Journal of Classification 21:185-206 (2004)
DOI: 10.1007/s00357-004-0016-y

Percept Variance, Subadditivity and the Metric Classification of Similarity, and Dissimilarity Data

David B. MacKay

Kelley School of Business, Indiana University

Bryan Lilly

College of Business Administration, University of Wisconsin Oshkosh

Abstract: Percept variance is shown to change the additive property of city-block distances and make city-block distances more subadditive than Euclidean distances. Failure to account for percept variance will result in the misclassification of city-block data as Euclidean. A maximum likelihood estimation procedure is proposed for the multidimensional scaling of similarity data characterized by percept variance. Monte Carlo and empirical experiments are used to evaluate the proposed approach.

Keywords: City-block metric; Euclidean metric; Subadditivity, Multidimensional scaling; Probabilistic scaling; Thurstonian modeling; PROSCAL.

1. Introduction

There is a long history, not all of it successful, in using measures of fit to indicate the metric that underlies similarity and dissimilarity data. Minkowski power metric or L_p norm models have been studied most often with the greatest attention being paid to the Euclidean ($p = 2$) metric

---

This research is based in part on work supported by a research grant from the Kelley School of Business at Indiana University.

Author's Address: David B. MacKay, Marketing Department, Kelley School of Business, Indiana University, Bloomington, IN 47405, USA, tel: (812) 855-1009, e-mail: mackay@indiana.edu

$$d_{ij} = \left(\sum_{k=1}^{r} \left(x_{ik} - x_{jk} \right)^2 \right)^{1/2} \tag{1}$$

and the city-block ($p = 1$) metric

$$d_{ij} = \sum_{k=1}^{r} \left| x_{ik} - x_{jk} \right| \quad , \tag{2}$$

where x_{ik} is the coordinate of object i on dimension k. Recent reviews of the literature have been provided by Borg and Groenen (1997), Brusco (2001) and MacKay (2001). Arabie's (1991) paper summarizes much of the earlier city-block literature.

There is a substantial body of experimental evidence for the use of a city-block metric in situations where psychological theory predicts its presence. Percepts described by separable (unrelated) dimensions, for example, are commonly thought to be more appropriate for city-block metrics (*viz.* Garner 1974; Shepard 1991). At the same time, there are several experimental studies that have demonstrated situations where data expected to be described by a city-block metric have been better described by a Euclidean metric (Dunn 1983; Glazer and Nakamoto 1991; Lee and Pope 2003; Melara 1989; Nosofsky 1985; Nosofsky 1986).

Two reasons seem to dominate the discussion of why city-block spaces may be misclassified. The first is the presence of noise (Lee and Pope 2003; Shepard 1986), low discriminability (Nosofsky 1985; Tversky and Gati 1982), instability (Eisler and Knöppel 1970), inter-subject variability (Dunn 1983; Schönemann 1994) and percepts that are not "clumped" in a set (Glazer and Nakamoto 1991) – conditions consistent with high variability data. (The absence of "clumping" is consistent with high variability data because variance, which shares the distance properties of symmetry and non-negativity, will cause some algorithms to spread out the estimated locations of the percepts.) The second is the occurrence of local optima, caused in part by the use of gradient based algorithms or poor initial configurations, which prevent algorithms from consistently achieving optimal results with the city-block metric (*viz.* Arabie 1973; Brusco 2001; Hubert, Arabie and Hesson-Mcinnis 1992; Shepard 1974).

A third reason, less frequently mentioned, is that there are important percept properties not accounted for in our metric space modeling (Ashby and Lee 1993). Failure to account for these properties may result in misidentifying the most appropriate metric. Tversky and his colleagues have argued that unaccounted for percept properties are sufficient cause for abandoning distance models (Tversky 1977; Tversky and Gati 1982). Others (Dunn 1983; Ennis 1988; Shepard 1986) have suggested that unaccounted for percept properties which are potentially compatible with distance models may be responsible for the disconcerting success of the Euclidean metric in describing distance relationships. Percept variance, first introduced by Thurstone's 1927 publications on discriminal processes, is an example of this kind of property. The discriminal process suggestion is appealing because it is consistent with many of the "high variability" explanations that have been offered. The discriminal process approach is also, as Takane (1978) notes, "intuitively more attractive" than a lot of the other suggestions that have been hypothesized as means for introducing variance into distance models.

Arguments for discriminal processes causing the misclassification of city-block processes as Euclidean are largely conjectural. Ennis (1988) looks at the issue in most depth and shows that data which are city-block on a trial by trial basis can, in the presence of perceptual variance, appear to be Euclidean when their expected values are investigated. This is a valuable observation but no link is made to the axiomatic foundations of Minkowski spaces and the reasons for the observation remain unexplored.

We now look at the issue of percept variance in more depth and show that percept variance can, by changing the additivity properties of city-block and Euclidean data, make city-block data appear to be more Euclidean than Euclidean data. As a result, once fairly modest levels of variation occur, city-block data will, when used with deterministic models, be classified as Euclidean.

To correctly identify city-block data, operational models are needed for evaluating city-block and Euclidean data in the presence of percept variation. We discuss a number of multidimensional Thurstonian models that have been proposed for the analysis of similarity and dissimilarity data in city-block and Euclidean metrics. A simple process is proposed for extending one class of dissimilarity data models to similarity data. Issues of operationalization are discussed. Experimental results, empirical and Monte Carlo, are presented that contrast the success of the probabilistic Thurstonian models in identifying the city-block metric with the success of conventional non-metric models that are deterministic.

2. Additivity and Probabilistic Distances

Numerous authors have discussed the properties of Minkowski distances. A well written summarization is provided by Borg and Gr oenen (1997). A key property is additivity. For metrics, the additivity function

$$a(x, y, z) = d(x,y) + d(y,z) - d(x,z) \tag{3}$$

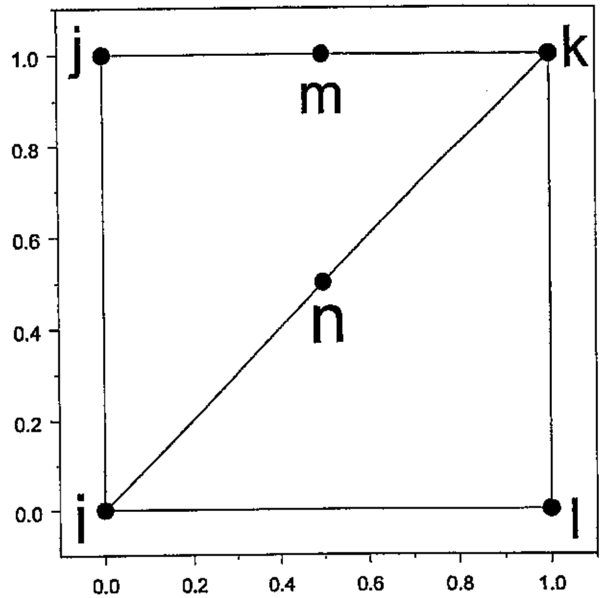

Figure 1. Unit Square

of distances d connecting points x, y, and z is said to be additive if $a()$ is zero, subadditive if $a()$ is positive and superadditive if $a()$ is negative. To illustrate the relation of additivity to the choice of a city-block or Euclidean metric, consider the unit square of Figure 1 with one midpoint along the line connecting points j and k and another midpoint along the line connecting points i and k.

