PSYCHOMETRIKA—VOL. 57, NO. 4, 599–614
DECEMBER 1992

THE IRRELEVANCE OF DISTRIBUTIONAL ASSUMPTIONS ON REACTION TIMES IN MULTIDIMENSIONAL SCALING OF SAME/DIFFERENT JUDGMENT TASKS

GERT STORMS AND LUC DELBEKE

KATHOLIEKE UNIVERSITEIT LEUVEN, BELGIUM

Takane and Sergent developed a model (MAXRT) for scaling same/different judgments and response times (RTs) simultaneously. The model assumes that RTs are distributed lognormally. Our experiment showed that the RT distribution of the judgments might be task dependent. It is shown that lognormal RTs provide a far better fit than exponential, normal, and Pareto distributed RTs (with the same means and variances), but that the final parameter estimates from the data set with lognormal RTs hardly differ from the alternatively distributed RTs. Finally, despite the robustness of the distributional assumption of the RTs with respect to the parameter estimates, it is shown that RTs have an informational value that is not contained in the same/different judgments alone.

Key words: response times, distributions, maximum likelihood scaling, likelihood ratio χ^2, Shapiro-Wilk test, AIC statistic, pairwise judgments.

Despite the popularity of response times (RTs) as a dependent variable in psychological research, surprisingly few studies have used RTs as input data for scaling. A remarkable exception to this lack of interest in latencies is Takane and Sergent's (1983) MAXRT model. Whereas earlier applications of multidimensional scaling to RTs merely used this dependent variable as an alternative to direct similarity ratings or confusion data (Brown & Andrews, 1968; Harrington & Brown, 1972; Kak & Brown, 1979; Melara, 1989; Young, 1970), the MAXRT model introduced a method that takes full account of the bivariate nature of same/different judgments and RT data (Takane & Sergent, 1983). The model assumes that RTs are distributed lognormally. The first question this paper addresses concerns the adequacy of this distributional assumption. Next, the effect of a violation of this assumption is investigated by means of a simulation study. Finally, the data of a same/different task are analyzed with and without the RT data to obtain a clear view of the additional information provided by the RTs.

Preliminaries

A Short Review of Takane and Sergent's (1983) MAXRT Model

Given the importance of Takane and Sergent's model for the research reported in this paper, a short review will be presented. The model assumes that a set of n stimuli has some parametric representation, from which a dissimilarity between two stimuli (i and j) is uniquely defined:

$$d_{ij} = d(\Theta_i, \Theta_j), \tag{1}$$

The authors thank Jan Beirlant, Geert De Soete, Yoshio Takane, the three reviewers, and the Editor, for their useful comments on earlier versions of this manuscript. We are also indebted to Noël Bovens for his help in programming the experiments.

Requests for reprints should be sent to Gert Storms, University of Leuven, Tiensestraat 102, B-3000 Leuven, BELGIUM.

0033-3123/92/1200-90297$00.75/0

where Θ_i and Θ_j are parameter vectors characterizing the stimuli, and d is a function that expresses their combination rule. MAXRT allows the application of three dissimilarity models: (a) the Minkowski power distance model (Kruskal, 1964a, 1964b), (b) the linear model (Medin & Shaffer, 1978), and (c) Tversky's (1977) feature matching model. We restrict ourselves to the first model, where the stimuli are represented as points in an M-dimensional space, and where d_{ij} is defined as

$$d_{ij} = \left\{ \sum_{m=1}^{M} |\theta_{im} - \theta_{jm}|^u \right\}^{1/u}, \tag{2}$$

with θ_{im} the coordinate of stimulus i on dimension m, and $u \geq 0$. We use the Euclidean distance only ($u = 2$). MAXRT further assumes that d_{ij} is error-perturbed by the following process:

$$\lambda_{ijkr} = d_{ij} e_{ijkr}, \tag{3}$$

where e_{ijkr} is the error random variable operating at replication r, and where

$$\ln e_{ijkr} \sim \mathcal{N}(0, \sigma_k^2). \tag{4}$$

The subscript k represents an experimental session. The parameter b_k symbolizes a threshold value, and Y_{ijkr} is the random variable for the same/different judgment, defined as

$$\begin{aligned} Y_{ijkr} &= 1 \quad \text{when} \quad \lambda_{ijkr} < b_k, \quad \text{(same judgment)} \\ &\;\;\; 0 \qquad\qquad\quad \lambda_{ijkr} \geq b_k. \quad \text{(different judgment)} \end{aligned} \tag{5}$$

The probability function of the variable Y_{ijkr} is

$$\begin{aligned} Q_{ijk} &= P(Y_{ijkr} = 1) \\ &= \int_{-\infty}^{v_{ijk}} \phi(z)\, dz = \Phi(v_{ijk}), \end{aligned} \tag{6}$$

$$1 - Q_{ijk} = P(Y_{ijkr} = 0),$$

where

$$v_{ijk} = \frac{\ln b_k - \ln d_{ij}}{\sigma_k}, \tag{7}$$

and $\phi(\cdot)$ and $\Phi(\cdot)$ are, respectively, the density and the distribution function of the standard normal distribution. MAXRT further assumes that the RTs T_{ijkr} are inversely related to the absolute value of the difference between the decision statistic ($\ln d_{ij}$) and the criterion ($\ln b_k$). (For more details on this traditional assumption in signal detection literature, see Murdock, 1985). It is also assumed that this relation is different, depending on the value of Y_{ijkr}. For different judgments, it is assumed that

$$\ln T_{ijkr} \sim \mathcal{N}\{p_k(\ln d_{ij} - \ln b_k) + a_k;\, q_k^2\}, \tag{8}$$

and for same responses,

$$\ln T_{ijkr} \sim \mathcal{N}\{p_k(\ln b_k - \ln d_{ij}) + a_k;\, q_k^2\}, \tag{9}$$

where p_k is assumed to be negative. The model further assumes that $q_k = -p_k\sigma_k$. The joint density function of T_{ijkr} and Y_{ijkr} is written as

$$g(t_{ijkr}, y_{ijkr}) = \{g^{(s)}(t_{ijkr})Q_{ijk}\}^{y_{ijkr}}\{g^{(d)}(t_{ijkr})(1 - Q_{ijk})\}^{1 - y_{ijkr}},$$

where $g^{(s)}(t_{ijkr})$ and $g^{(d)}(t_{ijkr})$ are the lognormal density functions for, respectively, same and different judgments. The joint likelihood can be stated as

$$L = \prod_k \prod_{i,j} \prod_r g(t_{ijkr}, y_{ijkr}).$$

The model parameters can be divided into the parameters (θ_{im}) of the representation model and the parameters of the response model (b_k, p_k, a_k, and q_k^2). The parameters are estimated by maximizing the log-likelihood function, ln L.

