

**MULTIDIMENSIONAL
MODELS OF
PERCEPTION AND
COGNITION**

Edited by

F. Gregory Ashby

University of California at Santa Barbara



1992

LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
Hillsdale, New Jersey Hove and London

2 A Probabilistic Multidimensional Scaling Approach: Properties and Procedures

Joseph L. Zinnes
Temple University

David B. MacKay
Indiana University

Multidimensional Scaling has had a relatively long history, starting perhaps with the foundational paper by Young and Householder (1938), but, as evidenced by this book, only recently does it seem to be making contact with other areas of psychology. We like to think that this has resulted from probabilistic, multidimensional approaches gaining wider acceptance.

Our aim in this chapter is to review some of the main features of the probabilistic model that we have been working with for some time, covering some technical details, but mainly focusing on some of its properties and empirical applications. Our hope is that this introduction will make the probabilistic Multidimensional Scaling approaches more amenable and interesting to the reader, and even, perhaps, to encourage the reader to explore further their potentialities as explanatory theories.

FOUNDATIONS

This section shows how the general model has been applied to different experimental tasks involving similarity or preference judgments. In the following two sections some of the more important properties of the model are highlighted, as indicated by various simulations and empirical findings.

The General Model

As do many other authors in this volume, we assume that the perceptual aspects of each stimulus can be represented by a random vector having a multivariate

normal distribution. Specifically, for stimulus S_j , $j = 1, \dots, m$, we let $\mathbf{X}_j = (X_{j1}, \dots, X_{jr})$ be the corresponding r -dimensional random vector and assume that it has an r -variate normal distribution with mean vector $\boldsymbol{\mu}_j = (\mu_{j1}, \dots, \mu_{jr})$ and covariance matrix $\boldsymbol{\Sigma}_j$. To simplify the discussion, we treat only the case in which the components $X_j = (X_{j1}, \dots, X_{jr})$ are independently distributed, and therefore we assume that the covariance matrix $\boldsymbol{\Sigma}_j$ is just the diagonal matrix $\mathbf{D}(\sigma_{jh}^2)$, $h = 1, \dots, r$. This does not result in any appreciable loss of generality, but we do not pursue these details here (see chap. 1).

Similarly, it will be helpful to assume that all random vectors $\mathbf{X}_1, \dots, \mathbf{X}_m$ are independent, although this assumption also can be relaxed with little additional effort.

There are two possible interpretations of the variance parameter σ_{jh}^2 , depending on the level of the analysis to be performed. For individual analyses, the variances $\sigma_{j1}^2, \dots, \sigma_{jr}^2$ reflect the degree of unfamiliarity or uncertainty that the individual has concerning the nature of stimulus S_j on each of its r dimensions. For group analyses, the variances instead indicate how heterogeneous the people in the group are with respect to their perception of stimulus S_j . In the following sections, we focus mainly on the individual and, therefore, generally refer to the variances as an uncertainty or unfamiliarity parameter.

Preferences can be incorporated into the general model by associating random vectors with individuals as well as with stimuli. Thus, for person P_i , $i = m + 1, \dots, m + n$, we let $\mathbf{X}_i = (X_{i1}, \dots, X_{ir})$ be the associated r -dimensional random vector, and assume, as we did with the stimulus vectors, that it has an r -variate normal distribution with mean vector $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{ir})$ and covariance matrix $\boldsymbol{\Sigma}_i = \mathbf{D}(\sigma_{ih}^2)$. The person or ideal random vector \mathbf{X}_i is intended to represent the most preferred or ideal stimulus for person P_i . This aspect of the general model is, of course, precisely equivalent to the unfolding model of Coombs (1964).

These multivariate normal distribution assumptions, first given by Hefner (1958), are a direct multidimensional generalization of Thurstone's pair comparison model (1927a). Hefner's extension, while direct, was by no means obvious. Earlier attempts to generalize Thurstone's single-dimensional model (Klingberg, 1941; Richardson, 1938; Torgerson, 1951) assumed that the distances between pairs of points, not the points themselves, have a univariate normal distribution. This assumption is quite compelling because it means that interpoint distances can be treated as single-dimensional stimuli. The single-dimensional model of Thurstone can, therefore, be applied directly to judgments of pairs of interpoint distances to estimate their mean value, at least up to an additive constant. Thus, a multidimensional approach was achieved by using purely single-dimensional methods.

There are, however, some problems with this approach: The interpoint distances are only determined up to an additive constant; that is, they are only measured on an interval scale. Unlike the single-dimensional case, the effects of

an additive constant in the multidimensional case are not benign. Adding a constant to a set of interpoint distances can change drastically the nature of the configuration of the points involved and, even further, their dimensionality as well. In fact, for m points, if the additive constant is sufficiently large, all the interpoint distances will tend to be equal, and the points themselves, therefore, will tend to be located at the vertices of a regular polyhedron in $m - 1$ dimensions. Hefner's approach, in addition to being intuitively attractive, avoids the additive constant problem. It is possible, with Hefner's assumptions, to obtain maximum likelihood (ML) estimates of all the relevant parameters—the means and variances of stimulus and ideal points—and to do this without encountering serious nonuniqueness issues.

However, a price has to be paid for the conceptual simplicity of the Hefner model, and this, no doubt, accounts for its unpopularity. The functions that make up each term of the likelihood function are, under the Hefner model, complex functions and, worst still, cannot be expressed in closed form. Despite this, it has been possible to develop reasonably simple, fast, and moderately accurate algorithms for obtaining the ML estimates for the location and variance parameters.

It will take us too far afield to cover in detail all the ML algorithms. Instead, we describe some of the main aspects of these algorithms and give references where more detailed information can be obtained.

Generally speaking, the ML algorithms that we have used have three key ingredients. First, they use simple approximations of the probability density functions that compose each term of the likelihood function. This aspect of the algorithm is essential, because the series expansions for these functions converge quite slowly for many values of the parameters to be estimated. Iterative procedures, which must of necessity be used to maximize the likelihood function, would have no hope of terminating in a reasonable period of time if, on each iteration and for each term of the likelihood function, it were necessary to evaluate a slowly converging series expansion.

Second, these ML algorithms use simple expressions for obtaining good initial estimates for the parameters to be estimated. Having good initial estimates not only speeds the rate at which the iterative process converges but it also makes it more likely that the convergence will terminate at a global maximum rather than at a local maximum. To obtain these initial parameter values, we have found that it has been sufficient to make just one simplifying assumption: the variances of the points—both the stimulus and ideal points—are assumed to be small relative to the size of most interpoint distances.

Third, the ML algorithms utilize general-purpose optimization programs (Chandler, 1969; International Mathematical and Statistical Libraries [IMSL], 1979) to control the iterative process. The value of using these programs is that they do not require a knowledge of the first or second derivatives of the function to be maximized. Given the complexity of the likelihood functions that arise

when the Hefner model is used, it would not be practical to use iterative procedures, such as Steepest Descent Methods, that require analytical expressions for the derivatives.

Special Cases

Special cases of the Hefner model can be defined that are analogous to Thurstone's case 3 and case 5. However, due to the multidimensional nature of the Hefner model, it is necessary to break down the case 3 and case 5 assumptions into two subcases or conditions: an isotropic condition and an anisotropic condition. The distinction between these conditions has to do with the variance of the components of \mathbf{X}_j over the r dimensions. If these variances are not correlated and do not change from dimension to dimension, the space is said to be isotropic. Otherwise, the space is anisotropic. Thus, all the directions of an isotropic space are equivalent, at least with respect to variability of the stimulus or ideal point. In contrast, anisotropic space exhibits directionality effects.

This leads us to the following definition.

Definition 2.1. Assume that the random vectors $\mathbf{X}_j = (X_{j1}, \dots, X_{jr})$ satisfy the Hefner model; that is, they have the multivariate normal distribution $N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$.

