Recognition models of alphanumeric characters

Perception & Psychophysics
1981, 29 (3), 234-246

Recognition models of alphanumeric characters

GIDEON KEREN
Institute for Perception TNO, Soesterberg, The Netherlands

and

STAN BAGGEN
Institute for Perception Research, Eindhoven, The Netherlands

Several methods to study the recognition and similarity of alphanumeric characters are briefly discussed and evaluated. In particular, the application of the choice-model (Luce, 1959, 1963) to recognition of letters is criticized. A feature analytic model for recognition of alphanumeric characters based on Tversky's (1977) features of similarity is proposed and tested. It is argued that the proposed model: (a) is parsimonious in that it utilizes a relatively small number of parameters, (b) is psychologically more meaningful compared with other approaches in that it is attempting to study underlying processes rather than just reveal a similarity structure, (c) yields predictions that have a high level of fit with the observed data. Possible implications from the use of the model for future research are briefly discussed.

The interest in the present study stems from two different sources: One is the theoretical analysis of similarity relations which have been often treated in the psychological literature as distance between stimuli in a Euclidean space. Recently, Tversky (1977) has challenged that approach by pointing out that some basic assumptions of spatial models may not be warranted. Tversky (1977) proposed a set-theoretical approach to similarity by which objects are represented as collections of features, and similarity is described as a feature matching process. This analysis avoids some of the deficiencies of the geometrical models, and may also turn out to be psychologically more meaningful, since it is testing some possible underlying processes of similarity judgments.

The other source is related to experimental work on perception of printed verbal material which often requires some measures of interletter similarity. Information of similarity between letters may serve as a useful source for understanding letter recognition and the factors that affect letter recognition performance (and possible confusions) in experimental studies that use letters as stimuli. One approach to estimating letter-similarity parameters appears to come from the application of Luce's choice-model (Luce, 1959, 1963). Townsend (1971), for instance, reported a confusion matrix derived from tachistoscopic letter-recognition performance of six subjects, and has shown the choice-model to be the most representative index of interletter similarity among several competing models. Gilmore, Hersh, Caramazza, and Griffin (1979) and Holbrook (1975) have supported this conclusion. There are, however, several difficulties with the application of the choice model to interletter similarity—to be discussed below—which raise questions regarding the superiority and validity of this model.

The goal of the present research was to develop an alternative model for interletter similarity that avoids certain deficiencies of the choice model and, at the same time, will have what we believe to be more psychological content. Despite the fact that the model to be proposed shares some common features with traditional geometrical models, it is mainly based on the ideas presented by Tversky (1977) and, as such, is not vulnerable to most of the criticisms raised by Tversky with regard to geometrical models.

In the next two sections, we briefly describe the set-theoretical approach as proposed by Tversky followed by a summary of different attempts to deal with interletter similarity with emphasis on the choice model. In subsequent sections, the application of the choice model is critically evaluated and an alternative model is proposed and tested. Some general comments with regard to the study of interletter similarity are left for the final section.

Tversky's Contrast Model

Tversky (1977) has challenged the dimensional-metric assumptions that underlie the geometric approach to similarity and has proposed an alternative set-theoretical approach. In particular, Tversky (1977)

The research reported in the present paper was done while the first author was on a research fellowship at the Institute for Perception Research, Eindhoven, Holland. Both authors would like to acknowledge the help and support they received from the Institute. They are also indebted to Herman Bouma and Don Bouwhuis for helpful comments on earlier drafts of this paper. Requests for reprints should be sent to Gideon Keren, Institute for Perception TNO, Kampweg 5, Soesterberg, Holland.

0031-5117/81/030234-13$01.55/0

has questioned the validity of the following assumptions: Consider a metric distance function, δ, which is a scale that assigns to every pair of points a nonnegative number, called their distance. Geometrical models assume the following with regard to δ:

Minimality:

$$\delta(a,b) \geq \delta(a,a) = 0;$$

Symmetry:

$$\delta(a,b) = \delta(b,a);$$

The triangle inequality:

$$\delta(a,b) + \delta(b,c) \geq \delta(a,c).$$

Tversky pointed out several examples from the experimental psychological literature which violate these assumptions. Even in the specific context of our paper, which deals with the recognition of alphanumeric characters, it is easy to point out several examples which are incongruent with the above three assumptions. For instance, in confusion matrices generated in letter-recognition studies, the off-diagonal entries, which represent similarity of different letters, may sometimes be significantly larger than some diagonal entries, which represent the similarity of a letter with itself.[1] Such observations violate the first assumption (minimality). For example, in the confusion matrix reported by Gilmore, Hersh, Caramazza, and Griffin (1979), the letter Q is more often recognized as an O ($p = .207$) than as a Q ($p = .184$), the letter M is more often recognized as an H ($p = .391$) or an N ($p = .165$) than as itself ($p = .110$), and the letter W is more often recognized as an H ($p = .354$) than as itself ($p = .180$).

The second assumption, namely that of symmetry, is refuted by simply observing the letter confusion matrices reported by Gilmore et al. (1979) and Townsend (1971) and the numeral confusion matrices reported by Van Nes and Bouma (1980) and the one that was obtained by the present authors and appears in Table 1.

The third assumption cannot be refuted with ordinal or even interval data (Tversky, 1977). However, informally it implies that if a is quite similar to b, and b is quite similar to c, then a and c cannot be very dissimilar. A brief scan of the data by Gilmore et al. (1979) suggests that this implication does not necessarily hold. For instance, the letter A is confused quite often with R ($p = .224$) and R is confused quite often with P ($p = .109$), but A is rarely confused with P ($p = .023$).

In an alternative approach proposed by Tversky (1977), objects a and b are characterized by sets of features denoted A and B, respectively, and the similarity between a and b—S(a,b)—is expressed as a function of their common ($A \cap B$) and distinctive features (A-B and B-A). Tversky's theory is based on a set of qualitative assumptions about the observed similarity ordering, and yields an interval similarity scale S, which preserves the observed similarity order, and a scale f, which reflects the salience (and hence the contribution) of any particular feature and is defined on the relevant feature space such that:

$$S(a,b) = \theta f(A \cap B) - \alpha f(A\text{-}B) - \beta f(B\text{-}A), \quad (1)$$

where $\theta, \alpha, \beta > 0$. The contrast model, as it has been termed by Tversky, describes the similarity between objects a and b as a linear combination of the measures of their common and distinctive features. The model defines a family of scales that are characterized by the parameters θ (weighting common features), and α and β (weighting distinguishing features).

The contrast model differs from the traditional geometric models of similarity in two major respects: First, the model does not depend on any of the three assumptions mentioned earlier. Second, unlike most geometrical models whose goal is to reveal the similarity structure between the objects under study, the contrast model is also concerned with cognitive aspects, underlying similarity judgments, in particular the process by which information about common and distinctive features is aggregated, and the relative conspicuity of these features. The importance of the psychological interpretation is emphasized in Tversky and Gati (1978) and will also be addressed in later sections of the present paper.