The Assumption of Lognormally Distributed RTs

The conditional lognormality of the RTs is assumed to follow from the lognormality of λ_{ijkr}. Takane and Sergent (1983) give two reasons to justify the assumed lognormality of the RTs: (a) it is a positively skewed distribution, which is typical of most empirically obtained RT data, and (b) Chocholle (1940) reported that for simple RTs, the standard deviation is roughly proportional to the mean, which makes the lognormal distribution a suitable one. Both arguments are not quite convincing. Firstly, as Luce (1986) remarked, the lognormal distribution is not the only parametric distribution that is positively skewed; other positively skewed distributions that are used in the RT literature are the exponential distribution (e.g., Kohfeld, Santee, & Wallace, 1981; Scheiblechner, 1979), the gamma distribution (e.g., Anderson & Bower, 1973; Townsend & Ashby, 1983), the generalized gamma distribution (McGill & Gibbon, 1965), nonidentified distributions but with an exponential tail (e.g., Green & Luce, 1967, 1971; Luce & Green, 1970), convolutions of normal and exponential distributions (e.g., Ratcliff, 1978; Ratcliff & Murdock, 1976), the double monomial distribution (Snodgrass, Luce, & Galanter, 1967) and the Weibull distribution (Ida, 1980). Secondly, the proportionality of standard deviations and means is also a characteristic of the exponential and the gamma distribution, and it is questionable whether this proportionality phenomenon can be generalized to RTs registered in more complex tasks (e.g., choice tasks).

The Experiment

To investigate the RT distribution, an experiment was designed that was similar to the experiment of Takane and Sergent (1983).

Experimental Procedure

Subjects. Five subjects, aged between 19 and 28, were paid to participate in the experiment. The results of the sixth subject, drawn randomly from Sergent and Takane's (1987) experiment (further denoted by TS), were also analyzed.

Stimuli. Two different stimulus sets were used. Three subjects (LA, AB, and DC) were presented schematic faces; two subjects (JM and WL) were shown pairs of geometric figures. The schematic faces were different from the stimuli used by Takane and Sergent (1983; Sergent & Takane, 1987) and were constructed with the Flury and Riedwyl (1981, 1983) procedure for representing multivariate data. Only three param-

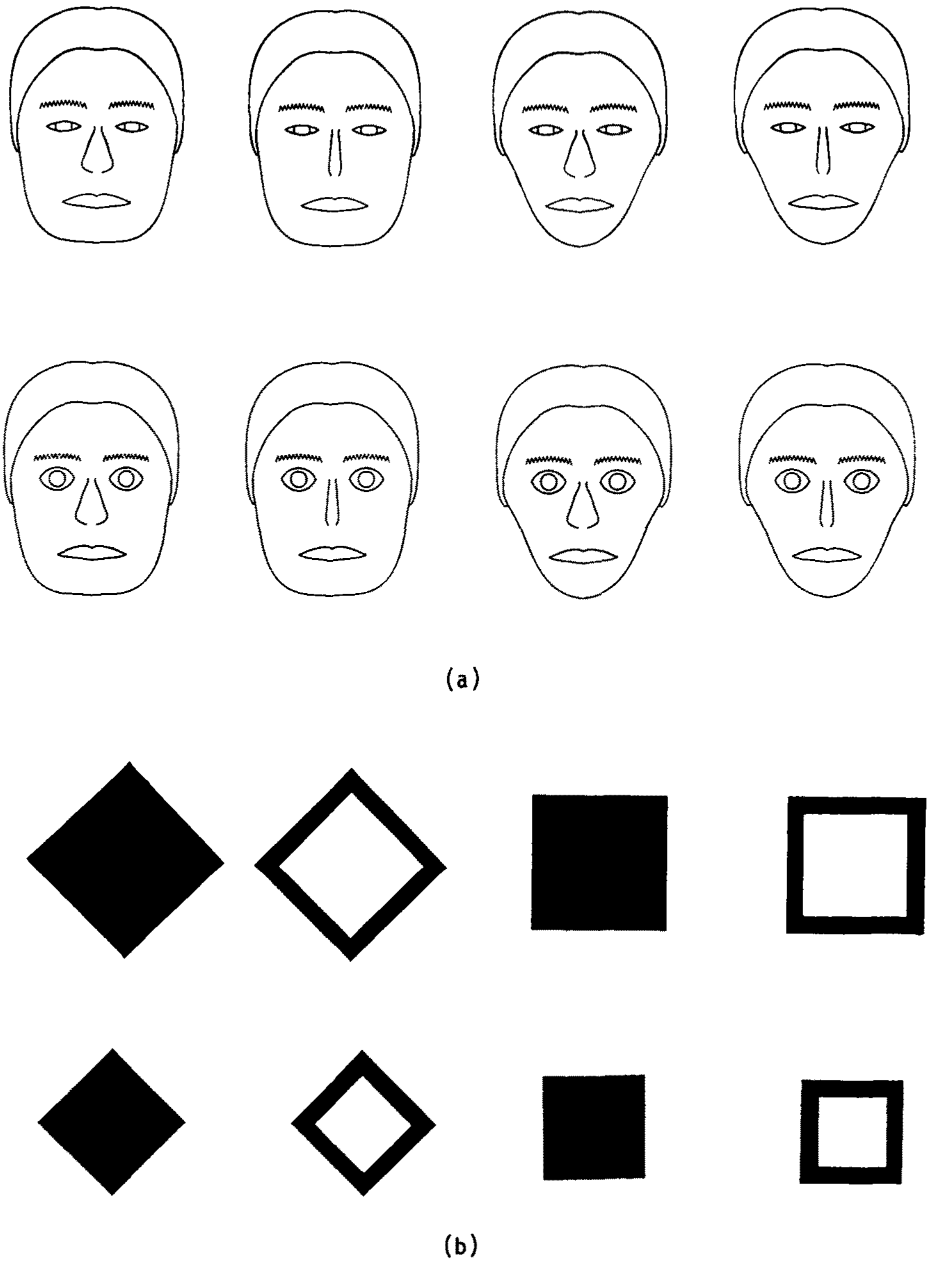

FIGURE 1
(a) Face stimuli; (b) geometric figure stimuli.

eters were varied: *eyes*, *jaws*, and *nose*. The perceptually most salient parameters were chosen (De Soete, 1986). The eight different stimuli were constructed by factorially combining two levels of the three varying features (parameter values 0.2 and 0.8). The second stimulus set consisted of eight symmetric geometric figures, constructed by factorially combining two levels of three features: *form* (square or diamond), *size* (large or small), and *filling*. The two stimulus sets are presented in Figure 1 ((a) and (b), respectively).