Minkowski distances assume "segmental additivity." In Figure 1, this means that $a(j, m, k)$ and $a(i, n, k)$ are zero. More interesting, for the purpose of metric identification, is a corner triple function such as $a(i, j, k)$ which is additive for the city-block metric and subadditive for the Euclidean metric. This additive/subadditive relationship of city-block and Euclidean metrics does not, of course, hold only for corner triples. The value of $a()$ for city-block distances should never be greater than it is for Euclidean distances. City-block and Euclidean metrics should have the same value of $a()$ only when x, y, and z are on a straight line.

If discriminal processes are indeed present, what are the implications for additivity? The essence of discriminal process models is the assumption that the points between which distances are defined are not fixed (deterministic) but normally distributed. As a result, the distances are no longer fixed but probabilistic. Instead of talking about the distance between two points we need to talk about sample distances (momentary psychological values), distance distributions and expected distances.

Expected distances $E(d_{ij} / \boldsymbol{\mu}, \mathbf{S}, r, p)$ are a function of the centroids $\boldsymbol{\mu}$, variances $\mathbf{S}$, dimensionality r and metric p of the space. Let's assume a simple variance structure in which the off-diagonal values of $\mathbf{S}$ are zero and the diagonal values are all the same. Substituting expected distances for deterministic distances in [3], the additivity of city-block and Euclidean metrics for the corner triple of Figure 1 can be calculated over increasing magnitudes of s_{jj}; $j = 1,2$. (Expected distance formulas are given in the Appendix.) Looking at the results shown in Figure 2, we see that when $s_{jj} = 0$, we have the anticipated deterministic results of $a() = 0$ for the city-block metric and $a() = 2 - \sqrt{2}$ for the Euclidean metric. By the time the variance has increased to only 0.1, we find that the subadditivity of the city-block metric now *exceeds* that of the Euclidean metric. For deterministic models that do not recognize percept variation, the city-block data will now appear to be "more Euclidean" than the Euclidean data.

Figure 2 leads us to expect that in the presence of discriminal processes, traditional multidimensional scaling (MDS) algorithms will correctly identify Euclidean data as Euclidean and incorrectly identify city-block data as Euclidean. These expectations will be investigated in Section 4. Similar expectations were reported by Lee and Pope (2003) in a study of how individual differences affect averaged distance matrices.

3. Multidimensional Thurstonian Models of Dissimilarity and Similarity Judgments

Thurstonian models with multidimensional discriminal processes have been modeled in a number of ways for a variety of data types. Here, our concern is with similarity and dissimilarity data types. The first use of a Thurstonian MDS program for dissimilarity data was that reported by MacKay and Zinnes (1981), Zinnes and MacKay (1981). They proposed a maximum likelihood (ML) model for Euclidean distances in which the variances of the percepts could differ from percept to percept but were identical on all dimensions. Their work drew heavily from earlier studies by Hefner (1958) and Zinnes and Wolff (1977) on Thurstonian models for same-different judgments in which the probability density functions (PDFs) of distances were modeled as functions of the non-central chi-square distribution. Their algorithm, not described until Zinnes and MacKay (1983), used least squares procedures to obtain initial values and a quasi-Newton method to find final values. An alternating ML procedure was used in which variances $\mathbf{S}$ and centroids $\boldsymbol{\mu}$ were sequentially reestimated.

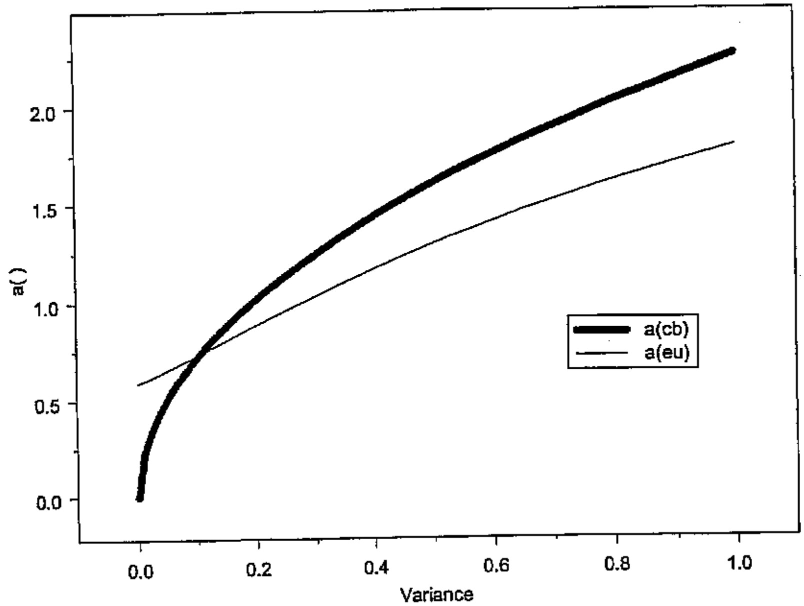

Figure 2. Corner triple subadditivity for the Euclidean and city-block metrics.

Ennis, Palen and Mullen (1988) proposed a least squares (LS) Thurstonian model for similarity data that used a general similarity function of the form $s(d) = exp(-d^a)$, $a > 0$ to relate distances to similarities. When squared distances were used, $s(d) = exp(-d^2)$, a closed form solution for the expected similarities was found. Numerical integration was suggested for other situations. An attractive feature of their implementation was its ability to handle non-zero covariances. A modified Levenberg-Marquardt (steepest descent) algorithm was used to minimize the sum of squared differences between the similarity data and the expected similarities.

Ennis and Johnson (1993) provided a theoretical basis for extending the Ennis et al. (1988) model by deriving the expected similarities for city-block distances. The extension is based upon the recognition that the PDFs of city-block distances follow a folded normal distribution.

