# Distributional Expectations and the Induction of Category Structure

Michael J. Flannagan, Lisbeth S. Fried, and Keith J. Holyoak

University of Michigan

Previous research on how categories are learned from observation of exemplars has largely ignored the possible role of prior expectations concerning how exemplars will be distributed. The experiments reported here explored this issue by presenting subjects with category-learning tasks in which the distributions of exemplars defining the categories were varied. In Experiments 1 and 2 the distributional form of a category was found to affect speed of learning. Learning was faster when a category's distribution was normal than when it was multimodal. Also, subjects in the early stages of learning a multimodal category responded as if it were unimodal. These results suggested that subjects enter category-learning tasks with expectations of unimodal, possibly normal, distributions of exemplars. Experiments 3 and 4 attempted to manipulate subjects' prior expectations by varying the distribution of exemplars in the first of two consecutive category-learning tasks. Learning a multimodal category was influenced by the shape of a previously learned distribution and was facilitated when the earlier distribution was either multimodal or skewed, rather than normal. These results are interpreted as support for a dual-process model of category learning that incorporates the effects of prior expectations concerning exemplar distributions.

them.

It is generally agreed that human acquisition of knowledge involves both processes based on analyses of stimulus information and processes based on a learner's prior assumptions and expectations. It is therefore surprising that research on the induction of categories-one of the most basic types of knowledge acquisition-has largely ignored the potential impact of a learner's expectations on what is learned. Numerous studies have documented that subjects presented with a series of exemplars of perceptual categories can acquire information sufficient to accurately classify further novel exemplars. Theorists have proposed several alternative conceptions of how such knowledge is represented, including prototypes (Posner & Keele, 1968; Reed, 1972), feature frequencies (Hayes-Roth & Hayes-Roth, 1977; Neumann, 1977; Reitman & Bower, 1973), and stored instances (Brooks, 1978; Medin & Schaffer, 1978). However, all of these proposals have assumed that what is learned is solely a function of the presented examples-none postulate any role of prior expectations.<sup>1</sup> Yet in view of the pervasive influence of assumptions and expectations in many aspects of cognition, it would be a surprising exception if the inductive processes involved in learning categories were in fact limited to the passive encoding

A preliminary report of results from the first three experiments was given at the 1981 meeting of the Psychonomic Society.

typically employed stimuli intended to be very unfamiliar to subjects (e.g., random dot patterns). Such stimuli ensure that the

of the presented exemplars or some summary representation of

One reason that little stress has been placed on the importance

of the learner's expectations may be that experimental work has

categories to be learned are novel to subjects, but they also make it unlikely that subjects will enter the learning situation with any but the vaguest expectations about the feature dimensions and values that characterize the categories to be learned. Prior expectations might more likely have an impact if subjects were asked to learn, for example, a novel subordinate category of a familiar superordinate (e.g., learning to recognize a new breed of dog from a series of exemplars).

It is possible, however, that people bring important expectations to bear on the learning process even for highly unfamiliar types of stimuli. These expectations will necessarily involve aspects of the categories more abstract than particular feature dimensions and values. The present study provides evidence that category learning is influenced by expectations regarding the form of the *distribution* of category exemplars over a feature space. Our focus on distributions follows from the central role ascribed to distributional knowledge by the *category density model* (Fried, 1979; Fried & Holyoak, 1984). Unlike previous models of category learning, the category density model explicitly assumes that various types of prior expectations (concerning the set of potentially

This research was supported by National Science Foundation Grant BNS-7904730 to Lisbeth Fried and Keith Holyoak. Keith Holyoak was also supported by a National Institute of Mental Health Research Scientist Development Award (1-K02-MH00342). Michael Flannagan was supported by a National Institute of Mental Health training grant (MH-16892).

We would like to thank Lawrence Barsalou, Douglas Medin, and Thomas Wallsten for numerous helpful comments on an earlier version of this article.

Correspondence concerning this article should be addressed to Michael Flannagan, Human Performance Center, University of Michigan, 330 Packard Road, Ann Arbor, Michigan 48104.

<sup>&</sup>lt;sup>1</sup> Although prior expectations have not been explicitly included in previous models of category representation, it is worth noting that they are at least implicitly involved in all prototype models. An attempt to represent a category by a single prototypical example implies a unimodal distributional expectation because it forces an essentially unimodal representation. The role of prior expectations that we propose differs from such a conception mainly in the variety and flexibility of prior expectations assumed to be possible.

relevant feature dimensions, the range of permissible values on the dimensions, the number of categories to be learned, as well as the form of the distribution of exemplars over the feature dimensions) influence the induction process. Although our aim in the present article is a very general one—to identify influences of distributional expectations on induction—the density model will be used as a specific instantiation of the kind of theory that can account for our findings.

Like various other models of category learning, the density model assumes that category exemplars can be represented as configurations of feature values isomorphic to points in a multidimensional feature space. Unique to the density model is its central assumption that the learner uses the presented instances as a *sample* to induce a density function over the feature space for the *population* of potential category members. If the density function has a simple form (e.g., a multidimensional normal distribution), it can potentially be represented by a small set of parameters (e.g., means and variances). On the other hand, if the function is complex, its representation may be no more concise than a histogram of presented instances. But in any case, the density model treats the learning process as the acquisition of knowledge about the distribution of category exemplars over a feature space.

A second assumption of the category density model concerns the decision rule used to classify novel instances on the basis of distributional knowledge. The model assumes that classification decisions are based on a *relative likelihood rule*, which essentially states that the probability of classifying an instance into a particular category is proportional to the relative likelihood that the item was generated by that category's distribution relative to the distributions of the alternative categories. In the present experiments, in which subjects made "member" versus "nonmember" judgments for single categories, the relative likelihood rule predicts that the probability of classifying an item as a category member will increase monotonically with the subjective likelihood that the item was drawn from the category's distribution.

A third assumption of the model is that the learner brings to the acquisition process prior expectations about the form of the category distributions to be learned. In particular, Fried and Holyoak (1984) argue that normal distributions may provide good approximations to the distributions of many natural perceptual categories, which seem to consist of a dense region of "typical" exemplars surrounded by sparser regions of atypical exemplars (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). People may therefore expect artificial categories also to be of this form. Because multinormal distributions are sufficiently described by the mean and variance of each feature dimension (plus covariances if the dimensions are not assumed to be statistically independent), people may approach a category-learning task by immediately using sampled exemplars to estimate these population parameters. Fried and Holyoak present a formal model, embodied in a computer simulation, of how the parameters describing multinormal categories can be induced by using presented items to update current estimates, without necessarily retaining traces of exemplars in permanent memory.

The category density model provides a framework within which it is natural to hypothesize that the learner's expectations about the *type* of category distribution to be presented can influence what is learned. The default for such expectations may be the normal distribution, though one can imagine situations in which specialized knowledge might lead people to expect other alternatives. Whatever the form of a distributional expectation, it should have observable effects on learning. For example, if people expect perceptual categories to be normally distributed (or at least unimodal and symmetric), they may immediately begin to use exemplars to estimate parameters appropriate for describing normal distributions (i.e., means and variances). If the presented category distribution is in fact of the assumed form (i.e., multinormal), learning will proceed efficiently. However, if the actual distribution is markedly non-normal, the subjective representation of the category that the learner induces will be seriously in error. For example, suppose the actual distribution is bimodal on each continuous feature dimension, with extreme values more likely than central ones. If people assume normality and simply estimate means and variances (which is of course possible for any distribution), a likelihood decision rule will lead them to accept items with unlikely central values (i.e., those near the estimated mean) as category members rather than the items with more likely extreme values.

Indeed, without some additional source of information, a strict parameter-revision process would be unable to detect even gross violations of prior assumptions, and thus could never accommodate an unexpected distribution. Although we are proposing that category learning is influenced by prior expectations, we do not wish to suggest that erroneous expectations can never be overcome. One mechanism for detecting such errors would be to estimate an additional parameter representing the "surprisingness" of the presented exemplars given the current distributional parameters. More generally, the frequent occurrence of items with supposedly unlikely extreme values, and the infrequent occurrence of items with supposedly likely central values, could be used as a cue that the entering assumption regarding the form of the distribution was incorrect.