Apparatus. All possible pairs of different stimuli were photographed twice (left/right permutation). The pairs were projected on a transparent screen with a random access projector (KODAK S-RA 2500). The presentation sequence was random. Responses were registered by two push buttons connected to an Apple II^+ microcomputer. After every trial, the identification number of the stimulus pair, the response (same/different), and the response time was recorded.

Procedure. Subjects were acquainted with the stimuli in a learning session similar to the six experimental sessions. Every session started with 24 training trials that were not included in the data set. Next, 448 pairs were presented; half of them were different pairs and half were same pairs. Every same pair was presented 28 times; all the 56 different pairs were shown 4 times in every session. Subjects were tested individually. They were seated in front of a screen (at 114 cm distance) with a chin and forehead rest used to fixate the head. During the experiment the room was darkened. The search time to select the next stimulus was fixed at six sec, and was followed by a warning tone of 0.5 sec. One second later the stimulus pair was presented. The faces, when projected, subtended a visual angle of 3°50′ in width and 4°42′ in height. Two subjects used the left button for same responses; three subjects used the left button for different responses. The stimuli remained on the screen until a response was given, but with a maximum presentation time of 2.5 sec. Subjects were asked to respond as accurately and as quickly as possible. A session lasted between 75 and 120 minutes.

Results

Because we wanted to focus our attention on the impact of the assumed lognormal distribution of the RTs in MAXRT, only one representation model was selected for the analysis of the data: the Minkowski power distance function with Euclidean distances. Furthermore, due to the incompatibility of the lognormal distribution with zero dissimilarities in the Minkowski power distance model, only the data of the different pairs were analyzed, and symmetry of the dissimilarities was assumed ($d_{ij} = d_{ji}$). (The latter two restrictions are also present in Takane & Sergent, 1983; and Sergent & Takane, 1987.) Solutions were obtained for spaces with different dimensionalities. The optimal number of dimensions, determined with the likelihood ratio χ^2, were 4, 4, 3, 3, 3, and 2 for subjects TS, LA, AB, and DC (who compared the faces), and JM and WL (who worked with the geometric figures), respectively.

Testing the lognormal assumption. The MAXRT model makes clear assumptions concerning the distribution of the RTs (see (8) and (9)). They imply that response times are distributed differently for every stimulus pair (i, j) in every session k. But due to the time consuming nature of the experiment, maximally eight RTs from every distribution could be observed. Takane and Sergent (1983) verified the adequacy of the lognormal assumption by means of a normal quantile plot of the log RT data per session by pooling the data according to the model. The normalized log residuals were plotted against the normal quantile scores. Visual inspection of the plot lead them to conclude that the lognormal distribution gave a reasonably good fit in all sessions for the two subjects, "although some irregularities (occurred) near the extreme ends of the distribution" (Takane & Sergent, 1983, p. 419).

Due to the central importance of the lognormal assumption in the research reported here, a more rigorous technique to verify this assumption was adopted. We again pooled the data (but only the different responses for different pairs) per session, and then tested the normality of these data with the Shapiro-Wilk test (Royston, 1982; Shapiro & Wilk, 1965) which provides the most powerful omnibus test of normality known (D'Agostino & Stephens, 1986; Pearson, D'Agostino, & Bowman, 1977). The data for two out of six subjects (TS and AB) showed no evidence for rejecting the lognormal hypothesis. The RTs for subjects DC and LA reject this null hypothesis for one and for three sessions, respectively, and for two subjects (JM and WL) who carried out the relatively simpler task (comparing geometric figures) the lognormal hypothesis could be rejected for all six of the sessions.

The Shapiro-Wilk test was also applied to the unpooled data, leaving maximally eight replications per stimulus pair. Using this much weaker test 168 times per person (28 stimulus pairs × 6 sessions), the null hypothesis could still be rejected far more than could be expected by chance. (At $\alpha = .05$, only 8 or 9 out of 168 replications can be expected to be significant, while we observed 17, 18, 15, 8, 24, and 26 rejections for subjects TS, LA, DC, AB, WL, and JM, respectively.)

It is remarkable that, for the pooled data as well as for the unpooled RTs, clearly more rejections were observed for the subjects who compared the geometric figures, indicating that the RT distribution of same/different judgments might be task dependent.

Goodness-of-fit. A second, but indirect way of checking the validity of the MAXRT assumptions concerning the RTs is the following. If the model holds, then the log RTs are normally distributed with mean

$$\mu_{ijk} = \hat{p}_k(\ln \hat{d}_{ij} - \ln \hat{b}_k) + \hat{a}_k,$$

for different responses, and

$$\mu_{ijk} = \hat{p}_k(\ln \hat{b}_k - \ln \hat{d}_{ij}) + \hat{a}_k,$$

for same responses, with standard deviation $\hat{q}_k$, as determined by the parameters. Therefore, it follows that the sum of the squared standard normal deviations

$$\sum_k \sum_{i,j} \sum_r \frac{(\ln T_{ijkr} - \mu_{ijk})^2}{q_k^2},$$

is approximately chi-square distributed, with degrees of freedom equaling the number of components ($N = 1344$), minus the number of free parameters. The critical values (at $\alpha = .01$) are 1419.5 for the four-dimensional, and 1423.6 for the three-dimensional model. The obtained values for all subjects reject the model (1470.1, 1420.3, 1538.1, 1589.4, 1604.3, and 1626.2 for TS, LA, DC, AB, JM, and WL, respectively), but the data do not allow us to attribute this lack-of-fit to the distributional assumption alone. Takane (personal communication, June, 8, 1990) suggested that the lower fit for the geometric data (subjects JM and WL) could be due to an inappropriate stimulus display time. In Sergent and Takane (1987), different presentation times had to be tried out for every subject until a good fit was obtained for the data from geometric stimulus sets. Anyway, it is obvious that the presented stimulus material alone cannot account for the differences in goodness-of-fit; only for one of the three subjects who were presented schematic faces was an acceptable goodness-of-fit value obtained.

