Array Models for Category Learning

W. K. ESTES

Harvard University

A family of models for category learning is developed, all members being based on a common memory array but differing in memory access and decision processes. Within this framework, fully controlled comparisons of exemplar-similarity, feature-frequency, and prototype models reveal isomorphism between models of different types under some conditions but empirically testable differences under others. It is shown that current exemplar-memory models, in which categorization judgments are based on similarities of perceived and remembered category exemplars, can be interpreted as generalized likelihood models but can be modified in a simple way to yield pure similarity models. Distance-based exemplar models are formulated that provide means of investigating issues concerning deterministic versus probabilistic decision rules and links between categorization and properties of perceptual dimensions. Other theoretical issues discussed include aspects of similarity, the role of memory storage versus computation in category judgments, and the limits of applicability of array models. © 1986 Academic Press, Inc.

My objective in this study is to develop baseline models representing principal approaches to category learning. The function of these models is to provide a framework for processing category learning data in order to estimate theoretically interesting quantities and to allow tests of alternative hypotheses about mechanisms and processes.

Much of the substantial volume of research on categorization conducted within the framework of cognitive psychology has turned on attempts to evaluate the relative merits of three types of models (Medin & Smith, 1984; Millward, 1980; Reed, 1973; Smith & Medin, 1981). (1) In exemplar-memory models, the learner stores mental representations of exemplars, grouped by category, then classifies new instances on the basis of their similarity to the remembered ensembles. (2) In feature-frequency (or cue-validity) models, the learner records the relative frequencies of occurrence of individual features of exemplars, then classifies new instances on the basis of estimates of the likelihood that the vector of features in a test pattern arose from each alternative category.

Preparation of this report and the research described were supported in part by NSF Grant BNS 80-26656, USPHS Grant MH37208, and a Fellowship from the John Simon Guggenheim Foundation. I am indebted to Martha J. Gordon, Robert M. Nosofsky, and Hazel Rovno for assistance with computations, and to Gordon and Nosofsky for many stimulating discussions of the work while in progress. Request for reprints should be sent to William K. Estes, Department of Psychology and Social Relations, Harvard University, 33 Kirkland Street, Cambridge, MA 02138.

0010-0285/86 $7.50
Copyright © 1986 by Academic Press, Inc.
All rights of reproduction in any form reserved.

(3) In prototype models, the learner forms an abstract representation of each category represented in a series of learning experiences, then classifies new instances on the basis of their distances from the category prototypes in a psychological space.

The many attempts to achieve differential empirical tests of these models have not proven as instructive as we might like for two reasons. One is that only rarely have tests compared explicitly formulated models that differed only with respect to one critical assumption. When the models are incompletely specified or differ in multiple respects, results of tests cannot be very informative. The other problem is that the tests typically depend on data from subjects who learn categorizations in the experimental situation, but the learning process generally is not represented in the models being compared; and differential predictions from models may vary greatly as a function of the stage of learning at which comparisons are made.

In an attempt to improve on this situation, I have sought to formulate a family of models broad enough to encompass the three types but organized within a framework providing a common basic form of memory representation and a common learning process. Within this framework, different assumptions about memory access and decision processes can generate models of the different types as special cases and set the stage for controlled comparisons of models that differ only in narrowly specified respects. In return for the simplifications and idealizations needed to achieve a tractable family of models, we can hope not only to progress toward sharper tests of alternative conceptions but also to achieve deeper understanding of the different types of models and their interrelationships.

As a preliminary to presentation of the theoretical framework, it is useful to specify the task domain to which the family of models should be applicable. In general, I shall follow Murphy and Medin (1985) in distinguishing between categorization and concept formation. Categorization refers to an individual's ability to assign objects or other stimulus patterns to categories for which there is some way of characterizing correct performance; concept formation refers to a mental representation of a category or set of categories that is presumed to underly both categorization performance and other related behaviors such as typicality ratings. In the work described in this article, I began with the simplest cases of what seemed to be basic types of category learning, primarily for reasons of tractability, but, as will be seen, the results prove to bear also on issues of mental representation.

Within the domain of categorization tasks, I consider both deterministic and nondeterministic, or probabilistic, situations. Here, *deterministic* is not used in a philosophical sense, but rather refers to the class of tasks in which a 100% success rate can in principle be achieved on the basis of the information available to the learner. In situations that I term probabilistic, perfect performance may or may not be attainable by an agency with complete information, but it is not attainable on the basis of the information available to the learner. Another major distinction relates to category size. Combining that dimension with deterministic/probabilistic situations yields three classes of tasks that have been prominent in research. (1) Deterministic categorizations of small sets of exemplars, for example, legal moves of chess pieces, metallic and nonmetallic chemical elements, inert gases, are common in ordinary life, but in psychological research on categorization they are generally treated as a degenerate case, equivalent to simple paired-associate (or "verbal discrimination") learning, since no element of abstraction is required for successful performance. (2) Deterministic categorization based on large or infinite sets, illustrated in ordinary life by games, edible mushrooms, list-programming languages, has been the prime focus of research on categorization and concept formation, with principal interest in the learners' ability to classify newly encountered exemplars of categories. (3) Categorization of large or infinite sets defined only probabilistically ("fuzzy sets") is exemplified by the assignment of patients to diseases or machine malfunctions to causes on the basis of symptom patterns or of job applicants to prognostic categories on the basis of test and interview data; generally, perfect performance is unattainable, but there may be some definable maximum based on optimal use of information from observation of exemplars. Models that can account for learning of Type 3 necessarily entail, as special cases, more restricted models applicable to Types 1 and 2. Thus I shall concentrate attention on models and experimental paradigms for Type 3 situations.

THE ARRAY FRAMEWORK

The models to be considered are all developed within the framework of an array representation of category memory. In common with a number of general approaches to memory (Bower, 1967; Norman & Rumelhart, 1970; Underwood, 1969), I assume that when a learner observes exemplars of categories, information about each exemplar is stored in the form of a vector of feature or attribute values.[1] For example, in a set of experiments designed for convenient applicability of array models (Estes, 1986), categories were diseases, exemplars were charts of hypothetical patients, and features were high or low values of symptoms. With the symptom values denoted 1 and 0, a learner's memory representation of a sequence of trials might take the form

[1] I shall make no distinction among the terms *feature, attribute,* and *dimension* (in conformity with the practice of Murphy & Medin, 1985, among others).

A 11011101
A 10101111
B 10000001

where the rows correspond to individual exemplars, and A and B denote category tags.

Array representations are more general than might be thought on first impression. Categories and exemplars used in research are nearly always readily describable in terms of features or attributes.[2] Therefore, information about them that in some theories would be entered in a propositional network (for example, Elio & Anderson, 1981) can be expressed as well in the array format; indeed, it needs to be for purposes of the computations required to deal with fuzzy categories. In principle, features can have more than two values, or even values on continuous dimensions, and the number of features per exemplar may vary within or across categories. However, to keep derivations manageable, I limit consideration in this article to cases in which features are binary valued (except where specified otherwise) and constant in number over all categories and exemplars within an experiment.

For an array model to become a model for category learning, assumptions must be made as to how information from perceived exemplars is entered in the memory array. In natural situations, there is no reason to believe that all of the available information in exemplar occurrences is effectively perceived and entered in memory. One type of selection is systematic, the perceptual or attentional learning (Gibson, 1969; Trabasso & Bower, 1968) that tends to lead the learner to attend and encode components or aspects of stimulus displays that are relevant to task demands. The other type is random, the consequence of confronting a limited-capacity processing system with an information overload (Bower, 1972; Estes, 1959; Norman & Rumelhart, 1970). Attentional learning is undoubtedly an important component of category learning in natural situations. However, current theories of attentional learning need considerable formal development before they could be incorporated into models for category learning without making the whole task unmanageable. Thus in the present work I take explicit account only of random selectivity, which must be assumed to be an inescapable aspect of the learning process.

Following earlier stimulus-sampling models for learning (Estes, 1959), I shall examine the implications of two specific assumptions about selection for storage—random selection of elements (features) of an exemplar display and random selection of patterns as units. These two mechanisms

[2] Even when categories seem to be characterizable only in qualitative or subjective terms, methods of multidimensional scaling can generate descriptions in terms of attributes or dimensions in a psychological space that are amenable to array representation.

are by no means on a par theoretically, however. The former is assumed to be a basic property of the initial storage of exemplar information in the memory array. Thus in each of the specific models developed, there is defined a probability, p, assumed for simplicity to be the same for all exemplars and constant over trials, that each feature of a perceived exemplar is entered in the memory array. Except in the special case of $p = 1$, fragmentary storage of exemplar patterns will sometimes occur and may set limits on attainable levels of categorization performance.

Even though a representation of an exemplar is successfully entered as a vector of feature values in the memory array, it may not be retrievable for comparison with a test pattern unless it has become associated with an effective retrieval cue, a process that may depend on both selective attention and rehearsal. In another study (Estes, 1986), I have begun to investigate the conditions under which an exemplar representation acts as a retrievable unit, but not enough is yet known about unitization to enable a formal representation. Therefore, when comparing exemplar-memory models with other types in this article, I make the simplifying assumption that all stored exemplar patterns are retrievable. Though the assumption seems strong, it does not appear that relaxing it would materially affect the conclusions that are drawn about relationships among models.

The issue of early versus late computation has defined a major branch point in the evolution of categorization models and will therefore do the same in my development of array models. The term *late computation* characterizes models in which the learner, at the time of a decision on categorization, is conceived to consult all of the relevant information in memory and enter it into whatever computations are required to generate a decision. This approach has been associated mainly with exemplar-memory models (Brooks, 1978; Hintzman & Ludlam, 1980; Medin & Shaffer, 1978). In contrast, *early computation* characterizes cue-validity, prototype, and schema-abstraction models in which the mental computations basic to categorization are largely accomplished during the course of learning with only the results of these being consulted at the time of the decision. For example, in prototype models, abstract representations of categories are developed during the course of learning, being updated as information comes in on each trial, but when a test exemplar is to be categorized, it is compared only to the current prototypes of the alternative categories.

With regard to adaptiveness, early computation would seem to have an advantage in allowing more rapid generation of responses, since only the results of previously accomplished computations need be consulted at the point of decision. On the other hand, early computation entails the discarding of information that does not enter into the construction of prototypes or other computed results, and if environmental contingencies change over the course of time, situations may arise in which the discarded information would be useful if a shift in strategy or tactics on the part of the learner were called for. In late-computation models, full information about previously experienced exemplars is preserved (except when forgetting is allowed for, as in Hintzman & Ludlam, 1980) and is available for utilization if selective attention shifts from one to another aspect of the stored information during the course of learning (Nosofsky, 1984a, 1984b). Indeed, Marr (1976) has offered it as a design principle of intelligent systems that any decision that commits the system irreversibly to one course of information processing should be delayed as long as possible.

Baseline-Exemplar, Prototype, and Feature Models

In the following presentation of the family of array models, I start with the late-computation branch. I focus first on the basic exemplar-memory model and show that by a reorganization of the memory array it can be converted into an early-computation model that is indistinguishable in many of its empirical implications. Then I take up examples of prototype and feature-frequency models of the early-computation variety and examine conditions under which various subsets of the models are distinguishable or equivalent in their empirical implications.

The basic exemplar-memory model. The simplest exemplar model to be considered is closely related to the one developed by Medin and Schaffer (1978), the principal restriction being the assumption that all attributes, or features, of an exemplar are equally salient and receive equal attention from the learner.[3] On any learning trial, each feature of the exemplar presented has probability p of being encoded in memory, and therefore each exemplar in a sequence has some likelihood of being stored, completely or in part, in the memory array along with its category tag. To illustrate the assumed categorization process, let us imagine that a learner has stored the sequence

10	A
11	A
00	B

and then is presented with pattern 10 and asked to assign it to the proper category. In the model, the test pattern would be compared with each of those stored in memory; the similarity of the test pattern to each remem-

[3] With this restriction, the expressions for categorization probability are the same in this model as in the corresponding special case of Medin and Schaffer's model. However, the exemplar model displayed here includes assumptions about memory storage during learning and memory access during generation of a categorization response, which remain unspecified in Medin and Schaffer's model.

bered pattern would be computed and the similarities summed for each category.

The rule for computing similarity is to assign a value of 1 for a comparison of a perceived and a remembered feature if there is a match and a value s ($0 \leq s \leq 1$) if there is a mismatch[4] and take the product of these values. In this example, comparison of the test pattern to the first row of the array yields $1 \times 1 = 1$, to the second row $1 \times s = s$, and to the third row $s \times 1 = s$. The sum of the similarities to category A is $1 + s$ and to category B the sum is s. If the test pattern were 01, which had not occurred previously, the comparison process would proceed similarly, yielding similarities of s^2, s, and s for the three rows and sums of $s + s^2$ and s for categories A and B, respectively.

The probabilities of assigning a pattern to category A or B are in the ratio of the summed similarities of the pattern to the stored exemplars of each category. Thus, in the example, the probability of categorizing pattern 10 as an A is $(1 + s)/(1 + 2s)$, and the probability of categorizing 01 as an A is $(s + s^2)/(2s + s^2) = (1 + s)/(2 + s)$. It will be seen that the relative probabilities of correct categorization of patterns 10 and 01 depend on the value of the similarity parameter, s. If s were equal to 1, meaning that the two possible feature values were not discriminated by the learner, both probabilities would be 2/3, depending only on the numbers of exemplars stored in the two categories. If s were equal to 0, meaning that the two possible feature values were perfectly discriminated, the probabilities would be 1 for pattern 10 and 1/2 for pattern 01, and the probability would be higher for 10 than 01 for all intermediate values of s. It will be generally true, as in this example, that if only a few exemplar representations have been stored, then, other things being equal, an exemplar that has occurred once previously will have higher probability of correct categorization than an examplar presented for the first time, because only the old exemplar has maximal similarity to some element of the memory array. As the number of stored exemplars increases, this difference decreases toward 0. Repetitions continue to be important, however. Probability of correct categorization of a test pattern increases, on the average, as a function of its frequency of occurrence during the preceding series, because the number of stored patterns to which it is maximally similar increases with repetition.

