COGNITIVE SCIENCE 16, 81–121 (1992)

A Two-Stage Model of Category Construction

WOO-KYOUNG AHN AND DOUGLAS L. MEDIN
University of Michigan

The current consensus is that most natural categories are not organized around strict definitions (a list of singly necessary and jointly sufficient features) but rather according to a family resemblance (FR) principle: Objects belong to the same category because they are similar to each other and dissimilar to objects in contrast categories. A number of computational models of category construction have been developed to provide an account of how and why people create FR categories (Anderson, 1990; Fisher, 1987). Surprisingly, however, only a few experiments on category construction or free sorting have been run and they suggest that people do not sort examples by the FR principle. We report several new experiments and a two-stage model for category construction. This model is contrasted with a variety of other models with respect to their ability to account for when FR sorting will and will not occur. The experiments serve to identify one basis for FR sorting and to support the two-stage model. The distinctive property of the two-stage model is that it assumes that people impose more structure than the examples support in the first stage and that the second stage adjusts for this difference between preferred and perceived structure. We speculate that people do not simply assimilate probabilistic structures but rather organize them in terms of discrete structures plus noise.

People both learn about preexisting categories and create new categories on their own. For example, one may construct categories such as, "my favorite movies," or "different types of college students I have known." Certainly, formal education provides or imposes other categories. Furthermore, there may be intermediate cases. For example, children may learn many natural object categories by being taught, but Nelson (1974) raised the alternative

We would like to thank Judy Florian, Mark Gluck, Evan Heit, Keith Holyoak, Steven Sloman, Edward E. Smith, Edward Wisniewski, and two anonymous reviewers for their helpful comments on earlier drafts. In addition, we thank Brad Whitehall for running CLUSTER/2 on the stimuli used in the current experiments and Kate McKusick and Young-pa So for helping us run COBWEB/3.

This research is supported by Grant NSF-BNS-8918701 to Douglas L. Medin. Parts of this article were drawn from a doctoral dissertation submitted to University of Illinois (Ahn, 1990b). Parts of this work were presented in the 11th Annual Conference of the Cognitive Science Society held in Ann Arbor (Ahn & Medin, 1989).

Correspondence and requests for reprints should be sent to Woo-kyoung Ahn, Department of Psychology, Cognition and Perception Programs, University of Michigan, 330 Packard Road, Ann Arbor, MI 48104-2994.

possibility that children construct their own categories first and then learn which labels apply to them rather than using the labels to form the categories.

Category construction needs to be highly constrained because the number of ways of partitioning unclassified objects grows exponentially with a linear increase in the number of objects. For example, one can partition 3 objects in 5 ways, 4 objects in 15 ways, 5 objects in 52 ways, and 10 objects in more than 100,000 ways. Considering the number of objects in the world, it is clear that our categories constitute only the tiniest subset of all possible partitionings: Therefore, the question of what principles determine how categories are constructed is an important one that ought to give us some insight into how the mind works.

Presumably, human categorization reflects the interaction of human goals and conceptual capabilities with the information and structure available in the environment. Indeed, the historical shift from the idea that categories follow strict definitions (the so-called classical view, Bruner, Goodnow, & Austin, 1956; Katz & Postal, 1964) to the idea that concepts are structured around prototypes that are only generally true of category examples (the probabilistic view) was motivated by a detailed analysis of natural object categories (Rosch, 1975; Rosch & Mervis, 1975; Smith, Shoben, & Rips, 1974; Smith & Medin, 1981). Associated with this analysis has been the idea that mental representations of categories closely mirror the structure afforded by properties of category examples. For example, the ideas that natural object categories are organized around prototypes and that membership judgments are based on the similarity of examples to prototypes are nicely compatible with fuzzy or probabilistic category structures.

Although much attention has focused on classification processes associated with preexisting categories, there has been a recent interest in principles of category construction. Different theories of category construction have tended to agree on two central assumptions. One assumption is that the computations associated with category construction (and category representations, where this description is appropriate) mirror the structure afforded by examples. That is, one goal has been to account for, or to be able to reproduce, fuzzy or family resemblance (FR) categories. An allied assumption has been that principles of category construction at the level of cultures, as reflected in natural object categories, also apply at the level of individuals. Specifically, it has been assumed that given the opportunity, individuals will create FR categories (Rosch, 1975).

Very few category-construction or free-sorting experiments have been carried out to see whether and when people create FR categories. The evidence that exists suggests that people do not find it natural to sort by an FR principle (e.g., Medin, Wattenmaker, & Hampson, 1987). Before discussing this evidence, we need to describe category-construction theories in more detail because they help define just exactly what is meant by FR.

The organization of this article is as follows. We first review current category-construction theories. Next, we describe a two-stage model of FR sorting which implies that category representations do more than mirror the structure of examples. Finally, we evaluate how well these theories account for when FR sorting will and will not occur in experiments where people are asked to create categories.

CATEGORY-CONSTRUCTION THEORIES

Similarity-Based Models

Traditionally, similarity among objects has been considered the basis for category construction: Objects are thought to belong to the same class because they are similar to each other (Rosch, 1978). Although Rosch did not offer a specific similarity-based model for category construction, Rosch suggested that people would sort examples into categories so as to maximize within-category similarity and minimize between-category similarity (Rosch, 1975). Several similarity-based clustering models have been developed in statistics and pattern recognition. Basically, they represent similarity in terms of distance in some multidimensional space. The dimensions of the space correspond to the dimensions from which the exemplars are composed. The overall similarity between objects is an inverse function of their distance in the dimensional space. Although these models have not generally been proposed as psychological process models, they do represent ways of instantiating similarity-based clustering.

In general, there are two classes of similarity-based models: One creates hierarchical categories and the other creates nonhierarchical categories. Nonhierarchical categories are created by repeating the process of selecting prototypes, assigning exemplars based on the similarity to the prototypes, and recalculating the prototypes. Hierarchical categories are created either by treating all the exemplars as belonging to one category and splitting it until all the categories have only one member (the so-called divisive method) or by treating each example as a category itself and combining the two most similar categories until there is only one category (the so-called agglomerative method; see Anderberg, 1973; Massart & Kaufman, 1983, for reviews).

These models are statistical tools designed to place objects into clusters suggested by the data and have not yet been proposed as psychological models. Although the actual processing assumptions may not be psychologically valid, the basic philosophy behind these methods is the same as Rosch's: Categories are formed as a function of between-category and within-category similarity.

Predictibility-Driven Models

Another class of category-construction models focuses on maximizing the inference potential of categories. These models will be called predictibility-driven models. The idea is that the better we can predict unknown features based on category membership, the more advantageous it is to create such a category (Anderson, 1990). For example, it may be advantageous to create a category of "my friends" because knowing that a person is a friend, one can predict a variety of behaviors such as willingness to loan books or read drafts of papers. On the other hand, "red things" may not be a very useful category because knowing that an object is red does not tell us much else about the object.

In general, the predictibility-driven models will be sensitive to the number of features shared among objects, and therefore, they will tend to make the same predictions as the similarity-based models. The more specific nature of these models will be investigated in our experiments by running simulation programs on particular sets of exemplars. This section will present two predictibility-driven models existing in the field; Anderson's rational model of induction and Fisher's COBWEB.

Anderson's Model (1988, 1990, 1991). Anderson pointed out that one can specify an ideal algorithm that keeps track of all possible partitions to select the partition with the maximum predictability. But as he pointed out, that kind of model is implausible as a model of human behavior because of computational limitations as the number of objects and/or features grows large. This ideal algorithm could not even be run on a computer because there are too many possible partitionings. As an alternative, Anderson proposed an iterative algorithm in which new exemplars are considered incrementally and classified into the category that maximizes the predictability of the resulting partitioning. For each new example, two kinds of probabilities are calculated; the probabilities of the new object coming from old categories and the probability of creating a new category. The probability of the object coming from an existing category (P_K) is operationally defined as follows:

$$P_K = \frac{cn_K}{(1-c)+cn} \prod_{i=1} \frac{C_{Ki}+1}{n_K+m_i}$$

where n is the number of objects so far, n_K is the number of objects in category K so far, C_{Ki} is the number of objects in category K so far with the same value on the ith dimension as the object to be classified, m_i is the number of values on dimension i, and c is a free parameter representing the probability that any two objects will be in the same category, which is called the coupling probability.

The probability of the object coming from a new category (P_o) is

$$P_o = \frac{1-c}{(1-c)+cn} \prod_{i=1} \frac{1}{m_i}$$

The new object is assigned to existing categories if P_o is smaller than P_K or to the new category if P_o is greater than P_K.

Anderson showed that his model will construct FR categories and will tend to construct categories at the basic level (i.e., the most useful level in the generalization hierarchy of categories, Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Anderson's model is consistent with this phenomenon in that it creates basic-level categories with the widest range of the coupling parameter (c in the preceding equations) values.

COBWEB. Fisher (1987) developed a clustering system, COBWEB, using category utility, a measure proposed by Gluck and Corter (1985) as an index of the basic level. Category utility is a function of the base rate of features, category validity (i.e., probability of a feature given a category), and cue validity (i.e., probability of a category given a feature). COBWEB takes each example incrementally and compares the category utility of placing the example in each of the existing categories with the category utility of creating a new category. The example is placed in the category (either new or old) that results in the highest category utility.

Both Anderson's model and COBWEB are incremental clustering algorithms and produce different types of partitionings depending on the input order. COBWEB tries to reduce this order effect by splitting and merging the old partitionings every time a new example is entered. Due to computational complexity, however, calculating category utility over all the old examples is done only approximately, and therefore, the system may remain sensitive to input order.

Neither of these models was proposed as a process model of human category construction. The basic processes of the two models are the same except for the difference in the clustering criteria. Although the specifics of the clustering criteria differ, the main idea for developing the criteria is essentially the same: maximizing predictibility.

Maximizing Comprehensibility

Michalski and Stepp (1983) proposed a clustering system, CLUSTER/2, which tries to construct concepts that can be described in simple terms. Such classifications would facilitate comprehension of the observations and the subsequent use of them. Unlike the models described in the earlier section, CLUSTER/2 does not use a measure of overall similarity as a basis for categorization. Instead, the main goal of CLUSTER/2 is to generate hierarchical categories that can be described by a single conjunctive concept. Categories at the same level of the hierarchy should have logically disjoint descriptions

and optimize a set of clustering criteria, one of which is the simplicity of a concept. The simplicity of a concept is a function of the number of descriptors for the concept. Other clustering criteria include the ratio between the number of observed objects of a concept and the number of possible but unobserved objects covered by the concept, and the intercluster difference called the discrimination index (the number of variables having different values in every cluster description; see Michalski & Stepp, 1983).

Because CLUSTER/2 produces disjoint categories described with a single conjunctive concept, the system seems to prefer categories with defining features. CLUSTER/2, therefore, is unlikely to produce FR categories. The way it deals with examples that do not admit simple definitions will be shown as we describe specific experiments.

INCREMENTAL VERSUS SIMULTANEOUS SORTING

The models reviewed so far can be classified in terms of whether they use incremental or simultaneous clustering methods. The two predictibility-driven models presented earlier are incremental clustering models where examples are sorted one by one. But the similarity-based models and CLUSTER/2 use a procedure in which examples are to be sorted all at once (i.e., simultaneous sorting). Also, the current experiments and the empirical studies that will be described later use simultaneous sorting tasks. In this article, we compare results from the sequential sorting models with the results from simultaneous sorting tasks. Therefore, it seems important to clarify the differences between the two types of sorting tasks and justify our comparison.