Reliability of the RTs. A necessary condition for an acceptable goodness-of-fit value for the RTs is that the mean log RTs for the stimulus pairs in different sessions should have comparable values. MAXRT assumes a linear relation (see (8) and (9)), and consequently a high correlation between the mean log RTs for the different pairs among the six sessions. For every subject, an intraclass coefficient (Shrout & Fleiss, 1979) was calculated, which can be interpreted as a general correlation between these mean log RTs. Considering the sessions as a random variable, and the 28 different pairs as a fixed variable, results in the adjustment of the ICC(2-1) coefficient (Shrout & Fleiss, 1979) for a mixed model (Kirk, 1982):

$$\mathrm{ICC} = \frac{MS_{sp} - MS_{sp\times s}}{MS_{sp} + \left(\left(\frac{kn}{n-1}\right) - 1\right)MS_{sp\times s} + \left(\frac{k}{n-1}\right)(MS_s - MS_{sp\times x})},$$

where

k = the number of sessions (=6),
n = the number of stimulus pairs (=28),
MS_{sp} = the mean squares between the stimulus pairs,
MS_s = the mean squares between the sessions,
$MS_{sp\times s}$ = the mean squares of the interaction term: stimulus pairs × sessions.

The ICCs give us an idea of the reliability of the RTs. We obtained the values .63, .92, −.02, .32, .25, and .12, for subjects TS, LA, AB, DC, JM, and WL, respectively. The data of LA seem to be very reliable, as are the data of TS, but the correlations for the other subjects are low.

The relation between RTs, errors, and similarity judgments. The choice process, as formulated by Takane and Sergent (1983, see (6) and (7)) and the assumptions concerning the RTs ((8) and (9)) lead to the prediction that log RTs for stimulus pairs increase as the number of errors for these pairs increases. For none of the subjects was a monotone increasing relation observed, but only the data of WL strongly deviate from a general increasing relation. Figure 2a shows this relation for subject WL and for subject JM (randomly chosen from the other five subjects). (The number of errors in the experiment was quite low. All subjects made between 2.5 and 6% errors, and the distribution of these errors was highly similar to the data of Takane & Sergent, 1983.)

The similarity between the pairs of faces were judged by twenty other subjects on a seven-point rating scale. The correlations of these direct judgments with the dissimilarities from the MAXRT analyses are .78 (LA), .79 (DC), and .13 (AB). The high intercorrelation for the first two subjects are in accordance with other studies where a good resemblance between scaling solutions of direct similarity ratings and RT based measures were found (Behrman & Brown, 1968; Brown & Andrews, 1968; Podgorny & Garner, 1979; Young, 1970). The low correlation for subject AB may be indicative of a different subjective stimulus space.

The results of the analyses of the experimental data do not give a clear answer to the question of the importance of the lognormal assumption in MAXRT. The data provide goodness-of-fit values that vary widely, and several other measures indicate that the adequacy of the model differs for different subjects. The test of the lognormality of the RTs, however, clearly showed that this assumption is not always met by the data. To obtain a clearer view on the effect of violations of the lognormal assumption in MAXRT, a simulation study was carried out.

A Simulation Study

The Generation of Data

In this simulation study, data were generated in the following way. For all six subjects, sixteen simulations were run. The same/different response was generated

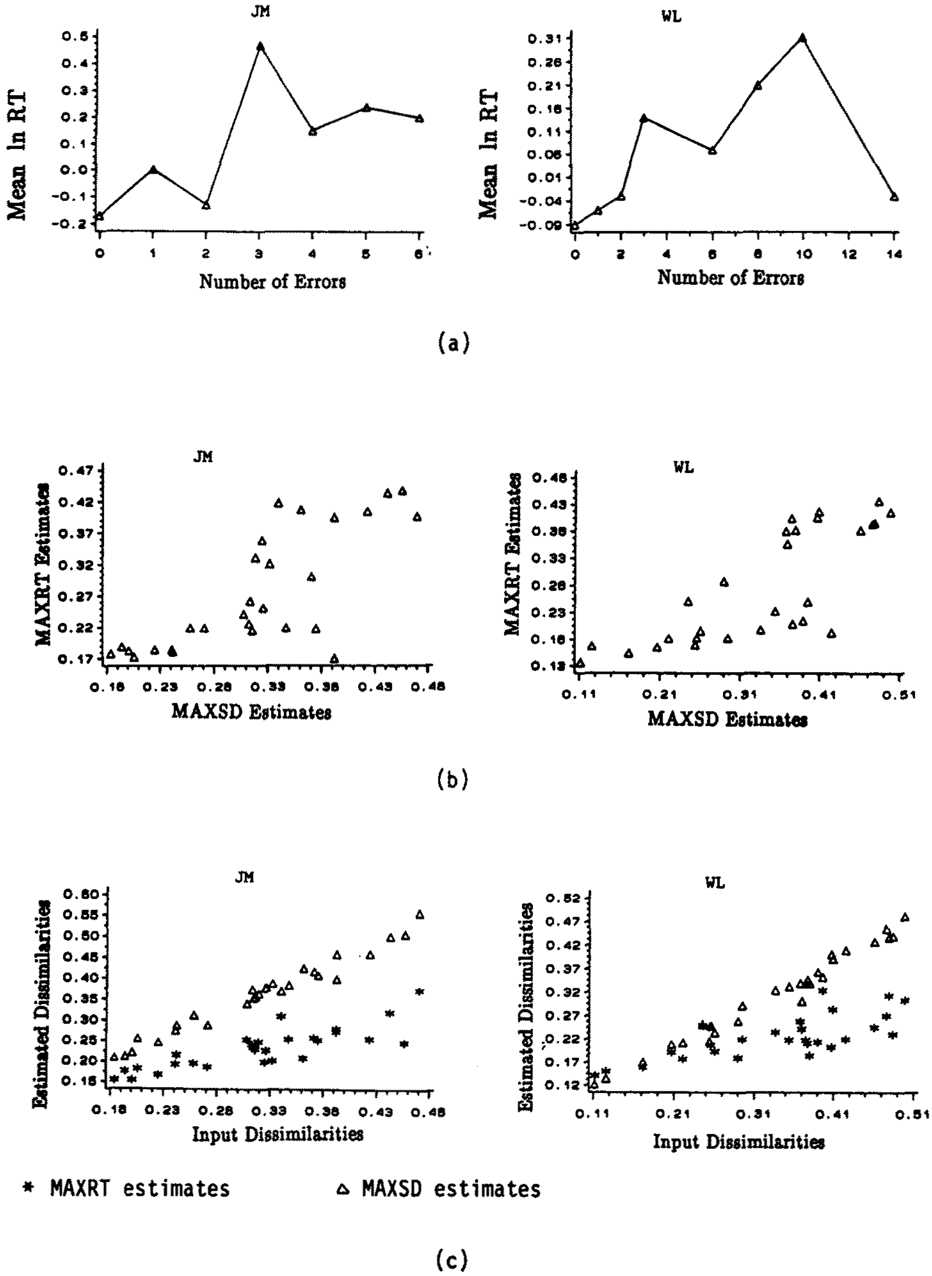

FIGURE 2
For representative subjects JM and WL, (a) mean RT for every number of errors; (b) MAXRT and MAXSD estimates; (c) true dissimilarities and the estimates from MAXRT and MAXSD.

