Japanese Psychological Research
1980, Vol. 22, No. 4, 188–196

MAXIMUM LIKELIHOOD ESTIMATION IN THE GENERALIZED CASE OF THURSTONE'S MODEL OF COMPARATIVE JUDGMENT[1]

YOSHIO TAKANE[2]

McGill University

In light of the empirical evidence against simple scalability a generalized case of Thurstone's model of comparative judgment was proposed, in which covariance terms in discriminal processes are not necessarily assumed zero or constant. In order to avoid the indeterminacy (over-parametrization) problem of the model, the covariance matrix was assumed to have a prescribed rank. A parameter estimation procedure based on the maximum likelihood principle was developed and implemented in the form of a FORTRAN program. An example was given to illustrate the empirical relevance of the proposal.

A bulk of recent literature on human choice behavior concerned the role of similarity among choice alternatives in judgmental processes. Similar stimuli are more comparable than dissimilar stimuli, so that choice probabilities involving similar pairs tend to be more extreme than those involving dissimilar pairs.

Two of the most representative models of choice behavior, Case V of Thurstone's model of comparative judgment (Thurstone, 1927) and Luce's choice model (Luce, 1959), rest on the premise that all aspects of a stimulus pertinent to choice probabilities are representable by a single number which is valid no matter with which stimuli that particular stimulus is compared. This is the notion which Krantz (1967) called " simple scalability ". Many counter examples, however, have been noted against simple scalability (Debreu, 1960; Krantz, 1967; Tversky & Russo, 1969; Rumelhart & Greeno, 1971). Specifically it has been shown by these authors that similarity among stimuli plays an important role in choice processes. Similar stimuli (those having many features in common or those varying along a single dimension) are easier to compare, and consequently more discriminable than dissimilar stimuli (those having little in common or those differing simultaneously on more than one dimension). Thus, when stimuli with varying degrees of similarity are included in an experiment binary choice probabilities predicted from models of simple scalability are typically less extreme than the observed for similar pairs, while the reverse is true for dissimilar pairs.

Restle's (1961) choice model, which incorporates similarity among choice alternatives, has been formulated as a generalization to Luce's choice model to cope with the above situation. His model has been further generalized to non-binary situations by Tversky (1972a, b). Curiously, however, an analogous extension has not been realized in the conventional Thurstonian framework. Or to be more precise Thurstone's original formulation was sufficiently general, but the parameter estimation problem associated with it has been long neglected. In fact the model is too general in the sense that the effective number of parameters $[(n+4)(n-1)/2]$ exceeds the number of observed choice

[1] The research reported here was supported by grant A. 6394 to the author by Natural Sciences and Engineering Research Council of Canada.

[2] Requests for reprints should be addressed to Yoshio Takane, Department of Psychology, McGill University, 1205 Avenue Docteur Penfield, Montreal, Quebec, Canada, H3A 1B1.

frequencies $[n(n-1)/2]$. Although a general parameter estimation procedure has been developed for Thurstone's unrestricted case of the model (Arbuckle & Nugent, 1973), the problem of overparametrization has not been faced squarely.

In this paper we discuss one plausible way to constrain parameters in Thurstone's model along with a maximum likelihood estimation procedure for this proposed model.

Preliminary Analysis

Let $X_i \sim N(\mu_i, \sigma_i^2)$ denote the discriminal process corresponding to stimulus i. Thurstone's general model of comparative judgment postulates that

$$p_{ij} = \Pr(X_i > X_j) = \Pr(X_i - X_j > 0) = \mathrm{F}\left(\frac{\mu_{i-j}}{\sigma_{i-j}}\right) \quad (1)$$

where p_{ij} is the probability that stimulus i is chosen over stimulus j, F is the standard normal cumulative distribution function, and

$$\mu_{i-j} = \mu_i - \mu_j \quad (2)$$

$$\sigma_{i-j}^2 = \sigma_i^2 + \sigma_j^2 - 2\sigma_{ij} = \mathrm{V}(X_i - X_j) \quad (3)$$

with $\sigma_{ij} = \mathrm{Cov}(X_i, X_j)$ and $\sigma_i^2 = \sigma_{ii}$.