The Choice Model and Other Approaches to Interletter Similarity and Confusions

Researchers interested in alphabetic pattern recognition have tried to investigate the underlying similarity structure of letters. One common method within the geometrical approach has been the use of different multidimensional scaling techniques. Gilmore et al. (1979), for example, have tried to assess the underlying similarity structure by using hierarchical clustering (Johnson, 1967) and the multidimensional scaling procedure TORSCA (Young & Torgerson, 1967). According to Gilmore et al. (1979), the hierarchical clustering technique did not reveal whether a common dimensional structure underlay the perceptual space. The TORSCA procedure yielded a five-dimensional metric space solution, but the authors claimed they could not find a readily apparent interpretation for those dimensions.

Another approach has been the use of mathematical models of recognition in an attempt to predict experimentally obtained confusion matrices. Townsend (1971) obtained two alphabetic confusion matrices (under two different experimental conditions) in

order to test the power of three mathematical models of recognition. Two of the models were based on the concept of a finite number of sensory states labeled by Townsend as the "all-or-none activation" model (which assumes only a perfect perception or a random guessing state following the stimulus presentation) and the "overlap activation model" (which assumes that pairwise similarity can affect interletter confusability and, consequently, the possibility of partial information states). The third model tested by Towsend was the choice model (Luce, 1959, 1963). Townsend's analyses suggest that the choice model is probably the most valuable one. This conclusion has since been endorsed by several investigators (e.g., Gilmore et al., 1979; Holbrook, 1975), who concluded that "the most reliable and valid of the available measures (of interletter similarity) appears to be Townsend's (1971) set of similarity parameters based on the Luce choice model" (Holbrook, 1975, p. 532). Consequently, the choice model became a popular measure of interletter similarity and has been often used as a standard to which alternative approaches have been compared (e.g., Geyer & DeWald, 1973; Holbrook, 1975). In the following sections we provide a brief description of the choice model along with some evaluation and critical comments regarding its performance.

The choice model (Luce, 1963) assumes the existence of two ratio scales, with which one may describe the "response strength" of each possible response pair as:

$$\eta: S \times R \rightarrow R^+$$

$$\beta: \quad R \rightarrow R^+,$$

where S is the set of stimuli, R the set of responses, and R^+ is the set of positive real numbers. η may be regarded as a measure of similarity between any pair of stimuli, while β is usually interpreted as a response bias. The response strength s_{ij} of response j to stimulus i is expressed by

$$s_{ij} = \beta_j \cdot \eta_{ij}. \tag{2}$$

According to Luce (1959, 1963), one can obtain a confusion matrix by normalizing each row of the response strength matrix such that the numbers in each row add to one and can thus be considered as probabilities. Each entry in the confusion matrix P then refers to the probability of a response (corresponding to the column) to a given stimulus (corresponding to the row).

The experimental data to serve as input for an analysis with the choice model, are usually summarized in a confusion matrix. One is then faced with the problem of estimating the response biases β_j and the similarities η_{ij} from the probabilities p_{ij}. Townsend (1971) has used the following estimators:

$$\hat{\eta}_{ij} = \left[\frac{p_{ij}\, p_{ji}}{p_{ii}\, p_{jj}}\right]^{1/2} \tag{3}$$

$$\hat{\beta}_j = \frac{1}{N} \sum_{K=1}^{N} \left[\frac{p_{jj}\, p_{kj}}{p_{jk}\, p_{kk}}\right]^{1/2}. \tag{4}$$

The popularity of the choice model stems mainly from its success in producing highly accurate predictions, is exemplified, for instance, in the reports of Gilmore et al. (1979) and Townsend (1971). Nevertheless, one may point out several problems with regard to the choice model, which may cast some doubt on its appeal. Some of these issues are summarized below. It should be emphasized that the following critical comments are not intended to evaluate the choice model in general but, rather, are restricted to the application of the model within the specific context of the present paper. Nevertheless, some of those comments may be generalized to other applications as well.

(1) It is often the case that certain cells in a confusion matrix contain values of zero (e.g., Gilmore et al., 1979; Townsend, 1971). There may be several explanations for that finding, such as insufficient number of observations (which probably explains the large number of empty cells in Townsend's data) or a strong response bias. Zeros in the confusion matrix are disturbing, since any value of zero in Equation (3) or Equation (4) will yield zero or undefined estimates for either $\hat{\eta}$ of $\hat{\beta}$. Zeros in the numerators of Equation (3) or Equation (4) would lead to underestimation of the parameters, while the effect of zeros in the denominators is indeterminate. There are several technical methods to overcome that problem. Gilmore et al. (1979), for instance, used an iterative least squares procedure to solve for the similarity matrix and bias vector by substituting a variable probability value in the empty elements of the confusion matrix.[2]

(2) The choice model assumes minimality and is thus unable to handle experimental data in which off-diagonal entries exceed the diagonal entries and are too large for the choice model to handle. If the number of cells that violate the minimality assumptions is small, it is possible to use several technical solutions: One possibility is to interchange the maximum value in a row and the diagonal value of that row in the confusion matrix (P) in order to obtain a workable similarity matrix (η). The values that were interchanged are replaced back, after the (theoretical) confusion matrix is recalculated from the similarity matrix and the response bias vector. Another alternative is to set all the values in the similarity matrix such that $\eta_{ij} = \min(1, \eta_{ij})$, thus postulating that an object cannot be identified as being more similar to another object than to itself. Such solutions do not have theoretical justification and are thus to a great extent arbitrary.

(3) The choice model explicitly assumes two ratio scales: The scale η is interpreted as a measure of similarity (determined by sensory influences), whereas β is associated only with the response system (which does not depend on sensory effect).[3] Although the two are assumed to be independent, it may still be questionable whether they are, indeed, separable, a question which, at least to a certain extent, is empirical in nature. Our observations lead us to doubt the validity of this assumption. For instance, we have analyzed the data reported by Gilmore et al. (1979), and noticed a correlation of $-.733$ between the response bias vector and the average similarity (for each stimulus) produced by the choice model.[4] An additional related issue is the interpretation of the response bias. It is common to assume that response bias may be correlated with the relative frequency of occurrence of the stimuli under study, but there has been little explicit discussion of what exactly the psychological nature and conceptual meaning of response bias is. If, indeed, response bias is, among other things, a function of frequency, then it would be reasonable to expect that it should be stable across different experimental situations using the same stimuli. In particular, the response bias vectors for different confusion matrices of alphabetic characters (even if the type fonts used are somewhat different from condition to condition) should be similar. This at least is not the case with regard to the data reported by Gilmore et al. (1979) and Townsend (1971), in which the corresponding response bias vectors have a quite different pattern. Moreover, the response bias is assumed to be common across subjects; we have carefully inspected the individual data on the basis of which Table 1 was constructed, and were unable to reveal any salient response bias across subjects. Rather, it seems as if the pattern of biases (if it exists) is unique to each individual.

