

## Pattern Recognition and Categorization<sup>1</sup>

STEPHEN K. REED

*Case Western Reserve University*

Four experiments are reported which attempt to determine how people make classifications when categories are defined by sets of exemplars and not by logical rules. College students classified schematic faces into one of two categories each composed of five faces. One probability model and three distance models were tested. The predominant strategy, as revealed by successful models, was to abstract a prototype representing each category and to compare the *distance* of novel patterns to each prototype, emphasizing those features which best discriminated the two categories.

Bruner (1957) assumed that all perceptual experience is the end product of a categorization process. Whatever is perceived achieves its meaning from a class of percepts with which it is grouped. By a category he meant a rule for classifying objects as equivalent. The rule must specify the critical properties of the stimuli, the manner in which these are combined, the weights assigned various properties, and the acceptance limits for each category. Bruner also believed that there was no reason to assume that laws governing inferences of this kind were discontinuous as one moved from perceptual to more conceptual activities. In contrast, Neisser (1967) has argued that some kinds of perceptual experience (such as visual tracking) may be more analogue than categorical. None the less, other forms of perception are clearly categorical since pattern recognition primarily involves the classification of stimuli into mutually exclusive categories.

The purpose of this article is to investigate how classification is accomplished. Given two classes of patterns defined by sets of exemplars, we may ask what kind of decision processes do Ss use to classify new pat-

<sup>1</sup>This article is based on a dissertation submitted to the Graduate School, University of California, Los Angeles, in partial fulfillment of the requirements for the PhD degree. I would like to thank my Chairman, Professor Morton P. Friedman, for his guidance and encouragement throughout my graduate study and the members of my Committee for their suggestions concerning the improvement of the dissertation. Daniel Rourke's contribution to the scaling study was greatly appreciated. This research was supported in part by an NSF Traineeship and by NIMH grant MH-07809 to Professors E. C. Carterette and M. P. Friedman.

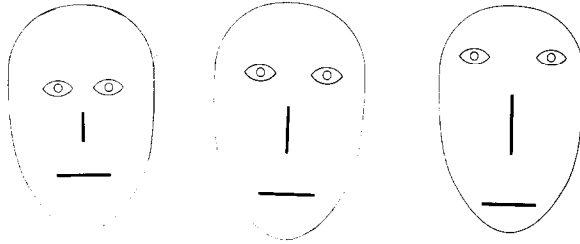


FIG. 1. Examples of schematic faces representing the three values of each dimension.

terns. In answer to this question, the formulation of a decision rule should specify Bruner's four criteria.

Since it is often difficult to specify the critical features, patterns selected for the present research were constructed to vary along a well-defined set of dimensions. Figure 1 shows the stimulus patterns used. The patterns consisted of Brunswik faces which differed in the height of the Forehead, the distance between the Eyes, the length of the Nose, and the height of the Mouth. Each feature could take on one of three values, as illustrated by the three examples. Classification problems always had two categories and five faces always belonged to a category. Category faces were selected randomly under the constraint that the two categories could be separated by a linear discriminant function (Nilsson, 1965). The constraint of linear separability provided the basis for a "category" in that the two categories occupied different areas in a multidimensional space.

Figure 2 shows an example of a classification problem. The category faces were either all presented simultaneously to Ss for their viewing throughout the experiment (Experiments I and II) or had to be initially learned in a standard paired-associates experiment (Experiment III).

We shall focus on the rules used by Ss in classifying new stimulus patterns which do not match any of the category patterns. A number of different theories are formulated as mathematical models to determine which could best predict Ss' classifications. Two general types of models are formulated: probability and distance models. A review of previous studies using such models and a discussion of their relevance to feature and template theories is contained in Reed (1970).

Each model assumes that a pattern can be represented in terms of a list of feature values. Let  $X = (x_1, x_2, \dots, x_d)$  be a pattern in which each  $x_m$  stands for a particular value of the  $m$ th feature. For example, the left face shown in Fig. 1 would be represented by  $X = (\text{low forehead, narrow eyes, short nose, high mouth})$ . Each value of a particular feature (hereafter a "cue") has a certain validity with respect to each category.

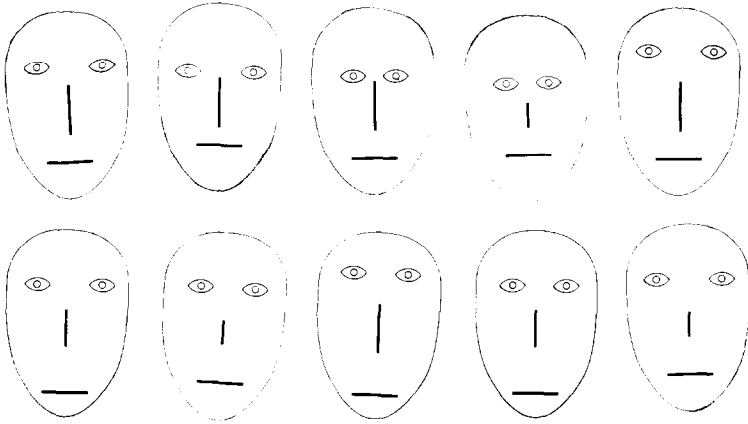


FIG. 2. Composition of the two categories used in problem 1: (upper five faces) Category 1; and (lower five faces) Category 2.

In the case of two categories, the validity of cue  $x_m$  will be  $p(\text{Category 1}|x_m)$  or  $p(\text{Category 2}|x_m)$  depending on which category is chosen. Since the sum of the two conditional probabilities is unity, a cue having a high validity with respect to one category has to have a low validity with respect to the other category. The cue "long nose" may have high validity with respect to Category 1 but a low validity with respect to Category 2 (see Fig. 2 for an example). The probability of a category as a function of the different cue values is the essence of *probability* models.

A second general class of models are *distance* models. An important problem is to define the circumstances under which we can expect to maximize differences in the predictions of the two models. Imagine a category of five faces consisting of three faces with a short nose, one face with a medium nose, and one face with a long nose. The issue is whether people would more likely consider a face with a medium nose as a member of this category than a face with a long nose. According to a probability model, no distinction is made between the two faces since the probability of a medium nose equals the probability of a long nose. According to a distance model, however, a face with a medium nose would more likely belong to that category than a face with a long nose since the average distance between that feature and the other patterns is less in the former case. In other words, the concept of stimulus generalization is incorporated into a distance model.