stochastically, based on (6) and (7). As the values of the parameters b_k and σ_k, the estimates from the three-dimensional solutions for the experimental data were used. For the mean log RTs, the parameter values a_k and p_k were also taken from the estimates of the experiment. A *first series* of four simulations used the estimates from the experiment as values for the d_{ij} parameters. (This series is labeled *estimates* in the tables.) The four generated data sets in this series only differed in the distribution of the RTs. A first set was in perfect accordance with the assumptions of the MAXRT model:

the RTs were distributed *lognormally* as described by (8) and (9). Now, if X is normally distributed with mean μ_X and standard deviation σ_X, then $Y = e^X$ is distributed lognormally with

$$\mu_Y = e^{\mu_X + (\sigma_X^2/2)}, \quad \text{and} \quad \sigma_Y^2 = e^{(2\mu_X + \sigma_X^2)}(e^{\sigma_X^2} - 1).$$

In the second set, RTs had the same mean and variance, but were distributed according to the *two-parameter exponential* distribution:

$$p_X(x) = \left(\frac{1}{\delta}\right) \exp\left\{\frac{-(x-\theta)}{\delta}\right\},$$

with $x > \theta$ and $\delta > 0$ (Johnson & Kotz, 1970). In this distribution, $E(X) = \theta + \delta$ and VAR $(X) = \delta^2$. Consequently, δ was set equal to the standard deviation of the RTs in the first set, while θ was set equal to the difference of the mean of the first set and δ.

The RTs in the third set were distributed *normally*. This symmetric distribution does not correspond to the empirical fact that RTs tend to be positively skewed, and these data were included only to study the impact of a relatively severe violation of the lognormal assumption. The means and the standard deviations were based on (8) and (9). The RTs in the fourth set were generated according to a *Pareto* distribution, which is positively skewed, but which can also be extreme because it can have a heavy right tail. The density function is

$$p_X = \frac{ak^a}{x^{a+1}},$$

where $a > 0$, $x \geq k$ and $k > 0$ (Johnson & Kotz, 1970). Because

$$\mu_X = \frac{a^k}{a-1}, \quad \text{and} \quad \mathrm{VAR}_X = \frac{ak^2}{(a-1)^2(a-2)},$$

the same mean and standard deviation can be obtained by making

$$a = 1 + \left(1 + \left(\frac{\mu_X^2}{\mathrm{VAR}_X}\right)\right)^{1/2},$$

and

$$k = \frac{\mathrm{VAR}_X\ (a-1)^2(a-2)}{a}.$$

The distribution functions for these four sets are plotted in Figure 3 for the parameter values of subject LA, first session, stimulus pair 1-2. The figure is highly similar for other subjects, sessions, and stimulus pairs, due to the equating of the means and standard deviations.

For reasons of generalizability, a *second*, *third*, and *fourth* series of four data sets per person were generated. They only differed from the *first series* in the choice of the dissimilarity matrix used to generate the data. In the *second series* (labeled *cube* in the tables), the d_{ij} values were derived from a three-dimensional configuration where eight stimuli were located at the eight vertices of a cube. The dissimilarities for the *third series* (labeled *tetra*) were derived from a tetrahedron configuration, where pairs of stimuli were close to each other on the four corners, thus forming a stimulus config-

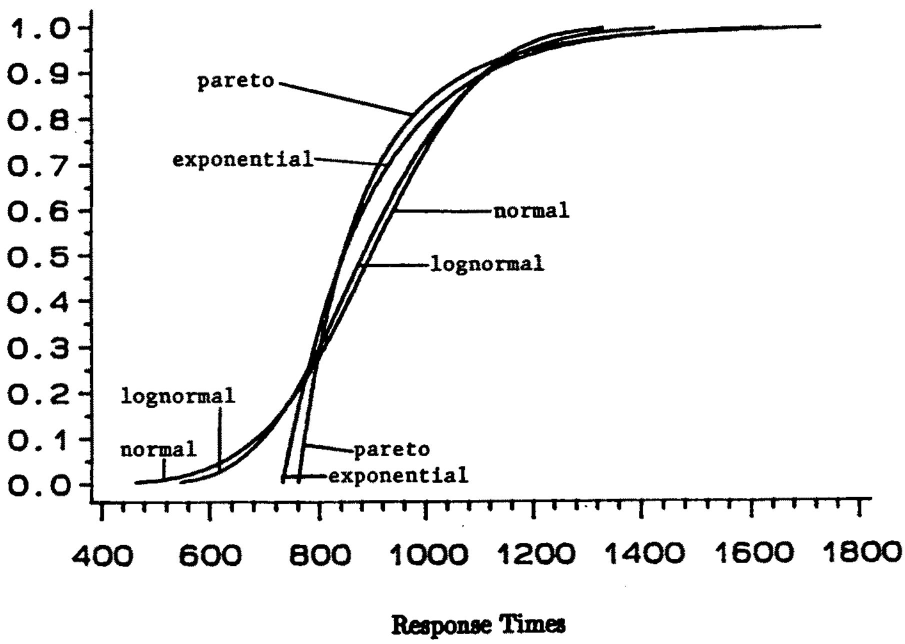

FIGURE 3
The distribution functions used to generate the RTs.

uration with quasi-dense regions. The d_{ij}'s for the *fourth series* (labeled *random*) were derived from a random stimulus configuration, again in a three-dimensional space.

Results

Goodness-of-fit of the RTs. The goodness-of-fit measure of the RTs, as described earlier, was calculated for all the MAXRT analyses on the simulated data sets. Table 1 shows these χ^2 values. The critical value with $a = .05$ is 1387.06. It is obvious that the lognormal RT data provide far better fits than the alternatively distributed RTs. The fits of the exponential and Pareto RTs are fairly similar, while the normal data provide the worst values. However, only four data sets resulted in a rejection of the model at $\alpha = .05$. Summing all the RT data across subjects and underlying stimulus configurations that were generated with the same distribution, only the normally generated data reject the hypothesis of lognormal RTs at $\alpha = .05$.