As evidence against an initially assumed distribution accumulated, it would be desirable to decrease the influence of that assumption in favor of a more flexible, distribution-free learning process. This could be done either by gradually decreasing the weighting attached to the distribution initially assumed or by switching entirely to a representation based on the exemplars actually being presented. In either case, the influence of a more purely data-driven process would emerge. Previous tests of the category density model indicate that such a process may in fact be available to supplement parameter revision, because some learning takes place even when subjects do not know the number of categories to be learned, in which case parameter revision is impossible (Fried & Holyoak, 1984). Furthermore, Malmi and Samson (1983) have found that subjects are capable of retaining fuller representations of exemplar distributions than could be provided by mean and variance information. One mechanism that could account for such findings would involve simply the storing in memory traces of some or all category instances. If these were used to construct a frequency distribution over the feature space, the resulting "histogram" would be expected to eventually approximate the true population distribution. At some point the learner might be able to identify a new (non-normal) function that describes the distribution, and to switch once again to a process of using exemplars to update a parameter set (one more appropriate than that used initially). The range of functions that can be represented parametrically in category-learning tasks is an open empirical issue. However, even in cases for which a fully parametric representation is impossible, a learner may be able to go beyond an instance storage strategy by approximating a complex function with a combination of simpler ones. A bimodal distribution, for example, might be approximated as the disjunctive combination of two normals.

Various versions of this general "dual process" model could be formulated, leading to specific predictions about when violations of expectations will be detected and about what will be learned about intially unexpected distributions. For our present purposes, the central prediction is that if a person enters the learning situation with an erroneous expectation regarding the form of the distribution to be learned, what is induced from exemplars will at least initially be systematically distorted in the direction of the entering expectations. As a result, achievement of a criterion of objective accuracy will be delayed relative to learners who enter the task with more appropriate distributional expectancies. The present experiments were designed to determine whether such distortions and differences in learning rates actually occur.

#### Experiment 1

To test the above predictions, we placed subjects in a categorylearning task and varied the number of exemplars presented for learning and whether or not the category definition was consistent with what we hypothesized to be their prior expectation. Following Fried and Holyoak (1984), we adopted the normal distribution as a plausible candidate for people's "default" expectation, based on its widespread applicability to real world distributions. For a distribution that could be expected to violate people's expectations, we chose a distinctly non-normal, multimodal function that we will refer to as the U distribution. It is described in more detail below.

The methodology of Experiment 1 is closely related to that of a series of experiments by Neumann (1977). Several similarities and differences are worth noting. Neumann's experiments, like those to be reported here, were concerned with how people learn a category from a series of exemplars when those exemplars form a multimodal distribution. The main finding in those experiments was that it was possible for subjects to learn multimodal distributions, though only under some of the conditions that were tested. In interpreting his results, Neumann argued that learning in such a situation is not influenced by assumptions about the general shape of distributions of exemplars. This is contrary to the position we will take, and this issue will be discussed further below in light of the present results. Methodological similarities between Neumann's experiments and our own include the similarity between the U distribution used here and the U and V distributions used by Neumann, as well as the general form of the stimuli in this experiment, which were loosely based on the abstract stimuli used in his Experiment 3 (Neumann, 1977).

Some significant methodological differences between Neumann's experiments and those reported here involve factors that in Neumann's experiments could be expected to strongly promote an "instance storage" approach to the task. He showed subjects a small number of exemplars (eight), and the attributes of his stimuli varied rather coarsely across only five levels. Each exemplar was presented for a fairly long time (10 s), and subjects were told to try to remember exactly the exemplars that they were shown. Later, subjects made old/new judgments for a set of test stimuli. The combination of all of these factors makes it plausible that subjects were both able and motivated to remember each of the specific stimuli that they saw during learning. To the extent that such a strategy is successful, it will necessarily lead to the veridical learning of multimodal distributions that Neumann sometimes observed. But this type of learning may be very different from what takes place in stiuations involving many more exemplars that vary on essentially continuous dimensions conditions that are probably important in at least some forms of real-world concept learning.

In addition to assessing the effect of number of exemplars, the present experiments investigated acquisition of multimodal categories in a situation less favorable to memory for specific stimuli than was the case in Neumann's experiments. In the present experiments many exemplars were presented, each for only a brief interval, and stimuli were varied on more nearly continuous, 10-level dimensions. Also, the task was not presented as one requiring memory for specific exemplars.

#### Method

Subjects. Subjects were 48 undergraduates at the University of Michigan who were selected from the psychology department paid subject pool. They were each paid \$3 plus a variable bonus for their participation.

Stimuli. A set of 1,000 abstract visual patterns, consisting of all combinations of 10 levels on each of three feature dimensions, was used. The patterns were presented on the screen of a standard color video monitor by an Apple II microcomputer, using that machine's low-resolution graphics facilities. The general form of the stimuli is illustrated in Figure 1. Each stimulus consisted of a white background measuring 22 cm horizontally by 18 cm vertically. Three different-colored rectangles were superimposed on this background. Each of these varied on one of its two dimensions, taking on 1 of 10 equally spaced levels, to produce one of the three stimulus feature dimensions.

One dimension (referred to as "F" for "Frame") was the height of a red rectangle that was centered vertically and horizontally within the white background. Its horizontal dimension was a constant 14.0 cm, and its vertical dimension varied between 9.5 and 16.0 cm. A horizontal black line 0.4 cm wide was centered vertically within the red rectangle, bisecting it. Another dimension (referred to as "H" for "Horizontal") was the width of a pink rectangle that appeared just above the black line, centered horizontally. It measured a constant 2.75 cm vertically and varied between 1.5 and 11.75 cm horizontally. The other dimension (referred to as "V" for "Vertical") was the height of a yellow rectangle that appeared just below the black line, centered horizontally. It measured a constant 2.75 cm horizontally and varied between .75 and 4.0 cm vertically. The pink and yellow rectangles were thus always contained within the borders of the red rectangle, which was in turn contained within the unvarying white background area.

Three category-defining functions were used for learning: "normal high-mean" (NH), "normal low-mean" (NL), and "U-shaped" (U). Each was obtained by independently combining three identical, appropriately shaped functions defined over each of the three stimulus dimensions. In order to apply these functions to the stimulus space, the levels of each dimension were assigned values from 1 to 10, with 1 always designating the shortest or narrowest level. Because the NH and NL categories were defined by continuous functions, whereas the stimuli actually varied discretely, the integral of the defining function plus or minus 0.5 unit around each level was used to determine the mass function value for that level.



Figure 1. A "blockist" painting composed of Level 5 on each dimension.

The NL function had a mean of 4 and a standard deviation of 1. The NH function had a mean of 7 and a standard deviation of 1. The functions were truncated below 0.5 and above 10.5 and renormalized, because values outside of those limits did not correspond to any levels of the stimulus dimensions. The U mass function was

$$p(x) = \begin{cases} .0032|x - 5.5|^3 + .0008|x - 5.5| \\ x = 1, 2, \dots, 10 \\ 0 & \text{otherwise,} \end{cases}$$

where x is a stimulus dimension level according to the convention described above. This function (the exact form of which was chosen primarily for ease of calculation) yields a markedly bimodal distribution on each dimension. Because the three dimensions were statistically independent, the overall distribution of the category actually had eight modes. All three of the density functions are illustrated in Figure 2.

A set of 125 stimuli was chosen for use in the test phase. To ensure that the test set was approximately uniformly and independently distributed over the three stimulus dimensions, the complete stimulus space was partitioned into 125 cells by cutpoints that divided each stimulus dimension into five equal intervals. The combination of these cutpoints yielded 125 divisions of the complete stimulus space, each representing 8 possible stimuli. Within each division, 1 stimulus was chosen randomly to be in the test set.

Design. The two principal factors in the design were number of exemplars experienced during learning (20 or 150), and shape of the distribution to be learned (NH, NL, or U). Each of the 48 subjects was assigned to one of the 6 resulting cells: 6 in each of the two NH cells, 6 in each of the NL cells, and 12 in each of the U cells.

*Procedure.* Each subject participated individually in a 45-min session. Subjects were seated in a sound-attenuating booth in front of a color video monitor and a set of response buttons. All instructions were presented to subjects under computer control, as text on the monitor. There

were several pages of instructions, which the subjects could read through at their own pace.