A method of converting similarity measures into distance measures has been devised by Shepard (1962a, 1962b) and was later modified by Kruskal (1964). Shepard proposed that distance and similarity should be monotonically related such that the greater the similarity of two

stimuli the smaller should be the distance between them in some multi-dimensional space. The exact form of the function relating distance and similarity is not specified *a priori*, but is found empirically for each set of stimuli. Distances are represented by the Minkowski  $r$ -metric given in Eq. 1.

$$d_{ij} = \left( \sum_{m=1}^d |x_{im} - x_{jm}|^r \right)^{1/r}, \quad (1)$$

where

$d_{ij}$  distance between stimuli  $i$  and  $j$   
 $m$  subscript for an orthogonal axis of the space ( $m = 1, 2, \dots, d$ )  
 $x_{im}$  projection of stimulus  $i$  on axis  $m$

Two popular spatial models have been the city block model with  $r = 1$  and the Euclidean model with  $r = 2$ . The city block model is applicable when the stimuli differ with respect to obvious and compelling dimensions such as size and brightness (Attneave, 1950). In this case a subject's ratings might behave as if they were the sum of the differences on the separate dimensions. The Euclidean metric, on the other hand, is more applicable when the separate dimensions are not obvious and the subject is more likely to judge the overall distance directly. Both models are used in formulating possible classification strategies. A more detailed statement of the various models follows, and a mathematical formulation is given in the appendix.

### *Cue Validity Model*

The specific probability model considered here will be called a *cue validity model* and is similar to one proposed by Beach (1964a, 1964b). According to the model, the  $S$  calculates the average cue validity for each category and assigns the pattern to the category with the highest average cue validity. We will consider a special case of the cue validity model in which the number of cues being compared across the two categories varies from one to  $d$ . In the one cue case, the cue with the highest validity in favor of Category 1 is compared with the cue with the highest validity for Category 2. In the two cue case, the two cues with the highest validity in favor of Category 1 are compared with the two cues with the highest validity in favor of Category 2. In the  $d$  cue case, all cues are compared.

### *Proximity Algorithm*

The remaining models are formulated as classification models based on the perceived distance between pairs of patterns. We assume that each

pattern can be represented as a point in a multidimensional space, with each axis of the space corresponding to a feature of the pattern. Let us denote the patterns in Category 1 by  $X_{1i}$ ,  $i = 1, N_1$ , and the patterns in Category 2 by  $X_{2i}$ ,  $i = 1, N_2$ .  $N_1$  is the total number of patterns in Category 1 and  $N_2$  is the total number of patterns in Category 2. The S's task is to classify new patterns,  $X_j$ . The distance models represent possible decision strategies that S might use to carry out the task.

If the patterns are thought of as being represented in a multidimensional space, the first procedure simply counts the number of patterns in the immediate area of the test pattern. If the majority of the patterns belong to Category 1, the test pattern is placed in Category 1. If the majority of the patterns belong to Category 2, the test pattern is placed in Category 2. The simplest rule in this class is the *proximity algorithm* which states: decide that a test pattern belongs to Category 1 if it is nearer to the nearest known member of Category 1 than to the nearest known member of Category 2.

One would expect such a decision rule to produce desirable results when the members of the different categories occupy widely separated regions in the multidimensional space. The weakness of such a classification rule is that a single, stray pattern could result in a large number of misclassifications. The natural extension is to base the decision rule on the nearest  $k$  patterns to the test pattern. Nilsson (1965) calls this rule the Fix and Hodges method (Fix and Hodges, 1951). The test pattern is classified by first pooling the patterns in the two training subsets and finding those  $k$  patterns which are closest to  $X_j$ . The decision rule is based on whether the majority of these patterns belong to Category 1 or 2. The test pattern is classified into Category 1 if the majority of patterns in the subset belong to Category 1 and into Category 2 if the majority of patterns in the subset belong to Category 2. Subset sizes of  $k = 1, 3$ , and 5 are considered in testing the model.

#### *Average Distance Model*

An *average distance model* assumes that Ss utilize all the patterns in each category in making their decisions instead of just a local subset. The rule states, "decide a pattern  $X_j$  belongs to Category 1 if the average distance of  $X_j$  from members in Category 1 is less than or equal to the average distance of  $X_j$  from members in Category 2. Otherwise place  $X_j$  in Category 2."

#### *Prototype Model*

A similar formulation is a *prototype* model. The prototype ( $P$ ) of a category represents its central tendency and can be defined as that

pattern which has for each component  $x_m$  the mean value of the  $m$ th component of all other patterns in that category. Thus,  $P = (\bar{x}_1, \bar{x}_2, \dots, \bar{x}_d)$ . The decision rule is to "decide a pattern  $X_j$  belongs to Category 1 if the distance of  $X_j$  from the Category 1 prototype is less than or equal to the distance from the Category 2 prototype. Decide  $X_j$  belongs to Category 2 otherwise."

The above decision rule is not mathematically identical to the Average Distance decision rule. It makes a difference whether one calculates the average distance to the patterns in a category or the distance to the average pattern in a category. A simple illustration is the case of two one-dimensional patterns represented on a straight line a distance  $d$  apart. The prototype  $P$  is the midpoint of the line. If a new pattern with value  $P$  is now presented, the distance from the prototype is zero but the average distance from the two patterns is  $d/2$ .

### *Weighted Features*

The final two models are a restatement of the average distance model and the prototype model except for the differential weighting of features. The basic idea is that some features are more important than others in distinguishing between categories. The weighted features-distance models thus combine aspects of both the cue validity and distance models. We use a normative approach to find the weights since estimation would require a number of parameters for complex stimuli making it difficult to obtain a unique solution.

A normative approach for estimating weights has been discussed by Sebestyen (1962). Sebestyen begins by calculating the weights so as to maximally cluster patterns within a category; the less the variability along a dimension, the greater that dimension is weighted. The problem with such an approach is that a feature may have low variability within a category and yet it may not be very useful in distinguishing between classes. A more reasonable approach makes use of class separating transformations (Sebestyen, 1962, pp. 37-43).

The optimal transformation to be considered here is one that not only clusters patterns of the same class, but also separates those that belong to different classes. The problem is to find the linear transformation that maximizes mean-square between set distances while holding constant the sum of mean-square between set and within set distances. The set of weights which satisfy this transformation are used to differentially weight the dimensions when calculating distances. The psychological implications of the model are that Ss differentially emphasize those features of a pattern which best enable them to discriminate among different categories. The procedure for calculating the weights and a

mathematical formulation of the *weighted features average distance model* and the *weighted features prototype model* are given in the Appendix.

Altogether, 18 models were compared. The models can be divided into four general classes: (1) the Cue Validity model, four versions depending on whether the decision is based on one, two, three, or all four cues; (2) the Proximity Algorithm, six versions depending on the metric ( $r = 1$  or  $r = 2$ ) and the size of the subset ( $K = 1, 3, \text{ or } 5$ ); (3) the Average Distance rule, four versions depending on the metric and weighted vs nonweighted features; and (4) the Prototype rule, four versions depending on the metric and weighted vs nonweighted features. Four experiments were conducted to compare the alternative models. Experiment I investigated the classification rules used by Ss when the category patterns were simultaneously present and Experiment II examined the effect of category composition on Ss' decision rules. The effects of memory were tested in Experiment III which required that Ss learn the category patterns in a preliminary paired associates experiment. Since predictions for the first three experiments were based on the physical distance between patterns, Experiment IV involved determining whether the use of psychologically scaled distances would improve the fit of the models.