On one hand, incremental sorting has advantages such as computational tractability and its ability to update the knowledge base as new examples are seen (Anderson, 1988, 1990; Fisher, 1987; Fisher & Langley, 1986; Langley, Kibler, & Granger, 1986). On the other hand, incremental systems can become too sensitive to skewed input order and produce biased partitionings. To handle this problem, Anderson and Matessa (1991) developed a hierarchical algorithm, which turned out to be somewhat more independent of presentation order. But (as they noted), no rational grounds (e.g., increased predictability) exist for using the hierarchical algorithm over the nonhierarchical one. COBWEB's solution to the order effect was to merge or split existing categories every time a new instance was entered. Therefore, COBWEB's partitionings may be comparable to simultaneous sorting.

In addition, even in simultaneous sorting situations, subjects may examine examples one by one. Therefore, simultaneous sorting can be considered a special case of incremental sorting where input order is randomized across subjects. In order to make sequential models as comparable to simultaneous models as possible with respect to our sorting task, sequential sorting models were tested with randomized presentation order over 50 trials in the current study. Using multiple presentation orders for the models may serve as a rough approximation to simultaneous sorting. This procedure will be fully described in the following.

TABLE 1
Abstract Notation of Stimuli Used in Medin, Wattenmaker, and Hampson (1987), Experiments 1, 2, and 3

	D1	D2	D3	D4		D1	D2	D3	D4
E1	1	1	1	1	E6	0	0	0	0
E2	1	1	1	0	E7	0	0	0	1
E3	1	1	0	1	E8	0	0	1	0
E4	1	0	1	1	E9	0	1	0	0
E5	0	1	1	1	E10	1	0	0	0

EMPIRICAL RESEARCH ON CATEGORY CONSTRUCTION

Only a few experiments have used sorting tasks (either sequential or simultaneous) to evaluate category-construction theories. The general finding in free-sorting experiments is that people tend to sort exemplars on the basis of values along one dimension (Imai & Garner, 1965, 1968; Medin, Wattenmaker, & Hampson, 1987).

For example, Medin, Wattenmaker, and Hampson (1987) found that subjects did not create categories according to the FR principle even when the examples were structured around prototypes. Subjects in their experiments received several exemplars, including two prototypes that differed from each other along all the component dimensions, and additional exemplars that were minimal distortions of a given prototype. The abstract notation for the 10 stimulus items presented in those experiments is shown in Table 1. (The first column indicates the exemplar label and the next four columns indicate the value on each dimension for the example.) The 10 exemplars are arranged in two columns, indicating FR partitioning for this set.

Rosch (1975) suggested that when asked to sort examples linked by overall similarity (as in these stimuli), people would tend to organize categories around the prototypes and produce FR categories. However, almost all of the subjects in Medin, Wattenmaker, and Hampson's experiment created categories on the basis of a single dimension. For example, if the first dimension was used by a subject, E1, E2, E3, E4, and E10 were grouped together and the rest were grouped together. This tendency for unidimensional (1-D) sorting held across a variety of stimuli, instructions, and procedures.

FR sorting did not emerge even when 1-D sorting was prevented. In one study, Medin, Wattenmaker, and Hampson (1987) used trinary-valued stimuli and required people to sort examples into two categories (see Table 2

TABLE 2
Abstract Notation of Stimuli Used in Medin, Wattenmaker, and Hampson (1987), Experiment 4

	D1	D2	D3	D4		D1	D2	D3	D4
E1	0	0	0	0	E7	1	1	2	2
E2	0	0	0	1	E8	1	2	1	2
E3	0	0	1	0	E9	2	1	2	1
E4	1	0	1	0	E10	2	2	2	1
E5	1	1	0	0	E11	2	2	1	2
E6	0	1	0	1	E12	2	2	2	2

for the abstract notations of the stimuli). Under these conditions, a small number of FR sortings was observed, but none of the subjects' descriptions were consistent with an FR explanation. Several sorting strategies were observed and the ones used by a majority of the subjects suggested that subjects first looked for a primary dimension to divide most of the stimuli. For example, they used just a single dimension (e.g., "long tail in general" vs. "short tail in general"), primary dimension plus conjunction (e.g., "four legs, or eight legs and a square head"), or primary dimension plus disjunction ("triangle-shaped head or more than eight legs"). These descriptions suggest that people tend to create categories with defining features and if the strategy does not work for some exceptional examples, they tend to patch them up by adding descriptions for the leftover examples. This kind of strategy could, in some cases, yield categories with FR structure.

The preceding analysis assumes that the descriptions given by the participants reveal the process by which they constructed categories. It is possible, however, that the FR sorting was "genuine" and that the descriptions were generated after the sorting. Consequently, one needs a more direct experimental comparison to distinguish true FR sorting (i.e., sorting corresponding to the similarity-based or predictability-driven models just described) from FR sorting that emerges as a by-product of other category-construction principles.

In order to provide a proper experimental contrast, we need to formalize the idea that people are looking for more simple structure than the examples afford. The general notion is that when faced with probabilistic structures, people construct a simple core or primary basis for classification and then adjust for examples that do not conform to the core. We first describe this two-stage model and then outline the experimental contrasts aimed at evaluating it.

A TWO-STAGE MODEL FOR CATEGORY CONSTRUCTION

The two-stage model of category construction was developed to capture people's tendency to construct categories by focusing on a primary feature, as shown in the Medin, Wattenmaker, and Hampson (1987) experiments. In

the model's first stage, the most salient dimension of the exemplars (e.g., size) is selected. (What determines salience of dimensions is beyond the scope of the model.) Then the exemplars are divided into the desired number of groups (e.g., 2) according to extreme values along the selected dimension (e.g., small vs. large). If the dimension is not continuous, then the most common values are selected to be defining features of initial categories. For example, if there are four squares, four circles, and two triangles, and the task is to create two categories, then the two categories in the first stage would consist of four squares in one group and four circles in the other group.

The second stage involves classifying the remaining exemplars that do not have an extreme value on the dimension (e.g., medium), or the ones that do not have the most common values. These remaining exemplars are categorized into one of the initially created groups depending on their overall similarity to each group. The judgement of overall similarity involves all the dimensions in the exemplars.

A few additional comments on the assumptions about the second stage are needed. The main point of the two-stage model is that people prefer imposing a simple structure to judging overall similarity in category construction. Although we initially assumed that people would classify the remaining exemplars based on their overall similarity to the initially created categories, the model allows various second-stage strategies. Besides comparing the exemplars based on overall similarity in the second stage, the subjects may classify the remaining exemplars based on the similarity along only the most salient dimension selected in the first stage. For example, suppose the subjects initially created small versus large categories and the remaining exemplars had medium size. Then subjects may compare the similarity of medium to large and medium to small, and place the remaining exemplars in the more similar category. The various strategies that may be used in the second stage do not seem to be important for the description of the two-stage model because the primary claim of the model is that people first create classical categories and then somehow deal with remaining exceptions.

The two-stage model readily explains Medin, Wattenmaker, and Hampson's (1987) 1-D sorting results. The sets of exemplars used in their experiments required an FR sorting to be based on characteristic features (i.e., features that are generally true for a category but not sufficient for the category). According to the two-stage model, these exemplars do not even have to pass through the second stage of the model to be classified into two groups because all the exemplars consist of binary values. Whichever dimension is selected as the most salient one, exemplars with characteristic values of a contrasting category on the salient dimension will be grouped with the members of the contrasting category, resulting in 1-D sorting.

This explanation would suggest that the basis for category construction is some kind of interaction of processing biases with the structure of the stimuli.

TABLE 3
Abstract Notation of Set A

	D1	D2	D3	D4		D1	D2	D3	D4
E1	0	0	0	0	E6	2	2	2	2
E2	0	0	0	1	E7	2	2	2	1
E3	0	0	1	0	E8	2	2	1	2
E4	0	1	0	0	E9	2	1	2	2
E5	1	0	0	0	E10	1	2	2	2

Surprisingly, the two-stage model also predicts that, under certain circumstances, apparent FR sorting will be obtained. We say "apparent" FR sorting, because, according to the two-stage model, the processing principles going into category construction are identical to those in which FR sorting does not obtain. For example, the model predicts that Set A shown in Table 3 can lead subjects to partition exemplars according to the FR principle when asked to categorize the exemplars into two groups.

Set A differs from the exemplars used in Medin, Wattenmaker, and Hampson's (1987) experiment in that the categories created based on the FR principle (e.g., E1, E2, E3, E4, E5 vs. E6, E7, E8, E9, E10) have sufficient featurs that do not appear in potential contrasting categories. Therefore, in this set, knowing an exemplar has a 0 or a 2 value is sufficient to know to which category the exemplar belongs. With these exemplars, the model predicts that, no matter which dimension is chosen as the most salient dimension, FR categories will be created.

To illustrate more specifically how the model works, suppose the first dimension is chosen as the most salient dimension. Then, E1, E2, E3, and E4 are classified as one category and E6, E7, E8, and E9 are classified as another category. In the second stage, because E5 is more similar to E1, E2, E3, and E4 than to E6, E7, E8, and E9, it is categorized with E1, E2, E3, and E4. Similarly, E10 is categorized with E6, E7, E8, and E9. Therefore, as a result of the second stage, the model generates FR categories for this set.

If the two-stage model is an accurate description of the basis for category construction, the exemplars with sufficient features should lead subjects to create categories based on FR structure. If the task requires subjects to create any number of categories, then the two-stage model predicts that 1-D categories will be created from Set A, and therefore, the set will be partitioned into three groups.

Note that 1-D categories can be created from Set A if subjects, in the second stage, were to use the strategy of assigning the remaining exemplars based on similarity along the most salient dimension. The remaining exemplars (E5 and E10) after the initial categorization have the same value (i.e., Value 1) along the most salient dimension. Therefore, regardless of whether

Value 1 is judged to be more similar to Value 0 or 2, by this strategy they both would be grouped into the same category, resulting in 1-D sorting.

To summarize, the two-stage model argues that people have a strong bias to create classical categories based on a single dimension. When this bias interacts with various structures of stimulus sets and task demands, different kinds of category structure can be produced. For example, categories with FR structure can be created when people are forced to classify exemplars that do not fit the definitions of the classical categories. The current experiments contrast predictions of the two-stage model with predictions generated by the similarity- and predictability-based clustering models as well as Michalski and Stepp's (1983) descriptive simplicity model.

EXPERIMENT 1

Although the Medin, Wattenmaker, and Hampson (1987) studies served to motivate the two-stage model, their empirical evidence was more suggestive than definitive. The present studies vary the structure of examples in such a way as to contrast predictions concerning when FR sorting will be most likely to occur. People's descriptions of their basis for classification will be taken as relevant observations, but the primary data come from constructed categories. Medin, Wattenmaker, and Hampson also did not collect subjects' similarity judgments, and therefore, their results cannot be compared with predictions of those theories that use similarity as a basis for sorting. The present studies use similarity judgments to generate predictions for similarity-based clustering models. The similarity judgments also ensure that 1-D sorting does not emerge simply because there is one dimension that is far more salient than the others.

The current experiment tests alternative category-construction models. As noted previously, the two-stage model argues that a fair amount of FR partitioning will be observed from Set A because this set has sufficient features in the resulting FR categories. When compared with the exemplars used in Medin, Wattenmaker, and Hampson (1987), this set differs from their exemplars not only with respect to the structure of exemplars, but also with respect to the number of matching features between the two resulting FR categories. Consequently, even if there is some increase in FR sorting, it can be attributed to the increased between-category difference, and not the structural difference. So, in Experiment 1, three more sets of exemplars were developed in which within- and between-category similarity and the structure of potential categories (i.e., existence of sufficient features) was varied. These manipulations allowed us to test if the structure of exemplars, rather than similarity relations per se, determine the creation of FR categories.