The correlation between the similarity and response bias suggests that the two are not unrelated and perhaps that response bias may also be affected by some stimulus properties. If this is, indeed, the case, one may be tempted to postulate two different and distinguishable bias sources: (1) a bias which, due to differences among stimuli in terms of their tendency to be confused, is determined by the qualifications of the stimuli, and (2) a "pure" bias that is due to differences among responses in terms of their tendencies to be produced. Goldstein (Note 1), citing Holman, has shown that any set of data that can be modeled by using two bias components (one for stimulus and one for response) can be modeled equally well by using just a single response bias. Thus, Goldstein (Note 1) argued that "mathematically there is no unique solution, in the metric model, for representing the asymmetry of the matrix in terms of bias functions" (p. 65).

A final possibility is that the response bias has no precise conceptual meaning and simply serves as an error term in order to obtain a symmetrical similarity matrix (η), in which case it does not have any theoretical importance. Though some of the above comments suggest that this may indeed be the case, such a contention has to be tested more rigorously. It would be fair to conclude, however, that the meaning and interpretation of the response bias is as yet an unresolved issue.

(4) The choice model contains $[n(n+1)/2]-1$ free parameters to be estimated from the data. In the present context of alphabetic characters, 350 parameters have to be estimated in order to predict the 676 cells of the confusion matrix. The large number of parameters estimated by the choice model probably accounts for its superior performance, but obviously there is still a lot to be desired in terms of parsimony. The final outcome of the model is an estimated similarity matrix (which may be accurate or not), but little insight is gained with regard to the underlying similarity structure.

There are two general criteria on which to test the quality of a model: One is its predictability or goodness of fit, that is, how close the predictions are to the observed data. However, one may always increase the number of parameters to be estimated and consequently gain additional precision. The second criterion is parsimony and the explanatory power of the model, that is, the extent to which the model is interpretable and able to explain the phenomenon under investigation. In that respect, a relatively small number of parameters is more likely to be interpretable and conceptually meaningful. While the choice model performs excellently in terms of the first criterion, it does rather poorly in terms of the second one.

(5) Another aspect of the choice model, which is also shared by multidimensional scaling methods, is the relative lack of psychological substance in the sense of revealing the processes underlying similarity judgments. The search for the underlying similarity structure in terms of dimensions (if those can be specified) is important but should not be considered as an end in itself.

An attempt to study information processing in recognition of letters has been made by Geyer and Dewald (1973). These investigators have postulated a rather elaborate information processing system for recognition. Unfortunately, some of the general assumptions underlying their approach—that is, that the display of arrays of symbolic stimuli results in the existence of stimulus-evoked (SE) lists, that preliminary preattentive processes organize coherent segments in the visual field to separate SE lists, and that SE lists are then compared with a set of lists in long-term memory—have really not been tested directly in their study and thus should be considered only as conjectures. Also, some of the explicit assumptions made by Geyer and Dewald (1973) are questionable. In particular, the assumption that the mechanism of

stimulus degradation operates such that features can be "lost" but, on the other hand, cannot appear spuriously in a degraded SE list, is inaccurate. It is, indeed, more often the case that, with highly similar letter pairs, the letter that contains more segments is more frequently confused with the letter that contains fewer segments than vice versa. For instance, E is more frequently confused with F than conversely, as indicated from the data reported by Townsend (1971). However, inspection of existing data (e.g., confusion matrices of letters and numerals) suggests that the rule cannot be stated deterministically. A probabilistic statement, that it is more likely for a letter with more segments (or features) to be confused with its counterpart with fewer segments than vice versa, seems to be more accurate. Such an argument has been raised by Van Nes and Bouma (1980), who reported confusion matrices of segmented numerals (like those that appear on calculator displays) and concluded that a digit not recognized correctly is more often perceived as one with a configuration simpler than that of the presented digit than the other way around.[5] The model to be proposed in the next section explicitly deals with processes of adding or deleting a specific feature (or segment) and provides estimates of the likelihood for the occurrence of such events.

A final note with regard to Geyer and Dewald's (1973) study is concerned with the feature list they recommend. We are aware of the difficulties involved in defining the feature list and will have more to say on the topic below. We are not convinced, however, that the feature list they proposed is indeed the most valid one. Holbrook (1975) compared several feature lists and has argued that the one proposed by Geyer and Dewald (1973) did not show an improvement over Gibson's (1969) list and was actually poorer. We have tried to incorporate the Geyer and Dewald list in the model to be described below and obtained rather poor results, thus confirming Holbrook's comments.

The purpose of the above discussion was to point out some deficiencies of previous approaches to study interletter similarity, in particular the choice model, and at the same time to emphasize some of the objective difficulties inherent in such an investigation. We now turn to the development of our model in which an attempt was made to overcome some of the problems discussed earlier. This is followed by a test of the model, using some empirical data. Finally, we discuss some unresolved problems with the hope of encouraging future research.

The Proposed Model

The following is a description of a feature-analytic model, which may be considered as a modified version of the approach proposed by Tversky (1977). Among other things, the proposed model does not assume minimality, symmetry, and the triangle inequality, which, as suggested by Tversky, are often violated in similarity data.

A preliminary step of the present model is to obtain a set of distinguishing features by which each stimulus may be uniquely described. Thus, the input to the model consists of a characterization of the stimuli in terms of the defining features. The difficulties involved in revealing the underlying set of defining features have already been discussed and are common to any feature-analytic approach. Since different sets of features have to be defined for different sets of stimuli, we defer the discussion on this issue to the experimental section, where we describe in detail the method used to obtain the feature set.

The starting point of the model is the assignment of a dissimilarity measure d_{ab}, for each pair of stimuli (a,b). Note that no assumptions are made with regard to symmetry, that is, we do not necessarily assume $d_{ab} = d_{ba}$. Using A, B, C to denote the set of defining features for stimuli, a, b, c, respectively, and following Tversky's (1977) approach, we assume that similarity judgments, and in turn our dissimilarity measure d_{ab} for the ordered pair (a,b) are a function of three components: the features that are common to the two sets of defining features, that is, $A \cap B$; those that define A but not B, that is, A-B; and those that define B but not A, that is, B-A. Each of those three components makes its own contribution to the dissimilarity measure, d_{ab}. Assuming a nonnegative scale, f, we define the dissimilarity between two objects a and b as

$$d_{ab} = \theta f(A \cap B) + \alpha f(A\text{-}B) + f(B\text{-}A), \qquad (5)$$

where the parameter θ stands for the relative weight of the common features, α stands for weights associated with distinctive features, and f is a scale that reflects the salience, or importance, of each particular feature to the similarity between objects in the context of both common and distinctive features. Thus, f is a scale of the overall prominence of various features, while θ and α reflect the process by which the information contained in different features is aggregated, depending on whether they are common or distinctive in each case.