The session consisted of three parts: an instruction phase, a learning phase, and a test phase. During the instruction phase, the task was introduced to the subject by means of a cover story that represented the stimuli as paintings by a group of abstract artists who worked in a "blockist" style. That the cover story was contrived was of course transparent to subjects; they were simply encouraged to approach the task as if the story were true. Subjects were told that during the learning phase they would see a series of paintings all by one artist, identified to the subjects as "Vango" (referred to here as the "target artist"), and that in the test phase they would have to identify other works by this artist when they were mixed with similar works by other artists.

The instructions emphasized that the distractor paintings would be very similar to those of the target artist, because all of the other artists worked in the same extremely spare, constrained style. Subjects were told that during the instruction phase they should simply observe the examples as carefully as possible in order to learn to recognize the target artist's style.

A bonus system was used to help maintain subjects' motivation. They were told that they could increase their pay for the experiment by a modest amount if they were accurate in recognizing the paintings. Because the objective definition of the target artist's style was probabilistic, subjects' answers could not be simply categorized as right or wrong, and the bonus system was actually based on a somewhat complex rule. In order not to affect subjects' performance except by generally increasing motivation, no feedback was given about the bonus until subjects had completed the experiment.

Before the learning phase began, subjects were introduced to the relevant stimulus dimensions and the full range of possible variation in the stimulus set. They were told that the school of artists being considered worked in an unusually simple and constrained style, so that when one of them created a painting, there were only three aspects of it that could be varied. The three stimulus dimensions and their ranges of variation were then described.

In the learning phase, a series of patterns was generated online according to the appropriate defining function (NH, NL, or U), with each pattern presented for a 3-s duration. A brief tone signaled the onset of each



Figure 2. Probability mass functions used to define the NL, NH, and U categories in Experiments 1-3.



Figure 3. Categorization responses in each condition of Experiment 1, shown as a function of "composite" stimulus dimension.

pattern. The interstimulus interval was 2 s. Depending on the experimental condition, either 20 or 150 patterns were presented. At the end of the learning phase, subjects were given a 5-min break.

In the test phase, the 125 standard test patterns were presented in a randomized order. Each pattern remained on the screen until the subject pressed one of two buttons, indicating whether the pattern was judged to be more likely a work of the target artist or of some other artist. Instructions encouraged the subjects to take as much time as necessary and to be as accurate as possible.

# **Results and Discussion**

Figure 3 provides a summary of the pattern of responses for each combination of distribution (NH, NL, or U) and number of exemplars viewed (20 or 150). The proportion of stimuli that were judged to be by the target artist is shown as a function of a variable referred to as *composite* stimulus dimension. The effect of this variable was constructed by simply averaging the main

# Table 1Trend Analyses for Experiment 1

	Trend				
Condition	Linear	tª	Quadratic	t*	
NL					
20	-3.61	5.71*	-3.13	7.83*	
150	-3.22	5.08*	-3.71	9.25*	
NH					
20	5.79	9.15*	-1.56	3.89*	
150	8.79	13.88*	-3.10	7.75*	
U					
20	2.82	5.12*	-1.52	4.36*	
150	2.42	4.39*	0.25	0.72	

Note. NL = normal low-mean, NH = normal high-mean, U = U-shaped distribution. 20 and 150 refer to the number of exemplars experienced during learning phase.

df = 180. \* p < .001.

effects of the three individual stimulus dimensions (F, H, and V). Figure 3 thus provides a simple summary of the effect of stimulus level on subjects' responses that ignores any differences in the main effects of the stimulus dimensions as well as any interactions between them.<sup>2</sup> To provide a quantitative description of the patterns of responses shown in Figure 3, linear and quadratic trends based on standard coefficients of orthogonal polynomials were computed and are given in Table 1.<sup>3</sup>

*Normal functions.* For all conditions involving normal defining functions, the quadratic trends were negative, suggesting that subjects in all of those conditions acquired a qualitatively accurate representation of the shape of the distribution that they were shown, even after only 20 exemplars. These negative quadratic

<sup>2</sup> To test for possible interactions between the three dimensions, analyses of variance (ANOVAS) were performed for the response variable using the three stimulus dimensions as factors. Because the standard set of test stimuli consisted of only 125 of the 1000 possible combinations of the 10 levels of these three dimensions, a design based on the full number of levels would have many empty cells. For this reason, the number of levels on each dimension was reduced to 5 by combining adjacent pairs of levels. This resulted in a completely crossed and balanced  $5 \times 5 \times 5$ design, with each subject contributing one observation to each cell. Separate analyses were performed for each of the six learning treatments created by combining the three defining functions and the two levels of learning. Although these analyses yielded evidence for significant interactions among the stimulus dimensions for some learning conditions, the mean squares for these terms were considerably smaller than those for the main effects, often by more than an order of magnitude. Also, informal examination of the data failed to reveal any interesting patterns of interaction. For these reasons, and because only main effects are of primary relevance to our predictions, we will not discuss these interactions further.

<sup>3</sup> Because all of the results in the present experiments involved proportions, and because those proportions were occasionally near zero, all ANOVAS were performed on arcsine transformed data as well as on the untransformed proportions. The results of these analyses were virtually identical. For consistency with the figures, all trends and tests reported are for the untransformed proportions. trends would of course also occur if subjects learned nothing about the exemplars they viewed and were simply biased to respond positively to the center of the stimulus range. However, the high- and low-mean normal conditions were significantly differentiated by their linear trends (which were appropriately negative for the NL conditions and positive for the NH conditions) after both 20 and 150 exemplars, indicating that subjects were in fact learning parameters specific to the distributions they observed.

There is a suggestion in the NH conditions that responses after 150 exemplars more accurately reflected the defining distribution than responses after 20 exemplars, since the quadratic trend for the former condition was significantly stronger than for the latter condition, t(180) = 2.23, p < .05. However, the corresponding difference for the NL conditions was not significant, t(180) = .82, p > .40. The main conclusion we wish to draw here is that subjects' responses after viewing either 20 or 150 exemplars in the normal distribution conditions are consistent with learning the general shape of the distribution viewed, and those responses are not attributable to a simple central response bias.

Although it is not of central interest here, it is worth commenting on an asymmetry between the NL and NH conditions that is evident in Figure 3. Responses in the NL condition are most positive in the vicinity of the mode of the defining distribution and decrease in an orderly manner with distance from the mode. Although the NH condition shows a pattern that is roughly similar, there is only a suggestion of a decrease above the mode. A possible explanation for this difference involves the almost certain lack of linear correspondence between intervals on the physical and perceptual dimensions of these stimuli. Steps between adjacent physical levels always involved equal absolute differences in size, meaning that steps between numerically higher levels involved smaller relative increases than those between numerically lower levels. Although we have no independent evidence regarding how levels of these stimuli are perceived or discriminated by subjects, it is reasonable to suppose that equally discriminable differences between them are more closely related to relative than to absolute differences in size. If that is true, subjects in the NH condition might have trouble discriminating the stimulus levels above the mode from the mode itself. Confusions among these levels during either observation of exemplars or decisions about test items could lead to the observed elevation of responses to the upper stimulus levels.

U-shaped function. In sharp contrast with the results for the normal conditions, responses in the U-20 and U-150 conditions were not consistent with veridical learning. To reflect the general shape of the concept-defining distribution in these conditions, responses should exhibit a positive quadratic trend. Instead, the U-20 condition showed a significant negative trend and the U-150 condition showed no significant quadratic trend. In addition, there were positive linear trends in the data for both of the U conditions, whereas the best correspondence to the defining distribution would involve no overall linear trend. These linear trends may be due to differences in discriminability at different levels of the stimulus dimensions as suggested above in connection with performance with normal distributions. For example, suppose that during the encoding of a series of exemplars occurrences of the two lowest levels of a dimension are tabulated separately, whereas occurrences of the two highest levels are assigned to a single combined high level. For the U distribution this would lead to the combined high level having a higher count of occurrences than the most frequent low level. Assuming the same pattern of discriminability in the encoding of test items, each of the two highest levels should receive more positive responses than either of the lower levels.

The quadratic trends are readily interpretable in terms of the dual process category-learning model outlined above. Although the pattern of responding in the U-20 condition cannot in any simple way be attributed to the distribution of stimuli that were actually seen, it is clearly not random, suggesting that it reflects some systematic bias on the part of the observers. Interpreted in terms of our model, the negative quadratic trend after 20 exemplars (a relatively small amount of experience) is due to observers' initial bias to fit distributions of exemplars with normal distributions. The essentially accurate performance after 20 exemplars in the normal conditions is of course consistent with such a bias; it is only when a significantly non-normal distribution of items is encountered, as in the case of the present U distribution, that the category representation is systematically distorted.