MacKay (1989) extended the ML model by using quadratic forms in normal variables distributions to accommodate anisotropic space variances that differed from dimension to dimension and had non-zero valued covariances. The earlier gradient based procedure was dropped in favor of a direct search method. MacKay (2001) developed exact and approximate PDFs for the city-block metric which permitted their use in a ML framework. (ML equations are in the Appendix.) Linear-exponential response models were also added to extend the ways percept dissimilarity responses could be related to the psychological distance estimates of the ML model.

Response models also provide a simple approach for extending ML models designed for dissimilarity judgments to similarities. Two functions, the exponential decay function $s = g(d) = exp(-d)$ and the Gaussian function $s = g(d) = exp(-d^2)$ dominate the literature as means for relating percept similarity ($0 < s = 1$) to psychological distance (Shepard 1974; Nosofsky 1985). The exponential function is usually associated with city-block distances and the Gaussian function is usually associated with Euclidean distances. However, some authors, such as Shepard (1987), have proposed that the exponential decay function is universal.

To transform the $f(d \,/\, \boldsymbol{\mu}, \mathbf{S}, r, p)$ PDF for dissimilarities to the $f(s \,/\, \boldsymbol{\mu}, \mathbf{S}, r, p)$ PDF for similarities we use the function

$$f(s \,/\, \boldsymbol{\mu}, \mathbf{S}, r, p) = \left| \frac{\partial}{\partial s} g^{-1}(s) \right| f\left(g^{-1}(s) \mid \boldsymbol{\mu}, \mathbf{S}, r, p\right) \qquad (4)$$

which, for the exponential decay relationship gives us

$$f\left(s \mid \boldsymbol{\mu}, \mathbf{S}, r, p\right) = \left| 1/s \right| f\left(-\ln(s) \mid \boldsymbol{\mu}, \mathbf{S}, r, p\right) \qquad (5)$$

and, for the Gaussian relationship gives us

$$f\left(s \mid \boldsymbol{\mu}, \mathbf{S}, r, p\right) = \left| 1/\left(2s\sqrt{-\ln s}\right) \right| f\left(\sqrt{-\ln s} \mid \boldsymbol{\mu}, \mathbf{S}, r, p\right). \qquad (6)$$

This permits the use of similarities in a ML framework where the objective is to maximize $\ln L = \sum_s \ln f(s \mid \boldsymbol{\mu}, \mathbf{S}, r, p)$.

4. Comparative Results of Probabilistic and Deterministic Models

Three experimental comparisons were made to assess the metric predictions of probabilistic Thurstonian models and deterministic nonmetric models. The first comparison is a Monte Carlo study that assesses the abilities of the different models to correctly determine a Euclidean or city-block metric when using similarity data characterized by different discriminal dispersion magnitudes. The second comparison, also a Monte Carlo study, looks at metric recovery for discretely valued dissimilarity data from three percept sets of different sizes. The third comparison is an empirical comparison involving incomplete data for multiple subjects.

Two popular deterministic nonmetric methods, KYST (Kruskal, Young, and Seery 1977) and SYSTAT (www.systat.com) were compared on error free simulated data sets in Euclidean and city-block metrics. The programs gave nearly identical results when a Euclidean metric was used but not when a city-block metric was used. KYST had lower stress solutions about seventy percent of the time and, as a result, KYST was chosen for the remaining analyses. For a discussion of more recent, but less available, city-block scaling methods see Borg and Groenen (1997).

PROSCAL (www.proscal.com) was used for the probabilistic analyses. Selection of PROSCAL over the programs of Ennis et al. was due to a preference for the consistency of ML estimates in situations where the error distribution is not multivariate normal.

Due to the well known local optima difficulties associated with nonmetric multidimensional algorithms (Arabie 1973; Borg and Groenen 1997), 100 random starting configurations were used for all of the following KYST analyses. To avoid premature termination, the following termination criteria of KYST - STRMIN, SRATST and ITERATIONS - were set to 0.0000001, 0.9999999 and 200. While local optima are possible with PROSCAL, they appear to be less frequent - due we suspect to the use of alternating maximum likelihood estimation and the avoidance of a gradient based optimization algorithm. Therefore, multiple random starting configurations were not used with PROSCAL. A single least squares based initial solution was used instead.

4.1 Similarities Comparison

Patterned after Shepard's (1962) introductory publication on nonmetric multidimensional scaling, coordinates were randomly generated and analyzed in a two dimensional space and adjusted so that the mean interpoint Euclidean distance was unity. These coordinates became the parametric centroids of the Thurstonian model that was used to generate the sample data. (Shepard used a single configuration from a prior study but we wanted to evaluate multiple configurations and drew coordinates from a uniform distribution instead. Another departure from Shepard's study was the use of 24 random points instead of 15. Sample size effects are looked at in the next comparison.)

To evaluate the effect of subadditivity on metric estimation, a simple Thurstonian model was used in which the percept variances were the same for all percepts on all dimensions. Five levels of variance were investigated: 0.0001, 0.001, 0.01, 0.1 and 1.0. (Not being concerned with Thurstonian models, Shepard assumed that all percepts had zero variance.) For each level of variance, coordinates were independently sampled on a trial-by-trial basis for each judgment from the parametric configurations. Euclidean and city-block distances were then computed from the sampled coordinates and transformed into similarities by using the Gaussian distribution for Euclidean distances and the exponential distribution for city-block distances. The process was repeated 100 times, providing 200 sets of similarity data for each variance condition. (This trial-by-trial sampling process should not be confused with the procedures of related studies – Ashby, Maddox and Lee (1994), Lee and Pope (2003) – where object coordinates are sampled once for each subject and then applied to all within subject trials without resampling.)

Each set of similarities was evaluated using a Euclidean and city-block metric by KYST and PROSCAL. One hundred random starts were used for each KYST analysis. The random start solution with the lowest stress was chosen. (Following all of the prior studies we have found on metric selection with similarity/dissimilarity data, "Stress1" was used as our measure of fit. A brief investigation of two, three, four and five parameter polynomial functions showed that Stress1 performed better at selecting the correct metric. Stress2 was, on the average, about the same as Stress1 but Stress1 was more consistent over the different variance conditions.)

To determine if subadditivity, which increases as percept variance increases, is associated with city-block data being misclassified as Euclidean, the percentage of city-block data sets having lower stress when evaluated with the city-block metric was calculated. The same was done for the Euclidean data. Results are presented in Figure 3. For the two lowest variance values, KYST does very well at estimating the city-block metric – 98 and 99% correct. As the variance magnitude (and subadditivity) increases, the hit rate rapidly declines. For the two highest variances, all of the city-block data sets are classified as Euclidean. As expected from the discussion of Section 2, the Euclidean data are always classified as Euclidean, even when the variance levels are high.