The fact that the comparatal dispersion, σ_{i-j}, is closely related to similarity between stimuli i and j has been pointed out by Sjöberg (1975), who obtained both direct rating judgments of the difference in discriminal processes, $X_i - X_j$, and similarity ratings between the two stimuli, and confirmed that the variance of $X_i - X_j$ closely agreed with average similarity rating between the two stimuli (Sjöberg, 1968; Sjöberg & Capozza, 1975). Indeed, if similarity—hence, covariance—increases, the comparatal dispersion should decrease. That σ_{i-j} has distance properties (i.e., $\sigma_{i-j} \geqslant 0$ and $\sigma_{i-j} = 0$ if and only if $i=j$, $\sigma_{i-j} = \sigma_{j-i}$, and $\sigma_{i-j} + \sigma_{j-k} \geqslant \sigma_{i-k}$) has been noted by Halff (1976).

Furthermore, let the squared distance between two random variables be defined by the expected squared difference between them. Then

$$\begin{aligned} d_{ij}^2 &= \mathrm{E}[(X_i - X_j)^2] \\ &= \mathrm{E}[(\mu_i - \mu_j) + (X_i - \mu_i) - (X_j - \mu_j)]^2 \\ &= (\mu_i - \mu_j)^2 + \mathrm{E}[(X_i - \mu_i) - (X_j - \mu_j)]^2 \\ &= (\mu_i - \mu_j)^2 + (\sigma_i^2 + \sigma_j^2 - 2\sigma_{ij}) \\ &= (\mu_i - \mu_j)^2 + \sigma_{i-j}^2 . \end{aligned} \quad (4)$$

That is, σ_{i-j}^2 corresponds to that portion of the squared distance between two random variables (as defined above) that cannot be accounted for by the difference in their expected values.

Thurstone's Case V is obtained by setting $\sigma_{i-j} = 1$ for all pairs of stimuli. Now if stimuli differ among themselves in their degree of similarity, this assumption cannot be justified. For example, when two stimuli are similar, σ_{i-j} is expected to be less than unity. Then setting $\sigma_{i-j} = 1$ would underestimate μ_{i-j}/σ_{i-j}, consequently p_{ij} is predicted to be less extreme than its observed counterpart. If, on the other hand, two stimuli are dissimilar, just the opposite will occur.

It is possible to indirectly estimate σ_{i-j} by measuring a discrepancy between observed and predicted p_{ij} after applying Case V. Let $\hat{p}_{ij}$ and $\tilde{p}_{ij}$ denote the observed and the predicted choice probabilities from Thurstone's Case V, respectively. Then

$$\tilde{p}_{ij} = \mathrm{F}(\hat{\mu}_i - \hat{\mu}_j),$$

or

$$\tilde{z}_{ij} = \mathrm{F}^{-1}(\tilde{p}_{ij}) = \hat{\mu}_i - \hat{\mu}_j , \quad (5)$$

where $\hat{\mu}_i$ is the estimated scale value of stimulus i. Suppose that Case V is substantially incorrect, and that $\hat{p}_{ij}$ has to be predicted by Thurstone's general model, namely

$$\hat{p}_{ij} = \mathrm{F}\left(\frac{\hat{\mu}_i - \hat{\mu}_j}{\sigma_{i-j}}\right),$$

or

$$\hat{z}_{ij} = \mathrm{F}^{-1}(\hat{p}_{ij}) = \frac{\hat{\mu}_i - \hat{\mu}_j}{\sigma_{i-j}} . \quad (6)$$

TABLE 1
Observed and predicted z_{ij} (above diagonal), and estimates of σ_{i-j} through Equation (7) (below diagonal)