Tversky (1977) has used two different parameters—α and β—for weighting A-B and B-A, respectively. In the present model, a single parameter (α) is used for the distinguishing features, and the second parameter is assumed to be constant and equal to 1. The formal justification for this modification is as follows: Assume, for the sake of simplicity, that $\theta = 0$, and let γ and β be two different parameters for the distinguishing features A-B and B-A, respectively. Then our measure of dissimilarity should be

$$d_{ab} = \gamma f(A\text{-}B) + \beta f(B\text{-}A),$$

which can be rewritten as

$$d_{ab} = \frac{\gamma}{\beta}[\beta f(A\text{-}B)] + [\beta f(B\text{-}A)] = \alpha g(A\text{-}B) + g(B\text{-}A),$$

where $\alpha = \gamma/\beta$ and $g(A) = \beta f(A)$. It is important to note that the scale f in Equation 5 satisfies feature additivity, that is, the function f applied to any subset of features is a linear combination of this function applied to each element in that particular subset. Also note, in Equation 5, that $\theta = 0$ implies $d_{ab} \geqslant 0$ for all a,b, that is, minimality holds and $d_{aa} = 0$. If $\alpha = 1$, the implication is that symmetry holds, that is, $d_{ab} = d_{ba}$. However, no specific value is assumed for either θ or α, since both are parameters that have to be estimated.[6] In all the applications to be described in the next section, the values obtained for θ were nonzero and those for α different from 1, thus suggesting that neither minimality nor symmetry hold for the data under consideration.

As a next step, we assume the existence of a similarity measure, S, which has a 1-to-1 relationship with the distance measure defined above such that[7]

$$S_{ab} = \exp(-d_{ab}). \tag{6}$$

From the similarity measures, a confusion matrix—expressed in probabilities (p_{ab})—is derived by normalizing each row of the similarity matrix such that the probabilities add to 1.0. Thus,

$$p_{ab} = \frac{S_{ab}}{\sum_{b'} S_{ab'}} \tag{7a}$$

and

$$p_{ba} = \frac{S_{ba}}{\sum_{a'} S_{ba'}}. \tag{7b}$$

For application of the model, each stimulus has to be uniquely defined in terms of features. At the final stage, the weights (for common and distinguishing features) are estimated such that a certain criterion expressing the deviations between the predicted and observed confusion matrices is minimized. There is no normative rule to suggest which criterion should be used. The present model adopts the customary criterion:

$$\frac{\Sigma (O\text{-}E)^2}{\Sigma (E\text{-}\bar{E})^2}, \tag{8}$$

where O and E are the observed and expected (predicted from the model) probabilities, respectively. The minimization is achieved through an optimization procedure termed STEPIT (Chandler, Note 2). Briefly, this procedure employs a multidimensional space with separate orthogonal dimensions for each variable that has to be estimated. A point in this space is defined by an n-dimensional vector, the n elements corresponding to the n values of those variables. Optimization proceeds by choosing a starting value for each of the n variables which, in combination, define the initial point in the space. The procedure continues by searching for neighboring points in the space such that the criterion that has been chosen is being minimized. In our applications, the criterion has been converging within the first 700-800 function computations, with negligible improvement beyond that. For more details of the optimization procedure, the reader is referred to Hook and Jeeves (1961) and Chandler (Note 2).

In the following sections, we describe the application of the proposed model on some empirical data. Several additional details regarding the model are dealt with in later sections.

EXPERIMENTAL TESTS OF THE PROPOSED MODEL

We decided to test the model in two stages: Since the model's performance is influenced by the extent to which a well-defined feature set can be found, and because of the difficulties involved in revealing the appropriate feature set, it seemed desirable to start with data in which a reliable and valid set of features might be detected without major difficulties. To that end, it was decided to start the initial test on segmented numerals, for which a self-evident feature set is readily available. The second test of the model was conducted by applying the model to interletter confusion data; that is described later.

Segmented Numeral Confusion Matrices

Segmented numerals are rectilinear configurations composed of subsets of seven segments.[8] The number 8 (see Figure 1) is composed of all the seven segments; all other digits consist of subsets of these segments. For convenience, we have numbered those segments from 1 to 7 as illustrated in Figure 1. Since these segments seem to be the most natural features, they were included in our feature-set as described below.

Some data on the discriminability of segmented numerals does exist. Van Nes and Bouma (1980) report two confusion matrices of segmented numerals that were obtained under different conditions: one with a relatively large observation distance (16 m) and no time limitations, and the other under para-

foveal (eccentric) vision and brief time exposures (100 msec). Unfortunately, both sets of data are based on a rather small number of observations and thus raise questions regarding the reliability of the data. Indeed, the confusion matrices they report contain a relatively large number of empty cells, which increases concern. Consequently, we decided to conduct our own experiment with a large number of observations.

Method

Apparatus and Stimuli. A Scientific Prototype Model N-1000 three-field tachistoscope was employed with Sylvania FT45/CWX lamps. Luminances of the three fields were set at 4.0 fL, as measured with an Ilford S.E.I. photometer. A small black fixation cross appeared constantly in Field 1 (the adaptation field), except when the stimulus display or a blank were presented in Fields 2 and 3, respectively.

The stimuli consisted of 10 black segmented numerals, 0-9, mounted on white vinyl cards. They were constructed from Cello-Tak transfer type (No. 3024) straight lines, .012 cm in width. All numerals were 1.8 cm in vertical length (which subtended 47 min of visual angle) and 1.1 cm in horizontal length (which subtended 30 min of visual angle). Two cards were constructed for each numeral: one in which the center of the numeral appeared 2.6 cm left and one in which it appeard 2.6 cm right of the fixation cross center.

Subjects. Eight undergraduate students (five females) from Erindale College, University of Toronto, participated in the experiment. Each served as a subject for two sessions and was paid $5 for her (his) services. All had normal or corrected-to-normal vision.

Procedure. Each session consisted of 300 trials (the first 100 served as practice) and lasted for approximately 1 h. Stimulus exposure times varied from subject to subject; they were determined during the practice session such that they met the criterion of 70% correct responses. In order to avoid guessing, it was decided not to set the criterion below 70%, since the stimulus population (and consequently possible responses) consisted of only 10 items. The range of stimulus exposure durations varied between 2 msec for the fastest subject and 4.4 msec for the slowest.