Misclassification due to the subadditivity that results from percept variance should go away if percept variance is accounted for in the scaling of the similarity data. To compare Euclidean and city-block solutions using the Thurstonian PROSCAL model, use was made of Bozdogan's (1987) CAIC criterion.

The CAIC criterion is but one of many approaches now being used to account for complexity when selecting a model. (See the March, 2000 issue of the *Journal of Mathematical Psychology* for a review of contemporary model selection approaches.) CAIC penalizes likelihoods by the number of parameters estimated in the model and takes the form

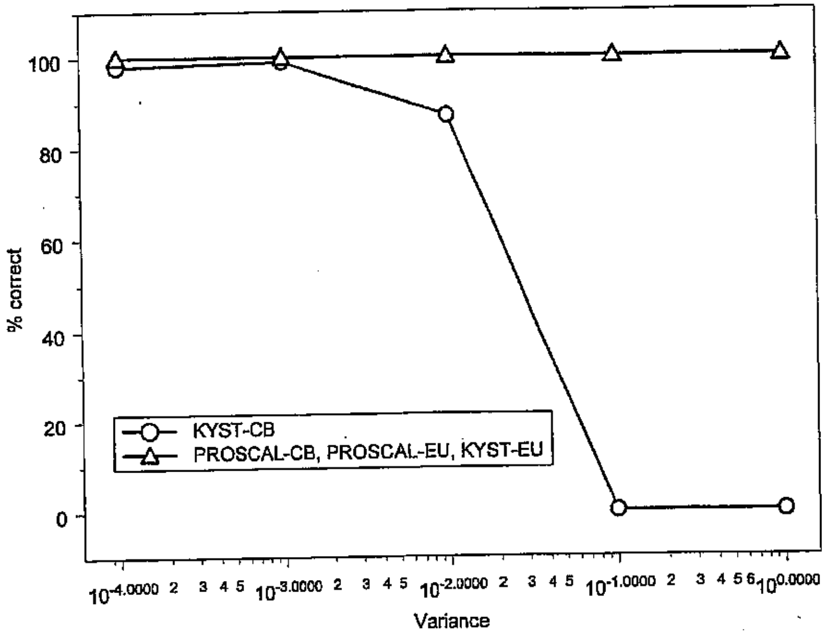

Figure 3. Metric classifications for similarity data.

$$CAIC = -2 \ln L + cK$$

where L is the likelihood, K is equal to the number of independently adjusted parameters, and c is the cost of adding a parameter to the model. The number of independently adjusted parameters K for the isotropic Euclidean space is

$$K = m + q + p - r - \frac{r(r-1)}{2} - 1 \qquad (7)$$

for m coordinates, q unique variances, p response model coefficients and r dimensions. The last three terms are subtracted for the centering, rotation and scale invariance of the solution. The rotational invariance term is omitted when the city-block metric is used and when an anisotropic Euclidean space is used. (The directionality of anisotropic space solutions fixes the orientation of the solution.) For the cost, $c = ln(S) + 1$, where S is the sample size. The metric with the lower criterion value is selected.

Figure 3 shows that PROSCAL had 100% correct metric classification for both metrics over all variance levels. These results suggest that the subadditivity "reversal" caused by the presence of percept variance was not a cause of classification error when a probabilistic model was used. The results do not indicate that one will always identify the correct metric when using a probabilistic model, even if the underlying processes are Thurstonian. The following comparisons are designed to give some guidance as to what might be expected under different conditions.

4.2 Discretely Valued Dissimilarities Comparison

The first comparison answered the initial question about the effect of subadditivity on metric classification but the success it demonstrated for the probabilistic model may not be enjoyed under less favorable circumstances. In this study, dissimilarities were used instead of similarities. Dissimilarities do not have the first comparison's diagnostic advantage of different response functions being associated with city-block and Euclidean metrics. To investigate the robustness of the proposed ML model we deliberately misspecified the distributional form in this comparison by transforming the continuous dissimilarities to a discrete nine point scale. The nine point transformation was made by dividing the range of each simulated subject's responses into nine equally spaced intervals. The continuous dissimilarities were generated in the manner described in Section 4.1.

Three levels for the number of percepts were used in this study – 8, 16 and 24. Percept variance was held constant at 0.001 (the highest value for which the nonmetric models did well in the previous comparison) for all percepts on all dimensions. As before, the data were generated and analyzed in a two-dimensional space. Results for the city-block metric are shown in Figure 4.

As expected from Section 4.1, with the low variance level, both the deterministic and probabilistic models did a good job of predicting the city-block metric when the number of percepts was 24. The success rate held up for the probabilistic method when the number of percepts dropped to 16 but the nonmetric method's success rate dropped drastically. Both methods dropped when only eight percepts were present. The inability to make a successful classification for the nonmetric procedure with eight percepts is due in large part to both metrics achieving a stress of zero. This was not the case with 16 percepts. Tied criteria were not an issue with eight percepts for the probabilistic model

Classification results with the Euclidean metric data are shown in Figure 5. As the number of percepts diminishes, the 100% correct classification results of Section 4.1 diminish as well. However, the results of the probabilistic model hold up much better than the results of the deterministic model.

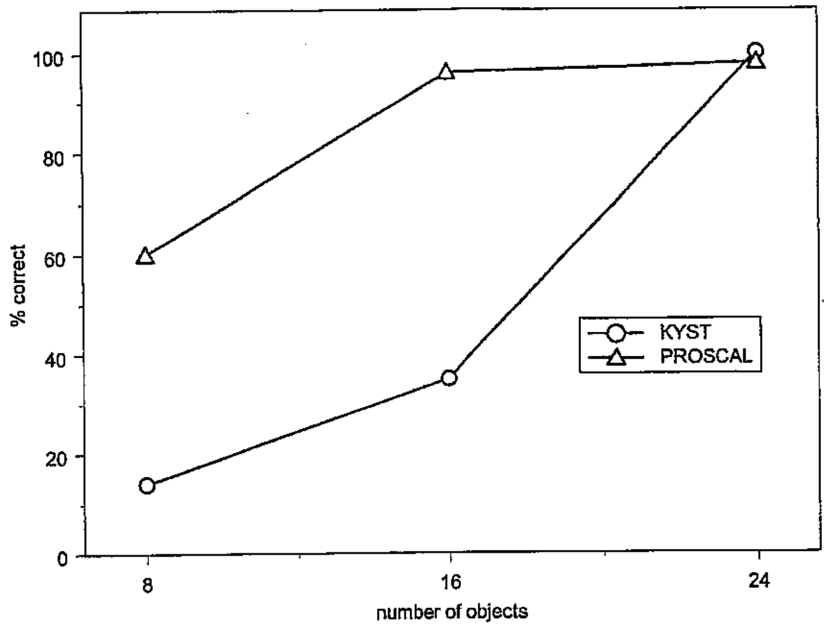

Figure 4. City-block classification for simulated integer valued city-block dissimilarities.

