togerson's method, subjects are instructed to select one of two objects which is most similar to the third (preselected) object. In Richardson's method, the subject is instructed to identify the most similar pair and the most different pair from three objects.

In modelling these two tasks, it is assumed that a stimulus is represented mentally by a momentary psychological magnitude that has been drawn at random from a univariate normal distribution. This assumption implies that stimulus magnitudes may change from trial to trial as the subject obtains new mental representations for a particular stimulus. In Richardson's method, the subject is asked to make two decisions within a trial about the three stimuli: namely, which two stimuli are most alike and which two are most different. It is not unreasonable to expect that a subject

†Requests for reprints.

may resample the stimulus distributions to make these two decisions and, consequently, may be working with different information about the stimuli when making these two decisions. In previous work (Ennis *et al.*, 1988) it was assumed that the mental representations of the objects remain at fixed values during a trial for both methods. If this assumption is violated when the subject is under the instructions for Richardson's method, as may occur if the subject resamples during a trial, it is possible that the same two objects may appear to be most similar and most different. Since this outcome is not unlikely with Richardson's method, and may occur with other discrimination paradigms, it is useful to understand more about its probability of occurrence and how to estimate and use it experimentally.

2. Assumptions

A triad is composed of three stimulus objects, S_i, S_j and S_k, $i \neq j \neq k \neq i$. Each object is assumed to be represented mentally by corresponding univariate psychological magnitudes which may vary from moment to moment but which are assumed to be drawn randomly from normal distributions with means μ_i, μ_j and μ_k and variances σ_i^2, σ_j^2 and σ_k^2. The distributions of momentary psychological magnitudes are assumed to be independent. If x_i, x_j and x_k are the psychological magnitudes used to decide which two objects are most alike and x_i^*, x_j^* and x_k^* are the magnitudes used to decide which two objects are most different, there is some probability that

$$|x_i - x_j| < |x_i - x_k| \text{ and } |x_i - x_j| < |x_j - x_k| = P(a_{ij})$$

and that

$$|x_i^* - x_j^*| > |x_i^* - x_k^*| \text{ and } |x_i^* - x_j^*| > |x_j^* - x_k^*| = P(d_{ij}).$$

Similarly, one may define $P(a_{ik})$, $P(d_{ik})$, $P(a_{jk})$ and $P(d_{jk})$. If the subject is not aware that the above outcome is possible, it would lead to what may be called a decision conflict for the subject. Represent its probability by P_{con}.

3. Probability of a decision conflict

A subject will be faced with a decision conflict whenever the stimulus pair selected to be most alike is also the pair selected to be most different. The probability that S_i and S_j are most alike is $P(a_{ij})$ and the probability that they are most different is $P(d_{ij})$, so the probability that S_i and S_j are most alike *and* most different is $P(a_{ij}) \cdot P(d_{ij})$. In formulating the model for Richardson's method of triads (Ennis *et al.*, 1988) when the momentary psychological magnitudes are assumed to remain at fixed values, this probability was zero. If resampling within a trial occurs, this probability may be different from zero. Since the same two stimuli may be most alike and most different in three ways, the probability of a decision conflict is

$$P_{\text{con}} = P(a_{ij}) \cdot P(d_{ij}) + P(a_{ik}) \cdot P(d_{ik}) + P(a_{jk}) \cdot P(d_{jk}). \tag{1}$$

Within-trial resampling in Richardson's method of triads

The equations for $P(a_{ij})$ and $P(d_{ij})$ will be given. The other four probabilities, $[P(a_{ik}), P(d_{ik}), P(a_{jk})$ and $P(d_{jk})]$ can be obtained similarly.

Let $(x_i - x_j) = u$; $(x_i - x_k) = v$; and $(x_j - x_k) = w$.

Since x_i, x_j and x_k are random normal deviates, then u, v and w are also random normal deviates.

$$P(a_{ij}) = P(|u| < |v| \text{ and } |u| < |w|)$$

$$= P(|v - w| < |v| \text{ and } |v - w| < |w|),$$

which implies that both v and w have the same sign.

Therefore,

$$P(a_{ij}) = P(v > w/2 \text{ and } v < 2w \text{ and } v > 0, w > 0)$$

$$+ P(v > 2w \text{ and } v < w/2 \text{ and } v < 0, w < 0).$$

This equation involves the sum of two twofold integrals involving a bivariate normal distribution with a particular vector of means and variance–covariance matrix for the variables v and w.

$$P(a_{ij}) = \int_0^{\infty} \int_{w/2}^{2w} f(v, w)\, dv\, dw + \int_{-\infty}^{0} \int_{2w}^{w/2} f(v, w)\, dv\, dw.$$

Let $(x_i^* - x_j^*) = r$; $(x_i^* - x_k^*) = s$; and $(x_j^* - x_k^*) = t$.

$$P(d_{ij}) = P(|s - t| > |s| \text{ and } |s - t| > |t|),$$

which implies that s and t have opposite signs.

Therefore,

$$P(d_{ij}) = P(s - t > s \text{ and } s - t > -t \text{ and } s > 0, t < 0)$$

$$+ P(t - s > -s \text{ and } t - s > t \text{ and } s < 0, t > 0)$$

$$= P(t < 0 \text{ and } s > 0) + P(t > 0 \text{ and } s < 0)$$

$$= \int_{-\infty}^{0} \int_{0}^{\infty} f(s, t)\, ds\, dt + \int_{0}^{\infty} \int_{-\infty}^{0} f(s, t)\, ds\, dt.$$

The function, f, is a bivariate normal distribution which has the general form,

$$f = \frac{\exp\{-0.5[(\mathbf{z} - \boldsymbol{\mu})'\mathbf{V}^{-1}(\mathbf{z} - \boldsymbol{\mu})]\}}{(2\pi)|\mathbf{V}|^{1/2}}.$$

For $f(v, w)$, $\mathbf{z}' = (v, w)$; for $f(s, t)$, $\mathbf{z}' = (s, t)$; and for both functions,

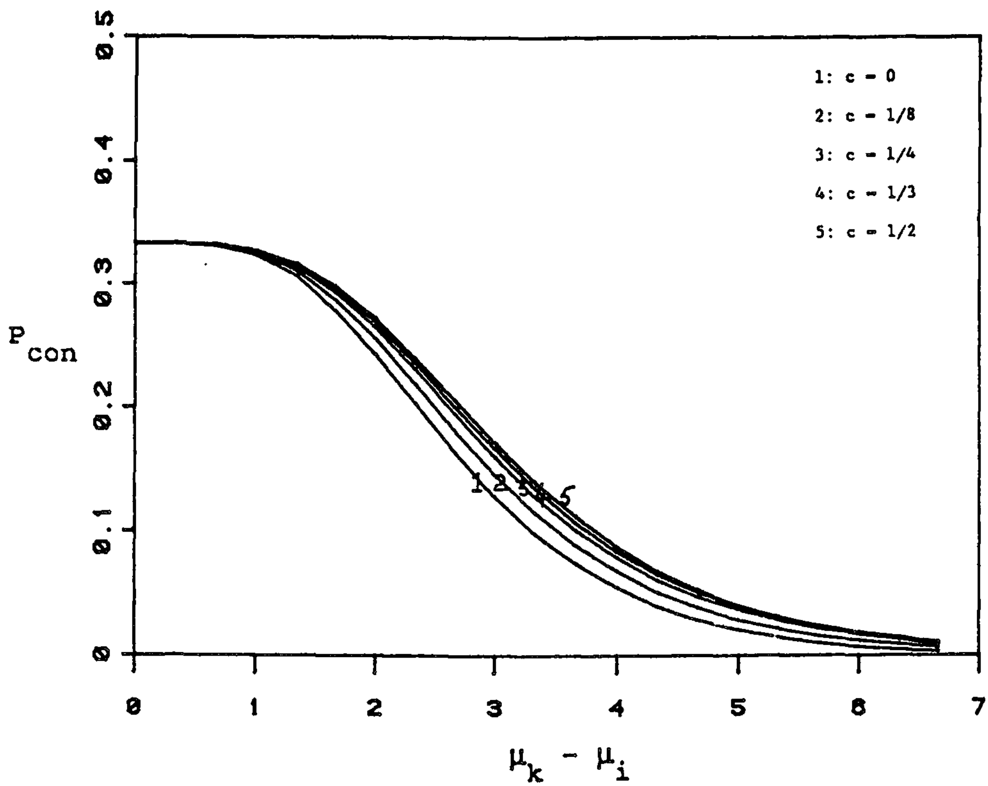


Figure 1. The relationship between P_{con} and $(\mu_k-\mu_i)$ for values of $c=0$, 1/8, 1/4, 1/3 and 1/2 (or 1, 7/8, 3/4, 2/3 and 1/2) where $c=(\mu_j-\mu_i)/(\mu_k-\mu_i)$, $\mu_k>\mu_j>\mu_i$ and $\sigma_i=\sigma_j=\sigma_k=1.0$.

$$\mu'=[(\mu_i-\mu_k),(\mu_j-\mu_k)]; \quad \text{and} \quad \mathrm{V}=\begin{bmatrix} \sigma_i^2+\sigma_k^2 & \sigma_k^2 \\ \sigma_k^2 & \sigma_j^2+\sigma_k^2 \end{bmatrix}.$$

The variance–covariance matrix, V, can be derived from the expected values for the variances and covariances of v and w or s and t.

4. Evaluation of the model

Equation (1) was evaluated numerically on a Gould 32/97. Relative error in each integration was less than 0.0001. In Fig. 1, P_{con} is plotted as a function of $(\mu_k-\mu_i)$ and the relative position of μ_j on the unidimensional continuum, assuming that $\mu_k>\mu_j>\mu_i$ and that $\sigma_i=\sigma_j=\sigma_k=1.0$. It can be seen that the probability of obtaining a decision conflict is determined mainly by the degree of overlap of the stimulus distributions,, and depends to a lesser extent on the distance between the middle stimulus and either of the two end stimuli. In the case where two of the stimuli have identical means ($c=0$ or 1 in Fig. 1), we have the triangular method (Frijters, 1979).

5. Suggestions for experimental work

Assume that a subject resamples within a trial while under the instructions for Richardson's method of triads, and develops a strategy for dealing with cases in which it appears that the same two stimuli are both most alike and most different. The resulting decision probabilities may be different from those obtained from the Thurstonian model for this method previously described by Ennis *et al.* (1988). The experimenter may wish to compare estimates of the means and variances obtained from both Torgerson's and Richardson's methods under Thurstonian assumptions. Assuming that there are consistent differences between these methods which are due to the effect of resampling in Richardson's method, then it would be of interest to determine the probability of a conflict and to confirm this value experimentally. Assuming that the values of the means and variances for an array of stimuli are known (using Torgerson's method, for instance), it should be possible to compute the probability of a conflict from the theory given in this paper. The subject could then be presented with triads of the stimuli of interest and instructed to make 'most similar' and 'most different' pair selections in different trials. These trials could be paired to simulate resampling within a trial, but the subject would be instructed to expect possibly different stimuli from trial to trial. From a sufficiently large number of trials, the probability of occurrence of pairs of trials in which the same stimuli are selected as most similar and most different could be computed. This value could be compared with the theoretical value obtained as previously described.

Acknowledgements

The authors thank C. Lilly who supported two of the authors (KM and JERF) through the Philip Morris Visiting Scientist Program which facilitated interdisciplinary collaboration. Thanks are extended also to E. Gee and M. Waugh for management support and computer resources.

References

Ennis, D. M., Mullen, K. & Frijters, J. E. R. (1988). Variants of the method of triads: Unidimensional Thurstonian models. *British Journal of Mathematical and Statistical Psychology*, **41**, 25–36.

Frijters, J. E. R. (1979). Variations of the triangular method and the relationship of its unidimensional probabilistic models to three-alternative forced choice signal detection theory models. *British Journal of Mathematical and Statistical Psychology*, **32**, 229–241.

Received 22 February 1988; revised version received 26 June 1989

Fechner-Thurstone Models

Daniel M. Ennis

PHILIP MORRIS RESEARCH CENTER

Kenneth Mullen

DEPARTMENT OF MATHEMATICS AND STATISTICS

UNIVERSITY OF GUELPH

Abstract

A *Fechner-Thurstone* model is a model that relates the outcome of a psychological task to a psychophysical transformation and the parameters of stimulus and psychological probability density functions. It is assumed that the psychophysical transformation is a *one-to-one* function, the domain of which is the vector space of stimulus magnitudes and the range of which is the vector space of psychological magnitudes. Since this function is one-to-one, it is invertible. A very general equation is derived for the probability density function (pdf) of the momentary psychological magnitudes based on any psychological pdf and any one-to-one function of stimulus magnitudes from any stimulus pdf. The psychophysical transformation may be one-to-one only over the domain of the stimulus pdf. This general model is applied specifically to the method of paired comparisons. Models for paired comparisons are derived based on assumed lognormally distributed stimulus values on a physicochemical continuum, a log or power psychophysical transformation and added psychological variance from normally distributed values. Fechner-Thurstone models can contribute to the unification of the scaling approaches of Fechner, Stevens and Thurstone by providing a basis for relating parameters which describe the stimulus, the psychophysical transformation and perceptual variance to a subject's response in a particular task. The parameters of a sample problem are estimated using the Levenberg-Marquardt algorithm for nonlinear least squares estimation.

Introduction

Attempts to derive valid relationships between the physical world and a mental representation of it originate, at least formally, with Fechner (1860). In this paper, we consider any model which relates a physicochemical attribute of a stimulus object with its mental representation or percept as a Fechnerian model. A logarithmic transformation is an example of one of several Fechnerian models that can be studied. Unlike Fechnerian models, in Thurstone's (1927) conception of mental scaling, a knowledge of the physical attributes of the stimulus object is unnecessary since the scaling is achieved by estimating the parameters of assumed (normal) distributions of psychological magnitudes based on modelling which connects the subject's choice probabilities with these parameters. In order to use Thurstonian modelling validly, it is necessary to assume that the variation leading to different momentary psychological magnitudes is perceptual. To allow this assumption, experimenters must ensure that physicochemical stimulus variance does not exist, otherwise a more general model is needed. For some sense modalities, such as those involving the chemical senses, this assumption is almost always impossible to justify, as exact stimulus control cannot be achieved. This fact and a lack of formal models to address this problem motivated the present work, although it is recognized that the models developed may be applied also to other modalities, such as audition and vision.

The purpose of this paper is to begin to address the problem of connecting stimulus probability density parameters through a psychophysical transformation to probabilistic perceptual models. Thus, this effort can be seen as an attempt to connect Fechnerian modelling with Thurstonian modelling. Different assumptions about the psychophysical transformation, the

Fechnerian model, connecting the physicochemical measures to the corresponding perceptual measures will be discussed. Consequently, the models discussed could be called *Fechner-Thurstone* models.

Fechner-Thurstone Models: Assumptions and Generalities

There is some number of sets of stimulus objects, typically represented as S_i, and within each set there are objects with a common attribute that can be measured on a single physicochemical continuum. A stimulus magnitude, φ_i, is the value on this continuum for a particular stimulus object. Let $f_1(\varphi_i)$ be some probability density function (pdf) of φ_i. Following a psychophysical transformation, the stimulus magnitude is represented mentally by a psychological magnitude, ψ_i. Let $g(\varphi_i)$ be any *one-to-one* function of φ_i which can operate on the entire domain of f_1. The function g is the psychophysical transformation. Since g is one-to-one, it is invertible because for $g(\varphi_1) = g(\varphi_2)$, the only solution is $\varphi_1 = \varphi_2$. All monotonic functions are one-to-one. The pdf of ψ_i, $h(\psi_i)$, is

$$f_1 \circ g^{-1}(\psi_i)\, |dg^{-1}(\psi_i)/d\psi_i|,$$

where $f_1 \circ g^{-1}(\psi_i)$ is a composition function. The psychological magnitude, ψ_i, is a parameter in a probability density function of momentary psychological magnitudes, $x_i|\psi_i$, which may occur if central or peripheral noise is present, for instance. Let $f_2(x_i|\psi_i)$ be any pdf of $x_i|\psi_i$. Since the pdfs of ψ_i and $x_i|\psi_i$ are h and f_2, the pdf of x_i is

$$\int_D f_2(x_i|\psi_i)h(\psi_i)d\psi_i \qquad (1)$$

where D is the domain of h.

The function f_1 may be a multivariate pdf of a vector of physicochemical measures instead of the single variable, φ_i. The function g, a one-to-one function, would map vectors from the domain space of f_1 to either single values (if g is a real-valued function) or vectors (if g is a vector-valued function) in the range space of g. Since g is a one-to-one function, g would have an inverse. Similar steps to those taken in deriving (1) could be taken to derive the pdf of $\mathbf{x}_i$ (the vector equivalent of x_i, the momentary psychological magnitude). Unidimensional and multidimensional probabilistic models, in which the pdf of x_i or $\mathbf{x}_i$ is assumed to be normal, have received some attention in recent years for a wide variety of subject tasks. Some examples of these models are given in Ashby & Perrin (1988), Bradley (1976), De Soete, Carroll, & DeSarbo (1986), Ennis (1988), Ennis & Mullen (1986), Ennis, Palen & Mullen (1988), Ennis, Mullen, & Frijters (1988), Frijters (1979), Iverson (1987), MacKay (1989), Mullen & Ennis (1987), Mullen, Ennis, de Doncker, & Kapenga (1988), Zinnes & MacKay (1983), and Zinnes & MacKay (1987). In these models, no attention is given to stimulus parameters or to processes which might have led to the percepts. In this sense, the probabilistic models are much more restrictive than the Fechner-Thurstone models, which would require very special assumptions to ensure normally (uni- or multidimensional) distributed momentary psychological magnitudes.

Once the pdf of x_i or $\mathbf{x}_i$ is known, the models relating response probabilities associated with various tasks to the parameters of f_1, f_2, and g can be determined. These tasks include m-alternative forced choice (such as the method of paired comparisons and the method of constant stimuli), triads, identification, categorization and preferential choice.

A General Form for Paired Comparisons

The pdf of x_i, from (1), is

$$\int_D f_2(x_i|\psi_i)h(\psi_i)d\psi_i$$

Consider two independent psychological magnitudes, x_1 and x_2. The joint pdf of x_1 and x_2 is

$$\left[\int_D f_2(x_1|\psi_1)h(\psi_1)d\psi_1\right]\left[\int_D f_2(x_2|\psi_2)h(\psi_2)d\psi_2\right]$$

If $y = x_1 - x_2$, the joint pdf of x_1 and y is

$$\left[\int_D f_2(x_1|\psi_1)h(\psi_1)d\psi_1\right]\left[\int_D f_2(x_1 - y|\psi_2)h(\psi_2)d\psi_2\right]$$

If C is the domain of x_i, the pdf of y is

$$\int_C \left[\int_D f_2(x_1|\psi_1)h(\psi_1)d\psi_1\right]\left[\int_D f_2(x_1 - y|\psi_2)h(\psi_2)d\psi_2\right] dx_1.$$

In a paired comparison task, where the probability that $x_2 > x_1$ is of interest, the binary choice probability is Pr(y < 0) or

$$\int_{-\infty}^{0} \int_{C} \left[\int_{D} f_2(x_1|\psi_1)h(\psi_1)d\psi_1 \right] \left[\int_{D} f_2(x_1 - y|\psi_2)h(\psi_2)d\psi_2 \right] dx_1 dy$$

$$= \int_{C} \left[\int_{D} f_2(x_1|\psi_1)h(\psi_1)d\psi_1 \right] \left[\int_{D} h(\psi_2) \cdot \int_{-\infty}^{0} f_2(x_1 - y|\psi_2)dy d\psi_2 \right] dx_1 \qquad (2).$$

Special Cases

Assume that the stimulus magnitudes are lognormally distributed with means μ_i (corresponding to stimulus set S_i), and variance σ^2_{1i}. One reason for the choice of a lognormal probability density function is that two forms of g will be considered, both of which require the domain of f_1 to be positive. Following a psychophysical transformation, the stimulus magnitude is represented mentally by a psychological magnitude, ψ_i which is itself assumed to be the mean of a normal distribution of psychological magnitudes. A randomly sampled psychological magnitude, x_i, is assumed to be drawn from this distribution of psychological magnitudes with mean ψ_i and variance σ^2_{2i} and is represented as $x_i|\psi_i$.

The Logarithmic Psychophysical Transformation

In this special case, $g(\varphi_i) = k\log \varphi_i$. By assumption, $\log \varphi_i \sim N(\mu_i,\sigma^2_{1i})$ ($\sim$ N means distributed normally). Since $\psi_i = k\log \varphi_i$, then $\psi_i \sim N(k\mu_i, k^2\ \sigma^2_{1i})$. Given a particular value of ψ_i, x_i is a normal deviate from a probability density function with mean ψ_i and variance σ^2_{2i}.

Since $\psi_i \sim N(k\mu_i, k^2\sigma^2_{1i})$, then x_i is normally distributed with mean $k\mu_i$ and variance $(k^2\sigma^2_{1i} + \sigma^2_{2i})$. If P_{S_2,S_1} is the binary choice probability in a paired comparison,

$$P_{S_2,S_1} = \Phi\left[\frac{(\mu_2 - \mu_1)k}{\{k^2(\sigma^2_{11} + \sigma^2_{12}) + \sigma^2_{21} + \sigma^2_{22}\}^{0.5}}\right], \tag{3}$$

where $\Phi(z)$ is the area under the normal pdf from $-\infty$ to z.

The Power Psychophysical Transformation

By assumption, $\log \varphi_i \sim N(\mu_i, \sigma^2_{1i})$, $g(\psi_i) = \varphi_i^{\beta}$ and $x_i|\psi_i \sim N(\psi_i, \sigma^2_{2i})$. From (2), employing the functions f_1, f_2 and g corresponding to these assumptions,

$$P_{S_2,S_1} = \int_{-\infty}^{0}\int_{-\infty}^{\infty}\left[\int_0^{\infty}\frac{\exp\{-(x_1-\psi_1)^2/(2\sigma^2_{21}) - (\log\psi_1 - \beta\mu_1)^2/(2\beta^2\sigma^2_{11})\}}{2\pi\sigma_{11}\sigma_{21}\psi_1\beta}d\psi_1\right]$$

$$\times\left[\int_0^{\infty}\frac{\exp\{-(x_1 - y - \psi_2)^2/(2\sigma^2_{22}) - (\log\psi_2 - \beta\mu_2)^2/(2\beta^2\sigma^2_{12})\}}{2\pi\sigma_{12}\sigma_{22}\psi_2\beta}d\psi_2\right] dx_1 dy.$$

$$= \int_{-\infty}^{\infty}\left[\int_0^{\infty}\frac{\exp\{-(x_1-\psi_1)^2/(2\sigma^2_{21}) - (\log\psi_1 - \beta\mu_1)^2/(2\beta^2\sigma^2_{11})\}}{2\pi\sigma_{11}\sigma_{21}\psi_1\beta}d\psi_1\right]$$

$$\times\left[\int_0^{\infty}\frac{\exp[-(\log\psi_2 - \beta\mu_2)^2/(2\beta^2\sigma^2_{12})]}{\sqrt{(2\pi)}\sigma_{12}\psi_2\beta}\cdot\Phi\left[-\frac{(x_1 - \psi_2)}{\sigma_{22}}\right]d\psi_2\right] dx_1 \tag{4}$$

Function Evaluations and Parameter Estimation

Equation (4) was evaluated numerically on both Gould 32/97 and Trace Multiflow computers using Fortran subroutines for numerical quadrature [Piessens, deDoncker-Kapenga, Überhuber & Kahaner (1983)]. These programs are contained in the IMSL Library (IMSL, 1987). Transformations to achieve (0,1) limits of integration were used for the outer integral following a separation of the integral into two parts ($-\infty$ to 0 and 0 to ∞). The upper bounds on the inner integrals (theoretically ∞) were set to $[\exp(\mu_i + 4\sigma_{1i})]^{\beta}$. These bounds corresponded to 4 standard deviation log units on the physicochemical continuum (assumed to be lognormal) which was found to be sufficient to give at least fifth decimal accuracy in the computed probabilities.

The physicochemical values presumed to be transformed to mental representations can be measured, in many cases, with a high degree of reliability. Examples are: the partial pressure or concentration of a compound, the amplitude of a tone, the lengths of lines and angles in geometric figures and light intensity. In other cases, the experimenter may not be able to identify and measure a particular attribute, or set of attributes, that are relevant in making a particular kind of judgment. In some cases, particularly in experiments in the chemical senses, the parameters of the stimulus pdf (f_1) can be measured, but stimulus variance cannot be eliminated. In this kind of application, the use of a Fechner-Thurstone model can greatly reduce the number of parameters to be estimated compared to Thurstonian modelling, and also provide information on the nature of the psychophysical transformation. An illustration of this type of model fitting will be given next in which β and σ_{2i} will be estimated. It is recognized that the parameters of f_1, the stimulus pdf, could be estimated also (for cases in which they cannot be measured easily) thus yielding the experimenter with a basis for uncovering stimulus attributes that, through an appropriate transformation, determine psychological magnitudes. This type of modelling is a relatively straightforward extension of the much simpler case that we will consider and can be implemented using the same model-fitting techniques.

A matrix of choice probabilities was computed for 10 stimulus distributions for which the geometric means were 0.0 to 0.9 in 0.1 increments and for which there was a common standard deviation of 0.2. The power exponent, β, in the function g was set to 0.6 and the psychological standard deviation, σ_{2i} was 0.2. It was assumed that subjects would select the stimulus with the greatest momentary psychological magnitude. These 45 probabilities are, therefore, a function of the stimulus parameters, the transformation function and variance at the psychological level. The Levenberg-Marquardt algorithm (Marquardt, 1963) was used to obtain nonlinear least squares estimates of these parameters. Let **a** be a vector containing the parameters to be estimated.

Define

$$q_{ij}(\mathbf{a}) = p_{ij} - P_{S_i,S_j}, \; i > j,$$

where p_{ij} is the observed proportion of judgments for which stimulus objects from S_i are chosen. Let **q**(**a**) be a vector with typical element $q_{ij}(\mathbf{a})$. The value to be minimized is the residual sum of squares, $\mathbf{q}(\mathbf{a})^t\mathbf{q}(\mathbf{a})$. If $\mathbf{a}^0$ is an initial estimate of **a**, a series of approximations are computed as

$$\mathbf{a}^{n+1} = \mathbf{a}^n - [\alpha_n \mathbf{D}_n + \mathbf{J}_n^t \mathbf{J}_n]^{-1} \mathbf{J}_n^t \mathbf{q}(\mathbf{a}^n),$$

where

J_n is the Jacobian matrix evaluated at a^n,

D_n is a diagonal matrix with entries equal to the diagonal of $J_n^t J_n$, and

α_n is the Marquardt parameter, a positive constant.

J_n was approximated using finite differences in double precision. The Marquardt parameter (α_n), initially 0.01, was quadrupled if the residual sum of squares increased from one iteration to the next and was halved if it decreased.

Random initial parameter values were used and several starting configurations led to parameter estimates that agreed to at least the third decimal place in each case. The results of one such analysis were: $\beta = 0.60063$ and $\sigma_{2i} = .20029$. There did not appear to be more than one solution which differed from the above case by more than 0.001 in β or σ_{2i} when different starting configurations were used.