At the beginning of each session, the subjects were first familiarized with the set of numerals by being presented with each number for 2 sec. They were then instructed to initiate each trial by pushing a button when they had the fixation cross in clear focus and were ready. They were told that one of the digits would appear either to the left or to the right of the fixation point for a very brief time and that their task was to identify the digit and report it. The subjects were also instructed to report the digit name, even if they were somewhat uncertain. In case they were unsure with regard to two possible numbers, they were to report the one that seemed to be more likely. However, they were discouraged from "pure" guessing and were given the choice to respond "illegible" whenever they felt they had seen absolutely nothing.

After the subject initiated the trial, the fixation field was replaced by a white blank field for 100 msec, followed by the stimulus display, which remained for the exposure time determined for each subject. Finally the blank field appeared again for 100 msec and was then replaced by the fixation field, which remained until the beginning of the next trial. Feedback of whether the response was correct or not was given during the first 100 practice trials but not during the remaining 200 experimental trials.

Results

Subjects responded "illegible" on less than 4% of the trials; these trials were omitted from the final analysis. The confusion matrix, based on the rest of the experimental trials, is presented in Table 1.

Several important comments should be made with regard to the data. First, the confusion matrix is obviously not symmetrical. Moreover, it is difficult to generalize the direction of asymmetry: In some instances, it is more often the case that numerals with the larger number of segments are confused with numerals consisting of fewer segments than vice versa (e.g., 7-1, 8-0, 4-1). In other instances, however, the order is reversed (e.g., 6-8, 3-9, 5-9). Second, it seems as if the seven segments (numbered 1 to 7) from which the numerals were constructed (see Figure 1) constitute an adequate feature set for describing the stimuli and their underlying perceived similarity. This conclusion is supported by inspection of Figure 1, in which probability of confusions between a given pair of digits (in either direction) is plotted as a function of the number of distinguishing elements (or difference segments) which separate a given pair. For example, there is only one distinguishing element between the pairs 1-7, 3-9, 8-0, etc. There are two distinguishing elements in pairs such as 3-8, 1-4, 9-4, and so on. As can be seen, the percentage of confusions between pairs of numerals, averaged over all observational conditions, is negatively related to the number of distinguishing elements characterizing those pairs. A similar finding is reported by Van Nes and Bouma (1980). Finally, it is important to notice that inspection of the individual data did not reveal consistent response biases across subjects.

Table 1
Confusion Matrix for Segmented Numerals Obtained From the Experiment

	Response									
Stimulus	1	2	3	4	5	6	7	8	9	0
1	.877	.007	.007	.022	.004	.015	.060	.000	.004	.004
2	.014	.782	.047	.004	.036	.047	.014	.029	.007	.018
3	.029	.029	.681	.007	.018	.000	.040	.029	.152	.015
4	.149	.022	.004	.732	.004	.011	.030	.007	.041	.000
5	.014	.026	.043	.014	.669	.079	.007	.007	.126	.014
6	.025	.014	.007	.011	.097	.633	.004	.155	.011	.043
7	.269	.004	.021	.021	.007	.000	.667	.000	.004	.007
8	.011	.028	.028	.018	.018	.070	.011	.577	.067	.172
9	.025	.029	.111	.046	.082	.011	.021	.082	.550	.043
0	.018	.004	.007	.011	.007	.018	.025	.071	.021	.818

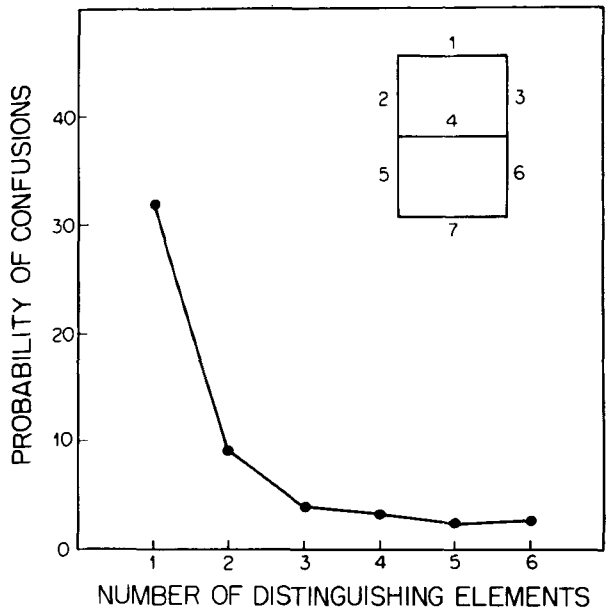

Figure 1. Probability of confusing any two digits as a function of the number of distinguishing segments.

The data in Table 1 were used as input for both the choice model and the model proposed in the present paper. To accommodate the data in Table 1 for the choice model analysis, the few empty cells were replaced with probabilities of p = .002 and the whole row was renormalized such that the sum of the probabilities would still equal 1.00.

The analysis of the data by the choice model yielded response biases that were moderately close to 1.00, except for the numeral 1, for which the response bias was 2.2, thus reflecting the relatively high probabilities of confusing 7 with 1 and 4 with 1, but not vice versa. Two criteria were utilized to test the adequacy of the model in terms of goodness of fit: One was the optimization criterion used for our model (see Equation 8), and the other was Pearson's correlation coefficient, r, where all cells of one matrix (obtained data) were correlated with the corresponding cells of the other (theoretical predictions matrix). The values for these two criteria were .009 and .995, respectively, thus indicating a high correspondence between the observed confusion matrix and the one predicted by the choice model. The same two criteria were also calculated with respect to the main diagonal values of the two confusion matrices: the value for the optimization criterion was .012 and the correlation coefficient was .994, suggesting that the goodness of fit of the diagonal values is slightly less than that of the entire matrix.

For application of our model, a set of defining features had to be determined. The seven segments from which the numerals were constructed (see Figure 1) were used as the feature set. Two additional global features—digit open to the right (e.g., 6) or left (e.g., 9)—were added, thus totaling nine features. The choice of the seven segments and the additional two global features is to some extent intuitive and hence justified mainly in terms of face validity. Thus, the existence of a more adequate set of features cannot be ruled out.

Two versions of the model were tested; these will be referred to as Versions 1 and 2. Version 1 included 11 parameters that had to be estimated: The salience of the nine features, one parameter for common features (θ) and one for the distinguishing features (α). The reason for using a single parameter α for all features was to economize in terms of parameters that have to be estimated. Version 2 was identical to Version 1 except that a different parameter, α_i, was attached to each of the features, so that the total number of parameters to be estimated in this case was 19. The same criteria of model performance used for the choice model were applied to both versions, and are summarized in Table 2a.