4.3 Empirical Comparison

The literature on integral and separable dimensions, briefly referred to in Section 1, states that "for stimuli chosen to vary along dimensions that seemed especially 'obvious', 'compelling,' 'analyzable' or … 'separable,'" (Shepard 1991) a city-block metric is more consistent with the data than a Euclidean metric. When separate dimensions are not obvious, then "the subject might be more likely to judge the over-all difference directly (and)… the Euclidean model (would) … show more promise" (Torgerson 1958). The association of a city-block metric with 'obvious' dimensions goes back to Attneave (1950).

Different literatures have exploited the separable/city-block and integral/Euclidean distinction. Within the psychophysical literature, the experimental emphasis has been on stimuli that differ with respect to shape or color. The decision making literature has investigated the association of city-block/Euclidean metrics with analytic/non-analytic information processing (Alba and Hutchinson 1987; Baumgartner 1993; Glazer and Nakamoto 1991). Factors that contribute to analytic processing are thought

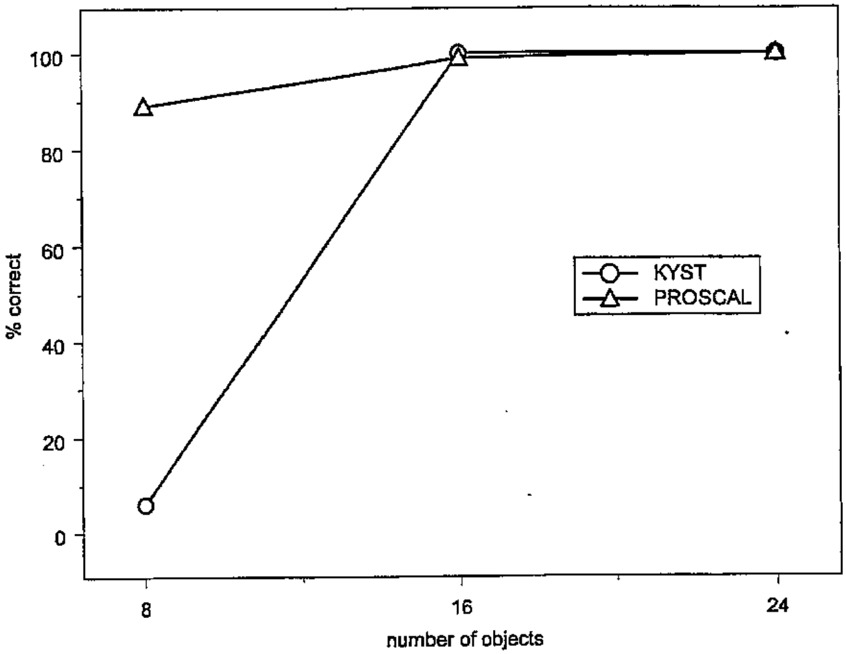

Figure 5. Euclidean classification for simulated integer valued Euclidean dissimilarities

to include stimulus simplicity, motivation, lack of time pressure, subject expertise and previous analytic processing of the same attribute information.

A decision making context was chosen for the design of this experiment in which the stimuli were twelve paper stocks used in the publication of resumes. The paper stocks differed in two respects that are expected to be separable – weight (three levels: 0.126, 0.220 and 0.374 oz./sheet) and color (four shades of white to medium gray: 0%, 10%, 20% and 30% tint). Thirty-two student subjects participated in the study. Subjects were told the study was about how experts evaluated products. A $50 prize for the best performance was provided to enhance motivation. At the beginning of the study, the subjects read a short essay on expert decision making. The essay stressed the characteristics of accuracy and consistency. The subjects were then quizzed on their understanding of the essay. The next task introduced the subjects to the paper stocks. Subjects were told that the paper stocks differed in weight and color and were given ten minutes to examine them. After examining the paper stocks, dissimilarity judgments on a 21 point scale were obtained for 45 of the possible 66 pairs of judgments. An incomplete design was used to reduce fatigue. To reduce time pressure, subjects were given 30 minutes to complete this task. Ross' (1934) method of ordering paired comparisons was used for ordering the stimuli and a cyclic design (Borg and Groenen, 1997) was used for determining the 45 judged pairs.

The sharp decline in classification success of the previous comparison when going from 16 to 8 percepts and the use of incomplete data in this experiment led to the decision to undertake a group, as opposed to individual, analysis. The nonmetric analysis was again based upon 100 random initial configurations. Values of Stress1 for the city-block and Euclidean metrics were 0.292 and 0.246, indicating a Euclidean metric and integral dimensions. The high stress values are indications of the difficulty deterministic models have dealing with probabilistic data.

For the probabilistic analyses, city-block analyses were conducted for the four combinations of a Case V/Case III variance structure (where the variances are the same/different for all percepts) and an isotropic/anisotropic variance structure (where the variances are the same/different for all dimensions). Using the CAIC criterion, Case V solutions were preferred to Case III solutions and isotropic solutions were preferred to anisotropic solutions. The Case V isotropic city-block solution, which had the lowest CAIC criterion value, is shown in Figure 6a. The circle about percept H indicates the estimated magnitudes of the percepts' standard deviations $(\hat{\sigma} = 0.184)$. The CAIC criterion was 928 for the city-block solution and 986 for the corresponding Euclidean solution, indicating a city-block metric and separable dimensions.

Since the solution in Figure 6a departs slightly from the lattice structure of the physical stimuli, constrained city-block and Euclidean solutions were estimated in which percepts with the same color were constrained to have the same mean and percepts with the same weight were constrained to have the same mean. The constrained solutions estimate only seven, instead of 24 coordinates. CAIC values improved for both metrics, 919 for the city-block metric and 960 for the Euclidean metric. The constrained cty-block solution, with a slightly higher standard deviation$(\hat{\sigma} = 0.195)$, is shown in Figure 6b. Both solutions, being city-block, correctly aligned their dimensions in accord with the color and weight of the stimuli.