Four sets of exemplars varying along four dimensions were shown in Figure 1 under Sets A, B, C, and D. Sets A and D had more between-category

Between-Category Difference

Within-Category Similarity	High	Low
High	**Set A** E1 0000 E6 2222 E2 0001 E7 2221 E3 0010 E8 2212 E4 0100 E9 2122 E5 1000 E10 1222	**Set B** E1 0000 E6 2222 E2 0001 E7 2220 E3 0020 E8 2212 E4 0100 E9 2022 E5 2000 E10 1222
Low	**Set D** E1 0010 E6 1221 E2 0031 E7 2242 E3 0100 E8 2214 E4 1300 E9 2422 E5 3003 E10 4122	**Set C** E1 0010 E6 1221 E2 0011 E7 2212 E3 0100 E8 2211 E4 1100 E9 2122 E5 1001 E10 1122

Figure 1. Abstract notation of Sets A, B, C, and D.

difference than Sets B and C, and Sets A and B had more within-category similarity than Sets C and D. Two experiments differing in task demands were conducted using these sets.

Experiment 1a

The model's predictions on the four sets were compared with subjects' sorting behavior in Experiment 1a. Before the models' predictions, specific descriptions of materials and procedure used in the experiment are necessary to explain how predictions of each model were actually made.

Method

Material. Four sets of 10 stimuli (Sets A, B, C, and D) were used (see Figure 1). The actual dimensions used were size, number of arms, shape of edge, and color. In deciding values for each dimension, a pilot study was conducted to create roughly equal intervals between the two adjacent values on the same dimension (i.e., approximately logarithmic steps for a continuous dimension). Furthermore, an attempt was made to equate the salience among the dimensions.

In Sets A, B, and C, there were three values in each dimension: 3 cm, 5 cm, and 8 cm for the size dimension; green, red, and yellow for the color dimension; four, five, and seven for the number-of-arms dimension; dotted

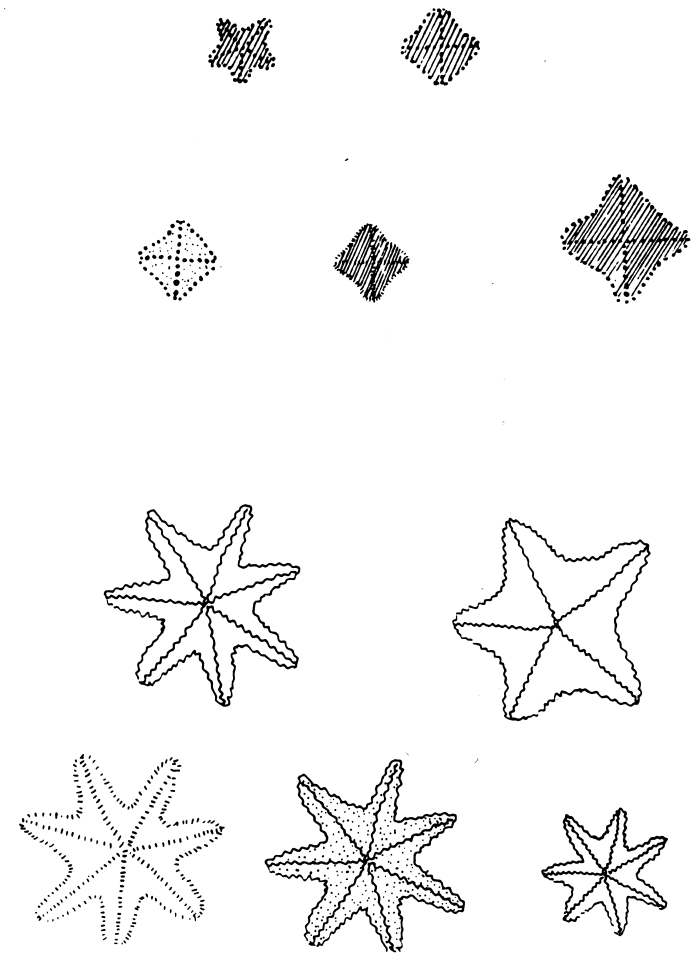

Figure 2. Set A used in Experiment 1

line, wavy line, and cut line for the shape-of-edge dimension. When all these dimensions were combined for Set A, the stimuli looked like starfish as shown in Figure. 2. (In Figure 2 three kinds of patterns were used to indicate the color dimension.)

Set D had more values than other sets: For the number-of-arms dimension, the values were 4, 5, 6, 7, and 8. For the edge dimension, five types of edges were developed, which were the three values used in Set A plus a cross-hatched edge and a squiggly edge. For the size dimension, the values were 3 cm, 4 cm, 5 cm, 6.5 cm, and 8 cm. For the color dimension, the values were red, purple, yellow, brown, and green. In assigning these values to the abstract notation of the exemplars in Figure 1, Value 0 in the abstract nota-

tion was assigned to the smallest (i.e., 4 arms, and 3 cm) or the first value (first type of edge or color) in the dimensions, Value 2 in the abstract notation was assigned to the largest (i.e., 8 arms, and 8 cm) or the last value (the last type of edge and color) in the dimensions. Values 3, 1, and 4 in the abstract notation were assigned to intermediate values in this order.

Design and Procedure. Each subject received a set of 10 randomly mixed examples that were mounted on 9.3 cm × 7.3 cm cards. Subjects were asked to categprize the exemplars into two groups in a way that seemed natural to them. They were also told that there could be different numbers of exemplars in the two groups and that there was no one correct sorting. After they created categories, the subjects were asked to write how they came up with their partitioning. There were four groups of 20 subjects, each receiving either Set A, B, C, or D.

Subjects. Subjects were undergraduate students at the University of Illinois, participating in partial fulfillment of course requirements for introductory psychology.

Predictions of the Models

Similarity-Based Model. A separate experiment was conducted to obtain judged overall similarities of the exemplars used in Experiment 1 for predictions of the similarity-based models. Fourteen subjects were asked to judge the overall similarity of each pair of exemplars on a 9-point scale, 1 being *very different* and 9 being *very similar*. They made similarity judgments on all possible pairs within each set.[1] The subjects were asked to consider all the dimensions of the exemplars when they made their similarity judgments. Four different random orders were used.

To obtain the similarity-based model's predictions, we tested the subjects' similarity ratings on a SAS clustering program, using the VARCLUS procedure. VARCLUS was chosen because it seems to be the most typical nonhierarchical clustering algorithm described in the earlier section. Other statistical clustering programs perform multidimensional scaling analysis, which requires more rigorous assumptions about computing overall similarities (e.g., feature weighting). We avoided these programs in order to minimize any assumptions about similarity computation and to use empirically obtained overall similarity as direct input for a clustering program.

[1] In addition to the four sets of exemplars, Set E was included for purposes not relevant to the present study and was not a part of Experiment 1 because this set was not diagnostic in comparing the four models. The abstract structure of Set E was 0 0 0 0, 0 0 0 1, 0 0 1 0, 0 1 0 0, 1 0 0 0, 2 2 2 2, 2 2 2 3, 2 2 3 2, 2 3 2 2, and 3 2 2 2. With the five sets of exemplars, there were 225 possible pairs for similarity judgment.

VARCLUS selects prototypes of each category and groups each object with each prototype with which it has the highest correlation. After the first assignment, the second phase proceeds in which each variable in turn is tested to see if assigning it to a different cluster increases the amount of variance explained. If a variable is reassigned during this phase, the prototypes of the clusters involved are recomputed before the next example is tested.

We ran VARCLUS on each individual's similarity ratings and not on average ratings because using an average of subjects' similarity ratings for clustering does not necessarily mean that the average subject would produce the same clusters. Even when each subject rated similarities in such a way that the categories produced by the similarity-based model would be 1-D, if each subject picked up different dimensions for the basis of similarity judgment, the average of those ratings could result in FR categories.

For Sets A and B, every subject's ratings produced FR clustering. For Set C, the ratings of 42.8% of the subjects and for Set D, the ratings of 71.4% of the subjects resulted in FR clustering. Therefore, the result shows an index of likelihood of obtaining FR partitionings from each set: Sets A and B are the most likely, Set D next, and Set C is the least likely to yield an FR structure.

Predictibility-Driven Model: Anderson's Model. The predictions of Anderson's model for the four sets of exemplars are given in Table 4. Because the model is sensitive to input order, the simulation was run on 50 trials with randomized input orders. The stimuli used in Experiment 1 consist of discrete values and continuous values, which the model handles differently. For continuous values, prior beliefs on means and variance must be specified (see Anderson, 1991, for details). As in Anderson (1991), the prior means were set to be the halfway point of the range and the prior variance was set to be the square of a quarter of the range. In addition, the coupling probability, c, was set to be either .3 or .4, which are the values used in all of Anderson's simulations (Anderson, 1990; 1991). The partitionings for each coupling parameter are shown under $c = .3$ and $c = .4$.

Table 4 shows the list of various types of partitionings produced by the simulation. Each type is separated by parentheses. The numbers in the parentheses indicate the number of examples in each partitioning. For example, (1 4 5) indicates that there are three groups in this partitioning, one with a single example, one with four, and one with five examples. As it turns out, when the partitioning was (5 5), the clusters always corresponded to FR categories. When the number of examples does not exceed five, the partitioning is always subdividing one of the two FR categories. Naturally, when the number exceeds five, the partitioning does not accord with the FR principle. The numbers under "%" are the percentages of trials generating each partitioning among the 50 trials. To summarize Table 4, FR cateogries were

TABLE 4
Predictions of Anderson's (1991) Model

	c=.4		c=.3	
	Partitioning	%	Partitioning	%
Set A	(5 5)	94	(5 5)	56
	(1 4 5)	6	(1 4 5)	24
			(1 1 3 5)	20
Set B	(5 5)	78	(5 5)	36
	(1 4 5)	8	(1 4 5)	30
	(1 0)	2	(1 1 3 5)	22
	(4 6)	12	(1 0)	2
			(6 4)	8
			(1 3 6)	2
Set C	(5 5)	28	(1 2 2 5)	28
	(2 3 5)	20	(1 2 2 2 3)	24
	(1 4 5)	34	(1 1 2 2 4)	16
	(1 4 1 4)	6	(1 2 3 4)	16
	(2 2 3 3)	4	(2 3 3 2)	8
	(1 2 3 4)	4	(2 3 5)	8
	(1 3 6)	2		
	(3 7)	2		
Set D	(5 5)	48	(5 5)	26
	(1 4 5)	42	(1 4 5)	30
	(1 1 4 4)	8	(1 1 3 5)	20
	(1 2 3 4)	2	(1 1 1 3 4)	6
			(1 2 3 4)	2
			(1 1 4 4)	10
			(1 1 2 3 3)	2
			(2 3 5)	4

the modal partitionings for Sets A and B, and some FR sortings emerged for Sets C and D, with Set D more likely than Set C to yield FR categories.

Predictibility-Driven Model: COBWEB. For the predictions of COBWEB, we tested our data set on COBWEB/3 (McKusick & Thompson, 1990), which was basically the same features as COBWEB except that COBWEB/3 can handle continuous values. Because neither can process examples with mixed-dimension types, which is the case in the stimuli used in Experiment 1, COBWEB/3 has tested on examples with only nominal dimensions. (See Experiment 2 for a test on continuous dimensions.)