The fit between the theoretical values as determined by the proposed model—for both Version 1 and Version 2—and the observed data was indeed encouraging and of the same magnitude as the choice model. The major advantage of our model is that the high degree of agreement between the theoretical and obtained data is achieved by estimating a number of parameters that is substantially smaller than the corresponding number used in the choice model.

Additional insight into the proposed model (Version 1) can be obtained by inspection of the particular estimated values for each parameter, f( ), reflecting the relative importance of each segment. The parameter values for the seven segments (numbered in accordance with Figure 1) were: f(1) = .997, f(2) = 1.675, f(3) = .715, f(4) = 2.060, f(5) = 1.703, f(6) = .524, f(7) = 2.084. Note that the lowest weight is given for segment 6, which indeed is the least informative (or the most redundant) segment, since it is contained by all the numerals except the numeral 2. Segments 1 and 3 have also relatively low weights, probably because each is contained in all but two numerals. A plausible implication of these results is that the perceptual system—at least under the present simplistic conditions, and perhaps because of them—is highly sensitive to redundancy and is able to tune itself toward the more informative segments. In summary, segments 4 and 7 seem to be the most conspicuous ones, while segments 6 and 3 are the least.

The estimate for θ (the parameter for common features) was $-.436$. The fact that θ turned out to be negative is not surprising, since common features would tend to increase similarity (and the probability of confusions) and hence decrease the dissimilarity measure. Since θ is not 0, it also implies that minimality does not necessarily hold. The estimate for the single parameter α was .102, implying that symmetry does not hold, at least for the present data.

Table 2
Values of Two Criteria, the Optimization Criterion (Equation 8) and Pearson's r, Used to Evaluate the Fit of the Different Models Applied to (a) Segmented Materials and (b) Letters, Based on the Confusion Matrix Reported by Gilmore et al. (1979)

	Entire Matrix		Diagonal Values		
	Optimization Criterion	Pearson's r	Optimization Criterion	Pearson's r	Number of Parameters
			(a) Numerals		
Choice Model	.009	.995	.012	.994	54
Proposed Model–Version 1	.011	.994	.117	.983	11
Proposed Model–Version 2	.006	.997	.048	.991	19
			(b) Letters		
Choice Model	.011	.995	.013	.994	350
Proposed Model	.115	.936	.261	.917	29

The parameter estimates of the scale f under Version 2 were very similar to those obtained under Version 1. The different α estimates for the first seven features (i.e., the seven line segments), express the ratio of the likelihoods of two different operations or processes: that of adding (or inventing) a particular line segment that was not physically presented and that of deleting (or missing) a segment that was physically presented. $\alpha > 1$ implies that, on the average, it is more likely that the segment will be added or invented, while $\alpha < 1$ implies that it is more likely that a segment will be missed or deleted. If $\alpha = 1$, then each of the two processes is equally likely to occur and, consequently, symmetry is implied. None of the seven estimated αs was equal to 1 (which is congruent with the nonsymmetrical nature of the data), and most of the αs (but not all) were smaller than 1, thus providing some indirect support for the conclusion reached by Van Nes and Bouma (1980), namely, that "a digit not correctly recognized is more often perceived as one with a configuration simpler than that of the presented digit than the other way around" (p. 469).

In summary, the proposed model stood up to its first test and performed rather well. In fact, we argue that our model has outperformed the choice model, for, despite the fact that both models performed equally well in terms of goodness of fit, our model had two advantages over the choice model: (1) It has achieved the same level of goodness of fit by using only a fifth (Version 1) or a third (Version 2) of the number of parameters used by the choice model. (2) Those parameters are interpretable and provide some insight into plausible underlying cognitive processes rather than merely making predictions. We now turn to a more stringent test of the model by applying it to alphabetic confusion data. This test is more stringent in two respects: The number of cells to be predicted is increased from 90 to 650, and there is no readily available feature set that will define the stimuli.

Interletter Confusion Matrices

Unfortunately, there are few interletter confusion matrices reported in the current literature. We are aware of only two such reports on capital English letters, those by Gilmore et al. (1979) and Townsend (1971). The data reported by Townsend are unfortunately based on a relatively small number of observations, which casts some doubts on their reliability. In addition, and as a consequence, the confusion matrix contains a large number of empty cells. Accordingly, we decided to use the data reported by Gilmore et al. (1979) for testing our model. It is important to note that Gilmore et al. (1979) used stimuli that were composed of bright dots presented on a computer-controlled cathode-ray tube (CRT) and that letters generated by such a system tend to be rather different from the typographically produced letters employed by Townsend (1971).[9] Hence, the following discussion is restricted to that particular set of letters. Some comments concerning generality are postponed for the final discussion.

The first step for the present application was a search for a satisfactory set of defining features. Previous researchers who have used different multidimensional scaling methods all report the familiar difficulty that the dimensions isolated were not identifiable. Townsend (1971) suggested that "this result points up the difficulty inherent in using scaling techniques as a detective device unless the scaled dimensions either are very elementary and obvious or unless they turn out to be equivalent to some previously hypothesized dimensions" (p. 47). Similarly, Gilmore et al. employed both the multidimensional scaling procedure TORSCA (Young & Torgerson, 1967) as well as a hierarchical clustering scheme (Johnson, 1967), and commented that "we are faced with a situation where a perceptual structure can be specified but the physical dimensions related to it cannot be labeled" (p. 430). Our approach to the problem is somewhat different, since we con-

sider multidimensional scaling to serve as an exploratory tool rather than as an end in itself. In other words, the role of multidimensional scaling in the present context is to aid in identifying an optimal defining feature set, which will then serve as an input for the model. Consequently, we do not necessarily have to adhere strictly to results obtained from multidimensional scaling analysis, and thus maintain the liberty for minor modifications which will yield interpretable dimensions and a well-defined feature list.

We have tried several multidimensional techniques and have concluded that additive trees—a hierarchical clustering scheme recently proposed by Sattath and Tversky (1977)—was the most appropriate analysis for our purposes: It produced an excellent fit, as measured by Kruskal's (1964) stress −.034, and at the same time yielded a representation that was readily interpretable. A set of 14 features composed on the basis of the results from the output of ADDTREE (the computer program for additive trees—for a detailed description see Sattath & Tversky, 1977) is presented in Table 3 (note that this feature list applies only to the specific letters used by Gilmore et al., 1979, and may not necessarily generalize to other type fonts). Table 3 also provides the definition of each letter in terms of the 14 features. There is a noticeable and important difference between previous feature sets that appear in the literature (Geyer & Dewald, 1973; Gibson, 1969; Laughery, 1971) and the one proposed here. The former have used mainly what may be labeled "basic features," namely, specific line segments, and have included relatively few "global features" which refer to properties of the entire gestalt, such as openness, symmetry, etc. In contrast, the list proposed in Table 3 is mainly composed of global features and relatively few basic features. We speculate that this feature list reflects more accurately the nature of the perceptual process where global features play the more important role.