## EXPERIMENTS I AND II

### *Subjects and Procedure*

Experiments I and II are presented together since the procedure was nearly identical in the two experiments. Forty-seven Ss served in Experiment I as part of a requirement for an introductory psychology course. The experiment was run in a room which enabled as many as 10 Ss to participate in a single session. Ten  $2 \times 3$  in. photographs of the schematic faces were Xeroxed on a sheet of paper, with the upper row of five faces labeled "Category 1" and the lower row of five faces labeled "Category 2." The category faces were identical to the ones shown in Fig. 2. Each S was given a sheet of paper containing the category faces, which remained in front of him throughout the experiment. They were instructed that they would be shown faces similar to these on the screen at the front of the room and that their task was to decide whether each new face belongs to category one or category two. The Ss had 2 min to study the faces in the two categories and were then shown 10 practice faces. The experiment required that they classify 25 faces, chosen to distinguish among alternative models. Each face was shown on the screen for 12 sec so that Ss had 12 sec to make each decision.

Experiment II was an attempt to discover whether the results of Experiment I were due to the particular composition of the categories. The design of the experiment was like that of Experiment I except that different patterns composed the two categories, and Ss were given two problems instead of one. The faces presented in the two problems are shown in Reed (1970). Fifty-two Ss served in the experiment. Half of the Ss were given Problem 2A first and half were given Problem 2B first. Each problem consisted of classifying 10 practice patterns followed by 24 test patterns. The first 20 test patterns were used to evaluate the alternative models. The last 4 test patterns consisted of the prototype pattern for each category and a corresponding control pattern whose average distance from the five faces in Category 1 and the five faces in Category 2 was approximately the same as the prototype pattern. Thus the Prototype model would predict that more Ss should correctly classify a prototype pattern than the corresponding control pattern, while the Average Distance model would predict that an equal number of Ss should correctly classify the two patterns (cf., Posner and Keele, 1968). After the experiment, Ss answered questions about how they classified the patterns and what features they emphasized.

### *Results and Discussion*

The classification predictions of each of the models were calculated using the feature values (physical measurements) of the test patterns. The alternative models were compared by calculating for each model the percentage correct predictions as to how Ss would classify the patterns. In making predictions for an entire group of Ss, the best any model can do is to always pick for each test pattern the category selected by the majority of Ss (Hunt, Marin, and Stone, 1966). This upper bound is therefore determined by how well Ss agree in making their classifications. For Problem 1, the best predictor would correctly predict 73% of Ss' classifications by always choosing that category selected by the majority of Ss. Since a detailed presentation of the results is given in Reed (1970), only a summary of the main findings is presented here. The clearest aspect of the results was the superiority of the distance models over the probability model. Percentage of correct predictions of the four versions of the Cue Validity model were all less than the 50% chance level. In contrast a number of distance models come close to the upper bound of 73%. The Average Distance and Prototype models did better than the Proximity Algorithm indicating that Ss were utilizing most of the available information rather than basing their decisions on a local subset of patterns.

The most successful model was the Weighted Average Distance model



TABLE 1  
Best Fitting Models for Each of the Four Problems

	<i>p</i>
Problem 1	
(1) Weighted Average Distance ( $r = 1$ )	—
(2) Weighted Average Distance ( $r = 2$ )	.80
(3) Weighted Prototype ( $r = 1$ )	.30
(4) Average Distance ( $r = 2$ )	.28
Problem 2A	
(1) Weighted Average Distance ( $r = 1$ )	—
(2) Weighted Prototype ( $r = 1$ )	.84
(3) Cue Validity (one cue)	.71
(4) Prototype ( $r = 2$ )	.07
Problem 2B	
(1) Cue Validity (one cue)	—
(2) Proximity Algorithm: $k = 1$ ( $r = 1$ )	.22
(3) Prototype ( $r = 2$ )	.06
Problem 3	
(1) Weighted Prototype ( $r = 1$ )	—
(2) Weighted Average Distance ( $r = 1$ )	.23
(3) Cue Validity (one cue)	.21

( $r = 1$ ) which correctly predicted 70% of *Ss* classifications. This model was compared with the alternative models to determine which models made significantly worse predictions. The normal deviate  $z$  was calculated to test for the significance between two correlated proportions. Table 1 shows the models which could not be rejected at the  $p = .05$  level as predicting significantly worse than Weighted Average Distance model. The models are ordered according to their success and the probability values corresponding to the obtained  $z$  score is shown for each model. With one exception, all other models could be rejected at the  $p = .001$  level.

Table 1 also shows the best predicting models for Problems 2A and 2B. The results obtained for Problem 2A are very similar to the results obtained for Problem 1 in that the Weighted Average Distance and Weighted Prototype models again made the best predictions. With one exception, all models not listed in Table 1 could be rejected at the  $p = .001$  level. The results for Problem 2B were quite different in that the one cue Cue Validity model and the Proximity Algorithm (subset size = 1) made the best predictions. These two models have in common the fact that they require *Ss* to integrate very little information in making their decision. The *Ss* using the one cue decision rule would have to base their

decision on the single, most valid cue whereas Ss using the Proximity Algorithm would have to base their decision on the single category pattern which best matched the test pattern. A possible post-hoc explanation for these findings can be given by referring to Fig. 3 showing the prototype faces for Problems 2A and 2B. The reader will hopefully agree that the category prototypes are much more similar for Problem 2B than for Problem 2A. The greater similarity of the two prototypes in Problem 2B may have made it difficult for Ss to use a prototype decision rule, causing them to use a decision rule requiring very little integration of information.

This explanation would be more appealing if it could demonstrate that many Ss were actually using a prototype decision rule in Problems 1 and 2A. The results indicated that the Weighted Average Distance and Weighted Prototype models best predicted Ss' classification. Although the two models are not mathematically equivalent, their predictions are highly correlated making it difficult to distinguish between the two models. However, two additional converging operations suggested that a prototype rule was the strategy predominantly used. One way of testing the two models is to determine whether Ss are more accurate in classifying the prototype pattern than a control pattern which is the same average distance from the category patterns (Posner & Keele, 1963). The last four patterns that the Ss classified in Problems 2A and 2B were the two prototype patterns and the corresponding control patterns. The results are shown in Table 2. In all four cases, the prototype was cor-

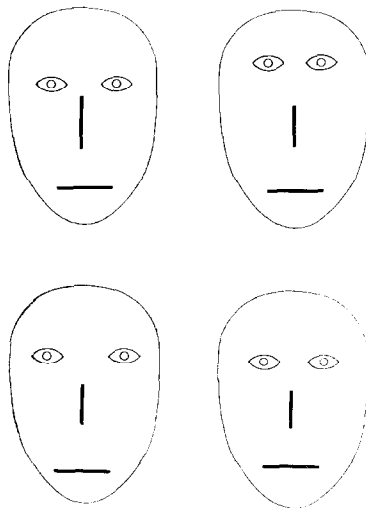


FIG. 3. Prototype faces for Problems 2A (above) and 2B (below).