Model parameters were also estimated for the more realistic case in which the choice probabilities would not be error-free. In the case of each probability in the previously mentioned matrix, a random deviate was added from a normal distribution with zero mean and variance equal to $[(P_{S_2,S_1})(1 - P_{S_2,S_1})]/200$. These choice probabilities correspond to the experimental situation where 200 observations per comparison have been obtained and the model is appropriate. A matrix of this kind was analyzed. Estimates of the parameters were: $\beta = 0.506$ and $\sigma_{2i} = 0.168$. These values can be compared with the actual values of 0.6 and 0.2 for β and σ_{2i}, respectively. The residual sum of squares (0.028) was consistent with the error added to the error-free matrix. Thus it appears that if the model applies to a data-set, estimates of the parameters can be obtained which will be as reliable as the error inherent in the choice probabilities. Exact recovery of the parameter values can be obtained for error-free data.

In deriving equation (3), it was pointed out that if stimulus values are lognormally distributed and a log psychophysical transformation is assumed, the momentary psychological magnitudes will be distributed normally with variances that depend on k and on both the stimulus and perceptual variances. If the stimulus variances and means are known, and if a common perceptual variance is assumed, it will be necessary to estimate only two parameters: k and σ_{2i}. The same parameter estimation procedure just described for (4) can be used to estimate the parameters of (3). Equation (3) itself can be computed using one of the many library routines available for $\Phi(z)$. For a particular matrix of choice probabilities, the best fitting transformation function of the two described in this paper, the power or log transformation, can be determined by comparing the respective residual mean squares.

Discussion

In some sense modalities, such as in the chemical senses, it is difficult, if not impossible, to eliminate stimulus variance. Where stimulus variance exists one cannot, without a formal model, distinguish between stimulus and psychological contributions to the choice probabilities. The purpose of this paper was to attempt to address this problem by deriving models that provide a means of relating stimulus and perceptual parameters to experimental results such as choice probabilities in a paired comparison task. These models have been termed *Fechner-Thurstone* models because they build on the ideas of Fechner with regard to psychophysical transformations from physicochemical values to mental representations and Thurstone's ideas concerning perceptual variance and scaling without regard for the physical continuum. Some computational aspects of the models are given using numerical quadrature. For a matrix of 45 choice probabilities derived from 10 lognormal stimulus distributions, the power exponent, β, and the perceptual standard deviation, σ_{2i}, were recovered using the method of nonlinear least squares for both error-free and error-laden data.

An important feature of this class of models is that since they may make use of known information about the physicochemical characteristics of the stimuli, they require very few parameters to model choice probabilities. For instance, in the sample problem only two parameters were needed to model 45 paired comparison probabilities. The number of parameters would always be two, irrespective of the number of stimuli compared. A Thurstone Type V model would have required 9 parameters (one for each mean except one which could have been set to zero) and this number of parameters would increase with the number of stimuli used. In a Thurstone Type III model, with unequal perceptual variance, one would need to estimate 18 parameters compared with only 11 parameters for this model (all 10 variances would be required because the relationship between stimulus variance and perceptual variance could not be assumed to be known). Of course, if one had lognormal stimulus variance and the psychophysical transformation was other than logarithmic, the Thurstone models would be inappropriate because the momentary psychological magnitudes would not be normally distributed. This is not to say, of course, that the Thurstone models would not fit data very well if the departure from normality was not too great, but the Fechner-Thurstone models discussed in this paper, if the assumptions made are applicable, would always do a better job of modelling the data with fewer parameters.

References

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

Bradley, R.A. (1976). Science, statistics and paired comparisons. *Biometrics 32,* 213.

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. (1988). Confusable and discriminable stimuli: Comments on Nosofsky and Shepard. *Journal of Experimental Psychology: General, 117,* 408-411.

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

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

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

Fechner, G.T. (1860). *Elements of Psychophysics.* In D.H. Howes & E.C. Boring, (Eds.), H.E. Adler, trans. New York: Holt, Rinehart and Winston, 1966.

Frijters, J.E.R. (1979). Variations of the triangular method and the relationship of its unidimensional probabilistic models to three-alternative forced-choice signal detection theory models. *British Journal of Mathematical and Statistical Psychology, 32,* 229-242.

IMSL (1987). *IMSL Library* (Version 1.0). Houston TX: Author. (Developed by IMSL, Inc., 7500 Bellaire Blvd., Houston, TX 77036)

Iverson, G.J. (1987). Thurstonian Psychophysics: Case III. *Journal of Mathematical Psychology, 31,* 219-247.

MacKay, D.B. (1989). Probabilistic multidimensional scaling: An anisotropic model for distance judgments. *Journal of Mathematical Psychology, 33,* 187-205.

Marquardt, D.W. (1963). An algorithm for least squares estimation of nonlinear parameters. *SIAM Journal of Applied Mathematics, 11,* 431-441.

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, 44*, 1169-1175.

Piessens, R., de Doncker-Kapenga, E., Überhuber, C.W., & Kahaner, D.K. (1983). *Quadpack*. New York: Springer-Verlag.

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.

CHAPTER 4

MULTIDIMENSIONAL PROBABILISTIC MODELS

The effect of dimensionality on results from the triangular method. 1985. *Chemical Senses* **10**, 605-608.

A multivariate model for discrimination methods. 1986. *Journal of Mathematical Psychology* **30**, 206-219.

Theoretical aspects of sensory discrimination. 1986. *Chemical Senses* **11**, 513-522.

A multidimensional stochastic theory of similarity. 1988. *Journal of Mathematical Psychology*, **32**, 449-465.

Confusable and discriminable stimuli: Comment on Nosofsky (1986) and Shepard (1986). 1988. *Journal of Experimental Psychology: General* **117**, 408-411.

A comparison of selected probabilistic multidimensional models of identification with respect to perceptual dependence. 1991. *Journal of Mathematical Psychology, submitted.*

Chemical Senses Vol.10 no.4 pp.605-608, 1985

Short Communication

The effect of dimensionality on results from the triangular method

Daniel M.Ennis and Kenneth Mullen[1]

Philip Morris Research Center, Commerce Road, Richmond, VA 23261, USA, and [1]Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario, Canada N1G 2W1

Abstract. Methods to measure differences between complex stimulus sets (such as foods and beverages) are numerous. One of the most commonly used procedures in food and beverage sensory research is the triangular method. A comparison of unidimensional and multidimensional normal models for the triangular method using Monte Carlo simulation showed that the expected subject response distribution depends not only on the size of the unidimensional discriminal distance between stimulus sets, but also on the number of dimensions for which the discriminal distance is zero in each case. Since the number of dimensions for which these conditions apply are usually unknown in complex systems, the power of the triangular method will be unknown. These findings may have important implications for the interpretation of results from many methods which involve a comparison of distance estimates in a multidimensional space.

The triangular method (Amerine *et al.*, 1965; Harrison and Elder, 1950) is a difference testing method which is widely used in food and beverage sensory research in academic and industrial laboratories. In the triangular method, the subject is instructed to select out of three stimuli (two randomly drawn from one stimulus set and one from another stimulus set) the one which is perceptually different from the other two. Unlike m-alternative forced choice methods (such as the method of paired comparisons) in which the subject selects a particular stimulus on the basis of a specified sensory continuum, in the triangular method the subject selects the attribute(s) on which to base the decision.

The normal Thurstone-Ura model for the triangular method has already been described (Frijters, 1979a) in which the subject's response distribution has been related to perceived stimulus dissimilarity on a unidimensional continuum. The sensory values were assumed to be drawn from normal density functions of equal variance. A comparison (Frijters, 1979b) of the normal Thurstone-Ura model for the triangular method and the 3-alternative forced choice model (3-AFC), in which the subject selects the 'strongest' or 'weakest' stimulus on a specified unidimensional sensory continuum, lead to a resolution of the 'paradox of the discriminatory non-discriminators' (Byers and Abrams, 1953; Gridgeman, 1970) for unidimensional continua. This 'paradox' arose when subjects could not identify the odd stimulus in simple and complex systems using the triangular method, but were successful in identifying the 'strongest' or 'weakest' stimulus in a 3-AFC procedure. In the triangular method, the subject uses *distances* between momentary sensory values to make a decision. In the 3-AFC method, the subject uses *absolute magnitudes* of the momentary sensory values to make a decision. The effect of the variance of the sensory values on the subject's success in identifying the 'correct' stimulus is different for the two methods because higher probabilities of correct response are found for the 3-AFC method over a wide range of perceived stimulus dissimilarities (discriminal distance, d').

In experiments designed to test the null hypothesis concerning discrimination between

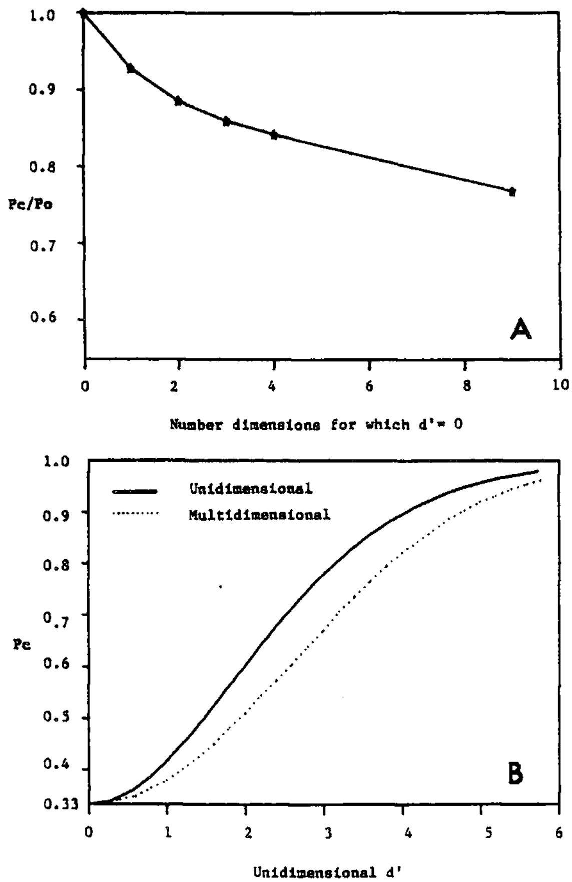


Fig. 1. The effect of increasing stimulus complexity on discrimination in the triangular method relative to the unidimensional case at constant d' (**A**) and across a range of d's (**B**). The experimental conditions, assumptions and decision rules for the unidimensional case were given by Frijters (1979a). The effect of 1, 2, 3, 4 and 9 additional dimensions for which d' =0 in each case was simulated for the case in which unidimensional d' =2.0 (**A**). A Monte Carlo method was employed to provide random deviates from independent normal distributions corresponding to these experimental conditions in which 50 000 trials were evaluated to provide proportion correct estimates for each point. The decision rules for determining a correct response from a unidimensional continuum were applied to the euclidean distances between stimuli in the multidimensional space. The results are plotted as a fraction of the proportion correct response expected for the unidimensional case (Pc/Po, where Pc is the proportion correct for a particular number of dimensions, and Po is the proportion correct in the unidimensional case when d' =2.0). The origin on the y-axis corresponds to the expected value for guessing. In the case of four additional dimensions in which d' =0 in each case (**B**), proportion correct response was estimated for unidimensional values of d' from 0 to 5.8 (0.2 increments) based on 50 000 simulated trials for each point.

complex (multidimensional) stimulus sets, it could be argued that the triangular method enjoys advantages over the paired comparison or 3-AFC methods. The basis for this argument is that the triangular method does not require experimenter knowledge of the sensory dimensions on which the stimulus sets differ and yields an overall measure of discrimination based on the subject's selection and weighting of particular sensory attributes. The problem with this argument is that in multidimensional systems, the subject's ability to discriminate between the stimulus sets will depend not only on the discriminal distance (d') between the means of the sets on continua for which $d' > 0$, but also on the number of dimensions for which, in each case, $d' = 0$ (Figure 1). Figure 1A shows that when the stimulus sets are compared in the triangular method in which the mean of the sensory values on a particular unidimensional continuum differ by 2.0 standard deviations (i.e., $d' = 2.0$), the probability of correctly identifying the odd stimulus (expressed as a fraction of the unidimensional case) decreases with the number of other dimensions for which $d' = 0$ in each case. Figure 1B shows that the effect of the additional dimensions varies with discriminal distance for the case where there are four additional dimensions for which $d' = 0$ in each case.

In deciding which stimulus to choose as the odd one, it was assumed that the subject estimates the multidimensional euclidean distance between the pairs of stimuli in each trial. When these distances are estimated, the subject compares them and selects the stimulus which is most different from the other two as the odd stimulus. An effect of increasing dimensionality is an increase in the variance of the interstimulus distances. The result of this added variance will be an increasing likelihood of choosing the odd stimulus incorrectly. [A corollary of this result is that when the added dimensions actually contain useful information to aid in discrimination (i.e., $d'_1 > 0$), diminishing returns for each added dimension will be evident (if $d'_1 = \text{constant}$)].

Unlike m-alternative forced choice methods (such as the paired comparison and 3-AFC methods), there are several methods used for measuring discrimination in multidimensional systems which require a comparison of distance estimates by the subject, such as the duo-trio and A-not-A methods. Like the triangular method, the power of these methods will be significantly affected by the complexity of the stimuli and will give results which depend on the nature of this complexity. We think that these results have interesting implications in psychometrics and should be of concern to others who draw inferences from sensory measurements. From a practical point of view, it seems reasonable to study the effect of a particular experimental variable in complex systems by measuring some response to the variable under trial conditions in which all other variables are kept 'constant'. These results show that for the triangular and similar methods, the multidimensional context itself is an important determinant of discrimination between stimulus sets. This means that the results obtained from these methods will be specific to the context in which the trial variables were evaluated. In the case of highly multivariate systems, discrimination between stimulus sets which differ in one dimension may require very large discriminal distances and/or very large sample sizes, if the null hypothesis is to be rejected with a reasonable degree of confidence. As it turns out, the argued strength of the triangular method may also be its weakness.

References

Amerine,M.A., Pangborn,R.M. and Roessler,E.B. (1965), *Principles of Sensory Evaluation of Food*, published by Academic Press, NY.

Byers,A.J. and Abrams,D. (1953). A comparison of the triangle and two-sample taste test methods. *Food. Technol.*, 7, 185-187.

Frijters,J.E.R. (1979a). Variations of the triangular method and the relationship of its unidimensional probabilistic models to three-alternative forced choice signal detection theory models. *Br. J. Math. Stat. Psychol.*, **32**, 229-241.

Frijters,J.E.R. (1979b). The paradox of discriminatory nondiscriminators resolved. *Chem. Senses Flavours*, **4**, 355-358.

Gridgeman,M.T. (1970). Re-examination of the two-stage triangle test for perception of sensory differenes. *J. Food Sci.*, **35**, 87-90.

Harrison,S. and Elder,L.W. (1950). Some applications of statistics to laboratory taste testing. *Food Technol.*, **4**, 434-439.

Received March 1985; accepted May 1985

Reprinted from JOURNAL OF MATHEMATICAL PSYCHOLOGY
All Rigts Reserved by Academic Press, New York and London

Vol. 30, No. 2, June 1986
Printed in Belgium

Theoretical Note

A Multivariate Model for Discrimination Methods

DANIEL M. ENNIS

Philip Morris Research Center

AND

KENNETH MULLEN

Department of Mathematics and Statistics, University of Guelph

We describe a multivariate model for a certain class of discrimination methods in this paper and discuss a multivariate Euclidean model for a particular method, the triangular method. The methods of interest involve the selection or grouping of stimuli drawn from two stimulus sets on the basis of attributes invoked by the subject. These methods are commonly used for estimation and hypothesis testing concerning possible differences between foods, beverages, odorants, tastants and visual stimuli.

Mathematical formulation of the bivariate model for the triangular method is provided as well as extensive Monte Carlo results for up to 10-dimensional cases. The effect of correlation structure and variance inequality are discussed. Results from these methods (as probability of a correct response) are not monotonically related to the distance between the means of the stimulus sets from which the stimuli are drawn but depend in a particular way on dimensionality, correlation structure, and the relative orientation of the momentary sensory values in a multidimensional space. The importance of these results to the validity of these methods as currently employed is discussed and the possibility of developing a new approach to multidimensional scaling on the basis of this new theory is considered. © 1986 Academic Press, Inc.

INTRODUCTION

In this paper we discuss a multivariate model for a particular class of discrimination methods. These methods all involve the selection or grouping of stimuli from two stimulus sets on the basis of attributes invoked by the subject. For the most commonly used method, we provide a formula which yields a subject's

We thank J. Tindall, J. Frijters, H. A. David, L. Eby, R. Pangborn, and J. and E. Kapenga for useful criticism and comments; A. Genz for the adaptive routine for numerical integration; M. Waugh, D. Zema, and D. Paulson for computer support; and Y. Lancaster and J. Day for help in preparing the manuscript. Send requests for reprints to Dr. Daniel M. Ennis, Philip Morris Research Center, Commerce Road, Richmond, VA 23261.

0022-2496/86 $3.00

probability of a correct response as a function of population parameters (variance, distance, correlation structure) in the bivariate case and extensive Monte Carlo simulation results for up to 10 dimensions. Our interest in this area is two-fold:

(a) Sensory discrimination methods are widely used to study the discriminability of stimuli which may be perceptually multivariate and, for some of the most commonly used methods, a theoretical framework for interpreting the results does not exist; and

(b) Since the methods of interest do not require experimenter foreknowledge of the attributes selected to make a decision, an understanding of the theoretical relationship between interstimulus distance in a multidimensional space and the subject's response may lead to a new approach to multidimensional scaling.

Examples of the kinds of methods to which the model might apply include the triangular method, the duo–trio method, the ABX method, and the multiple pairs methods. The triangular and duo-trio methods are very commonly used for estimation and hypothesis testing concerning possible differences between foods, beverages, odorants, tastants, and visual stimuli (Amerine, Pangborn & Roessler, 1965; Harrison & Elder, 1950). These methods were adopted without the development of theory and have remained largely outside the field of psychology until recently. In the triangular method, the subject is instructed to select out of three stimuli (two drawn from one stimulus set and one from another stimulus set) the one which is perceptually different from the other two. In the duo–trio method, one of the three stimuli is a designated standard and the subject's task is to identify which of the other two stimuli is perceptually most similar to the standard. Unlike m-alternative forced choice methods (Green & Swets, 1966), such as the method of paired comparisons, in which the subject selects a particular stimulus on the basis of a specified sensory continuum, in the triangular and duo–trio methods, the subject selects the attribute(s) on which to base the decision. The ABX method is similar to the duo–trio method but involves two standards. Multiple pairs methods require the subject to form two equally sized groups of similar stimuli.

THE MULTIVARIATE MODEL

Assumptions

(a) There are two sets of stimuli, S_x and S_y, and within each set, the stimuli are physicochemically identical. Both stimulus sets are sampled, and at least two stimuli are drawn from at least one of the stimulus sets. The stimuli, S_{xi} and S_{yj}, give rise to corresponding momentary sensory values of the respective magnitudes $\mathbf{x}_i$ and $\mathbf{y}_j$ where $\mathbf{x}_i' = (x_{i1}, x_{i2}, \ldots, x_{in})$ and $\mathbf{y}_j' = (y_{j1}, y_{j2}, \ldots, y_{jn})$ where $\mathbf{x}_1'$ indicates the transpose of the vector $\mathbf{x}_1$ and n is the number of sensory dimensions. The momentary sensory values are mutually independently distributed with $\mathbf{x}_i$ having density function $f(\mathbf{x})$ and $\mathbf{y}_j$ having density function $f(\mathbf{y})$.

(b) The probability densities $f(\mathbf{x})$ and $f(\mathbf{y})$ are multivariate normal distributions with means $\boldsymbol{\mu}_x$ and $\boldsymbol{\mu}_y$, where $\boldsymbol{\mu}'_x = (\mu_{x1}, \mu_{x2}, \ldots, \mu_{xn})$ and $\boldsymbol{\mu}'_y = (\mu_{y1}, \mu_{y2}, \ldots, \mu_{yn})$ and variance–covariance matrices $\mathbf{V}_x$ and $\mathbf{V}_y$.

(c) If $\boldsymbol{\mu}_x$ and $\boldsymbol{\mu}_y$ are in standard units, the distance between them is

$$\delta = \left[\sum_{k=1}^{n} |\mu_{xk} - \mu_{yk}|^{\gamma} \right]^{1/\gamma} \qquad \gamma \geqslant 1.$$

(d) In a particular trial, a correct overt response will be obtained if the subject forms two groups in which stimuli drawn from the same set are in only one group. The subject determines group membership by minimizing within group sensation distances.

(e) There are no response preferences due to spatial or temporal positions of the stimuli.

A Multivariate Euclidean Model for the Triangular Method

Unidimensional models for the triangular method have already been developed. The normal Thurstone–Ura model for the triangular method was described by Frijters (1979) in which the subject's response distribution was related to perceived stimulus dissimilarity on a unidimensional continuum. The sensory values were assumed to be drawn from normal density functions of equal variance. Frijters developed the psychometric basis on which the triangular method could be introduced as a signal detection theory method for stimuli with univariate momentary sensory values. Derivation of the relationship between the probability of a correct response (P_c) and the difference between the population means of the density functions representing the stimulus sets (d') had been published earlier by David and Trivedi (1962), which was based on Ura's work (1960).

In the multivariate Euclidean model for this method, $\gamma = 2$ in assumption (c) and assumption (d) will be:

In a particular trial, a correct overt response will be obtained if

$$\text{(i)} \quad \sum_{k=1}^{n} (x_{1k} - x_{2k})^2 < \sum_{k=1}^{n} (x_{1k} - y_k)^2 \textit{ and } \sum_{k=1}^{n} (x_{1k} - x_{2k})^2 < \sum_{k=1}^{n} (x_{2k} - y_k)^2$$

for triangles composed of S_{x1}, S_{x2} and S_y; or if

$$\text{(ii)} \quad \sum_{k=1}^{n} (y_{1k} - y_{2k})^2 < \sum_{k=1}^{n} (y_{1k} - x_k)^2 \textit{ and } \sum_{k=1}^{n} (y_{1k} - y_{2k})^2 < \sum_{k=1}^{n} (y_{2k} - x_k)^2$$

for triangles composed of S_{y1}, S_{y2}, and S_x.

The distance between $\boldsymbol{\mu}_x$ and $\boldsymbol{\mu}_y$ will be represented as d'.

Throughout the rest of this paper we will discuss a mathematical form of this model for the bivariate case, the results of extensive Monte Carlo simulation of the model for up to 10 dimensions, and explain the relevance of this work to current uses of the method and possible new applications. We have chosen the triangular method for illustrative purposes realizing that much of what has been learned with this method will apply to other methods as well.

Mathematical Formulation of the Bivariate Case

Let

$$(x_{1k}-x_{2k})=u_k; \qquad (x_{1k}-y_k)=v_k; \qquad (x_{2k}-y_k)=w_k; \qquad k=1,2.$$

A correct decision will be made if (assumption d(i))

$$\sum u_k^2<\sum v_k^2 \qquad \textit{and} \qquad \sum u_k^2<\sum w_k^2.$$

If $\boldsymbol{\mu}_x=(0,0)$ then the means of the distributions of u_1 and u_2 are zero. If $\boldsymbol{\mu}_y=(\mu_1,\mu_2)$ then the means of v_1 and v_2 are $-\mu_1$ and $-\mu_2$, respectively. $\mathbf{V}$ is the variance–covariance matrix of the joint distribution of u_1, u_2, v_1, v_2. The probability of a correct response, P_c,

$$=2\int_{-\infty}^{\infty} dv_1 \int_{-\infty}^{\infty} dv_2 \left[\int_{2\pi/3}^{4\pi/3} \left\{\int_0^1 f(v_1, v_2, \theta, w)\, dw\right\} d\theta\right]$$

where

$$f(v_1, v_2, \theta, w)=\frac{[w(r_2-r_1)+r_1](r_2-r_1)\exp\{-0.5(\mathbf{z}-\boldsymbol{\mu})'\,\mathbf{V}^{-1}(\mathbf{z}-\boldsymbol{\mu})\}}{(2\pi)^2|\mathbf{V}|^{1/2}}$$

$$r_1=(v_1^2+v_2^2)^{1/2}$$

$$r_2=-2[v_1(\cos\theta\cos\theta_0-\sin\theta\sin\theta_0)+v_2(\sin\theta\cos\theta_0+\cos\theta\sin\theta_0)]$$

$$\cos\theta_0=\frac{v_1}{(v_1^2+v_2^2)^{1/2}}, \qquad \sin\theta_0=\frac{v_2}{(v_1^2+v_2^2)^{1/2}}$$

$$z_1=\{w(r_2-r_1)+r_1\}\{\cos\theta\cos\theta_0-\sin\theta\sin\theta_0\}+v_1$$

$$z_2=\{w(r_2-r_1)+r_1\}\{\sin\theta\cos\theta_0+\cos\theta\sin\theta_0\}+v_2$$

$$z_3=v_1, \qquad z_4=v_2$$

$$\boldsymbol{\mu}'=(0,0,-\mu_1,-\mu_2).$$

(A proof is available from the authors on request)

Numerical Integration and Monte Carlo Simulations

Numerical integration for the bivariate case and Monte Carlo simulations of the general multivariate model were conducted on a Gould 32/97 and a DEC 2060 computer. Numerical integration was accomplished using the IMSL Fortran-callable subroutine DMLIN (IMSL, 1984), and these results were found to be identical to those obtained using a much faster adaptive routine by Genz and Malik (1980). The bivariate model was evaluated over a broad range of parameters and was found to agree closely (differing slightly in the third decimal place) with simulations involving 100,000 triangles per estimate.

The computer simulations, on which the results that follow are based, involved the evaluation of the effect of different values of d' on P_c for different numbers of independent sensory variables (up to 10), the effect of correlation between the dimensions and inequality of variance on P_c in the bivariate case, and an estimate of the power $(1.0-\beta)$ of the triangular method as dimensionality increases. The triangles (three simulated momentary sensory values per triangle) were formed by selecting deviates, using the IMSL subroutines GGNML and GGNSM, from populations with specified parameters. In the case of each estimate 100,000 triangles were sampled; power was estimated by simulating 1000 experiments involving 200 triangles in each case, using a one-tailed α of 0.05.

Different Numbers of Independent Sensory Variables

The variance of each of the independent variables sampled was unity. The unidimensional and multidimensional distances between the means of the stimulus sets are in univariate standard units. Figure 1 shows the relationship between P_c

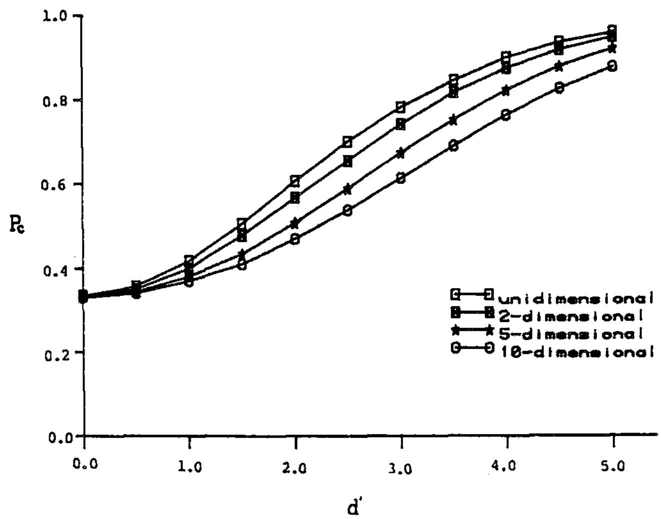


FIG. 1. Probability of a correct response (P_c) as a function of distance (d') for 1-, 2-, 5-, and 10-dimensional stimulus sets.