The log likelihood of the data. In comparing the maximized log-likelihood values for the different simulations, the same pattern was observed as for the goodness-of-fit values for the RTs: lognormal RTs resulted in far superior values, Pareto and exponentially distributed RTs produced mutually comparable values, and normal RTs resulted in the worst likelihood. It is also interesting to compare the maximized likelihood of the empirical data with the values provided by the simulated data of Series 1 (since they have common underlying parameters). For all subjects the likelihood of the empirical data were inbetween the value for the lognormally generated data set and the values for the exponentially and Pareto distributed data.

Goodness-of-recovery. In examining the effect of the assumption of lognormal RTs in MAXRT, the goodness-of-fit is not the only measure of importance. Since MAXRT is a model designed for scaling a fixed set of stimuli, the estimated distances

TABLE 1

Goodness-of-Fit of the RTs and Squared Goodness-of-Recovery Correlations [in Brackets]

Series	Estimates		Cube		Tetra		Random	
TS								
lognormal RT	316.9	[.99]	528.9	[.95]	459.8	[.99]	947.4	[.99]
exponential RT	1309.7	[.99]	1357.7	[.94]	1346.3	[.99]	1348.0	[.97]
normal RT	1358.6	[.96]	1368.6	[.95]	1379.5	[.96]	1380.0	[.97]
pareto RT	1333.8	[.97]	1362.0	[.94]	1463.2*	[.99]	1376.0	[.98]
LA								
lognormal RT	328.9	[.99]	1065.4	[.95]	1039.9	[.99]	1002.7	[.99]
exponential RT	1322.5	[.99]	1319.4	[.94]	1303.4	[.98]	1297.9	[.97]
normal RT	1372.6	[.99]	1344.5	[.98]	1357.7	[.99]	1378.5	[.86]
pareto RT	1332.6	[.99]	1299.0	[.95]	1306.0	[.97]	1287.1	[.99]
AB								
lognormal RT	982.6	[.75]	1290.8	[.94]	559.2	[.95]	363.1	[.87]
exponential RT	1270.3	[.81]	1335.8	[.94]	1275.4	[.87]	1270.7	[.88]
normal RT	1359.5	[.67]	1363.2	[.34]	1391.2*	[.85]	1392.3*	[.64]
pareto RT	1213.0	[.74]	1343.9	[.93]	1225.8	[.87]	1271.4	[.66]
DC								
lognormal RT	550.3	[.96]	515.8	[.95]	321.1	[.97]	560.7	[.98]
exponential RT	1280.3	[.85]	1324.2	[.94]	1297.8	[.97]	1314.4	[.97]
normal RT	1361.2	[.88]	1374.4	[.93]	1354.0	[.94]	1397.1*	[.94]
pareto RT	1236.6	[.96]	1263.0	[.94]	1285.2	[.94]	1318.3	[.97]
JM								
lognormal RT	394.4	[.97]	917.8	[.95]	317.2	[.97]	463.7	[.97]
exponential RT	1329.5	[.91]	1348.5	[.94]	1291.7	[.97]	1332.2	[.97]
normal RT	1378.1	[.94]	1351.6	[.94]	1358.4	[.96]	1364.9	[.94]
pareto RT	1360.5	[.92]	1351.0	[.95]	1323.5	[.91]	1332.8	[.97]
WL								
lognormal RT	357.8	[.98]	1310.4	[.94]	612.8	[.98]	287.6	[.97]
exponential RT	1315.4	[.97]	1337.3	[.94]	1301.1	[.96]	1314.3	[.95]
normal RT	1351.5	[.89]	1334.5	[.94]	1340.6	[.95]	1366.7	[.97]
pareto RT	1288.6	[.94]	1333.9	[.95]	1316.5	[.98]	1294.3	[.96]

* $p < 0.05$

were also examined, leading to the final scaling solutions. The estimated distances were computed under differently assumed RT distributions. In the simulation study, the underlying dissimilarities were known, and consequently, the correlation between the true d_{ij}'s and the $\hat{d}_{ij}$'s estimated from the MAXRT analyses could be calculated. The squared goodness-of-recovery correlations are shown in Table 1. In all the simulations, with the exception of the normally distributed RTs in the second series (cube) for subject AB, unexpectedly high values were obtained, regardless of the underlying d_{ij} matrix. In general, slightly higher values were observed for the lognormal data, and the normal RTs resulted again in little worse; but, surprisingly, the differences are almost negligible.

Likelihood ratio χ^2 and the AIC-statistic. Violations of the assumption concerning the RT distribution should also be evaluated by studying the effect on the χ^2 test with nested models. The data for the simulations were generated, based on a three-dimensional stimulus configuration. When comparing three- and four-dimensional solutions for these data with a χ^2 statistic, the data should not favor the four-dimensional solution. The χ^2 statistic was calculated for all the simulations (2 [log $L(\hat{\omega}_4)$ − log $L(\hat{\omega}_3)$], where log $L(\hat{\omega}_n)$ represents the log likelihood of the n-dimensional solution, with a difference of four in free parameters). Four rejections at the 1%, and five at the 5% significance level are close to the expected number based on chance. The very small number of rejections (five out of 96 hypothesis tests) does not allow an interpretation based on the generating RT distributions. However, since the sum of χ^2 statistics is again χ^2 distributed, all the three- and four-dimensional solutions could be combined (regardless of the subject and the underlying stimulus configuration) for every RT distribution, resulting in a χ^2 test with 96 free parameters (24 times 4). This test incorrectly favors the four-dimensional solution for the exponentially and for the normally distributed RTs at $\alpha = .05$.

The 96 pairs of models were also evaluated using the AIC-statistic (Akaike, 1974):

$$\text{AIC }(\hat{\omega}_n) = -2 \ln L(\hat{\omega}_n) + 2n_\omega,$$

where n_ω is the number of free parameters in the model. The model with the smallest value is favored by the statistic. The n_ω equals 18 and 22 for the three- and four-dimensional solutions, respectively. The AIC-statistic rejected the three-dimensional model in five simulations, exactly the same data where the χ^2 test ($\alpha = .01$) rejected this model.

From MAXRT to MAXSD?

Analyses of the simulated data sets showed that, although the goodness-of-fit measures were much higher for lognormal RT data than for exponential, normal, and Pareto distributed RTs, the differences in goodness-of-recovery were remarkably small. Therefore, it can be concluded that the underlying RT-distribution has little effect on the final maximum likelihood parameter estimates, given that the model assumes a lognormal distribution, and that this distribution is characterized by the mean and the variance. The next question is whether an analysis of same/different judgments alone, without the RT-information, results in a qualitatively inferior scaling solution.