TABLE 2  
 Percentage of Correct Classifications of the Prototype and Control Patterns

Problem	Pattern	Correct	Difference ( <i>p</i> )
2A	Prototype 1	92%	.02
	Control 1	75%	
2A	Prototype 2	88%	.001
	Control 2	46%	
2B	Prototype 1	67%	.03
	Control 1	50%	
2B	Prototype 2	85%	.01
	Control 2	67%	
3	Prototype 1	94%	.001
	Control 1	62%	
3	Prototype 2	94%	.001
	Control 2	50%	

rectly categorized significantly more often than the control pattern. The Ss were more accurate in classifying the prototypes for Problem 2A, as one would expect since the prototypes in Problem 2B were more similar to each other. These findings lend support to the Prototype model as opposed to the Average Distance model.

Further support for the Prototype model comes from S's verbal reports following the experiment. The models were translated into verbal descriptions and Ss were asked to choose which description best described the strategy they used in solving the two problems. The verbal descriptions corresponded to the Cue Validity model, the Prototype model, the Average Distance model, and the Proximity Algorithm. Table 3 lists the strategies and the percentage of Ss who said they used each strategy. The Ss could also list alternative strategies and two Ss indicated that they based their decisions in Problem 2A on the distance between the eyes and nose.

The verbal reports give additional support to the Prototype model as 37% of the Ss claimed they based their decisions on an abstract image while only 11% of the Ss claimed that they compared the test patterns to all the faces in the two categories. Five Ss reported using different strategies in the two problems and each case involved a shift to an easier strategy in Problem 2B. Perhaps a greater number of Ss would have reported using a different strategy for each problem if Ss had been questioned after each problem rather than at the end of the experiment.

TABLE 3  
 Percent of the Subjects Who Reported Using the Different Strategies  
 in Experiment II and III

Strategy	Experiment (%)	
	II	III
(1) Prototype: I formed an abstract image of what a face in category one should look like and an abstract image of what a face in category two should look like. I then compared the projected face with the two abstract images and chose the category which gave the closest match.	37	58
(2) Proximity Algorithm: I compared the projected face with all the faces in two categories looking for a single face which best matched the projected face. I then chose the category in which that face appeared.	32	10
(3) Cue Validity: I looked at each feature on the projected face and compared how many times it exactly matched a feature in each of the two categories. I then chose the category which gave the highest number of matches.	20	28
(4) Average Distance: I compared the projected face with each of the five faces in Category 1 and with each of the five faces in Category 2. I then chose the category in which the faces were more like the projected face, basing my decision on all faces in the two categories.	11	4

The verbal reports did indicate, however, that there are individual differences in the strategies people use. Although an attempt was made to determine whether the models and verbal reports agreed for individual Ss, the comparison was difficult to make since each S classified only 20 test patterns and usually a number of models did equally well on an individual basis. A possible solution (suggested by the reviewer) to be used in future analyses would be to divide the Ss into groups based upon their verbal reports and test the models separately for each group.

#### EXPERIMENT III

The purpose of the first two experiments was to investigate how Ss would categorize patterns when the exemplars of the two categories were all presented simultaneously to S. A more representative procedure of the pattern recognition process would involve the successive presentation of patterns to Ss so that they would have to learn the patterns belonging to the two categories. The learning procedure would be more

representative in that human pattern recognition presumably develops from the successive learning of individual patterns. The specific question we wish to investigate is whether Ss' decision processes in classifying patterns change as a function of the memory demands of the task.

The additional memory demands imposed by the learning requirement may make it difficult to use certain strategies. The Average Distance rule or the Proximity Algorithm, for example, both require that Ss retain the description of all 10 individual patterns. If Ss could form a prototype for each category, they would have to retain the descriptions of only the two prototypes. Thus the different decision strategies impose different memory demands on the categorizer. In order to compare the success of the models in the learning and nonlearning tasks, the patterns used previously in Problem 2A were used in Experiment 3.

### *Procedure*

Fifty Ss served in the experiment. The Ss consisted of both men and women who were fulfilling an introductory psychology course requirement. The subjects were told that they would be shown slides of 10 schematic faces and that their task was to learn which 5 faces belong to Category 1 and which 5 faces belong to Category 2. The Ss were informed about the features which would vary and they were encouraged to look for similarities among the faces in each category since they would be asked to classify new faces after the learning task. The learning experiment consisted of 1 practice trial and 12 learning trials/pattern. The order of presentation was organized into 12 successive blocks with each block consisting of the 10 faces presented for 6 sec with a feedback light turned on for the last 2 sec. The intertrial interval was 6 sec. A Splindler-Sauppes random access projector was used to present the slides.

The total errors per subject ranged from 2 to 51 with a mean of 18.6 errors/subject. The error probability on trial 12 was 0.11 indicating that learning was not yet complete, but performance was well above the 50% chance level. After the learning phase, Ss classified the 24 test patterns presented in Problem 2A. The test patterns were presented in the same order as in Problem 2A, but were presented for 8 sec instead of 12 sec. The shorter presentation time was used in Experiment III because Ss did not have to consult the sheet of exemplars and therefore made their classifications quite rapidly.

### *Results*

The classification results are based on all 50 Ss as it did not seem necessary to reject Ss because of failure to learn. The most successful

model was the Weighted Prototype model ( $r = 1$ ) which predicted 63% of the classifications out of a possible 68%. All models except the Weighted Average Distance model ( $r = 1$ ) and the one cue Cue Validity model could be rejected at the .05 level for predicting significantly worse than the Weighted Prototype model. It is again necessary to use additional converging operations to distinguish between the Prototype and Average Distance models. It was pointed out earlier that a Prototype decision rule would be easier to implement as Ss would not have to remember all the category patterns as is required in the use of the Average Distance rule. Two sources of evidence support this interpretation. First, as indicated in Table 3, 58% of the Ss reported using the Prototype rule while only 4% reported using the Average Distance rule. Another rule that requires Ss to memorize all the category patterns is the Proximity Algorithm. While 32% of the Ss reported using the Proximity Algorithm in Experiment 2, only 10% of the Ss reported using this strategy in Problem 3. Twenty-eight percent of the Ss reported using the Cue Validity rule and six Ss said they based their decision on the distance between the eyes and nose.