This simple exemplar model may seem too conceptually rudimentary

[4] The asymmetry in treatment of matches and mismatches is only apparent. We could define similarity parameters s_1 and s_2, say, for matches and mismatches, respectively, but since they would enter into expressions for relative likelihood and categorization probability only as ratios, one of the values can be chosen arbitrarily. The choice of $s_1 = 1$ for matches simplifies comparisons of the generalized feature-frequency model and the exemplar model.

to be the basis of category learning since it includes no process of abstraction. However, the appearance is deceiving. Because new instances are categorized on the basis of their similarities to collections of remembered patterns, generalization does occur in effect, and the resulting performance is often hard to discriminate from that of prototype or schema-abstraction models (Busemeyer, Dewey, & Medin, 1984; Elio & Anderson, 1981). Further, the lack of dependence on any specific mechanism of abstraction means that the exemplar model is capable of learning any type of categorization task regardless of how the categories were actually generated. An individual who processes information in accord with the assumptions of the exemplar model can readily learn a variety of rule-defined categorizations (Nosofsky, 1984a, 1984b) and can even do quite well with a task in which contingencies between feature patterns and categories are so complex as to be beyond description in any simple verbal rule (Estes, 1986, Experiment 2).

The idea that the whole memory array is scanned on every trial does raise conceptual problems, however. It is apparent that if the array is searched serially, or by means of a capacity-limited parallel process (Townsend, 1974), processing time should increase indefinitely over the course of learning as the size of the memory array grows. This implication of the model has not been tested formally to my knowledge but seems unlikely to be borne out. Further, there is some empirical evidence suggesting that the whole array does not enter into comparisons on each trial (Estes, 1986).

A weighted-vector exemplar model. The assumption that categorization is accomplished by comparing test patterns to the contents of a memory store does not necessarily require that the comparisons be achieved by a scan through a chronologically ordered list of remembered patterns. To illustrate an alternative possibility, suppose that a learner has encountered the following sequence,

A	10
B	01
A	10
B	00
A	11
B	01
B	10

the pattern 10 having been presented first as an instance of category A, then the pattern 01 as an instance of category B, and so on. It is possible that, rather than continuing to work with the array as chronologically ordered, the processing system reorganizes the contents of memory into a canonical array that might take the form in this example

	A	B
10	2	1
01	0	2
11	1	0
00	0	1

in which each distinct exemplar pattern is represented only once but together with information concerning its frequency of occurrence in both of the categories. Clearly the result of computing the total similarity of a test pattern to the A and B columns of the canonical array will yield the same result as computing the total similarity to the A and B elements of the chronological array. It makes no difference, for example, whether a test pattern 10 is compared to the two A instances in the chronological array and the resulting similarities added or is compared to the single entry for 10 in the canonical A array and the result multiplied by 2. In effect, the reorganization converts the exemplar model from a late- to an early-computation model, since once a canonical array is set up it needs only to be updated on each trial.

Since the two forms of the exemplar model are equivalent in their implications for learning and transfer data, the task of ascertaining which is closer to the way the human system actually operates will evidently have to wait on information coming from reaction time measurements or other kinds of auxiliary information. For present purposes, since the canonical form is more convenient to deal with analytically, I shall use it as a basis for comparisons with other models in the remainder of this article.

Also, it will facilitate the exposition to make comparisons in terms of a single-task design. The one to be employed comprises two alternative categories, A and B, with the exemplars of each generated by combinations of two binary-valued features. All of the theoretical results presented generalize readily to larger numbers of features. The representations given in Table 1 can be taken to portray either the design of the experiment, in which case the cell entries are the probabilities with which the exemplar patterns occur in the two categories, or the canonical memory array resulting from a learning series, in which case the cell entries are the expected relative frequencies with which the memory vectors are stored in the columns of the array. Except when expressly stated otherwise, the two categories are assumed to be sampled with equal probabilities when sequences of exemplars are generated for presentation to a learner.

Problems of very different difficulty can be generated depending on the particular values assigned to the cell entries in Table 1. As presented, the design is general enough to allow for either independent or correlated features, and the distinction will prove to be of major importance. For

TABLE 1
Design (and Memory Array) for the General Case of Two Binary-Valued Features

	Category	
Pattern	A	B
1 0	u	x
0 1	v	y
1 1	w	z
0 0	$1\text{-}u\text{-}v\text{-}w$	$1\text{-}x\text{-}y\text{-}z$

simplicity, I will start with the simple special case of independent feature values shown in Table 2. In this case, the first feature in an exemplar has probability θ of taking on Value 1 and $1 - \theta$ of Value 0 in category A and the second-feature probability $1 - \theta$ of Value 1 and θ of Value 0 in category A. The probabilities of the feature values in category B are just the complements of these. Thus the exemplar pattern 10 has probability of θ^2 occurring on a category A trial and probability $(1 - \theta)^2$ on a category B trial, and so on. Applying the rules for computing similarities to Table 2, we find that if a learner were presented with pattern 10 after a long series of learning trials in the situation, the total similarity of this pattern to the memory array for category A would be

$$\theta^2 \cdot 1 + 2\theta(1 - \theta) \cdot s + (1 - \theta)^2 \cdot s^2 = [\theta + (1 - \theta)s]^2,$$

since in computing total similarity, the similarities of pattern 10 to the four patterns of the array are 1, s^2, s, and s, from top to bottom, and their relative frequencies are θ^2, $(1 - \theta)^2$, $\theta(1 - \theta)$, and $(1 - \theta)\theta$. Similarly, the total similarity of pattern 10 to the memory vectors of category B is

$$(1 - \theta)^2 + 2\theta(1 - \theta) \cdot s + \theta^2 \cdot s^2 = (1 - \theta + \theta s)^2;$$

and therefore the probability of correct categorization if the pattern were presented as an exemplar of category A would be

TABLE 2
Design of a Categorization Task with Two Independent, Binary-Valued Features

	Category	
Pattern	A	B
1 0	θ^2	$(1\text{-}\theta)^2$
0 1	$(1\text{-}\theta)^2$	θ^2
1 1	$\theta(1\text{-}\theta)$	$(1\text{-}\theta)\theta$
0 0	$(1\text{-}\theta)\theta$	$\theta(1\text{-}\theta)$

$$P_{10}(\mathrm{A}) = \frac{[\theta + (1 - \theta)s]^2}{[\theta + (1 - \theta)s]^2 + (1 - \theta + \theta s)^2}, \tag{1}$$

with a value ranging from ½ for $s = 1$ to 1 for $s = 0$.

In an actual application, the value of θ is of course prescribed by the experimenter, and if an estimate of s is available in advance (having perhaps been determined from data of a previous experiment), the asymptotic probability of categorizing pattern 10 as an A can be predicted from this expression. In the following sections, however, Eq. (1) is of interest mainly for purposes of comparisons with corresponding expressions derived for other models. Categorization probabilities for the other patterns in Table 2 are given for completeness in Appendix 1, and it can be assumed that all results given for relationships between models hold for all patterns.

In contrast to the almost universal practice in the earlier literature of learning theory, it is rarely feasible to derive closed expressions for theoretical learning curves in array models. The standard procedure is, rather, to calculate theoretical probabilities trial-by-trial by means of a computer program. For the special case of the exemplar model with the storage parameter p equal to unity, the procedure would be to replace the cell values in Table 1 with actual relative frequencies, updating the entries as the current exemplar pattern is stored on each trial. The categorization probabilities would be computed on each trial as in the derivation of Eq. (1), except that the similarities between the current exemplar and elements of the array would be weighted by the current relative frequencies rather than by the asymptotic values given in Table 1. For the general case when p is less than unity, the procedure is basically the same (see Appendix 1).

If the frequencies of occurrence of the two categories during learning were unequal, say category A having some probability π_1, and category B probability $\pi_2 = 1 - \pi_1$ of being represented on any trial, then the relative frequencies of occurrence of the exemplar patterns would be modified, and, when the table is interpreted as representing the memory array, all entries in column 1 of Table 2 would be multiplied by π_1 and the entries in column 2 by π_2. Then the expressions for categorization probability would be modified in the obvious way—Eq. (1), for example, becoming

$$P_{10}(\mathrm{A}) = \frac{\pi_1[\theta + (1 - \theta)s]^2}{\pi_1[\theta + (1 - \theta)s]^2 + \pi_2(1 - \theta + \theta s)^2}. \tag{2}$$

Analogous results obtain, of course, for the other patterns.

When π_1 and π_2 are equal, Eq. (2) reduces to Eq. (1), but when they are unequal, the probability of a category given an exemplar, and therefore also the probability of a categorization response to the exemplar, depends on the category probabilities (base rates) as well as on information from the exemplar pattern. In the extreme case when patterns are perfectly discriminated ($s = 0$) and features are entirely uninformative ($\theta = ½$), categorization probability depends wholly on the π values: $P_{10}(A) = \pi_1$, and $P_{10}(B) = \pi_2$.

A distance-based exemplar model. The exemplar model developed above follows the precedent of Medin and Schaffer (1978) in the computation of similarities. The similarities of the features of a test exemplar to those of each remembered exemplar are multiplied, then the results are added across all of the remembered exemplars of a category. This mixture of two ways of combining similarities and has the effect of making it difficult to achieve controlled comparisons between exemplar and prototype models (in which combination rules generally are additive). However, we can make use of a simple relation between similarity and distance in a psychological space (Nosofsky, 1984a, 1984b; Shepard, 1958) to formulate a version of the exemplar model that depends entirely on additive combination rules.

The key to this formulation is the function

$$s_{ij} = e^{-cd_{ij}}, \tag{3}$$

where s_{ij} denotes similarity between units i and j and d_{ij} the distance between them in a similarity space. Since the present treatment is limited to binary-valued features, we need deal with only two distances, the distance between a perceived and a remembered feature that match and the distance between a perceived and a remembered feature that mismatch. Only relative distances matter, so the distance in the case of a match can be set equal to zero. Further, since distances will appear only in ratios, the unit of distance is arbitrary, and without loss of generality we can set the distance for a mismatch equal to unity. Hence, the distance d_{xy} between two exemplar representations is simply the number of mismatches. For example, test exemplar 100111 and stored exemplar 001001 differ by 4 and their computed similarity is e^{-4c}.

The average similarity of a test exemplar to a memory array is computed by summing the distances between the exemplar pattern and each member of the array and then transforming the result to a measure of similarity as in Eq. (3). Denoting the number of stored A exemplars by N_A, cumulated distance between exemplar x and category A by D_{xA} and their similarity by S_{xA}, these assumptions can be summarized by

$$D_{xA} = (1/N_A) \sum_{y \text{ in } A} d_{xy} \tag{4}$$

and

$$S_{xA} = e^{-cD_{xA}}. \tag{5}$$

The probability of assigning exemplar x to category A when A and B are the alternatives is computed as in the standard exemplar model

$$P_x(A) = \frac{S_{xA}}{S_{xA} + S_{xB}} = \frac{1}{1 + \frac{S_{xB}}{S_{xA}}}, \tag{6}$$

but then by substitution from Eq. (5) can be expressed in terms of distances

$$P_x(A) = \frac{1}{1 + \frac{e^{-cD_{xB}}}{e^{-cD_{xA}}}} = \frac{1}{1 + e^{-c(D_{xB} - D_{xA})}}. \tag{7}$$

It should be mentioned that the definition of d_{xy} is based on the assumption that distances are measured in a city-block metric (Fig. 1). This assumption is implicit in the Medin and Schaffer model and is generally considered to be appropriate for separable, or analyzable, stimulus dimensions (Garner, 1974). A generalization of Eq. (7) to a broader class of models can be obtained by defining d_{ij} as

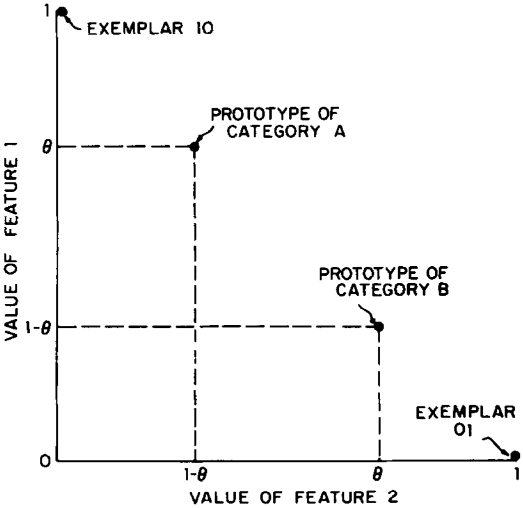

FIG. 1. Average prototypes of categories for the task design of Table 2, together with the closest exemplars. The city-block distance from test exemplar 10 to the category A prototype is the distance (1 − θ) along the vertical axis plus the distance (1 − θ) along the horizontal axis; the distance to B is θ along the vertical plus θ along the horizontal.

$$d_{xy} = \left(\sum_{i=1}^{N} \left| d_{x_i y_i} \right|^{r} \right)^{\frac{1}{r}} \tag{8}$$

where N is the number of features (the dimensionality of the simularity space) and r is a positive constant. For $r = 1$, this function reduces to the city-block metric and for $r = 2$ to the equally familiar Euclidean metric. In this presentation I limit consideration to the special case of $r = 1$, in which case the model can appropriately be termed the city-block-distance (or CBD) exemplar model. It is related to an average distance model proposed by Reed (1972) but differs in basing responses on similarities rather than directly on distances.