MAXSD

First, the MAXRT model was simplified to MAXSD (*max*imum likelihood algorithm for *s*ame/*d*ifferent data). Keeping (1) to (7), but dropping the modeling of the RTs ((8) and (9)) gave the following density function:

$$g(y_{ijkr}) = Q_{ijk}^{y_{ijkr}}(1 - Q_{ijk})^{(1 - y_{ijkr})}.$$

The log likelihood of the total set of observations then becomes

$$\ln L = \sum_k \sum_{i,j} \{n_{ijk}^{(s)} \ln Q_{ijk} + n_{ijk}^{(d)} \ln (1 - Q_{ijk})\},$$

where $n_{ijk}^{(s)} = \sum_r y_{ijkr}$ and $n_{ijk}^{(d)} = \sum_r (1 - y_{ijkr})$.

The log-likelihood function was maximized using Fisher's scoring algorithm (De Soete & Carroll, 1983; Takane & Sergent, 1983), where estimates of a parameter γ are updated as follows:

$$\gamma^{(t+1)} = \gamma^{(t)} + \lambda^{(t)} H(\gamma^{(t)})^{-1} g(\gamma^{(t)}),$$

where $\lambda^{(t)}$ is the stepsize and $\gamma^{(t)}$ is the parameter value at iteration t, $g(\gamma^{(t)})$ is the *gradient*, and $H(\gamma^{(t)})$ is the *information matrix*.

The elements of the gradient for the parameters d_{ij} are

$$\frac{\partial \ln L}{\partial d_{ij}} = \sum_k \left(\frac{n_{ijk}^{(s)}}{Q_{ijk}} - \frac{n_{ijk}^{(d)}}{1 - Q_{ijk}}\right)\left(\frac{-\phi(v_{ijk})}{\sigma_k d_{ij}}\right).$$

For the thresholds b_k one gets

$$\frac{\partial \ln L}{\partial b_k} = \sum_{i,j} \left(\frac{n_{ijk}^{(s)}}{Q_{ijk}} - \frac{n_{ijk}^{(d)}}{1 - Q_{ijk}}\right)\left(\frac{\phi(v_{ijk})}{\sigma_k b_k}\right),$$

and for the σ_k's, the gradient is

$$\frac{\partial \ln L}{\partial \sigma_k} = \sum_{i,j} \left(\frac{n_{ijk}^{(s)}}{Q_{ijk}} - \frac{n_{ijk}^{(d)}}{1 - Q_{ijk}}\right)\phi(v_{ijk})\left(\frac{\ln d_{ij} - \ln b_k}{\sigma_k^2}\right).$$

In the information matrix, the diagonal values are:

$$-E\left(\frac{\partial^2 \ln L}{\partial d_{ij} \partial d_{ij}}\right) = \sum_k \left(\frac{N_{ijk}}{Q_{ijk}} - \frac{N_{ijk}}{1 - Q_{ijk}}\right)\left(\frac{(\phi(v_{ijk}))^2}{\sigma_k^2 d_{ij}^2}\right),$$

$$-E\left(\frac{\partial^2 \ln L}{\partial b_k \partial b_k}\right) = \sum_{i,j} \left(\frac{N_{ijk}}{Q_{ijk}} - \frac{N_{ijk}}{1 - Q_{ijk}}\right)\left(\frac{\phi(v_{ijk})^2}{\sigma_k^2 b_k^2}\right),$$

$$-E\left(\frac{\partial^2 \ln L}{\partial d_{ij} \partial d_{ij}}\right) = \sum_{i,j} \left(\frac{N_{ijk}}{Q_{ijk}} - \frac{N_{ijk}}{1 - Q_{ijk}}\right)\left(\frac{\phi(v_{ijk})(\ln b_k - \ln d_{ij})}{\sigma_k^2}\right)^2.$$

All other entries are zero, except the following three second-order partial derivatives

$$-E\left(\frac{\partial^2 \ln L}{\partial d_{ij} \partial b_k}\right) = \left(\frac{N_{ijk}}{Q_{ijk}} - \frac{N_{ijk}}{1 - Q_{ijk}}\right)\frac{-(\phi(v_{ijk}))^2}{\sigma_k^2 d_{ij} b_k},$$

$$-E\left(\frac{\partial^2 \ln L}{\partial d_{ij} \partial \sigma_k}\right) = \left(\frac{N_{ijk}}{Q_{ijk}} - \frac{N_{ijk}}{1 - Q_{ijk}}\right)\frac{\phi(v_{ijk})^2(\ln b_k - \ln d_{ij})}{\sigma_k^2 d_{ij}},$$

$$-E\left(\frac{\partial^2 \ln L}{\partial b_k \partial \sigma_k}\right) = \sum_{i,j} \left(\frac{N_{ijk}}{Q_{ijk}} - \frac{N_{ijk}}{1 - Q_{ijk}}\right)\frac{-\phi(v_{ijk})^2(\ln b_k - \ln d_{ij})}{\sigma_k^3 b_k}.$$

MAXSD and the Data of the Experiment

MAXSD was applied to the same/different data of the experiment. The estimated dissimilarity parameters $\hat{d}_{ij}$ from MAXSD, and the $\hat{d}_{ij}$ from MAXRT for two representative subjects are presented in Figure 2b. The correlations between these estimates for every subject were .63, .78, .43, .83, .61, and .86 for TS, LA, AB, DC, JM, and WL,

TABLE 2

Partial and Semi-partial Correlations Between the "True" d_{ij} and the Estimates from MAXRT

	Semi-Partial	Partial
TS	.78	.91
LA	.62	.99
AB	.57	.94
DC	.65	.79
JM	.57	.88
WL	.57	.93

respectively. As can be seen, a substantial part of the variance of the $\hat{d}_{ij}$'s from MAXRT can be accounted for by the estimates from MAXSD, but the correlation is far from perfect, indicating that the RTs did have a substantial influence on the estimates.