TABLE 1

Estimated Probability of a Correct Response, P_c, with the Triangular Method as a Function of Discriminal Distance, d'

Euclidean d'	Number of dimensions					
	1	2	3	4	5	10
0.0	0.334	0.335	0.334	0.335	0.335	0.330
0.1	0.335	0.335	0.332	0.335	0.334	0.334
0.2	0.339	0.333	0.339	0.336	0.334	0.334
0.3	0.340	0.339	0.340	0.336	0.338	0.336
0.4	0.345	0.344	0.341	0.340	0.340	0.337
0.5	0.358	0.351	0.348	0.349	0.344	0.341
0.6	0.366	0.361	0.356	0.353	0.352	0.348
0.7	0.376	0.368	0.361	0.359	0.358	0.349
0.8	0.387	0.379	0.372	0.369	0.365	0.355
0.9	0.402	0.390	0.379	0.377	0.372	0.360
1.0	0.418	0.401	0.394	0.384	0.381	0.370
1.1	0.432	0.414	0.403	0.399	0.392	0.374
1.2	0.451	0.428	0.417	0.409	0.402	0.382
1.3	0.470	0.445	0.432	0.420	0.412	0.392
1.4	0.487	0.460	0.444	0.433	0.425	0.402
1.5	0.506	0.479	0.459	0.448	0.435	0.410
1.6	0.527	0.495	0.477	0.461	0.452	0.423
1.7	0.547	0.511	0.491	0.475	0.465	0.432
1.8	0.565	0.530	0.510	0.493	0.479	0.445
1.9	0.584	0.547	0.525	0.509	0.494	0.458
2.0	0.607	0.568	0.542	0.522	0.509	0.471
2.1	0.625	0.582	0.560	0.539	0.526	0.481
2.2	0.641	0.604	0.575	0.558	0.542	0.493
2.3	0.663	0.622	0.591	0.574	0.557	0.511
2.4	0.683	0.641	0.614	0.588	0.572	0.524
2.5	0.700	0.655	0.628	0.607	0.589	0.538
2.6	0.719	0.677	0.647	0.624	0.606	0.551
2.7	0.733	0.692	0.662	0.644	0.626	0.568
2.8	0.748	0.709	0.681	0.659	0.641	0.582
2.9	0.765	0.727	0.697	0.675	0.655	0.598
3.0	0.782	0.741	0.713	0.694	0.673	0.613
3.1	0.797	0.759	0.730	0.709	0.691	0.627
3.2	0.810	0.774	0.746	0.726	0.708	0.643
3.3	0.821	0.788	0.763	0.739	0.722	0.657
3.4	0.834	0.803	0.778	0.756	0.735	0.674
3.5	0.846	0.817	0.791	0.770	0.751	0.689
3.6	0.858	0.829	0.804	0.784	0.766	0.703
3.7	0.869	0.843	0.819	0.797	0.782	0.717
3.8	0.879	0.853	0.830	0.812	0.795	0.732
3.9	0.888	0.864	0.844	0.826	0.808	0.746
4.0	0.899	0.874	0.855	0.837	0.822	0.762
4.1	0.907	0.883	0.866	0.849	0.832	0.774
4.2	0.913	0.895	0.877	0.858	0.845	0.788
4.3	0.922	0.903	0.885	0.871	0.855	0.802
4.4	0.928	0.910	0.895	0.881	0.868	0.814
4.5	0.935	0.918	0.902	0.888	0.879	0.826
4.6	0.939	0.924	0.910	0.898	0.887	0.837
4.7	0.945	0.930	0.916	0.906	0.894	0.847
4.8	0.951	0.937	0.926	0.915	0.904	0.859
4.9	0.956	0.943	0.932	0.921	0.912	0.868
5.0	0.959	0.948	0.938	0.929	0.921	0.877

Note. Each estimate is based on 100,000 simulated triangles.

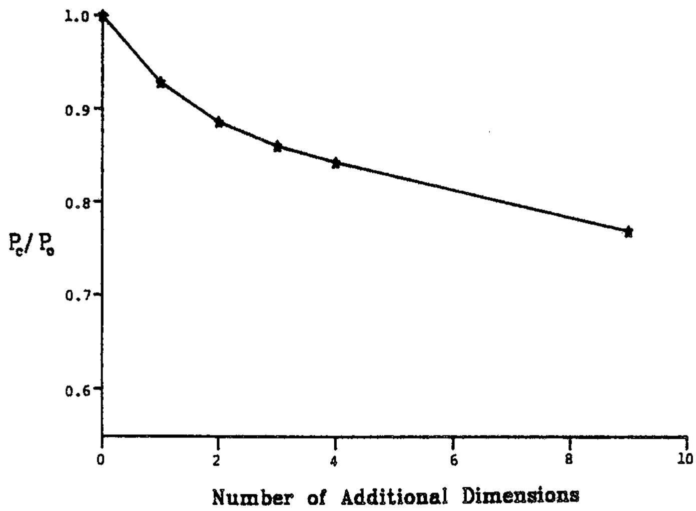


FIG. 2. The effect of increasing stimulus dimensionality on the probability of a correct response (P_c) relative to the unidimensional probability (P_0) when unidimensional $d' = 2$.

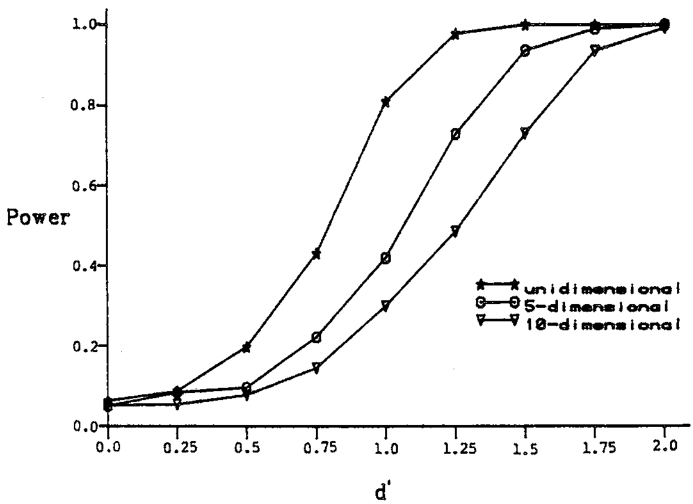


FIG. 3. The power ($1.0 - \beta$) of the triangular method to detect differences between unidimensional and multidimensional stimulus sets when $\alpha = 0.05$ (one-tailed). Each sample contains 200 triangles.

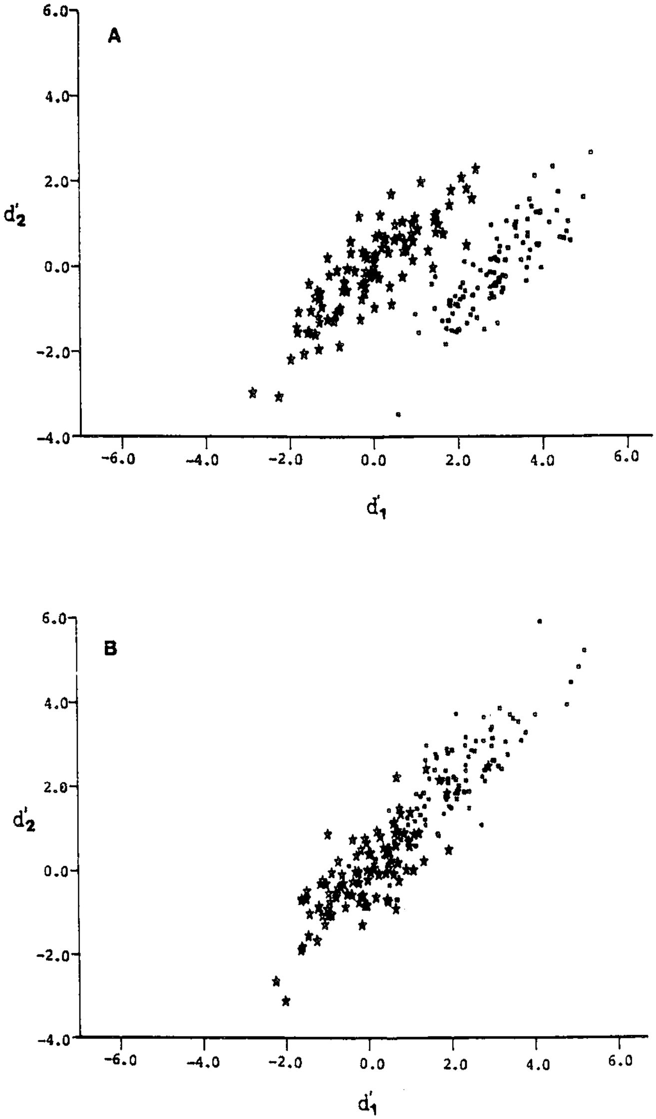


FIG. 4. (A–E) Randomly sampled coordinates from stimulus sets which differ by $d' = 3.0$ and within which the correlation coefficient between variables is 0.8.

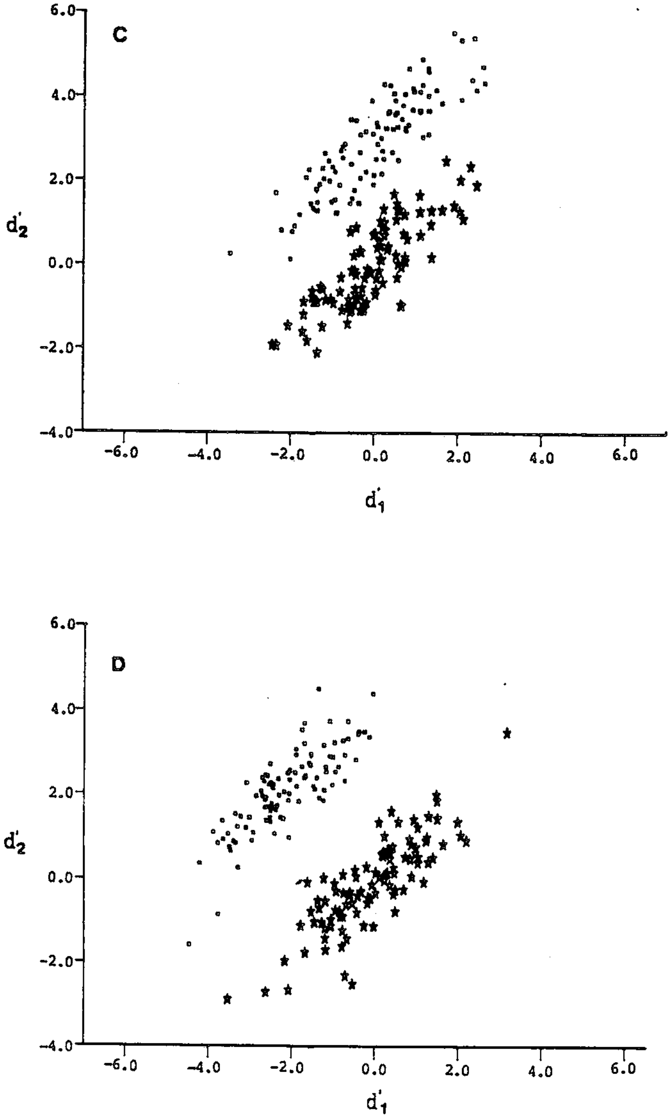


FIG. 4—*Continued.*

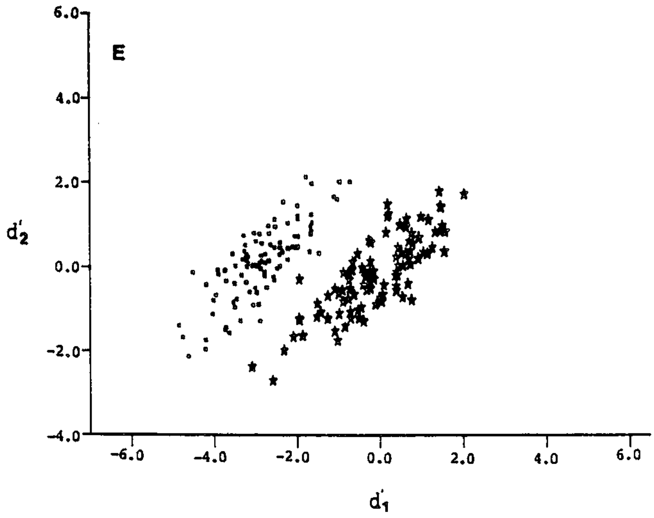


FIG. 4.—*Continued.*

and d' as the number of variables increases from 1 to 10 and Table 1 presents estimates of P_c for a range of d' and for 1, 2, 3, 4, 5, and 10 independent variables. For a given d', the effect of increasing dimensionality is to reduce P_c. This is a real effect of dimensionality on results from this method, the consequence of which cannot be captured on some new unidimensional continuum derived for each multivariate case. If stimuli have multivariate sensory attributes but the stimulus sets differ, on the average, on only one of those dimensions, P_c will decay as a function of dimensionality as shown in Fig. 2 where P_c/P_0 (P_0 is P_c in the unidimensional case) is plotted against the number of added dimensions on which there is no difference between the stimulus sets. This means that P_c is not monotonically related to d' for stimulus sets which differ in dimensionality. These results also anticipate Fig. 3 which shows that the power of the triangular method also depends on dimensionality.

CORRELATED VARIABLES (TWO DIMENSIONS)

When the variables are independent (correlation coefficient, $\rho = 0.0$), the probability of a correct response will depend on the distance between the means of the stimulus sets and the number of variables involved in the distance estimate. It will not depend on the relative contribution of different dimensions to d'; in other words, it will not depend on the relative orientation of the stimulus sets in a multidimensional space. When the variables are correlated within a stimulus set, P_c will depend on the degree of correlation and on the relative direction of the difference between sets. In order to illustrate this effect, consider Figs. 4A – E in which 100 stimuli coordinates have been randomly drawn in each case from stimulus sets which differ by 3.0 standard units ($d' = 3.0$) in two dimensions and for which $\rho = 0.8$ within each set. In Fig. 4A, the means of the stimulus sets differ on the x-axis only (0°); in 4B, they differ equally on x- and y-axes (45°); in 4C, on the y-axis only (90°); in 4D, on x- and y-axes equally (135°); and in 4E, on the x-axis only (180°). Figure 5 is a plot of P_c as a function of orientation when $d' = 3.0$, the number of dimensions is 2, and $\rho = 0.0$, 0.4, and 0.8 within each stimulus set. As expected, when $\rho = 0.0$, P_c is constant. However, as ρ increases, the effect of the direction of d' increases. These results show that P_c may not vary monotonically with the distance between stimulus sets but will depend not only on dimensionality, but also on the particular relative orientation of the perceived attributes in a multidimensional space. In addition to possible difficulties with hypothesis testing concerning the difference between stimulus sets for which these parameters are unknown, these results also suggest the possibility of learning something about the multidimensional representation of objects for which pairwise P_c values are known.

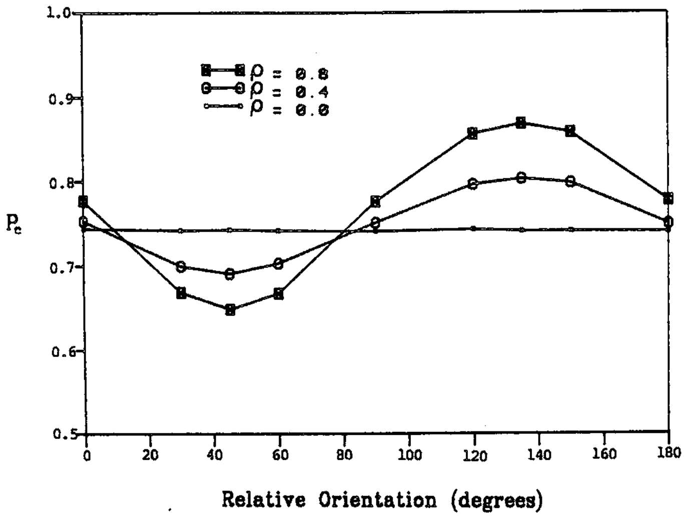


FIG. 5. Probability of a correct response (P_c) for different relative orientations of stimulus coordinates in a 2-dimensional space when correlation coefficients (ρ) are 0.0, 0.4, and 0.8 and discriminal distance (d') between stimulus sets = 3.0.

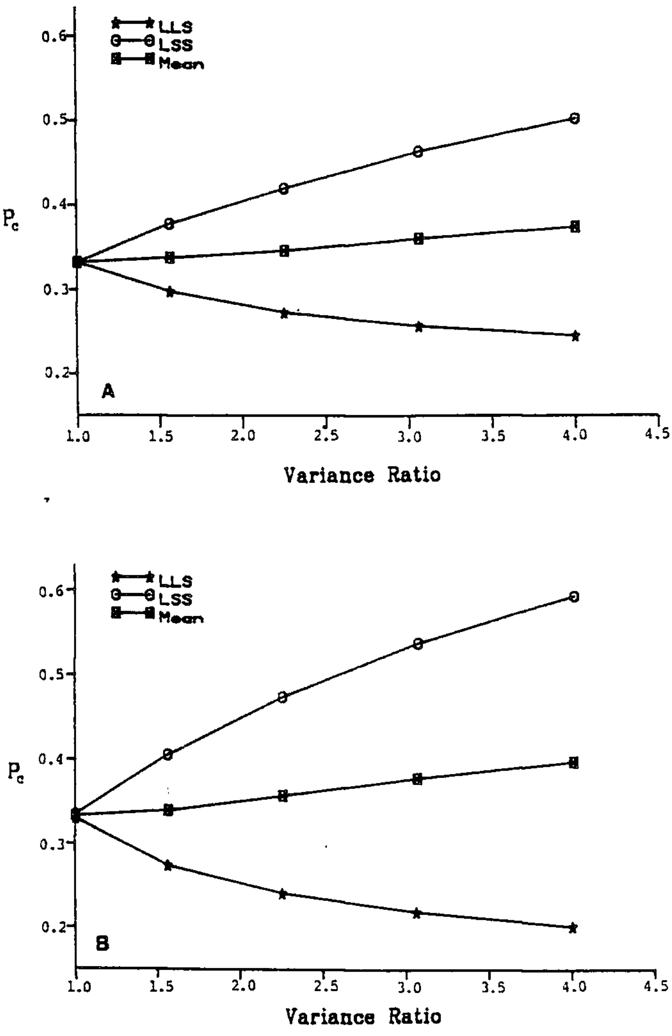


FIG. 6. (A and B) The effect of unequal variances (expressed as the ratio of the variances) on the probability of a correct response (P_c) in 1- and 2-dimensional cases (A and B, respectively) in which the odd stimulus is drawn from the stimulus set with the large variance(s) (LSS), the small variance(s) (LLS), and their mean.

UNEQUAL VARIANCES (TWO DIMENSIONS)

For the sake of simplicity, variances within the 2-dimensional stimulus sets will be assumed to be equal (i.e., both will be large or both will be small). The triangular method specifies an experimental conditions in which two stimuli will be drawn from one stimulus set and one for the other for half of the triangles, and visa versa for the other half. When the variances of the stimulus sets are unequal, these two experimental sets will lead to different values of P_c when $d' = 0$, as shown in Figs. 6A and B for unidimensional and 2-dimensional cases. When the single stimulus is drawn from the population with the larger variance, P_c increases with increasing difference between the variances; this effect is more pronounced in the 2-dimensional case than in the unidimensional case. In contrast, when the single stimulus is drawn from the population with a lower variance, P_c decreases. The average results show that when both experimental conditions are balanced, P_c gradually increases with an increase in the relative size of the larger variance leading to an increased likelihood of committing a Type I error as shown in Table 2.

Although these results have consequences for hypothesis testing, they also offer the interesting possibility of investigating stimulus dimensionality by selectively changing the relative variance of the sensory attributes of the stimulus sets by allowing the subject more exposure to one stimulus set than the other. By conducting experiments of this type, it may be possible to determine dimensionality since the number of dimensions determines the way in which variance inequality affects the results.

TABLE 2

The Effect of Increasing the Variance Ratio of Sampled Populations Using the Triangular Method on Real α Levels When a One-Tailed α of 0.05 Is Assumed

	Unidimensional		2-dimensional	
Variance ratio	P_c	Real α	P_c	Real α
1.0	.332	.05	.333	.05
1.56	.338	.06	.340	.08
2.25	.347	.11	.358	.19
3.06	.361	.21	.379	.40
4.00	.374	.34	.400	.63

Note. Each experiment involves 200 triangles.

DISCUSSION

In this paper we discussed a multivariate model for a general class of discrimination methods. For one of these methods, the triangular method, we presented a multivariate Euclidean model, gave the mathematical form of the model in the bivariate case and extensive Monte Carlo results for a selection of multivariate cases. When the number of sensory dimensions is greater than 1, the probability of a correct response, P_c, for the triangular and the other discrimination methods in this class is not monotonically related to d', the Euclidean distance between the means of the stimulus sets. Although this may present the experimenter with some difficulties in using these methods for hypothesis testing and estimation when certain parameters such as relative variance, dimensionality, and correlation structure are not known, it may also present the opportunity to develop a new approach to multidimensional scaling. This would require the mathematical formulation of the general multivariate model for at least one of these methods and the basis for obtaining a least constrained configuration to correspond with the experimentally determined pairwise P_c values.

REFERENCES

AMERINE, M. A., PANGBORN, R. M., & ROESSLER, E.B. (1965). *Principles of sensory evaluation of food.* New York: Academic Press.

DAVID, H. A. & TRIVEDI, M. C. (1962). *Pair, triangle and duo–trio tests.* Technical Report 55, Department of Statistics, Virginia Polytechnic Institute, Blacksburg, VA.

FRIJTERS, J. E. R. (1979). Variations of the triangular method and the relationship of its unidimensional probabilistic models to three-alternative forced choice signal detection theory models. *British Journal of Mathematical and Statistical Psychology* **32**, 229–241.

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 Computational and Applied Mathematics* **6**(4), 295–302.

GREEN, D. M. & SWETS, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

HARRISON, S. & ELDER, L. W. (1950). Some applications of statistics to laboratory taste testing. *Food Technology* **4**, 434–439.

IMSL (1984). IMSL Library (Version 9.2), developed by IMSL, Inc., 7500 Bellaire Blvd., Houston, TX 77036.

URA, S. (1960). Pair, triangle and duo–trio tests. *Reports of Statistical Application Research, Union of Japanese Scientists and Engineers* 7, 107–119.

RECEIVED: September 27, 1985

Printed by the St. Catherine Press Ltd., Tempelhof 41, Bruges, Belgium

Chemical Senses Vol.11 no.4 pp.513–522, 1986

Theoretical aspects of sensory discrimination

Daniel M.Ennis and Kenneth Mullen[1]

Philip Morris Research Center, Commerce Road, Richmond, VA 23261, USA and [1]Department of Mathematics and Statistics, University of Guelph, Guelph, Ontario N1G 2W1, Canada

Abstract. Discrimination methods are commonly used in research on the chemical senses. Chemosensory stimulation often leads to multivariate sensations. A multivariate theory for a class of discrimination techniques is described here along with a discussion of the practical implications of the theory to experimenters. One of the most important findings is that the probability of a correct response for the discrimination tasks modelled is not monotonically related to the Euclidean distance between the means of the populations from which the stimuli are drawn. There are two important consequences which flow from this finding. First, the power of these methods cannot be determined without specifying values for the multivariate parameters. Second, the traditional assumption in multidimensional scaling that proximity measures and multivariate distances are monotonically related is invalid. The theory presented here is the basis for a new approach to multidimensional scaling using these discrimination methods without invoking the assumption of monotonicity. The conceptual basis is established to extend the model to include stimulus variation and an explicit assumption, based on Stevens' power function, about the nature of the relationship between the stimulus continuum and the sensation continuum.

Introduction

One approach towards a fundamental understanding of how chemosensory stimulants are perceived is to build and test a theory which reliably models the effect of the perceptual parameters on performance. Very often the sensations resulting from chemosensory stimulation by odorants and tastants are multivariate and, consequently, a viable model must specify the relationship between the multivariate parameters and the subject's response. The multivariate parameters of interest here are: (i) the multidimensional distance between stimuli; (ii) the relative orientation of the stimuli to each other in a space of a certain dimensionality; (iii) the degree of correlation between the variables perceived by the subject for a particular stimulus; and (iv) their relative variances. The multivariate parameters may be estimated on the basis of various behavioral data (e.g. probability of a correct response, probability of confusing one stimulus with another, identification errors, and so on) assuming that the other factors influencing performance in an experiment are controlled by the experimenter. Knowing these parameters will provide insights into how the stimuli may be perceived by the subject. An improved understanding of the chemical and physical basis for olfactory and gustatory quality will result from relating sensation magnitudes for a stimulus to its physicochemical properties.

Since Shepard (1962, 1963) and Kruskal (1964a,b) made their original contribution, much of the theory of multidimensional scaling has been based on the assumption that proximity measures and perceptual distances are monotonically related (maintain a rank-order relationship). Thus, the behavioral response (a measure of proximity) is assumed to depend on only a single parameter, the multidimensional distance, δ. Building on this theory, different assumptions have been made about the nature of the distance metric (for instance, city-block or Euclidean), the relative weights given to different dimensions by subjects in determining δ, and even stronger assumptions about the relationship between δ and the proximity measure.

Recently it has been shown (Ennis and Mullen, 1986) that the monotonicity assumption is invalid under many multivariate conditions when methods are used which involve grouping stimuli on the basis of their similarity. Although the monotonicity assumption may hold up in certain cases, it should certainly not be invoked as a general rule. A theory which relates the multivariate parameters to the measure of proximity would free us from the need to assume any particular relationship between δ and the behavioral response.

Multivariate theory of the type discussed here can be derived and evaluated experimentally for a wide variety of different methods. We chose to focus our attention on a particular class of grouping techniques, which include the triangular and duo–trio methods, because they are extensively used for hypothesis testing and estimation in taste and olfaction. The triangular and duo–trio methods were introduced as sensory discrimination methods about 45 years ago and have been used to measure and test for possible differences between food and beverage formulations in addition to their application to more fundamental research on olfaction and taste. In the triangular method, the subject is instructed to select out of three stimuli (two drawn randomly from one stimulus set and one from another stimulus set) the one which is perceptually different from the other two. In the duo–trio method, one of the three stimuli is a designated standard and the subject's task is to identify which of the other two stimuli is perceptually most similar to the standard. These methods remained largely outside the field of psychology until Frijters (1979) linked the triangular method with the broader field of signal detection theory through his work on the unidimensional, or Thurstone–Ura, model. The statistical and psychological framework for making this contribution has been established earlier by David and Trivedi (1962) and was based on Ura's ideas (1960).

The ideas developed here are quite general and can also be applied to other behavioral measures, such as confusion matrices and identification errors.

The general multivariate model

The methods of interest all involve the grouping of stimuli drawn from two stimulus sets on the basis of attributes invoked by the subject. Within each stimulus set, the stimuli are physicochemically identical. It is assumed that the subject minimizes within group sensation distances as a basis for grouping the stimuli and that 'distance' is exactly specified. As Frijters (1979) and others assumed for the unidimensional model, it is assumed that all of the variation in making discrimination judgments results from error by the subject in estimating sensation magnitudes of physicochemically identical stimuli. An extension of the model to include stimulus variation will be considered below. Unlike m-alternative forced choice methods (Green and Swets, 1966), such as the method of paired comparisons, all of the methods in the class of interest assume that the subject selects the attribute(s) on which to base the decision. In addition to the triangular and duo–trio methods, the model also applies to the ABX method and multiple pairs methods. The ABX method is similar to the duo–trio method but involves two standards. The subject's task is to identify the standard (A or B) to which the third stimulus (X) is most similar. Multiple pairs methods require the subject to form two equally sized groups of similar stimuli. For instance, the subject's task may be to form two groups of four stimuli which had been drawn from different stimulus sets.