An initial comparison of the CBD with the standard model can be made in terms of Table 2. The cell entries represent the probabilities, or long-term relative frequencies, of the patterns in the two categories, and therefore also the relative frequencies with which the distance between a test pattern and the patterns of each array enter into the summation of total distance between the test pattern and the category. For test pattern 10 and category A, the four distances d_{xy} are 0, 2, 1, and 1, so Eq. (4) takes the form

$$\begin{aligned} D_{x\mathrm{A}} &= \theta^2 \cdot 0 + (1 - \theta)^2 \cdot 2 + \theta(1 - \theta) \cdot 1 + (1 - \theta)\theta \cdot 1 \\ &= 2(1 - \theta), \end{aligned}$$

and the similarity of this exemplar to the category is, from Eq. (5),

$$S_{x\mathrm{A}} = e^{-2c(1-\theta)}.$$

Similarly,

$$D_{x\mathrm{B}} = 2\theta$$

and

$$S_{x\mathrm{B}} = e^{-2c\theta}.$$

Entering the two total similarities into Eq. (6) yields

$$P_{10}(\mathrm{A}) = \frac{1}{1 + e^{-2c(2\theta-1)}}. \tag{9}$$

This function is obviously different in form from the corresponding function in the standard model, Eq. (1), but the material question is whether the relationships between $P_{10}(\mathrm{A})$ and θ are distinguishably different in the two cases. Numerical computations show that the agreement is very close when similarities are high and differences are relatively small even at very low levels of similarity. For example, if $s = .9$, we can choose a value of c (equal to .10) for which the values of $P_{10}(\mathrm{A})$ are .510, .520, and

.530 at θ values of .6, .7, and .8, respectively, computed from Eq. (9) compared to .511, .521, and .532 from Eq. (1). If $s = .1$ and $c = 1.7$, the corresponding values are .664, .796, and .885 from Eq. (9) and .659, .796, and .896 from Eq. (1). It seems clear that with the one free parameter of each model estimated from data, we could not expect to discriminate the two versions of the exemplar model in the independent-feature situation. The question of whether the same would be true if there were more features per exemplar or more than two values of each feature requires further investigation.

A city-block-distance (CBD) prototype model. Now we are in a position to formulate a prototype model that differs from an exemplar model in only one respect: Categorization will depend on the distances of test exemplars to average prototypes rather than cumulated distances to the individual exemplar representations stored in memory arrays. It will be assumed that from the entries in the memory arrays, the learner forms a mental representation of the relative frequency with which each feature occurs in each category (exactly as in a feature-frequency model) and that the vectors of average feature values computed from these representations constitute the average prototypes.

To illustrate again in terms of Table 2, it is apparent from the cell entries that the relative frequencies of Features 1 and 2 in category A are θ and $1 - \theta$, respectively, and in the notation used these are also the average feature values. Hence the average prototype of category A is θ, $1 - \theta$, and, similarly, the average prototype of category B is $1 - \theta$, θ. For test exemplar 10, the distance $D_{10,A}$ from the A prototype is $(1 - \theta) + (1 - \theta) = 2(1 - \theta)$ and the distance $D_{10,B}$ from the B prototype is 2θ (Fig. 1). Entering these values in Eq. (7), we find the probability of assigning the exemplar to category A to be

$$P_{10}(A) = \frac{1}{1 + e^{-2c(2\theta - 1)}}, \qquad (10)$$

which is identical to the corresponding expression for the CBD exemplar model. Clearly, similar equalities obtain for the other exemplar patterns of Table 2 and, further, for any categorization task defined by equally frequent categories and independent, binary-valued features.

Thus, the different processing operations assumed in the exemplar and prototype models do not necessarily lead to different predictions about categorization probabilities. The specification of city-block distance is critical to the results on equivalence, and we should not expect to find similar equivalence of the two types of models in general when the restriction to city-block distance is removed. Neither, however, should we assume that exemplar and prototype models will yield distinguishably different predictions for any type of situation unless the models being compared are actually specified and differential implications derived.

Feature models. In order to parallel the presentation of exemplar models, I will start with what may qualify as the simplest baseline model of the family exemplified by the feature-frequency model of Franks and Bransford (1971) and the cue-validity model of Reed (1972). The memory array generated by a sequence of learning trials has the same form as for the exemplar models, and again it is assumed that, on each trial, each feature of the exemplar presented has some fixed probability, p, of being stored. However, in the feature model, the array is not accessed by rows. Rather, a memory representation is formed, and updated trial-by-trial, for the proportion of occurrences of each feature value in each column of the array. In effect (and actually in a computer program used to generate theoretical predictions), two counters for 1's and two counters for 0's are associated with each feature (each column of the array), one counter of each pair being incremented when the feature value is stored on a category A trial and the other when the value is stored on a category B trial. The contents of these counters are converted to proportions, which represent the learner's current estimates of the probability of occurrence of each feature value in each category. I will denote by $f_{\mathrm{X}i}$ this representation of a specified value for feature i on category X trials. In this notation, the likelihood function for category A and exemplar x on any trial is

$$L(\mathrm{A},x) = \pi_{\mathrm{A}} f_{\mathrm{A}1} f_{\mathrm{A}2} \cdots f_{\mathrm{A}N}, \tag{11}$$

and for category B,

$$L(\mathrm{B},x) = \pi_{\mathrm{B}} f_{\mathrm{B}1} f_{\mathrm{B}2} \cdots f_{\mathrm{B}N}, \tag{12}$$

where π_x, as before, denotes the probability that category X is represented on any trial, N is the number of features, and f_{A_i} or $f_{\mathrm{B}i}$ denotes the learner's estimate of the probability of the observed value of the i^{th} feature of the exemplar in category A or B, respectively. If a learner had observed and stored the exemplar patterns

10 A
11 A
00 B
11 B
01 A

and then were tested with pattern 10, the likelihoods that this pattern arose from categories A and B would be

$$L(\mathrm{A},10) = (3/5)(2/3)(1/3)$$

and

$$L(\mathrm{B},10) = (2/5)(1/2)(1/2).$$

On the basis of Bayes' theorem, the probability that category A gave rise to exemplar x in the general case is

$$P_x(\mathrm{A}) = \frac{L(\mathrm{A},x)}{L(\mathrm{A},x) + L(\mathrm{B},x)}, \tag{13}$$

which in the illustration would be

$$\frac{(3/5)(2/3)(1/3)}{(3/5)(2/3)(1/3) + (2/5)(1/2)(1/2)} = \frac{(3/9)}{(3/9) + (1/4)} = 4/7.$$

In this simple feature learning model, the learner's probability of assigning a test exemplar x to a category X is taken to be equal to the estimated probability, $P_x(\mathrm{X})$, that the exemplar was generated from category X, so in the example probability of categorization response A is 4/7.

Conditions of Equivalences between Models

Similarity in a feature-frequency model. At first thought, it might seem reasonable to compare this simple representative of the feature-frequency model family to the basic exemplar model. However, the two models lack comparability because the similarity parameter, s, of the exemplar model has no counterpart in the feature model. To remedy this lack of comparability, a problem overlooked in previously reported tests of feature, or cue-validity models versus other types, I will introduce a feature-frequency model that is strictly comparable with the basic exemplar model as regards the meaning of the parameters.

Although traditionally feature-frequency models have not taken account of gradations of similarity, there is no reason why they cannot do so. A way of accomplishing the task can be illustrated in terms of the memory array of the example used in the calculation just given for a learner presented with test exemplar 10. In the estimation of the likelihood of this pattern in category A, comparison of the first feature of pattern 10 with the first elements of the A exemplars yields $f_{\mathrm{A}1} = 2/3$. This calculation assumes that a feature value 1 in the test exemplar is perfectly discriminated from a value 0 in the memory array and a 0 in the test exemplar from a 1 in memory. Just as in the exemplar model, we could, more generally, allow for the possibility that perceived and remembered feature values can have various degrees of similarity, or confusability. Following the same line of implementation of this idea as in the exemplar model, we would, when counting the number of matches between a perceived feature value and a column of the memory array, enter a 1 if the perceived value were 1 and the memory value 1 (or 0 and 0), but a quantity s $(0 \leqslant s \leqslant 1)$ if the perceived value were 1 and the memory value 0 (or

0 and 1) (see Footnote 4). Then, in the illustration above, comparison of the first feature of the test exemplar 10 with the first elements of the A exemplars would yield $f_{A1} = (2 + s)/(3)$, the second feature of 10 with the second elements, $f_{A2} = (1 + 2s)/(3)$, and the other two comparisons, $f_{B1} = f_{B2} = (1 + s)/(2)$. In this case, the estimated probability of category A given the test pattern would be

$$P_{10}(A) = \frac{\dfrac{2+s}{3} \cdot \dfrac{1+2s}{3}}{\dfrac{2+s}{3} \cdot \dfrac{1+2s}{3} + \dfrac{1+s}{2} \cdot \dfrac{1+s}{2}},$$

and we note that if $s = 0$, this expression reduces to 8/17, the result obtained previously, as it should.

This feature-frequency model can be instructively compared with the basic exemplar model, since both models use the same array information and both take account of variations in similarity between perceived and remembered feature values; they differ only with respect to how the information is applied to the task of categorization. A convenient first step is to consider the independent-feature design of Table 2, for which categorization probabilities have already been derived for the exemplar model. To apply the feature model, we again interpret the entries in Table 2 as the relative frequencies of occurrence of the different exemplar patterns in the memory arrays for the two categories. (Each of these values would be multiplied by n, the total number of learning trials, to obtain expected frequencies, but since n would divide out of all expressions for probability estimates, it is ignored.) Here, if s were equal to 0, the estimate of f_{A1} would be $\theta^2 + \theta(1 - \theta) = \theta$, the sum of the relative frequencies given in the first column of Table 2 for patterns having a 1 in the first position, and the estimate of f_{A2} would be $\theta^2 + \theta(1 - \theta) = \theta$, the sum of relative frequencies for patterns having a 0 in the second position. However, removing the restriction $s = 0$, we obtain

$$f_{A1} = \theta^2 + \theta(1 - \theta) + (1 - \theta)^2 s + (1 - \theta)\theta s,$$

the last two items being the relative frequencies of patterns having a 0 in the first position, each weighted by the similarity parameter, s. Similarly, we obtain

$$f_{A2} = \theta^2 + \theta(1 - \theta) + (1 - \theta)^2 s + (1 - \theta)\theta s.$$

When the corresponding expressions are obtained for category B, and the appropriate substitutions are made in Eq. (13), the result, after some algebraic manipulation, proves to be

$$P_{10}(\mathrm{A}) = \frac{[\theta + (1 - \theta)s]^2}{[\theta + (1 - \theta)s]^2 + (1 - \theta + \theta s)^2},$$

in exact agreement with Eq. (1), derived above for the exemplar model. Similar computations for the other three exemplar patterns show that the models agree exactly in each case. Therefore, for this case of two independent, binary-valued features, the two models yield the same predicted categorization probabilities for all values of s. Even though the basis of categorization seems intuitively quite different in the two models, there clearly are many situations in which they cannot be distinguished on the basis of empirical data.

The independence condition. How general is this result? Does the equivalence in this case depend on the particular symmetries in Table 2, or are the models always equivalent, or only equivalent for categories based on independent feature distributions? We can progress toward an answer by analyzing the general class of categorization tasks for patterns of two binary-valued features, represented in Table 1, the only restriction on the pattern relative frequencies (the cell entries) being that they sum to unity for each category. For this general class, it can be shown (Appendix 2) that if and only if the features are independent, that is, in terms of Table 2,

$$(u + w)(1 - v - w) = u,$$
$$(x + z)(1 - y - z) = x,$$

and so on, the exemplar and feature models are equivalent.

The results on equivalence of the basic exemplar and feature-frequency models for independent-feature situations hold even when the storage parameter, p, has a value intermediate between 0 and 1, as shown in Appendix 3. The canonical memory array has to be enlarged to include partial patterns (Table A1), but nonetheless the categorization probabilities for all possible test patterns are the same for both models. The proof of equivalence given in Appendix 3 holds only for asymptotic probabilities, but in computer simulation of the models the equivalence is close beyond the very early trials of a learning series. In both models, the asymptotic level of categorization performance is determined mainly by the value of the similarity parameter, s, and the rate of approach to asymptote by the storage probability, p, as illustrated in Fig. 2. Under some circumstances, the asymptotic level is actually independent of p (so long as it is greater than 0), and in general it is much more sensitive to variation in s than to comparable variation in p (Appendix 1).