The flat, relatively neutral pattern obtained in the U condition after 150 exemplars represents an intermediate state of learning in which, for these subjects as a group, neither the original assumption of normality nor a veridical representation built by the flexible learning process is predominant. The flat response pattern might have resulted either from an equal mixture of both influences or from a homogeneous neutral state in which the original assumption of normality had been abandoned but no new representation had been created. Examining individual subjects provides little evidence for a mixture at the level of the group. Although quadratic trend values for individual subjects did differ (six were positive, and six were negative), the dispersion of those values was about the same as that of the values for subjects in the 20-exemplar condition. If the 150-exemplar group had been a mixture of some subjects who had learned the distribution veridically and others who were still strongly influenced by the assumption of normality, the dispersion of scores should have been higher than in the presumably homogeneous 20-exemplar group.

The interpretation of the 150-exemplar results as representing a transitional stage of learning suggests that after even more experience people's responses will eventually reflect the distribution actually seen. Experiment 2 tested this prediction by presenting 600 exemplars from the U distribution, distributed over several sessions.

### **Experiment** 2

#### Method

Subjects. Eight students from the psychology department paid subject pool who had not served in the first experiment were paid \$8 plus a variable bonus for their participation.

*Procedure.* Subjects participated in four sessions, each on a separate day. A total of 600 training exemplars were presented. Because of scheduling conflicts the days were not always consecutive, but in all cases the four sessions were completed within 1 week. Each session was a slightly modified version of the procedure for the U-150 condition from Experiment 1.

Two minor changes were made in the procedure from the first experiment. First, because repetition of the full set of instructions was considered unnecessary, subjects were given only greatly condensed versions in the second, third, and fourth sessions. Instructions in the first session were the same as in the first experiment. Second, both the learning trials and test trials for each session were split roughly in half and presented in interleaved blocks. Thus in this experiment, each session consisted of four blocks: 75 learning trials, 63 test trials, 75 more learning trials, and finally 62 more test trials. In the earlier procedure, the same stimuli would have been presented as a single learning block of 150 trials, followed by a single test block of 125 test trials. This change was made in an attempt to alleviate the monotony of the procedure.

#### Results

The pattern of responses for the final session is shown in Figure 4, using the same conventions as in Figure 3. As can be seen in the figure, the pattern of responses after several hundred exemplars matches the general shape of the U distribution. The course of learning over the four sessions is summarized by the quadratic trends shown in Table 2. These trends are significantly positive (indicating veridical learning) for each session, and show a significant increase over the four sessions, t(189) = 2.25, p < .05.<sup>4</sup>

#### Discussion

The results of Experiments 1 and 2 support the hypothesis that people bring to a category-learning situation an expectation about the shape of the distribution that exemplars of a category will form. Furthermore, it appears that that expectation is for a unimodal distribution, possibly the normal. Two aspects of the results are of particular interest. First, it appears that the shape



Figure 4. Categorization responses in the final session of Experiment 2 (U distribution with 600 training exemplars), shown as a function of "composite" stimulus dimension.

Table 2Trend Analyses for Experiment 2

Session	Quadratic trend	t*
1	1.34	4.06*
2	2.08	6.30*
3	2.50	7.57*
4	2.26	6.85*

df = 189.

\* *p* < .001.

of the distribution of exemplars affects ease of learning, because a normally distributed category (even with a noncentral mean) is learned in far fewer trials than one based on the U distribution. Second, subjects who see a small number of exemplars of the U distribution exhibit a unimodal response pattern that systematically violates the distribution of exemplars actually seen. Both of these results are easily explained as consequences of a normal or other unimodal expectation, suggesting that distributional expectations of the learner play a role in category learning.

An alternative interpretation of these results, which does not involve the influence of a distributional expectation, is suggested by a hypothesis offered by Neumann in accounting for a similar finding (Neumann, 1977). Neumann presented subjects with exemplars of categories that followed multimodal distributions similar to the U distribution used here. He observed that subjects' responses under some conditions did reflect the true shape of those distributions, but under other conditions displayed a peak in the center of the stimulus range, in violation of the distribution of exemplars. One explanation that Neumann considered was similar to the one presented here in that it involved the influence of a factor independent of the stimuli and of the way individual stimuli are encoded (his "local-versus-global schema hypothesis"; Neumann, 1977, p. 190). However, Neumann pointed out that the central peak in responses could also be explained by a model that does not involve such a mechanism. He proposed one such model, the "interval-storage" hypothesis, and ultimately concluded that it gave a better account of when the central peak in responses did or did not occur in his experiments than did the local-versus-global schema hypothesis.

The interval-storage hypothesis can account for a nonveridical central peak in a person's representation of a category while assuming that a representation is built up from a simple tabulation of observed frequencies. The hypothesis assumes that whenever a stimulus is observed, the frequencies of the attributes that it exhibits are simply incremented appropriately in some

<sup>&</sup>lt;sup>4</sup> The match of responses to the U distribution in the first session of Experiment 2 is surprisingly good in comparison to the results of the U-150 condition of Experiment 1. The only difference in procedure between those sessions was the presentation of trials in two blocks in Experiment 2, suggesting that the improved performance might be due to the alternation of learning and test trials. However, the quadratic trend for the first half of the first session of Experiment 2 was already 1.40, not substantially different from the value for the entire session. An alternative possibility is that the context of a 4-day experiment affected the subjects' approach in the initial session of Experiment 2, possibly increasing their willingness to concentrate on a demanding task.

memory representation. A record is thus built up that should faithfully reflect the distributions of attributes in the observed stimuli: Unimodal distributions should give rise to unimodal representations, bimodal distributions should give rise to bimodal representations, and so forth. The model explains why learning is sometimes not veridical by suggesting that the levels on a stimulus dimension that a subject experiences are encoded not as points but as intervals of variable width. When the interval is narrow, the resulting internal representation closely matches the observed distribution. When the interval is wide enough, however, representations of two different modes may overlap. Because instances at both true modes of a bimodal distribution could contribute to the region of overlap, that region might develop the highest internal frequency count and become a spurious mode. A unimodal representation would thus emerge from experience with a bimodal distribution.

The present results, demonstrating an effect of amount of experience on whether or not a multimodal distribution is learned, present a problem for the interval-storage hypothesis, which has no mechanism by which different amounts of experience might affect the shape of the representation of a distribution. Whether or not a pattern of overlap produces a spurious mode should be independent of the number of observations involved, unless some provision is made for amount of experience to change the width of the encoding interval. If the interval-storage model is modified to include the assumption that the encoding interval shrinks with greater amounts of experience with a set of stimuli, then it would predict less overlap and consequently more accurate representation after greater numbers of exemplars. We have no direct evidence as to whether or not observers' discrimination of stimulus levels improved with experience in our tasks, although the fact that the low and high means of the normally distributed categories were well differentiated after only 20 observations argues against this hypothesis. However, because the suggestion that even perceptual discrimination along single stimulus dimensions improves with experience is not unreasonable (e.g., Gibson, 1953), the modified interval-storage hypothesis must be considered viable.

The present results, as well as those of Neumann (1977), indicate that observers can learn categories defined by multimodal distributions under at least some conditions. Our results and his also indicate that when observers fail to learn a multimodal distribution, their concept of the distribution will not necessarily be simply random but may exhibit a systematic, unimodal pattern. The existence of that pattern suggests, but does not require, the influence of a prior expectation such as that proposed by the category density model. Perhaps the most direct test of whether or not such a hypothesis is necessary would involve assessing the effect of prior exposure to a non-normal distribution on subsequent learning. If people do have default expectations for normal distributions, prior exposure to a non-normal category may reduce those expectations and facilitate subsequent learning of other non-normal categories, even if different stimuli are used. This approach is taken in the next experiment.

#### Experiment 3

Experiment 3 used a transfer paradigm to discover what effect learning one of the two distribution types used in Experiment 1 has on subsequent category-learning performance. Subjects were given two successive category-learning tasks based on very different stimuli, which sometimes did and sometimes did not involve the same distributional structure. In an attempt to minimize superficial similarities between stimulus sets, the previous stimuli, which varied on three dimensions of spatial extent, were supplemented with a new set of stimuli that varied on three dimensions of a different quality-numerosity. These stimuli were intended to be sufficiently different to prevent any transfer of stimulus-specific responding from the first to the second task, whereas the dimensional structure of the two sets allowed each to be used with the same category-defining distributions. In terms of the distributions used, there were four transfer conditions: Subjects learned either a normal or a U distribution and then transferred to either the same or the alternative distribution. In all cases the stimulus set was changed between the first and second categories.