The assumptions underlying the model are as follows [Ennis and Mullen (1986)]:

(a) There are two sets of stimuli, S_x and S_y, and within each set the stimuli are physicochemically identical. Both stimulus sets are sampled, and at least two stimuli are drawn from at least one of the stimulus sets. The stimuli, S_{xi} and S_{yj}, give rise to corresponding momentary sensory values of the respective magnitudes $\mathbf{x}_i$ and $\mathbf{y}_j$ where $\mathbf{x}'_i = (x_{i1}, x_{i2}, \ldots, x_{in})$ and $\mathbf{y}'_j = (y_{j1}, y_{j2}, \ldots, y_{jn})$ where $\mathbf{x}'$ indicates the transpose of the vector, $\mathbf{x}$ and n is the number of sensory dimensions. Although the stimuli are physicochemically identical within each stimulus set (S_x or S_y), the multivariate momentary sensory values corresponding to stimuli from the same set, or the same stimulus at another point in time, will not be identical because of variation in the subject's sensation magnitude estimates. The momentary sensory values are mutually independently distributed with $\mathbf{x}_i$ having density function $f(\mathbf{x})$ and $\mathbf{y}_j$ having density function $f(\mathbf{y})$.

(b) The probability densities $f(\mathbf{x})$ and $f(\mathbf{y})$ are multivariate normal distributions with means $\underline{\mu}_x$ and $\underline{\mu}_y$, where $\underline{\mu}'_x = (\mu_{x1}, \mu_{x2}, \ldots, \mu_{xn})$ and $\underline{\mu}'_y = (\mu_{y1}, \mu_{y2}, \ldots, \mu_{yn})$, and variance–covariance matrices $\mathbf{V}_x$ and $\mathbf{V}_y$. By specifying different variance–covariance matrices, stimulus sets with particular correlation structures can be considered. For instance, in the case of stimulus sets with two sensory dimensions, the correlation coefficient, ϱ_1, relating the dimensions for the first stimulus set might be 0.8, and the correlation coefficient, ϱ_2, relating the dimensions for the second stimulus set might be 0.6. Similarly, cases where the variances on different dimensions are unequal can be considered.

(c) If $\underline{\mu}_x$ and $\underline{\mu}_y$ are in standard units, the distance between them is

$$\delta = [\sum_{k=1}^{n} |\mu_{xk} - \mu_{yk}|^{\gamma}]^{1/\gamma} \qquad \gamma \geq 1.0$$

The same values of δ can be obtained for different dimensional contributions to $\underline{\mu}_x$ or $\underline{\mu}_y$. In other words, δ can be computed for different orientations of the stimulus sets to each other in a multidimensional space. When $\gamma = 2$, the above distance corresponds to the familiar Euclidean distance. However, other proximity measures may be appropriate.

(d) The particular procedure adopted by the subject to determine group membership is called the decision rule. In a particular trial, a correct overt response (an overt response occurs when a particular decision rule has been invoked) will be obtained if the subject forms two groups in which stimuli drawn from the same set are in only one group. The subject determines group membership by minimizing within group sensation distances.

(e) There are no response preferences due to spacial or temporal positions of the stimuli.

Multivariate Euclidean models

The triangular method

In the multivariate Euclidean model for this method, $\gamma = 2$ in assumption (c), and assumption (d) will be:

in a particular trial, a correct overt response will be obtained if

(i) $\sum^{n}(x_{1k} - x_{2k})^2 < \sum^{n}(x_{1k} - y_k)^2$ and $\sum^{n}(x_{1k} - x_{2k})^2 < \sum^{n}(x_{2k} - y_k)^2$ for triangles composed of S_{x1}, S_{x2} and S_y; or if

(ii) $\sum^{n}(y_{1k} - y_{2k})^2 < \sum^{n}(y_{1k} - x_k)^2$ and $\sum^{n}(y_{1k} - y_{2k})^2 < \sum^{n}(y_{2k} - x_k)^2$ for triangles composed of S_{y1}, S_{y2} and S_x.

This means that if the perceived Euclidean distance between stimuli drawn from the same stimulus set is smaller than both of the other two distances, the subject will make a correct choice by grouping the two physicochemically identical stimuli and declaring the other stimulus to be different from the other two.

Let

$$(x_{1k} - x_{2k}) = u_k;\ (x_{1k} - y_k) = v_k;\ (x_{2k} - y_k) = w_k;\ k=1,\ldots,n$$

where n is the number of sensory dimensions.

A correct decision will be made if [from assumption d(i)]

$$\sum^{n} u_k^2 < \sum^{n} v_k^2 \text{ and } \sum^{n} u_k^2 < \sum^{n} w_k^2$$

If $\underline{\mu}'_x = (0,0,\ldots,0)$ then the means of the distribution of $u_1, u_2,\ldots,u_n$ are each zero. If $\underline{\mu}'_y = (\mu_1, \mu_2,\ldots,\mu_n)$ then the means of the distributions of $v_1, v_2,\ldots,v_n$ are $-\mu_1$, $-\mu_2,\ldots, -\mu_n$ respectively, as are the means of $w_1, w_2,\ldots,w_n$. $\mathbf{V}$ is the variance–covariance matrix of the joint distribution of $u_1,u_2,\ldots,u_n,v_1,v_2,\ldots v_n$. It has been shown (Mullen and Ennis, 1986) that the probability of a correct response, P_c, corresponds to twice the hypervolume inside the n-dimensional hypersphere $\sum^{n} u_k^2 = R^2$ centered at $(0,0,\ldots,0)$, (where $R^2 = \sum^{n} v_k^2$) and outside another n-dimensional hypersphere of the same radius, centered at $(v_1,v_2,\ldots,v_n)$ weighted by the multivariate normal distribution, or

$$P_c = 2 \int \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int\int \cdots \int_{C} f(u_1,u_2,\ldots,u_n,v_1,v_2,\ldots,v_n)du_n,\ldots, du_2,du_1,dv_n,\ldots,dv_2,dv_1$$

where C is the region for which $[\sum^{n} u_k^2 < \sum^{n} v_k^2 \text{ and } \sum^{n}(u_k - v_k)^2 > \sum^{n} v_k^2]$ and

$$f(u_1,u_2,\ldots,u_n,v_1,v_2,\ldots,v_n) = \frac{\exp[-0.5(\mathbf{z}-\underline{\mu})'\ \mathbf{V}^{-1}(\mathbf{z}-\underline{\mu})]}{(2\pi)^n|\mathbf{V}|^{1/2}}$$

$\mathbf{z}' = (u_1,u_2,\ldots,u_n,v_1,v_2,\ldots, v_n)$ and

$\underline{\mu}' = (0,0,\ldots,0,-\mu_1,-\mu_2,\ldots,-\mu_n)$.

The duo–trio method

As in the triangular method, $\gamma = 2$, but assumption (d) is:

in a particular trial, a correct overt response will be obtained if

(i) $\sum^{n}(x_{1k}-x_{2k})^2 < \sum^{n}(x_{1k}-y_k)^2$ if S_{x1} is the standard, or if

(ii) $\sum^{n}(x_{1k}-x_{2k})^2 < \sum^{n}(x_{2k}-y_k)^2$ if S_{x2} is the standard.

Corresponding assumptions are made when S_{y1} or S_{y2} are designated standards. The decision rule means that if the perceived distance between the designated standard and the stimulus drawn from the same set is smaller than the distance between the standard and the third stimulus, the subject will make the correct grouping.

In terms of u,v, and w, a correct decision will be made if

$$\sum^{n} u_k^2 < \sum^{n} v_k^2 \text{ or } \sum^{n} u_k^2 < \sum^{n} w_k^2$$

depending on whether S_{x1} or S_{x2} is the designated standard. As shown in Mullen and Ennis (1986), P_c corresponds to the volume inside the n-dimensional hypersphere $\sum^{n} u_k^2 = R^2$ centered at (0,0,. . .,0)(where $R^2 = \sum^{n} v_k^2$) weighted by the multivariate normal distribution, or

$$Pc = \int \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \int\int \ldots \int_{C} f(u_1,u_2. . .,u_n,v_1,v_2,. . .,v_n)du_n,. . .,du_2,du_1,dv_n,. . .,$$

$$. . .,dv_2,dv_1$$

where C is the region for which $[\sum^{n} u^2 < \sum^{n} v^2]$
and
$f(u_1,u_2,. . .,u_n,v_1,v_2,. . .,v_n)$, $\mathbf{z}'$, and $\underline{\mu}'$ are as defined for the triangular method.

Evaluation of the models

A two-stage process has been adopted in evaluating the models for discrimination methods. In the first stage, the effect of the parameters of the models (δ, variance, correlation structure, dimensionality) on P_c are estimated using Monte Carlo Simulation. These simulations have been conducted on two computers — a Gould 32/97 and a DEC 2060 — using IMSL (1984) Fortran-callable subroutines (GGNML and GGNSM) which, when compared, give identical results. To minimize error, these simulations have been conducted with 100 000 simulated trials per estimate. The second stage involves derivation and evaluation of mathematical forms of the models. The integrals given earlier were evaluated using different approaches. Numerical integration in Cartesian coordinates has been used to evaluate the integrals in the bivariate case using IMSL DBLIN and an adaptive routine by Genz and Malik (1980). These results compare very well with simulation (Mullen and Ennis, 1986) and with a more generalizable form in polar coordinates (Ennis and Mullen, 1986). A complete derivation and evaluation, in spherical coordinates, of the multivariate Euclidean models for the triangular and duo–trio methods is in progress.

Examples

The following are some specifications for possible multivariate parameters and their corresponding probabilities of correct response for the triangular and duo–trio methods. These are all bivariate cases in which the variances for the multivariate normal distribution f(x) are σ_1^2 and σ_2^2, and for f(y) are σ_3^2 and σ_4^2. The means of f(x) and f(y) are (0,0) and (μ_1, μ_2) respectively, and the correlation coefficients relating the dimensions are ϱ_1 and ϱ_2 respectively.

(a) $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, $\sigma_3^2 = 1$, $\sigma_4^2 = 1$, $\varrho_1 = 0.0$, $\varrho_2 = 0.0$, $\mu_1 = 2$, $\mu_2 = 0$, $\delta = 2.0$
P_c (triangular method) = 0.57, P_c (duo–trio method) = 0.71

(b) $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, $\sigma_3^2 = 4$, $\sigma_4^2 = 4$, $\varrho_1 = 0.0$, $\varrho_2 = 0.0$, $\mu_1 = 0$, $\mu_2 = 0$, $\delta = 0.0$

P_c (triangular method) = 0.59, P_c (duo–trio method) = 0.72

(c) $\sigma_1^2 = 4$, $\sigma_2^2 = 4$, $\sigma_3^2 = 1$, $\sigma_4^2 = 1$, $\varrho_1 = 0.0$, $\varrho_2 = 0.0$, $\mu_1 = 0$, $\mu_2 = 0$, $\delta = 0.0$

P_c (triangular method) = 0.20, P_c (duo–trio method) = 0.35

(d) $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, $\sigma_3^2 = 1$, $\sigma_4^2 = 1$, $\varrho_1 = 0.8$, $\varrho_2 = 0.6$, $\mu_1 = 3$, $\mu_2 = 2$, $\delta = 3.6$

P_c (triangular method) = 0.75, P_c (duo–trio method) = 0.86

The first example is the simplest case where variances are constant, the dimensions perceived for each stimulus are not correlated and only one dimension contributes to the distance. Under these conditions, the same value of δ produced by different dimensional contributions [for instance, ($\sqrt{2}$,$\sqrt{2}$) or (0,2)] will result in the same probability of a correct response. Examples (b) and (c) are interesting because in each case the correlation coefficients and the means are zero but the variances on each of the two dimensions for the two stimuli are different. It can be clearly seen that the two presentations possible (S_{x1}, S_{x2} and S_y) or (S_{y1}, S_{y2} and S_x) will give very different results for both methods. Example (d) where variances are constant and the means and correlation coefficients are non zero is probably fairly typical of an actual experimental case. When correlation coefficients are non zero, the relative dimensional contributions to δ will affect the probability of a correct response. The consequences of this are considered in the next section.

P_c is not a monotonic function of δ

Table I gives a selection of multivariate parameter values and corresponding values of P_c for the triangular and duo–trio methods. It can be seen that the probability of a correct response, P_c, varies widely at a constant sensory Euclidean distance, δ, of 3.0 between the population means for the stimuli. Note particularly how P_c is affected by different dimensional contributions to δ when the correlation coefficient between dimensions is 0.8. Compare these results where δ is 3.0 with example (d) above where δ was 3.6 and P_c in the triangular method was 0.75 — an intermediate value for the Table I entries. Previous research (Ennis and Mullen, 1985, 1986), based on extensive Monte Carlo simulation for the triangular method, showed that P_c depends in a particular way on several parameters other than δ. Similar, but unpublished, results for the duo–trio method were obtained. For both methods, as the number of independent sensory dimensions increases, P_c decreases at constant δ. When the sensory dimensions are correlated, P_c depends on the correlation structure and also on the relative contributions of different dimensions to δ.

In view of these theoretical findings, it may seem surprising that the assumption of monotonicity is central to traditional multidimensional scaling. In some cases it may be a reasonable assumption and lead to useful multivariate representations of proximity measures. However, this research points to a need for a deeper understanding of the nature of the relationship between the results from discrimination tasks and the stimulus and psychological factors that determine them. This knowledge would extend our ability to understand the way in which stimuli with multivariate sensory attributes are perceived. Instead of invoking an assumption that is applied to the results from a wide variety of methods, this research suggests that each method involves a model that provides a unique relationship between the multivariate parameters and the behavioral response and that special multivariate specifications are needed to support

Table I. The effect of multivariate parameters on P_c for the triangular and duo–trio methods when δ = 3.0, and variance = 1.0 on all dimensions

Dimensionality	Correlation structure	Dimension contributions to δ	P_c
Triangular method			
1	NA	NA	0.78
2	orthogonal	any	0.74
3	orthogonal	any	0.71
4	orthogonal	any	0.69
5	orthogonal	any	0.67
10	orthogonal	any	0.61
2	$\varrho_1 = 0.8$, $\varrho_2 = 0.8$	$\mu_1 = 3.00$, $\mu_2 = 0.00$	0.78
2	$\varrho_1 = 0.8$, $\varrho_2 = 0.8$	$\mu_1 = 2.60$, $\mu_2 = 1.50$	0.67
2	$\varrho_1 = 0.8$, $\varrho_2 = 0.8$	$\mu_1 = 2.12$, $\mu_2 = 2.12$	0.65
2	$\varrho_1 = 0.8$, $\varrho_2 = 0.8$	$\mu_1 = 1.50$, $\mu_2 = 2.60$	0.67
2	$\varrho_1 = 0.8$, $\varrho_2 = 0.8$	$\mu_1 = 0.00$, $\mu_2 = 3.00$	0.78
2	$\varrho_1 = 0.8$, $\varrho_2 = 0.8$	$\mu_1 = -1.50$, $\mu_2 = 2.60$	0.86
2	$\varrho_1 = 0.8$, $\varrho_2 = 0.8$	$\mu_1 = -2.12$, $\mu_2 = 2.12$	0.87
2	$\varrho_1 = 0.8$, $\varrho_2 = 0.8$	$\mu_1 = -2.60$, $\mu_2 = 1.50$	0.87
2	$\varrho_1 = 0.8$, $\varrho_2 = 0.8$	$\mu_1 = -3.00$, $\mu_2 = 0.00$	0.78
Duo–trio method			
1	NA	NA	0.88
2	orthogonal	any	0.85
3	orthogonal	any	0.83
4	orthogonal	any	0.81
5	orthogonal	any	0.79
10	orthogonal	any	0.75
2	$\varrho_1 = 0.8$, $\varrho_2 = 0.8$	$\mu_1 = 3.00$, $\mu_2 = 0.00$	0.87
2	$\varrho_1 = 0.8$, $\varrho_2 = 0.8$	$\mu_1 = 2.60$, $\mu_2 = 1.50$	0.80
2	$\varrho_1 = 0.8$, $\varrho_2 = 0.8$	$\mu_1 = 2.12$, $\mu_2 = 2.12$	0.78
2	$\varrho_1 = 0.8$, $\varrho_2 = 0.8$	$\mu_1 = 1.50$, $\mu_2 = 2.60$	0.80
2	$\varrho_1 = 0.8$, $\varrho_2 = 0.8$	$\mu_1 = 0.00$, $\mu_2 = 3.00$	0.87
2	$\varrho_1 = 0.8$, $\varrho_2 = 0.8$	$\mu_1 = -1.50$, $\mu_2 = 2.60$	0.91
2	$\varrho_1 = 0.8$, $\varrho_2 = 0.8$	$\mu_1 = -2.12$, $\mu_2 = 2.12$	0.92
2	$\varrho_1 = 0.8$, $\varrho_2 = 0.8$	$\mu_1 = -2.60$, $\mu_2 = 1.50$	0.91
2	$\varrho_1 = 0.8$, $\varrho_2 = 0.8$	$\mu_1 = -3.00$, $\mu_2 = 0.00$	0.87

ϱ_1 and ϱ_2 are correlation coefficients relating the two sensory dimensions characterizing the two stimulus sets, respectively.

a monotonic relationship between the subject's response and the multivariate distance. Estimates of the multivariate parameters can be obtained for a given method by finding a multivariate representation of the stimuli which minimizes stress (disagreement between real and theoretical results) for a set of pairwise P_c values using the theory set out here. This configuration would therefore be obtained without assuming that distances and proximities are monotonically related. This would result in a more fundamental theory of multidimensional scaling.

Aside from their potential use in providing multidimensional representations of stimuli, discrimination techniques can be and, in fact, are used extensively as hypothesis testing methods (is one stimulus significantly different sensorially from another?). The dependence of P_c on dimensionality, the variance–covariance matrices of the stimulus sets, and on the orientation of the stimulus sets to each other in a multidimensional space (relative weights on different dimensions) means that, for a particular comparison in the absence of knowledge about these parameters, the power (1.0-Type II error) of these methods will be unknown and the size of P_c may bear no obvious relationship to the size of δ. These findings suggest that discrimination methods such as the triangular and duo–trio methods should not be relied upon for hypothesis testing, unless a great deal is known about the multidimensional nature of the stimuli of interest. The theory also provides a way of comparing existing and alternative methodologies so that an optimum method can be chosen for a particular set of multivariate parameters since some discrimination techniques will be more sensitive to changes in the multivariate parameters than others.

Extensions of the model

The model that has been discussed so far is applicable to situations where there is no physicochemical variation and the unit of measurement has been the standard deviation on the sensation continuum. It was assumed that the source of this error was in the individual subject's perception of the stimulus, which may vary from moment to moment. In situations where stimulus variation is negligible, this assumption is perfectly valid. However, it is not always possible to control the stimulus so exactly. If one were not concerned about separating stimulus error from subject error, the general discrimination model could be rewritten to include an assumption about stimulus variation and the unit for δ, measured on the sensation continuum, would contain both components. All of the previous theory would then continue to apply, assuming that the distribution

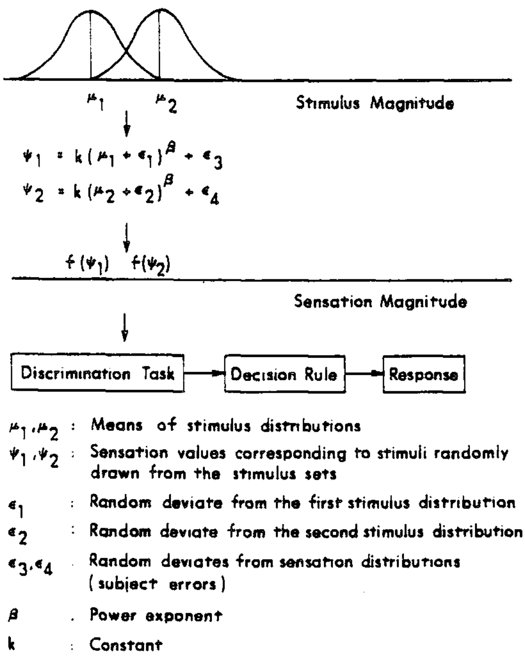


Fig. 1. The conceptual framework for a psychophysical-decision model for discrimination methods.

assumptions were still valid, and δ would simply take on this new meaning.

It would be valuable to be able to separate stimulus from subject error for several reasons. First, error on some stimulus continuum (concentration of a compound, Instron measurements, spectral absorbance, etc.) can be estimated for a particular stimulus. Second, the nature of the distribution on the stimulus continuum may be different from that due to subject error (one may be normal, the other may be log normal, for instance). Third, if the relative size of stimulus to subject variance is known, more replications can be made at the point where variance is highest. It becomes apparent that to include stimulus variation would require a model that relates the stimulus continuum to the sensation continuum. Although many models of this type are conceivable, there is considerable support for the power function (Stevens, 1975) which states that $\psi = k\phi^{\beta}$ where ψ is the sensation magnitude, ϕ is the stimulus magnitude, β is the power exponent (which depends on the nature of the stimulus) and k is a constant.

Figure 1 shows these ideas schematically in the unidimensional case; the multivariate case is conceptually similar. Computer simulations of this model, which will be the subject of a future publication, have shown the interesting possibility that one may be able to estimate β for different kinds of stimuli using one of the discrimination methods given earlier. These results could then be compared with magnitude estimates or cross-modality matching experiments which assume that either magnitude estimates or matched intensities of a stimulus correspond exactly to the sensation magnitudes of the stimulus of interest. This assumption could then be tested without using direct methods for estimating sensation magnitudes and provide a new way of estimating psychophysical functions.

Conclusion

The results from experiments involving discrimination tasks depend on many variables, not all of which are under the control of the experimenter. The models presented here should help to establish the nature of the relationship between the proximity measure obtained from these experiments and a selection of important multivariate parameters. Many of these parameters have not previously been formally included in models for discrimination, with the exception of the multivariate distance, δ. Computer routines have now been written for various methods (including the triangular and duo–trio methods) by the authors to calculate P_c as a function of the multivariate parameters for cases of low dimensionality. These programs also allow the experimenter to compare different methods under particular multivariate scenarios. We are currently working on a fast routine for grouping techniques applicable to the n-dimensional case which will be used as the basis for applying a new theoretical approach to multidimensional scaling.

Acknowledgements

We are especially grateful to Mr C.Lilly and the Philip Morris Visiting Scientist Program for facilitating our collaboration. We thank J.Tindall, J.Frijters, P.Johnson, H.A.David, R.Pangborn and J. and E.Kapenga for useful criticism and comments; A Genz for the adaptive routine for numerical integration; M.Waugh, D.Zema and D.Paulson for computer support and C.Nilles for manuscript assistance.

References

David,H.A. and Trivedi,M.C. (1962) Pair, triangle and duo–trio tests. Technical Report 55, Department of Statistics, Virginia Polytechnic Institute, Blacksburg, Virginia.

Ennis,D.M. and Mullen,K. (1985) The effect of dimensionality on results from the triangular method. *Chem. Senses*, **10**, 605–608.

Ennis,D.M. and Mullen,K. (1986) A multivariate model for discrimination methods. *J. Math. Psychol.*, **30**, 206–219.

IMSL (1984) IMSL Library (Version 9.2), developed by IMSL, Inc., 7500 Bellaire Blvd., Houston, Texas 77036.

Frijters,J.E.R. (1979) Variations of the triangular method and the relationship of its unidimensional probabilistic models to three alternative forced choice signal detection theory models. *Br. J. Math. Stat. Psych.*, **32**, 229–241.

Genz,A.C. and Malik,A.A. (1980) Remarks on algorithm 006: an adaptive algorithm for numerical integration over an N-dimensional rectangular region. *J. Comput. and Appl. Math.*, **6**, 295–302.

Green,D.M. and Swets,J.A. (1966) *Signal Detection Theory and Psychophysics*, John Wiley and Sons, New York, pp. 43–45.

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

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

Mullen,K. and Ennis,D.M. (1986) Mathematical formulation of multivariate Euclidean models for discrimination methods. *Psychometrika*, in press.

Shepard,R.N. (1962) The analysis of proximities: multidimensional scaling with an unknown distance function, I, II. *Psychometrika*, **27**, 125–140; 219–246.

Shepard,R.N. (1963) Analysis of proximities as a technique for the study of information processing in man. *Human Factors*, **5**, 19–34.

Stevens,S.S. (1975) *Psychophysics*, John Wiley and Sons, New York.

Ura,S. (1960) Pair, triangle and duo–trio tests, *Rep. Stat. Appl. Res., JUSE*, **7**, 107–119.

Received March 1986; accepted June 1986

Reprinted from JOURNAL OF MATHEMATICAL PSYCHOLOGY
All Rigts Reserved by Academic Press, New York and London

Vol. 32, No. 4, December 1988
Printed in Belgium

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.

INTRODUCTION

The goal of this paper is to describe a multidimensioal 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.

0022-2496/88 $3.00

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, ${}_iP_{jk}$ 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 ${}_iP_{jk}$ 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 $\mathbf{x}$ and $\mathbf{y}$ where $\mathbf{x}' = (x_1, x_2, ..., x_n)$, $\mathbf{y}' = (y_1, y_2, ..., y_n)$; $\mathbf{x}'$ indicates an n-dimensional row vector and n is the number of psychological 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 $\boldsymbol{\mu}_y$ 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 $\mathbf{z} = \mathbf{x} - \mathbf{y}$.

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

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

The distance between population means is

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

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

$$g(d) = \exp(-d^{\alpha}), \qquad \alpha \geqslant 0.$$

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

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

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

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

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

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

where

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

and $\mathbf{I}$ 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, τ.

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

The formula for calculating $f(\boldsymbol{\mu}, \mathbf{V}, \tau)$ is

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

τ 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 τ 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,

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

where $P(R_j \mid S_i)$ is the probability that stimulus S_i leads to response R_j; b_j and b_k are response bias parameters, $0 \leqslant b_j \leqslant 1$; 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} $(k = 1, ..., m)$. 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,

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

where σ_1^2 and σ_2^2 are the variances of the distributions from which x_1 and x_2 were drawn, respectively; σ_3^2 and σ_4^2 are the variances of the distributions from which y_1 and y_2 were drawn, respectively; ρ_1 is the correlation coefficient between the dimensions of $h(\mathbf{x})$ and ρ_2 is the correlation coefficient between the dimensions of $h(\mathbf{y})$.

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 stimui, 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 δ. 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 δ than it does to d, since it is not expected to vary from trial to trial.

When modelling the relationship between δ and $f(\boldsymbol{\mu}, \mathbf{V})$, it is instructive to consider, for a particular similarity function and metric (α and γ), the effect of the

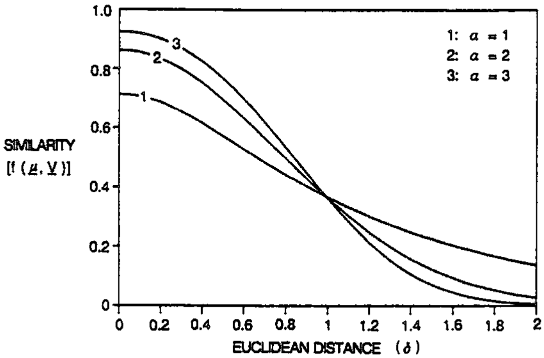


FIG. 1. Expected value of similarity as a function of the Euclidean distance between the means of the distributions of psychological magnitudes for values of α of 1, 2, and 3 in the similarity function $g(d) = \exp(-d^{\alpha})$.