	LJ	HW	CD	JU	CY	AF	BB	ET	SL	Estimated μ values from Thurstone's Case V
LJ		466 (290)	515 (442)	668 (698)	779 (989)	723 (679)	641 (808)	478 (420)	271 (216)	505
HW	62		227 (152)	527 (409)	628 (699)	478 (389)	431 (518)	054 (130)	054 (−174)	215
CD	86	67		305 (257)	443 (547)	227 (237)	249 (366)	054 (−022)	031 (−225)	063
JU	105	77	84		681 (290)	−021 (−020)	075 (109)	−338 (−279)	−641 (−482)	−194
CY	127	111	124	43		−443 (−310)	−238 (−181)	−502 (−569)	−641 (−772)	−484
AF	94	81	105	93	70		183 (129)	−271 (−259)	−515 (−462)	−174
BB	126	120	147	146	76	71		−564 (−388)	−824 (−591)	−303
ET	88	241	−41	82	113	95	69		−327 (−203)	085
SL	80	−137	−731	75	120	90	72	62		288

The unit of measurement is .001.

Legend: $\sigma_{i-j} \setminus \hat{z}_{ij} \, (\tilde{z}_{ij})$

(Note that we are using $\hat{\mu}$ obtained from Case V in the above equation, which, strictly speaking, is not quite right. If Case V is not correct, the estimate of μ should also be modified. This is one of the reasons why this is an approximate method.) Then

$$\frac{\hat{z}_{ij}}{\tilde{z}_{ij}} = \sigma_{i-j} \tag{7}$$

should give an estimate of σ_{i-j}.

This method of estimating σ_{i-j} was applied to the data collected by Rumelhart and Greeno (1971). The data pertain to pairwise preference judgments on nine stimuli obtained from 234 subjects. The stimuli consist of three groups of people, three politicians (Lyndon Johnson (LJ), Harold Wilson (HW) and Charles de-Gaulle (CD)), three atheletes (Johnny Unitas (JU), Carl Yastzremski (CY) and A. J. Foyt (AF)), and three actresses (Brigitte Bardot (BB), Elizabeth Taylor (ET) and Sophia Loren (SL)). The stimuli were deliberately chosen in such a way that they differ in their degree of similarity. People in a same professional group are expected to be more similar to each other than those in different groups.

Table 1 shows the observed $\hat{z}_{ij}$ (obtained by applying F^{-1} to the observed $\hat{p}_{ij}$) and the predicted $\tilde{z}_{ij}$ (obtained by applying Thurstone's Case V) in parenthesis in the upper diagonal portion of the table. (The scale values from Thurstone's Case V are listed in the right margin of the table.) Entries below the main diagonal are the estimates of σ_{i-j} obtained by Equation (7). It can be observed that the estimates of σ_{i-j} are generally smaller for pairs whose members belong to a same group than those for pairs whose members come from different professional groups.

In order to highlight the above observation metric multidimensional scaling (Takane, Young, & de Leeuw, 1977) was applied to the estimated σ_{i-j}. In this analysis four entries in Table 1 (enclosed by a dotted line) were treated as missing data. For these estimates $\hat{z}_{ij}$'s were so close to zero ($\hat{p}_{ij} \approx 1/2$) that they were deemed unreliable. Note that in three cases estimates are negative. While this should not happen theoretically, it happens in practice. Figure 1 shows a two-dimensional stimulus configuration obtained from the MDS analysis of σ_{i-j}. It can be clearly seen that the three groups of people form three distinct clusters in the multidimensional space, indicating that σ_{i-j} is indeed a measure of dissimilarity between stimuli.

Maximum Likelihood Estimation

Motivation

We have seen that σ_{i-j} serves as a measure of dissimilarity between i and j. Being a measure of dissimilarity it also serves as a measure of comparability between the two stimuli. Thus, there seems enough justification to estimate σ_{i-j} in Thurstone's model.