A second converging operation which supported the prototype rule was the finding that Ss were more accurate in classifying the prototype pattern than an equidistant control pattern. The prototype and control patterns were the same as those used in Problem 2A. The category one prototype was correctly classified by 94% of the Ss while the corresponding control pattern was classified by only 62% of the Ss. The category two prototype was also correctly classified by 94% of the Ss while 50% of the Ss correctly classified the control pattern. Both these results are significant at the  $p = .001$  level. Table 2 summarizes these findings.

Another method of comparison is to calculate for each model the rank order correlation between the favorability of Category 1 and the number of Ss choosing category 1. For example, the greater the cue validity is for Category 1, the greater should be the number of Ss choosing Category 1 if that model is appropriate. Alternatively, the closer a test pattern is to Category 1 as opposed to Category 2, the greater should be the number of Ss choosing Category 1 if a distance model is appropriate. The favorability of a category, as determined by the four cue Cue Validity model and the various varieties of the Prototype and Average Distance model, was correlated with the number of Ss choosing that category in each of the problems. The general findings can be summarized in the following way: (1) The distance models generally led to significant, positive correlations with the best correlation being  $r = 0.88$  for Problem 1,  $r = 0.63$  for Problem 2A,  $r = 0.63$  for Problem 2B, and  $r = 0.64$  for Problem 3. (2) With the exception of Problem 2B, the four cue Cue Validity model resulted in negative correlations. (3) No consistent differences

could be found between the two metrics or between the Average Distance and Prototype model. (4) Consistently better correlations did result when the features were differentially weighted. A more detailed presentation of these findings is given in Reed (1970).

The emphasis thus far has been on the relative predictions of the various models and not on the degree of success of the best predicting model. The models, in fact, always fell short of the maximal possible prediction which could have been achieved by classifying each pattern into the category chosen by the majority of Ss. The actual percentage of correct predictions compared to the maximal possible were 70%/73% for Problem 1, 66%/71% for Problem 2A, 62%/65% for Problem 2B, and 63%/68% for Problem 3. It is likely that one of the main reasons for the relatively poor fit of the models on an absolute basis was that distances were calculated from the physical values of each feature and these physical distances were perhaps a poor indicant of perceived similarities. The final experiment investigates whether the predictions can be improved by using psychologically scaled distances.

#### EXPERIMENT IV

There are two possible advantages of using scaled distances rather than physical distances. First, previous research has shown that the features of the faces differ in saliency (Friedman, Reed, & Carterette, 1971) and these differences would presumably influence S's similarity judgments. Secondly, even if all features were equally salient, it is doubtful whether physical distances could be translated directly into psychological distances. For example, ratios of the physical distance rather than absolute differences in physical distance might be a more important determinant of perceived similarity. In order to determine whether psychological scaling would improve the fit of the models, patterns used in Problem 2A were used in the scaling experiment. Since the same patterns were used in Problem 3, the scaled distances were used to reformulate the predictions of the distance models as applied to both problems.

#### *Procedure*

Six Ss participated in the experiment. All Ss were male undergraduate students at UCLA. Five of the six Ss were currently enrolled in a Perception Laboratory course. The Ss were told that they would be paid \$2/hr for participation in a scaling experiment and that a \$2 bonus would be given to those Ss who performed well on the task.

In the scaling experiment Ss judged the similarity of 679 pairs of faces. Photographs ( $2 \times 3$  in.) of the faces were Xeroxed onto legal-sized

paper. There were three pairs of faces on a page. The first 496 pairs represented all possible pairs of the 32 faces used in Problem 2A. The last 183 pairs were included to provide data for a factorial, analysis-of-variance analysis of the similarity ratings. This additional data was needed for future work and is not discussed here.

The S's task was to judge the similarity of each pair of faces on a scale ranging from "1" to "20." Both ends of the scale were anchored. Ss were told that a rating of 1 should be assigned only to identical faces and a rating of 20 should be assigned to faces that were as dissimilar as the left and right faces shown in Fig. 1. All Ss completed the scaling experiment in either two or three sessions and returned the following week for a classification experiment. The procedure of the classification experiment was identical to the procedure used in Experiment II except that Ss repeated Problem 2A three times and were told to use a different strategy each time. The results of the classification experiment are discussed in Reed (1970). The findings presented here are concerned with the more central issue of whether scaled distances would improve the fit of the models.

### *Results*

The TORSCA-9 multidimensional scaling program (Young, 1968) was used to analyze the similarity ratings. The basic objective of the program was to combine previous developments in scaling theory into a single program. The program is based on the Shepard-Kruskal formulation, but includes additional refinements such as Cliff's (1966) method of rotating the spatial configuration into congruence with a specified configuration. If the space in which the computations are being performed is Euclidean, the derived configuration can be orthogonally rotated to a least-squares fit with an hypothesis matrix. In this case, the hypothesis matrix is the physical values of the four features for each of the 32 scaled patterns. After rotation, the axes of the scaled configuration should be aligned with the physical dimensions.

Scaled solutions were first obtained for the individual Ss in three, four, and five dimensions. Table 4 shows the resulting stress measures for both the Euclidean and city block metrics. The stress values for the two metrics are very close, with the city block metric usually being 1% higher in stress. In addition, the range in stress values for the six Ss appears to be quite similar. For example, the 4 dimensional Euclidean solution produces stress values of 14, 18, 16, 14, 16, and 15% for each of the six Ss. Since the patterns varied on four physical dimensions and the five-dimensional stress values were not much lower than the four-dimensional stress values, it was decided to use the four-dimensional solution. The



TABLE 4  
Stress Measures of the Seven Scaling Solutions as a Function of  
Dimensionality and Metric

Subject	Metric <i>r</i>	Dimensions (%)		
		5	4	3
1	1	12	15	20
	2	11	14	19
2	1	16	19	24
	2	15	18	23
3	1	14	17	21
	2	13	16	20
4	1	12	15	21
	2	11	14	20
5	1	15	18	20
	2	14	16	21
6	1	13	16	20
	2	12	15	19
Mean ratings =	1	11	13	18
	2	10	12	16

Euclidean metric was chosen for subsequent analysis because the Euclidean solutions were slightly lower in stress and the configuration could be rotated to correspond with the physical dimensions of variation. This rotation is necessary for the Weighted Feature models which depend on knowing the value of each dimension in addition to the overall distance between two patterns. The feature weights were the same weights calculated from the physical values that were used previously for Problems 2A and 3. It was therefore necessary to align the scaled solution with the physical dimensions so that the scaled dimensions could be appropriately weighted. An alternative procedure would be to derive a new set of weights from the scaled dimensions.