The criterion of independence can be met even in cases that would be characterized as rule-defined categorizations. Suppose, for example, that

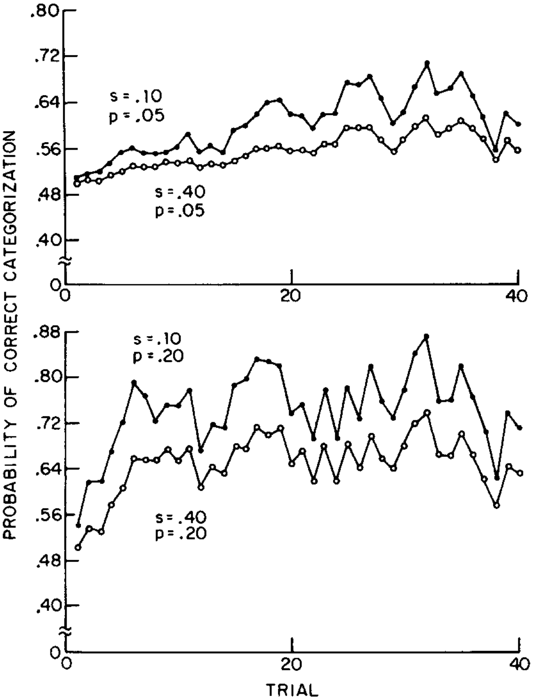

FIG. 2. Illustrative learning curves computed from the standard exemplar or feature models, equivalent in this case, for the task design of Table 2 with θ set equal to 0.8. Comparison of the curves within each panel shows the effect of variation in the similarity parameter, *s*; comparison of corresponding curves between panels shows the effect of variation in the storage parameter, *p*.

the cell entries (exemplar probabilities) in Table 1 had the following values:

½	0
0	½
½	0
0	½

Categorization would, then, depend entirely on the first feature, being A if the first feature had value 1 and B if it had value 0; and the second feature would be entirely invalid. However, independence would be satisfied, and thus both models would yield the same categorization probabilities. Substituting these exemplar probabilities into Table 1, and then into the expressions for $P_{10}(\mathrm{A})$ in the exemplar and feature models (Appendix 2, Eqs. (A4) and A5)), for example, yields

$$P_{10}(\mathrm{A}) = \frac{1}{1 + s} \tag{14}$$

for both models. If the learner distinguishes perfectly between a feature value 1 in a perceived exemplar and a value 0 in the memory array (or a 0 and a 1), that is, $s = 0$, then the pattern 10 is always assigned (correctly) to category A; but as s departs from 0, categorization performance declines, approaching chance as s approaches 1.

In contrast to this instance of a rule defined categorization for which the exemplar and feature models are equivalent, cases can readily be constructed in which their predictions differ widely. An example is obtained by setting the exemplar probabilities in Table 1 to the values

$$\begin{array}{cc} \frac{1}{2} & 0 \\ \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \\ 0 & \frac{1}{2} \end{array}$$

Now we have a task in which A is always the correct category if the values of the two features differ and B is correct if they are the same. For this case, the categorization probability of pattern 10 is equal to ½ for the feature model, regardless of the value of s (and the same result holds for the other patterns, since both features have zero validity), but for the exemplar model,

$$P_{10}(\mathrm{A}) = P_{01}(\mathrm{A}) = \frac{1 + s^2}{1 + 2s + s^2}$$

and

$$P_{11}(\mathrm{A}) = P_{00}(\mathrm{A}) = \frac{2s}{1 + 2s + s^2} .$$

All of these probabilities reduce to those of the feature model when $s = 1$, but the first two diverge to 1 and the second two to 0 as s decreases toward 0. This example illustrates the general point that the difference between the categorization probabilities predicted by the two models (for any given value of $s < 1$) increases as the intercorrelation of the feature values increases from independence to perfect correlation.

Metric restrictions and the role of feature validity. Although the independence condition is critical for equivalences between standard forms of the models (that is, forms with no restrictions on the similarity space in which comparison are made), its role is replaced by another condition in the case of distance-based models. We have seen that in the independent-feature design of Table 2, the versions of the exemplar model and the average-prototype model based on a city-block metric are equivalent. However, independence of features is not critical, and, in fact, the CBD exemplar model can be shown to be equivalent to the CBD average-prototype model for the more general case of Table 1, in which independence is not assumed (Appendix 4). In both models, the categorization probability for exemplar pattern 10, for example, proves to be

$$P_{10}(\mathrm{A}) = \frac{1}{1 + \mathrm{e}^{-c(u+y-v-x)}}, \tag{15}$$

in the notation of Table 1. The equivalence does not hold, however, if any metric other than city-block is assumed for the similarity space (as demonstrated for the case of a Euclidean metric in Appendix 4).

In most situations, the individual features of category exemplars carry some validity (that is, have nonzero correlations with categories). However, it is possible to design experiments in which all individual features are invalid and only combinations of features carry information about category assignment; and category learning has been demonstrated under this condition (Estes, 1972; Estes, 1986, Experiment 2). Referring to Table 1, it can be seen that both features are invalid if the following equalities hold

$$u + w = x + z$$

and

$$v + w = y + z,$$

the first and second lines equating the probabilities of Value 1 for the first and second features, respectively, in categories A and B. If the second equation is subtracted from the first, the result is

$$u - v = x - y,$$

and when this equality holds, Eq. (15) reduces to

$$P_{10}(\mathrm{A}) = \frac{1}{1 + \mathrm{e}^{-c(0)}} = ½;$$

and the same result holds for the other pattern probabilities. Thus, in both the CBD exemplar model and the CBD average-prototype model, learning can occur only if one or more of the individual features are at least partially valid predictors of categories.

If the restriction to a city-block metric is removed, then a distance-based exemplar model can predict learning even when all features are invalid, provided only that some combinations of features are at least partially valid predictors (Appendix 4).

Similarity and Likelihood

Generalized likelihood. In the exemplar models discussed so far in this article, as in Medin and Schaffer (1978), the characterization as "similarity-based" is somewhat misleading, since there is an implicit role of likelihood in the use made of relative frequencies of categories as well as of exemplars and features within categories. This fact may be apparent from Eq. (2), but it can be pointed up in terms of the following illustrative case: Suppose that after three learning trials the memory array is

A	B
1 1	0 0
	0 1

and a test is given with a new pattern, 10. In terms of the simple exemplar model, the similarity of test exemplar 10 to category A is s; the similarity to category B is $s + s^2$; and the probability of categorization of the pattern as an exemplar of A is

$$P_{10}(\mathrm{A}) = \frac{s}{2s + s^2} = \frac{1}{2 + s} .$$

Although the pattern 10 seems more similar to category A than to category B, the probability of assigning it to category B is greater than the probability of assigning it to A (unless s is equal to 0) simply because there are more exemplars stored in the B array. In effect, the model combines information about category frequencies, or probabilities, with information about similarities when it generates categorization responses.

It is easy to show that if effects of similarity are eliminated, the exemplar model becomes a simple pattern-likelihood model. Referring to Table 2 and Eq. (1), for example, if we let s equal 0, the categorization probability for pattern 10 reduces to

$$P_{10}(\mathrm{A}) = \frac{\theta^2}{\theta^2 + (1 - \theta)^2} .$$

Since θ^2 is the probability of occurrence of pattern 10 on category A trials and $(1 - \theta)^2$ the probability on category B trials, and the categories are equally likely, this expression for $P_{10}(\mathrm{A})$ is simply the Bayesian probability of category A given the information that the test exemplar is pattern 10; and analogous results are readily derived for the other patterns. This reduction of the relative similarity expressions of the exemplar model to category likelihoods when s equals 0 does not depend on feature independence or the restriction to two-feature patterns, but, rather, is completely general.

It is important to note also that there is no discontinuity when s equals 0 in the transition from an exemplar-similarity to an exemplar-likelihood model. As s becomes very small, the relative similarity expressions become closer to category likelihoods. Thus, the relative similarities, or categorization probabilities, of the exemplar model can be viewed as generalized expressions for category likelihood—the likelihood of a category given not the objective test pattern, but rather the test pattern as perceived and encoded by the learner.

A true similarity model. The fusing of exemplar similarity and category frequency information in current exemplar models can be eliminated by the simple tactic of normalizing the measures of exemplar–category similarity, that is, dividing each summed similarity by the number of comparisons entering into it. In the illustrative example, the similarity of pattern 10 to category A would remain unchanged at the value s but the similarity to B would become $(s + s^2)/(2)$. The new probability of categorizing pattern 01 as an A would be

$$P_{10}(\mathrm{A}) = \frac{s}{s + \dfrac{s + s^2}{2}} = \frac{2}{3 + s}\,.$$

Now categorization probability depends only on the average similarities of the test exemplar to the alternative categories. In this example, if s approaches 1, so that all patterns are maximally similar, $P_{10}(\mathrm{A})$ approaches 1/2, rather than 1/3 as in the standard exemplar model, since the different frequencies of stored patterns in the two categories exert no influence. The implications of the two models are also quite different if s approaches (but does not reach) 0, so that all patterns are maximally dissimilar. In the standard model, $P_{10}(\mathrm{A})$ approaches 1/2, since there is no difference in the total similarity to the two categories and the category frequency information is not allowed to operate; in the revised model, however, $P_{10}(\mathrm{A})$ approaches 2/3 since the average similarity of pattern 10 to category A is approximately double the average similarity to B at small values of s.

This modification of the basic exemplar model qualifies as a pure similarity-based model, since categorization is independent of category base rates. It is apparent that both standard and pure-similarity versions of the distance-based exemplar and prototype models could be formulated. In all cases, the standard model and the pure-similarity version yield identical predictions when categories are equally probable, but they diverge as category probabilities deviate from equality, so differential empirical tests of the two versions should be straightforward to accomplish.

Two aspects of similarity. Virtually all forms of learning require the learner to discriminate stimuli or stimulus patterns; and up to a point this generalization is true of category or concept learning, since exemplars of different categories must be discriminated in order to be appropriately classified. Thus category learning, like other forms, must in part be a matter of overcoming similarities. However, the development of exemplar models has brought out the point that there is another aspect to similarity, for new exemplars of categories can sometimes be appropriately assigned on the basis of their similarity to remembered instances previously encountered. In terms of the similarity parameter s of the specific models developed in this study, a low value of s always favors performance on old exemplars, with s equal to 0 being optimal; however, a zero value of s means chance performance on new exemplars, and since the same is true when s equals 1, evidently some intermediate value must be optimal. No general statement can be made about the precise level of similarity that is optimal, and results for different cases turn out to be surprising in some instances.

To investigate the way the optimal level of similarity depends on specific conditions, we can start with the general two-feature categorization situation of Table 1 and consider the memory array that would arise if some particular exemplar pattern, say pattern 10, happened not to be presented on the first $n - 1$ trials of the learning series, the result of which is illustrated in Table 3, based on a derivation given in Appendix 5. In terms of the simple exemplar model, the expected similarity of exemplar 10 to the array for category A on its first occurrence would be

$$\mathrm{Sim}(10,\mathrm{A}) = [vs^2 + (1 - u - v)s]k$$

and the similarity to the array for category B

$$\mathrm{Sim}(10,\mathrm{B}) = [ys^2 + (1 - x - y)s]k,$$

TABLE 3
Memory Array Resulting from Design in Table 1 Given No Occurrence of Pattern 10 over First n Trials[a]

	Category	
Pattern	A	B
1 0	0	0
0 1	$\frac{v}{1\text{-}u}$	$\frac{y}{1\text{-}x}$
1 1	$\frac{w}{1\text{-}u}$	$\frac{z}{1\text{-}x}$
0 0	$\frac{1\text{-}u\text{-}v\text{-}w}{1\text{-}u}$	$\frac{1\text{-}x\text{-}y\text{-}z}{1\text{-}x}$

[a] All entries multiplied by k (defined in text) are expected frequencies of patterns in the memory array.

where

$$k = \frac{(n - 1)}{2 - u - x} .$$

As usual, the predicted likelihood of the learner's assigning the pattern to category A would be equal to the similarity to A over the sum of similarity to A and similarity to B. The quantities u, v, x, and y in these expressions are the pattern probabilities from Table 1.

Now several different results can be obtained for different combinations of the pattern probabilities. First of all, if the features are independent, then it can be shown that the probability of correct categorization of pattern 10 on its first occurrence increases from a minimum at $s = 1$ to a maximum as s approaches 0, illustrated in the Task I column of Table 4. If features are not independent, then quite different results occur in different cases. In one, illustrated in the Task II column of Table 4, the categorization probability for pattern 10 increases from near chance at $s = 1$ to a much higher value as s approaches 0, whereas under the kind of arrangement illustrated in the Task III column of Table 4 the 10 categorization probability decreases from near chance at $s = 1$ to near 0 as s approaches 0. However, in each of the latter two cases, it should be noted that there is a discontinuity at $s = 0$. For the case in the middle column of Table 4, the categorization probability increases as s ap-

TABLE 4
Categorization Probability at Different Values of s in the Exemplar Model for Three Categorization Tasks[a]

	Task					
	I		II		III	
Pattern	A	B	A	B	A	B
1 0	.56	.06	.10	0	.10	0
0 1	.06	.56	0	.60	.90	.20
1 1	.19	.19	.90	.40	0	0
0 0	.19	.19	0	0	0	.80
			$P_{10}(A)$			
s						
1	.32		.47		.47	
.5	.38		.56		.33	
.05	.48		.68		.05	
.01	.50		.69		.01	
0	.50		.50		.50	

[a] The task designs are summarized in the upper portion of the table in the format of Table 1; categorization probabilities are the cell entries in the lower portion.

proaches 0, but then drops to chance, .50, if s actually equals 0. Similarly, in the case shown in the third column of Table 4, the categorization probability decreases to a very low value as s approaches 0, but would jump to the chance level of .50 if s actually reached 0. With regard to optimization, in the independent feature case, it is clear that the learner is best off with s as small as possible, which is, of course, also optimal for performance on old exemplars. In Task II of Table 4, the same is true, but in Task III best performance on new exemplars would result from s equal to 1 and poorest for s values that become very small, though not actually equal to 0.

Few data are available to allow assessment of how human learners actually adjust to these optimization conditions.[5] The one substantial set of data available at the time of writing was generated by subjects learning categorizations of bar-chart patterns into categories defined by independent feature distributions (Estes, 1986). For this study, it was found that when the parameters of s and p of the exemplar model were estimated from the learning data for four groups of subjects, separate estimates being obtained for each of four successive blocks of 80 trials, the overall best fit to the learning data was obtained when s was set equal to .50 and p to .05. However, with p fixed a .05 and s estimated separately by blocks, the best estimates of s, averaged over the four groups, proved to be .60, .21, .16, and .16 for the four successive blocks of trials. Clearly the effective similarity value, as indexed by the value of s, decreased substantially as learning progressed.