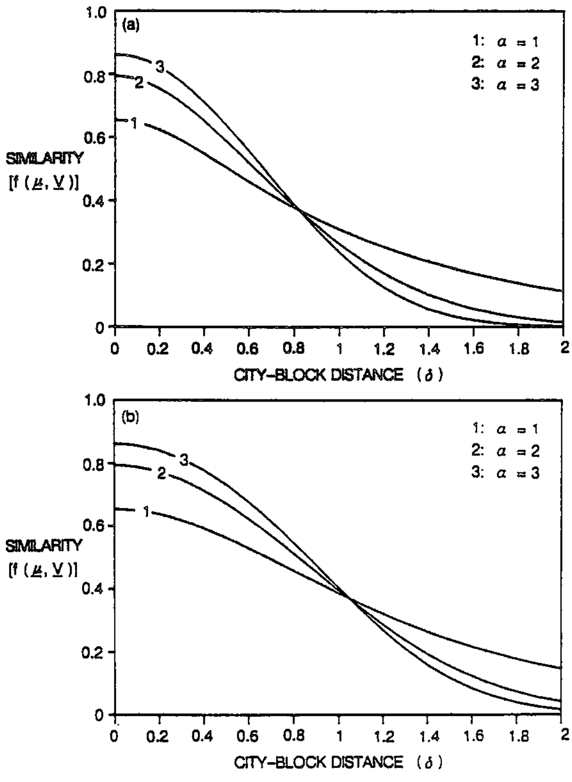


FIG. 2. Expected value of similarity as a function of the city-block distance between the means of the distributions of psychological magnitudes for values of α of 1, 2, and 3 in the similarity function $g(d) = \exp(-d^{\alpha})$. (a) Means differ on one axis only; (b) means differ equally on both axes.

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 δ and $f(\boldsymbol{\mu}, \mathbf{V})$ 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 $f(\boldsymbol{\mu}, \mathbf{V})$ 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 $f(\boldsymbol{\mu}, \mathbf{V})$ and δ 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 ($\alpha = 1$)

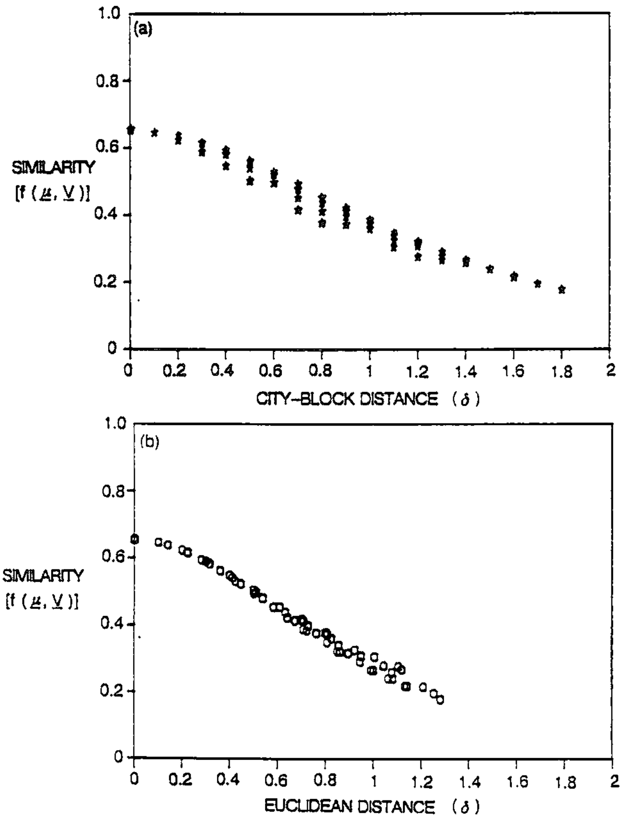


FIG. 3. Expected value of similarity between pairs of 16 stimuli plotted against the city-block and Euclidean distances (δ) between the means of the distributions of psychological magnitudes. An exponential decay function has been used to describe the relationship between city-block distance (d) and 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 δ (the distance between population means) and $f(\boldsymbol{\mu}, \mathbf{V})$ (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 $\ln[f(\boldsymbol{\mu}, \mathbf{V})]$ and δ^2 to the data for both metric forms of δ. 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

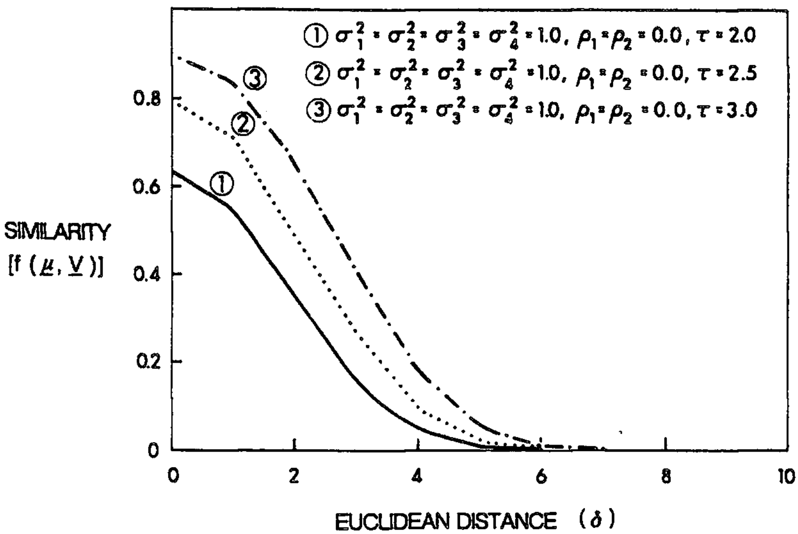


FIG. 4. Expected value of similarity as a function of the Euclidean distance between the means of the distributions of psychological magnitudes for different values of τ when a step judgment function is assumed.

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 δ and $f(\boldsymbol{\mu}, \mathbf{V})$ for step functions where τ 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, τ 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 $f(\boldsymbol{\mu}, \mathbf{V})$ as a function of δ for a given $\mathbf{V}$ matrix, for instance, is to change τ. This will also have the effect of changing the predicted probability of declaring identical objects "same."

Stochastic Multidimensional Scaling

Assuming that α and γ 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 ($\boldsymbol{\mu}$) and the variance–covariance matrix of the difference between psychological values ($\mathbf{V}$). 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 $f(\boldsymbol{\mu}, \mathbf{V})$ 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 $\mathbf{a}$ 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 $\mathbf{a}$, it is simple to compute $\boldsymbol{\mu}_{ij}$ and $\mathbf{V}_{ij}$ (the means of differences and variance–covariance matrix of differences for stimuli S_i and S_j) and, consequently, $f(\boldsymbol{\mu}_{ij}, \mathbf{V}_{ij})$ can then be computed from Eq. (2). The function to be minimized is

$$q(\mathbf{a}) = \sum_i \sum_j \left[P_{ij} - f(\boldsymbol{\mu}_{ij}, \mathbf{V}_{ij})\right]^2, \qquad j \leqslant i,$$

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 1

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	−0.77	−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).

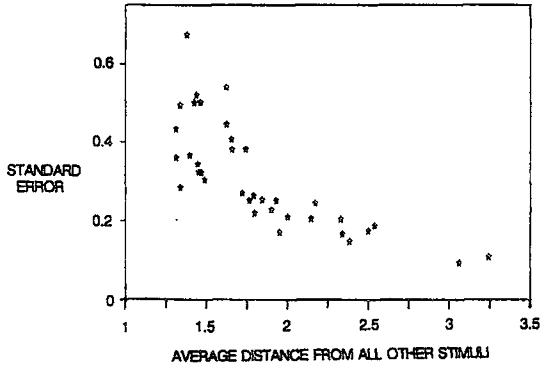


FIG. 5. The standard error on either dimension for 36 Morse code signals plotted in relation to the average Euclidean distance between each stimulus mean and the 35 other stimulus means.

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 α and γ 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 be 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 $\mathbf{x}' = (x_1, x_2, ..., x_n)$, $\mathbf{y}' = (y_1, y_2, ..., y_n)$; $\mathbf{x}'$ 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 $\boldsymbol{\mu}'_x = (\mu_{x1}, \mu_{x2}, ..., \mu_{xn})$ and $\boldsymbol{\mu}'_y = (\mu_{y1}, \mu_{y2}, ..., \mu_{yn})$, 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

$$\begin{aligned} g(d) &= \exp(-d^{\alpha}) \\ &= \exp\left(-\left[\sum_{k=1}^{n} |z_k|^{\gamma}\right]^{\alpha/\gamma}\right) \\ &= G(\mathbf{z}), \end{aligned}$$

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

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

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

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

$$\begin{aligned} f(\boldsymbol{\mu}, \mathbf{V}) &= (2\pi)^{-n/2}\,|\mathbf{V}|^{-1/2} \\ &\quad \times \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \exp[-0.5(\mathbf{z}'\mathbf{J}\mathbf{z} - \mathbf{z}'\mathbf{b} - \mathbf{b}'\mathbf{z} + \boldsymbol{\mu}'\mathbf{V}^{-1}\boldsymbol{\mu})]\,dz_1\cdots dz_n, \end{aligned}$$

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:

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

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

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

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

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

$$\begin{aligned} \mathbf{C}^{-1}\mathbf{J}\mathbf{C} &= \mathbf{D} \text{ and } \mathbf{C}^{-1} = \mathbf{C}^{\mathrm{T}}, \\ \mathbf{z}'\mathbf{J}\mathbf{z} &= \mathbf{z}'\mathbf{C}\mathbf{C}^{-1}\mathbf{J}\mathbf{C}\mathbf{C}^{-1}\mathbf{z} \\ &= \mathbf{z}'\mathbf{C}\mathbf{D}\mathbf{C}^{-1}\mathbf{z} \\ &= (\mathbf{z}'\mathbf{C}\mathbf{D}^{1/2})(\mathbf{D}^{1/2}\mathbf{C}^{\mathrm{T}}\mathbf{z}). \end{aligned}$$

Let

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

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

$$\begin{aligned} dt_1 \cdots dt_n &= (\lambda_1 \lambda_2 \cdots \lambda_n)^{1/2}\, dz_1 \cdots dz_n \\ &= |\mathbf{D}|^{1/2}\, dz_1 \cdots dz_n \\ &= |\mathbf{J}|^{1/2}\, dz_1 \cdots dz_n. \end{aligned}$$

$$\begin{aligned} f(\boldsymbol{\mu}, \mathbf{V}) &= (2\pi)^{-n/2}\, |\mathbf{V}|^{-1/2} \cdot |\mathbf{J}|^{-1/2} \\ &\quad \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp[-0.5(\mathbf{t}'\mathbf{t} + \boldsymbol{\mu}'\mathbf{V}^{-1}\boldsymbol{\mu} - \mathbf{b}'\mathbf{J}\mathbf{b})]\, dt_1 \cdots dt_n \\ &= (2\pi)^{-n/2}\, |\mathbf{V}|^{-1/2}\, |\mathbf{J}|^{-1/2} \exp[-0.5(\boldsymbol{\mu}'\mathbf{V}^{-1}\boldsymbol{\mu} - \mathbf{b}'\mathbf{J}\mathbf{b})] \\ &\quad \times \int_{-\infty}^{\infty} \exp(-t_1^2/2)\, dt_1) \int_{-\infty}^{\infty} \exp(-t_2^2/2)\, dt_2) \cdots \\ &\quad \times \int_{-\infty}^{\infty} \exp(-t_n^2/2)\, dt_n) \\ &= (|\mathbf{V}|\, |\mathbf{J}|)^{-1/2} \exp[\boldsymbol{\mu}'(2\mathbf{J}^{-1} - \mathbf{I})\boldsymbol{\mu}], \end{aligned}$$

where

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

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 coment: 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., & Frijters, 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

Printed by Catherine Press, Ltd., Tempelhof 41, B-8000 Brugge, Belgium

Journal of Experimental Psychology: General
1988, Vol. 117, No. 4, 408–411

0096-3445/88/$00.75

COMMENTS

Confusable and Discriminable Stimuli: Comment on Nosofsky (1986) and Shepard (1986)

Daniel M. Ennis
Philip Morris Research Center, Richmond, Virginia

Nosofsky (1986) modeled identification and categorization performance with highly similar stimuli by using a model involving a "Gaussian" relationship between similarity and Euclidean distance. Nosofsky found that this model fitted his data better than a model involving similarity as an exponential decay function of city-block distance. Nosofsky's stimuli evoked independent perceptual dimensions. Shepard (1986) conjectured that perceptual "noise" may have contributed to the results of Nosofsky's experiments because, in the absence of such noise, previous research suggested that similarity is best modeled as an exponential decay function of city-block distance for stimuli that evoke independent perceptual dimensions. By using a multivariate model of similarity, in which perceptual variation is included, this article provides a possible reconciliation of the kind of result found by Nosofsky (1986) and Shepard's (1986) theory concerning the relationship between similarity and perceptual distance.

In a comment on an article by Nosofsky (1986), Shepard (1986) discussed a theory of identification learning in which the conditional probability that one stimulus, S_i, would elicit the response (R_j) corresponding to another stimulus, S_j, is a monotonically decreasing function of the distance between the points representing the stimuli in a multidimensional space. Because, as Shepard (1987) noted, generalization and similarity arise from the same basic processes, similarity may also be a monotonically decreasing function of distance. On the basis of work with dissimilar or discriminable stimuli, Shepard (1986, 1987) suggested that this function may be an exponential decay function and that the metric for separable stimuli (where the perceptual dimensions are unrelated) may be city block. The term *discriminable* will be used to describe stimuli that cannot be confused because of variation in the mental representations of the stimulus objects. The term *confusable* will be used to describe stimuli for which perceived similarity may vary from moment to moment because of this variation. Pairs of discriminable and confusable stimuli exist on a continuum of similarity, and hence the distinction between discriminable and confusable stimuli is necessarily arbitrary. Nosofsky discussed identification and classification performance by using highly similar or confusable stimuli and reported a "Gaussian" function that modeled the relationship between the Euclidean distance separating the stimulus points and similarity. Shepard (1986) conjectured that "asymptotic performance following protracted discrimination training with highly similar stimuli may be limited by irreducible noise in the perceptual/memory system" (p. 60). In the case of confusable stimuli, he suggested that internal noise may make "the otherwise sharply peaked gradient of generalization . . . more nearly Gaussian" (p. 60).

Mathematical modeling of tristimulus grouping techniques (Ennis & Mullen, 1986a, 1986b; Mullen & Ennis, 1987) with confusable stimuli had led to the development of tools applicable to modeling the similarity of pairs of confusable objects. This work was conducted independently of theoretical developments on identification and categorization performance. The comment by Shepard (1986) and further elaboration of his theoretical position (Shepard, 1987) were important in the development of a stochastic similarity model that I will discuss in this article. This model will be shown to be useful in providing a possible reconciliation of Nosofsky's (1986) findings and Shepard's (1986, 1987) basic hypotheses concerning similarity. This article is intended as a brief comment on a specific issue raised by Nosofsky (1986) and Shepard (1986, 1987). A more complete treatment of the multidimensional similarity model used here can be found in an article by Ennis, Palen, and Mullen (in press). Theoretical work on other multidimensional probabilistic models are of interest (Ashby & Perrin, 1988; De Soete, Carroll, & DeSarbo, 1986; Hefner, 1958; Zinnes & MacKay, 1983).

Similarity

Stimuli S_x and S_y give rise to momentary psychological values of the respective magnitudes **x** and **y**, and n is the number of psychological dimensions. The momentary psychological values are assumed to be mutually independently distributed and drawn from multivariate normal distributions with means μ_x and μ_y and variance–covariance matrixes $\mathbf{V}_x$ and $\mathbf{V}_y$.

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

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

I thank R. Shepard, R. Nosofsky, G. Ashby, K. Mullen, J. Frijters, and P. Arabie for comments and discussions, which proved to be very helpful in developing the ideas leading to this article.

Correspondence concerning this article should be addressed to Daniel M. Ennis, Philip Morris Research Center, P. O. Box 26583, Richmond, Virginia 23261.

Assume that the similarity of the momentary psychological values is a function of d, for example, $g(d)$. The function g corresponds to Nosofsky's (1986) η and Shepard's (1986, 1987) f. Nosofsky (1986) and Shepard (1986, 1987) both used functions of the form

$$g(d) = exp(-d^{\alpha}) \qquad \alpha \geq 0.$$

Nosofsky (1986) and Shepard (1986, 1987) treated d in a deterministic manner because it is the distance between the points representing the stimuli. In terms of the present discussion, this would correspond to the distance between the means of the multivariate normal distributions of momentary psychological values, δ, rather than the momentary values themselves. For confusable stimuli, this is an important distinction, and it will be seen in the next section how the distribution of momentary psychological values affects the relationship between averaged similarity over trials and the distance between the population means. The distance between population means is

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

The subject uses d to make trial-by-trial similarity judgments but never has direct access to δ.

A Multivariate Model for the Similarity of Confusable Stimuli

$\mathbf{V}$ is the variance–covariance matrix of the difference between psychological values, $\mathbf{z}$. μ is a vector of differences between the means of the momentary psychological values, μ_x and μ_y.

A prime sign (′) will be used to represent n-dimensional row vectors. Let $f(\mu, \mathbf{V})$ be the expected value of g for given values of α and γ.

$$f(\mu, \mathbf{V}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \frac{exp\,[-0.5\,(\mathbf{z}-\mu)'\,\mathbf{V}^{-1}\,(\mathbf{z}-\mu)]}{(2\pi)^{n/2}\,|\,\mathbf{V}\,|^{1/2}} \times exp\,(-d^{\alpha})dz_1dz_2\ldots dz_n. \qquad (1)$$

Unlike the monotonic relationship between similarity and distance proposed by Shepard (1986, 1987), and occurring within a trial in this model, there will not be a monotonic relationship between $f(\mu, \mathbf{V})$ and δ, except in special cases (e.g., when $\mathbf{V}$ is an identity matrix). This point has already been discussed for various multidimensional stochastic models (e.g., Ashby & Perrin, 1988; Ennis & Mullen, 1986b).

The Relationship Between $f(\mu, \mathbf{V})$ and δ

Equation 1 was evaluated numerically on a Gould 32/97 computer by using an adaptive routine by Genz and Malik (1980). When n, the number of dimensions, = 2,

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

where σ_1^2 and σ_2^2 are the variances of the distributions from which x_1 and x_2, respectively, were drawn; σ_3^2 and σ_4^2 are the variances of the distributions from which y_1 and y_2, respectively, were drawn; and ρ_1 is the correlation coefficient between the dimensions of $f(\mathbf{x})$, and ρ_2 is the correlation coefficient between the dimensions of $f(\mathbf{y})$.

Figure 1 shows the relationship between δ and $f(\mu, \mathbf{V})$ for different values of α when $\gamma = 2$ (Euclidean metric), $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = 0.2$, and $\rho_1 = \rho_2 = 0.0$. Because the variances are equal and the covariances are zero, different orientations of the means to each other will lead to the same values of $f(\mu, \mathbf{V})$. However, this is not true of the city-block metric. Figure 2a shows the same relationship when $\gamma = 1$ (city-block metric) and the difference in means is on one axis only; Figure 2b shows results for the same city-block distances when the two dimensions contribute equally to the distance. Because I have assumed a multivariate normal distribution for the psychological values in which variances are equal and covariances are zero, the nature of the relationship between δ and $f(\mu, \mathbf{V})$ will be monotonic for the cases discussed and tend to be Gaussian in form, even though the relationship between similarity and distance within a trial may be an exponential decay function. Although this result confirms Shepard's (1986, 1987) conjecture as one possible explanation for Nosofsky's (1986) findings concerning the function g, other forms of g also provide a modified Gaussian relationship between δ and $f(\mu, \mathbf{V})$.

If the similarity function within a trial was an exponential decay function and the subject used a city-block metric to obtain distances, which of the two metrics, city block or Euclidean, would best fit the relationship between similarity averaged over trials $f(\mu, \mathbf{V})$ and δ? I selected 16 hypothetical stimulus means in two dimensions (in a grid similar to Nosofsky's, 1986, stimuli) from a set of means for which $f(\mu, \mathbf{V})$ was in the range of 0.5 to 1.0 when it was assumed that the standard errors on each dimension for each stimulus distribution were 0.2 and the correlation coefficients between dimensions for each stimulus were 0.0. Figures 3a and 3b show the relationship between δ and $f(\mu, \mathbf{V})$ for the set of 16 stimuli.

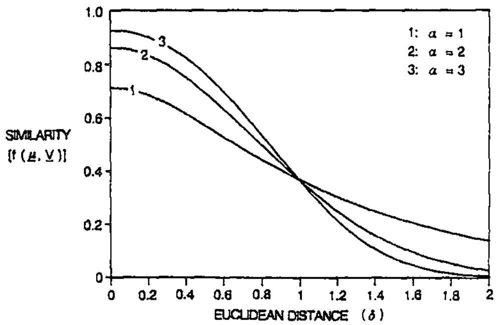


Figure 1. Expected value of similarity as a function of the Euclidean distance between the means of the distributions of psychological magnitudes for values of α of 1, 2, and 3 in the similarity function $g(d) = \exp(-d^{\alpha})$.

Inspection of these figures and a least squares analysis (a linear regression analysis of $\log_e[f(\mu, V)]$ and δ^2) suggest that the Euclidean metric leads to a fit of the data ($r^2 = .98$) that is at least as good as the city-block metric ($r^2 = .95$). It is possible, consequently, to qualitatively reconcile Nosofsky's findings with those of Shepard's (1986, 1987) regarding the form of the metric.

Conclusion

In this article, I gave an equation that combines a model of the process that produces the momentary within-trial psychological magnitudes with a model of similarity perceived by subjects on a trial-by-trial basis. This equation was evaluated for cases in which the distance metric is either Euclidean or city block and in which the exponent varies in the function relating within-trial distance to similarity. This evaluation shows that it is possible to produce a modified Gaussian function relating the expected value of 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. Shepard's (1986, 1987) conjecture concerning perceptual noise in Nosofsky's (1986) experiments is supported by the evaluation, but other similarity–distance functions may also lead to similar results. Nosofsky's findings regarding the form of the metric (Euclidean) are consistent with Shepard's (1986, 1987) theory that the appropriate metric may be city block for stimuli with separable dimensions, provided that Shepard's (1986, 1987) theory is applied at the individual trial level. When working with confusable stimuli, it will be important to develop approaches for estimating the parameters of the distributions of psychological magnitudes as well as the parameters of the similarity function. A possible approach to this problem is discussed in Ennis et al. (in press).

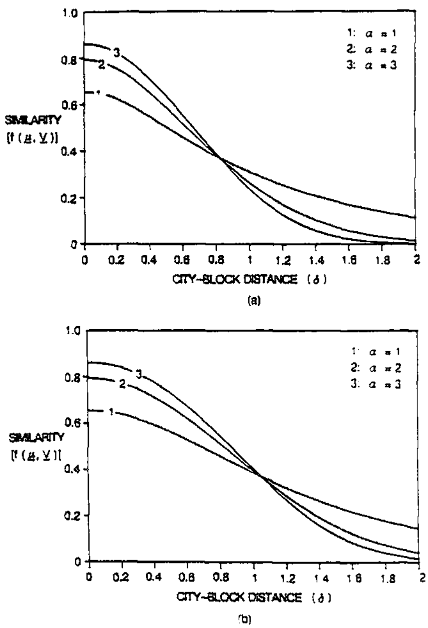


Figure 2. Expected value of similarity as a function of the city-block distance between the means of the distributions of psychological magnitudes for values of α of 1, 2, and 3 in the similarity function of $g(d) = \exp(-d^{\alpha})$. (a) Means differ on one axis only; (b) means differ equally on both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

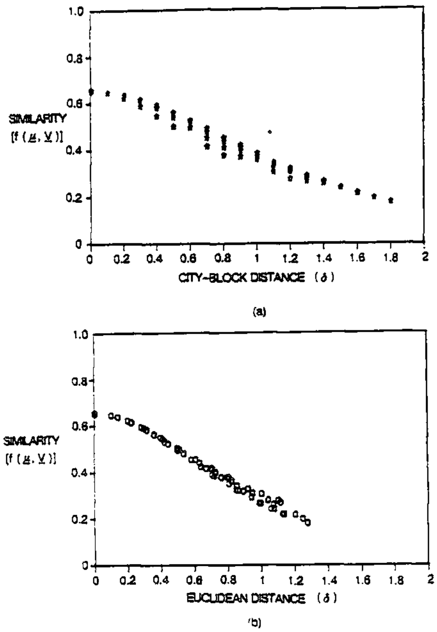


Figure 3. Expected value of similarity between pairs of 16 stimuli plotted against the city-block and Euclidean distances (δ) between the means of the distributions of psychological magnitudes. An exponential decay function has been used to describe the relationship between city-block distance and within-trial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

Because all pairs of stimuli exist on a continuum of similarity, certain deterministic models of similarity and identification performance should be viewed as special cases of corresponding stochastic models. When stimulus pairs are sufficiently discriminable and self-similarity estimates are identical, the effect of the variance–covariance matrix will diminish to a point at which similarity can be modeled as a function of μ or δ.

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., & 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., Palen, J., & Mullen, K. (in press). A multidimensional theory of similarity. *Journal of Mathematical Psychology.*

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, Ann Arbor.

Mullen, K., & Ennis, D. M. (1987). Mathematical formulation of multivariate Euclidean models for discrimination methods. *Psychometrika, 52*, 235–249.

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

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 sciences. *Science, 237*, 1317–1323.

Zinnes, J. L., & MacKay, D. B. (1983). Probabilistic multidimensional scaling: Complete and incomplete data. *Psychometrika, 48*, 27–48.

Received January 4, 1988
Revision received April 12, 1988
Accepted April 18, 1988 ■

A Comparison of Selected Probabilistic Multidimensional Models of Identification with Respect to Perceptual Dependence

DANIEL M. ENNIS

Philip Morris Research Center, Richmond, Virginia

AND

F. GREGORY ASHBY

Department of Psychology, University of California, Santa Barbara

ABSTRACT

Probabilistic models of identification are compared with regard to their sensitivity to perceptual dependence or the degree to which the underlying psychological dimensions are correlated. Three types of models are compared: Signal detection models, a probabilistic multidimensional scaling model, and probabilistic models based on the Shepard-Luce choice rule. The signal detection models were found to be most sensitive to perceptual dependence, especially when there is considerable distributional overlap. The choice rule based on the city-block metric and an exponential decay similarity function was found to be particularly insensitive to perceptual dependence. These theoretical results may play an important role in studying different decision rules employed at different stages of identification training.

Correspondence should be addressed to Daniel M. Ennis, Philip Morris Research Center, P.O. Box 26583, Richmond, VA 23261.

INTRODUCTION

Probabilistic multidimensional models have recently been used to account for a wide variety of psychophysical and perceptual phenomena (Ashby & Gott, 1988; Ashby & Perrin, 1988; Ashby & Townsend, 1986; De Soete, Carroll, & DeSarbo, 1986; Ennis, 1988; Ennis & Mullen, 1986a,b; Ennis, Palen, & Mullen, 1988; Graham, Kramer, & Yager, 1987; Hirsch, Hylton, & Graham, 1982; Mullen & Ennis, 1987; Mullen, Ennis, deDoncker, & Kapenga, 1988; Olzak & Wickens, 1983; Zinnes & MacKay, 1983, 1987). In developing these models, it is assumed that the perceptual effect of a stimulus or its memory trace can be represented as a point in a multidimensional space. Depending on the model, the location of some or all of the stimulus percepts, ideal points (in preference models) or memory traces (in identification models) are assumed to vary from trial to trial according to some multivariate probability distribution. It has been common to assume that this distribution is multivariate normal.