Then, for the isotropic condition,

(i) case 3i holds if and only if

$$\boldsymbol{\Sigma}_j = \mathbf{D}(\sigma_j^2);$$

(ii) case 5i holds if and only if

$$\boldsymbol{\Sigma}_j = \mathbf{D}(\sigma^2).$$

For the anisotropic condition,

(iii) case 3a holds if and only if

$$\boldsymbol{\Sigma}_j = \mathbf{D}(\sigma_h^2), \quad h = 1, \dots, r;$$

(iv) case 5a holds if and only if

$$\boldsymbol{\Sigma}_j = \mathbf{D}(\sigma_h^2), \quad h = 1, \dots, r.$$

Figure 2.1 illustrates these four cases graphically. As shown, the equal likelihood contours for all the points in the isotropic condition (cases 3i and 5i) are circles, whereas in the anisotropic condition (cases 3a and 5a) they are ellipses. The graphical distinction between cases 3 and 5 is equally transparent. In case 3, the sizes of the circles or ellipses differ from point to point, but in case 5 they do not.

The isotropic condition, cases 3i and 5i, is discussed in this and the following sections. The anisotropic condition is taken up in the third section.

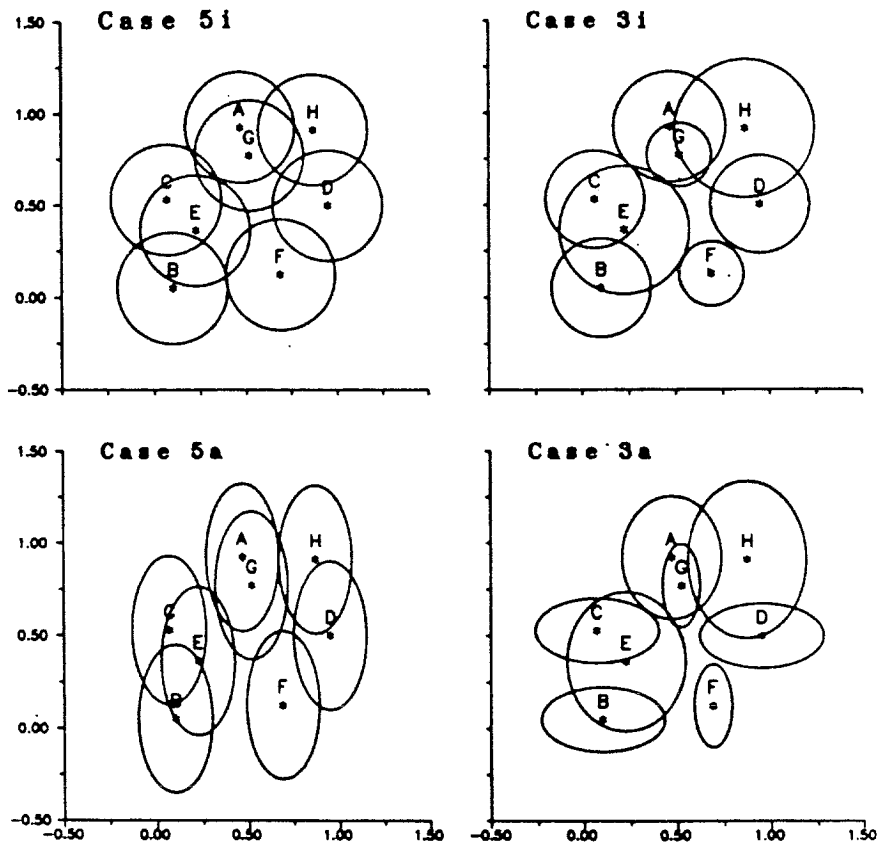


FIG. 2.1. Isotropic and anisotropic cases of the Hefner model.

Similarity

Perhaps, the simplest application of the Hefner model is in the similarity domain. Although there are a number of different similarity or dissimilarity judgments that subjects can be asked to make, depending on whether the stimuli are presented in pairs, triples, or quadruples, we consider just one of these judgments: the direct dissimilarity judgment of a pair of stimuli (MacKay & Zinnes, 1981; Zinnes & MacKay, 1983). This should suffice to illustrate some of the main issues involved in applying the Hefner model to similarity data.

The decision rule that we use for similarity judgments is the natural one. It is assumed that a subject's dissimilarity judgment of a pair of stimuli S_j and S_k directly corresponds to the Euclidean distance between these two stimuli.

$$d_{jk} = [(\mathbf{X}_j - \mathbf{X}_k)'(\mathbf{X}_j - \mathbf{X}_k)]^{1/2}, \quad (1)$$

as perceived by the subject on each trial. Since the distance d_{jk} is a random variable, its value, and thus the subjects' dissimilarity judgment, can be expected to change from trial to trial.

The observed dissimilarity judgment d_{jk} needs to be distinguished from the actual Euclidean distance δ_{jk} , defined by

$$\delta_{jk} = [(\boldsymbol{\mu}_j - \boldsymbol{\mu}_k)'(\boldsymbol{\mu}_j - \boldsymbol{\mu}_k)]^{1/2}. \quad (2)$$

Unlike d_{jk} , the Euclidean distance δ_{jk} is not a random variable and therefore does not change from trial to trial. In the Hefner model, it is also important to distinguish δ_{jk} from $E(d_{jk})$, the expected value of d_{jk} . This latter distance, which can be conceptualized as the average of an infinite number of replications of the stimulus pair S_j and S_k , is not, in general, equal to δ_{jk} , nor is it, in most cases, even a good estimate of δ_{jk} . This important feature of the Hefner model is discussed more fully in the second section.

In the simplest case, all stimulus pairs from a given stimulus set $\{S_1, \dots, S_m\}$ are presented to the subjects, and, thus, under cases 3i or 5i, the likelihood function to be maximized equals

$$L = \prod_{k=j+1}^m \prod_{j=1}^m f(d_{jk}), \quad (3)$$

where $f(d_{jk})$ is the probability density function (pdf) of the random variable d_{jk} . The probability density function $f(d_{jk})$ is closely related to the noncentral chi-square distribution, which we define next.

Definition 2.2. Let z_1, \dots, z_ν be independent standard normal variables, and let $\delta_1, \dots, \delta_\nu$ be constants. Then the sum

$$\sum_{j=1}^{\nu} (z_j + \delta_j)^2$$

has the noncentral chi-square distribution $\chi'^2(\nu, \lambda)$ with ν degrees of freedom and noncentrality parameter λ equal to

$$\lambda = \sum_{j=1}^{\nu} \delta_j^2.$$

For future reference, we note that the mean of the noncentral chi-square distribution equals

$$E(\chi'^2) = \nu + \lambda, \quad (4)$$

and the variance equals

$$\text{Var}(\chi'^2) = 2(\nu + 2\lambda). \quad (5)$$

It can be readily seen, from the definition of the noncentral chi-square distribution and from the preceding expressions for the mean and variance, that the noncentral chi-square distribution reduces to the central or standard chi-square distribution when $\lambda = 0$. However, unlike the central chi-square distribution, the pdf of the noncentral chi-square distribution cannot be expressed in closed form. We say more on this later.