An important distinction should be made between (at least) two possible interpretations of a feature list. In one context, the defining feature list may be derived directly from the subject by simply asking him to describe each letter, feature by feature. Such a procedure would yield what we may label a *descriptive feature list*, which is generated consciously and is likely to consist mainly of "basic features." In a different context, one may have an interest in the feature list used by the perceptual system for recognition purposes, as, for instance, in reading or in tachis-

Table 3
Feature List Used for Analyzing the Data of Gilmore et al. (1979)

	Feature*													
Letter	1	2	3	4	5	6	7	8	9	10	11	12	13	14
A			+	+	+			+				+		
B	+	+		+	+				+		+			
C	+								+		+			+
D	+	+							+		+			
E	+			+							+			+
F	+			+										
G	+			+					+					+
H			+	+			+	+			+	+		
I						+					+	+		+
J						+			+					+
K	+						+	+		+	+			
L	+					+								+
M			+					+				+	+	
N			+				+	+					+	
O		+							+		+	+		
P	+			+	+				+					
Q		+							+					
R	+			+	+			+	+	+				
S	+								+	+				+
T						+						+		
U			+				+		+			+		
V							+			+		+	+	
W			+				+					+	+	
X							+	+		+	+	+	+	
Y						+	+			+		+	+	
Z										+				+
f( )	.99	1.36	2.53	.34	.80	2.36	.32	.30	1.34	2.94	1.21	1.01	.44	1.75

**(1) Facing to the right, (2) whole figure closed, (3) parallel vertical lines, (4) horizontal line in center of letter, (5) upper part of letter-closed element, (6) single straight vertical line, (7) open toward top, (8) open toward bottom, (9) circular elements, (10) diagonal lines (excluding those that are between parallel straight lines), (11) horizontal symmetry, (12) vertical symmetry, (13) an open angle of approximately 45 deg in a vertical position (Λ or V), (14) nonclosed letters standing on a broad base.*

toscopic recognition experiments. The list in this context may be termed a *recognition feature list*; and it can be derived only by indirect methods—some of which were discussed briefly above—since it is presumably unconscious. The point to be emphasized here is that these two feature lists are not necessarily identical, and are most probably not. We suspect that some of the difficulties, mentioned earlier, in revealing the *recognition feature list* stem from the fact that it is apparently unconscious, and from the tendency of many researchers to treat it as if it were a *descriptive feature list.*

The confusion matrix reported by Gilmore et al. (1979) and the feature list in Table 3 were used as input to our model. In the present application, we assigned a different α—the distinguishing features parameter—to each of the 14 features and added one parameter for the common features (θ). Thus, the parameters to be estimated totaled 29.

The measures of fit were the same as those used for the segmented numerals employed in the present application and are summarized in Table 2b, along with the values obtained for the performance of the choice model as applied to the same data. In terms of goodness of fit, the choice model outperforms our model by any of the measures used. That observation, however, is not surprising, given that the numbers of parameters used by the two models differ by a factor of 12! Thus, it is questionable whether the performance of the choice model that is utilizing 350 parameters should be used as a standard or base rate. At any rate, there is a high agreement between the observed and predicted values for the proposed model (as indicated from the values in Table 2b), even though it is somewhat below the corresponding agreement for the choice model. Additional precision for the proposed model could have been obtained simply by adding additional features, though probably those would be of decreasing importance. For purposes of parsimony, it seemed preferable to use only the 14 main features that resulted from the additive tree analysis.

The different estimated values for the f scale of the 14 features are given at the bottom of Table 3. They indicate the relative salience or importance, of the information value carried by the different features. In addition, estimates were obtained for each of the 14 αs corresponding to the 14 features. Not all the features are described in terms of concrete physical segments (e.g., symmetry, figure facing to the right, etc.), but the interpretation of the α is rather similar: $\alpha > 1$ implies that it is more likely that the feature will be added to a character rather than deleted from a character where it was actually present; $\alpha < 1$ implies the converse. None of the estimated αs was equal to 1, thus implying that none of the features is perfectly symmetrical. As in the case of numerals, a majority of the estimated αs were smaller than 1.0 and none was equal to 1.0.

The estimated value of the parameter of the common features (θ) was $-.489$. Incidentally, the relatively small values of θ obtained for both the numeral and letter data suggest that common features play a minor role compared with distinguishing features. There are some other studies which have implicitly suggested a similar conclusion (e.g., Medin & Schaffer, 1978; Tversky, 1977). The question of the relative importance of common vs. distinguishing features in similarity judgments and confusion matrices is beyond the scope of the present paper, but is an important empirical question which should be further investigated.

GENERAL DISCUSSION

The results described in the previous sections, in which the two models—the choice model and the one we propose—were applied to two different sets of data, support our contention about the superiority of our model. The choice model performed slightly better than our model in terms of goodness of fit, a result which was to be expected in light of the number of parameters used by each model: The choice model used a larger number of parameters by a factor of 5 (or 3) for the numeral data and by a factor of 12 for the letter data. It was also demonstrated that the few parameters used in our model are readily and naturally interpretable and may thus provide a first insight into underlying processes of similarity judgments. Finally, our model makes minimal assumptions with regard to the nature of the data. In contrast, the choice model, like many other geometrical models, assumes minimality, symmetry, and the triangle inequality, which in our model are tested rather than assumed.

That is not to say that the proposed model is without problems. As mentioned, the major difficulty in using the model (which applies to Tversky's contrast model in general) is associated with the obstacles involved in revealing the appropriate feature structure, which usually cannot be readily specified. At present, this problem can be partially solved by the use of multidimensional techniques and intuition, but other possibilities, such as experimental work, should not be excluded. Once a satisfactory feature set is found, as was the case with segmented numerals, the model is extremely powerful both in explanatory power (e.g., by assessing the relative importance of each feature as well as by describing the process by which information regarding different features is aggregated) and in terms of goodness of fit.

A final comment that we would like to make, one which is actually beyond the scope of the model, concerns the nature of the existing data on alphanumeric characters, in particular with regard to their reliability. It is somewhat disturbing that confusion matrices of both uppercase letters (Townsend, 1971, Conditions I and II; Gilmore et al., 1979) and seg-

mented numerals (Van Nes & Bouma, 1978, Conditions I, II, III; Table 1 in the present paper) reported in the literature show very little resemblance among themselves. There are, of course, several plausible reasons for that state of affairs: One may be the typographical differences in the stimuli used (e.g., Townsend vs. Gilmore et al.). Another could be the particular experimental conditions in which the data were collected (e.g., Townsend Condition I vs. Case II; Van Nes & Bouma, Condition I vs. Condition II vs. Condition III). The importance of such factors has been recently emphasized by Garner and Haun (1978), who have distinguished between a state limitation, in which a stimulus does not adequately get into or through the organism, and a process limitation, which occurs when stimuli or their features are not adequately differentiated. Finally, the data may simply be unreliable because of an insufficient number of observations (which may be the case in the Townsend, 1971, and the Van Nes & Bouma, 1980, data) or because of inappropriate experimental conditions (e.g., the comment by Mewhort and Dow, 1979, who assert that similarity may be confounded with brightness in the data reported by Gilmore et al.).