Despite their often identical assumptions about the perceptual representation, these models differ in how the subject is assumed to use the perceptual information to select a response. Although the models have been applied to a wide variety of experimental paradigms, three types of decision functions stand out. These can be classified as: Signal detection rules; Distance-based ordinal rules; and Shepard-Luce choice rules. Three types of parameters characterize the multivariate normal distribution: location parameters (i.e., the means), spread parameters (i.e., the variances), and association parameters (i.e., the correlations or covariances). Manipulating the values in any one of these classes will presumably have some effect on the predictions of the various models. Of the three types of parameters, however, the least is known about the effects of manipulating the perceptual correlations. In this article we compare the identification probabilities of these three classes of models as a function of the degree of perceptual correlation. For the parameters considered, all of these models make different predictions. However, of the three types, the signal detection models are more sensitive to perceptual correlation than the other two.

OVERVIEW OF THE MODELS

To compare the various classes of models, consider an identification experiment with *n* stimuli. On each trial, a single stimulus is presented and the subject is asked to identify it uniquely.

Multidimensional signal detection (MSD) models assume that the subject divides the perceptual space into regions and assigns a response to each region (Ashby, 1988; Ashby & Gott, 1988; Ashby & Perrin, 1988; Ashby & Townsend, 1986; Graham, Kramer, & Yager, 1987; Hirsch, Hylton, & Graham, 1982; Olzak & Wickens, 1983). On each trial, the subject determines in which region the stimulus representation happens to fall and then gives the associated response. Several versions of this model can be formulated depending on how the subject divides the perceptual space into response regions. In this article, we consider three possibilities.

In the *minimum distance model*, the subject is assumed to respond according to the nearest perceptual mean. In the two stimulus case, this rule defines the decision bound as that set of points that are equidistant from the two means. Note that this bound is always linear and it bisects and is orthogonal to the chord connecting the perceptual means. Minimum distance bounds depend on interstimulus mean distances and so MSD models of this type are related to the simple Euclidean multidimensional scaling model (e.g., Kruskal, 1961a,b; Shepard, 1962a,b; Torgerson, 1958). They are also related to prototype models of categorization (e.g., Posner & Keele, 1968, 1970; Reed, 1972; Rosch, 1973; Rosch, Simpson, & Miller, 1976) and to parallel distributed memory models that assume matched filtering or cross-correlation (Ashby & Gott, 1988; Hinton & Anderson, 1982). *General linear models* constrain the decision bounds to be linear, but they place no constraints on their slope or intercept. Finally, the *optimal model* places the decision bound so that overall identification accuracy is maximized. With only rare exceptions, this model predicts that the decision bounds are nonlinear.

Models involving distance-based ordinal rules, which we will call **probabilistic multidimensional scaling (PMDS) models**, assume that the subject selects a response on the basis of the distances between appropriate perceptual effects, rather than by forming decision bounds (De Soete, Carroll, & DeSarbo, 1986; Ennis & Mullen, 1986; Mullen & Ennis, 1987; Mullen, Ennis, deDoncker & Kapenga, 1988; Zinnes & MacKay, 1983, 1987). During identification, some of these perceptual effects might be samples from a memory representation and, when judging preference, some might be samples from a representation of the ideal stimulus. After stimulus presentation in an identification task, this class of models assumes that the subject recalls an instance of each stimulus alternative, determines the similarity of each of these memory traces to the stimulus, and then gives the response associated with the nearest or most similar trace. Typically, the perceptual effects of stimulus presentation and the memory of each stimulus alternative are each represented by a multivariate normal distribution. Thus, for example, the events that occur on a trial when a stimulus is presented are modeled by: 1) taking a sample from the perceptual distribution corresponding to the stimulus, 2) taking a sample from each memory distribution, 3) computing the distances between the perceptual sample and each memory sample, and finally 4) giving the response associated with the minimum of these distances. In most models, the same distribution is assumed to describe the perceptual and memory representations of each stimulus alternative.

In **stochastic choice (SC) models**, the decision function does not lead to an explicit choice among alternatives, but instead provides a probability that a particular choice will be made (Ennis, Palen, & Mullen, 1988). The models that have been investigated to date have all been based on the Shepard-Luce biased choice model (Luce, 1963; Shepard, 1957), but it should be possible to develop models based on other choice formulations. The Ennis et al. model assumes a multivariate normal distribution for the perceptual representation of each stimulus. The η_{ij} similarity terms from the biased choice model are defined as $\eta_{ij} = \exp(-d_{ij}^{\alpha})$, for some constant α and where d_{ij} is the distance between a random sample from the S_i perceptual distribution and one from the S_j. The identification probabilities equal the expected value of the choice rule ratio. Two common ways in which the models within this class differ are in the metric which defines interstimulus distances and in the way in which similarity is defined (i.e., in the value of α; see, e.g., Ennis, 1988; Nosofsky, 1985, 1986, 1988; Shepard, 1987, 1988).

Note that SC models differ fundamentally from MSD and PMDS models in the sense that SC models postulate a probabilistic decision process whereas MSD and PMDS models postulate deterministic responding. In MSD models, a given perceptual effect always leads to the same response, at least in the absence of criterial noise. In PMDS models, a given set consisting of a perceptual effect and *n* memory effects always leads to the same response. In SC models, however, no deterministic response rule is specified. All we can determine is the probability that a given response will be made.

Multidimensional signal detection (MSD) models assume that subjects set up response regions with decision bounds which they use to make subsequent decisions about presented stimuli. Presumably, the location of these decision bounds evolve gradually from trial-by-trial feedback. This implies that subjects have been trained with examples of the stimuli. In PMDS models, however, the assumptions underlying the decision processes reflect only the behavior of subjects in isolated trials. No training to form decision bounds is assumed. If a subject is asked to identify a stimulus, these models assume that a comparison made between a presented stimulus and a memory representation corresponds to a comparison only between a single sample drawn from the various distributions in question. The subject is not assumed to have any knowledge of or control over the process that evokes the particular values from these distributions. In contrast, the MSD models, by assuming a knowledge of response regions, make stronger assumptions about what the subject has learned about the stimuli during training.

A FORMAL DESCRIPTION OF THE MODELS

Assumptions Common to all Models

We will focus on a two stimulus identification task with stimulus objects, S_A and S_B and responses, R_A and R_B. All models we consider assume that the perceptual effects of each stimulus can be represented as multivariate normal deviates with means $\underset{\sim}{\mu}_A$ and $\underset{\sim}{\mu}_B$ and variance-covariance matrices $\mathbf{V}_A$ and $\mathbf{V}_B$. Random values from these distributions vary from trial to trial in a manner determined by the parameters of the multivariate normal density functions, f_A and f_B. For example, when the perceptual representations are composed of *n*-tuples (or the vectors have *n* elements each),

$$f_i(\mathbf{x}) = \frac{\exp\{-0.5(\mathbf{x}-\underset{\sim}{\mu}_i)^t \mathbf{V}_i^{-1}(\mathbf{x}-\underset{\sim}{\mu}_i)\}}{(2\pi)^{n/2}|\mathbf{V}_i|^{1/2}}, \; i = A, B.$$

In all numerical examples, we will assume that $n = 2$. In this case, the multivariate normal becomes a bivariate normal. In scalar form

$$f_i(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}} \times$$
$$\exp\left\{-\frac{1}{2(1-\rho^2)}\left(\left[\frac{x_1-\mu_1}{\sigma_1}\right]^2 - 2\rho\left[\frac{x_1-\mu_1}{\sigma_1}\right]\left[\frac{x_2-\mu_2}{\sigma_2}\right] + \left[\frac{x_2-\mu_2}{\sigma_2}\right]^2\right)\right\},$$

where μ_1 and σ_1^2 (μ_2 and σ_2^2) are the mean and variance of the first (second) element of the random vector, **x**, respectively; ρ is the correlation coefficient between the first and second elements; and $\mathbf{x}^t = (x_1, x_2)$. Following Ashby and Townsend (1986), we will use ρ as a measure of perceptual dependence.

A stimulus object, S_i, is presented to a subject and this object evokes a perceptual effect which is also distributed as a multivariate normal deviate with mean $\underset{\sim}{\mu}_i$ and variance-covariance matrix $\mathbf{V}_i$. This probability density function is f_i. In an identification task the presented stimulus, S_i, is either S_A or S_B. In other tasks, however, S_i may be different from S_A or S_B.

Tasks which involve comparisons among psychological magnitudes from three distributions occur frequently in experimental paradigms in psychology and marketing. Examples include Torgerson's method of triads (Torgerson, 1958), the ABX design, the duo-trio tristimulus discrimination paradigm, and preferential choice experiments. In Torgerson's method of triads, subjects are instructed to select from two alternative stimuli, the one which is most similar to a third, designated stimulus. In the ABX design, subjects are given two designated stimuli, S_A and S_B, and are instructed to identify which of S_A or S_B is most similar to a third stimulus, S_X. In the duo-trio method, subjects are instructed to select, from two alternatives, the stimulus that is most similar to a designated stimulus. In this method, the designated stimulus is identical to one of the alternatives. The duo-trio method is a special case of Torgerson's method of triads. If one of the alternatives is not a physical stimulus, but an ideal point (actually, ideal distribution), then the task of choosing a preferred stimulus from two alternatives on the basis of similarity to the ideal point, is equivalent to Torgerson's method of triads.

The Multidimensional Signal Detection Model

Let **x** be a random value from f_i and denote the region in the perceptual space assigned to response R_j by D_j. Then in all MSD models, the probability of responding R_j on trials when stimulus S_i is presented is equal to the proportion of the perceptual distribution associated with stimulus S_i that lies in the D_j response region. Formally,

$$P(R_j|S_i) = \int\int\cdots\int_{D_j} \frac{\exp\{-0.5(\mathbf{x}-\boldsymbol{\mu}_i)^t \mathbf{V}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\}}{(2\pi)^{n/2}|\mathbf{V}_i|^{1/2}} d\mathbf{x} \qquad (1),$$

Row vectors will be represented with a t superscript [for example, $(\mathbf{x}-\boldsymbol{\mu}_i)^t$ above], column vectors will not have a superscript. Different versions of the MSD model assume different shapes for the region D_j. In this paper we consider three cases: 1) the **ideal observer**, 2) the **general linear classifier**, and 3) the **minimum distance classifier.**

The **ideal observer** is assumed to maximize overall identification accuracy. Assume that each stimulus is presented with equal probability on each trial. The ideal observer responds R_A whenever $f_A(\mathbf{x}) > f_B(\mathbf{x})$ and R_B whenever $f_B(\mathbf{x}) > f_A(\mathbf{x})$. Therefore, D_A is the region of the perceptual space for which $f_A(\mathbf{x}) > f_B(\mathbf{x})$. If there are more than two stimuli, the ideal observer gives the response associated with the greatest likelihood. Note that in the two-stimulus case, another interpretation is that response R_A is given whenever $h(\mathbf{x}) = -\ln[f_A(\mathbf{x})/f_B(\mathbf{x})] < 0$ and response R_B is given when $h(\mathbf{x}) > 0$. Integration over the region for which $h(\mathbf{x}) < 0$ may be accomplished by defining a function which takes on the value 0 when $h(\mathbf{x}) > 0$ and 1 otherwise. (The probability that $h(\mathbf{x}) = 0$ is zero.) Such a function is $0.5\{1 - \text{sgn}[h(\mathbf{x})]\}$, where sgn($a$) (the signum function evaluated at a) = -1 for $a \le 0$ and +1 for $a > 0$. Equation 1 can now be written as an integral over R^n as

$$P(R_j | S_i) = \int_{R^n} \frac{\exp\{-0.5(\mathbf{x}-\boldsymbol{\mu}_i)^t \mathbf{V}_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i)\}}{(2\pi)^{n/2}|\mathbf{V}_i|^{1/2}} 0.5\{1 - \text{sgn}(h(\mathbf{x}))\}\, d\mathbf{x}. \quad (2).$$

The contour for which $h(\mathbf{x}) = 0$ is called the decision bound because it separates the perceptual space into two regions. On one side of this bound, $h(\mathbf{x}) < 0$ and response R_A is given and on the other side $h(\mathbf{x}) > 0$ and response R_B is given.

For the ideal observer, the decision bound is quadratic in $\mathbf{x}$, but for the **general linear classifier** the decision bound is constrained to be linear. Thus, there exists a vector of constants $\mathbf{a}$ and a scalar b such that

$$h(\mathbf{x}) = \mathbf{a}^t\mathbf{x} + b.$$

Because $\mathbf{x}$ has a multivariate normal distribution, $h(\mathbf{x})$ has a univariate normal distribution. On trials when stimulus S_i is presented, $h(\mathbf{x})$ has mean $\mathbf{a}^t\boldsymbol{\mu}_i + b$ and variance $\mathbf{a}^t\mathbf{V}_i\mathbf{a}$. Therefore, when each stimulus is presented with equal probability, the overall probability of an error is given by

$$\varepsilon = P[h(\mathbf{x}) > 0|S_A]/2 + P[h(\mathbf{x}) < 0|S_B]/2$$

$$= 1/2\,\Phi\left[\frac{\mathbf{a}^t\boldsymbol{\mu}_A + b}{(\mathbf{a}^t\mathbf{V}_A\mathbf{a})^{0.5}}\right] + 1/2\,\Phi\left[\frac{-\mathbf{a}^t\boldsymbol{\mu}_B - b}{(\mathbf{a}^t\mathbf{V}_B\mathbf{a})^{0.5}}\right] \qquad (3)$$

where $\Phi(z)$ is the standard normal (i.e., Z) cumulative distribution function [i.e., $\Phi(z) = P(Z < z)$].

In this paper we consider the most accurate general linear classifier, that is, the model with the most accurate of all possible linear bounds (see, e.g., Ashby & Gott, 1988). In other words, we wish to identify that model with vector $\mathbf{a}$ and scaler b which minimizes ε. Unfortunately, no analytic solution is known. However, it can be shown that for the most accurate linear classifier, there exists a constant κ in the interval $0 \le \kappa \le 1$ such that the following constraint is satisfied (Anderson, 1962; Fukunaga, 1972; Peterson & Mattson, 1966)

$$\mathbf{a} = [\kappa\mathbf{V}_A + (1-\kappa)\mathbf{V}_B]^{-1}(\boldsymbol{\mu}_B - \boldsymbol{\mu}_A).$$

Once $\mathbf{a}$ is known, b can be calculated as

$$b = \frac{-[\kappa(\mathbf{a}^t \mathbf{V}_A \mathbf{a})\, \mathbf{a}^t \boldsymbol{\mu}_B + (1-\kappa)(\mathbf{a}^t \mathbf{V}_B \mathbf{a})\mathbf{a}^t \boldsymbol{\mu}_A}{\kappa(\mathbf{a}^t \mathbf{V}_A \mathbf{a}) + (1-\kappa)(\mathbf{a}^t \mathbf{V}_B \mathbf{a})}.$$

Therefore, for a given value of κ, both **a** and b can be determined and then Equation 3 can be used to find the probability of an error, ε. The most accurate linear bound can be identified by using a numerical minimization routine to find that value of κ which minimizes ε.

For the **minimum distance classifier** the decision bound is the set of all points equidistant from the two distribution means. This bound is always linear and bisects and is perpendicular to the line connecting the two means. The minimum distance classifier is a special case of the general linear classifier in which $\mathbf{a} = \boldsymbol{\mu}_B - \boldsymbol{\mu}_A$ and $b = 1/2(\boldsymbol{\mu}_A^t \boldsymbol{\mu}_A - \boldsymbol{\mu}_B^t \boldsymbol{\mu}_B)$. For both models, the probability of responding R_j on S_i trials can be computed from Equation 2 once the appropriate definition of $h(\mathbf{x})$ is given.

Probabilistic Multidimensional Scaling Models

In addition to the perceptual distribution of a presented stimulus, these models assume that on every trial the subject recalls a memory trace of each stimulus alternative. In this article however, we consider only models that assume that the distribution of perceptual effects equals the distribution of memory traces for each stimulus. Thus in the two stimulus identification task, three random samples are generated on each trial: one from the distribution associated with the presented stimulus, $\mathbf{x}$, and one from each of the distributions corresponding to the memory traces, $\mathbf{x}_i$ and $\mathbf{x}_j$.

Presentation of stimulus S_i will lead to response R_i if

$$|\mathbf{x} - \mathbf{x}_i| < |\mathbf{x} - \mathbf{x}_j|,$$

and to response, R_j if

$$|\mathbf{x} - \mathbf{x}_j| < |\mathbf{x} - \mathbf{x}_i|.$$

Let $(\mathbf{x} - \mathbf{x}_i) = \mathbf{u}$ and $(\mathbf{x} - \mathbf{x}_j) = \mathbf{v}$. The identification probability is identical to the probability of a correct response in the duo-trio method (Ennis & Mullen, 1986; Mullen & Ennis, 1987),

$$P(R_i \mid S_i) = \int_C f(\mathbf{u},\mathbf{v})\, d(\mathbf{u})d(\mathbf{v}) \qquad (4)$$

where C is the region for which $|\mathbf{u}| < |\mathbf{v}|$

and

$$f(\mathbf{u},\mathbf{v}) = \frac{\exp\{-0.5(\mathbf{z}-\underset{\sim}{\mu})^t\ \mathbf{V}^{-1}(\mathbf{z}-\underset{\sim}{\mu})\}}{(2\pi)^n\ |\mathbf{V}|^{1/2}},$$

$$\mathbf{z} = (\mathbf{u},\mathbf{v})$$

$$\underset{\sim}{\mu} = [(\underset{\sim}{\mu}_i - \underset{\sim}{\mu}_i), (\underset{\sim}{\mu}_i - \underset{\sim}{\mu}_j)],$$

$$\mathbf{V} = \begin{bmatrix} 2\mathbf{V}_i & \mathbf{V}_i \\ \mathbf{V}_i & \mathbf{V}_i + \mathbf{V}_j \end{bmatrix}.$$

Note that if n is 2, there are 4 elements in $\mathbf{z}$.

The decision rule in the PMDS model was stated earlier in terms of distance comparisons between a probe and the two memory distributions. This decision rule could also have been stated in terms of a comparison between the similarity of the probe and the two memory random values. Provided that similarity is defined as a monotonic function (or a one-to-one function) of distance, then identical similarities implies that the distances are also identical. If the similarity function is monotonically decreasing, then an inequality relating two similarities always implies the reverse inequality relating the corresponding distances. Hence, a probabilistic identification model based on similarity in which the probe is identified with the most similar memory value is identical to the PMDS model just described.

In addition to identification, the PMDS model can be applied to many paradigms in psychology and marketing. For instance, if the means and variance-covariance matrices for three stimulus distributions may all be different, this model is the multidimensional

Thurstonian variant of Torgerson's method of triads (Torgerson, 1958). If one of the distributions is interpreted as an ideal point distribution, the model is also a general preference model. If two of the distributions are considered to be memory representations and the third a stimulus distribution, the model is an identification model. Special cases arise when two of the distributions are identical as occurs in the model for the ABX and duo-trio tristimulus discrimination paradigms.

The $2n$-fold integral (Equation 4) has been converted to fixed limits of integration in spherical coordinates (Mullen, Ennis, deDoncker, & Kapenga, 1988). Equation 4 has also been simplified to a single integral irrespective of the dimensionality of the vector space of psychological magnitudes (Mullen & Ennis, 1990). In this form, the probability of responding R_i is

$$P(R_i \mid S_i) = 0.5 - \frac{1}{\pi}\int_0^{\infty} \frac{\sin \theta(t)}{t \ \rho(t)}\, dt,$$

where,

$$\theta(t) = 0.5 \sum_{k=1}^{r} [m_k \tan^{-1}(\delta_k t) + \omega_k^2 \delta_k t(1 + \delta_k^2 t^2)^{-1}],$$

$$\rho(t) = \left[\prod_{k=1}^{r} (1 + \delta_k^2 t^2)^{m_k/4}\right] exp\{0.5 \sum_{l=1}^{r} (\omega_l \delta_l t)^2 / (1 + \delta_l^2 t^2)\},$$

δ_k are the eigenvalues of $\mathbf{V} \begin{bmatrix} \mathbf{I} & 0 \\ 0 & -\mathbf{I} \end{bmatrix}$,

m_k are the degrees of freedom of the kth non-central chi-square,

ω_k is the non-centrality parameter of the kth chi-square and

r is the number of distinct eigenvalues.

The steps needed to obtain the non-centrality parameters, ω_k, are given in Mullen and Ennis (1990).

The single integral form is generally about 3,000 times faster to compute to the same accuracy as adaptive numerical integration of Equation 4 when n equals 2. These developments have overcome earlier difficulties in computing Equation 4 accurately in reasonable computer time.

Stochastic Choice Models

In an earlier paper on a stochastic theory of similarity (Ennis, Palen, & Mullen, 1988) it was pointed out that the Shepard-Luce choice rule could be formulated in stochastic terms by computing the expected value of the choice rule ratio. Specifically, it was proposed that

$$P(R_j \mid S_i) = E\left[\frac{\beta_j g(d_{ij})}{\sum_{k=1}^{m} \beta_k g(d_{ik})}\right] \qquad (5),$$

where

β_j is the j^{th} response bias parameter,

m is the number of memory representation alternatives,

$$d_{ij} = [\sum_{r=1}^{n} |x_{ir} - x_{jr}|^{\gamma}]^{1/\gamma} \qquad \gamma \geq 1,$$

$$= (\mathbf{u}^{*} \bullet \mathbf{1})^{1/\gamma},$$

where

$\mathbf{u}^{*}$ is a row vector with typical element $|u_i|^{\gamma}$, $\mathbf{1}$ is a column vector of 1's, and

$$g(d_{ij}) = \exp(-d_{ij}^{\alpha}), \ \alpha \geq 0.$$

Note that d_{ii} is the distance between two random values from the same distribution and, therefore, may be nonzero.

An integral expression for Equation 5 was not given in Ennis, Palen, & Mullen (1988). This equation will now be given. Consider the case when $m = 2$, i.e., when there are only two alternatives to choose from. In making an identification decision within a trial using the Shepard-Luce choice rule, it is assumed that the subject's probability of choosing R_j when the stimulus is S_i is a function of the two distances d_{ij} and d_{ii}. The same notation from the PMDS models can be used, where $\mathbf{u} = \mathbf{x} - \mathbf{x}_A$, $\mathbf{v} = \mathbf{x} - \mathbf{x}_B$, $\mathbf{z} = (\mathbf{u},\mathbf{v})$, $\mathbf{V}$ is the variance-covariance matrix of the joint distribution of $\mathbf{u}$ and $\mathbf{v}$, defined earlier, and $\boldsymbol{\mu}$ is the vector of mean differences previously defined also. To obtain the expected value of the choice rule, it is necessary only to integrate over the vector space composed of all $2n$-tuples represented by $\mathbf{z}$, weighting each element of the space by its probability of occurrence (given by the multivariate normal density function), or

$$P(R_A | S_i) = \int_{R^n}\int_{R^n} \frac{\exp\{-0.5(\mathbf{z}-\boldsymbol{\mu})^t \ \mathbf{V}^{-1}(\mathbf{z}-\boldsymbol{\mu})\}}{(2\pi)^n \ |\mathbf{V}|^{1/2}} \ \frac{\beta_A g(d_{iA})}{\beta_A g(d_{iA}) + \beta_B g(d_{iB})} \, d\mathbf{z}. \quad (6)$$

If $\alpha = 1$ and $\gamma = 1$, g is an exponential decay function and the metric of d is city-block. If $\alpha = 2$ and $\gamma = 2$, g is a Gaussian function and the metric of d is Euclidean. Both of these cases will be considered and will be called the city-block/exponential decay and Euclidean/Gaussian cases.

THE MODELS COMPARED WITH RESPECT TO PERCEPTUAL DEPENDENCE

As stated earlier, we are mainly interested in assessing the degree to which the various models are sensitive to variations in perceptual dependence. Components of a stimulus are perceived independently if their perceptual effects are statistically independent (Ashby & Townsend, 1986). With bivariate normal perceptual distributions, perceptual independence occurs if and only if the correlation coefficient equals zero. Therefore in this section we examine the predicted identification accuracy of the various models as a function of ρ, the correlation coefficient.

There are a number of methods for testing whether a pair of stimulus components are perceived independently. Ashby and Townsend (1986) developed several tests that can be applied to data from certain identification tasks. Perhaps the simplest and most powerful test, however, involves an experimental paradigm called the concurrent rating task (Ashby, 1988; Hirsch, Hylton, & Graham, 1982; Olzak, 1986).

Consider a stimulus ensemble constructed by factorially combining two levels of two components, X and Y. Then the four stimuli are X_1Y_1, X_1Y_2, X_2Y_1 and X_2Y_2. In the previous applications of the concurrent rating task, the two components have been sine-wave gratings of different frequency and the various levels have been different contrasts. A stimulus consists of two superimposed gratings of a certain frequency (the same frequencies are used in all trials) but variable contrast from trial to trial. On each trial, the subject is shown one of the four stimuli. The subject's task is to give two possible rating responses, one for each component. The rating corresponding to the *i*th component within a trial (in this example, $i = 1$ or 2) is a number between 1 and k and reflects the subject's confidence that the *i*th component was present at the highest possible level (i.e., at the highest possible contrast). The data in such an experiment is conveniently catalogued in a $k \times k$ matrix for each stimulus. The entry in row i and column j is the frequency with which the subject responded i on component X and j on component Y.

A natural way to model the subject's performance in this task is to assume that a perceptual dimension exists for each stimulus component and that the subject constructs k-1 criteria on each dimension. Response R_i is given on component X if the percept has a value on the dimension associated with component X that falls between the *i*th-1 and the *i*th criterion. Under these assumptions, Ashby (1988) showed that the perceptual independence of components X and Y guarantees that the ratings on the two components are uncorrelated. Further, if uncorrelated ratings are found for all possible criterion placements, then perceptual independence is implied. In the case of a perceptual dependence, Ashby (1988) suggested a simple generalization of the tetrachoric correlation coefficient that provides accurate estimates of the perceptual dependence, as measured by the correlation coefficient, ρ_{XY}.

This test of perceptual independence has not yet been widely used outside the spatial vision area. It should be possible, however, to use the method to assess the degree of perceptual dependence between *any* pair of stimulus components. This in turn would allow one to conduct a series of experiments which vary the average degree of perceptual dependence.

We derived the predicted accuracy on S_A trials, $P(R_A|S_B)$, for each of the six models described above in a total of 36 different conditions. In each of these, $\underset{\sim}{\mu}_A^t = (0,0)$ and $V_A = \begin{bmatrix} 4 & 4\rho_A \\ 4\rho_A & 4 \end{bmatrix}$ or $\begin{bmatrix} 1 & \rho_A \\ \rho_A & 1 \end{bmatrix}$. The correlation coefficient, ρ_A, was assigned the values 0, 0.4, -0.4, 0.8 or -0.8. The S_B mean was set to $\mu_B^t = (0.707, 0.707)$ or to (1.414, 1.414) and V_j was either $\begin{bmatrix} 4 & 4\rho_B \\ 4\rho_B & 4 \end{bmatrix}$ or $\begin{bmatrix} 1 & \rho_B \\ \rho_B & 1 \end{bmatrix}$. Smaller means and larger variances occurred in the same combination, leading to two types of conditions, one of high overlap and one of moderate overlap. The S_B correlation, ρ_B, was either 0, 0.4, -0.4, 0.8 or -0.8 and was allowed to differ from ρ_A but ρ_A and ρ_B were constrained to be either both positive or both negative. This constraint reduces the total number of conditions from 72 ($6 \times 6 \times 2$), if all possible combinations were considered, to 36. Nine different cases are implied by the values of ρ_A and ρ_B, i.e. 0,0; 0,0.4; 0,0.8; 0.4,0; 0.4,0.4; 0.4,0.8; 0.8,0; 0.8,0.4; 0.8,0.8. Another nine cases are implied by $-\rho_A$, $-\rho_B$.