The method of estimating σ_{i-j} given in the previous section is a quick and dirty method. It has several undersirable properties. First, the estimated σ_{i-j} is unstable for values of $\hat{p}_{ij}$ near 1/2. The corresponding $\hat{z}_{ij}$ is close to zero in this case, so that a small change in $\hat{z}_{ij}$ may change σ_{i-j} drastically. Secondly, no estimate of σ_{i-j} can be obtained when $\hat{p}_{ij}$ is exactly one-half. In this case $\hat{z}_{ij}=0$ and σ_{i-j} in (7) is not definable (i.e., σ_{i-j} is infinitely unstable in this limiting situation). Thirdly, the estimate of σ_{i-j} may be negative. As we have seen in Rumelhart and Greeno's example, negative estimates of σ_{i-j} can be obtained when the signs of $\hat{z}_{ij}$ and $\hat{z}_{ij}$ are opposite. Since σ_{i-j} is by definition non-negative, this is not a desirable outcome. (A negative estimate of σ_{i-j} may occur frequently when $\hat{p}_{ij}$ is close to 1/2.) Finally, and most importantly, we have to assume that the estimates of μ's obtained from Case V remain intact to be used in (6) for deriving (7). This is obviously not strictly true. Furthermore, the estimates of μ's from Case V are likely to be distorted to start with due to the violation of Case V assumptions. In Fig. 1 the estimates of scale values derived under Case V are shown alongside the stimulus points. It may be readily seen that a certain direction (from upper left to lower right) in this two-dimensional stimulus configuration is highly correlated with the estimated μ's.

This may be an artifact caused as a consequence of using distorted μ's in both (5) and (6). That is, some bit of information which should be captured in μ's might have been mixed up in the estimates of σ_{i-j}, making a direction in the multidimensional space highly correlated with preference values of stimuli.

It seems necessary to develop an estimation procedure which is free from all of the above inconveniences. It is not too difficult to construct a maximum likelihood estimation procedure. However, the problem is how to reduce the number of parameters in the general case of Thurstone's model (Takane, 1975). The $n(n-1)/2$ observed choice probabilities are the minimum sufficient statistics (Wilks, 1962) for p_{ij}, and consequently no model with a larger number of parameters can possibly do better than this model.

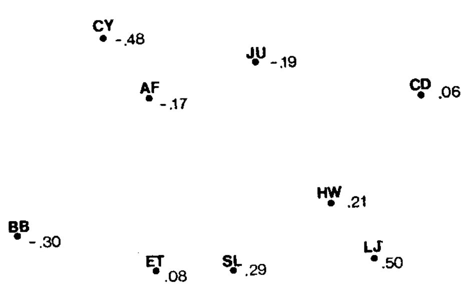

FIG. 1. Multidimensional stimulus configuration derived from estimated σ_{i-j} as defined in (7).

We propose the following model to avoid the over-parametrization of Thurstone's general model:

$$\sigma_{i-j}^2 = \sigma_i^2 + \sigma_j^2 - 2\sum_{a=1}^{A} b_{ia}b_{ja}, \quad (i, j = 1, \ldots, n) \tag{8}$$

where

$$\sigma_i^2 = \sum_{a=1}^{A} b_{ia}^2, \qquad (i = 1, \ldots, n). \tag{9}$$

That is, the covariance matrix between discriminal processes is assumed to have a prescribed rank ($=A$). The constraint on the covariance matrix expressed in (8) is also convenient to ensure the non-negative definiteness of a covariance matrix. We call this proposed model the "factorial model".

There is an interesting correspondence between the above proposal and the analysis we had undertaken in the previous section. Let $\boldsymbol{\Sigma}_{i-j}$ be the $n \times n$ matrix of σ_{i-j}^2. It was tacitly assumed in the previous section that the set of estimated σ_{i-j} had a two-dimensional representation in the Euclidian space. This is equivalent to saying that matrix $\boldsymbol{S}$ obtained by applying the Young-Householder (1938) transformation to $\boldsymbol{\Sigma}_{i-j}$ has rank 2. That is,

$$\boldsymbol{S} = -\frac{1}{2}\boldsymbol{J}\boldsymbol{\Sigma}_{i-j}\boldsymbol{J} = \boldsymbol{X}\boldsymbol{X}', \tag{10}$$