Obviously, researchers should make all possible effort to avoid unreliability which is due to either insufficient observations and/or possible confoundings. On the other hand, the physical properties of the stimuli, that is, typographical differences and the type font used as well as the nature of conditions by which data are collected, should be subjects for future research. For instance, current proposals for feature lists are limited in their applicability to a given font type. Future investigations should be addressed to the possibility of revealing a set of defining features that underlies different (though similar) type fonts. Naturally, such an undertaking would require a large investment in data collection.

It would also be of interest to find out how performance is affected by different physical experimental conditions, given a particular type font, that is, to investigate the effect of state vs. process limitation on the obtained confusion matrix. The work of Garner and Haun (1978) is a first step in this direction, but additional research is obviously needed. In that regard, the proposed model may be a useful tool: Since, in such experiments, the stimulus population and consequently the defining feature list would stay unchanged from one condition to the other, the model can be applied to different sets of data obtained under different experimental conditions. Studying the different weights for features, common (θ) and distinct (α), which the model will assess for those various conditions may provide preliminary cues for the differences underlying state vs. process limitation conditions. We are currently exploring some of these issues and will report our findings at a later stage.

REFERENCE NOTES

1. Goldstein, L. M. *Three studies in speech perception: Features, relative salience, and bias* (*Working Papers in Phonetics*, October 1977). Los Angeles: UCLA, 1977.

2. Chandler, J. P. *Subroutine Stepit* (Program OCPE66, Quantum Chemistry Exchange). Bloomington: Indiana University, 1965.

REFERENCES

GARNER, W. R., & HAUN, F. Letter identification as a function of type of perceptual limitation and type of attribute. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, **4**, 199-209.

GEYER, L. H., & DEWALD, C. G. Feature lists and confusion matrices. *Perception & Psychophysics*, 1973, **14**, 471-482.

GIBSON, E. *Principles of perceptual learning development*. New York: Appleton-Century-Crofts, 1969.

GILMORE, G. C., HERSH, A., CARAMAZZA, A., & GRIFFIN, J. Multidimensional letter similarity derived from recognition errors. *Perception & Psychophysics*, 1979, **25**, 425-431.

HOLBROOK, M. B. A. A comparison of methods for measuring the interletter similarity between capital letters. *Perception & Psychophysics*, 1975, **17**, 532-536.

HOOK, R. T., & JEEVES, T. A. "Direct search"—solution of numerical and statistical problems. *Journal of the Association for Computing Machinery*, 1961, **8**, 212-229.

JOHNSON, S. C. Hierarchical clustering schemes. *Psychometrika*, 1967, **32**, 241-254.

KRUSKAL, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 1964, **29**, 1-27.

LAUGHERY, K. R. Computer simulation of short-term memory: A component decay model. In G. T. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 6). New York: Academic Press, 1971.

LUCE, R. D. *Individual choice behavior*. New York: Wiley, 1959.

LUCE, R. D. Detection and recognition. In R. D. Luce, R. R. Bush, & S. E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1). New York: Wiley, 1963.

MEDIN, D. L., & SCHAFFER, M. M. Context theory of classification learning. *Psychological Review*, 1978, **85**, 207-238.

MEWHORT, D. J. K., & DOW, M. L. Multidimensional letter similarity: A confound with brightness? *Perception & Psychophysics*, 1979, **26**, 325-326.

SATTATH, S., & TVERSKY, A. Additive similarity trees. *Psychometrika*, 1977, **42**, 319-345.

TOWNSEND, J. T. Theoretical analysis of an alphabetic confusion matrix. *Perception & Psychophysics*, 1971, **9**, 40-50.

TVERSKY, A. Features of similarity. *Psychological Review*, 1977, **84**, 327-352.

TVERSKY, A., & GATI, I. Studies of similarity. In E. Rosch & S. B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, N.J: Erlbaum, 1978.

VAN NES, F. L., & BOUMA, H. On the legibility of segmented numerals. *Human Factors*, 1980, **22**, 463-474.

YOUNG, F. W., & TORGERSON, W. S. TORSCA: A FORTRAN IV program for Shepard-Kruskal multidimensional scaling analysis. *Behavioral Science*, 1967, **12**, 498.

NOTES

1. More precisely, since subjects' responses are basically different from the physical stimuli, the main diagonal entries represent the extent to which the physical stimuli and the responses are congruent in the sense that they refer to the same thing. In all of the following discussion, it is important to distinguish between the physical stimuli and the response system as they occupy different domains (see also Footnote 6). In terms of notation, the

first subscript will always refer to the stimulus domain and the second subscript to the response domain.

2. Alternative solutions to the estimation problem have been developed by Don Bouwhuis and the present authors, and can be obtained from us upon request.

3. This distinction is analogous to the one that is made within the framework of signal detection theory between perceptual sensitivity and an independent decision-making process.

4. The correlation was calculated in the following manner: For each letter in the similarity matrix (Table 2 in Gilmore et al., 1979), we added the elements of the corresponding row (excluding the self-similarity, which for each letter is 1.0). We thus obtained a vector containing the 26 sums corresponding to each letter. The correlation between this vector and the response bias vector yielded $r = -.733$.

5. Inspection of the data by Van Nes and Bouma (1978) as well as our own data using similar numerals (see Table 1) suggests several violations of that rule.

6. Since our dissimilarity measure is not necessarily symmetrical, we explicitly distinguish between, for instance, letters as stimuli and letters as possible responses, which occupy different domains in the pyschological space.

7. Equation 6 was constructed such that it will contain certain desired properties, in particular the independence of features, which consequently requires additivity in the distance space and multiplicity in the similarity space. A detailed account for the construction of Equation 6 can be obtained from the authors upon request.

8. Recently, the use of this type of numerals has continuously increased and has been adopted by several industries in such items as calculator displays, digital watches, meters at gasoline pumps, etc.

9. After completion of this project, we became aware of a recent paper by Mewhort and Dow (1979), who argued that the results of Gilmore et al. (1979) might be confounded with brightness. We do not exclude such a possibility, but even if Mewhort and Dow's claim is correct, it should have little effect on our results, and in particular should not affect the comparison with the choice model.

(Received for publication June 20, 1980;
revision accepted December 17, 1980.)