MSD and PMDS models yield identification probabilities that are invariant to numerical multiples of the vectors representing the psychological magnitudes. For example, when the covariance matrices associated with stimuli S_A and S_B both equal the identity matrix multiplied by the same scalar, then models in both classes predict that performance depends only on the standardized distance between means. Specific values of the means and variances are not important. However, the SC models give identification probabilities that depend on the absolute values of the elements of these vectors. This is because the judgment function used in the SC models ($e^{-d^{\alpha}}$) depends on the actual magnitudes of $\mathbf{x}$, $\mathbf{x}_A$ and $\mathbf{x}_B$ rather than their relative values. In the MSD and PMDS models, subjects make decisions based on the relative sizes of either likelihoods or distances which will retain the same rank order when all the vectors from the three distributions are multiplied by the same constant (or, equivalently, when the means and variances of these distibutions are multiplied by corresponding constants). Because of this difference, comparisons among the models were made on the basis of the relative differences implied by the variance-covariance structures investigated within a particular model using $\rho_A = \rho_B = 0$ as the reference. A comparison of the absolute identification probabilities will be given also.

Tables 1 and 2 show how the six models in three types (MSD, PMDS, and SC) compare with regard to sensitivity to perceptual dependence. Note that when there is considerable distributional overlap, (see Table 1), the MSD models are the only ones to show sensitivity. Almost all of the identification probabilities were in the 0.5 - 0.7 range, so there was no obvious constraint (e.g., ceiling effects) preventing sensitivity to correlation structure (as would occur if these values were close to 1). When the distributional overlap is moderate, (see Table 2), the differences among the models diminishes somewhat as the PMDS and SC(Euclidean/Gaussian) models begin to respond to perceptual dependence. The SC (City-block/

exponential decay) model showed very little sensitivity to perceptual dependence for either of the parameter sets studied. The decrease in sensitivity displayed by the optimal model in the case of moderate distributional overlap is due to a ceiling effect. In fact, when the S_A correlation was -.8 and the S_B correlation was -.8, the optimal model predicted an accuracy of .987, suggesting that even at moderate levels of confusability it is extremely sensitive to perceptual dependence. The lower sensitivity displayed by the PMDS and the SC models can not be attributed to a ceiling effect since these models predict lower overall accuracy than the MSD models.

When the distributional overlap was reduced even further, the PMDS and SC models did not show an increased sensitivity to perceptual dependence. In addition, the identification probabilities predicted by the optimal MSD model were almost all greater than .9. Thus a comparison of these models at low levels of confusability is not very meaningful.

With the exception of the best linear classifier, note that negative correlations gave higher identification probabilities than positive correlations. When the f_A mean is in the first quadrant, negative correlations generally lead to less overlap of the distributions than when the correlations are positive. An exception occurs for the ideal observer in the case of high distributional overlap when $\rho_A = 0$ and $\rho_B = 0.4$. When $\rho_B = -.4$ the optimal bound is concave down and when $\rho_B = +.4$ the optimal bound is concave up. Note that the line $y = x$ connects the two means. Because the distributional overlap is so large, the optimal bound crosses the $y = x$ line at about the same place when the S_j correlation is both positive and negative. Therefore the overlap of f_A into the R_B response region is less when the S_B

correlation is positive (because the bound is concave up in this case) and so a higher accuracy is predicted on S_A trials when the S_B correlation is positive. On the other hand, in these same conditions, accuracy is higher on S_B trials when the S_B correlation is negative. In fact, it is large enough so that overall accuracy (the average of S_A and S_B trials) is greater when the S_B correlation is negative.

Similar effects occur with the best linear classifier, only in this case the effects are more extreme. It can be seen from Table 1 that in several cases accuracy is higher on S_A trials when both correlations are positive. In each of these cases however, overall accuracy is higher when both correlations are negative.

Note that the predictions of the minimum distance model depend only on the S_A correlation. In this model the decision bound depends only on the perceptual means. An S_A perceptual dependence does affect predicted accuracy on S_A trials however, because it affects the proportion of samples from f_A that will fall on the side of the bound associated with response R_A.

From Tables 1 and 2, it can be seen that there is an extremely high degree of similarity between the sensitivity of the PMDS model and the Gaussian-Euclidean SC model to perceptual dependence. In fact, the predictions of these two models are more similar than the predictions of the two SC models. These tables show the general trend that when both correlations are negative, identification accuracy is higher then when both are positive. It is also clear from these tables that the relative sensitivity of the SC (city-block/exponential decay) model to perceptual dependence is very weak compared to MSD models and that PMDS and SC (Euclidean/Gaussian) models are intermediate in sensitivity.

In Figures 1 and 2, the MSD (ideal observer) model has been compared to the other five models by plotting the predictions for the cases given in Table 1 and 2. These plots clearly show the greater sensitivity of the MSD models to perceptual dependence and the relative insensitivity of the other models, especially the city-block/exponential decay SC model. The degree of overlap of the distributions also affects responsiveness to perceptual dependence, as these figures show.

PERCEPTUAL DEPENDENCE AND THE SHAPE OF THE SIMILARITY FUNCTION

The identification models discussed differ in several important ways that might give rise to their differences in sensitivity to perceptual dependence. In the MSD models, the subject is assumed to be trained sufficiently to establish decision boundaries. In the case of the ideal observer, the positioning of these boundaries would be determined very precisely by the variance-covariance matrix, and thus would be likely to depend strongly on the perceptual dependencies (correlation coefficients) that influence the shape of the perceptual distributions. In the PMDS and SC models, subjects make decisions based only on random values from the distributions of interest and are not assumed to have any knowledge concerning the location of possible decision boundaries. In the PMDS models, the subject chooses the alternative that, on a particular trial, yields the smallest distance between stimulus and memory momentary values. The absolute size of this distance is immaterial to the decision. In the SC models, however, the degree of similarity of the momentary values is important. In these models, the probability of choosing an alternative on a given trial depends in a monotonically decreasing manner on the distance between the momentary stimulus and memory values. In principle, this function could range from linearity to a step function. The particular form of the similarity function chosen in the SC models is $g(d) = \exp(-d^{\alpha})$ that, when α is very large, can approach a step function. If the similarity function were linear, then its expected value would not depend on the variance-covariance matrix, but would be a linear function of the mean. However, as α increases in the above expression for $g(d)$, sensitivity to perceptual dependence should increase. Taking the basic form, $g(d) = \exp(-d^{\alpha})$, α was varied from 1 to 25 in unit increments. These functions include exponential decay, Gaussian, and a series that, at $\alpha = 25$, approaches a step function. Figure 3a shows the form of g when α is 1, 2, 6 and 25. Figure 3b gives the difference between the values of $E[g(d)]$ in two perceptual dependence cases for the two levels of distributional overlap. When α is 1, the exponential decay case, the effect of perceptual dependence is very small. As α increases, especially if the level of distributional overlap is not great, the difference between the perceptual dependence cases increases and ultimately saturates. The maximum difference is seen when a step judgment function is operative.

The MSD and PMDS models involve step function decision rules of either likelihoods or distances. Unlike the SC models, they provide a definitive account of a subject's decision given certain information. Although the results in Figures 3a and 3b might provide a basis for comparing identification models of the SC type with regard to perceptual dependence, we realize that they do not provide a satisfactory quantitative explanation for the much higher sensitivity of the MSD models to perceptual dependence.

DISCUSSION

The probabilistic multidimensional models considered in this paper differ with regard to assumptions about how momentary psychological magnitudes are used by subjects within a trial. In the case of the signal detection models, it was assumed that subjects formed response regions following extensive training with the stimuli. These response regions are assumed to be used by subjects within a trial to make identification decisions concerning a presented stimulus by giving the response that corresponds to the greatest likelihood as determined by the multivariate normal probability density function. If it is assumed that subjects do not form response regions, but respond to momentary memory and stimulus values within a trial, several other models can be proposed. In the probabilistic multidimensional scaling models, it is assumed that subjects choose an identification response within a trial that is consistent with the minimum Euclidean distance between the psychological magnitude for the stimulus and the memory representations. The stochastic choice rule models assume that the subject uses the Shepard-Luce choice rule within a trial and these models can be formulated with different metrics and similarity functions (only the Euclidean/Gaussian and city-block/exponential decay were considered in this paper).

In general, but particularly when memory and stimulus probability density functions are very similar, the multidimensional signal detection models appear to be more sensitive to perceptual dependence than the other models considered. Some responsiveness to perceptual dependence was predicted as the means of the probability density functions became more different in the case of the PMDS model and the Euclidean/Gaussian case of the SC models. However, the city-block/exponential decay SC model showed virtually no sensitivity to perceptual dependence for any of the parameter values studied.

An ideal application of these results would be in an experiment that manipulated the degree of perceptual dependence as an independent variable. If this were possible, a number of conditions could be evaluated in which the degree of perceptual dependence varied over a wide range. One could then examine identification accuracy in much the same way as in Tables 1 and 2. Sensitivity to perceptual dependence would be strong evidence in favor of the MSD hypothesis that subjects construct decision bounds. On the other hand, a finding that identification accuracy is only slightly affected by the degree of perceptual dependence would support the PMDS models and the Euclidean-Gaussian SC model. Finally, complete insensitivity to perceptual dependence would provide strong support for the city block-exponential SC model. In addition, these experiments might help to formally specify the effect of training on identification confusions and to provide a basis for determining the most likely metric and decision rule adopted by subjects at a particular stage in their training.

ACKNOWLEDGMENTS

We thank Robert Melara, Jerome Busemeyer, and an anonymous reviewer for comments on an earlier draft of this paper. We also thank Ed Gee, Jim Myracle, and Cliff Lilly for support through Philip Morris programs that promote research collaborations.

REFERENCES

Anderson, T.W. (1962). *An introduction to multivariate statistical analysis.* New York: Wiley.

Ashby, F.G. (1988). Estimating the parameters of multivariate signal detection theory from simultaneous ratings on separate stimulus components. *Perception and Psychophysics,* **44,** 195-204.

Ashby, F.G., & Gott, R.E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **14,** 33-53.

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

Ashby, F.G., & Townsend, J.T. (1986). Varieties of perceptual independence. *Psychological Review,* **93,** 154-179.

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. (1988). Confusable and discriminable stimuli: comments on Nosofsky (1986) and Shepard (1986). *Journal of Experimental Psychology, General*, **117**, 408-411.

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

Fukunaga, K. (1972). *Introduction to statistical pattern recognition.* New York: Academic Press.

Graham, N., Kramer, P., & Yager, D. (1987). Signal detection models for multidimensional stimuli: Probability distributions and combination rules. *Journal of Mathematical Psychology*, **31**, 366-409.

Hinton, G.E., & Anderson, J.A. (1982). *Parallel models of associative memory.* Hillsdale, N.J.: Erlbaum.

Hirsch, J., Hylton, R., & Graham, N. (1982). Simultaneous recognition of two spatial-frequency components. *Vision Research*, **22**, 365-375.

Kruskal, J.B. (1961a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, **29**, 1-27.

Kruskal, J.B. (1961a). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, **29**, 115-129.

Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207-238.

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*, **44**, 1169-1175.M

Mullen, K. & Ennis, D.M. (1990). A simple multivariate probabilistic model for preferential and triadic choices. *Psychometrika, in press.*

Nosofsky, R.M. (1985). Overall similarity and the identification of separable-dimension stimuli: A choice model analysis. *Perception and Psychophysics*, **38**, 415-432.

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

Nosofsky, R.M. (1988). On exemplar-based exemplar representations: Comment on Ennis (1988). *Journal of Experimental Psychology, General*, **117**,

Nosofsky, R.M. (1990). Relations between exemplar-similarity and likelihood models of classification. *Journal of Mathematical Psychology, in press.*

Rolzak, L.A. (1986). Widely separated spatial frequencies: Mechanism interactions. *Vision Research*, **26**, 1143-1153.

Olzak, L.A., & Wickens, T.D. (1983). The interpretation of detection data through direct multivariate frequency analysis. *Psychological Bulletin*, **93**, 574-585.

Peterson, D.W. & Mattson, R.L. (1966). A method of finding linear discriminant functions for a class of performance criteria. *IEEE Transactions on Information Theory*, **IT-12**, 380-387.

Posner, M.I., & Keele, S.W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, **77**, 353-363.

Posner, M.I., & Keele, S.W. (1970). Retention of abstract ideas. *Journal of Experimental Psychology*, **83**, 304-308.

Reed, S.K. (1972). Pattern recognition and categorization. *Cognitive Psychology*, **3**, 382-407.

Rosch, E. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*. New york: Academic Press.

Rosch, E., Simpson, C., & Miller, R.S. (1976). Structural bases of typicality effects. *Journal of Experimental Psychology: Human Perception and Performance*, **2**, 491-502.

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

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

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

Shepard, R.N. (1988). Time and distance in generalization and discrimination: Comment on Ennis (1988). *Journal of Experimental Psychology, General*, **117**,

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

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.

Table 1: A comparison of identification probabilities for cases involving two highly overlapping distributions with particular correlation coefficients. Six models are compared: MSD (I.O.), ideal observer; MSD (B.L.), best linear classifier; MSD (M.D.), minimum distance; PMDS; SC (E/G), Euclidean/Gaussian; SC (C/E), city-block/exponential decay. The first row of numbers for each model are identification probabilities ($x10^3$); the second row are values relative to the (0,0) correlation case.

CASES (High Overlap)
Correlation Coefficients

Model	*0,0*	*0,.4*	*0,-.4*	*0,.8*	*0,-.8*	*.4,0*	*-.4,0*	*.4,.4*	*-.4,-.4*	*.4,.8*	*-.4,-.8*	*.8,0*	*-.8,0*	*.8,.4*	*-.8,-.4*	*.8,.8*	*-.8,-.8*
MSD	599	684	504	583	587	536	749	584	627	536	550	818	856	754	835	574	712
(I.O.)	100	114	84	97	98	89	125	97	105	89	92	137	143	126	140	96	119
MSD	603	602	559	668	536	587	687	583	627	611	550	526	829	555	822	574	712
(B.L.)	100	100	93	111	89	97	114	97	104	101	91	87	137	92	136	95	118
MSD	599	599	599	599	599	583	628	583	628	583	628	574	712	574	712	574	712
(M.D.)	100	100	100	100	100	97	105	97	105	97	105	96	119	96	119	96	119
PMDS	518	512	517	498	509	524	527	515	522	496	510	551	556	537	547	512	532
	100	99	100	96	98	101	102	99	101	96	98	106	107	104	106	99	103
SC	516	512	513	499	512	522	523	515	520	498	509	551	555	536	547	511	530
(E/G)	100	99	99	97	99	101	101	100	101	97	99	107	108	104	106	99	103
SC	514	510	515	501	510	519	519	513	520	501	510	530	528	521	529	511	520
(C/E)	100	99	100	98	99	101	101	100	101	98	99	103	103	102	103	100	101

Table 2: A comparison of identification probabilities for cases involving two moderately overlapping distributions with particular correlation coefficients. Six models are compared: MSD (I.O.), ideal observer; MSD (B.L.), best linear classifier; MSD (M.D.), minimum distance; PMDS; SC (E/G), Euclidean/Gaussian; SC (C/E), city-block/exponential decay. The first row of numbers for each model are identification probabilities (x10^3); the second row are values relative to the (0,0) correlation case.

CASES (Moderate Overlap)
Correlation Coefficients

Model	*0,0*	*0,.4*	*0,-.4*	*0,.8*	*0,-.8*	*.4,0*	*-.4,0*	*.4,.4*	*-.4,-.4*	*.4,.8*	*-.4,-.8*	*.8,0*	*-.8,0*	*.8,.4*	*-.8,-.4*	*.8,.8*	*-.8,-.8*
MSD	842	847	850	859	890	804	896	801	902	818	936	815	961	796	967	772	987
(I.O.)	100	101	101	102	106	96	106	95	107	97	111	97	114	95	115	92	117
MSD	846	846	830	905	899	806	925	805	902	842	938	710	967	739	984	774	997
(B.L.)	100	100	98	107	106	95	109	95	107	100	111	84	114	87	116	91	118
MSD	841	841	841	841	841	803	902	803	902	803	902	774	987	774	987	774	987
(M.D.)	100	100	100	100	100	95	107	95	107	95	107	92	117	92	117	92	117
PMDS	712	693	731	671	751	705	730	864	748	660	767	708	759	687	775	660	795
	100	97	103	94	105	99	103	96	105	93	108	99	107	96	109	93	112
SC	702	683	722	664	740	693	717	676	731	657	753	693	740	675	755	652	774
(E/G)	100	97	103	95	105	99	102	96	104	94	107	99	105	96	105	93	110
SC	658	649	666	641	673	655	661	647	671	636	676	649	666	642	673	632	678
(C/E)	100	99	101	97	102	99	100	98	102	97	103	99	101	98	102	96	103

Figure 1. The probability of correct identification for six probabilistic identification models in the case of high distributional overlap [means of (0,0) and (0.707, 0.707)]. The multidimensional signal detection model, MSD (ideal observer) has been plotted against: A, the MSDBL (best linear classifier); B, the MSDMD (minimum distance rule); C, the PMDS model; D, the SCEG (Euclidean/Gaussian) model; and E, the SCCE (city-block/exponential decay) model.

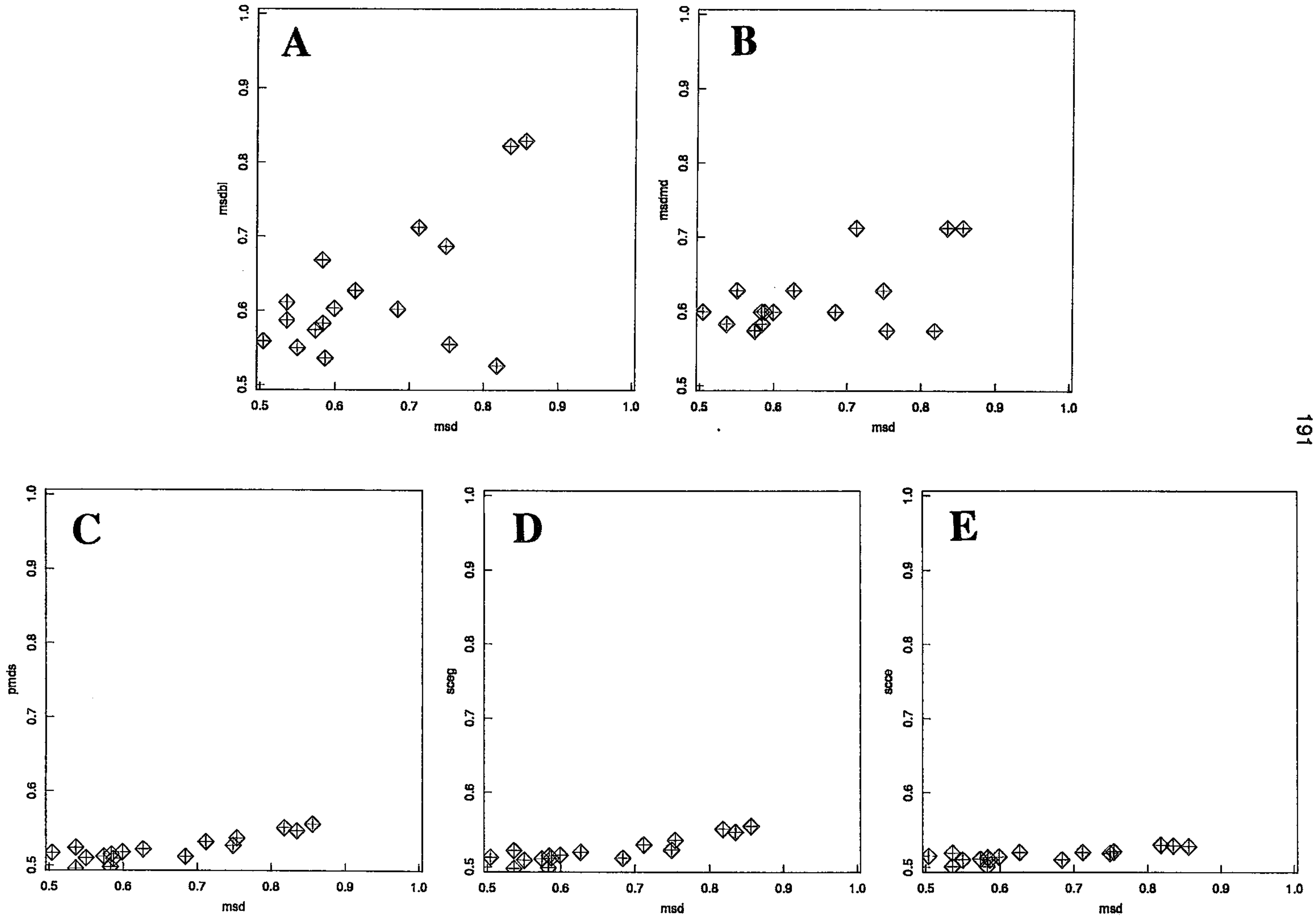

A
B
C
D
E
msdbl
msdmd
pmds
sceg
scce
msd
0.5
0.6
0.7
0.8
0.9
1.0

Figure 2. The probability of correct identification for six probabilistic identification models in the case of moderate distributional overlap [means of (0,0) and [1.414, 1.414)]. (See Figure 1 for captions.)

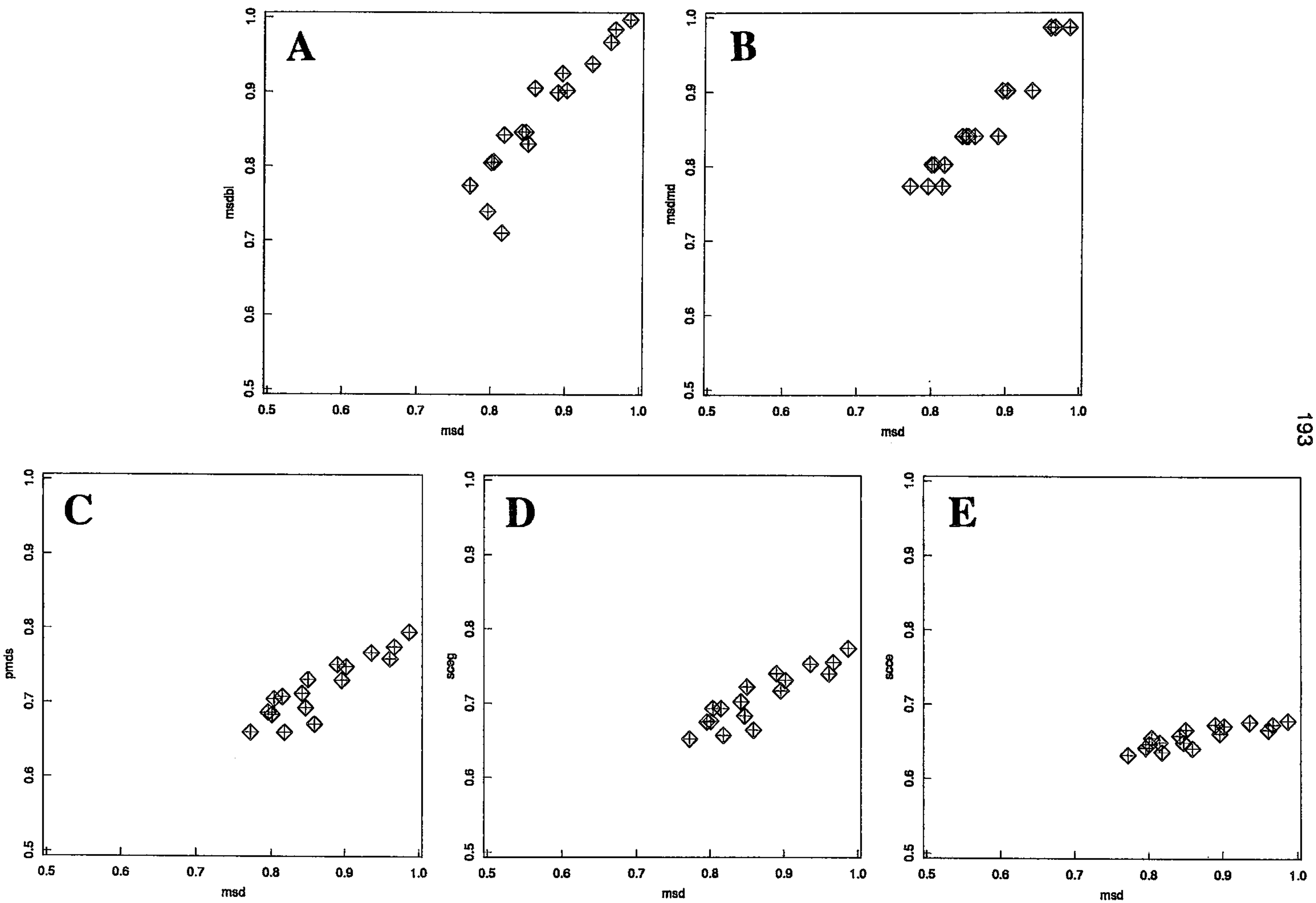
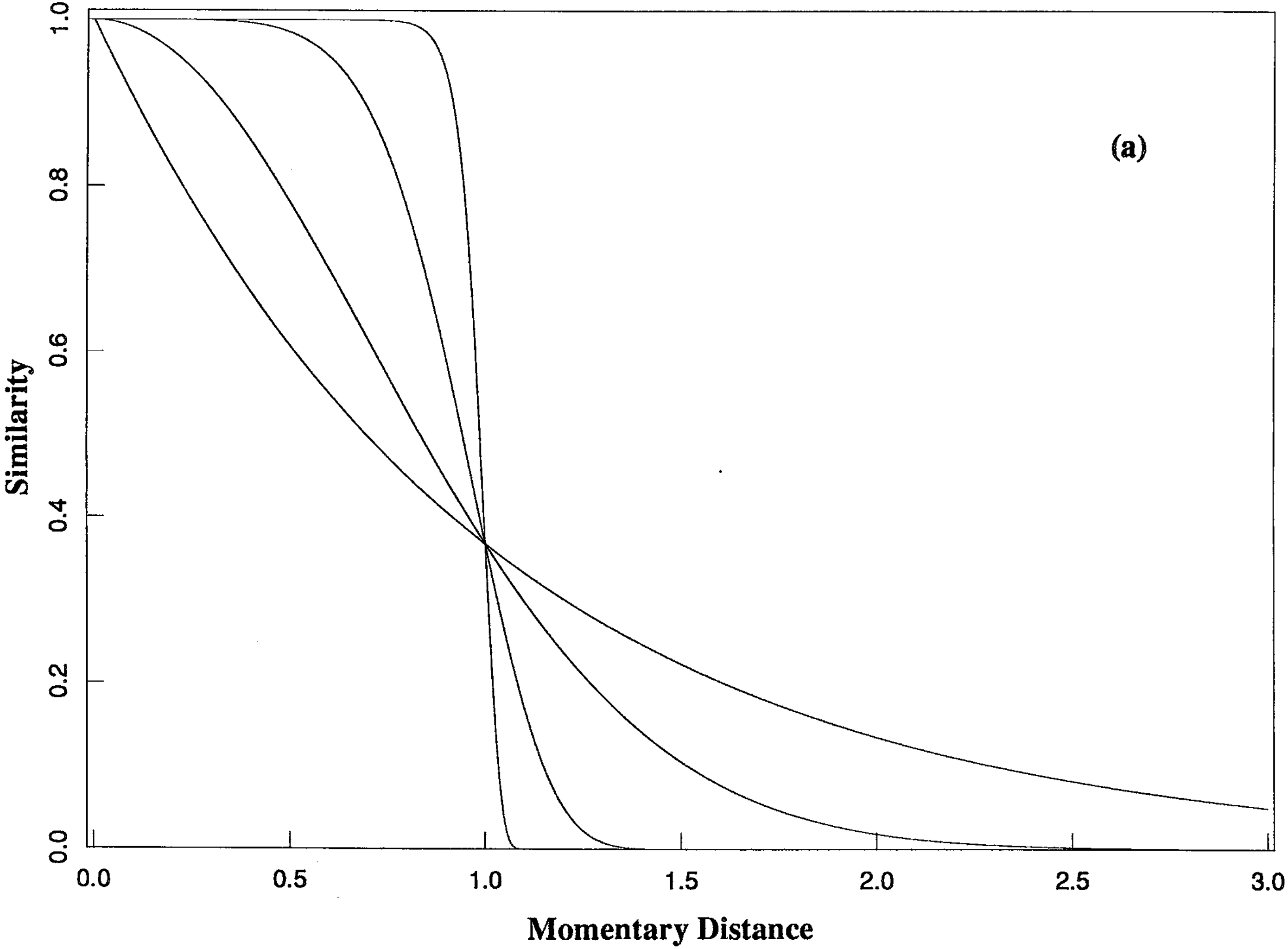
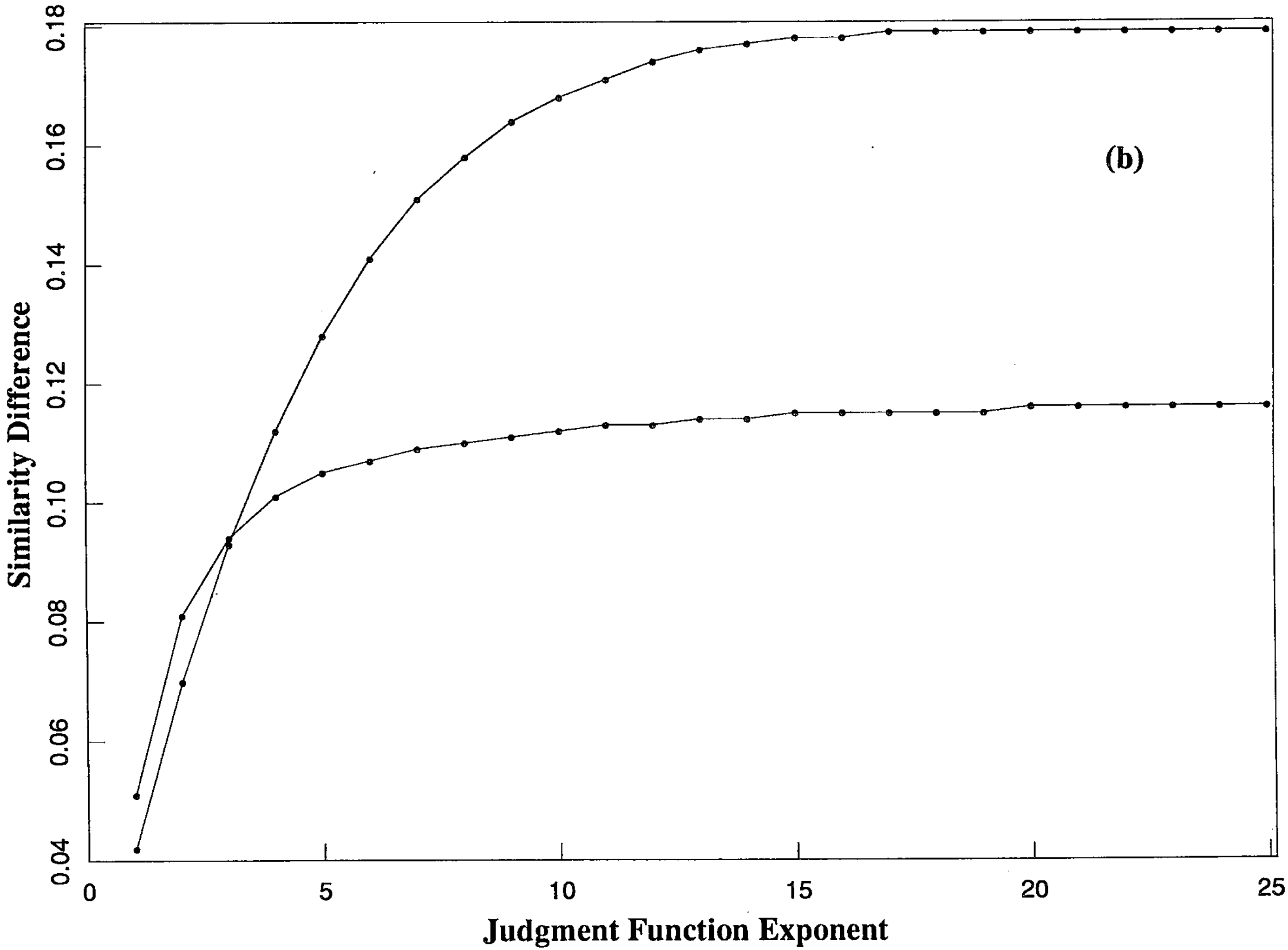