where $\boldsymbol{J}$ is a centering matrix of order n and $\boldsymbol{X}$ is an $n \times 2$ matrix of stimulus configuration (whose origin is set at its centroid). Let $\tilde{\boldsymbol{\Sigma}}_{i-j}$ be the $n \times n$ matrix of σ_{i-j}^2 having the structure defined in (8). If we apply the same transformation to $\tilde{\boldsymbol{\Sigma}}_{i-j}$, we obtain

$$\tilde{\boldsymbol{S}} = -\frac{1}{2}\boldsymbol{J}\tilde{\boldsymbol{\Sigma}}_{i-j}\boldsymbol{J} = \boldsymbol{JCJ} = \boldsymbol{B}^*\boldsymbol{B}^{*\prime}, \tag{11}$$

where $\boldsymbol{C}$ is the matrix of σ_{ij} and $\boldsymbol{B}^* = \boldsymbol{JB}$ ($\boldsymbol{B}$ is a matrix of b_{ia}). If the double-centered covariance matrix ($\boldsymbol{JCJ}$) has rank 2, then $\boldsymbol{B}^*\boldsymbol{B}^{*\prime}$ is equivalent to $\boldsymbol{XX}'$ in (10), and $\boldsymbol{X} = \boldsymbol{B}^*\boldsymbol{T}$ for some orthogonal transformation matrix $\boldsymbol{T}$. The fact that $\boldsymbol{JCJ}$ is as good as $\boldsymbol{C}$ (and consequently, that $\boldsymbol{B}^*$ is as good as $\boldsymbol{B}$) may be readily seen by pointing out that $\boldsymbol{\Sigma}_{i-j}$ (and $\tilde{\boldsymbol{\Sigma}}_{i-j}$) is invariant over the transformation of $\boldsymbol{C}$ of the form:

$$\hat{\boldsymbol{C}} = \boldsymbol{C} + \boldsymbol{la}' + \boldsymbol{al}'$$

where $\boldsymbol{a}$ is an arbitrary n-component vector and $\boldsymbol{l}$ a vector of ones. That is, the rows and columns of $\boldsymbol{C}$ are determined only up to an additive constant.

Estimation Procedure

Let Y_{ij} denote the observed frequency with which stimulus i is chosen over stimulus j. Then the probability of that event occurring Y_{ij} times out of N_{ij} replicated observations follows a binomial distribution,

$$\Pr(Y_{ij} \mid p_{ij}) \propto p_{ij}^{Y_{ij}}(1 - p_{ij})^{N_{ij} - Y_{ij}},$$

where p_{ij} is given in (1). The likelihood function for the total set of observations can be stated as

$$L \propto L' = \prod p_{ij}^{Y_{ij}}(1 - p_{ij})^{N_{ij} - Y_{ij}} \tag{12}$$

where the product may be taken over the pairs of i and j for which Y_{ij}'s are actually observed. Note that in defining (12) we have assumed that the difference processes, $X_i - X_j$, at two distinct occasions are statistically independent. Note also that we are assuming that two discriminal processes involved in a single judgment have a non-zero covariance. A minimal condition for the independence of the difference processes involved in two distinct judgments has been given in Takane (1978).

Taking the log of L' in (12) we have

$$\ln L' = \sum [Y_{ij} \ln p_{ij} + (N_{ij} - Y_{ij}) \ln (1 - p_{ij})]. \tag{13}$$

We wish to determine μ's and b_{ia}'s in such a way that this quantity is maximized. Fisher's scoring algorithm seems useful to solve a set of likelihood equations derived by differentiating $\ln L'$ with respect to the model parameters. This method has been successfully used in many situations (see Takane, 1978, 1979, 1980, for example) similar to the present case. Fisher's scoring algorithm is an iterative procedure, which updates the parameter values by

$$\boldsymbol{\theta}^{(q+1)} = \boldsymbol{\theta}^{(q)} + \alpha^{(q)} \boldsymbol{I}(\boldsymbol{\theta}^{(q)})^{-1} \boldsymbol{s}(\boldsymbol{\theta}^{(q)}) \quad (14)$$

where the parenthesized superscript indicates an iteration number, $\boldsymbol{\theta}$ is a vector of parameters, $\boldsymbol{s}$ is Fisher's scoring vector, $\boldsymbol{I}$ is Fisher's information matrix, and α is a step size. The updating formula (14) is iteratively applied until all parameter values are stabilized.