The results of the scaling experiment indicated that the mean ratings produced lower stress values than any of the individual solutions (Table 4). Although this result was not expected, there was some indication that the mean ratings were less "noisy" and more consistent than the individual ratings. Only 5% of the mean ratings violated the triangle inequality compared to a range of 10 to 32% over the six individual sets of ratings (Beals, Krantz, & Tversky, 1968). The use of the general distance matrix based on the mean ratings resulted in better predictions as to how

the six Ss would classify patterns than the use of the individual distance matrices. It would appear that mean ratings removed some of the error variance from the data, which resulted in a better distance representation for the general solution.

These distances were therefore used to test whether scaled distances would improve predictions for Problems 2A and 3. Table 5 shows a comparison between the Euclidean distance models based on the previous, physical distances and the scaled distances. In all cases, the scaled distances improved the predictions of the models. In Problem 2A, the Average Distance and Prototype models with weighted features predicted 70% of the classifications out of a possible 71%. In Problem 3, the Prototype models predicted 67% of the classifications out of a possible 68%. Thus, not only did the use of scaled dimensions improve predictions, but these improvements approached the maximal possible number of correct predictions.

The use of scaled distances also improved the correlations between the "favorability of Category 1" and the "number of Ss choosing Category 1." These results are presented in Table 6. The correlations listed under Physical features were those presented previously as a part

TABLE 5  
Percentage of Correct Predictions of the Euclidean Distance Models  
Based on Physical and Scaled Dimensions

Model	Physical (%)	Scaled (%)
Problem 2A		
Weighted Average Distance	62	70
Weighted Prototype	62	70
Average Distance	57	63
Prototype	63	69
Proximity: $k = 1$	62	64
$k = 3$	56	66
$k = 5$	45	60
Best Predictor	71	71
Problem 3		
Weighted Average Distance	61	67
Weighted Prototype	61	67
Average Distance	58	63
Prototype	59	67
Proximity: $k = 1$	59	62
$k = 3$	56	61
$k = 5$	49	59
Best Predictor	68	68

TABLE 6  
 Correlations Between "Number of Subjects Choosing Category 1" and "Favorability of Category 1" as Calculated from Euclidean Distance Models Based on Physical and Scaled Dimensions

	Physical	Scaled
Problem 2A		
Weighted Average Distance	0.62	0.83
Weighted Prototype	0.63	0.83
Average Distance	0.53	0.84
Prototype	0.46	0.80
Problem 3		
Weighted Average Distance	0.64	0.85
Weighted Prototype	0.64	0.83
Average Distance	0.43	0.77
Prototype	0.42	0.73

of Experiments II and III. As shown in Table 6, the use of scaled dimensions resulted in much higher correlations. The four models resulted in approximately the same correlations in Problem 2A, but the weighted feature models resulted in slightly higher correlations in Problem 3. This latter finding confirms the results presented in the previous experiments in which the weighted feature models usually resulted in higher correlations.

One consequence of the improved predictions of the distance models is that the one cue Cue Validity model now predicts significantly worse than the weighted feature Average Distance and Prototype models. A test of significance for correlated proportions revealed that the one cue model predicted significantly worse for both Problem 2A ( $z = 2.52, p = .01$ ) and Problem 3 ( $z = 2.67, p = .01$ ). This finding contrasts with earlier results in which the one cue model could not be rejected.

In conclusion, the best fitting models for Problems 2A and 3 are the Average Distance model with weighted features and the Prototype model, with or without weighted features. Further distinctions cannot be made since these three models make essentially the same predictions for this particular problem. However, all other models predicted significantly worse than these three models. In addition, the results of the previous experiments supported the Prototype model over the Average Distance model and suggested a differential weighting of features.

The choice of a particular metric did not appear to be a crucial variable in either the scaling experiment or the classification experiments. The overall findings are inconsistent and strong support could not be given to either the Euclidean or city block metric. Since the scaling

solution based on the Euclidean metric can be rotated to the physical axes and this solution led to near optimal predictions, it is recommended that this approach be used in future work.

#### DISCUSSION

The experimental findings, taken as a whole, support the Weighted Features Prototype model. The results suggest that *the predominant strategy was to form an abstract image or prototype to represent each category and to classify test patterns on the basis of their similarity to the two prototypes*. Similarity was determined, however, by placing a greater emphasis on those features which best discriminated between the two categories.

What are the theoretical implications of these findings? One important theory is the feature differentiation theory. It should be emphasized that this theory is a theory of perceptual learning which is defined as "an increase in the ability to extract information from the environment, as a result of experience and practice with stimulation coming from it" (Gibson, 1969, p. 3). Perceptual learning occurs through filtering out irrelevant information and learning to attend to the relevant dimensions. The problem is how do Ss learn to distinguish one object from another. What was not made explicit is how does categorization relate to this theory. This point is briefly discussed by Gibson (1969):

It would seem, superficially, that practice in categorizing should lead only to more generalized, less specific perceptions and thus result in the opposite of perceptual learning. But it is also possible that members assigned to a category actually share, potentially, some feature—either a minimal distinctive one or some higher order one—that distinguishes the category from other categories. If this is picked up with practice and processed without verbal intervention, perceptual learning has taken place (Gibson, 1969, p. 190).

It is unlikely, however, that all patterns within a category share a common feature. This is true in many concept learning experiments in which S's task is to find the dimension which will enable him to correctly classify the stimuli. Once he has discovered this dimension, classification is trivial as he can respond on the basis of this one cue. Formulations of the pattern recognition problem differ in that stimuli are typically composed of a number of partially valid cues rather than a single, perfectly valid cue. A decision rule is therefore required which will integrate the available information. This problem is not discussed by Gibson and, in fact, it might be said that the present study begins where that theoretical formulation ends. The categorization models assume that the relevant features have been identified and Ss in the present study

were told at the beginning of the experiment what were the relevant features.

The fact that the weighted feature models led to better predictions suggests that some feature learning may occur with respect to determining the most informative features for a particular task. Yonas (1969) has expressed this view as an explanation for the improvement shown by Ss in a letter classification task. Even if Ss had already learned the basic features of letters, additional learning could involve the selection of an optimal set of features for distinguishing between the positive and negative sets. The selection of relevant dimensions limited the success of using similarity ratings to predict how Ss would make classifications in a study by Shepard, Hovland, and Jenkins (1961). The present findings suggest that distance models can be used to predict how Ss will make classifications if the dimensions are differentially weighted to account for the greater emphasis given to the more relevant dimensions.

The results of these experiments are favorable to the view that the formation of a prototype is a means of classification (Attneave, 1957; Evans, 1967; Posner, 1969). Although these findings are not the first to suggest the formation of a prototype, their main contribution is that they provide a more rigorous quantitative test of competing theories than has been reported previously. Furthermore, the results suggest that while the Prototype model was comparatively successful, other strategies were also used. This was particularly evident in Problem 2B in which the one cue and Proximity Algorithm were the most successful models. It is therefore hoped that the approach used here will provide a broader framework in which to study the effects of instructions, subject variables, stimulus variables, and category composition on decision rules. A subject variable of particular importance is "age" and it would be of interest to study how children's categorization develops chronologically and how this could be related to a central computing space (Pascual-Leone & Smith, 1969).