As in Anderson's model, COBWEB (and COBWEB/3) is sensitive to input order. For example, when the two most extreme cases of input order were used, different hierarchical structures were produced. In one case (Input Order 1), two prototypes (e.g., 0 0 0 0 and 2 2 2 2) were entered first, followed by eight distortions of the two prototypes with two groups of dis-

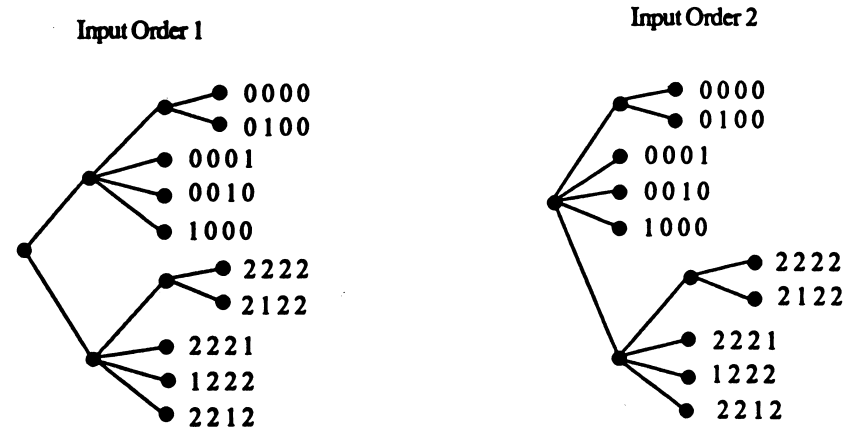

Figure 3. Partitionings of COBWEB for Set A with two input orders.

tortions evenly intermixed (e.g., 0 0 0 1, 2 2 2 1, 0 0 1 0, 2 2 1 2, etc.). In the other case (Input Order 2), one prototype was entered first, followed by four distortions of the prototype and then the other prototype was entered, followed by four distortions of the prototype (e.g., 0 0 0 0, 0 0 0 1, 0 0 1 0, ...2 2 2 2, 2 2 2 1, 2 2 1 2,...etc.). These two input orders led to different types of partitionings as shown in Figure 3. In Figure 3, dots indicate each node and the most abstract node in a hierarchy is placed at the left side of the hierarchy. Nodes at the same level also have the same vertical position in the figure.

For a more thorough test, we added a randomization procedure to COBWEB/3 and tested it on 50 trials with randomized input order. We examined only the types of partitionings at the most abstract level in the hierarchy because more specific levels had many more than two partitionings, which may not be comparable to the current experimental results. At the most abstract level, FR partitionings were almost always produced from Sets A and B (92% of the trials from Set A and 88% from Set B). For Set C, COBWEB produced 28 types of partitionings with three or more than three categories in each type. Naturally, there was no FR partitioning from Set C. From Set D, 24 types of partitionings were produced and except for the FR partitioning (26% of the trials), 23 types of partitionings had three or more than three categories. In sum, COBWEB's prediction was the same as Anderson's model; for Sets A and B, FR categories were the modal partitionings, and the likelihood of producing FR sorting for Set C was the least of all four sets.

CLUSTER/2. For CLUSTER/2, dimension types have to be specified. Dimensions 1 and 2 were linear whereas 3 and 4 were nominal. The parameters used were the default values that are most frequently used in their sys-

tem.[2] With these parameters, CLUSTER/2 generated three clusterings of each set of exemplars, differing in the number of clusters in each clustering. Because the system has no preference among these clusterings, only those clusterings with two clusters will be used for comparison with the results obtained in the experiment, in which subjects were asked to categorize the exemplars into two categories.

Table 5 shows the categorization made by CLUSTER/2 for each set of exemplars. As shown here, CLUSTER/2 did not generate any FR sorting. All the resulting partitionings were 1-D. Furthermore, the system generated categories of very different sizes, which seems psychologically implausible. (Note that previous experiments showed that people had a tendency to create equal-sized categories, Handel & Imai, 1972; Imai, 1966). Although we have not fully explored the parameter space, it is very unlikely to produce FR categories by adjusting parameters, because the most important clustering criterion of this system is simplicity of description, and FR categories cannot be described in simple terms.

Two-Stage Model. The two-stage model predicts that FR categories will be created only from the sets with sufficient features in the potential FR categories (i.e., Sets A, C, and D in Figure 1). The reason that sufficient features are necessary for the creation of FR categories can be illustrated by examining Set B. This set does not have sufficient features in the resulting FR categories and therefore, according to the two-stage model, FR categories will not be created from this set.

Suppose the first dimension was selected as the most salient dimension in Set B. Then E1, E2, E3, and E4 will be placed into Category 1, and E5, E6, E7, E8, and E9 will be placed into Category 2 in the first stage. The remaining example, E10, will be put in Category 2 if subjects categorize it based on overall similarity, or it will be put in either Category 1 or 2 if subjects categorize it based on the similarity along the most salient dimension. Whichever strategy is used, the resulting categories are not FR categories because E5, which has more similarity to Category 1 than Category 2, is placed into Category 2 in the first stage.

There can be some differences among the sets with sufficient features. Sets C and D are less likely to produce FR categories than Set A because overall similarities between the remaining exemplars and the initially created categories are smaller in Sets C and D. Also, Sets A, C, and D can produce 1-D sorting if subjects sort the remaining examples based on similarity only along the most salient dimension.

[2] The default values were Mink = 2, Maxk = 4, covertype = disjoint, H1 = 3, H2 = 2, H3 = 3, Cbase = 2, probe = 2, NIDspeed = fast, maxheight = 99, minsize = 4, beta = 3.0, LEF = [(sparseness = 0.3) (simplicity = 0.3)]. See Michalski and Stepp, 1983, for more details on the parameters.

TABLE 5
Category Construction Made by CLUSTER/2

	Category A	Category B
Set A	0 0 0 0	2 2 2 1
	0 0 1 0	0 0 0 1
	1 0 0 0	
	0 1 0 0	
	2 2 1 2	
	2 2 2 2	
	2 1 2 2	
	1 2 2 2	
Set B	0 0 0 0	0 0 0 1
	0 0 2 0	
	0 1 0 0	
	2 0 0 0	
	2 2 2 2	
	2 2 2 0	
	2 2 1 2	
	2 0 2 2	
	1 2 2 2	
Set C	0 0 1 0	1 2 2 1
	2 2 1 2	0 0 1 1
	0 1 0 0	2 2 1 1
	1 1 0 0	1 0 0 1
	2 1 2 2	
	1 1 2 2	
Set D	0 0 1 0	0 0 3 1
	0 1 0 0	2 2 4 2
	1 3 0 0	
	3 0 0 3	
	1 2 2 1	
	2 2 1 4	
	2 4 2 2	
	4 1 2 2	

Summary of Predictions. So far, specific predictions of each model on the four sets of exemplars have been described (see Table 6 for the summary of these predictions). All of the models, with the exception of CLUSTER/2, predict FR sorting from Set A. Both the similarity-based model and the predictibility-driven model predict that Set C is less likely to produce FR categories than Set B. The two-stage model predicts the opposite effect because Set B does not have sufficient features. This prediction for Set B provides the sharpest contrast between the two-stage model and the similarity-based and predictibility-driven models.

TABLE 6
Summary of Predictions of Various Category-Construction Models

	Two-stage	Similarity	Anderson	CLUSTER/2
Set A	FR, 1-D	FR	FR	1-D
Set B	1-D	FR	FR	1-D
Set C	FR, 1-D	FR	FR or Others	1-D
Set D	FR, 1-D	FR	FR or Others	1-D

Note. FR=family resemblance sorting; 1-D=unidimensional sorting; Others=other types of sorting.

TABLE 7
Types of Sorting in Each Set in Experiment 1

	Set A (%)	Set B (%)	Set C (%)	Set D (%)
FR	55	0	35	20
1-D	45	100	10	65
Others	0	0	55	15

Note. Numbers indicate percentages of sorting types in each condition in each set; FR=family resemblance sorting, 1-D=unidimensional sorting, Others=other responses.

Results

Scoring Method. For Sets A, B, and C, subjects' sortings were considered as 1-D if all exemplars in one of the categories had the same value along any dimension. For Set D, two types of sortings were considered as 1-D; either if one of the two categories had only one value in a dimension, or if one of the two categories had all the exemplars with two adjacent values on a continuous dimension and one of the adjacent values was an extreme value on the dimension. The two FR categories for each set are shown in Figure 1 as two columns, each column representing one of the two categories. Subjects' sortings were considered as FR only when they were the same as the ones in this figure.

Analysis of Types of Sorting. Table 7 shows a summary of the results for each set. As predicted by most of the models, a fair amount of FR sorting was observed in Set A. For this set, 55% of the subjects produced FR categories, and 45% of the subjects produced 1-D sorting.

The most critical test in comparing the previous category-construction models is between Sets B and C. For Set B, all of the subjects produced 1-D sorting. For Set C, 35% of the subjects produced an FR sorting, 55% produced 1-D sorting, and 10% produced other responses. The Fisher's exact test (two-tailed) was used to test the difference between Sets C and B in the proportions of FR sorting. The difference was highly significant, $p < .001$. The difference between Set C and Set A was not significant, $\chi^2(1, N = 40) = 1.616$, $p > .10$.

From Set D, 20% of the subjects produced FR categories, 65% of the subjects produced 1-D categories, and 15% created other kinds of sorting. The proportion of a FR sorting from Set D was not significantly different from Set C, $\chi^2(1, N = 40) = 1.129$, $p > .10$, but was significantly different from Set A, $\chi^2(1, N = 40) = 5.227$, $p < .05$. The Fisher's exact test (two-tailed) indicated that differences between the proportion of an FR sorting from Set D and that from Set B was marginally significant, $p = .053$.

Analysis of Protocols. Subjects' descriptions of how they came up with their categories were also analyzed. Only the descriptions of those who generated FR categories are reported here because the main interest of the current experiments is whether various models accurately describe how people create FR categories.

Almost all of the subjects who created FR categories used only one or two dimensions in category constructions. Thirteen subjects simply mentioned one dimension they used (e.g., "large vs. small") and did not explain what they did with the medium value in the dimension. Four subjects mentioned two dimensions they used (e.g., "more defined and larger" vs. "less defined and smaller"). Three subjects mentioned three dimensions but the descriptions were not clear enough (e.g., "colors first, shapes second, dot configurations last"). Two subjects' descriptions were too vague to be classified. One subject's description was exactly the same as the two-stage model's description of the category-construction process. The description was: "First I broke down into three groups by number of points on each object. Second, I broke down the middle group by size." No subject provided a description of sorting by overall similarity.

Discussion

In the Medin, Wattenmaker, and Hampson experiments (1987), no FR sorting was observed when independent dimensions were used and when subjects were free to create categories of any size. In the present experiment, between-category similarities were increased by developing exemplars in such a way that the categories based on the FR principle had sufficient features. For the first time, a fair amount of FR sorting was observed from stimuli consisting of independent features (i.e., Set A). This result was predicted by the two-stage model, the similarity-based model, and the predictibility-driven model. However, the explanations for the FR sorting were different across the models: According to the similarity-based model and the predictibility-driven model, it was due to increased between-category

difference. Considering the decreased proportions of FR sorting in Set D relative to that in Set A where the between-category difference is constant, this explanation may be plausible.