Fisher's scoring vector is the vector of first derivatives of $\ln L'$ with respect to $\boldsymbol{\theta}$, namely

$$\boldsymbol{s}(\boldsymbol{\theta}) \equiv \frac{\partial \ln L'}{\partial \boldsymbol{\theta}} = \Sigma\left(\frac{Y_{ij} - N_{ij} p_{ij}}{p_{ij}(1 - p_{ij})}\right)\left(\frac{\partial p_{ij}}{\partial \boldsymbol{\theta}}\right), \quad (15)$$

where the elements of $\partial p_{ij}/\partial \boldsymbol{\theta}$ are given by

$$\frac{\partial p_{ij}}{\partial \mu_k} = (\delta_{ik} - \delta_{jk}) f(z_{ij})/\sigma_{i-j}, \quad (16)$$

$$\frac{\partial p_{ij}}{\partial \sigma_k} = -(\delta_{ik} + \delta_{jk}) f(z_{ij}) z_{ij} \sigma_k / \sigma_{i-j}^2, \quad (17)$$

and

$$\frac{\partial p_{ij}}{\partial b_{ka}} = \delta_{ik} f(z_{ij}) z_{ij} (b_{ja} - b_{ia})/\sigma_{i-j}^2 + \delta_{kj} f(z_{ij}) z_{ij} (b_{ia} - b_{ja})/\sigma_{i-j}^2. \quad (18)$$

In the above formulae $\delta_{..}$ is a Kronecker delta, f is the density function of the standard normal distribution, and $z_{ij} = \mu_{i-j}/\sigma_{i-j}$. Fisher's information matrix is the covariance of $\boldsymbol{s}(\boldsymbol{\theta})$, and takes the following form under the present circumstance:

$$\boldsymbol{I}(\boldsymbol{\theta}) = \Sigma\left(\frac{N_{ij}}{p_{ij}(1 - p_{ij})}\right)\left(\frac{\partial p_{ij}}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial p_{ij}}{\partial \boldsymbol{\theta}}\right)'. \quad (19)$$

When the information matrix is singular, we may use the Moore-Penrose inverse of $\boldsymbol{I}(\boldsymbol{\theta})$ in (14) (Ramsay, 1978).

Results and Discussion

A computer program has been written which incorporates the model proposed in the previous section as well as Thurstone's conventional cases (Case V and Case III). In this section we report some of the results we have obtained with Rumelhart and Greeno's (1971) data we used previously.

TABLE 2
The hierarchy of fitted models and a summary of goodness of fit statistics

	Null Model	
	−309.8	
	691.5	
	(36)	
Case III		**Factorial Model ($A=2$)**
−326.7		−313.2
685.4		670.5
(16)		(22)
Case V		
−350.6		
717.2		
(8)		

Legend: Log-likelihood (+const.)
AIC (+const.)
(d.f. of the model)

For the purpose of choosing the best fitting model, four different models have been fitted to the same set of data, and the results are reported in Table 2.

In the table the null model refers to a model in which no further structural assumptions (like the one in (1)) are imposed on the choice probabilities. It is well known that observed choice probabilities ($\hat{p}_{ij} = Y_{ij}/N_{ij}$) are the maximum likelihood estimates of population choice probabilities (p_{ij}) when no specific substructures are assumed of p_{ij}. Furthermore, as noted earlier, $\hat{p}_{ij}$ is the minimum sufficient statistic for p_{ij}. Hence, this model represents the most unrestrictive model conceivable.