Returning to our present concern, we wish to determine the pdf of the distance d_{jk} between stimuli S_j and S_k under the case 3i assumptions. To do this, we first define the standardized variable

$$z_{jkh} = \frac{d_{jkh} - \delta_{jkh}}{\sigma_{jk}}, \quad (6)$$

where

$$\begin{aligned} d_{jkh} &= X_{jh} - X_{kh}, \\ \delta_{jkh} &= \mu_{jh} - \mu_{kh}, \end{aligned} \quad (7)$$

and

$$\sigma_{jk}^2 = \sigma_j^2 + \sigma_k^2. \quad (8)$$

Therefore under the case 3i assumptions, z_{jkh} has the standard normal distribution $N(0,1)$. Thus, for the squared distance

$$d_{jk}^2 = \sum_{h=1}^r (X_{jh} - X_{kh})^2,$$

we can write

$$d_{jk}^2 = \sum_{h=1}^r d_{jkh}^2,$$

which, from Equation 6, gives

$$d_{jk}^2 = \sigma_{jk}^2 \sum_{k=1}^r \left(z_{jkh} + \frac{\delta_{jkh}}{\sigma_{jk}} \right)^2. \quad (9)$$

Equation 9 indicates that d_{jk}^2/σ_{jk}^2 has the noncentral chi-square distribution $\chi'^2(\nu, \lambda_{jk})$, where $\nu = r$ and

$$\lambda_{jk} = \sum_{h=1}^r \frac{\delta_{jkh}^2}{\sigma_{jk}^2} = \frac{\delta_{jk}^2}{\sigma_{jk}^2}. \quad (10)$$

For case 5i, the results can be simplified further, for then

$$\sigma_{jk}^2 = \sigma_j^2 + \sigma_k^2, \quad (11)$$

and since σ^2 can be set equal to an arbitrary value, it is customary to set it equal to $\frac{1}{2}$, giving

$$\sigma_{jk}^2 = \frac{1}{2} + \frac{1}{2} = 1. \quad (12)$$

Thus, under case 5i, d_{jk}^2 has a noncentral chi-square distribution with non-centrality parameter $\lambda_{jk} = \delta_{jk}^2$.

These results can be used to obtain the pdf of d_{jk} under cases 3i and 5i. Starting with the cumulative distribution function $F(d_{jk})$, it follows that

$$F(d_{jk}) = G\left(\frac{d_{jk}^2}{\sigma_{jk}^2}\right), \quad (13)$$

where $G(\cdot)$ is the noncentral chi-square cumulative distribution function (cdf). Differentiating Equation 13 by d_{jk} and using $g(\cdot)$ for the noncentral chi-square pdf gives

$$f(d_{jk}) = g\left(\frac{d_{jk}^2}{\sigma_{jk}^2}\right) \left(\frac{2d_{jk}}{\sigma_{jk}^2}\right), \quad (14)$$

which shows that the pdf of d_{jk} can indeed be expressed in terms of the pdf of the noncentral chi-square distribution function. Thus, to evaluate $f(d_{jk})$, it is sufficient to concentrate our attention on the function $g(\cdot)$.

As indicated earlier, the exact form of $g(\cdot)$ can be expressed only as an infinite series, whose rate of convergence depends on the values of ν and λ . When ν is large and λ is small, the series will tend to converge rapidly. However, for other values of ν and λ , the rate of convergence can be painfully slow. Fortunately, normal and central chi-square approximations have been worked out (Abdel-Aty, 1954; Patnaik, 1949; Sankaran, 1959), which, considering their simplicity, are amazingly accurate. In some applications (Zinnes & Griggs, 1974; Zinnes & Wolff, 1977), the normal distribution approximation was found to be superior, and in other applications (MacKay & Zinnes, 1986; Zinnes & MacKay, 1987) the central chi-square approximation was more useful. In still other cases (MacKay & Zinnes, 1981; Zinnes & MacKay, 1983), it turned out to be useful to piece together the normal approximation and the series expansion of the noncentral chi-square distribution. This was done by using the series expansion in the region where it converges rapidly—when $\delta_{jk}d_{jk}/\sigma_{jk}^2 < 2.55$ —and using the normal distribution approximation for the remaining regions, where it tends to be quite accurate.

Preferences

This section covers two types of preference data: binary choice data and preference ratio data. In both cases, the subject indicates which of a pair of stimuli he

or she prefers. For preference ratio judgments, the subject indicates, in addition, how much he or she prefers the more preferred stimulus over the less preferred stimulus. The reason for using the preference ratio judgment is to extract more information from the subject on each trial, thus minimizing the number of replications necessary to estimate the parameters at some desired level of accuracy.

As indicated earlier, preferences can be accommodated within the Hefner model by adding to it the assumptions of the Coombs unfolding model. The preferences of subjects are then assumed to be determined by the distance between the ideal points and the stimulus points. The smaller the distance d_{ij} between ideal point i and stimulus point j , the more desirable stimulus S_j is to person P_i . The probability p_{ijk} that person P_i chooses S_j over S_k is then equal to

$$p_{ijk} = P(d_{ij} \leq d_{ik}) , \quad (15)$$

which can also be written as

$$p_{ijk} = P\left(\frac{d_{ij}^2}{d_{ik}^2} \leq 1\right) \quad (16)$$

or as

$$p_{ijk} = P\left(\frac{d_{ij}^2/\sigma_{ij}^2}{d_{ik}^2/\sigma_{ik}^2} \leq \frac{\sigma_{ik}^2}{\sigma_{ij}^2}\right) . \quad (17)$$

For the case 3i model, both d_{ij}^2/σ_{ij}^2 and d_{ik}^2/σ_{ik}^2 have a noncentral chi-square distribution, and therefore the fraction on the left side of the inequality in Equation 17 consists of the ratio of two random variables having this distribution. This leads us to a consideration of the doubly noncentral F distribution (Johnson & Kotz, 1970, vol. 2).

Definition 2.3. Let $\chi_1'^2(\nu_1, \lambda_1)$ and $\chi_2'^2(\nu_2, \lambda_2)$ be two independent noncentral chi-square random variables. Then the ratio

$$\frac{\chi_1'^2/\nu_1}{\chi_2'^2/\nu_2}$$

has the doubly noncentral F distribution $F''(\nu_1, \nu_2, \lambda_1, \lambda_2)$ with ν_1 and ν_2 degrees of freedom and noncentrality parameters λ_1 and λ_2 . If $\lambda_2 = 0$, the ratio is said to have an ordinary noncentral F distribution.

Returning to Equation 17, we see that the left side of the inequality, under certain independence conditions, is distributed as a doubly noncentral F distribution $F''(r, r, \lambda_{ij}, \lambda_{ik})$, where the noncentrality parameters λ_{ij} and λ_{ik} are as defined in Equation 10. The independence condition refers to the independence of the numerator and denominator of this ratio. This could be a problem in the present case, since d_{ij} and d_{ik} both involve the same ideal point i . However, we can simplify matters for the present by assuming that person P_i selects two indepen-

dent samples from his or her ideal distribution, one of which is used to determine the distance d_{ij} and the other the distance d_{ik} . Under these conditions, d_{ij} and d_{ik} are independent random variables. Whether this assumption is plausible or reasonable in any given situation could very well depend on the specific details of the experimental procedure.

Making use of these results, we can write Equation 17 more simply as

$$p_{ijk} = P \left[F''(r, r, \lambda_{ij}, \lambda_{ik}) \leq \frac{\sigma_{ik}^2}{\sigma_{ij}^2} \right], \quad (18)$$

indicating that the values of the binary choice probabilities p_{ijk} can be determined by evaluating the cdf of the doubly noncentral F distribution.

The analysis of preference ratio judgments is similar to that of binary choices. It is assumed that person P_i reports the preference ratio R_{ijk} , where

$$R_{ijk} = \frac{d_{ik}}{d_{ij}} \quad (19)$$

when the stimulus pair S_j and S_k is presented. To arrive at the pdf $f_R(R_{ijk})$, we start with the cumulative distribution function

$$F_R(R_{ijk}) = H \left(R_{ijk}^2 \frac{\sigma_{ij}^2}{\sigma_{ik}^2} \right), \quad (20)$$

where $H(\cdot)$ is the cdf of $F''(r, r, \lambda_{ij}, \lambda_{ik})$. Differentiating Equation 20 with respect to R_{ijk} gives

$$f_R(R_{ijk}) = 2R_{ijk} \left(\frac{\sigma_{ij}^2}{\sigma_{ik}^2} \right) h \left(\frac{d_{ik}^2 / \sigma_{ik}^2}{d_{ij}^2 / \sigma_{ij}^2} \right), \quad (21)$$

where $h(\cdot)$ is just the pdf of the doubly noncentral F distribution $F''(r, r, \lambda_{ij}, \lambda_{ik})$. This shows that preference ratio data, as well as binary choice data, require working with the doubly noncentral F distribution.