Since the data for this comparison are incomplete and the percept variance is high, a Monte Carlo analysis was conducted to get some idea of the power for tests of a city-block vs. a Euclidean metric. Treating the estimates of Figure 6b as parameters, incomplete data sets of dissimilarity judgments were generated for three levels of replication: 8, 16, and 32

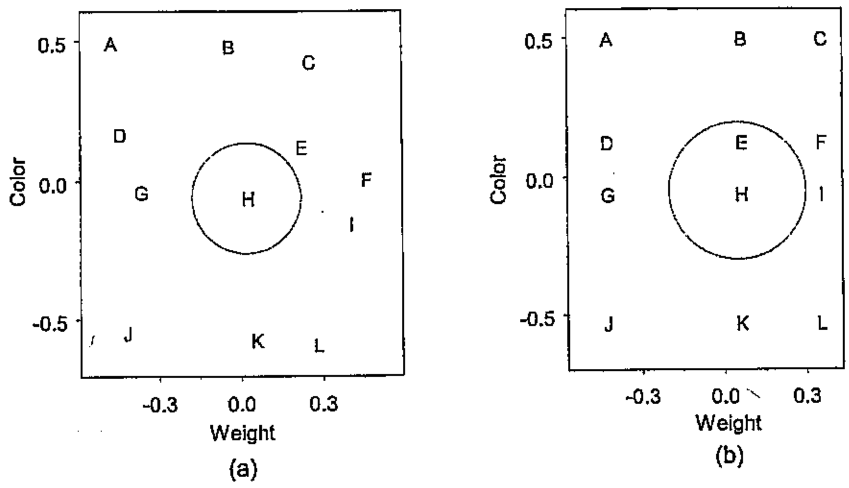

Figure 6. Unconstrained and constrained city-block configurations for the paper study.

replicates. One hundred samples were generated for each replication level. Using the CAIC criterion, the null hypothesis of a Euclidean metric was tested for each of the three conditions.

For 32 replicates, the actual number in the study, 89 percent of the simulated data sets were correctly classified as city-block. This number rapidly diminished as the number of replicates went down. The corresponding percents for 16 and 8 replicates were 64 and 31 percent. In limited-data high-variance situations, metric classification is very dependent upon the criterion used. Using an AIC criterion (Akaike 1974), which takes the form

$$AIC = -2\ln L + 2K$$

the correct percentage classifications rise to 92, 84 and 66 percent. This change is due to the less stringent penalty function built into the AIC criterion. Arguments for preferring CAIC, which is very similar to Schwarz' (1978) Bayesian criterion (BIC), have been presented by Bozdogan (1987).

This power analysis is designed to assist in the evaluation of the preceding empirical comparison and care should be taken in generalizing the results to other situations. Many factors influence the magnitudes of the percept variances. The values of the preceding study may vary significantly with different data types, instructions and experimental conditions. Different methodologies and different stimulus domains also bring their own sources of misspecification error.

5. Discussion

It has been shown that percept variation will result in expected distances that are more subadditive for the city-block metric than the Euclidean metric – a reversal of the deterministic expectation that the city-block metric is additive and the Euclidean metric is subadditive. City-block data thus appear to be more Euclidean than Euclidean data and deterministic nonmetric algorithms will mistakenly classify city-block data as Euclidean. Probabilistic MDS algorithms that explicitly account for percept variation have the potential of being able to correctly distinguish city-block from Euclidean metrics.

Properties of probabilistic distances need to be investigated further. This study has concentrated on the additivity property because the additivity properties of city-block and Euclidean metrics are very different but discriminal processes may also affect metric classification through other means. We know that the properties of probabilistic distances are not the same as deterministic distances. In Case III spaces and in anisotropic spaces, for example, we know that expected distances are not monotonically related to the distances among the centroids (Zinnes & MacKay 1992). It is also self-evident that for Thurstonian models, self-similarities need not be equal (a percept with a small variance is more similar to itself than a percept with a large variance), minimality need not exist (two different percepts may be more alike than one percept is to itself), and the triangle inequality may be violated. Thurstonian models can also be extended by adding devices such as response bias parameters (Ennis and Johnson 1994) to accommodate asymmetric data.

Very little is known about the properties of probabilistic distances that arise from Thurstonian (or non-Thurstonian) processes. Probabilistic distances may violate the properties of deterministic distances but the underlying spatial structure of the probabilistic distance models imposes constraints on the degree to which these violations may occur. The nature of these constraints is largely unknown.

Appendix: PDFs and Expected Values

It is assumed that the coordinates x_{ik} of stimulus i on dimension k are normally distributed with mean μ_{ik} and variance σ_{ik}^2 . For the Euclidean case, we can temporarily assume the covariances are zero and from [1] write

$d_{ij}^2 = \sum_{k=1}^{r} d_{ijk}^2$ where $d_{ijk} = x_{ik} - x_{jk}$. Thus,

$$d_{ijk} \sim N(\mu_{ijk}, \sigma_{ijk}^2)$$

where

$$\mu_{ijk} = \mu_{ik} - \mu_{jk}$$
$$\sigma_{ijk}^2 = \sigma_{ik}^2 + \sigma_{jk}^2 .$$

Then, letting

$$z_{ijk} = d_{ijk}^2 / \sigma_{ijk}^2$$

it is well known (Suppes and Zinnes, 1968) that z_{ijk} is distributed as a non-central chi-square χ'^2 distribution with one degree of freedom ? and non-centrality parameter

$$\lambda_{ijk} = \mu_{ijk}^2 / \sigma_{ijk}^2.$$

Thus,

$$d_{ijk}^2 \sim \sum_{k=1}^{r} \sigma_{ijk}^2 \chi'^2_{\nu=1,\lambda_{ijk}} .$$

To accommodate the situation where the covariances are not zero, we note (dropping the i, j subscripts) that the distribution of d^2 is a specific example of a quadratic form

$$Q(\mathbf{W}) = \sum_{k=1}^{r} \sigma_k^2 \left(w_k - \sqrt{\lambda_k}\right)^2$$

In normal variables distribution where $\mathbf{W}$ is an r-element vector and $w_k \sim N(0,1)$. Johnson and Kotz (1970, Chapter 29) provide the transformations for expressing the non-zero covariance case as a function of independent variables. The PDF f of d follows trivially as

$$f(d \mid \boldsymbol{\mu}, \mathbf{S}, r, p=2) = f(Q(\mathbf{W}))2d$$

where $f(Q(\mathbf{W}))$ is the PDF of the quadratic form.