The other three models assume some specific structures under p_{ij} as given in (1). Case V further assumes that $\sigma_{i-j} = 1$ for all pairs of i and j. Case III assumes $\sigma_{ij} = 0$ ($i \neq j$). Finally, the factorial model assumes

$$\sigma_{ij}=\sum_{a=1}^{A} b_{ia}b_{ja}\,.$$

All these three models are special cases of the null model. Case V in turn is a special case of Case III. Note that the factorial model is not a proper generalization of either Case III or Case V, because with the restriction stated in (9) conditions for Case III or Case V cannot be generated.

The likelihood of the most unrestrictive null model is naturally the highest among the fitted models. (See the top figure under each model, which gives the log-likelihood (plus constant) of the model). However, it uses 36 parameters ($=9\times 8/2$) to achieve this. Quite naturally a larger likelihood can be attained if a larger number of parameters are used to describe the data. Thus, in order to evaluate the performance of a model in relation to other models, a goodness of fit statistic, which explicitly takes into account the number of parameters in the model, is called for. The AIC statistic, proposed by Akaike (1974), satisfies this need. The AIC of model π is defined by

$$\mathrm{AIC}(\pi)=-2\ln L+2n_\pi \qquad (20)$$

where $\ln L$ is the log-likelihood of model π maximized over its parameters and n_π is the number of parameters in model π. The model which gives the minimum value of the AIC is considered the best fitting model.

The AIC values associated with the four fitted models are given as the second entries in the table (right below the log-likelihoods). The factorial model with two factors has the minimum AIC value of 670.5. Thus, according to the AIC criterion the factorial model is deemed the best fitting model.

The log likelihood of the factorial model ($=-313.2$) is slightly smaller than that of the null model ($=-309.8$). However, the former uses only 22 parameters while the latter uses 36 parameters. The 14 additional parameters in the null model do not significantly improve the goodness of fit of the model.

The 22 degrees of freedom for the factorial model are calculated as follows. There are 27 parameters estimated (9 μ's and 9×2 b_{ia}'s) of which five parameters may be arbitrarily chosen, one for the origin of μ's, two for the origins of b_{ia} ($a=1, 2$) and one for rotation, and the remaining one for the scale of μ's and b_{ia}'s. Note that there is one common scale factor for both μ's and b_{ia}'s. Thus, the degrees of freedom for the factorial model is, in general, given by $nA-A(A+1)/2-1$ where n is the number of stimuli and A is the assumed rank of the covariance matrix between discriminal processes.

The estimated b_{ia}'s are plotted in Fig. 2. The three groups of people again form three clusters, though they are not as distinct as in Fig. 1. However, unlike Fig. 1, no direction in the space seems to be confounded with preference values of stimuli (which are given alongside the stimulus points in the configuration). It may be pointed out that less preferred stimuli tend to be in the center of the configuration, while more preferred stimuli tend to be located outside. This means that more preferred stimuli tend to have larger discriminal dispersions. Whether this is mere coincidence or reflects some important psychological truth is, however, yet to be investigated.

In certain cases the restriction implied

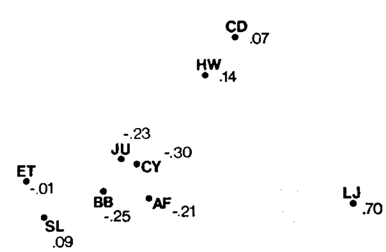

FIG. 2. The plot of an estimated square-root factor of the covariance matrix between discriminal process.

by (9) may be too strict, and we may wish to replace it by a somewhat weaker restriction that

$$\sigma_i^2 \geqslant \sum_{a=1}^{A} b_{ia}^2, \qquad (i=1,\ldots,n). \quad (21)$$

This is equivalent to hypothesizing a common factor analysis type of structures on the covariance between discriminal processes as opposed to a component analysis type of structures implied by (9). This restriction has a desirable property that both Thurstone's Case III and Case V can be subsumed under the factorial model. In the present case, however, the log likelihood of the factorial model with restriction (9) is already so close to that of the null model (the difference is only 3.4) that it is not likely that the factorial hypothesis with this weaker restriction can substantially improve the goodness of fit of the model. The restriction given in (21) may still have some general appeal.