A
B
C
D
E
msdbl
msdmd
pmds
sceg
scce
msd
0.5
0.6
0.7
0.8
0.9
1.0

Figure 3. The effect of the form of g, the judgment function, on the sensitivity of its expected value, *similarity*, to perceptual dependence. (a) Four judgments of the form, $g(d) = exp(-d^{\alpha})$, where $\alpha = 1, 2, 6$ and 25. (b) The difference between the similarities of two pairs of distributions: In the first pair, correlations between dimensions are both -0.8; in the second pair, correlation coefficients between dimensions are both +0.8. In both cases, $\delta = 1.0$. The curve with the higher asymptote corresponds to distributions with a standard deviation of 0.2, the lower curve corresponds to distributions with a standard deviation of 1.0.


(a)
Similarity
Momentary Distance
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.5
1.0
1.5
2.0
2.5
3.0


Similarity Difference
0.04
0.06
0.08
0.10
0.12
0.14
0.16
0.18
(b)
0
5
10
15
20
25
Judgment Function Exponent

CHAPTER 5

GENERAL DISCUSSION AND CONCLUSIONS

Dynamic changes in the chemistry and biology describing the interaction between an organism and its environment, suggest that the "psychological quantities" governing behavior should not be static. It is not surprising that static, or deterministic, models of perception should ultimately prove to be limited in explaining experimental results. These limitations should be expected to become most evident when the objects being studied are easily confused.

In food and beverage sensory evaluation, the duo-trio and triangular methods have been used extensively to determine differences between pairs of objects. Unidimensional modelling of these methods revealed that they are special cases of more general models of methods applied in the early development of multidimensional scaling. The duo-trio is a special case of Torgerson's method of triads and the triangular method is a special case of Richardson's method of triads. The models discussed by Torgerson (1958) for Torgerson's and Richardson's methods of triads, however, did not involve classical Thurstonian assumptions about the psychological magnitudes (normally distributed) but assumed, instead, that distances between values were normally distributed. Thurstonian models in unidimensional and multidimensional cases have now been derived and evaluated for these methods. In many of the models discussed in the thesis, the objects must be perceptually confusable or the measured response would be, trivially, 1.0. Triad methodology offers the opportunity to determine the parameters of distributions which may not overlap, but for which the within-trial distances between pairs in a triad are confusable. If resampling occurs within a trial, it was pointed out that the same two stimuli may appear to be most alike and most different in Richardson's method of triads. Since this may lead to unpredictable behavior by the subject, the experimenter should be aware of this possibility if using Richardson's method or its special case, the triangular method.

An interesting connection exists between triad models and those involving preferential choice and two-alternative identification. By replacing one of the stimuli in a triad by an ideal point, the triad model becomes a preference model in which the subject makes a decision on the basis of the relative distances between the stimuli and the ideal point. If two of the stimuli in a triad are replaced by memory exemplars, then the model becomes a two-alternative identification model and is a prototype for m-alternative identification. These three applications for the same model provided the motivation to find a mathematically and computationally simple form for the n-dimensional Thurstonian variant of Torgerson's method of triads. This form was obtained using existing theory on quadratic forms in normal variables. In future work, it would be satisfying if the same parameters fit data obtained from two or more of these methods (for example, triads and preferential choice).

The similarity of two objects has been modelled as a monotonically decreasing function of the perceptual distance between the objects (Shepard, 1987). Some multidimensional scaling models are based on this assumption. Shepard (1987) has argued in favor of an exponential decay similarity function as a general principle and, for stimuli for which the perceptual components can be attended to separately, he has suggested that the distance metric is city block (the L_1 norm). Some exceptions to these two ideas were shown to arise with confusable stimuli when a deterministic model is used. Specifically it has been found that for these types of stimuli, support for the Euclidean metric and a Gaussian similarity function was obtained. This paradox, called the *Shepard-Nosofsky* paradox in the thesis, can be resolved by using an appropriate probabilistic model to represent stimuli as distributions of multivariate normally distributed momentary values. The result of applying this model was to show that the exponential decay similarity function and the city block metric may model behavior at the individual trial level, although a deterministic analysis of expected values may not reveal this.

A closed form for the expected value of the Gaussian similarity function of Euclidean distance was derived. This equation provides predicted "same-different" proportions in the absence of response bias. This model was applied to artificial and real data sets on same-different judgments to estimate parameters (means and variances) using nonlinear least squares. It was found that perceptual variances for morse code signals are smaller for stimuli isolated from other stimuli and larger when stimuli are in close proximity to others. This difference in variance provides a way of modelling observed differences between stimuli in self-similarity. Isolated stimuli would have higher values for self-similarity than less isolated ones.

Not all probabilistic models respond equally to perceptual dependence. In general, models involving linear or close to linear judgment functions will be insensitive or relatively insensitive to correlated perceptual dimensions. Judgment functions closer to step functions will be far more sensitive to perceptual dependence. For this reason, the Gaussian similarity function was more sensitive than the exponential decay function in a probabilistic model. Comparisons were made among models which involve distance comparisons within a trial (either ordinal decision rules or monotonically decreasing similarity functions) and signal detection models (where the decision is based on the position of a vector relative to a decision boundary). The signal detection models were more sensitive to perceptual dependence than the distance-based models. Since these two classes of models differ with regard to assumptions which could be affected by training, perceptual dependence may be used as an experimental tool in studying the effect of training on decision rules employed by subjects. Distance-based rules may be used until response regions are learned. This hypothesis would imply that sensitivity to perceptual dependence should increase with training.

Most of the papers in the thesis have been concerned with *Thurstonian* models. This means that subjects' behavior is modelled in terms of the parameters of normally distributed (uni- or multidimensional) momentary psychological magnitudes and an appropriate judgment function. In some respects, this structure imposes fairly severe restrictions and provides no direct information on the nature of the function which transduces physicochemical stimulus information into perceptual information. One also has to assume that there is no stimulus variability, if one assumes that the noise present is perceptual. These limitations can be removed by building models that start by assuming the existence of a stimulus probability density function. To this level, a psychophysical transformation is applied and, subsequently, perceptual or psychological variance is added. It is thus possible to separate these three components which generate momentary psychological magnitudes. These models are called *Fechner-Thurstone* models. One does not require a normal probability density function for these magnitudes and, in fact, very special conditions would be required to create such a density function. It might be imagined that this type of approach to modelling psychological tasks would lead to a larger number of parameters than traditional Thurstonian modelling. On the contrary, it was shown that this type of model is highly parameter efficient because it makes use of known physicochemical information which can be measured directly, eliminating the need to estimate psychological scale values. Thus, for instance, a 9 parameter Thurstonian model of data from 10 stimuli was shown to be modelled by only 2 parameters with a Fechner-Thurstone model. These models also provide a means of determining the form and parameters values of psychophysical functions without having to employ magnitude estimation or other direct scaling techniques. These ideas were used to develop a new model for paired comparisons on the basis of this theory.

From a modelling standpoint, one can imagine that stimulation of the senses leads to psychological magnitudes that are used to make particular decisions which determine behavior. These magnitudes should not be viewed as static, but fluctuate from moment to moment. These ideas, of course, do not imply that psychological magnitudes or their fluctuations actually exist; they simply capture a particular way of thinking that might be useful in modelling behavior. It is clear that probabilistic models ought to be chosen over deterministic models when stimuli are confusable, because there is a need to formally include the effects of momentary fluctuations in the organism's perception of stimuli. An ever-present difficulty, which has been resolved in some cases in the thesis, is the computational intensity of many of these models. In fact, advances in computational speed and accuracy will be needed before the full potential of these models can be realized. These advances should be closely linked to computationally efficient mathematical forms of the models. The ability to explain complex behavior over many different types of tasks in a simple, comprehensive, consistent manner should provide the motivation for these needed developments.

References

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

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

CHAPTER 6

SUMMARY

CHAPTER ONE

The origin of the work reported in the thesis is traced to a single paper by Frijters (*British Journal of Mathematical and Statistical Psychology*, 32, 229-241) in which unidimensional models for the triangular method are discussed. This paper stimulated the author's interest in developing a multidimensional model for the same method. These ideas were then extended to the duo-trio method. This modelling work initiated an interest in developing probabilistic models, first of similarity, then of identification, triads in general, and preferential choice. The organization of the thesis around these ideas is presented.

CHAPTER TWO

A central idea of the thesis is the formal inclusion of momentary fluctuations in psychological magnitudes in models of perception. This chapter contains two papers which are overviews of the topic of probabilistic models by the author and others. Models for similarity, identification, triad discrimination, and preferential choice are given.

In the first paper, the expected value of a judgment function of the psychological magnitudes sets the stage for all of the models discussed in the paper. Similarity models and several kinds of identification models are presented with a short discussion of the *Shepard-Nosofsky* paradox. Following a discussion of nonlinear least squares estimation, *Fechner-Thurstone* models are introduced.

In the second paper in this Chapter, the connection between triad discrimination, preferential choice, and two-alternative identification is given. A computationally simple form for the 2n-fold integral expression which yields predicted responses for all three of these models is discussed. The relationship between this model and previous unidimensional and multidimensional models is explained.

CHAPTER THREE

Unidimensional triad theory for Richardson's and Torgerson's methods of triads is derived and evaluated. Parameter estimation is illustrated with a sample problem. The duo-trio method is shown to be a special case of Torgerson's method of triads and the triangular method is a special case of Richardson's method of triads. A special problem in Richardson's method arises when resampling within a trial occurs. In Richardson's method, the subject is instructed to select the pair in a triad which is "most similar", and then to select the pair which is "most different". It is possible, under resampling, for the same two objects to be both "most similar" and "most different". This problem is discussed and the probability of, what is called, a *decision conflict* is estimated using Monte Carlo simulation.

The probability of a decision conflict is derived in the second paper under the resampling assumption. Some suggestions for experimental work are given and the dependence of this probability on the distance between the extreme stimulus means in a triad is shown.

The third paper in this Chapter describes *Fechner-Thurstone* models. These models, not strictly unidimensional, are concerned with relating stimulus parameters through a psychophysical transformation to psychological values used to make a decision in a particular task. The stimulus magnitudes are assumed to follow a probability density function (pdf) and psychological noise is added from a different pdf. No particular form for these pdfs is required and the psychophysical transformation is also very general, provided it is a *one-to-one* function. An application to paired comparisons is given and parameter estimates of a sample problem are obtained for special cases of the pdfs and the psychophysical transformation.

CHAPTER FOUR

A Monte Carlo study of the triangular method under multivariate assumptions revealed that the probability of a correct response decreases with dimensionality for a fixed distance between stimulus means. A multivariate model for the triangular method is given in the second paper of this Chapter and the model is evaluated numerically in two dimensions. The effects of variance and correlation structure on the probability of a correct response are discussed. These evaluations clearly show that the probability of a correct response and the distance between stimulus means are not monotonically related (as required by traditional multidimensional scaling). Variance effects on the triangular method results are investigated, particularly the consequence of variance inequality on Type I errors.

The third paper in this Chapter discusses models of both the triangular and duo-trio methods with special emphasis on the nonmonotonic relationship between the probability of a correct response and the distance between the means of the stimulus distributions. A general scheme for modelling the results of discrimination tasks is given when stimulus and psychological variability is assumed to exist. These ideas were formally discussed as Fechner-Thurstone models in the third paper in Chapter 3.

The next paper in this Chapter is concerned with probabilistic models of similarity and provides a sketch of how these ideas may be used to model identification. The main contribution of this paper is to show how probabilistic models are needed to resolve a metric and judgment function paradox created when a deterministic model is used to represent data from confusable stimuli. A closed form for the probabilistic similarity model is given when the distance metric is Euclidean and the judgment function is Gaussian. Parameter estimation for artificial and real data sets (morse code signals) is discussed. The paper following the similarity paper makes specific comment on a resolution of the paradox arising from Shepard's and Nosofsky's deterministic modelling.

The sixth paper in this Chapter provides a comparison of six probabilistic models of identification with respect to their sensitivities to perceptual dependence, or the degree to which the perceptual dimensions are correlated. From this comparison, it is shown that signal detection models involving some form of decision boundary (linear or quadratic) are more sensitive to perceptual dependence than within trial distance-based models. These latter models may involve an ordinal decision rule (the decision is based only on an ordinal distance comparison) or a distance-based similarity function. The least sensitive model involved the exponential decay similarity function of city block distance. This result is interesting because this last model is based on assumptions concerning the distance metric and similarity function that Shepard has argued are universal principles for all sentient organisms.

CHAPTER FIVE

This Chapter contains the general discussion and conclusions of the thesis. The scope, strengths and limitations of the models described in the thesis are discussed.

CHAPTER 7

SAMENVATTING (SUMMARY IN DUTCH)

HOOFDSTUK EEN

De oorsprong van de studie waarover in dit proefschrift wordt gerapporteerd kan worden herleid tot één bepaald artikel van de hand van Frijters (*British Journal of Mathematical and Statistical Psychology*, **32**, 229-241) waarin unidimensionele modellen voor de driehoeksmethode worden besproken. Dit artikel wekte de belangstelling van de auteur voor het ontwikkelen van een multidimensioneel model voor dezelfde methode. Vervolgens werden dezelfde ideeën toegepast op de duo-trio methode. Deze modelstudies initieerde een belangstelling voor het ontwikkelen van waarschijnlijkheids modellen; eerst voor gelijkheid, dan voor identificatie, triaden in het algemeen, en preferentiële keuze. De indeling van dit proefschrift rond deze gedachten wordt besproken.

HOOFDSTUK TWEE

Een centraal idee in dit proefschrift is het formeel opnemen in perceptiemodellen van momentane fluctuaties van psychologische grootheden. Dit hoofdstuk bevat twee artikelen welke overzichten zijn van waarschijnlijkheidsmodellen van de auteur en anderen. Er worden gelijkheids, identificatie, triadische discriminatie en preferentiële keuze modellen gegeven.

In het eerste artikel wordt uiteengezet dat de verwachte waarde van een beoordelingsfunctie ten grondslag ligt aan alle modellen die in het artikel besproken worden. Gelijkheidsmodellen en verscheidene soorten van identificatiemodellen worden gepresenteerd met een korte bespreking van de *Shepard-Nosofsky* paradox. *Fechner-Thurstone* modellen worden geïntroduceerd na een discussie over nonlineaire kleinste kwadraten schatting.

In het tweede artikel van dit Hoofdstuk wordt het verband tussen triadische discriminatie, preferentiële keuze en 2-alternatieve identificatie besproken. Een rekenkundig eenvoudige vorm voor de 2n-voudige integrale uitdrukking welke de voorspelde responsies voor alle drie methoden voorspeld wordt besproken. De relatie tussen dit model en de voorafgaande unidimensionele en multidimensionele modellen wordt verklaard.

HOOFDSTUK DRIE

Een unidimensionele triaden theorie voor Richardson's en Torgerson's methoden van triaden wordt ontwikkeld en geëvalueerd. De schatting van de parameters wordt geïllustreerd aan de hand van een steekproef probleem. Er wordt aangetoond dat de duo-trio methode een speciaal geval is van Torgerson's methode van triaden en dat de driehoeksmethode een speciaal geval is van Richardson's methode van triaden. Een speciaal probleem onstaat in Richardson's methode wanneer "resampling" binnen een aanbieding toegestaan is. In Richardson's methode wordt de proefpersoon geïnstrueerd om het paar in een triade te selecteren dat "meest gelijk" is, en vervolgens het paar de selecteren dat "meest verschillend" is. Het is mogelijk onder de voorwaarde van "resampling" dat dezelfde twee objecten als zowel het "meest gelijk" als het "meest verschillend" worden beoordeeld. Dit probleem wordt besproken en de waarschijnlijkheid van wat genoemd is een *beslissingsconflict* wordt geschat door gebruik te maken van een Monte Carlo simulatie.

De waarschijnlijkheid van een beslissingsconflict wordt afgeleid in het tweede artikel onder de aanname van "resampling". Enkele suggesties voor experimenteel onderzoek worden gegeven en de afhankelijkheid van deze waarschijnlijkheid van de afstand tussen de extreme stimulus gemiddelden in een triade wordt aangetoond.

Het derde artikel in dit Hoofdstuk beschrijft *Fechner-Thurstone* modellen. Deze modellen, die strikt gesproken niet unidimensioneel zijn, relateren stimulus parameters door middel van een psychofysische transformatie aan de psychologische waarden die vervolgens gebruikt worden om een beslissing te nemen in een bepaalde taak. De stimulus grootheden worden geacht verdeeld te zijn volgens een waarschijnlijkheidsdichtheidsfunctie (pdf) en psychologische ruis wordt aan hen toegevoegd van een andere pdf. Er is geen bijzondere vorm van deze pdf's vereist en de psychologische transformatie is zeer algemeen, onder voorwaarde dat het een *een-op-een* functie is. Een toepassing hiervan wordt gegeven voor de methode van paarsgewijze vergelijking en parameter schattingen van een steekproef probleem worden verkregen voor speciale gevallen van pdf's en psychologische transformatie.

HOOFDSTUK VIER

Een Monte Carlo studie van de driehoeksmethode onder multivariate aannamen liet zien dat de kans op een juiste uitspraak afneemt met de dimensionaliteit voor een vaste afstand tussen stimulus gemiddelden. Een multivariaat model voor de driehoeksmethode wordt gegeven in het tweede artikel van dit Hoofdstuk en het model wordt numeriek geëvalueerd in twee dimensies. De effecten van de variantie en de correlatie structuur op de kans op een juiste uitspraak wordt besproken. Deze evaluaties laten duidelijk zien dat de kans op een juiste uitspraak en de afstand tussen de stimulus gemiddelden niet monotoon zijn gerelateerd (zoals vereist in traditioneel multidimensionele schaalmodellen). De effecten van variantie op resultaten van de driehoeksmethode worden onderzocht, in het bijzonder de gevolgen van ongelijkheid van varianties op Type I fouten.

Het derde artikel van dit Hoofdstuk bespreekt modellen zowel voor de driehoeksmethode en de duo-trio methode met speciale nadruk op de non-monotone relatie tussen de kans op een juiste uitspraak en de afstand tussen gemiddelden van stimulus verdelingen. Een algemeen schema voor het modeleren van resultaten van discriminatie taken wordt gegeven onder de aannamen dat er zowel sprake is van stimulus als van psychologische variabiliteit. Deze ideeën werden formeel besproken als Fechner-Thurstone modellen in Hoofdstuk 3.

Het volgende artikel in dit Hoofdstuk gaat over waarschijnlijkheids modellen van gelijkheid en geeft een overzicht over hoe deze ideeen gebruikt kunnen worden om identificatie to modelleren. De belangrijkste bijdrage van dit artikel is dat het aantoont hoe noodzakelijk waarschijnlijkheidsmodellen zijn om een metrische en beoordelingsfunctie paradox op te lossen, die ontstaat wanneer een deterministisch model wordt gebruikt om data van verwarbare stimuli te representeren. Een gesloten vorm voor het waarschijnlijkheids gelijkheidsmodel wordt gegeven in het geval dat de afstandsmetriek Euclidisch is en de beoordelingsfunctie Gaussiaans. Parameter schatting voor zowel kunstmatige als echte data set (morse code signalen) wordt besproken. In het artikel volgend op het gelijkheids artikel gaat met name in op de oplossing van de paradox die voortkomt uit Shepard's en Nosofsky's deterministisch modelleren.

Het zesde artikel in dit Hoofdstuk geeft een vergelijking van zes waarschijnlijkheidsmodellen met betrekking tot de gevoeligheid voor perceptuele afhankelijkheid, of the mate waarin de perceptuele dimensies zijn gecorreleerd. Door deze vergelijking wordt aangetoond dat signaal-detectie modellen die een of andere beslissingsgrens (lineair of kwadratisch) incorporeren veel gevoeliger zijn voor perceptuele afhankelijkheid dan binnen aanbieding afstandsmodellen. Deze laatste modellen kunnen een ordinale beslissingsregel hebben (d.i. de beslissing is alleen gebaseerd op ordinale afstands vergelijking) of een gelijkheidsfunctie gebaseerd op afstand. Het minst gevoelige model gaat uit van een exponentiële-verval gelijkheidsfunctie functie van "city-block" afstand. Dit resultaat is belangrijk omdat het laatste model gebaseerd is op aannamen met betrekking tot de afstandsmetriek en gelijkheidsfunctie waarover Shepard heeft beweerd dat deze universeel is voor alle waarnemende organismen.

HOOFDSTUK VIJF

Dit hoofdstuk bevat de algemene discussie en conclusies van dit proefschrift. De reikwijdte, sterke kanten en beperkingen van de modellen die in dit proefschrift beschreven zijn, worden besproken.

CURRICULUM VITAE

Daniel M. Ennis was born in Dublin, Ireland on January 15, 1950. He studied for the Leaving Certificate at Chanel College, Dublin which he completed in 1967. Daniel Ennis then studied at the National Botanic Gardens, Dublin for one year prior to entering University College, Dublin in 1968 to study for a degree in Agricultural Science (Horticulture). Upon completing his degree in 1972, he attended the University of Maryland, U.S.A. to study Food Science, and was awarded the M.S. in May, 1974 and the Ph.D. in December, 1975. Dr. Ennis' Ph.D. dissertation was on the microbial metabolism of synthetic polyamides.

Following post-doctoral employment with the United States Department of Agriculture, Dr. Ennis became an Assistant Professor in the Department of Food Scientist at the University of Guelph, Ontario, Canada. He taught and conducted research on food quality and began to develop an interest in sensory evaluation.

In 1978, Dr. Ennis joined the Research Department of Frito-Lay Inc., Dallas, Texas and in 1979 joined the research staff at Philip Morris in Richmond, Virginia. At Philip Morris, Dr. Ennis was extensively involved in all aspects of sensory and consumer research. This experience exposed him to the need for the development of a stronger theoretical foundation for the measurement of sensory perceptions. This interest, which began in 1984, has been pursued by Dr. Ennis since then. In the last two years, he has also begun to develop an interest in developing mathematical models of the molecular processes responsible for chemical sensing.