The Density Function of F''

The doubly noncentral F distribution $F''(\nu_1, \nu_2, \lambda_1, \lambda_2)$ has been studied extensively (Bulgren, 1971; Chou & Arthur, 1985; Price, 1964; Tang, 1938; Tiku, 1965). As one might expect from a consideration of the noncentral chi-square pdf, the pdf of the doubly noncentral F'' distribution cannot be expressed in closed form and, in fact, requires a doubly infinite series of terms, each term of which includes a number of complex functions. Obviously, to make practical use of this density function, one must use an approximation.

Two approaches have been used to approximate the pdf of the F'' distribution. Both involve using one of the approximations of the noncentral chi-square distribution discussed previously. In one application (Zinnes & Griggs, 1974), the

Abdel-Aty (1954) approximation was used to convert the noncentral chi-square distribution to a normal distribution, which results in transforming the F'' ratio to a difference of two normally distributed random variables. In another application (Zinnes & MacKay, 1987), the Patnaik (1949) approximation was used to convert a noncentral chi-square distribution into a central chi-square distribution. This results in the doubly noncentral F distribution being approximated by a central F distribution.

Although these approximations of the F'' distribution would seem to be overly simplistic, nevertheless, they are reasonably accurate. Figure 2.2, for example, shows the level of accuracy that is obtained when the central F approximation is used for F'' . Three different distributions are plotted in this figure, one for each of three different values of λ . The exact distributions (the solid lines) shown in this figure were determined by summing the series expansion for F'' . From the discrepancy between the exact and the approximate curves (the dashed lines), it is evident that the largest absolute error occurs in the middle of the distribution, especially when the distribution is highly skewed. But, in general, it can be seen that the exact and approximate lines follow each other quite closely, even in the tails of the distribution.

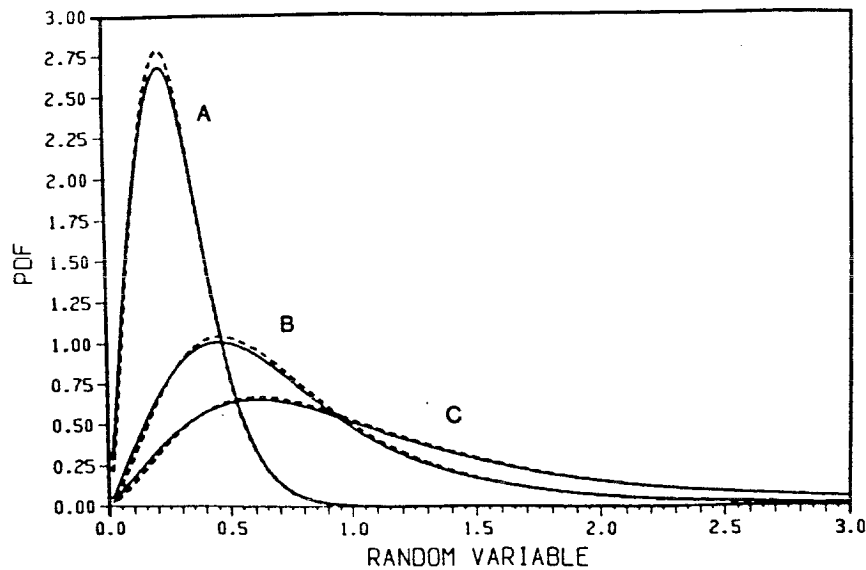


FIG. 2.2. The central F approximation of the double noncentral F distribution. Both degrees of freedom are equal to 2. The solid lines are the exact values, the dashed lines the central F approximation. For curve A, $\lambda_1 = 1, \lambda_2 = 30$; for curve B, $\lambda_1 = 1, \lambda_2 = 4$; for curve C, $\lambda_1 = 1, \lambda_2 = 1$.

PROPERTIES OF THE MODEL

The Hefner model has many interesting properties, even some counterintuitive properties, which distinguish it from other multidimensional models. In this section we cover perhaps its most controversial property, the "nonmonotonicity property."

The Nonmonotonicity Property

Most, if not all, of the multidimensional scaling models assume, either implicitly or explicitly, that the observed similarity judgment d_{jk} , or, perhaps, the average dissimilarity judgment $E(d_{jk})$, is monotonically related to the true distance δ_{jk} between stimuli S_j and S_k . For the Hefner model, even under case 3i, this monotonicity property does not, in general, hold. The relationship between $E(d_{jk})$ and δ_{jk} need not be monotonic. To show this, we consider first the condition under which monotonicity can be expected to hold. This will occur when the noncentrality parameters λ_{jk} are large. Since $\lambda_{jk} = \delta_{jk}^2 / \sigma_{jk}^2$, λ_{jk} will be large when δ_{jk}^2 is appreciably larger than σ_{jk}^2 . Under this condition, it has been shown (Patnaik, 1949) that $E(d_{jk})$ can be approximated by

$$E(d_{jk}) \cong \sigma_{jk} \cdot \left(\frac{2a - (1 + b)}{2} \right)^{1/2}, \quad (22)$$

where $a = \nu + \lambda_{jk}$ and $b = \lambda_{jk}/a$. As λ_{jk} becomes indefinitely large, Equation 22 indicates that then

$$E(d_{jk}) \cong \sigma_{jk} (\lambda_{jk}^{1/2}). \quad (23)$$

Substituting for λ_{jk} in Equation 23 and simplifying gives

$$E(d_{jk}) = \delta_{jk}. \quad (24)$$

In other words, when σ_{jk} is negligible relative to the size of the true distance δ_{jk} , the expected distance $E(d_{jk})$ will be indistinguishable from the true distance δ_{jk} , and therefore, the Hefner model, for this case, will necessarily satisfy the monotonicity condition.

However, consider next what happens when λ_{jk} is negligibly small. This occurs when the variance σ_{jk}^2 is substantially larger than the true distance δ_{jk}^2 . From Equation 4 and from the fact that d_{jk}^2 / σ_{jk}^2 has, under case 3i, a noncentral chi-square distribution, we have

$$E(d_{jk}^2) = \sigma_{jk}^2 (\nu + \lambda_{jk}),$$

or

$$E(d_{jk}^2) = \nu \sigma_{jk}^2 + \delta_{jk}^2. \quad (25)$$

Therefore, when σ_{jk}^2 dominates δ_{jk}^2 ,

$$E(d_{jk}^2) \cong v\sigma_{jk}^2. \quad (26)$$

Equation 26 shows that when the variance σ_{jk}^2 becomes appreciably larger than the distance δ_{jk} , the expected value of d_{jk}^2 will be determined almost entirely by the value of σ_{jk}^2 and will tend to be almost completely insensitive to the actual distance δ_{jk} . It is clear that, under these conditions, the monotonicity condition cannot be expected to hold. In the extreme case, $E(d_{jk})$ could, in fact, be indefinitely large, even when the true distance δ_{jk} is precisely equal to zero.

This property of the Hefner case 3i model may seem counterintuitive, but it actually is quite plausible. If two independent random samples are drawn from the same distribution, having a large variance, we would expect the distance or difference between these two random samples to be large most of the time. This nonmonotonicity principle also makes sense psychologically. If a person is highly uncertain about the perceptual aspects of a set of stimuli or his or her ideal stimulus, it is not unreasonable to expect that person's dissimilarity or preference judgment of the stimuli to be large and highly variable, even when, on the average, these stimuli are perceived as having identical positions on each of the relevant attributes.

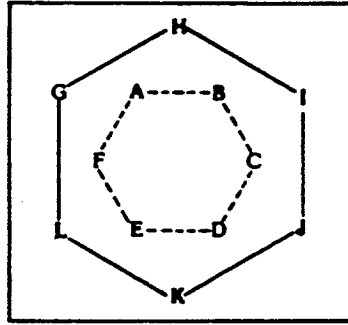
Does the lack of monotonicity in the Hefner model have practical consequences? To answer this question, we consider next two simulations and the results of an empirical study. In both simulations a typical nonmetric Multidimensional Scaling algorithm (KYST, Kruskal, Young, & Seery, 1973) was applied to simulated data generated by a "subject," who follows the Hefner case 3i model. The monotonicity property, it should be noted, is an intrinsic aspect of the nonmetric approaches.