One might well raise the question of why similarity should change in any regular way over the course of learning, since there is no obvious reasons why the properties of the features of stored exemplars should change over trials in the direction of greater discriminability. We may get a clue to an answer by looking at the situation from the viewpoint of signal detectability theory (Swets, 1964). Consider the problem for a learner when a feature of a perceived exemplar is being compared to an actually mismatching feature of a memory vector. In terms of detectability theory, the perceived feature gives rise on different occasions to a distribution of internal states and the same is true of the feature of the memory vector, and in general the two distributions would be expected to overlap. Consequently, on some occasions, when a comparison is made, the internal state activated by the memory element will fall within the range of those that might be activated by the perceived feature, so that

[5] Over an entire learning series, optimization would depend on how the learner weights the values of correct categorization of new and old exemplars. Testable predictions could readily be generated for conditions designed to vary the relative weights.

the perceived and judged feature may be judged to match even though they are actually different.

In terms of signal detectability theory, the way the individual deals with this uncertainty is to set some criterion, that is, some value on the continuum of internal states generated by feature comparisons, above which the perceived and remembered features are judged to match and below which they are judged to mismatch. Further, there is evidence from studies of signal detection and recognition (Healy & Kubovy, 1977; Murdock, 1974; Swets, 1964) that human observers are capable of learning to adjust criteria so as to increase rates of payoffs for correct decisions. Thus it seems a reasonable hypothesis that the systematic changes in the memory parameter s in my data reflect criterion shifts that the individuals learn to make during lengthy experience with the categorization task.

GENERAL DISCUSSION

Summary of Results on Interrelations among Models

The array framework has provided an effective basis for an organized attack on problems of comparability among category learning models. However, the models have proliferated even while being studied, so an overview of the family developed in the preceding sections may be useful. The following qualitative summary of interrelationships is supplemented in Appendix 6 with a summary of the principal predictive formulas and brief characterizations of the algorithms used to compute learning functions for the various models. However, the details of the particular cases chosen for analysis in this article are less important than the demonstration that by imbedding the models in a common framework, we can show when various particular models are or are not empirically distinguishable and can specify conditions under which we can test predictions from representatives of different model types that differ in only a single processing assumption.

The simplest, yet in a sense the most general purpose, member of the family is the basic exemplar-memory model, incorporating the assumptions that exemplar representations are stored in a chronologically ordered array and that test exemplars are compared to elements of the array via a serial search process. An individual who processes information in accord with this model can learn, at least to some degree, any categorization task, provided only that the features in terms of which he encodes category exemplars are correlated, individually or in combinations, with category occurrences. It would be straightforward to formulate variants of the model in which only recent (the last k) exemplars are searched or in which only exemplars with some specified criterion property are stored in the array, but these possibilities have not been formally explored.

The weighted-vector exemplar model differs from the basic model only in that the memory array is assumed to contain only a single representation of each exemplar pattern that has been encoded during a learning series, together with information about its frequency of occurrence in each category. Categorization of new test exemplars is based only on comparisons with the vectors in the canonical array, the result of a comparison between a perceived pattern and a memory vector being weighted by the relative frequency of the latter. These two forms of the exemplar model cannot be distinguished on the basis of categorization data.

The basic and the weighted-vector exemplar models both embody what Nosofsky (1984b, 1986) terms the "mapping hypothesis"—that is, the assumption that the probability of assigning an exemplar to a category is equal to the sum of the probabilities that it is identified (correctly or incorrectly) as any one of the remembered members of the category. In contrast, it is assumed in distance-based exemplar models that categorization depends only on the summed distance in a "similarity space" between the test pattern and the members of the memory array for each category. For appropriate corresponding values of their parameters, the simple exemplar-memory model and the distance-based exemplar models in some cases yield quite similar predicted values of categorization probabilities; the conditions remain to be determined under which the different forms of the exemplar model can be empirically distinguished when their parameters are evaluated from data.

A feature, or cue-validity, model, in which categorization depends on the computed likelihoods that the feature combination of a test exemplar would arise from sampling the alternative categories, has been shown to be very similar in its predictions, and in fact asymptotically isomorphic, to the standard exemplar-memory and weighted-vector models when features are independent. The sense of *independence* intended here is illustrated in Table 2: The probability of any combination of two or more features occurring in a category exemplar is equal to the product of their individual probabilities. When independence does not hold (i.e., features are correlated), the exemplar models generally predict higher performance levels than the feature model with the same parameter values.

The prototype models investigated within the array framework assume that categorization is based on comparison of the distances of a test exemplar from the average prototypes of the alternative categories in a similarity space. When the metric associated with the space is city-block distance, often assumed to be appropriate for separable stimulus dimensions (Garner, 1974; Nosofsky, 1984b), and categories are equally probable, the prototype model proves to be equivalent to the CBD exemplar model, regardless of independence or nonindependence of features. When the metric is Euclidean (often assumed appropriate for integral stimulus dimensions), differential empirical tests of the models may be accomplishable by categorization data.

Evidence on Component Processes

Retrieval of exemplar representations. Interpretations of categorization in terms of exemplar memory have been characterized by the assumption that stored exemplar representations are retrieved and compared with the perceived pattern of a test exemplar. The most direct support for this assumption has come from findings that old (that is, previously experienced) exemplars are categorized more accurately than new exemplars of presumably similar difficulty (Brooks, 1978; Elio & Anderson, 1981). However, those observations tell little about the generality of the process, and it should be noted that they all come from studies in which the number of different exemplars presented to a learner is relatively small. In experiments employing larger numbers of different patterns, I obtained evidence for exemplar retrieval only when exemplars were repeated within the learner's short-term memory span or when the number of repetitions was large (Estes, 1986). In the former case, the evidence took the form of an advantage for old exemplar patterns that included unique retrieval cues over old patterns that did not; in the latter case, the evidence was a selective facilitation of response to old patterns by instructions to attend to exemplar similarities. Thus, although there is good reason to believe that a process of comparing stored to perceived exemplar patterns does occur, more evidence is needed concerning its generality.

Use of feature and pattern frequencies. Although feature-frequency, or cue-validity, models of categorization have been out of favor since rather negative results of an empirical test were reported by Reed (1972), there is substantial reason to believe that information about feature frequencies is stored and used in recognition (Bower, 1972), and, at least when the number of exemplar patterns involved in a task is large and repetitions relatively rare, also in categorization. In a study meeting these stipulations, a feature-frequency model yielded accurate, parameter-free predictions of the asymptotes of learning functions and the results of transfer tests on selected exemplar patterns (Estes, Burke, Atkinson, & Frankmann, 1957).

A large literature on "memory for frequency" leaves no doubt that learners can store information about pattern frequencies (e.g., Hintzman, 1976; Underwood & Freund, 1970; Wells; 1974; Whitlow & Estes, 1979) but does not address the question of whether the information is used in categorization or other tasks outside the domain of recall and recognition tests. There is some evidence for an affirmative answer in experiments showing that when the number of distinct exemplar patterns is small and individual features are invalid, asymptotes of learning functions can be predicted by a simple pattern-frequency model (Estes, 1972). The same study indicates that when conditions early in learning direct attention toward features rather than patterns, learners continue to respond on the basis of feature frequencies even when categorization on the basis of pattern frequencies would be more efficient.

Comparisons to prototypes. Prototype models have been popular since the studies of Franks and Bransford (1971) and Posner and Keele (1971), but mainly because they offer explanations of such phenomena as better categorization of new prototypic than old nonprototypic patterns rather than because there is direct evidence for processes of judging distances from test patterns to prototypes. However, it has become apparent that alternative explanations in terms of feature-frequency (Reitman & Bower, 1973) or exemplar-memory (Busemeyer et al., 1984; Medin & Schaffer, 1978; Smith & Medin, 1981) mechanisms are available, so there is need for more direct tests of the processing assumptions of the prototype models. One procedure has been to present prototypic stimuli, or descriptions of them, to subjects in advance of a learning series and look for a facilitation of learning and transfer performance (Medin, Altom, & Murphy, 1984). However, any observed facilitation could be interpreted in an exemplar-memory model in terms of the idea that the prototypic pattern is stored in the memory array, quite possibly in a privileged position or with more effective retrieval cues than other observed exemplars, owing to the special attention it has received. It may be that compelling differential tests of prototype abstraction versus other mechanisms can be achieved only by well-controlled comparisons of models of the kind that can be accomplished within the array framework.

Deterministic vs probabilistic response rules. Rules for getting from memory representations to categorization responses vary widely over extant models. Prototype models have traditionally assumed a deterministic process, a test exemplar being assigned to the category from whose prototype its distance in the similarity space is smallest. Cue-validity models are usually classed as probabilistic, but the version formulated and tested by Reed (1972) assumed a deterministic rule. In the exemplar models of Medin and Schaffer (1978) and Nosofsky (1984b), response selection is probabilistic (categorization probabilities for an exemplar being in the ratio of its similarities to the alternative categories), and the same is true for the likelihood-based model of Fried and Holyoak (1984).

It is not feasible to decide empirically between deterministic and probabilistic rules by comparing models that differ in their response rules, because in every case the models differ in other respects. It would be possible to test alternative versions of a single model, for example, the standard exemplar model, with an alternative version in which a deterministic response rule replaces usual probabilistic rule. However, there would be difficulties. To apply the models in an experiment, the storage and similarity parameters have to be estimated from the data, and the difference in efficiency of the two response rules can trade off with values of the parameters (a higher value of p or lower value of s in the standard version compensating at least in part for the greater efficiency of the deterministic rule in the other). Further, even if a learner adopts a deterministic rule, it may not be possible to apply the rule infallibly, so errors in judging which category array is more similar to a test exemplar may produce the same kind of noise in the data that would result from a probabilistic decision rule.

A better framework for an approach to the problem is offered by the distance-based models of the array family. In all of these models, the expression for categorization probability takes the form

$$P_{\text{ex}}(\text{A}) = \frac{1}{1 + e^{-cd}}$$

where d denotes the difference in distances from the test exemplar to the arrays (in the exemplar models) or the prototypes (in the prototype models) of the categories A and B. If the parameter c has a relatively small value, $P_{\text{ex}}(\text{A})$ will vary between 0 and 1 as a function of the size of d, and we would speak of a probabilistic response process. If c is sufficiently large, however, the exponential term will have a value close to 0 if d is negative and a very large value if d is positive, yielding a value of $P_{\text{ex}}(\text{A})$ of approximately 1 in the former case and approximately 0 in the latter. Thus, for large c the response rule becomes indistinguishable from a deterministic rule. The relevance to the problem at hand is that we can fit the distance-based models to empirical learning functions and allow the data to speak to the issue via the value of c that proves to be required for best fit. It may turn out that either a deterministic or a probabilistic rule is generally the preferred assumption, or that the choice varies with conditions.

Descriptive Adequacy of the Models

The focus of this paper is on extending major classes of categorization models to handle learning as well as asymptotic performance and on understanding relationships among the classes of models. Our interest in this task is predicated, of course, on the assumption that these models are useful for the interpretation of empirical phenomena. With this thought in mind, I undertook, in collaboration with Robert Nosofsky, a survey of the relevant literature at the beginning of the theoretical studies leading to this paper. We found 10 studies published between 1961 and 1981, all of which presented data on the relative difficulty of different categorization tasks or of different category exemplars within a task, usually in terms of error proportions on a block of test trials given at the end of a learning series. In all cases where it had not been done by the authors, we fitted the Medin and Schaffer (1978) model[6] to the data (Nosofsky being responsible for the computations). In all cases, we obtained fits that looked intuitively satisfactory.

The notion of "satisfactory" can be made somewhat more specific. Four of the studies examined (Elio & Anderson, 1981; Medin, 1983 [the data was available to us in advance of publication]; Medin & Schaffer, 1978; Medin & Smith, 1981) were homogeneous enough with regard to design and number of data values to make an overall analysis feasible. The Medin and Schaffer model, with allowance for differential weighting of stimulus attributes, was fitted to all of the data sets with a least-squares criterion. The standard errors of estimate (standard deviations of difference between observed and theoretical response proportions) ranged from .019 to .070 with a mean of .048.

Data for trial-by-trial changes in response proportions during learning were not available from any of the published studies. Such data were obtained in experiments carried out in conjunction with the present study (Estes, 1986). In the first of two experiments, exemplars were generated from independent probability distributions of features; in the second experiment, individual features were uncorrelated with correct category assignment, but the features were not independent, and patterns of features were correlated with category assignment. This situation is simpler than those of the previous studies cited in that salience and diagnosticity were equated across features so that differential weighting of attributes was not needed. However, the versions of the models applied were also simpler, having only one of two free parameters rather than four or five. For four groups in Experiment 1, the average standard error of estimate was .052 for the exemplar model. To convey an impression of the goodness of fit indicated by these standard error values, the result for two of the groups (from Condition U of Experiment 1) is shown in Fig. 3, the standard errors for the upper and lower panels being .048 and .041, respectively. In Experiment 2 of that study, the model yielded a standard error of .067.

In summary, the appropriate cases of the family of array models describe learning data and asymptotic data, and tasks based on independent or on correlated features about equally well. The fits of the models are not, however, insensitive to variations in experimental procedures or instructions. In the study of Estes (1986), instructions to subjects to attend to exemplar similarities produced a shift toward a better fit for the exem-

[6] More precisely, the model fitted was the standard exemplar-memory model augmented by the addition of a weighting parameter for each stimulus attribute.