If prior expectations do have an effect on category learning, and if those expectations are influenced by the form of a distribution in a preceding task, learning at transfer will be different depending on which distribution was learned in the earlier task. Specifically, subjects attempting to learn the U distribution after learning a previous U distribution may learn more rapidly or veridically than subjects presented with the identical task following a normally distributed category. Similarly, the normal distribution may be learned more readily when it follows another normal distribution than when it follows the U distribution. Assuming that the stimulus sets used here are different enough to prevent any stimulus-specific transfer, the occurrence of such effects would indicate that prior expectations about the shape of category distributions do influence category learning.

## Method

Subjects. Sixteen undergraduate students from the University of Michigan psychology department subject pool, none of whom had participated in the previous experiments, served as subjects. Each was paid \$6 plus a variable bonus.

Stimuli. Two sets of stimuli were used. One set consisted of the stimuli used in the previous experiments (hereafter referred to as the "size" stimuli). A new set of stimuli (the "numerosity" stimuli) was constructed to be very different in appearance while still sharing the abstract stimulus space structure of three dimensions with 10 levels on each. This new set, like the first, consisted of patterns displayed on a color video monitor by an Apple II microcomputer using the machine's low-resolution graphics. Each stimulus consisted of a certain number of small rectangles presented against a black background. Each rectangle was 4 mm in height and 6 mm in width (the size of a single pixel in the Apple's low-resolution graphics as displayed on the monitor used).

The rectangles formed three groups, distinguished by color and location. The numbers of rectangles within the groups constituted the three dimensions of stimulus variation, corresponding to the sizes of the three large rectangles in the other set. As with the size stimuli, there were 10 possible levels on each dimension: The number in each group could vary between 2 and 20 by steps of 2. For convenience in identifying the various levels of each dimension and for constructing distributions by quantitative methods, the levels were assigned values from 1 to 10, with 1 designating the level with the largest number. The three colors used were red, white, and blue. The red rectangles always appeared within a rectangular area 11.2 cm high and 5.4 cm wide on the side of the monitor to the subject's left. The

white and blue rectangles were restricted to areas of the same dimensions in the middle and on the right of the screen, respectively. The vertical borders of the middle area were contiguous with the borders of the flanking areas. None of the borders actually appeared on the screen, so that the existence and shape of the areas could be detected only by the restrictions on where rectangles of the various colors appeared. The general appearance of these stimuli is illustrated in Figure 5.

The locations of rectangles within their appropriate areas were determined randomly with the restriction that no two could overlap. Locations of individual picture elements as well as various aspects of the entire display could thus be thought of as additional "noisy" dimensions superimposed on the relevant stimulus dimensions of numerosity. This system for locating picture elements was adopted as a practical measure because it was impossible to vary numerosity without also varying location or arrangement. The category-defining functions used here were the same as in the previous experiments (NH, NL, and U).

Design. Four subjects were assigned to each of the following four transfer conditions: normal distribution followed by normal (N-N), normal followed by U (N-U), U followed by normal (U-N), and U followed by U (U-U). All subjects saw a different set of stimuli in each task, and the order of stimulus sets was balanced across subjects, with 2 subjects serving in each order within each transfer condition. Subjects in transfer conditions involving a normal distribution were assigned to either low-or high-mean conditions in such a way as to balance that factor within each combination of transfer condition and order of stimuli. Because distributional mean was varied only between subjects, the N-N transfer condition always involved transfer between numerically equal means.

*Procedure.* The procedure was similar in most respects to that of the previous experiments. Differences in the procedure for this experiment were as follows. Each subject participated in a 2-hr session during which two complete category-learning tasks were presented, each similar to an entire session in the previous experiments. The distribution used to define the category that a particular subject was to learn in each task was determined by the transfer condition to which the subject was assigned.

Instructions to the subjects introduced the task by means of a cover story similar to the one used in the previous experiments. Subjects were again told that they would be seeing examples of paintings by a particular artist, and that they should observe these carefully so that they could later identify other works by that artist when these were presented mixed with similar works by other artists. To establish some connection between the two category learning tasks and possibly to enhance transfer effects from the first to the second, observers were told that they would be learning to recognize paintings from two different periods in the career of a single artist. They were told that although paintings from these two periods were very different in immediate appearance, they nevertheless all reflected certain aspects of the artist's style that remained constant throughout his career. Subjects were told that because of that continuity of style what they learned about the first set of paintings could be of some help in learning to recognize works from the second set. Because these instructions were more complex than the ones for the previous experiments, they were read to subjects by an experimenter rather than presented on the monitors. Subjects were allowed to interrupt and ask for clarifications.

To test for transfer effects in the second task it was of course necessary for subjects to have learned the distribution presented in the first half. Because subjects in Experiment 1 had been unable to learn the U distribution in a single session, two changes were made that seemed likely to lead to faster learning in the first task. Just as before, subjects saw one set of stimuli presented as exemplars for simple observation and another set presented as test trials requiring classification responses. In this experiment, these sets were presented in relatively small blocks of 25 stimuli each instead of the larger blocks used previously. This was done in the hope that relatively rapid alternation between observing and responding would keep subjects more alert, thereby leading to faster learning. Each block of 25 exemplars was followed by a block of 25 test trials. There were 10 blocks of each type.



Figure 5. An example of the "numerosity" stimuli used in Experiments 3 and 4.

A second change made to promote faster learning of the first category was the introduction of feedback. Immediately after each response, either the word "correct" or "incorrect" appeared on the subject's monitor screen. Because membership for the categories used in this task was probabilistic rather than all-or-none, the basis for giving feedback had to be somewhat arbitrary. For the purposes of feedback, the "correct" response to each stimulus was determined by how it was generated. Half of the 250 test stimuli used in each half of a session were generated randomly by the same algorithm used to produce the exemplar set and were designated as positive instances of the category. The remaining 125 stimuli were the same standard test stimuli used in the previous experiments. These stimuli were evenly distributed throughout the potential stimulus space and were designated as noninstances. To a subject who knew the defining distribution perfectly but who did not know which specific stimuli were drawn from it, it would therefore appear that the probability of a "member" response to a particular stimulus being considered "correct" was never zero or one, but increased with the likelihood of that stimulus having been generated by the category-defining distribution.

In addition to immediate feedback after each response, subjects were given summaries of their performance at the end of each block of 25 responses. The summary included the numbers of correct and incorrect responses in that block, the overall percentage correct for the session up to that point, and the amount of bonus awarded. As in the previous experiments, subjects received a variable bonus in addition to base pay. Instead of using the relatively complex rule for determining bonuses as in previous experiments, in which individual responses were never designated as correct or incorrect, the bonus here was determined by simply adding or subtracting money for nominally correct or incorrect answers.

Because special measures to ensure veridical learning were only necessary in the first task, feedback was not given during learning of the second category. However, the same blocking of observation and response stimuli was maintained.

#### Results

Test stimuli generated by the algorithm that produced the exemplar set were of course distributed very unevenly over the potential stimulus space, occurring only in regions of the space in which the nominally correct response was likely to be positive. Responses to those stimuli therefore provide relatively limited information about a subject's response tendencies throughout the stimulus space as a whole. Because of this, and because com-



Figure 6. Categorization responses for each distribution in each half of the learning phase of Experiment 3.

bining responses to member and nonmember stimuli would complicate our presentation of results in terms of a composite stimulus dimension, only responses to the 125 standard test stimuli are considered in the summaries reported below.