To find $E(d)$, we again go to Johnson and Kotz (1970) for the characteristic function of $Q(\mathbf{W})$ which then allows $E(d^2)$ to be calculated. Jensen and Solomon (1972) derive moments of transformations of the type $(Q(\mathbf{W})/E(d^2))^h$ which, letting $h = ½$, allows $E(d)$ to be derived. The result is the following:

$$E(d\mid \boldsymbol{\mu},\mathbf{S},r,p=2)=E(Q,1/2)\theta_1^{1/2}$$

where:

$$E(Q,1/2)=1-\frac{0.25\theta_2}{\theta_1^2}+\frac{0.375\theta_2^4(-1.25+\frac{4\theta_3}{3\theta_1})}{\theta_1^4}-\frac{0.9375\left(\frac{2.625\theta_2^3}{\theta_1^2}-\frac{4.66667\theta_2\theta_3}{\theta_1^2}+\frac{2\theta_4}{\theta_1}\right)}{\theta_1^3}+O(\theta_1^{-4})$$

$$\theta_s=\sum_{k=1}^{r}\sigma_k^{2s}(1+s\lambda_k).$$

For the city-block metric, we start with the observation that the difference in [2], unlike the Euclidean case [1], must be positive. Assuming, as with the Euclidean model, that the coordinates are normally distributed, the absolute differences of the city-block metric then follow a folded normal distribution. PDFs of one-dimensional folded normal distributions are provided by Leone, Nelson and Nottingham (1961) from which we get

$$f(d)=\frac{1}{\sqrt{2\pi}\sigma}\left[\exp\left(-\frac{(d-\mu)^2}{2\sigma^2}\right)+\exp\left(-\frac{(d+\mu)^2}{2\sigma^2}\right)\right]$$

PDFs of distances derived from multidimensional folded normal distributions are provided by MacKay (2001). The PDF has the following closed form in two dimensions:

$$f(d\mid \boldsymbol{\mu},\mathrm{S},r=2,p=1)$$

$$=\frac{1}{\sqrt{2\pi}a}\sum_{k_1=1}^{2}\sum_{k_2=1}^{2}\sum_{k_3=1}^{2}\frac{\Phi\left[\left(-1^{k_3}\mu_j\sigma_i^2(-1)^{k_2}\mu_i\sigma_j^2+c_{k_1}d\right)/(ab)\right]-(1/2)}{\exp\left[-1^{k_1+k_2}\mu_i(-1)^{k_1+k_3+1}\mu_j+d)^2/\left(2a^2\right)\right]}$$

where $a=\sqrt{\sigma_i^2+\sigma_j^2}$, $b=\sigma_i\sigma_j$, $c_1=\sigma_i^2$, $c_2=\sigma_j^2$ and Φ is the cumulative distribution function (CDF) of the standardized normal distribution. PDFs in higher dimensions are approximated. The approximations are reported to be quite good.

To obtain the expected values, an easy approach is to calculate the moment generating function

$$m(t) = \int_0^\infty \exp(td) f(d) dd$$

for the one dimensional case, then take the log for the cumulant generation function,

$$c(t) = \ln\left[\frac{1}{2}\exp\left(\frac{1}{2}t(-2\mu + \sigma^2 t\right) \left(1 + \exp(2\mu t) - \operatorname{erf}\left(\frac{\mu - \sigma^2 t}{\sqrt{2}\sigma}\right) + \exp(2\mu t)\operatorname{erf}\left(\frac{\mu + \sigma^2 t}{\sqrt{2}\sigma}\right)\right)\right],$$

and evaluate setting $t = 0$. Since the cumulants of a sum of independent random variables are equal to the sum of the cumulants, the mean (first cumulant) is readily found by summing over r dimensions to be

$$E(d \,|\, \boldsymbol{\mu}, \mathrm{S}, r, p = 1) = \sum_{k=1}^{r} \left[\frac{\sqrt{2/\pi}\,\sigma_k}{\exp(\mu_k^2 / 2\sigma_k^2)} + \mu_k \operatorname{erf}\left(\frac{\mu_k}{\sqrt{2}\sigma}\right)\right].$$

References

AKAIKE, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, 19, 716-723.

ALBA, J. W., and HUTCHINSON, W. (1987), "Dimensions of Consumer Expertise," *Journal of Consumer Research*, 13, 411-454.

ARABIE, P. (1973), "Concerning Monte Carlo Evaluations of Nonmetric Multidimensional Scaling Algorithms," *Psychometrika*, 38, 607-608.

ARABIE, P. (1991), "Was Euclid an Unnecessarily Sophisticated Psychologist?," *Psychometrika*, 56, 567-587.

ASHBY, F. G., and LEE, W. W. (1993), "Perceptual Variability as a Fundamental Axiom of Perceptual Science," in *Foundations of Perceptual Theory*, Ed. S. C. Masin, New York: North-Holland, 369-399.

ASHBY, F. G., MADDOX, W. T. , and LEE, W. W. (1994), "On the Dangers of Averaging Across Subjects when Using Multidimensional Scaling or the Similarity-Choice Model," *Psychological Science*, 5, 144-151.

ATTNEAVE, F. (1950), "Dimensions of Similarity," *The American Journal of Psychology*, 63, 516-556.

BAUMGARTNER, H. (1993), "An Exploratory Investigation of Holistic and Analytic Modes of Product Perception," *Advances in Consumer Research*, 20, 673-677.

BORG, I., and GROENEN, P. (1997), *Modern Multidimensional Scaling: Theory and Applications*, New York: Springer-Verlag.

BOZDOGAN, H. (1987), "Model Selection and Akaike's Information Criterion (AIC): the General Theory and its Analytical Extensions," *Psychometrika*, 52, 345-370.

BRUSCO, M. J. (2001), "A Simulated Annealing Heuristic for Unidimensional and Multidimensional (City-Block) Scaling of Symmetric Proximity Matrices," *Journal of Classification*, 18, 3-33.

DUNN, J. C. (1983), "Spatial Metrics of Integral and Separable Dimensions," *Journal of Experimental Psychology: Human Perception and Performance*, 9, 242-257.

EISLER, H., and KNOPPEL (1970), "Relative Attention in Judgments of Heterogeneous Similarity." *Perception and Psychophysics*, 8, 420-426.

ENNIS, D. M. (1988), "Confusable and Discriminable Stimuli: Comment on Nosofsky (1986) and Shepard (1986)," *Journal of Experimental Psychology: General*, 117, 408-411.