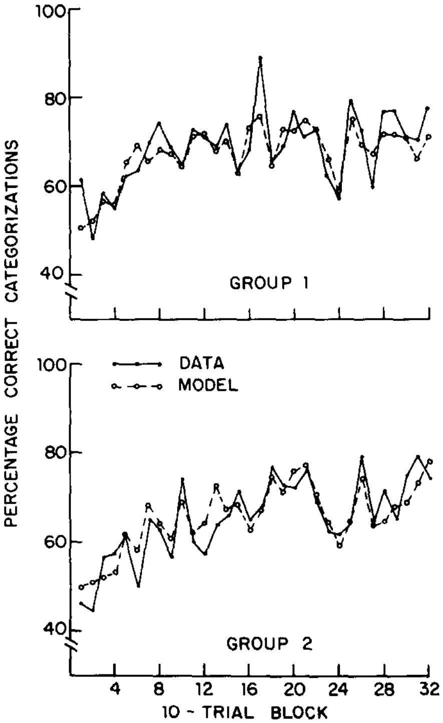

FIG. 3. Learning functions for an independent-feature task compared with predicted functions from the exemplar model. The data represent two groups differing only in the sequence of exemplar presentations, the sequence being the same for all subjects in each group (data from Estes, 1986). The parameter values are $s = .50$ and $p = .10$ for the upper panel, and $s = .45$ and $p = .05$ for the lower panel, p denoting pattern storage probability in each case.

plar model in the correlated feature situation. In Elio and Anderson (1981), sequencing of material intended to favor generalization strategies yielded slightly poorer fits of the exemplar model than control instructions (.069 vs .061 and .025 vs .019, respectively, in two experiments). On the other hand, in Smith and Medin (1981), supplying additional information to subjects in the form of prototype demonstrations did not change the fit of the exemplar model (as though the additional information were simply encoded into the memory array along with the results of observing exemplars on learning trials).

Do the array models ever really fail to describe relevant data? The only clear instance I know of occurred in a study by Gordon (1985, Experiment 3). That experiment differed from all of the others I have cited in that the category exemplars were not definable in terms of probability distributions of perceptual features but, rather, in terms of patterns of abstract properties of the exemplars. Not surprisingly, the exemplar model, interpreted in terms of exemplars as patterns of perceptual features, could not give a reasonable account of the data. Whether the model could be successfully applied with an appropriate redefinition of features is an open question. From the standpoint of predicting behavior, a limitation on all of the array models is that the investigator has to know what features the learner is using so that similarities or frequencies can be computed on the proper units.

Although it is generally considered desirable to describe data by means of a model having as few free parameters as possible, neither fitting data nor reducing the numbers of parameters are ends in themselves. One of the principal values derived from fitting a model to data arises from the transformation of raw statistics of the data to quantities that are more theoretically significant. Often this procedure provides the only way to tease out component processes in complex systems. Several illustrations of this point have arisen in applications of exemplar models. The parameters of the Medin and Schaffer (1978) version that represent dimensional weights allow freedom for data fitting, but also they can yield input to theoretical analyses. For example, Nosofsky (1984a, 1984b) was able to investigate the hypothesis that learners tend to shift their allocation of attention over stimulus dimensions toward the optimal distribution for a task by deriving the optimal distribution for the exemplar model, then fitting the model to categorization data and showing that the estimated weights did in fact shift toward the optimal distribution during learning. For another example, in the analysis of the learning data in the study of Estes (1986), it was found that the estimated value of the similarity parameter, s, of the exemplar or feature models declines systematically during learning, at least for independent-feature problems, but levels off at a value greater than zero. Thus it appears that one constituent of categorization learning is a systematic change in the learner's criterion for similarity judgments, a process whose adaptive value was discussed in an earlier section. This process is not manifest in the raw data, however, but becomes apparent only when the data have been transformed by means of a model to yield quantities interpretable as measures of effective similarity. We do not yet have a model of the process of adjusting criteria of similarity, but we are at least in a position to attack the problem now that we have a way of tracing the process by means of behavioral data.

Open Problems

Storage versus computation. Models of the exemplar and feature-frequency families differ considerably with respect to the information-processing load placed on the learner. In the latter type, feature-frequency representations are updated as each exemplar is processed, and only the computed quantities need be accessed when categorization judgments are called for. But in the former type, all exemplars encountered during learning and stored in the memory array must be accessed for comparison with the current test exemplar on each test trial. The demands on storage and retrieval processes can become quite large when learning runs to hundreds of trials, as is common in categorization learning experiments, whereas in the feature-frequency models the demands are virtually independent of the length of the learning series.

One possible shortcut in the process of memory access and comparison in the exemplar model suggested by Medin and Schaffer (1978) is that the learner might on any test trial retrieve only some one stored exemplar—the one most similar to the test pattern. However, it is not apparent how this shortcut could be accomplished without comparing the test pattern to all of those in storage in order to locate the most similar one. Another alternative to accessing the full memory array is offered by the weighted-vector exemplar models, in which the decision process involves only the representations of distinct exemplar patterns together with their relative frequency counts. However, even this process requires access to a large number of stored representations when the number of different patterns presented becomes large (for example, 256 in an experiment described by Estes, 1986) and might well entail as great a processing load as accessing the full memory array. In any event, it is not clear that the standard and the weighted-vector versions of exemplar models can be distinguished empirically, since they yield very similar predictions regarding categorization data. Conceivably reaction time measures would be useful, but this idea has yet to be explored.

Selecting and combining processes. How do the various processes discussed in this article enter into category learning? Memory for feature frequencies is a sufficient basis for categorization only when category exemplars are defined by independent features. In this case, very simple feature-frequency models have been shown to account for the detailed course of category learning. The demonstration has been accomplished only with the further restriction that features are equally diagnostic, but there is little reason to doubt that the models could be generalized to handle unequal diagnosticities just as has been done with exemplar models. The use of feature frequencies may be a preferred mode for human learners, at least when the number of patterns to be coped with is large, because the acquisition of frequency information is typically rapid (Estes, 1976; Whitlow & Estes, 1979) and, if not automatic, at least relatively light in demands on attentional capacity (Hasher & Zacks, 1979; Tzeng & Cotton, 1980).

Memory for relative frequencies of exemplar patterns provides an optimal basis for classifying previously experienced exemplars, but storing exemplar frequencies requires a steeply increasing processing load relative to feature frequencies as the number of different patterns becomes large. There is some evidence that human learners tend to rely on feature frequencies when exemplars are readily analyzable and individual features are diagnostic; use of memory for exemplar frequencies seems to come into play readily, however, when individual features are not diagnostic (Estes, 1972, 1986). The major limitation on usefulness of exemplar frequency is that it is of no help to the categorization of new exemplars. Two kinds of augmentation of simple exemplar-frequency models have been introduced for the purpose of circumventing this limitation. One is the use of observed exemplar frequencies for estimating parameters of probability distributions (Fried & Holyoak, 1984), applicable, of course, only when exemplars are generated by well-defined distribution functions. The other tactic, discussed in detail above, is to augment simple exemplar-frequency models so that estimates of category probabilities from exemplar information depend both on remembered frequencies and on similarities between perceived and remembered exemplar patterns.

The computation of generalized likelihoods from similarity judgments has the advantageous property of always being a useful mechanism regardless of the structure of the task. Generalized exemplar, or exemplar-similarity, models are advantageous also for learning categorizations that are spontaneous in the sense of not being defined for the learner by external feedback. In a situation studied by Fried and Holyoak (1984), for example, subjects observed sequences of stimuli and tried to sort them into categories, knowing only that categories existed but receiving no informational feedback as to the correctness of their responses from the standpoint of the experimenter. Learning did occur, in the sense that subjects tended to improve over trials at assigning stimuli to subsets corresponding to the experimenter-defined categories. However, the task had the special property that stimuli were sampled from well-defined distributions of familiar form whose means and variances could, in principle, have been estimated by the subjects from their observations. And, this estimation procedure would depend on the subjects being able to hypothesize the form of the distribution (normal in Fried & Holyoak's study).

Thus it is of interest to inquire whether similar learning should be possible in a broad class of situations where it would not be feasible to hypothesize the forms of distributions. If we were to try to answer this question by constructing a computer model that could accomplish the task, it would not be apparent how to make a start via the class of feature-frequency models; frequencies have to be defined relative to some criterion, and trying out criteria on a pure trial-and-error basis seems an impractically tedious approach. Proceeding in terms of an exemplar-similarity model appears more promising. If even a small number of representations of observed exemplar patterns can be held in working memory, it would be possible to form a tentative classification of these on the basis of similarities, then classify additional exemplars observed on the basis of similarity to these tentative categories. It would be a straightforward task to determine empirically whether various classification problems would be learnable by this model, possibly with iteration of the procedure, and then investigate whether learnability by the model would predict learnability by human subjects.

In general, it seems reasonable to conclude that the processes embodied in the feature-frequency, exemplar-frequency, and distance-based models are all available to human learners. It seems plausible that learners would have a basic priority, or preference order, but would shift from one mechanism to another in response to task demands. However, no formal model is yet available to specify when shifts should occur and how they are accomplished.

Model-based analyses have shown that category learning includes aspects of skill learning as well as knowledge acquisition. In the formulation of their exemplar model, Medin and Schaffer (1978) allowed for the possibility that learners would tend to weight different stimulus attributes in relation to their diagnosticities or saliences, and Nosofsky (1984a, 1984b, 1986) has added evidence that learners tend to shift dimensional weights toward the optimal values for a task during learning. Analyses discussed in this article and in Estes (1986) indicate that there is also a learning process with respect to the parameter s, representing effective similarity, which appears to have much the same character as learning that produces criterion adjustments in signal detection and recognition. The role of attentional learning can evidently be made negligible by eliminating variation in diagnosticity among features or dimensions in a task, but it may always be necessary for a learner to discover the appropriate value of s for a given situation. It would be expected, of course, that both aspects of skill learning would reach stable asymptotic levels with sufficient experience in a task.

The role of prior knowledge. In the present development, as in most of the related literature, it is assumed that the memory array is filled by the trial-by-trial output of feature analyses of presented category exemplars. However, it is possible that knowledge from other sources can be encoded into the array format. Though it is convenient to start with analyses based on experimenter-defined features, the problem of discovering what features are actually used by learners in natural situations can be approached by already available methods such as multidimensional scaling (Nosofsky, 1984b). I have no disagreement with the many investigators, for example, Murphy and Medin (1985) and Smith and Medin (1981), who stress the importance of the learner's prior knowledge in naturalistic categorization learning. However, it is one thing to recognize the importance of knowledge acquired outside of the experimental situation and another to understand how the knowledge is represented in the data structures of the cognitive system and brought to bear on particular category judgments. The only serious attack on this problem, that of Anderson (1976, 1983) in his ACT system, is based on the idea that knowledge is stored in memory in a network of propositions. That approach appears to offer promise in application to category learning (Elio & Anderson, 1981) but it remains to be shown how propositional knowledge feeds into mental mechanisms that compute probability estimates and similar quantities that are needed to learn fuzzy categorizations in noisy environments. Knowledge has to be coded in a form with the essential properties of the memory array before such computations can be accomplished, regardless of the specific mechanisms.

It may be noted, further, that regardless of the importance of general world knowledge in the formation of concepts, the implications of such knowledge need to be checked out for any specific situation before one can be confident that it leads to appropriate categorizations as defined by the task at hand. In terms of the array model, background knowledge might suggest the relevant attributes on which category exemplars should be encoded and the relative weights to be placed on different attributes or patternings of them, but there would remain the task of acquiring skillful categorization performance in the given situation. Interpreting this acquisition of performance capability may be the special contribution of models of the array family to the overall task of comprehending natural category learning.

APPENDIX 1

Categorization Probabilities in Standard and Weighted-Vector Exemplar Models for the Design of Table 2

Derivations similar to the one given in the text for categorization of pattern 10 (Eq. (1)) yield for the probabilities of assigning the other patterns in Table 2 to category A

$$P_{01}(A) = \frac{(1 - \theta + \theta s)^2}{(1 - \theta + \theta s)^2 + [\theta + (1 - \theta)s]^2}, \tag{A1}$$

and

$$P_{11}(A) = P_{00}(A) = ½. \tag{A2}$$

To obtain the probability of a correct categorization on a category A trial, we weight the values of $P_x(A)$ from Eqs. (1), (A1), and (A2) by the relative frequencies of the patterns, from column 1 of Table 2, yielding

$$P_A(A) = \theta(1 - \theta) + \frac{\theta^2[\theta + (1 - \theta)s]^2 + (1 - \theta)^2(1 - \theta + \theta s)^2}{[\theta + (1 - \theta s)]^2 + (1 - \theta + \theta s)^2}. \qquad \text{(A3)}$$

The corresponding expressions for category B are obtained similarly.

When the storage parameter p has a value less than unity, individual features of perceived exemplars are stored probabilistically, and consequently the memory array will contain fragments of exemplar patterns. Table A1 shows all of the memory vectors that might be stored in the two-feature example of Table 2 in canonical form, together with their asymptotic relative frequencies. Computation of similarities between test exemplars and vectors in the array proceeds as in the simpler case of Table 2 except when a feature of the test exemplar is compared with a null feature (denoted by a dash [-] in the array) of a remembered exemplar. Owing to noise in the system, it is possible for a perceived feature to have some degree of similarity greater than zero to a null feature; a parameter t will represent this similarity just as s does for comparison of perceived with actually stored features.

With this parameter added to the model, the similarities of test exemplars to the memory arrays are computed in the standard manner. For test exemplar 10, the similarities to the asymptotic A and B arrays are

$$\begin{aligned}\text{sim}(10,\text{A}) &= p^2[\theta^2 + (1 - \theta)^2s^2 + 2\theta(1 - \theta)s] \\ &+ 2p(1 - p)[\theta t + (1 - \theta)st] + (1 - p)^2t^2\end{aligned}$$

and

$$\begin{aligned}\text{sim}(10,\text{B}) &= p^2[(1 - \theta)^2 + \theta^2s^2 + 2\theta(1 - \theta)s] \\ &+ 2p(1 - p)[(1 - \theta)t + \theta st] + (1 - p)^2t^2.\end{aligned}$$

These expressions are not very edifying to look at, but they will be useful for comparisons with other models. Learning curves are computed trial-by-trial, just as in the simpler case of $p = 1$, with the exemplar presented on each trial being compared to the vectors of an array similar to Table A1 except that the cell entries are current actual relative frequencies of the patterns rather than asymptotic probabilities. The effect of reducing the value of the storage parameter p is to both slow the curve of learning and lower the asymptote, as illustrated in Fig. 2.