Simulation I: Hexagons

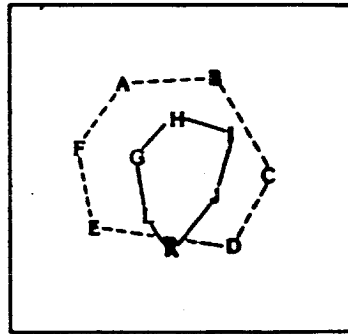
In this simulation, the data set consists of 30 replications of all the interpoint distances of a two-dimensional configuration containing 12 points. These 12 points were arranged in two hexagons, one inside the other. (See Figure 2.3a.) Two values of σ_{jk} were used: 2.5 and 1.0. The larger value was assigned to each of the six points forming the inner hexagon, the smaller value to each of the six points forming the outer hexagon. The outer hexagon was about 2 units wide.

To summarize, the data sets were generated by obtaining random samples from a Hefner case 3i subject. These data were then analyzed by the nonmetric approach and by using the maximum likelihood approach described previously.

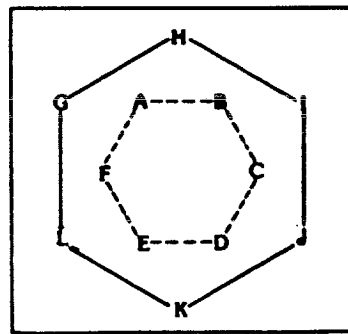
The recovered configurations obtained from the nonmetric and maximum likelihood approaches are shown in Figure 2.3b,c, respectively. It can be seen that the nonmetric solution interchanged the position of the two hexagons. The



(a)



(b)



(c)

FIG. 2.3. Simulation I: (a) original configuration, (b) solution obtained using a nonmetric procedure, and (c) solution obtained using Hefner model and Maximum Likelihood Estimation.

hexagon that initially was on the outside is now contained almost entirely inside the hexagon that initially was on the inside.

This result is as expected, because as shown in the previous section, the expected value of the interpoint distances will tend to reflect, almost entirely, the uncertainty values of the stimuli when these values are large relative to the true distance. Therefore, the inner hexagon, having large variances, will tend to be "perceived" by a nonmetric approach as a large hexagon, having large interpoint distances.

The solution obtained from the maximum likelihood approach, shown in Figure 2.3c is quite accurate. There are no detectable differences between the true and recovered configurations.

Simulation II: Expected Values

In this example, unlike the previous one, the simulated data consisted of the expected values of the interpoint distances (i.e., not the true interpoint distances) rather than a finite number of independently sampled replications. This was done to simplify issues somewhat. We wish to show in this example what happens to the nonmetric solution when the uncertainty values become quite large and when the nonmetric results cannot possibly be attributed to perturbations resulting from samples having a small size.

The configuration used in this example was generated by randomly selecting 10 points along the bell-shaped curve of the normal distribution. The configuration obtained is shown in Figure 2.4a. The values of the uncertainty parameter that were assigned to each of the 10 points varied between zero and an upper bound σ_B . Seven values of σ_B , ranging from 0.15 to 9.6, were used. This effectively resulted in seven simulations, all having the same initial stimulus configuration but differing in the range of variability assigned to each point in the configuration. The relative sizes of these uncertainty values can be appreciated by comparing them to the magnitude of the configuration, which, in each of the seven simulations, was standardized to have a variance of 1 on each of the two axes.

Figure 2.4 shows the configuration recovered from each of the seven simulations by using, in each case, a nonmetric approach. Panel (a) gives the true configuration. The points in each of the panels are labeled 1 through 9 followed by A. This labeling reflects the order of magnitude of the uncertainty parameter assigned to each point. Point 1 has the lowest value, point A the highest.

From the seven nonmetric solutions shown in Figure 2.4, it is evident that the nonmetric solution degenerates considerably as the degree of variability of the points increases. When the level of σ_B is quite high, the recovered configuration bears no resemblance to the true configuration. What actually happens is that the higher numbered points, those assigned higher uncertainty values, move to the

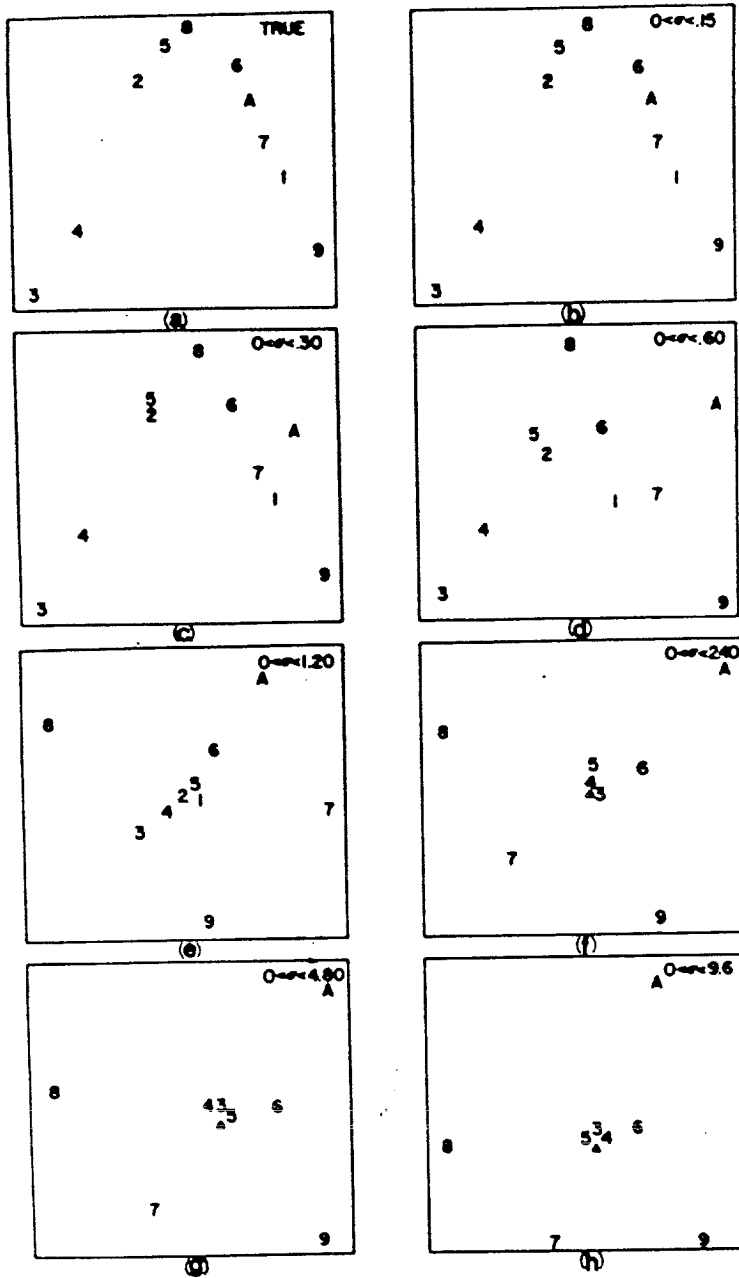


FIG. 2.4. Simulation II. The higher-numbered points have higher variances assigned to them. Point A has the highest variance. The delta symbol indicates a multiple point, consisting of points 1 and 2.

outside of the configuration as σ_B increases, and the remaining points gradually move toward the center.

These results show in a more extreme form what was evident in the previous example. When the uncertainty value of a point is large, its location in the configuration recovered by the nonmetric approach is determined almost entirely by the magnitude of its uncertainty value, not by its actual position in the true configuration. In the present example, this property generates what we call the "black hole" effect. Points having a small variance, relative to the other points, are gradually pulled into the "hole," even if they are initially quite some distance from it. Only those points with large variance succeed in not being drawn into the hole. They drift into outer space, so to speak.