According to the two-stage model, however, the FR sorting was associated with the presence of sufficient features and the decreased proportion of FR sorting in Set D simply derives from the decreased overall similarity of remaining exemplars to initially created categories. To test these two explanations, Sets B and C were compared. Set C had less within-category similarity than Set B, but it had sufficient features in the resulting FR categories. The results clearly supported the two-stage model's prediction: The creation of FR categories depended on the particular structure of exemplars, and not on overall similarity.

The two-stage model can also explain the 1-D sortings from Sets A, C, and D in terms of the different strategies used in the second stage of category construction. Unlike the similarity-based model or Anderson's model, the two-stage model claims that the 1-D sortings resulted, not from decreased overall similarity, but from the strategy of assigning the remaining exemplars based on similarity only along the most salient dimension. If the 1-D sortings were simply due to the strategies used in the second stage, all of the subjects should produce an FR partitioning when a task does not allow subjects to use the strategy of classifying the remaining exemplars based on the similarity along the most salient dimension. Experiment 1b tests how task demands can change the strategy used in the second stage of category construction.

Experiment 1b

In Experiment 1b, subjects were asked to create two equal-sized categories. For Set A, the two-stage model predicts that this task demand will keep subjects from using the strategy of assigning remaining exemplars based on the similarity along the salient dimension. More specifically, after the first stage, there are four exemplars in each category and two exemplars that are unclassified. If subjects assign both of the two remaining exemplars to one of the groups, it will result in unequal-sized categories. Consequently, the new task demand will force them to classify the remaining exemplars based on their overall similarity to each category, resulting in the creation of FR categories.

On the other hand, for Set B, the new task demand would not affect the *type of sorting*. Although 16 of 20 subjects in Experiment 1a created unequal-sized 1-D categories, it was predicted that forcing them to create equal-sized categories would still lead to the creation of 1-D categories. According to the two-stage model, the 1-D sorting from Set B in Experiment 1a is a consequence of the first stage. Suppose the first dimension were the most salient to a subject, resulting in a category with E1, E2, E3, and E4 and a category with E5, E6, E7, E8, and E9 after the first stage. Assigning

the remaining example (E10) to the category with four exemplars (E1, E2, E3, and E4) in order to create equal-sized categories will not produce FR categories because the category that initially had five exemplars had E5, which was more similar to the contrasting category.

The other models do not provide clear predictions of the effects of the new task demands. For the similarity-based model, it had already predicted that from Set B there would be two equal-sized FR categories. Similarly, the two predictibility-driven models already produced two equal-sized categories most of the time and those categories were always FR categories. It was impossible to make more precise predictions of how these two models would behave differently when they were forced to create equal-sized categories because these two models do not have a feature to specify the number of categories to be created. It seems clear, however, that there is no reason for the models to switch from FR sorting to 1-D sorting for Set B just because they were forced to make categories equal-sized. For CLUSTER/2, allowing complex descriptions of categories might lead to creation of equal-sized categories, but the change seems to be against the spirit of the model. To summarize, none of the other models make the same prediction as the two-stage model under this particular task demand.

Method

The procedure was the same as in Experiment 1a except for the instructions on the category size. Subjects were told that they should create two categories and that there should be an equal number of exemplars in each category. Only Sets A and B from Experiment 1a were used. Two groups of 10 subjects received either Set A or Set B.

Results and Discussion

The results were straightforward. As predicted, all subjects who received Set A produced FR categories, whereas all subjects who received Set B produced 1-D sorting. These results are consistent with the idea that the 1-D sorting created from Set A in Experiment 1a was attributable to various strategies used in the second stage, whereas the 1-D sorting from Set B in Experiment 1a derived from the structure of the exemplars.

General Discussion of Experiment 1

Summary of Results

Experiment 1 compared the two-stage model, the similarity-based model, the predictibility-driven model, and CLUSTER/2 in predicting when FR categories will be created. In Experiment 1a, the FR partitionings were obtained in substantial numbers. Effects of patterns of within- and between-category similarity were contrasted with the structural property of whether

or not sufficient features would be present in potential FR categories. The results showed that the structure of the exemplars was the critical determinant of the creation of the FR categories, as predicted by the two-stage model.

In Experiment 1b, when one of the strategies was prevented, which was assumed to lead to the generation of 1-D sorting in Set A, all of the subjects produced the FR categories. None of the subjects produced the FR categories when the same task demand was imposed on Set B. Again, the two-stage model predicts the effect because it implies that 1-D sorting derives from the first stage for this stimulus set. The results clearly show that FR partitioning can be obtained depending on the task demands and the structure of the given exemplars.

Generality of Procedure

In Experiment 1, the sorting task was constrained in that subjects were asked to create only two categories. According to the two-stage model, if any number of categories are allowed, subjects would not bother to go through the second stage and carry out only the first stage, creating 1-D categories regardless of structures of the categories. In an experiment not reported in this article, this instruction was given to subjects. No subject created FR categories using any of the exemplar sets. Instead, the vast majority (70%) created 1-D categories and the rest created categories that were essentially 1-D (see Ahn, 1990b, Experiment 3 for details.)

Generality of Stimulus Types

In Experiment 1, the stimuli consisted of two nominal or qualitative dimensions and two continuous or numerical dimensions. It is not yet clear how various types of dimensions might have interacted with the category-construction processes. Furthermore, the analysis of 1-D sorting indicated that not all dimensions were used to the same degree as a basis for 1-D sorting. The number-of-arms dimension was used most often (72.5% of 1-D sorting across all four sets of exemplars). Given this unbalanced figure, one may argue that the unequal salience of dimensions might have discouraged subjects from using overall similarity for sorting. However, the similarity judgments obtained in Experiment 1a suggest that the number-of-arms dimension was not the only dimension considered.

An alternative explanation is that a continuous dimension might have been used more often in dividing exemplars into two categories simply because the exemplars with an intermediate value on the continuous dimension could easily be placed with either of the two extreme values. In other words, it might be easier to group five-armed objects with either four-armed or seven-armed objects than to group yellow objects with green or red objects. The generality of the results across different types of stimuli was tested in Experiment 2.

EXPERIMENT 2

Experiment 2 was designed to obtain converging evidence for the two-stage model by varying the types of stimuli. One set of stimuli consisted of only continuous dimensions (continuous condition) and the other set consisted of only nominal dimensions (nominal condition). In Experiment 2, only Sets A, B, and C were used because Set D was not diagnostic in differentiating various clustering models.

As in Experiment 1a, subjects in Experiment 2 were asked to create two categories of any size. The materials and procedure used in Experiment 2 will be presented first, followed by the predictions of the various models for this experiment.

Method

Materials. Two types of stimuli were used: one consisting of continuous dimensions, and the other consisting of nominal dimensions. Set A consisting of continuous dimensions is shown in Figure 4. The exemplars are arranged in such a way that the upper five exemplars and the lower five exemplars each represent one of the resulting FR categories. The four dimensions and three values in these exemplars were as follows:

1. The height of a rectangle of a constant 2.54-cm width was one of the following three values: 3.18, 4.44, or 5.72 cm.
2. The shape on the top of the rectangle was a trapezoid of constant height (1.27 cm) and constant base width (4.44 cm); the shape was varied by varying the top width: 30, 1.57, or 3.50 cm.
3. A white arrow of constant (1.90 cm) length was superimposed on the bottom left of the rectangle and rotated about .30 cm from the left side and .64 cm from the base of the rectangle; the orientations of the arrow were 5°, 45°, or 85° clockwise from the horizontal.
4. The brightness of the rectangle varied in three steps.

Set A, consisting of nominal dimensions, is shown in Figure 5. They are arranged in such a way that the upper five exemplars and the lower five exemplars each represent one of the two resulting FR categories. The four nominal dimensions were: (1) overall shape, (2) shape of a top part, (3) shape of an inner part, and (4) types of pattern. The three values used in each dimension can easily be seen in Figure 5.

Using the same values and the same dimensions, Sets C and B were also developed for both the continuous dimensions and the nominal dimensions. In total, there were six sets of stimuli (two types of stimuli × three sets of stimuli).

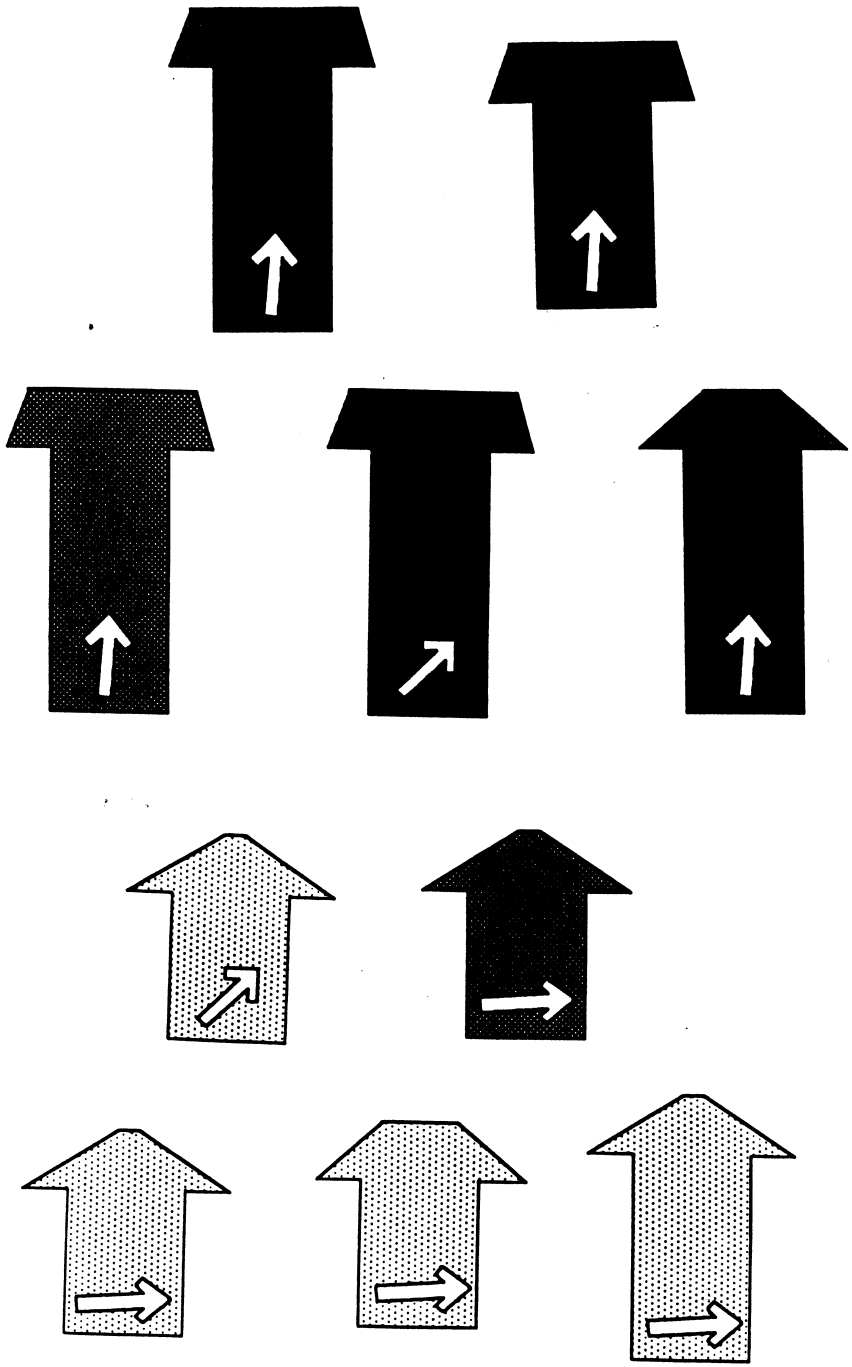

Figure 4. Materials used in continuous condition in Experiment 2.