Entering the similarities in the expression for categorization probability of the exemplar model,

TABLE A1
Memory Array Resulting from Independent Storage of Features

	Category	
Pattern	A	B
1 0	θ^2p^2	$(1\text{-}\theta)^2p^2$
0 1	$(1\text{-}\theta)^2p^2$	θ^2p^2
1 1	$\theta(1\text{-}\theta)p^2$	$(1\text{-}\theta)\theta p^2$
0 0	$(1\text{-}\theta)\theta p^2$	$\theta(1\text{-}\theta)p^2$
1 -	$\theta p(1\text{-}p)$	$(1\text{-}\theta)p(1\text{-}p)$
- 1	$(1\text{-}\theta)p(1\text{-}p)$	$\theta p(1\text{-}p)$
- 0	$\theta p(1\text{-}p)$	$(1\text{-}\theta)p(1\text{-}p)$
0 -	$(1\text{-}\theta)p(1\text{-}p)$	$\theta p(1\text{-}p)$
- -	$(1\text{-}p)^2$	$(1\text{-}p)^2$

$$P_{10}(\mathrm{A}) = \frac{\mathrm{sim}(10,\mathrm{A})}{\mathrm{sim}(10,\mathrm{A}) + \mathrm{sim}(10,\mathrm{B})},$$

and inserting selected values of the parameters yields the pattern of categorization probabilities as a function of p, s, and t shown in Table A2. When the parameter t, reflecting similarity of perceived features to null (missing) features in the memory representation is equal to 0, so that fragmentary stored exemplars have no effect, the asymptotic categorization probabilities are independent of the storage probability. But if t is even slightly greater than 0 (see the row for $t = .01$), lowering the storage parameter produces a reduction in categorization probability. To the extent that the comparison process is under the control of the learner, it would be good strategy to distinguish clearly between complete and fragmentary memory representations and to set a criterion such that matches between perceived features and elements of a memory representation are not accepted on insufficient evidence.

APPENDIX 2

The Conditions for Equivalence of Exemplar and Feature-Frequency Models

Derivation of the condition for equivalence is presented in terms of the general two-feature design of Table 1. Considering first the exemplar model, and taking pattern 10 for illustrative purposes, the similarity of this pattern to the category A array, in terms of the relative frequencies of stored exemplars given in the first column of Table 1, is

$$\mathrm{sim}(10,\mathrm{A}) = u + (1 - u - v)s + vs^2$$

and the similarity to the B array

$$\mathrm{sim}(10,\mathrm{B}) = x + (1 - x - y)s + ys^2,$$

yielding for the probability of categorizing exemplar 10 as an A

$$P_{10}(\mathrm{A}) = \frac{u + (1 - u - v)s + v\,s^2}{u + x + (2 - u - v - x - y)s + (v + y)s^2}. \tag{A4}$$

For the feature model, the representation of the relative frequency of values of 1 for Feature 1 in category A is

$$f_{\mathrm{A}1}^{(1)} = (u + w) + (1 - u - w)s$$

and category B

TABLE A2
Asymptotic Values of $P_{10}(\mathrm{A})$ in the Exemplar Model for Selected Parameter Values

	$s = .4$		$s = .1$	
t	$p = .2$	$p = .05$	$p = .2$	$p = .05$
1.0	.54	.51	.56	.51
0.5	.57	.52	.60	.53
0.1	.66	.57	.76	.61
0.01	.73	.69	.88	.82
0.0	.74	.74	.90	.90

$$f^{(1)}_{B1} = (x + z) + (1 - x - z)s.$$

Similarly, the representations for values of 0 for Feature 2 are

$$f^{(0)}_{A2} = (1 - v - w) + (v + w)s$$

and

$$f^{(0)}_{B2} = (1 - y - z) + (y + z)s.$$

The probability of category A given exemplar 10 for this model is, then,

$$P^{(A)}_{10} = \frac{f^{(1)}_{A1} f^{(0)}_{A2}}{f^{(1)}_{A1} f^{(0)}_{A2} + f^{(1)}_{B1} f^{(0)}_{B2}}, \tag{A5}$$

with the appropriate substitutions from the expressions just derived.

Now let us consider the numerator of Eq. (A5)

$$f^{(1)}_{A1} f^{(0)}_{A2} = [u + w + (1 - u - w)s]\,[(1 - v - w) + (v + w)s] \tag{A6}$$

and ask under what conditions it would be identical to the numerator of Eq. (A4). For equivalence to obtain, the coefficients of each power of s would have to be equal in the two expressions. If we multiply out the factors on the right of Eq. (A6), the product of terms not involving s is $(u + w)(1 - v - w)$, which, for the models to be equivalent, must be equal to u, the term not involving s in the numerator of Eq. (A4). Referring to Table 1, we note that $u + w$ is the probability of a 1 for Feature 1 in category A and $1 - v - w$ the probability of a 0 for Feature 2 in category A. If, and only if, the features are independent, the product of these is the probability of pattern 10 in category A, which is simply u. Proceeding similarly with the coefficients of s and s^2 in Eqs. (A4) and (A6), we find that in each case equality obtains only if independence holds. Derivations for the other patterns are analogous, so we conclude that the models are equivalent if and only if the features are independent. (However, the CBD exemplar model is equivalent to the feature model regardless of independence.)

APPENDIX 3

Equivalence of Exemplar and Feature Models in Independent-Feature Situations with Probabilistic Feature Storage

When Table 2 is expanded to represent the canonical memory array resulting from probabilistic feature storage, as in Table A1, computation of the learner's estimates of feature frequencies proceeds just as shown in the text for the case of Table 2, except that similarities of perceived to null features are taken into account via the parameter t. Referring to Table A1, the representation of relative frequency of the Value 1 for Feature 1 in category A is

$$\begin{aligned} f^{(1)}_{A1} &= c^2[\theta + (1 - \theta)s] + c(1 - c)[\theta + (1 - \theta)s] + (1 - c)t, \\ &= c[\theta + (1 - \theta)s] + (1 - c)t, \end{aligned}$$

and the representation for Value 0 for Feature 2 in category A is the same. Therefore, the learner's probability estimate for pattern 10 in category A is

$$\{c[\theta + (1 - \theta)s] + (1 - c)t\}^2.$$

On expansion, this expression is seen to be identical to that for the similarity of pattern 10 to category A in the exemplar model (Appendix 1); and a similar computation shows that the probability estimate for pattern 10 in category B for the feature model is identical to the similarity of 10 to B in the exemplar model. Therefore the probability of category A given exemplar 10 is the same in both models. Similar derivations yield analogous equivalences for the other patterns, and thus, given feature independence, the models are equivalent regardless of the feature storage probabilities or the similarity parameter values.

APPENDIX 4

Distance-Based Exemplar and Average-Prototype Models for the General Two-Feature Design of Table 1

First, CBD versions of both models are compared, then the city-block restriction is removed. Starting with the exemplar model, and using the test pattern 10 for illustrative purposes, we find by reference to Table 1 (and assuming equally probable categories) that the total distance of pattern 10 from the four representations in category A is

$$D_{10\mathrm{A}} = 0 \cdot u + 2 \cdot v + 1 \cdot [w + (1 - u - v - w)].$$

The distance from test exemplar 10 to the memory representation 10 is equal to 0; the distance to 01 is 2 (since both features mismatch) and this distance is weighted by relative frequency v; the distances to 11 and 00 are both 1 (since there is one mismatch in each case). The distances are weighted by the sum of the relative frequencies of these stored exemplars from the A column of Table 1. Simplifying, we have

$$D_{10\mathrm{A}} = 1 - u + v,$$

and, similarly,

$$D_{10\mathrm{B}} = 1 - \mathrm{x} + \mathrm{y}.$$

These quantities can now be entered in the general formula for categorization probability in distance-based models (Eq. (7)) to obtain the probability of categorizing exemplar 10 as an A:

$$P_{10}(\mathrm{A}) = \frac{1}{1 + e^{-c(u+y-v-x)}}, \qquad \text{(A7)}$$

where c is a scaling constant to be determined from the data.

In the average prototype model, the prototype for category A is $(u + w)$, $(v + w)$. The first component is the average value of the first feature of category A

$$(u + w) \cdot 1 + (1 - u - w) \cdot 0,$$

and the second component is the average value of the second feature in category A

$$(v + w) \cdot 1 + (1 - v - w) \cdot 0.$$

The distance of exemplar 10 from the prototype is the sum of the distance from 1 to the first component, $1 - (u + w)$, and the distance from 0 to the second component, $(v + w)$, and combining these yields

$$D_{10\mathrm{A}} = 1 - u + v.$$

Similarly, we obtain

$$D_{10\mathrm{B}} = 1 - x + y,$$

and entering these quantities in Eq. (7) yields the same expression for $P_{10}(\mathrm{A})$ that was derived for the exemplar model. Analogous results hold for the other exemplar patterns, so we conclude that the two models are indistinguishable on the basis of categorization probabilities.

To lift the restriction to a city-block metric, we define $d^{(1)}$ as the distance between a perceived exemplar and a stored pattern that mismatches in exactly one feature and $d^{(2)}$ as the distance when there are mismatches in both features. In applications of the model, these unit distances will be treated as parameters to be determined from the data. With these distances replacing the city-block distances of the CBD exemplar model, the expressions for summed distance from test pattern 10 to the two categories become

$$D_{10\mathrm{A}} = 0 \cdot u + d^{(1)} \cdot (1 - u - v) + d^{(2)} \cdot v$$

and

$$D_{10\mathrm{B}} = 0 \cdot y + d^{(1)} \cdot (1 - x - y) + d^{(2)} \cdot y,$$

and the probability of categorizing 10 as an A is

$$P_{10}(\mathrm{A}) = \frac{1}{1 + e^{-c[(y-v)d^{(2)} - (x+y-u-v)d^{(1)}]}} \tag{A8}$$

Examining the exponent in the denominator, we see that $P_{10}(\mathrm{A})$ is equal to 1/2 if the relative frequencies of the 10 and 01 patterns are equal in the two categories, that is, $u = x$ and $v = y$; however, unlike the CBD special case, $P_{10}(\mathrm{A})$ does not reduce to 1/2 when the individual features are invalid ($u + w = x + z$ and $v + w = y + z$). Thus the model can predict learning whenever some of the patterns, as distinguished from individual features, are at least partially valid.

It is not obvious how to achieve a comparable generalization of the prototype model, since we would have to define four, rather than two, basic distances, that is, the distances between each of the feature values of the test exemplar and each of the components of the category A and B prototypes. However, we can readily compare the two models for the case of a Euclidean distance metric. In the exemplar model, the summed distances from test pattern 10 to the categories are

$$D_{10\mathrm{A}} = 0 \cdot u + (\sqrt{2}) \cdot v + 1 \cdot (1 - u - v)$$

and

$$D_{10\mathrm{B}} = 0 \cdot x + (\sqrt{2}) \cdot y + 1 \cdot (1 - x - y)$$

yielding

$$P_{10}(\mathrm{A}) = \frac{1}{1 + e^{-c[(\sqrt{2})(y-v) + (u+v-x-y)]}},$$

whereas for the prototype model the distances are

$$D_{10\mathrm{A}} = \sqrt{(1 - u - w)^2 + (v + w)^2}$$

and

$$D_{10\mathrm{B}} = \sqrt{(1 - x - z)^2 + (y + z)^2},$$

yielding

$$P_{10}(\mathrm{A}) = \frac{1}{e^{-c[\sqrt{(1-x-y)^2+(y+z)^2} - \sqrt{(1-u-w)^2+(v+w)^2}]}}.$$

Clearly, the expressions for $P_{10}(\mathrm{A})$ are in general not equal for the two models, and the same would be true for the other exemplar patterns. Numerical computations are needed to determine whether there are task designs (that is, sets of values for the parameters in Table 1) for which predictions from the two models differ enough to be empirically distinguishable in practice.

APPENDIX 5

Derivation of Expected Memory Array for the Exemplar Model Given that One Exemplar Has Not Occurred

Starting with the pattern probabilities in Table 1, and assuming equally probable categories, we note that the probability that pattern 10 does not occur on any trial is $\frac{1}{2}(1 - u) + \frac{1}{2}(1 - x) = (2 - u - x)/(2)$; and since the probability of an A trial on which pattern 10 does not occur is $\frac{1}{2}(1 - u)$, the probability of category A given that 10 does not occur is $(1 - u)/(2 - u - x)$. Therefore, over the first $n - 1$ trials of an experiment, the expected number of category A trials given that pattern 10 has not occurred is $([n - 1][1 - u])/(2 - u - x)$. This quantity represents also the number of stored exemplars in the A array, and they will be divided among the last three rows of the array in proportion to the pattern probabilities v, w, and $1 - u - v - w$, yielding the entries in the A column of Table 3. The entries in the B column are obtained similarly.

APPENDIX 6

Computation of Categorization Probabilities for Array Models

This summary includes all of the models of the array family described in the article as applied to the task design of Table 1. Extensions to designs with exemplars defined by more than two features are straightforward. For brevity, the summary is limited to the case in which all perceived features are stored ($p = 1$). This case may be adequate for applications of the models when the number of features per exemplar is small, and reference to Table A1 and the associated discussion in Appendix 1 will indicate how to handle cases involving uncertain storage ($p < 1$).