Residential Locations

As a final example of the nonmonotonic property of the Hefner model, we consider an empirical study. In this experiment (MacKay and Zinnes, 1986), 20 subjects made preference judgments of 12 different apartment locations. These locations differed with respect to two variables: time (T) or distance from the central business district (CBD), where it was assumed the subjects worked, and the environmental quality (E) of the locations, as expressed by the population density of the location and by the availability of local services. Four levels of T (5, 10, 15, and 20 minutes) and three levels of E (3,000, 4,000, and 5,000 persons per square mile) were used, which, when combined factorially, resulted in 12 locations.

The specific judgment task used was a "preference ratio" judgment. Recall from the previous section that this task requires subjects to indicate not only the stimulus in each pair they prefer but also the degree of preference. A response of 2, for example, would indicate that the subject preferred one stimulus twice as much as the other. After completing the preference ratio judgments, subjects were asked to indicate what their ideal T and E values were.

The nonmonotonic solution is shown in Figure 2.5, and the maximum likelihood solution, based on the Hefner case 3i model, is shown in Figure 2.6. The 12 stimuli in these figures are labeled 0, 1 through 9, followed by A and B. The ideal points, one for each of the 20 subjects, are labeled C through V.

From Figure 2.5, it can be seen that the nonmetric solution exhibits no evidence of the factorial structure of the stimuli, which, in contrast, is clearly evident in the ML solution. It is also interesting to observe that the ideal points of the nonmetric solution are, for the most part, located far from the cluster of stimulus points. This is not consistent with the subjects' self-reported ideal T and E values. Most of the ideal values selected by the subjects were interior points, falling within the stimulus configuration, rather than at the edges of the configuration. This aspect of the data is well captured by the ML solution.

It is also interesting to compare the dimensionality test of the nonmetric and

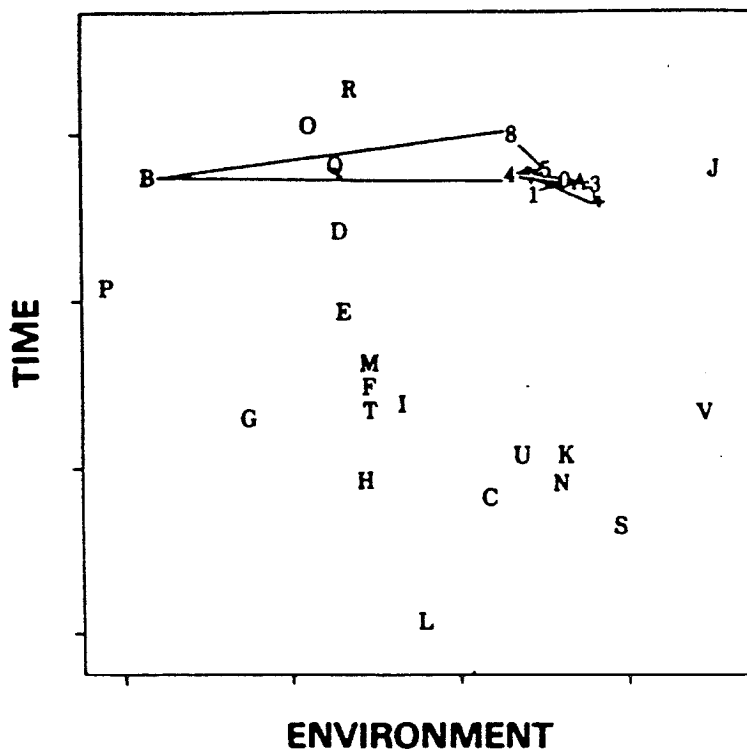


FIG. 2.5. Nonmetric solution of 12 residences and 20 ideal points. Lines connect adjacent residences. The unconnected points are the ideal points. The asterisk identifies points 2 and 7, the plus sign points 6 and 9.

the ML solutions. The likelihood ratio test for one- and two-dimensional solutions is statistically significant, while the test for two- and three-dimensional solutions is not, indicating that a two-dimensional solution is the most appropriate solution. Since there are no statistical tests for the nonmetric approach, it is customary to look at how the stress values decrease as the dimensionality of the solution increases. Stress is a measure of the degree to which the obtained solution does not account for the observed data. The stress values of the nonmetric solution decrease gradually from one to three dimensions and then less sharply from three to four dimensions. This "elbow" at three dimensions is usually taken as an indication of a three-dimensional solution. However, given the two-dimensional factorial structure of the stimuli, it is difficult not to conclude that the two-dimensional solution is, indeed, the more appropriate solution. There seems to be a tendency for nonmetric solutions to overestimate dimensionality.

where the standardized variable z_{jkh} is now equal to

$$z_{jkh} = \frac{(x_{jh} - x_{kh}) - \delta_{jkh}}{\sigma_{jkh}} \quad (28)$$

and

$$\sigma_{jkh}^2 = \sigma_{jh}^2 + \sigma_{kh}^2 . \quad (29)$$

Letting

$$\xi_{jkh} = \frac{\delta_{jkh}}{\sigma_{jkh}} \quad (30)$$

and defining the vectors

$$\mathbf{z}_{jk} = (z_{jkh}) , \quad \boldsymbol{\xi}_{jk} = (\xi_{jkh}) , \quad (31)$$

we can write Equation 27 more compactly as

$$d_{jk}^2 = (\mathbf{z}_{jk} + \boldsymbol{\xi}_{jk})' \boldsymbol{\Sigma}_{jk} (\mathbf{z}_{jk} + \boldsymbol{\xi}_{jk}) , \quad (32)$$

where $\boldsymbol{\Sigma}_{jk}$ is the diagonal matrix $\mathbf{D}(\sigma_{jkh}^2)$. Equation 32 indicates that the squared distance d_{jk}^2 , in an anisotropic space, can be expressed as a quadratic form containing normally distributed variables.

The research on quadratic forms has been reviewed by a number of writers (Johnson & Kotz, 1970; MacKay, 1989; Ruben, 1962). In the following sections, we avoid the algorithmic details of the anisotropic cases and content ourselves with describing some of the interesting properties and empirical results of cases 3a and 5a, especially how these cases differ from the isotropic cases.

Variance Properties of the Interpoint Distances

In an isotropic space, the variance of the interpoint distances exhibits what might be called a "quasi-Weber" effect. As the distance δ_{jk} between stimuli S_j and S_k increases, the variance of d_{jk} increases monotonically to an asymptotic value. In fact, the asymptote will be the sum of the variances σ_j^2 and σ_k^2 associated with S_j and S_k .

In contrast, the variance of the distance d_{jk} in an anisotropic space exhibit properties that depend on the specific orientation of the contours of equal likelihood. Consider the two-dimensional example shown in Figure 2.7a. If points i and j move apart from each other in the horizontal direction, the quasi-Weber effect is observed. This is shown by the solid line in Figure 2.8.

However, if the points in the anisotropic space, shown in Figure 2.7b, move apart in this same horizontal direction, their variance first increases slightly, but then decreases and gradually approaches an asymptote at a much lower level than previously. This is shown by the dashed lines in Figure 2.8.