ENNIS, D. M., and JOHNSON, N.L. (1993), "Thurstone-Shepard Similarity Models as Special Cases of Moment Generating Functions," *Journal of Mathematical Psychology*, 37, 104-109.

ENNIS, D. M., and JOHNSON, N. L. (1994), "A General Model for Preferential and Triadic Choice in Terms of Central F Distribution Functions," *Psychometrika*, 59, 91-96.

ENNIS, D. M., PALEN, J. J., and MULLEN, K., (1988), "A Multidimensional Stochastic Theory of Similarity," *Journal of Mathematical Psychology*, 32, 449-465.

GARNER, W. R. (1974), *The Processing of Information and Structure.* Potomac, MD:, Erlbaum.

GLAZER, R., and NAKAMOTO, K. (1991), "Cognitive Geometry: An Analysis of Structure Underlying Representations of Similarity," *Marketing Science*, 10, 205-228.

HEFNER, R. A. (1958), *Extensions of the Law of Comparative Judgment to Discriminable and Multidimensional Stimuli*, unpublished doctoral dissertation, Ann Arbor: University of Michigan.

HUBERT, L., ARABIE, P., and HESSON-MCINNIS, M. (1992), "Multidimensional Scaling in the City-Block Metric: A Combinatorial Approach," *Journal of Classification*, 9, 211-236.

JENSEN, D. R., and SOLOMON, H. (1972), "A Gaussian Approximation to the Distribution of a Definite Quadratic Form," *Journal of the American Statistical Association*, 67, 898-902.

JOHNSON, N. L., and KOTZ, S. (1970), *Continuous Univariate Distributions – 2*, New York: Wiley.

KRUSKAL, J. B., YOUNG, F. W., and SEERY, J. B. (1977), "How to use KYST-2a, A Very Flexible Program to Do Multidimensional Scaling and Unfolding," unpublished manuscript, Bell Laboratories.

LEE, M. D., and POPE, K. J. (2003), "Avoiding the Dangers of Averaging Across Subjects When Using Multidimensional Scaling," *Journal of Mathematical Psychology*, 47, 32-46.

LEONE, F. C., NELSON, L. S., and NOTTINGHAM, R. B. (1961), "The Folded Normal Distribution," *Technometrics*, 3, 543-550.

MACKAY, D. B. (1989), "Probabilistic Multidimensional Scaling: An Anisotropic Model for Distance Judgments," *Journal of Mathematical Psychology*, 33, 187-205.

MACKAY, D. B. (2001), "Probabilistic Multidimensional Scaling Using a City-Block Metric," *Journal of Mathematical Psychology*, 45, 249-264.

MACKAY, D. B., and ZINNES, J. L. (1981), "Probabilistic Scaling of Distance Judgments," *Geographical Analysis*, 34, 21-37.

MELARA, R. D. (1989), "Similarity Relations Among Synthetic Stimuli and Their Attributes," *Journal of Experimental Psychology: Human Perception and Performance*, 115, 212-231.

NOSOFSKY, R. (1985), "Overall Similarity and the Identification of Separable-Dimension Stimuli: A Choice Model Analysis," *Perception & Psychophysics*, 38, 415-432.

NOSOFSKY, R. M. (1986), "Attention, Similarity, and the Identification-Categorization Relationship," *Journal of Experimental Psychology: General*," 115, 39-57.

ROSS, R. T. (1934), "Optimum Orders for the Presentation of Pairs in the Method of Paired Comparisons," *The Journal of Educational Psychology*, 25, 375-382.

SCHÖNEMANN, P. H. (1994), "Measurement: The Reasonable Ineffectiveness of Mathematics in the Social Sciences," in *Trends and Perspectives in Empirical Social Research*, Eds., I. Borg and P. P. Mohler, New York: De Gruyter, 149-160.

SCHWARZ, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461-464.

SHEPARD, R. N. (1962), "The Analysis of Proximities: Multidimensional Scaling with an Unknown Distance Function. II," *Psychometrika*, 27, 219-246.

SHEPARD, R. N. (1974), "Representation of Structure in Similarity Data: Problems and Prospects," *Psychometrika*, 39, 373-421.

SHEPARD, R. N. (1986), "Discrimination and Generalization in Identification and Classification: Comment on Nosofsky," *Journal of Experimental Psychology: General*, 115, 58-61.

SHEPARD, R. N. (1987), "Toward a Universal Law of Generalization for Psychological Science," *Science*, 237, 1317-1323.

SHEPARD, R. N. (1991), "Integrality versus Separability of Stimulus Dimensions: From an Early Convergence of Evidence to a Proposed Theoretical Basis," in *Perception of Structure: Essays in Honor of Wendell R. Garner*, Eds., G. R. Lockheand and J. R. Pomerantz, Washington, D. C.: American Psychological Association, 53-71.

SUPPES, P., and ZINNES, J. L. (1963), Basic Measurement Theory," in *Handbook of Mathematical Psychology, Vol. I*, Eds., R. D. Luce et al., New York: Wiley, 1-76.

TAKANE, Y. (1978), "A Maximum Likelihood Method for Nonmetric Multidimensional Scaling: I. The Case in Which All Empirical Pairwise Orderings are Independent – Theory," *Japanese Psychological Research*, 20, 7-17.

THURSTONE, L. L. (1927a), "A Law of Comparative Judgment," *Psychological Review*, 34, 273-286.

THURSTONE, L. L. (1927b), "Psychophysical Analysis," *American Journal of Psychology*, 38, 368-389.

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

TVERSKY, A. (1977), "Features of Similarity," *Psychological Review*, 84, 327-352.

TVERSKY, A., and GATI, I. (1982). "Similarity, Separability, and the Triangle Inequality," *Psychological Review*, 89, 123-154.

ZINNES, J. L., and MACKAY, D.B. (1981), "Multidimensional Scaling Models: the Other Side," in *Multidimensional Data Representations: When & Why,* Ed. I. Borg, Ann Arbor: Mathesis Press, 517-542.

ZINNES, J.L., and MACKAY, D.B. (1983), "Probabilistic Multidimensional Scaling: Complete and Incomplete Data," *Psychometrika,* 48, 27-48.

ZINNES, J. L., and MACKAY, D.B. (1992), "A Probabilistic Multidimensional Scaling Approach: Properties and Procedures," in *Multidimensional Models of Perception and Cognition*, Ed. F. G. Ashby, Hillsdale: Erlbaum, 35-60.

ZINNES, J.L., and WOLFF, R.P. (1977), "Single and Multidimensional Same-Different Judgments," *Journal of Mathematical Psychology*, 16, 30-60.