Another important variable is how category composition influences the complexity of the problem and the decision rules Ss use. It was suggested that the greater complexity of Problem 2B may have made it more difficult for Ss to abstract a prototype. Taylor and Haygood (1968) found that Ss' ability to classify words into two semantic categories depended on how far the categories were separated on a Semantic Differential scale. The results showed that performance improved continuously as category separation increased from no separation and chance responding to maximal separation. These results are suggestive that information processing breaks down as categories overlap;

and this, in turn, may lead to the use of simpler decision rule. Further research is needed to determine how decision rules are influenced by task complexity.

It is also of interest to consider the generality of the categorization models. It was hoped that these "miniature experiments," particularly the learning experiment, would capture some of the processes involved in the real-life learning of patterns. It is difficult at this stage to test the correctness of this assumption, but the experiments may also have a value as approaches to more cognitive categorizing. Let us consider an oversimplified example in which I try to determine whether my own views are basically liberal or conservative. I am willing to label four prominent politicians, A, B, C, and D as "liberal" and four prominent politicians, W, X, Y, and Z as "conservative." I compare my own position on a number of important issues and decide that I am basically a liberal because my own views are most similar to politician C. In this example, the politicians are the patterns in the two categories; their stand on the various issues are the features; I am the test pattern which I classify using the Proximity Algorithm. Whether *S*'s decision strategies would be invariant across these two diversified tasks would be an important issue related to interactions between perceptual and conceptual development.

The final issue concerns the relationship of the models considered here to structural theories of pattern recognition (Sutherland, 1971). The models were all based on the same assumption—that the patterns could be adequately represented in terms of a vector of independent feature values. That this assumption is an oversimplification is indicated by those *Ss* who used the distance between the eyes and the nose as a basis for classification. This distance is determined by the relationship between two features; the height of the forehead and the length of the nose. More generally, there are many patterns for which the consequences of ignoring structural information will be much more serious and present techniques for testing for additivity should aid in identifying such patterns (Anderson, 1970; Tversky and Krantz, 1969). A more detailed discussion of the characteristics of the three types of models—probability, distance, and structural—and their interrelationships is given in Reed (1973). On the whole, however, the additive models examined here gave an adequate account of the data.

#### APPENDIX

A more precise mathematical formulation of the models is presented in this section.

*Decision Rule One (Cue Validity Model)*

$$\text{Decide } X_j \leftarrow \text{Category 1 if: } \sum_{m=1}^d \frac{p(\text{Category 1}/x_m)}{d} \geq \sum_{m=1}^d \frac{p(\text{Category 2}/x_m)}{d}.$$

$$\text{Decide } X_j \leftarrow \text{Category 2 if: } \sum_{m=1}^d \frac{p(\text{Category 1}/x_m)}{d} < \sum_{m=1}^d \frac{p(\text{Category 2}/x_m)}{d}.$$

The cue validities were calculated by considering both the frequency and the proportion of cues in the two categories. In order to make cue validities dependent on cue frequencies, cue validities were calculated from a weighted combination of prior and sample probabilities. Each cue has a prior probability of .5 since both categories are equally likely with respect to any cue before  $S$  sees the composition of the two categories. The prior probability will be weighted by  $p$  and the sample probability by  $1 - p$ , where  $p = 1/(1 + F)$ .  $F$  is the frequency with which the cue appears in the two categories and the larger the value of  $F$ , the more the sample probability is weighted.

An illustration of this approach can be seen in the following example. Consider two categories, each category being composed of five faces. In case one, one face in Category 1 has a long nose and zero faces in Category 2 have a long nose. Thus,  $p(\text{Category 1}|\text{Long Nose}) = 1/1 = 1$ . In case two, three faces in Category 1 have long noses and zero faces in Category 2 have a long nose. Thus,  $p(\text{Category 1}|\text{Long Nose}) = 3/3 = 1$ . If cue frequency is important, however, a long nose should be a more valid cue in the second case. Applying the frequency rule, the validity of a long nose in case one ( $F = 1$ ) is:

$$v = p(0.5) + (1 - p) \cdot (1.0) = (0.5) \cdot (0.5) + (0.5) \cdot (1.0) = 0.75,$$

since  $p = 1/(1 + F) = 0.5$ .

The validity of a long nose in case two ( $F = 3$ ) is:

$$v = (0.25) \cdot (0.5) + (0.75) \cdot (1.0) = 0.88,$$

since  $p = 1/(1 + F) = 0.25$ .

This version of a Cue Validity model therefore makes the validity of a cue dependent on its frequency.

*Decision Rule Two (Proximity Algorithm)*

If  $n_1$  is the number of the  $k$  closest patterns belonging to Category 1 and  $n_2$  is the number of the  $k$  closest patterns belonging to Category 2,

Decide  $X_j \leftarrow$  Category 1 if,  $n_1 > n_2$   
 Decide  $X_j \leftarrow$  Category 2 if,  $n_1 < n_2$   
 Continue testing if  $n_1 = n_2$ .

The rule is tested for  $k = 1, 3,$  and  $5$  and for the city block and Euclidean metrics.

*Decision Rule Three (Average Distance)*

Decide  $X_j \leftarrow$  Category 1 if,

$$1/N1 \sum_{i=1}^{N1} \left( \sum_{m=1}^d |x_{jm} - x_{im}|^r \right)^{1/r} \leq 1/N2 \sum_{i=1}^{N2} \left( \sum_{m=1}^d |x_{jm} - x_{im}|^r \right)^{1/r}.$$

Decide  $X_j \leftarrow$  Category 2 if,

$$1/N1 \sum_{i=1}^{N1} \left( \sum_{m=1}^d |x_{jm} - x_{im}|^r \right)^{1/r} > 1/N2 \sum_{i=1}^{N2} \left( \sum_{m=1}^d |x_{jm} - x_{im}|^r \right)^{1/r}.$$

The decision rule is investigated for  $r = 1$  and  $r = 2$ .

*Decision Rule Four (Prototype)*

Decide  $X_j \leftarrow$  Category 1 if,

$$\left( \sum_{m=1}^d |x_{jm} - p_{1m}|^r \right)^{1/r} \leq \left( \sum_{m=1}^d |x_{jm} - p_{2m}|^r \right)^{1/r}.$$

Decide  $X_j \leftarrow$  Category 2 if,

$$\left( \sum_{m=1}^d |x_{jm} - p_{1m}|^r \right)^{1/r} > \left( \sum_{m=1}^d |x_{jm} - p_{2m}|^r \right)^{1/r}.$$

Decision rule four is investigated for  $r = 1$  and  $r = 2$ .