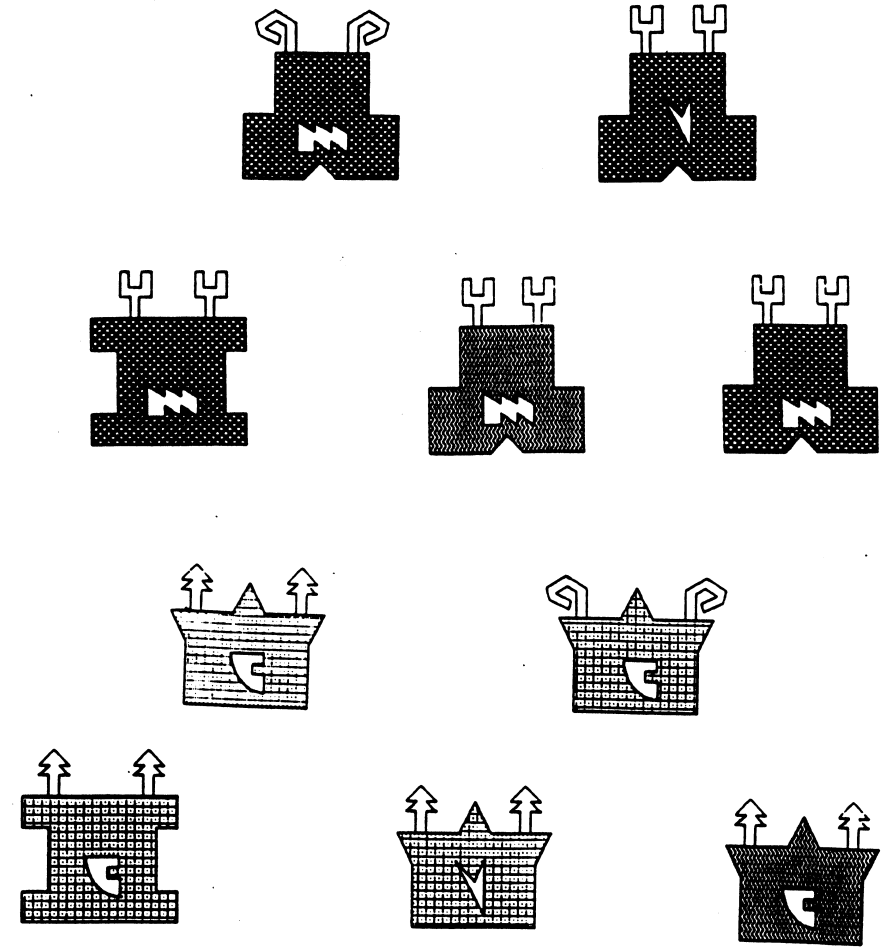

Figure 5. Materials used in nominal condition in Experiment 2.

Procedure. The same procedure as in Experiment 1a was used. Each subject received one of the six sets of examples. For each set, there were 20 subjects, resulting in 120 subjects in total. Subjects were undergraduate students at the University of Illinois, participating in partial fulfillment of course requirements for introductory psychology.

Predictions of Models

Two-Stage Model. Changing the types of dimensions could affect the strategy used on the second stage. In advance of the experiment, no precise prediction could be made on the general differences between the two conditions for Sets A and B. One possible difference may appear from Set C for

the following reasons: As shown in Figure 1, there were four 1's on each dimension in Set C, whereas there were three 0's and three 2's. If a dimension is nominal, people would tend to choose 1 as a primary feature for a category, simply because the Value 1 is the most frequent feature in the exemplars. However, when a dimension is continuous (e.g., size), not frequency but actual value (e.g., small, medium, or large) seems to determine which values would be used as defining values for each category. In Experiment 1, an intermediate value (e.g., medium size) on a continuous dimension was always assigned for the Value 1. In that case, although the Value 1 was the most frequent value in Set C, it was not chosen as a primary value in the first stage, but instead, extreme values (i.e., 0 and 2) were chosen as defining features for each category. According to the two-stage model, when the Values 0 and 2 were chosen as primary values, FR categories can be created from Set C. But if the Value 1 was chosen as a primary value for any of the two categories to be created, then there is no way of producing FR categories by classifying the remaining examples (the ones with either 0's or 2's) in the second stage. The Value 1 is more likely to be selected as the defining value if the primary dimension selected in the first stage is nominal than if it is continuous. Consequently, for Set C, subjects would create fewer FR categories in the nominal condition than in the continuous condition.

Similarity-Based Model. To derive predictions for the similarity-based model, we collected similarity judgment as follows. The procedure was the same as the one described in Experiment 1 for the prediction of the similarity-based model. Subjects were undergraduate students at the University of Michigan, participating in partial fulfillment of course requirements in introductory psychology. There were 12 subjects in the nominal condition and 11 subjects in the continuous condition.

The individual subjects' similarity ratings for each pair were entered as input to SAS clustering programs, using the VARCLUS procedure. For the continuous condition, every subject's ratings produced FR clustering from Sets A and B. For Set C, the ratings of 63.6% of the subjects resulted in FR clustering. The rest of the partitionings were either 1-D (1 subject) or other responses (e.g., grouping E1, E2, E3, and E4 in Figure 1 together and the rest of the exemplars together). For the nominal condition, ratings of all subjects except for one produced FR categories from Set A. From Set B, ratings of 58.4% of the subjects produced FR partitioning and the rest produced 1-D partitioning along the size dimension. From Set C, FR partitionings were reduced to 25% and the rest were either 1-D (41.6%) or other types (33.4%).

To summarize, the similarity-based model predicted almost 100% of FR sorting for Set A in both conditions and for Set B in the continuous condition. Also, in both conditions, Set B would be more likely to result in FR categories than Set C.

TABLE 8
Predictions of Anderson's (1991) Model on the Continuous Condition

	c=.4		c=.3	
	Partitioning	%	Partitioning	%
Set A	(5 5)	100	(5 5)	100
Set B	(5 5)	92	(5 5)	86
	(1 0)	4	(1 4 5)	12
	(6 4)	4	(1 1 4 4)	2
Set C	(5 5)	12	(1 2 2 2 3)	24
	(1 4 5)	20	(1 2 2 5)	20
	(1 2 3 4)	14	(2 3 5)	12
	(2 3 5)	12	(1 2 3 4)	10
	(2 2 3 3)	10	(1 4 5)	8
	(2 4 4)	8	(1 1 2 2 2 2)	8
	(3 3 4)	4	(2 4 4)	6
	(2 2 2 4)	2	(4 3 3)	2
	(1 2 2 5)	2	(2 2 3 3)	2
	(1 3 3 3)	2	(1 1 2 2 4)	2
	(1 1 4 4)	2	(1 3 3 3)	2
	(2 2 6)	2	(2 2 2 4)	2
	(7 3)	2	(1 1 2 6)	2
	(1 9)	2		
	(6 4)	2		
	(1 2 6 1)	2		
	(1 2 7)	2		

Predictibility-Driven Model: Anderson's Model. The simulation of Anderson's (1991) model was run 50 times with randomized input orders using two coupling parameters (either .3 or .4) as in Experiment 1. For the nominal condition, the input was three distinctive values on four dimensions as specified in Figure 1. For the continuous condition, prior belief on means and variances was set based on means and ranges of the actual values. (For the brightness dimension described in Experiment 2b, the values were the number of dots within the same area.) Tables 8 and 9 show the model's predictions for the continuous and the nominal conditions, respectively.

For the continuous condition, a majority of partitionings from Sets A and B were FR categories regardless of the coupling parameter. From Set C, most partitionings consisted of more than two categories. Also they were of various kinds, none of which were generated more than 25% of the trials. Still, most of the partitionings were not against the FR principle in that they were splitting FR categories into smaller categories.

TABLE 9
Predictions of Anderson's (in press) Model on the Nominal Condition

	c=.4		c=.3	
	Partitioning	%	Partitioning	%
Set A	(5 5)	32	(5 5)	6
	(1 1 5 3)	16	(1 4 5)	28
	(1 1 1 5 2)	14	(1 1 1 2 5)	14
	(1 4 5)	10	(1 1 1 3 4)	12
	(1 1 3 1 1 3)	6	(1 1 3 5)	12
	(1 1 1 1 1 2 3)	6	(1 1 1 1 2 4)	10
	(1 1 1 4 3)	6	(1 1 1 1 3 3)	6
	(1 1 1 1 2 4)	6	(1 1 4 4)	6
	(1 1 4 4)	2	(1 1 1 1 1 1 2 2)	4
	(1 1 1 1 1 1 2 2)	2	(1 1 1 1 1 2 3)	2
Set B	(5 5)	16	(5 5)	20
	(1 4 5)	16	(1 4 5)	22
	(1 1 3 5)	10	(1 1 3 5)	12
	(1 1 6 3)	8	(1 1 1 1 1 2 3)	12
	(1 9)	8	(1 1 1 2 5)	8
	(1 1 1 4 3)	6	(1 1 1 3 4)	6
	(1 1 4 4)	6	(1 1 1 1 1 1 2 2)	6
	(1 1 1 2 5)	6	(1 1 1 1 3 3)	4
	(1 1 1 1 1 2 3)	4	(1 1 4 4)	4
	(1 1 1 1 2 4)	4	(1 1 1 1 2 4)	2
	(1 1 8)	4	(1 1 8)	2
	(1 3 6)	4	(4 6)	2
	(4 6)	4		
	(1 1 1 1 3 3)	2		
	(1 1 1 1 1 1 2 2)	2		
Set C	(1 2 2 2 3)	24	(1 1 2 2 2 2)	100
	(1 1 2 2 4)	20		
	(1 1 2 2 2 2)	16		
	(1 2 2 5)	16		
	(1 2 3 4)	10		
	(2 3 5)	6		
	(2 2 3 3)	4		
	(1 4 5)	2		
	(2 2 6)	2		

The partitionings from the nominal condition were more unstable; there were about 10 different types of partitionings from all three sets except when the coupling parameter was .3 for Set C. In addition, a majority of partitionings consisted of singleton categories and more than two thirds of the partitionings were of more than two categories. These categories were mostly either FR categories of split-up FR categories. Because subjects in Experiment 2 were asked to create two categories, the coupling parameter was increased from Anderson's usual value, .3 or .4 to .5 in order to force

the model to generate two categories. The resulting categories were FR categories on all trials for Set A, 94% of the trials for Set B, and 58% of the trials for Set C.

To summarize, there was no 1-D sorting under any circumstances. Also, the coupling parameters and input order greatly affected the types and the number of partitionings. When the coupling parameter was adjusted in order to produce the same type of partitioning unaffected by input order, the FR partitionings were produced more frequently from Sets A and B than from Set C.

Predictibility-Drive Model: COBWEB. The predictions of COBWEB were generated by running COBWEB/3 on 50 randomized trials using actual values used in Experiment 2. In short, COBWEB/3 produced much more than 10 different types of partitionings from all three sets and none of them were FR. Because the assumptions of COBWEB/3 on continuous dimensions have not yet been extensively investigated, these predictions may not be generalizable as COBWEB's predictions. The predictions on nominal dimensions were discussed in Experiment 1. To summarize, Sets A and B resulted in FR partitionings and Set C did not result in any FR partitionings.