Learning phase. The results from the first category-learning task are displayed in Figure 6, and linear and quadratic trends for these data are given in Table 3 with corresponding *t* statistics. The results have been summarized separately for the first and second halves of that phase of the experiment to give a rough indication of the course of learning. The most important aspect of these results is that all three distributions had been learned by the time subjects were transferred to the second categorylearning task. This is indicated by the linear and quadratic trends for the second halves, which were all in the directions predicted for veridical learning of the various distributions. For the two normal distributions, both quadratic trends were negative, whereas the linear trend was positive for the NH distribution

 Table 3

 Trend Analyses for Learning Phase of Experiment 3

	Trend				
Distribution	Linear	ťª	Quadratic	ťª	
NL					
1	-5.33	6.22*	-2.67	4.92*	
2	-5.76	6.71*	-4.24	7.80*	
NH					
1	5.28	6.15*	-2.40	4.42*	
2	7.67	8.94*	-2.42	4.46*	
ບ					
- 1	-0.93	1.37	0.80	1.85	
2	-0.34	0.50	1.55	3.59*	

Note. NL = normal low-mean, NH = normal high-mean, U = U-shaped distributions. 1 and 2 refer to the first and second halves of the learning phase, respectively.

df = 36 for NL and NH distributions; df = 54 for U distribution. \* p < .001. and negative for the NL distribution. The quadratic trend for the U distribution was positive, and the linear trend was not reliably different from zero.

A secondary aspect of these data concerns the time course of learning for the various distributions. If multimodal distributions are in fact harder to learn, as the results of Experiment 1 indicated, responses to the two normal distributions should conform to the shape of the defining distribution relatively early, whereas learning of the U distribution should be relatively slow. Although the present data are not rich enough to provide a continuous record of the learning curves for the different distributions, splitting the learning phase into halves, as in Table 3, allows a coarse assessment of rates of learning. Because of the range of possible learning rates, this split allows only a weak test of the prediction that U distributions will take longer to learn than normal ones. Although the results are not conclusive, they are in the predicted direction in that the U is the only distribution for which there is not strong evidence of veridical performance in the first half of training. The increase in the quadratic trend between halves in the U condition, however, is not significant, t(54) = 1.23, p > .20.

Transfer phase. To assess possible transfer effects more sensitively, data from just the first half (Trials 1-125) of the second category-learning task were examined.<sup>5</sup> These data are from trials relatively close in time to the first category-learning task and may therefore show transfer effects that are later attenuated over the entire course of learning the new category. The data are shown in Figure 7, and trend tests are summarized in Table 4.

The patterns of responding in the N-U and U-U conditions were clearly affected by the type of distribution that was learned earlier. For the U-U condition the quadratic trend was significantly positive, consistent with veridical learning of a U distribution. However, for the N-U condition a significantly *negative* quadratic

<sup>&</sup>lt;sup>5</sup> The results were not substantially different when all 250 trials were included, although the difference between the N-U and U-U conditions was attenuated.



Figure 7. Categorization responses in each transfer condition in the first half (Trials 1-125) of the transfer phase of Experiment 3.

trend was obtained, opposite to that which should result from veridical learning. This quadratic trend indicates a systematically unimodal deviation from veridical performance, similar to that found in the U-20 condition of Experiment 1.

In contrast to the effect of prior distribution on subsequent learning of U-shaped categories, there was no evidence that the prior distribution had any effect on transfer to normal distributions. As Figure 7 indicates, both the NL and NH conditions produced response functions with single modes regardless of whether the initial category distribution had been normal or Ushaped.

Several secondary aspects of performance with the normal distributions were unexpected and are worth noting. First, although the NH conditions showed a strong positive linear trend, there was no indication of a corresponding negative trend in the NL conditions. The differences between the NL and NH conditions involved here appear to be exaggerated versions of the differences that were noted in Experiment 1, and that are evident

# Table 4

Trend Analyses for Transfer Phase of Experiment 3 (First 125 Trials Only)

Condition	Trend				
	Linear	t*	Quadratic	ť	
U-NL	-0.41	0.27	-2.81	2.97*	
N-NL	-0.24	0.16	-1.91	2.02	
U-NH	9.68	6.47**	-1.58	1.67	
N-NH	10.42	6.96**	-1.81	1.91	
U-U	-0.36	0.33	1.96	2.88*	
N-U	1.58	1.47	-2.38	3.50*	

Note. N = normal, L = low-mean, H = high mean, U = U-shaped distributions.

df = 36 for the U-NL, N-NL, U-NH, and N-NH conditions; df = 54 for the U-U and N-U conditions.

\* p < .01. \*\* p < .001.

to a lesser degree in the learning phase of this experiment. In each case the differences can be characterized as inappropriately positive responding at the higher levels of the stimulus dimensions and may be due to discriminability differences, as suggested in the discussion of Experiment 1. Second, as can be seen in the left panel of Figure 7, there was an unexpected effect of prior distribution on overall proportion of positive responses to the NL distribution. This effect was largely due to a single subject in the N-N condition who made an extremely low proportion of positive responses. Although this subject was an outlier in terms of overall proportion of positive responses, the distribution of those responses agreed with that of all other subjects given the NL distribution in that they were clustered around the NL mean.

# Discussion

The effect of the abstract distributional structure of the first category learned on the learning of a subsequent U distribution provides further evidence that prior expectations regarding category structure influence the learning of categories from exemplars. Unlike the pattern of results in Experiments 1 and 2, in which the number of exemplars viewed accounted for whether or not the U distribution was learned, these results cannot be explained by changes in discrimination with experience, as the modified form of Neumann's interval-storage hypothesis (Neumann, 1977) would suggest. Because stimulus sets were changed between halves of this experiment, all subjects went into the transfer phase having had no experience with the stimuli to be presented. Furthermore, experience with the earlier set, in terms of number of exemplars seen, was equal for all subjects.

Amount of experience with a specific set of stimuli, however, is not the only factor that might be expected to affect the width of the intervals involved in Neumann's model. For example, discrimination training on one stimulus dimension may affect the width of encoding intervals or generalization gradients on other dimensions by producing a "set to discriminate" (Thomas, Freeman, Svinicki, Burr, & Lyons, 1970). It is conceivable that some aspect of learning the U distribution, perhaps its general difficulty, caused subjects trained on the U to adopt such a set more strongly than those trained on the normal distributions, thus allowing their performance on the subsequent U distribution to reflect its multimodal character. However, if initial training on the U distribution affects later learning by narrowing encoding intervals, then prior learning of the U should lead to improved performance on any subsequent distribution. For example, in this experiment people who learned a normal distribution preceded by a normal spread their positive responses to parts of the stimulus space well beyond the range in which exemplars had any substantial chance of occurring. Presumably this was due to a failure to discriminate similar stimulus levels in the encoding of either exemplars or test stimuli. If so, prior exposure to the U should have improved that discrimination, resulting in less spread of positive responses for people learning a normal distribution preceded by the U. As can be seen in Figure 7, there was no indication of such an effect.

However, the failure to observe an effect of prior distribution on normal learning is not conclusive evidence that the encoding interval hypothesis is wrong. It could be argued that some aspect of encoding or decision making causes the transfer effect to be easier to observe for the U distribution than for the normal, and that an effect on normal learning did occur but was not detectable given the power of the experiment. Consequently, although the present results cast doubt on the hypothesis that changes in discrimination mediate the transfer effect, they do not definitively reject it.

Two hypotheses that do account for the apparent asymmetry of transfer between the U and normal distributions are suggested by the category density model. Both of these involve distributional schemata and account for asymmetry of transfer by assigning a special status to the schema for the normal distribution. The first hypothesis suggests that people approach a category-learning task with a repertoire of distributions that can be represented schematically. The normal distribution is always included in this set, whereas the complex U distribution typically is not. If the learner encounters a set of exemplars that correspond to a schematic distribution, such as the normal, learning is relatively rapid and veridical. If the exemplars do not fit any immediately available schematic distribution, the learner will eventually shift to a strategy of storing representations of individual instances in order to construct a nonparametric representation of the category density function.

To account for the asymmetric transfer effects obtained in Experiment 3, this hypothesis assumes that after learning a nonstandard distribution such as the U-shaped function, people have a parametric form of that newly learned distribution type available to them, at least for a time. People who learn a U distribution in one category-learning task will develop a schema appropriate for learning another U distribution at transfer. Accordingly, a U distribution will be learned much more readily following a U distribution than following a normal distribution. However, assuming the normal distribution corresponds to a "default" schema that is always available along with any additional schemata, a normally distributed category will be learned relatively easily regardless of the form of the prior category distribution. According to this hypothesis, prior exposure to a non-normal distribution will not eliminate a normal expectation, as was previously suggested, but will merely augment a schematic repertoire in which the normal will still be available.