MAXSD and the Data from the Simulation Study

To arrive at a clearer view of the quality of the $\hat{d}_{ij}$ estimates from MAXSD, this model was applied to the data from the simulations. In Figure 2c the true d_{ij} for two representative subjects are plotted with the estimates from MAXRT and MAXSD. The MAXSD estimates are more scattered, while the estimates from MAXRT show a clear linear relation. The squared goodness-of-recovery correlations based on the MAXSD model were .69, .99, .14, .62, .70, and .52 for TS, LA, AB, DC, JM, and WL, respectively. Comparing these correlations with the recovery indices of MAXRT (Table 1), it becomes clear that MAXSD does a lot worse for five out of six subjects. Still the recovery in MAXSD is quite high, taking into consideration the considerable number of ties in choice proportions on which the scaling was based. Partial and semi-partial correlations between the true $\hat{d}_{ij}$ and the MAXRT estimates, with the effect of the MAXSD estimates partialled out, are given in the first and second column of Table 2. Both correlations confirm the additional information of the RTs in the MAXRT model.

Conclusion

Detailed analysis of the data from the experiment showed that the MAXRT model for response times and same/different judgments provides widely varying goodness-of-fit measures for the data of different subjects. The Shapiro-Wilk test suggested that the RT distribution might be task dependent. On the basis of an elaborate simulation study, it was possible to show that the underlying RT distribution had a clear effect on the goodness-of-fit of the RTs and on the χ^2 test in comparing nested models, but it had little impact on the final distance parameters as estimated by the model. Despite this robustness of the distance parameters under differently distributed RTs, a comparison of MAXRT with a similar model, based only on the same/different judgments (and not on the RTs), proved that RTs do give some additional information to the judgments in a two-choice task.

References

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

Anderson, J. R., & Bower, G. H. (1973). *Human associative memory*. Washington: Winston & Sons.

Behrman, B. W., & Brown, D. R. (1968). Multidimensional scaling of form: A psychophysical analysis. *Perception & Psychophysics*, *4*, 19–25.

Brown, D. R., & Andrews, M. H. (1968). Visual form discrimination: Multidimensional analysis. *Perception & Psychophysics*, *3*, 401–406.

Chocholle, R. (1940). Variations des temps de réaction auditifs en fonction de l'intensité à diverses fréquences [Variations of auditory reaction times as a function of intensity at varying frequences]. *L'Année Psychologique*, *41*, 65–124.

D'Agostino, R., & Stephens, M. A. (1986). *Goodness-of-fit techniques*. New-York: Dekker.

De Soete, G. (1986). A perceptual study of the Flury-Riedwyl faces for graphically displaying multivariate data. *International Journal for Man-Machine Studies*, *25*, 549–555.

De Soete, G., & Carroll, J. D. (1983). A maximum likelihood method for fitting the wandering vector model. *Psychometrika*, *48*, 553–566.

Flury, B., & Riedwyl, H. (1981). Graphical representation of multivariate data by means of asymmetrical faces. *Journal of the American Statistical Association*, *76*, 757–765.

Flury, B., & Riedwyl, H. (1983). *Some applications of asymmetrical faces*. Unpublished manuscript, Institut für Mathematische Statistik und Versicherungslehre, Universitaet Bern.

Green, D. M., & Luce, R. D. (1967). Detection of auditory signals presented at random times. *Perception & Psychophysics*, *2*, 441–449.

Green, D. M., & Luce, R. D. (1971). Detection of auditory signals presented at random times III. *Perception & Psychophysics*, *9*, 257–268.

Harrington, L. F., & Brown, D. R. (1972). Visual pattern perception: Multidimensional analysis of discrimination difficulty for children. *Psychonomic Science*, *29*, 219–221.

Ida, M. (1980). The application of the Weibull distribution to the analysis of the reaction time data. *Japanese Psychological Review*, *22*, 207–212.

Johnson, N. L., & Kotz, S. (1970). *Distributions in statistics: Continuous univariate distributions-2*. Boston: Houghton Miffin.

Kak, A. V., & Brown, D. R. (1979). Visual pattern perception: Multidimensional analysis of development of children's reading skills. *Perceptual and Motor Skills*, *49*, 819–830.

Kirk, R. E. (1982). *Experimental design. Procedures for the behavioral sciences* (2nd ed.). Belmont, CA: Brooks/Cole.

Kohfeld, D. L., Santee, J. L., & Wallace, N. D. (1981). Loudness and reaction time: II. Identification of detection components at different intensities and frequencies. *Perception & Psychophysics*, *2*, 550–562.

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

Kruskal, J. B. (1964b). Multidimensional scaling: A numerical method. *Psychometrika*, *29*, 115–129.

Luce, R. D. (1986). *Response times. Their role in inferring elementary mental organization*. New York: Oxford University Press.

Luce, R. D., & Green, D. M. (1970). Detection of auditory signals presented at random times II. *Perception & Psychophysics*, *7*, 1–14.

McGill, W. J., & Gibbon, J. (1965). The general gamma distribution and reaction times. *Journal of Mathematical Psychology*, *2*, 1–18.

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

Melara, R. D. (1989). Similarity relations among synesthetic stimuli and their attributes. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 212–231.

Murdock, B. B. (1985). An analysis of the strength-latency relationship. *Memory and Cognition*, *13*, 511–521.

Pearson, E. S., D'Agostino, R. B., & Bowman, K. O. (1977). Tests for departure from normality: Comparison of powers. *Biometrika*, *64*, 231–246.

Podgorny, P., & Garner, W. R. (1979). Reaction time as a measure of inter and intra-object visual similarity. *Perception & Psychophysics*, *26*, 37–52.

Ratcliff, R. (1978). Theory of memory retrieval. *Psychological Review*, *85*, 59–108.

Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review*, *86*, 190–214.

Royston, J. P. (1982). An extension of Shapiro and Wilk's *W* test for normality to large samples. *Applied Statistics*, *31*, 115–124.

Scheiblechner, H. (1979). Specifically objective stochastic latency mechanisms. *Journal of Mathematical Psychology*, *19*, 18–38.

Sergent, J., & Takane, Y. (1987). Structures in two-choice reaction-time data. *Journal of Experimental Psychology: Human Perception and Performance*, *13*, 300–315.

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*, 591–611.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420–428.

Snodgrass, J. G., Luce, R. D., & Galanter, E. (1967). Some experiments on simple and choice reaction times. *Journal of Experimental Psychology*, *75*, 1–17.

Takane, Y., & Sergent, J. (1983). Multidimensional scaling models for reaction times and same-different judgments. *Psychometrika*, *48*, 393–423.

Townsend, J. T., & Ashby, F. G. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge: University Press.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.

Young, F. W. (1970). Nonmetric scaling for line length using latencies, similarity, and same-different judgments. *Perception & Psychophysics*, *8*, 363–369.

Manuscript received 10/29/91
Final version received 2/14/92