CLUSTER/2. CLUSTER/2 was tested on the two types of dimensions using the same default parameters as in Experiment 1. The predictions for the nominal condition were the same as for Experiment 1. The predictions for the continuous condition were also 1-D sorting as shown in Table 10.

Summary of Predictions. So far, we have described the predictions of various clustering models on stimuli consisting of either only nominal dimensions or only continuous dimensions. To summarize, the two-stage model predicts that if sufficient features exist in the resulting FR categories and they are the modal features, subjects will be able to create FR categories. However, if the resulting FR categories do not have sufficient features as in Set B, there cannot be any FR categories. The similarity-based model and Anderson's model both predict that the likelihood of obtaining FR categories would be the greatest for Set A, next greatest for Set B, and least for Set C. In addition, both models predict that the stimuli with continuous dimensions would be more likely to produce FR categories than the stimuli with nominal dimensions. COBWEB predicted no 1-D sorting from any set. CLUSTER/2 predicted 1-D sorting in all sets.

Results

Analysis of Types of Sorting. Table 11 summarizes the results from Experiment 2. For the continuous condition, the results replicated Experiment 1a. FR sortings were obtained most often from Set A (55%), and next from

TABLE 10
Category Construction Made by CLUSTER/2 for the Continuous Condition

	Category A	Category B
Set A	0 0 0 0	2 2 2 2
	0 0 1 0	2 2 2 1
	0 1 0 0	2 1 2 2
	1 0 0 0	1 2 2 2
		2 2 1 2
		0 0 0 1
Set B	0 0 0 0	0 0 0 1
	0 0 2 0	2 2 2 2
	0 1 0 0	2 2 1 2
	2 0 0 0	2 0 2 2
	2 2 2 0	1 2 2 2
Set C	0 0 1 0	0 0 1 1
	0 1 0 0	1 0 0 1
	1 1 0 0	1 2 2 1
		2 2 1 2
		2 2 1 1
		2 1 2 2
		1 1 2 2

TABLE 11
Proportions of Different Types of Sorting in Experiment 2

	Set A	Set B	Set C
Continuous Condition			
FR	55	0	20
1-D	45	95	60
Others	0	5	20
Nominal Condition			
FR	35	0	0
1-D	60	95	80
Others	5	5	20

Note. Numbers indicate percentages of sorting types in each condition in each set. FR=Family resemblance sorting; 1-D=unidimensional sorting; Others=other responses.

Set C (20%). There were no FR sortings for Set B. The Fisher's exact test (two-tailed) indicated that the difference between Set A and Set B was highly significant ($p<.001$), but that the difference between Set C and Set B was not significant ($p=.10$). Also, the difference between Set A and Set C was significant, $\chi^2(1, N=40)=5.227$, $p<.05$.

The proportions of 1-D sorting were 45% for Set A, 95% for Set B, and 60% for Set C. The difference between Sets A and C was not significant,

$\chi^2(1, N=40)=1.600$, $p<.10$. But the Fisher's exact test (two-tailed) indicated that the difference between Sets B and C with respect to 1-D sorting was significant, $p<.05$. Therefore, the nonsignificant difference between Set B and Set C with respect to FR sorting may derive from other responses from Set C.

For the nominal condition, 35% of the subjects who received Set A produced an FR sorting. None of the subjects who received either Set C or Set B produced an FR sorting. Instead, the overwhelming majority of subjects (95% in Set B and 80% in Set C) produced 1-D sorting. The Fisher's exact test (two-tailed) indicated that the difference between Set A and either Set B or Set C was significant, $p<.01$.

Proportions of FR sorting in the two conditions were compared. Although proportions of FR sorting from the nominal condition seem less than those from the continuous condition, none of the differences was significant. For Set A, the difference between the nominal condition and the continuous dimension with respect to FR sorting was not significant, $\chi^2(1, N=40)=1.616$, $p<.10$. For Set C, the Fisher's exact test (two-tailed) indicated the difference was not significant, $p=.106$. Proportions of 1-D sorting in the two conditions were also compared. The only significant difference between the two conditions were obtained from Set C (60% for the continuous condition and 95% for the nominal condition), $p<.05$.

Analysis of 1-D Sorting. For 1-D sortings, we examined which two values were placed together in the same category. (Recall that with three values, e.g., 0, 1, and 2, 1-D sorting consists of placing all exemplars with the same value, e.g., 0, in one category and all exemplars with either of the remaining values, e.g., 1 or 2, in the other category.) This analysis was carried out to examine why there was no FR sorting from Set C in the nominal condition. In all sets in both conditions except for Set C in the nominal condition, all of the subjects created 1-D categories by placing all exemplars with 0 (or 2, for some subjects) in one category and placing all exemplars with the remaining values in the other category (e.g., 2 and 1 or 0 and 1). In contrast, in Set C, 12 of 16 subjects who generated 1-D categories placed all exemplars with 1 in one category and placed all exemplars with the rest of the values (i.e., 0 and 2) in the other category. Therefore, the modal feature 1 in Set C was frequently chosen as the primary dimension on the first stage, which prevented subjects from generating FR categories.

Analysis of Protocols. No subject's description was consistent with the FR principle. In the continuous condition, the protocols of 15 subjects who produced FR categories were as follows. Two subjects mentioned only a single dimension, 6 subjects mentioned two dimensions, 4 subjects mentioned three dimensions. Three subjects' descriptions corresponded precisely

with the two-stage model's predictions (e.g., "First, I compared the little arrow. Then I compared the shading.")

In the nominal condition, the protocols of 7 subjects who produced FR categories were as follows. Two subjects mentioned only one dimension, and 2 subjects mentioned two dimensions. Three subjects' descriptions matched exactly with the two-stage model's predictions (e.g., "First, I looked for the shapes that were most similar. Then there were two cards left. I put those where their color was most similar.").

Discussion

The results from Experiment 2 using different types of stimuli were generally consistent with the results from Experiment 1a. Regardless of whether dimensions were nominal or continuous, FR sorting was obtained most frequently from Set A and no FR sorting was obtained from Set B.

The proportions of FR sorting from Set C varied depending on the type of stimuli. When continuous dimensions were used, the results were consistent with Experiment 1a: There were more FR sortings from Set C than from Set B (although the difference was only marginally reliable).

When nominal dimensions were used, no FR sorting from Set C was obtained as predicted by the two-stage model. The values overlapping across two FR categories (i.e., the Value 1's) were not intermediate ones and there were more exemplars with these values than the other two (i.e., the Values 0's and 2's). This explanation for 1-D sorting in the nominal condition was supported by details of 1-D sorting. The reason why there was no FR sorting in Set C consisting of nominal dimensions seems to be simply because the Value 1, which is the most frequent value in Set C, is used as the primary value on the first stage.

The proportion of FR sorting from Set A was somewhat less when nominal dimensions were used than when continuous dimensions were used. One possible reason is that nominal dimensions are usually parts (e.g., different types of legs, different types of skin, etc.) and as Tversky and Hemenway (1984) argued, people may be more reluctant to separate objects with the same part than to separate objects with the same property, such as size. In this case, subjects might have preferred the strategy of classifying the remaining exemplars based on the most salient dimension.

An additional experiment was carried out to test whether the particular strategy used in the second stage could explain the difference between the nominal condition and the continuous condition. As in Experiment 1b, 9 subjects were asked to create two equal-sized categories from Set A consisting of nominal dimensions (see Experiment 1b for the rationale for this procedure). Eight of the 9 subjects created an FR sorting, showing that the subjects in the nominal condition used the strategy of classifying the remaining exemplars based on the primary dimension more often than the strategy of assigning remaining exemplars based on the overall similarity.

None of the other clustering models can explain the results. The similarity-based model fails because it predicts more FR sorting from Set B in the continuous condition than from Set C. Also, the model has no way of explaining 1-D sorting from Set B by almost all subjects and 1-D sorting from Set C by almost two thirds of subjects in both conditions. The predictability-driven model clearly fails in explaining our results because few of its partitionings were 1-D sorting. CLUSTER/2 cannot explain the results because the system does not produce FR categories.

Overall, Experiment 2 demonstrated the generality of the two-stage model across different types of stimuli. The FR sorting that was observed fits better with the two-stage model than with alternative clustering models. The overall pattern of results is consistent with a two-stage an analysis where the second stage is affected by stimulus types, category structure, and task demands.

GENERAL DISCUSSION

Comparison with Previous Models

Theories of conceptual structure suggest that natural categories are fuzzy categories with an FR structure in which members in the same category are more similar to each other relative to members in the contrast categories. Previously, it was assumed that category representations mirrored this fuzzy structure and that people would prefer to construct FR categories. This approach was taken by the similarity-based model and the predictibility-driven model. Consequently, they predicted that the likelihood of FR sorting would increase as within-category similarity and between-category difference increase. This hypothesis, however, has not been supported by the current experiments because subjects tended to sort exemplars on a single dimension.

The new model proposed in this article sheds a different light on this issue: Even when people prefer classically defined categories, an FR structure may be obtained. If the values used for 1-D sorting do not appear in the contrast category, then sorting through two stages can result in FR partitioning. The current results clearly showed that sufficient features in resulting FR categories were the critical factors in producing FR categories and not the overall similarity.

The current results were also inconsistent with CLUSTER/2, a clustering model emphasizing simplicity as a clustering criterion because for all sets of exemplars, CLUSTER/2 predicted 1-D sorting and could not explain why FR sorting was obtained in certain sets.

Are the Stimuli and Task Demands Realistic?

The similarity-based model and predictibility-driven model were initially developed for incremental clustering, and hence, it might be unfair to compare

their predictions with our results where sorting was done simultaneously. We have already discussed justifications for comparing results from simultaneous sorting tasks with results from sequential sorting tasks but whether FR sorting will emerge under sequential clustering is, of course, an empirical issue.

One may also argue that the current experimental stimuli and tasks are too artificial in that subjects were forced to create two categories from a set of very homogeneous examples. This argument does not seem to be convincing. We have tested Anderson's rational model, which has been shown to explain a considerable variety of data obtained in numerous categorization experiments (Anderson, 1990). It was found that the model created two or more categories using the same coupling parameters that have been applied to these other experiments. In short, our stimuli do not seem to be out of line with other categorization experiments where the predictibility-driven models have been applied successfully.

Another possible objection is that the task is artificial because the subjects were asked to create only two categories. Prescribing a number of categories to be created may not be too artificial because in natural settings, prior knowlege may also indicate how many categories should be created. Creating two categories may be an especially common phenomenon: Mental patients can be diagnosed as either normal or abnormal, friends can be classified into introverts and extroverts, and so on. Still, in the current experiment, this task was imposed not because it is a natural phenomenon but rather because the two-stage model predicts that creation of FR structure occurs only when there are two categories to be created. As noted earlier, when subjects were allowed to create any number of categories, they did not create FR categories. Creation of FR structure occurs as an interaction of special task demands (i.e., creating two categories) and the structure of examples. The purpose of the special task demand was simply to demonstrate this aspect of the model.

The generality of our findings may be constrained by conflicting results obtained by Smith (1981) with different stimulus materials. The general consensus from developmental studies is that children's sorting is based on overall similarity, whereas adults' sorting is based on a single feature (Imai & Garner, 1968; Smith & Kemler, 1977). Smith (1981), however, attributed these findings to the use of only a small number of dimensions (e.g., two or three) in the stimuli. Smith argued that if examples had many dimensions (four in her experiments), adult subjects would also sort the examples based on overall similarity. Smith found results consistent with this prediction. The results conflict with ours because we also used four dimensions, yet did not obtain much FR sorting. Furthermore, the four continuous dimensions used in Experiment 2 were essentially the same as the ones in Smith's experiment except that we used the brightness dimension instead of the color dimension.