The prototype of each category is calculated by finding the mean value of all features for the patterns in that category. If  $x_{id}$  is the  $d$ th component of the  $i$ th pattern,

$$P_1 = (p_{11}, p_{12}, \dots, p_{1d}) = \left( \sum_{i=1}^{N1} x_{i1}/N1, \dots, \sum_{i=1}^{N1} x_{id}/N1 \right), \quad (2)$$

$$P_2 = (p_{21}, p_{22}, \dots, p_{2d}) = \left( \sum_{i=1}^{N2} x_{i1}/N2, \dots, \sum_{i=1}^{N2} x_{id}/N2 \right). \quad (3)$$



*Decision Rule Five (Average Distance-Weighted Features)*

Decide  $X_j \leftarrow$  Category 1 if,

$$1/N1 \sum_{i=1}^{N1} \left( \sum_{m=1}^d w_m^r |x_{jm} - x_{im}|^r \right)^{1/r} \leq 1/N2 \sum_{i=1}^{N2} \left( \sum_{m=1}^d w_m^r |x_{jm} - x_{im}|^r \right)^{1/r}.$$

The weights are found by computing a matrix  $B$  of between-set distances and matrix  $T$  involving distances between the total number of patterns in all sets. Both matrices  $B$  and  $T$  are in the form of covariance matrices with elements  $b_{m,n}$  and  $t_{m,n}$  for the  $m$ th and  $n$ th features given by Eqs. 4 and 5.  $N = N1 + N2$  and  $M = (N2)$ .

$$b_{m,n} = 1/(N1 \cdot N2) \sum_{i=1}^{N1} \sum_{j=1}^{N2} (x_{im} - x_{jm})(x_{in} - x_{jn}), \quad (4)$$

$$t_{m,n} = 1/M \sum_{i=1}^N \sum_{j=1}^N (x_{im} - x_{jm})(x_{in} - x_{jn}). \quad (5)$$

The weights are found by taking the eigenvector corresponding to the largest eigenvalue of the matrix  $BT^{-1}$ .

*Decision Rule Six (Prototype-Weighted Features)*

Decide  $X_j \leftarrow$  Category 1 if,

$$\left( \sum_{m=1}^d w_m^r |x_{jm} - p_{1m}|^r \right)^{1/r} \leq \left( \sum_{m=1}^d w_m^r |x_{jm} - p_{2m}|^r \right)^{1/r}.$$

Decide  $X_m \leftarrow$  Category 2 if,

$$\left( \sum_{m=1}^d w_m^r |x_{jm} - p_{1m}|^r \right)^{1/r} > \left( \sum_{m=1}^d w_m^r |x_{jm} - p_{2m}|^r \right)^{1/r}.$$

The weights are identical to the weights calculated for decision rule five.

## REFERENCES

- ANDERSON, N. H. Functional measurement and psychophysical judgement. *Psychological Review*, 1970, 3, 153-170.
- ATTNEAVE, F. Dimensions of similarity. *American Journal of Psychology*, 1950, 63, 516-556.
- ATTNEAVE, F. Transfer of experience with a class schema to identification learning of patterns and shapes. *Journal of Experimental Psychology*, 1957, 54, 81-88.
- BEACH, L. R. Cue probabilism and inference behavior. *Psychological Monographs*, 1964, 78(5), 1-20. (a)
- BEACH, L. R. Recognition, assimilation, and identification of objects. *Psychological Monographs*, 1964, 78(6), 21-37. (b)

- BEALS, R., KRANTZ, D., & TVERSKY, A. Foundations of multidimensional scaling. *Psychological Review*, 1968, **75**, 127-142.
- BRUNER, J. On perceptual readiness. *Psychological Review*, 1957, **64**, 123-152.
- CLIFF, N. Orthogonal rotation to congruence. *Psychometrika*, 1966, **31**, 33-44.
- EVANS, S. H. A brief statement of schema theory. *Psychonomic Science*, 1967, **8**(2), 87-88.
- FIX, E., & HODGES, J. Discriminatory analysis, nonparametric discrimination: consistency properties, University of California, Berkeley, Contract AF41(128)-37, February 1951.
- FRIEDMAN, M. P., REED, S. K., & CARTERETTE, E. C. Feature saliency and recognition memory for schematic faces. *Perception and Psychophysics*, 1971, **10**, 47-50.
- GIBSON, E. *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts, 1969.
- HUNT, E. B., MARIN, J., & STONE, P. *Experiments in induction*. New York: Academic Press, 1966.
- KRUSKAL, J. B. Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis. *Psychometrika*, 1964, **29**, 1-27.
- NEISSER, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.
- NILSSON, N. J. *Learning machines*. New York: McGraw-Hill, 1965.
- PASCUAL-LEONE, J., & SMITH, J. The encoding and decoding of symbols by children: A new experimental paradigm and neo-Piagetian model. *Journal of Experimental Child Psychology*, 1969, **8**(2), 328-355.
- POSNER, M. I. Abstraction and the process of recognition. Vol. 3 in G. H. Bower & J. T. Spence (Eds.), *Psychology of learning and motivation*, New York: Academic Press, 1969.
- POSNER, M. I., & KEELE, S. W. On the genesis of abstract ideas. *Journal of Experimental Psychology*, 1968, **77**(3), 353-363.
- REED, S. K. Decision processes in pattern classification. Technical Report No. 32, Perceptual Systems Laboratory, University of California, Los Angeles, 1970.
- REED, S. K. *Pattern recognition and information processing*. New York: Academic Press, 1973.
- SEBESTYEN, G. S. *Decision-making processes in pattern recognition*. New York: Macmillan Co., 1962.
- SHEPARD, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika*, 1962, **27**, 125-140. (a); II. *Psychometrika*, 1962, **27**, 219-246. (b)
- SHEPARD, R. N., HOVLAND, C. J., & JENKINS, H. M. Learning and memorization of classifications. *Psychological Monographs*, 1961, **75**, No. 13 (Whole No. 517).
- SUTHERLAND, N. S. Object recognition. Vol. 3 in E. C. Carterette and M. P. Friedman (Eds.), *Handbook of perception*, New York: Academic Press, 1971.
- TAYLOR, C. L., & HAYGOOD, R. C. Effects of degree of category separation on semantic concept identification. *Journal of Experimental Psychology*, 1968, **76**, 356-359.
- TVERSKY, A., & KRANTZ, D. H. Similarity of schematic faces: A test of interdimensional additivity. *Perception & Psychophysics*, 1969, **5**(2), 124-128.
- YONAS, A. The acquisition of information-processing strategies in a time-dependent task. PhD dissertation, Cornell University, 1969.
- YOUNG, F. W. TORSCA-9, a Fortran IV program for nonmetric multidimensional scaling. *Behavioral Science*, 1968, **13**(4), 343-344.