This radically different result between the isotropic and anisotropic cases

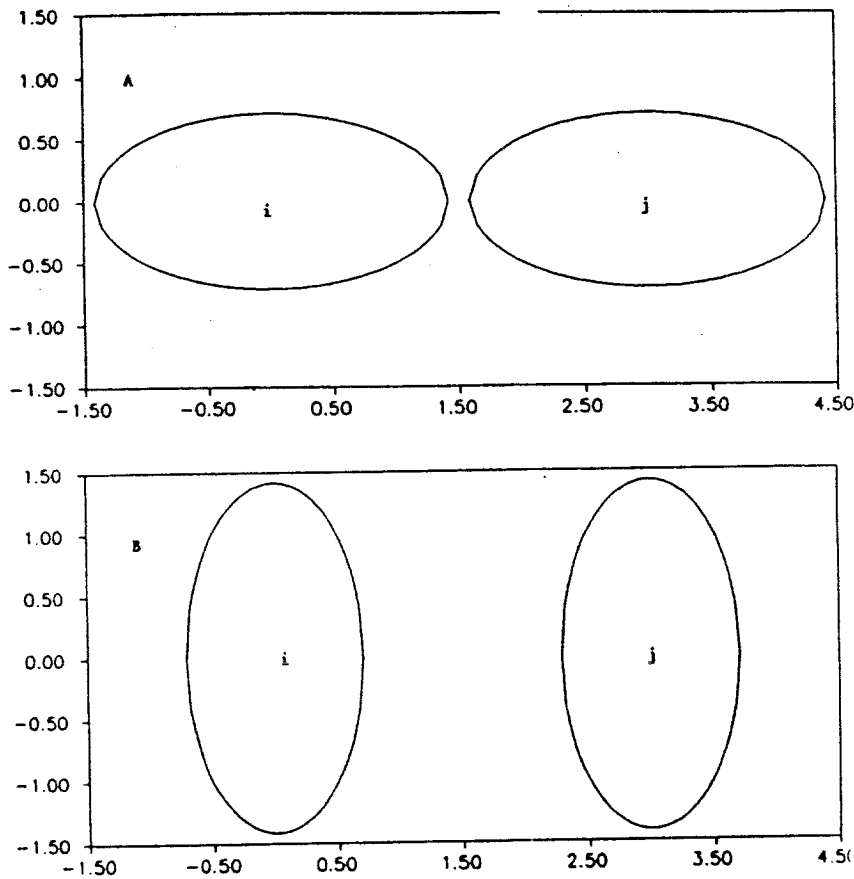


FIG. 2.7. An example of two points in an anisotropic space: (a) major axes of the equiprobability ellipses is on the line joining i and j ; (b) major axes are perpendicular to a line joining i and j .

explains why it frequently happens that both cases will result in approximately the same stimulus configuration, but nevertheless have drastically different expected distances and choice probabilities. In other words, similar configurations in anisotropic spaces do not imply similar observable properties.

Properties of Choice Probabilities

Choice probabilities in an anisotropic space also exhibit nonmonotonic properties. Consider three stimulus points j , k , and m , located on the vertices of an equilateral triangle in a two-dimensional space, and an ideal point located at the centroid of this triangle (see Figure 2.9). Since ideal point i is equally far from all

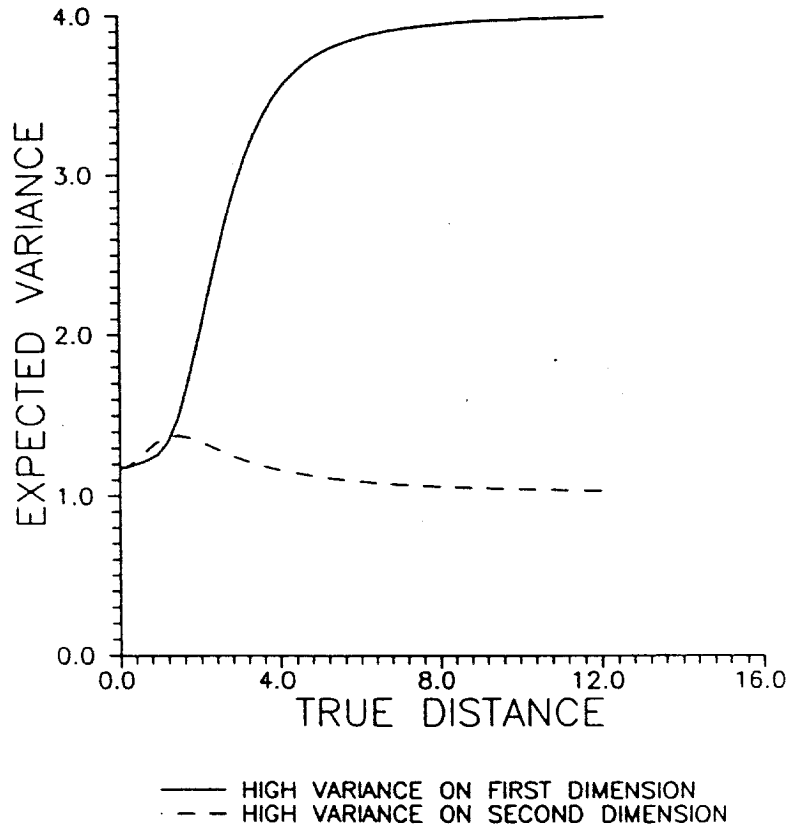


FIG. 2.8. The influence of the true distance and the variance structure on the expected variance of the distance between two points. Both the solid and dashed lines show the expected variance when the two points move apart from each other along the horizontal axis, but in the first case the points have the variance structure shown in Figure 2.7a, and in the second case the variance structure in Figure 2.7b.

three stimulus points, the choice probabilities p_j , p_k , and p_m will depend entirely on the variances associated with the three stimuli. If all three points have the same variances, and if the space is isotropic, then all three probabilities will be equal to $\frac{1}{3}$. Person P_i will be indifferent to the three stimuli. However, if the space is anisotropic, the results are quite different, even when all three points have the same variance structure. Consider the variance structure in Figure 2.9. In this example, all four points have precisely the same variance structure, which involves a large variance in the vertical direction and a smaller variance in the horizontal direction. In fact, in this example, the standard deviation on the vertical axis is 1.0 and on the horizontal axis is 0.05. Under these conditions,

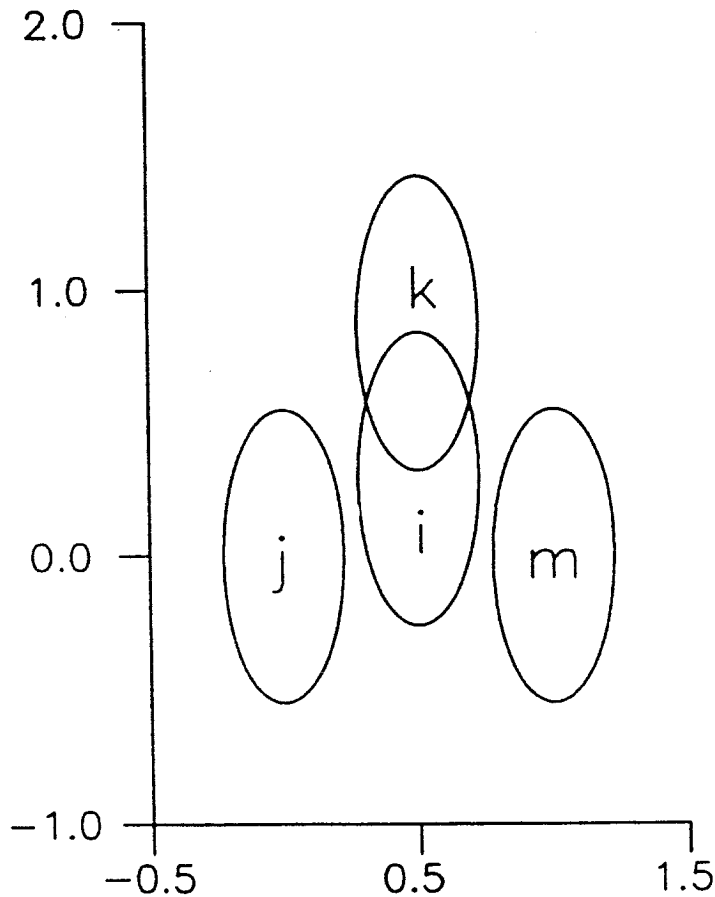


FIG. 2.9. An example of three points located at the vertices of an equilateral triangle and an ideal point located at the centroid of the triangle.

which are that of a case 5a Hefner model, the choice probability p_k is larger than both p_j and p_m , indicating that person P_i prefers S_k over the other two stimuli.

Consider what happens next when the variances in the vertical direction of all four points increases. One might expect the choice probability p_k to keep increasing and, therefore, person P_i to show increasing preference for stimulus S_k over the other two stimuli. This does not happen. As shown in Figure 2.10, p_k increases at first, but then decreases and approaches an asymptotic value of about 0.36. Thus, the choice probability p_k is not a simple monotonic function of the size of the variance in the vertical direction.