An alternative hypothesis is that the positive transfer from learning one U distribution to a second is not specific to the U distribution, but rather is due to a general strategy shift. Subjects may tend to approach all category-learning situations with a normal expectation, which they can weaken or abandon in favor of a distribution-free strategy such as instance storage if they discover that their expectation is being violated. The transfer effect on a new task may thus be due to carrying over a tendency to use the general distribution-free learning strategy rather than to applying a newly devised schematic representation. People who first learned a U distribution would go into a second task with little or no tendency to make assumptions about shapes of distributions, and thus should have relatively little difficulty learning another U-shaped distribution. Because they would also quickly learn a normal distribution if it were present, initial U learning would not impair learning of a normal distribution to any significant degree. In contrast, people who learned the normal first would go into the second task still assuming normality, and thus would have difficulty learning a U distribution.

# **Experiment 4**

This experiment was designed to help evaluate the above hypotheses by assessing whether the transfer effect observed in Experiment 3 was specific to the shape of the distribution learned first, as suggested by the first hypothesis, or rather would have generalized to any subsequent non-normal distribution, as suggested by the second. We assessed how well a U-shaped distribution was learned after learning either (a) a normal distribution, (b) the U-shaped distribution itself, or (c) a different non-normal distribution. If the effect of learning one distribution on learning of another is due to the acquisition of a new distributional schema, only a similarly shaped distribution will be facilitated when learned second. If, on the other hand, the effect is due to a shift to a more flexible general strategy, learning of a non-normal distribution should have a beneficial effect on subsequent learning of any other non-normal distribution.

To investigate transfer between two different non-normal distributions, the normal and U distributions used in the previous experiment were supplemented with a new distribution that was unlike either of them. Ideally, this new distribution would have been as different from both the normal and the U distributions as those two are from each other. The distribution actually used was chosen to approximate that ideal. Because the U distribution most notably violates the unimodality of the normal distribution, the other non-normal distribution was selected to conform to that aspect of normality while violating a different one, symmetry.<sup>6</sup>

<sup>&</sup>lt;sup>6</sup> As was pointed out in the discussion of Experiment 1, equal intervals on the physical dimensions of these stimuli are probably not preserved on the perceived dimensions. We are assuming that the correspondence between these dimensions is nevertheless close enough that our nominally symmetric and asymmetric distributions are perceived as such, an assumption that is supported by the fact that these distributions showed systematically different learning and transfer effects.

#### Method

Subjects. Twelve students from the University of Michigan psychology department subject pool, none of whom had participated in the previous experiments, served as subjects. Each was paid \$4 plus a variable bonus.

Stimuli. The sets of potential stimuli were the same as those used in Experiment 3, except that a minor modification was made to one dimension of the "size" stimuli. The H feature was made to vary between 3.5 and 13 cm in width, so that it was always wider than the V feature, which appeared directly below it. The constant width of the F feature was increased to 16 cm to accommodate the wider range of the H feature. These changes were made to eliminate what several subjects had described as a discontinuity in the variation of the H feature. Previously, at the two narrowest levels of that feature it was either narrower or the same width as the V feature below it. Subjects reported informally that the resulting patterns appeared to be special cases, because the H and V features did not then have the configural quality of forming a T, as they did at all other levels of H.

The three distributions used to define categories included the U distribution used previously. A single normal distribution with a mean in the center of the possible stimulus range (5.5 in terms of the convention described previously) was used instead of the two distributions with different means used previously. The standard deviation for this distribution corresponded to a step between two adjacent stimulus levels, just as before. The new form of distribution introduced in Experiment 4 was selected to be markedly asymmetric. This distribution was that of a modified geometric random variable. As with the U distribution, the exact form of this "G" distribution was chosen primarily for ease of calculation in generating exemplars. The probabilities of occurrence of the levels of a stimulus dimension are given by

$$p(x) = \begin{cases} ab(1-b)^{6-x} & x = 1, 2, \dots, 6\\ 0 & \text{otherwise,} \end{cases}$$

where x is a stimulus level according to the convention described earlier, b is a constant equal to two-thirds, and a is a constant for normalization. In this distribution, stimulus levels corresponding to values from 7 to 10 cannot occur; probability of occurrence is highest for Level 6 and declines for levels below 6. The shape of the distribution is illustrated in Figure 8. As with the other distributions, this unidimensional function was used to construct a function over three dimensions by applying it to each dimension independently.

Design. Four subjects were assigned to each of three conditions: normal distribution learning (N-U), U learning (U-U), and geometric learning (G-U), each followed by transfer to the U distribution. All subjects saw a different set of stimuli in each task, and the order of stimulus sets was balanced across subjects, with two subjects in each order assigned to each transfer condition.

*Procedure.* The procedure was the same as in the previous experiment, except that the second task was shortened. Instead of viewing 250 exemplars and making 250 responses in interleaved blocks of 25, subjects viewed only 25 exemplars, all in a single block. The stimuli subsequently presented for classification responses were the standard 125 test stimuli used in previous experiments. This change was made to maximize information about the early stages of learning the second category, where differences between the transfer conditions might be expected to be greatest, as was in fact observed in Experiment 3. Because of this change, sessions were shorter overall than in the previous experiment, lasting approximately 1 hr, 15 min.

## **Results and Discussion**

Learning phase. Results obtained in the first category-learning task are presented in Figure 9 and Table 5, summarized



Figure 8. The "geometric" (G) mass function used in Experiment 4.

separately for the two halves of that phase of the experiment. Table 5 displays the results of quadratic trend analyses for responses to the normal and U distributions. If the categories were learned veridically, then the quadratic trend should be negative for resonses to the normal distribution and positive for responses to the U distribution. Unlike the previous experiments, for the new central-mean normal there is no prediction of an overall linear trend. For the geometric distribution condition, Table 5 displays the results of a contrast designed to quantify the symmetry of responses around the mean of the geometric distribution (5.5). The coefficients of this contrast are simply the cubed deviations of the individual stimulus levels from 5.5, and the index produced is thus similar to the skew of the distribution of subjects' positive responses, except that it measures symmetry around a point fixed a priori (the mean of the presented exemplars) rather than around a fitted mean. If the G distribution is learned veridically, the symmetry index should be negative, reflecting the negative skew in the exemplar set.

Results for the normal and U distributions were similar to those seen in Experiment 3. As indicated by the quadratic trends reported in Table 5, the general shape of the normal distribution was learned even by the end of the first half. Responses to the U distribution show a quadratic trend that is significant in the second half, but does not quite reach significance in the first half. As in Experiment 3, the increase in the quadratic trend between halves was not significant, t(108) = 0.70, p > .40. As was pointed out in the discussion of Experiment 3, comparing halves of the learning phase provides only a coarse record of the learning curves in these conditions. However, the weak evidence that this analysis can provide is consistent with the prediction that a normal distribution will be learned more quickly than a non-normal one such as the U.

Making inferences from the symmetry measure about having learned in the G condition is somewhat more complicated. A significantly negative value on the symmetry measure might be proposed as a simple criterion for learning the G distribution. However, such a proposal involves the assumption that the zero point of the symmetry measure is a meaningful baseline. That point does in fact correspond to symmetry around the middle of the stimulus range on the objective stimulus dimensions, but as we have pointed out previously, the perceptual dimensions of these stimuli are probably not linearly related to the objective



Figure 9. Categorization responses in each transfer condition for each half of the learning phase of Experiment 4.

dimensions. Because of this, it is necessary to consider an interpretation of the symmetry measure that does not involve assumptions about the zero point. Two aspects of these results seem to allow such an interpretation. First, the symmetry index for the G condition did change significantly between the first and second halves, t(108) = 2.95, p < .01, and in fact changed in a negative direction. Second, if it can be assumed that the objectively symmetric normal distribution was subjectively represented as symmetric, subjects' responses to it can be used as a baseline against which to measure skew. When the symmetry measure was applied to the pattern of responses obtained for the normal distribution, the resulting values were satisfyingly stable over the course of learning, being 30.4 and 30.8 for the first and second halves, respectively. The symmetry value for the first half of the G condition did not differ reliably from the first-half normal value, t(108) = 1.05, p > .20, but the second-half G value was significantly different in the negative direction from the corresponding normal value, t(108) = 4.03, p < .001. This result suggests that subjects acquired some knowledge of skew during

 Table 5

 Trend Analyses for Learning Phase of Experiment 4

	Trend					
Dist.	Linear	ť	Quadratic	t <sup>n</sup>	Symmetry	ťª
N						
1	3.78	3.93***	-3.46	5.67***		
2	4.01	4.17***	-5.02	8.24***		
υ						
I	-0.74	0.77	1.19	1.96		
2	-2.02	2.10*	1.79	2.94**		
G						
1					19.14	2.50*
2					-12.69	1.66

Note. Dist. = distribution; N = normal, U = U-shaped, and G = geometric distributions; 1 and 2 refer to the first and second halves of the learning phase.

$$^{*} df = 108.$$

\* p < .05. \*\* p < .01. \*\*\* p < .001.

the course of learning the G category and is consistent with the suggestion that they began the task with an assumption of symmetry that was later abandoned.