One of the major implications of the present study relates to Takane's (1977, 1978) work on a maximum likelihood estimation procedure for nonmetric multidimensional scaling when dissimilarity measures are taken by pair comparisons of dissimilarities. A difficulty which has been encountered was the differential degree of comparability between dissimilarities. For example, dissimilarities are more easily discriminable when they are defined on a single common dimension than when they are defined on two or more different dimensions. Since the degree of comparability is related to σ_{i-j}, and now that the procedure is found feasible, which incorporates different σ_{i-j}'s in Thurstone's model, this feature may be incorporated into the above multidimensional scaling procedure.

References

Akaike, H. 1974 A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.

Arbuckle, L., & Nugent, J. H. 1973 A general procedure for parameter estimation for the law of comparative judgment. *British Journal of Mathematical and Statistical Psychology*, **26**, 240–260.

Debreu, G. 1960 Review of R. D. Luce, Individual choice behavior: A theoretical analysis. *American Economic Review*, **50**, 186–188.

Halff, H. M. 1976 Choice theories for differentially comparable alternatives. *Journal of Mathematical Psychology*, **14**, 244–246.

Krantz, D. H. 1967 Rational distance functions for multidimensional scaling. *Journal of Mathematical Psychology*, **4**, 226–245.

Luce, R. D. 1959 *Individual choice behavior: A theoretical analysis*. New York: Wiley.

Ramsay, J. O. 1978 Confidence regions for multidimensional scaling analysis. *Psychometrika*, **43**, 145–160.

Restle, F. 1961 *Psychology of judgment and choice*. New York: Wiley.

Rumelhart, D. L., & Greeno, J. G. 1971 Similarity between stimuli: An experimental test of the Luce and Restle choice models. *Journal of Mathematical Psychology*, **8**, 370–381.

Sjöberg, L. 1968 Unidimensional scaling of multidimensional facial expressions. *Journal of Experimental Psychology*, **78**, 429–435.

Sjöberg, L. 1975 Uncertainty of comparative judgments and multidimensional structure. *Multivariate Behavioral Research*, **11**, 207–218.

Sjöberg, L., & Capozza, D. 1975 Preference and cognitive structure of Italian political parties. *Italian Journal of Psychology*, **2**, 391–402.

Takane, Y. 1975 Generalized uni- and multidimensional Thurstonian scaling. Unpublished dissertation proposal. The University of North Carolina.

Takane, Y. 1977 Statistical procedures for nonmetric multidimensional scaling. Unpublished doctoral dissertation, The University of North Carolina.

Takane, Y. 1978 A maximum likelihood method for nonmetric multidimensional scaling: I. The case in which all empirical pairwise orderings are independent—theory and evaluations. *Japanese Psychological Research*, **20**, 7–17 and 105–114.

Takane, Y. 1979 Maximum likelihood multidimensional scaling from directional rankings of similarity. A paper presented at the Psychometric Society Meeting, Monterey.

Takane, Y. 1980 Multidimensional successive categories scaling: A maximum likelihood method. *Psychometrika*. (In press).

TAKANE, Y., YOUNG, F. W., & DE LEEUW, J. 1977 Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika*, **42**, 7–67.

THURSTONE, L. L. 1927 A Law of comparative judgments. *Psychological Review*, **34**, 273–286.

TVERSKY, A. 1972a Choice by elimination. *Journal of Mathematical Psychology*, **9**, 341–367.

TVERSKY, A. 1972b Elimination by aspects: A theory of choice. *Psychological Review*, **79**, 281–299.

TVERSKY, A., & RUSSO, E. 1969 Substitutability and similarity in binary choices. *Journal of Mathematical Psychology*, **6**, 1–12.

WILKS, S. S. 1962 *Mathematical statistics*. New York: Wiley.

YOUNG, G., & HOUSEHOLDER, A. S. 1938 Discussion of a set of points in terms of their mutual distances. *Psychometrika*, **3**, 19–22.

(Received March 21, 1980)