This example and the one in the previous section both illustrate some of the

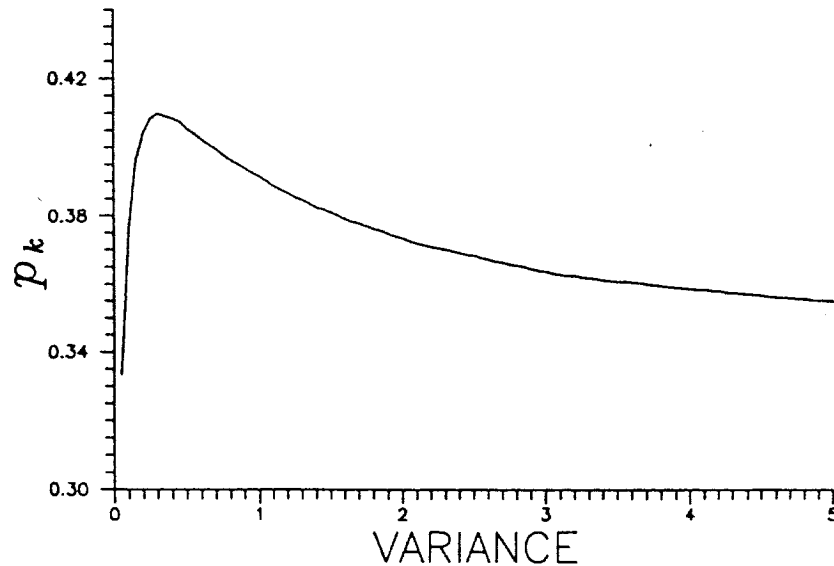


FIG. 2.10. Percent of time the ideal point i in Figure 2.9 is closest to point k , as the variance on the vertical axis increases from 0.05 to 5.0 while the variance on the horizontal axis stays fixed at 0.05.

nonintuitive properties of anisotropic spaces, which make those spaces so fascinating. To demonstrate the potential usefulness of anisotropic solutions, we consider next a simple empirical example.

An Empirical Example of an Anisotropic Space

In this experiment (MacKay & Dröge, 1990), 52 subjects judged the similarity of eight brands of toothpaste presented pairwise.

On the basis of past advertising claims, it was expected that the eight brands could be divided into three groups: (1) those brands stressing decay prevention (brands A and B), (2) those stressing breath freshening (brands C, D, E, and F), and (3) those stressing whitening (brands G and H). With this division, Likelihood Ratio Tests were carried out, in which a case 5i and a constrained case 3i solution containing three variances, one for each of the three groups, were compared. Likelihood ratio tests were also conducted for the constrained and unconstrained case 3i solution, where all stimuli have unique variances. The former test was significant but the latter was not, suggesting that the constrained case 3i solution adequately accounts for the data obtained. Similarly, results of Likelihood Ratio Tests for one, two, and three dimensions suggested that a two-dimensional solution was most appropriate.

Figure 2.11a shows the coordinates and the standard deviations of the con-

CASE III MODELS

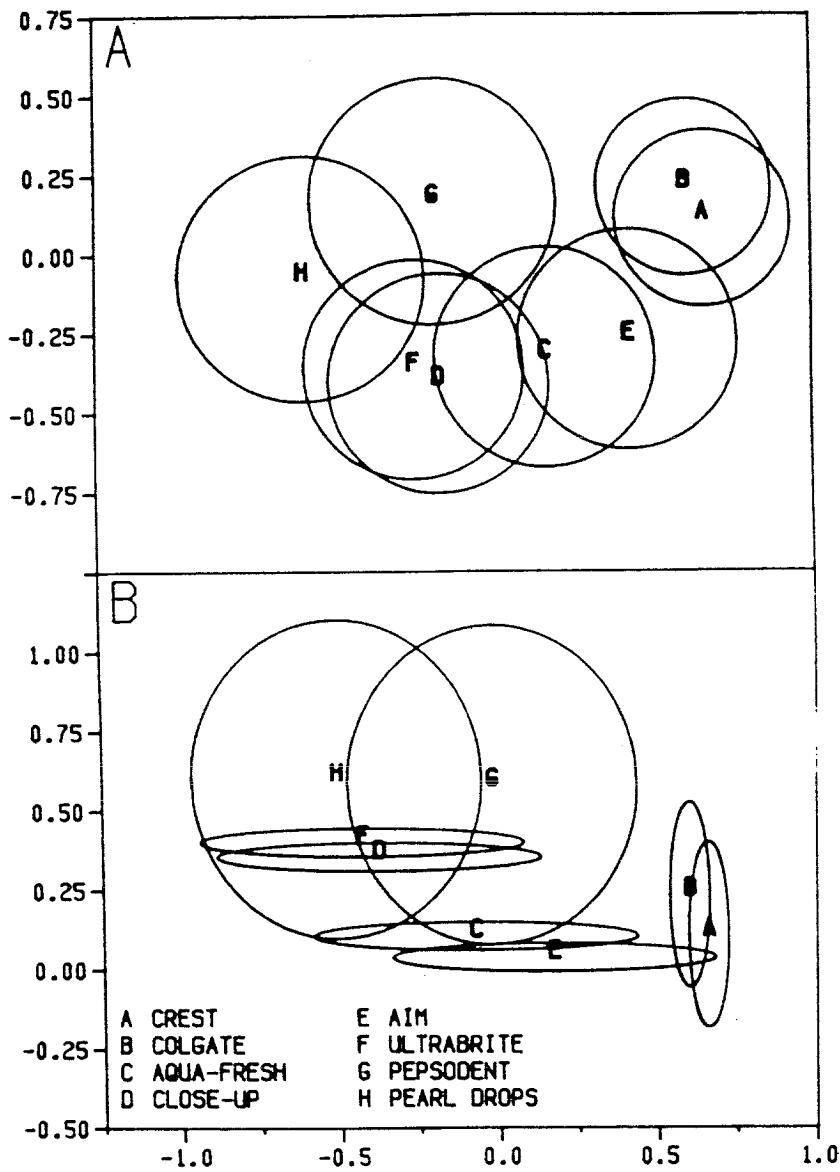


FIG. 2.11. Solution obtained for the eight toothpastes: (a) constrained case 3i solution, (b) case 3a solution.

strained case 3i solution. Generally, the results are as expected. The horizontal dimension appears to reflect decay prevention, the vertical dimension whitening and breath freshening. Also the relative sizes of variances seem plausible. Brands A and B have the smallest variance, and G and H have the largest, no doubt because brands A and B advertise heavily their decay prevention properties to college students, the subjects in this experiment. In contrast, brands G and H are positioned as a smoker's toothpaste for older consumers. Apparently, students are more concerned about decay prevention than they are about whitening and breath freshening.

However, more insight into these data can be obtained by obtaining the anisotropic solution, which is shown in Figure 2.11b. Figure 2.11a,b indicates that the isotropic and anisotropic configurations are similar, but that their variances are quite different. In Figure 2.11b, brands A and B, have small variances on dimension 1 (the decay prevention dimension) and large variances on dimension 2 (the whitening-breath freshening dimension). Just the opposite is the case for brands C, D, E, and H. These brands, which emphasize breath freshening, have a small variance on dimension 2 and a large variance on dimension 1. Brands G and H, which emphasize whitening, have nearly equal variances on both axes. Evidently, brands G and H are not well known or understood by the subjects.

The usefulness of the case 3a solution is also supported by a Likelihood Ratio Test. The comparison between a constrained case 3a and a constrained case 3i solution was statistically significant, thus providing justification for the extra variance parameters used in the case 3a solution.

These results, taken as a whole, seem to provide strong support for the appropriateness and the usefulness of the anisotropic solution. As we become more adept at working with these spaces and understanding them, it should be possible to make stronger connections with the theoretical developments in other psychological domains.