A systematic comparison between Smith's experiments and ours revealed four differences in experimental procedures and materials. First, in our experiments the subjects were always asked to create two categories, whereas in Smith's experiments, the subjects could create any number of categories as long as there were at least two groups. However, this difference does not explain similarity classifications in Smith's experiment because when subjects were asked to create any number of categories as discussed in Experiment 1, they still did not produce FR sorting (Ahn, 1990b). Second, in Smith's experiments, the differences between adjacent values do not seem to be psychologically as great as the ones used in the current experiments. In her experiments, the differences between the two adjacent values were too small to be easily detected (e.g., 0.63cm in the height dimension). If one does not look at these exemplars very carefully, some examples look almost identical, resulting in increased likelihood of sorting based on overall similarity. A third difference is that the structure of exemplars used in Smith's experiments did not distinguish sorting based on overall similarity and sorting based on two dimensions. Finally, the dimensions used in Smith's experiment were somewhat integral (Garner, 1974). In her stimuli, subjects could have treated the height and the width as a single dimension such as "the overall shape."

A further test was conducted to examine whether Smith's results were due to these differences (Ahn, 1990b). In this experiment, the structure of stimuli was exactly the same as that in Smith's experiment. In Ahn's experiments, however, when subjects were asked to *carefully* examine the materials or when the one-step dimensional differences were increased, modal sorting responses were based on a single feature. So, this result implies that subjects simply overlooked the dimensional differences and sorted based on combined values. Furthermore, protocols of subjects who produced similarity sorting mentioned only one or two dimensions as a basis of their sorting, indicating that their apparent similarity sorting might be, in fact, sorting based on a single (e.g., "overall shape") or conjunctive features. Consequently, the apparent discrepancy between our experiments and Smith's experiments disappears when we took a closer look: People still prefer unidimensional sorting regardless of the number of dimensions.

Related Evidence and Ideas

Several researchers have presented evidence and ideas related to the current data and the two-stage model. Medin, Wattenmaker, and Michalski (1987) observed similar behavior when subjects were asked to induce a set of rules from preclassified exemplars. They showed subjects pictures of trains that were preclassified into two groups and asked the subjects to come up with a rule that could be used to distinguish them. A frequent observation was that people first came up with an overly generalized rule (a rule which also covered some nonmembers), and then modified the rule to eliminate the

counterexamples by using negative properties. For example, a person might notice that all trains in our group have a triangle load but that two trains in the other group do also. In that case, the person might add to the initial rule another feature that does not apply to the contrasting trains, such as, "triangle load in nonlast car." If people came up with a rule that does not cover all the members, they patched up the rule by using disjunctive features. For example, if a person realized that only one group of trains had two cars but that this rule was not complete, then the person might notice that the remaining trains had jagged tops and came up with the rule, "two cars or a jagged top."

The processing strategies found in their experiments are described in a Patch model (Bettger, 1989; Medin, Wattenmaker, & Michalski, 1987). According to this model, people first select a feature by which most of the exemplars in one category can be described. This initially selected rule is patched up in various ways to be a coherent and consistent rule. The processes used in the Patch model are similar to the two-stage model.

In a similar vein, Michalski (1989) proposed a two-tiered concept representation. In this representation, concepts consist of the first tier, called the Base Concept representation and the second tier, called the Inferential Concept interpretation. The Base Concept representation consists of typical properties of a concept in an explicit, comprehensible, and efficient form. This tier seems analogous to the information used in the first stage of the two-stage model. The Inferential Concept interpretation consists of inference rules and meta-knowledge that define allowable transformations of the concept under different contexts, and handle special cases and exceptional instances. The second tier seems analogous to the second stage of the two-stage model.

Markman (1989) made a similar point suggesting that FR structures could result from stretching classically defined categories to include exemplars that lack defining features of a category but fit better in that category than in any other. Of course, natural categories are not constructed at a single setting in some close analog of our experimental procedures. Clearly, the two-stage model might not be the only method for creating an FR structure.

Perhaps FR categories can be obtained in knowledge-rich domains. If exemplars, differing in surface structure, share a hidden or deeper property, then 1-D sorting based on this hidden property can result in FR structure at the surface levels. Some researchers call these kinds of hidden properties "theories" (Markman, 1989; Murphy & Medin, 1985) and it has been argued that FR structures could be consequences of implicit theories that form the basis for categorization. For example, in Medin, Wattenmaker, and Hampson's (1987) experiments, subjects created FR categories from examples such as "Susan is outgoing, energetic, entertaining, and a daydreamer, Carrie is vigorous, active, courteous, and talkative, Miranda is sad, self-conscious,

inhibited, and a daydreamer, Dawn is withdrawn, solemn, soft-spoken, and spirited, etc." The subjects' protocols (e.g., one group's is "fun to be with" and the other group's is "not the sort of person to take to a party") indicate that their FR sorting is based on a dimension underlying the features used to describe the examples.

The two-stage model is a purely syntactic model and the current experiments used materials only from knowledge-poor domains in which people do not have prior knowledge on the domain-specific goal of category construction, relationships among features, and so on. The category-construction process in knowledge-rich domains might be quite different as observed in the Medin, Wattenmaker, and Hampson (1987) experiment (also see Ahn, 1990a, 1991). The two-stage model might not be literally translated to these situations. Even in knowledge-rich domains, however, one might still expect to see not a mirroring of structure, but a base (or core) plus noise strategy where people impose a certain amount of structure by discovering underlying theories.

Conclusions

The two-stage model successfully predicts when FR sorting will or will not occur. People seem to like structures with more organization than is present in the examples. But then, given the demands of the task, they have to figure out what to do with the examples that do not fit the simple structure. Processes associated with the second stage yield categories where the defining features become converted to typical features. In the same way, FR categories may represent a compromise between a preference for highly structured concepts and the necessity of mapping concepts onto real-world examples.

REFERENCES

Ahn, W. (1990a). Effects of background knowledge on family resemblance sorting, *Proceedings of the 12th Annual Conference of the Cognitive Science Society* (pp. 149–156). Hillsdale, NJ: Erlbaum.

Ahn, W. (1990b). *A two-stage model of category construction.* Unpublished doctoral dissertation, University of Illinois, Urbana.

Ahn, W. (1991). Effects of background knowledge on family resemblance sorting: Part II, *Proceedings of the 13th Annual Conference of the Cognitive Science Society* (pp. 203–208). Hillsdale, NJ: Erlbaum.

Ahn, W., & Medin, D.L. (1989). A two-stage categorization model of family resemblance sorting. *Proceedings of the 11th Annual Conference of the Cognitive Science Society* (pp. 315–322). Hillsdale, NJ: Erlbaum.

Anderberg, M.R. (1973). *Cluster analysis for applications.* New York: Academic Press.

Anderson, J.R. (1988). The place of cognitive architectures in a rational analysis. *Proceedings of the 10th Annual Conference of the Cognitive Science Society.*

Anderson, J.R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J.R. (1991). A rational analysis of categorization, *Psychological Review, 98,* 409–429.

Anderson, J.R., & Matessa, M. (1991). An incremental Bayesian algorithm for categorization. In D. Fisher & P. Langley (Eds.), *Concept formation: Knowledge and experience in unsupervised learning.* San Mateo, CA: Morgan Kaufman.

Bettger, J.G. (1989). *Rule induction: A comparison of human and computer simulation performance.* Unpublished master's thesis, University of Illinois, Champaign.

Bruner, J.S., Goodnow, J.J., & Austin, G.A. (1956). *A study of thinking.* New York: Wiley.

Fisher, D. (1987), Knowledge acquisition via incremental conceptual clustering. *Machine Learning, 2,* 139–172.

Fisher, D., & Langley, P. (1986). Methods of conceptual clustering and their relation to numerical taxonomy. In W. Gale (Ed.), *Artificial intelligence and statistics.* Boston, MA: Addison-Wesley.

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

Gluck, M., & Corter, J. (1985). Information, uncertainty, and the utility of categories. *Proceedings of the Seventh Annual Conference of the Cognitive Science Society* (pp. 283–288). Hillsdale, NJ: Erlbaum.

Handel, S., & Imai, S. (1972). The free classification of analyzable and unanalyzable stimuli. *Perception & Psychophysics, 12,* 108–116.

Imai, S. (1966). Classification of sets of stimuli with different stimulus characteristics and numerical properties. *Perception & Psychophysics, 1,* 48–54.

Imai, S., & Garner, W.R. (1965). Discriminability and preference for attributes in free and constrained classification. *Journal of Experimental Psychology, 69,* 596–608.

Imai, S., & Garner, W.R. (1968). Structure in perceptual classification. *Psychonomic Monograph Supplements, 2,* 153–172.

Katz, J.J., & Postal, P.M. (1964). *An integrated theory of linguistic descriptions.* Cambridge, MA: MIT Press.

Langley, P., Kibler, D., & Granger, R. (1986). Components of learning in a reactive environment. In R. Michalski, J. Carbonell, & T. Mitchell (Eds.), *Machine learning: A guide to current research.* Boston, MA: Kluwer.

Markman, E. (1989). *Categorization and naming in children: Problems of induction.* Cambridge, MA: MIT Press.

Massart, D., & Kaufman, L. (1983). *The interpretation of analytical chemical data by the use of cluster analysis.* New York: Wiley.

McKusick, K., & Thompson, K. (1990). *COBWEB/3: A portable implementation* (Tech. Rep. No. FIA-90-6-18-2). Moffett Field, CA: NASA Ames Research Center.

Medin, D.L., Wattenmaker, W.D., & Hampson, S.E. (1987). Family resemblance, concept cohesiveness, and category constructure. *Cognitive Psychology, 19,* 242–279.

Medin, D.L., Wattenmaker, W.D., & Michalski, R.S. (1987). Constraints in inductive learning: An experimental study comparing human and machine performance. *Cognitive Science, 11,* 319–359.

Michalski, R.S. (1989). Two-tiered concept meaning, inferential matching and conceptual cohesiveness. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning.* New York: Cambrige University Press.

Michalski, R.S., & Stepp, R.E. (1983). Learning from observation: Conceptual clustering. In R.S. Michalski, J.G. Carbonell, & T.M. Mitchell (Eds.), *Machine learning: An artificial intelligence approach.* Palo Alto, CA: Tioga.

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

Nelson, K. (1974). Concept, word, and sentence: Interrelations in acquisition and development. *Psychological Review, 81,* 267–285.

Rosch, E. (1975). Universals and cultural specifics. In R. Brislin, S. Bochner, & W. Lonner (Eds.), *Cross-cultural perspectives on learning.* New York: Halstead Press.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B.B. Lloyd (Eds.), *Cognition and categorization.* Hillsdale, NJ: Erlbaum.

Rosch, E., & Mervis, C.B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573–605.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8,* 382–439.

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

Smith, E.E., Shoben, E.J., & Rips, J.J. (1974). Structure and processes in semantic memory: A featural model for semantic decisions. *Psychological Review, 81,* 214–241.

Smith, L.B. (1981). Importance of the overall similarity of objects for adults' and children's classifications. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 811–824.

Smith, L.B., & Kemler, D.G. (1977). Developmental trends in free classification: Evidence for a new conceptualization of perceptual development. *Journal of Experimental Child Psychology, 10,* 502–532.

Tversky, B., & Hemenway, K. (1984). Objects, parts, and categories. *Journal of Experimental Psychology: General, 113,* 169–193.