Computation of categorization probabilities in all of the models is based on an array of the form of Table 1, but with observed frequencies as cell entries,

	A	B
10	T_1	T_2
01	T_3	T_4
11	T_5	T_6
00	T_7	T_8

where the T_i denote cumulative total occurrences of the row pattern in the column category up to the trial for which categorization probability is being computed. A default value of 1/2 is used for categorization probability on the first trial, when all of the T_i are equal to 0 (and in the general model with $p < 1$, on all trials until at least one of the T_i is greater than 0). For many purposes it is convenient to transform this array to one with entries t_i, obtained by dividing T_i by its current column total (i.e., the current total number of occurrences of the category). The quantities t_i serve as estimators of the conditional pattern probabilities whose asymptotic values are the cell entries in Table 1. For actual computations of categorization probabilities, a value of s must, of course, be specified. If no a priori value is available, the standard procedure is to compute probabilities for different s values and select the one that yields the best agreement between the model and data by a least-squares or other conventional criterion.

Exemplar models. For the standard exemplar model, the similarities of the current test exemplar to the A and B arrays are computed by expressions of the form

$$\mathrm{Sim}(10,\mathrm{A}) = T_1 + (T_5 + T_7)s + T_3 s^2$$

and

$$\mathrm{Sim}(10,\mathrm{B}) = T_2 + (T_6 + T_8)s + T_4 s^2,$$

and the results are entered in the general formula

$$P_{\text{ex}}(\text{A}) = \frac{\text{Sim(ex,A)}}{\text{Sim(ex,A)} + \text{Sim(ex,B)}}, \tag{A9}$$

where ex denotes the test exemplar pattern. The probability of a correct categorization on a category A trial is then obtained from

$$P_{\text{A}}(\text{A}) = u\,P_{10}(\text{A}) + v\,P_{01}(\text{A}) + w\,P_{11}(\text{A}) + (1 - u - v - w)\,P_{00}(\text{A}), \tag{A10}$$

and the probability for a category B trial is obtained similarly. Probability of a correct categorization on any trial, given no information as to which category will be represented, is, then, given by

$$P(\text{C}) = \pi_1\,P_{\text{A}}(\text{A}) + \pi_2\,P_{\text{B}}(\text{B}). \tag{A11}$$

The only change in procedure required for the pure-similarity form of the exemplar model is to replace the T_i in the expressions for Sim(ex,A) and Sim(ex,B) with the corresponding t_i.

For the general distance-based exemplar model, expressions of the type Sim(ex,A) are replaced by expressions for distances between the test exemplar and the memory arrays, for example,

$$D_{10\text{A}} = (T_5 + T_7)d^{(1)} + T_3d^{(2)}$$

and

$$D_{10\text{B}} = (T_6 + T_8)d^{(1)} + T_4d^{(2)},$$

and categorization probability is computed from expressions of the form

$$P_{10}(\text{A}) = \frac{1}{1 + e^{-c[(T_6+T_8-T_5-T_7)d^{(1)}+(T_4-T_3)d^{(2)}]}}. \tag{A12}$$

In this form, the distance-based model reflects category base rates much as does the standard model, but in the opposite direction. Also, as the number of trials, and therefore also the values of the T_i, becomes large, the exponent in the denominator of Eq. (A12) also becomes large, and consequently the categorization probability approaches unity if $D_{10\text{B}}$ is greater than $D_{10\text{A}}$ and zero otherwise, regardless of the value of c. To produce a version closer to the standard model, we can convert this "absolute distance" model to a "relative distance" model by replacing the T_i by corresponding t_i and weighting the exponential terms by the category probabilities, yielding categorization probabilities of the form

$$P_{10}(\text{A}) = \frac{1}{1 + (\pi_2/\pi_1)e^{-c[\{(t_6+t_8)d^{(1)}+t_4d^{(2)}\}-\{(t_5+t_7)d^{(1)}+t_3d^{(2)}\}]}}. \tag{A13}$$

(Alternatively, the π_i can be replaced by the current actual relative frequencies of the categories in a trial-by-trial computation.) Now, the categorization probability approaches unity or zero only for large values of c, and in any given case a value of c can be chosen that yields categorization probabilities close to those of the standard model. The model can be converted to a form comparable to the pure-similarity exemplar model simply by omitting the ratio π_2/π_1 from Eq. (A13) and the corresponding expressions for the other patterns. City-block-distance and Euclidean distance versions of these models are, of course, obtainable simply by replacing the unit distances $d^{(1)}$ and $d^{(2)}$ by 1 and 2, respectively, for the city-block and by 1 and $\sqrt{2}$ for the Euclidean version.

The feature-frequency model. For the feature-frequency model, trial-by-trial computations are based on a table of the form

	A		B	
	f_1	f_2	f_1	f_2
1	x_1	x_2	x_3	x_4
0	x_5	x_6	x_7	x_8

with the x_i denoting the cumulative frequency in each category of values of 1 and 0 for features 1 and 2. To compute categorization probability on any trial, these frequencies are used to generate the values of the learner's feature-probability estimates, f_{x_i} of Eqs. (11) and (12). Those quantities start with default values of 1/2, then are updated on each trial in terms of the current values of the x_i:

$$f_{A1} = \frac{x_1}{x_1 + x_5}$$

$$f_{B1} = \frac{x_3}{x_3 + x_7},$$

and so on. The new estimates are then entered in the expressions for likelihoods. If the test pattern is 10, for example, its likelihoods on category A and category B trials, respectively, are

$$L(\mathrm{A},10) = \pi_{\mathrm{A}} f_{\mathrm{A}1} (1 - f_{\mathrm{A}2})$$

and

$$L(\mathrm{B},10) = \pi_{\mathrm{B}} f_{\mathrm{B}1} (1 - f_{\mathrm{B}2}),$$

and the probability of categorizing it as an A is

$$P_{10}(\mathrm{A}) = \frac{L(\mathrm{A},10)}{L(\mathrm{A},10) + L(\mathrm{B},10)}.$$

For a priori predictions of asymptotic categorization probabilities, the asymptotic values of the $f_{\mathrm{A}i}$ and $f_{\mathrm{B}i}$ can be obtained from Table 1 as described in Appendix 1.

Prototype models. The average prototype of a category is a vector whose components are the average values of each feature in exemplars of the category, in the two-feature case,

$$\mathrm{Prot(A)} = (m_1, m_2)$$

and

$$\mathrm{Prot(B)} = (m_3, m_4),$$

where

$$m_1 = t_1 + t_5$$
$$m_2 = t_3 + t_5$$
$$m_3 = t_2 + t_6$$
$$m_4 = t_4 + t_6$$

in the notation defined at the beginning of this appendix. For the independent-feature case illustrated in Fig. 1, $m_1 = m_4 = \theta$ and $m_2 = m_3 = (1 - \theta)$.

For computation of trial-by-trial categorization probabilities, the values of m_i are updated on each trial, and for a priori predictions of asymptotic probabilities, the asymptotic values of the t_i, and hence the m_i, are obtainable from Table 1. Computation of distances between exemplars and prototypes depends on the metric assumed. For the city-block metric, distances are defined as in the city-block exemplar model, and, for example,

$$D_{10\mathrm{A}} = 1 - m_1 + m_2$$
$$D_{10\mathrm{B}} = 1 - m_3 + m_4,$$

and

$$P_{10}(\mathrm{A}) = \frac{1}{1 + e^{-c(m_1 + m_2 - m_3 - m_4)}} . \qquad \text{(A14)}$$

For the Euclidean metric,

$$D_{10\mathrm{A}} = \sqrt{(1 - m_1)^2 + m_2^2}$$
$$D_{10\mathrm{B}} = \sqrt{(1 - m_3)^2 + m_4^2}$$

and

$$P_{10}(\mathrm{A}) = \frac{1}{1 + e^{-c(\sqrt{(1 - m_3)^2 + m_4^2} + \sqrt{(1 - m_1)^2 + m_2^2})}} .$$

REFERENCES

Anderson, J. R. (1976). *Language, memory, and thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard Univ. Press.

Bower, G. H. (1967). A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 1, pp. 229–325). New York: Academic Press.

Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 85–123). Washington, DC: Winston.

Brooks, L. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization* (pp. 169–211). Hillsdale, NJ: Erlbaum.

Busemeyer, J. R., Dewey, G. I., & Medin, D. L. (1984). Evaluation of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **10**, 638–648.

Elio, R., & Anderson, J. R. (1981). The effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Human Learning and Memory,* **7**, 397–417.

Estes, W. K. (1959). Component and pattern models with Markovian interpretations. In R. R. Bush & W. K. Estes (Eds.), *Studies in mathematical learning theory* (pp. 9–52). Stanford, CA: Stanford Univ. Press.

Estes, W. K. (1972). Elements and patterns in diagnostic discrimination learning. *Transactions of the New York Academy of Sciences, Series II,* **34**, 84–95.

Estes, W. K. (1976). The cognitive side of probability learning. *Psychological Review,* **83**, 37–64.

Estes, W. K. (1986). Memory storage and retrieval processes in category learning. *Journal of Experimental Psychology: General,* **115**, 155–174.

Estes, W. K., Burke, C. J., Atkinson, R. C., & Frankmann, J. P. (1957). Probabilistic discrimination learning. *Journal of Experimental Psychology,* **54**, 233–239.

Franks, J. J., & Bransford, J. D. (1971). Abstraction of visual patterns. *Journal of Experimental Psychology,* **90**, 65–74.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **10,** 234–257.

Garner, W. R. (1974). *The processing of information and structure.* Potomac, MD: Erlbaum.

Gibson, E. J. (1969). *Principles of perceptual learning and development.* New York: Appleton-Century-Crofts.

Gordon, M. J. (1985). *Learning artificial analogs of natural categories.* Ph.D. dissertation, Harvard University.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General,* **108,** 356–388.

Healy, A. F., & Kubovy, M. (1977). A comparison of recognition memory to numerical decision: How prior probabilities affect cutoff location. *Memory & Cognition,* **5,** 3–9.

Hintzman, D. L. (1976). Repetition and memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 10, pp. 47–92). New York: Academic Press.

Hintzman, D. L., & Ludlam, K. (1980). Differential forgetting of prototypes and old instances: Simulation by an exemplar-based classification model. *Memory & Cognition,* **8,** 378–382.

Marr, D. (1976). Early processing of visual information. *Philosophical Transactions of the Royal Society of London. B,* **275,** 483–524.

Medin, D. L. (1983). Structural principles of categorization. In T. J. Tighe & B. E. Shepp (Eds.), *Perception, cognition, and development: Interactional analyses* (pp. 203–230). Hillsdale, NJ: Erlbaum.

Medin, D. L., Altom, M. W., & Murphy, T. D. (1984). Given versus induced category representations: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **10,** 333–352.

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

Medin, D. L., & Smith, E. E. (1981). Strategies and classification learning. *Journal of Experimental Psychology: Human Learning and Memory,* **7,** 241–253.

Millward, R. B. (1980). Models of concept formation. In R. E. Snow, P.-A. Federico, & W. E. Montagne (Eds.), *Aptitude, learning, and instruction: Vol. 1. Cognitive process analyses of aptitude* (pp. 245–275). Hillsdale, NJ: Erlbaum.

Murdock, B. B., Jr. (1974). *Human memory: Theory and data.* Potomac, MD: Erlbaum.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review,* **92,** 289–316.

Norman, D. A., & Rumelhart, D. E. (1970). A system for perception and memory. In D. A. Norman (Ed.), *Models of human memory* (pp. 19–64). New York: Academic Press.

Nosofsky, R. M. (1984a). *Attention, similarity, and the identification–classification relationship.* Ph.D. dissertation, Harvard University.

Nosofsky, R. M. (1984b). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **10,** 104–114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General,* **115,** 39–57.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology,* **77,** 353–363.

Reed, S. K. (1972). Pattern recognition and categorization. *Cognitive Psychology,* **3,** 382–407.

Reed, S. K. (1973). *Psychological processes in pattern recognition.* New York: Academic Press.

Reitman, J. S., & Bower, G. H. (1973). Storage and later recognition of exemplars of concepts. *Cognitive Psychology,* **4,** 194–206.

Shepard, R. N. (1958). Stimulus and response generalization: Deduction of the generalization gradient from a trace model. *Psychological Review,* **65,** 242–256.

Smith, E. E., & Medin, D. L. (1981). *Categories and concepts.* Cambridge, MA: Harvard Univ. Press.

Swets, J. A. (1964). *Signal detection and recognition by human observers.* New York: Wiley.

Townsend, J. T. (1974). Issues and models concerning the processing of a finite number of inputs. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 133–185). Hillsdale, NJ: Erlbaum.

Trabasso, T., & Bower, G. H. (1968). *Attention in learning: Theory and research.* New York: Wiley.

Tzeng, O. J. L., & Cotton, B. (1980). A study-phase retrieval model of temporal coding. *Cognitive Psychology,* **5,** 207–232.

Underwood, B. J. (1969). Attributes of memory. *Psychological Review,* **76,** 559–573.

Underwood, B. J., & Freund, J. S. (1970). Relative frequency judgments and verbal discrimination learning. *Journal of Experimental Psychology,* **83,** 279–285.

Wells, J. E. (1974). Strength theory and judgments of recency and frequency. *Journal of Verbal Learning and Verbal Behavior,* **13,** 378–392.

Whitlow, J. W., Jr., & Estes, W. K. (1979). Judgments of relative frequency in relation to shifts of event frequencies: Evidence for a limited-capacity model. *Journal of Experimental Psychology: Human Learning and Memory,* **5,** 395–408.

(Accepted March 25, 1986)