Transfer phase. Results for the second category-learning task are presented in Figure 10 and Table 6. Data from conditions N-U and U-U replicate the findings for the corresponding conditions in Experiment 3: Learning of the U distribution following another U was relatively veridical, whereas responses to the U after learning a normal distribution again showed a single central peak. The new condition, G-U, resulted in relatively veridical learning of the U distribution, and thus appears very similar to the U-U condition. The positive transfer effect between two distributions that bear no obvious resemblance to each other indicates that positive transfer does not depend on the specific shape of the distribution initially learned. The results of Experiment 4 thus favor the hypothesis that transfer is due to a general change in the strategy used to encode a set of exemplars rather than to generation of a schematic representation of the distribution type initially learned.

It should be pointed out, however, that these results do not rule out the possibility that the transfer effects are due to a schema that is general enough to include both the U and G distributions but that is still more constraining than our hypothesized assumption-free strategy. For example, although the U and G distributions were selected to be as different as possible, they still share the characteristic of "high density at the end(s)." If subjects do develop and use schemata of that level of generality, even transfer between the G and U distributions could be due to a common schema rather than a qualitative strategy shift. Study of transfer effects between a greater variety of distributions could in principle illuminate this issue further by delineating families of distributions within which positive transfer effects occur. The inclusiveness of those families could be used to infer limits on the power of any schemata responsible for transfer effects. One extreme possibility would be for all non-normal distributions to form one family. Any schema that could account for transfer within such a set would of course be too weak to be distinguished from a distribution-free strategy.



Figure 10. Categorization responses in each transfer condition for the transfer phase of Experiment 4, in which all subjects were exposed to a U distribution.

# General Discussion

The experiments reported here support the view that the induction of category structure is guided by expectations about the form of the distribution of category exemplars. Such expectations constitute a highly abstract influence on learning, one that has an impact even on the acquisition of categories based on unfamiliar patterns unlikely to elicit more concrete expectations. The overall pattern of results supports a dual-process model of category induction, involving a mechanism based on distributional schemata corresponding to predictable general forms of environmental organization and a mechanism that is relatively independent of distributional assumptions, which can be invoked to provide greater flexibility when the environment is not well described by an expected distributional schema.

More specifically, the results suggest that the normal distribution (or at any rate some unimodal and symmetric distribution) constitutes a general "default" expectation about distributional type, as proposed by Fried and Holyoak (1984). The privileged status of the normal distribution is supported by two major results. First, a category defined by a normal distribution was learned far more readily than one based on a severely nonnormal U-shaped distribution. Second, subjects exposed to a category defined by a U distribution initially tended to classify

 Table 6

 Trend Analyses for Transfer Phase of Experiment 4

Condition	Trend			
	Linear	t <sup>a</sup>	Quadratic	t*
N-U	3.34	3.83***	-1.87	3.40***
U-U	2.56	2.93**	1.39	2.53*
G-U	2.55	2.93**	2.01	3.64***

Note. N = normal, U = U-shaped, G = geometric distribution. • df = 108.

\* 
$$p < .05$$
. \*\*  $p < .01$ . \*\*\*  $p < .001$ .

novel patterns *as if* the distribution were normal. After viewing 20 exemplars, subjects tended to classify patterns with intermediate values on the variable dimensions as category members more often than they classified patterns with extreme values as members, even though the latter type of pattern was objectively more likely to be generated by the defining distribution.

Other results provide support for the existence of an alternative, more flexible learning mechanism. The qualitative form of the non-normal U distribution can be learned if a large number of exemplars are provided, even without explicit error correction. Furthermore, the U distribution can be learned even after relatively few learning trials if a prior category-learning task has placed the observer in a state of readiness for a non-normal distribution. This transfer effect did not appear to be specific to a particular distribution type, because in Experiment 4 we found that initial learning of a category defined by either a U distribution or a skewed geometric distribution facilitated subsequent acquisition of another category defined by a U distribution. The nonspecific nature of the transfer effect suggests that its basis is a shift in learning strategy, from a procedure for estimating parameters of a normal distribution to a procedure less closely tied to any particular distribution type. The overall dual process model of category acquisition would seem well suited to dealing with a world in which categories often, but now always, exhibit a distribution of exemplars that is approximately normal in form. At a theoretical level, the model integrates a schema-based mechanism of category acquisition (Fried & Holyoak, 1984) with an instance-based mechanism (Brooks, 1978; Jacoby & Brooks, 1984; Medin & Schaffer, 1978).

The present study leaves several significant questions about category induction unanswered. First, although the transfer results indicate that a subject's readiness to learn a normal or nonnormal distribution varies with past experience in a way that is consistent with the hypothesis of schematic distributional expectations, that hypothesis is not the only one that might account for those results. A theoretical alternative that merits further consideration is represented by the variations of Neumann's interval-storage hypothesis discussed above. That approach differs from ours by explaining biases toward unimodality and symmetry as effects of wide encoding intervals for individual stimuli rather than by invoking schemata for entire distributions of stimuli. Although we have argued that the hypothesis of a schematic, normal expectation is preferable to an encoding interval hypothesis for the present results, the hypotheses are similar enough that these experiments do not definitively discriminate between them. Because both hypotheses are able to account for biases toward unimodality and symmetry, and because in their most general forms both are somewhat flexible as to when such biases should occur, more precise quantitative results will be necessary to differentiate them. A related issue that may be easier to resolve, and that is perhaps ultimately more interesting in its implications for the usefulness of the hypothesis of schematic distributional expectations, is whether or not people can be biased to learn specific non-normal distributions. Would they, for example, under some circumstances impose a characteristic such as skew on an objectively symmetric distribution? The hypothesis that people use normal distributional schemata, unlike the interval hypothesis, could be extended naturally to account for such results by allowing non-normal schemata. The only distributional expectations suggested by the results reported here were for normal or nearly normal distributions, but that is not surprising given that the normal is a reasonable default and that the artificial stimuli used here were highly unfamiliar. In more familiar or meaningful contexts, in which people may have or think they have some knowledge of the processes generating a set of exemplars, they may make use of more specialized assumptions.

A second issue concerns how the shift in learning strategies hypothesized by the dual process model might occur. If the normal distribution constitutes a default expectation, just how is it overridden? A number of mechanisms are worth considering. As we suggested in the introduction to this article, people may tacitly estimate the "surprisingness" of each exemplar as it is presented, given their current estimates of distributional parameters. Alternatively, people may store a small number of instances in memory, perhaps always saving the most recent few stimuli in a "first in-first out" short-term store. These remembered instances could provide the equivalent of a crude "sample" histogram, which could be compared to the subjective "population" distribution implied by current parameter estimates. As the current or accumulated deviation between the sample and the population estimates increased, the learner might weaken or abandon the assumption of normality in favor of a more open strategy. An important issue for future research will be to determine the degree to which people are susceptible to distortions in their categorization judgments due to failures to detect deviations from the expected distribution type.

A third issue concerns how non-normal distributions are represented. The present study indicates that the severely non-normal U distribution can eventually be learned but leaves open the ultimate form of its representation. Our finding that transfer across successive category-learning tasks is not distribution-specific (Experiment 4) implies that subjects did not learn a general schema for U distributions in the course of learning a single category distributed in that manner. However, this result does not preclude the possibility that schemata for novel distribution types can in fact be learned with sufficient experience, especially for less complex non-normal distributions than the octomodal U distribution used in the present study. In addition to the extreme possibilities that people may represent a non-normal category as either an unsummarized set of stored instances or as parameters of a unitary distributional schema, there is the intermediate possibility that the representation may involve the disjunction of multiple schemata (e.g., a bimodal distribution). Which of these theoretical alternatives are in fact used to represent non-normal category distributions, and under what conditions, constitute prime questions for future research.

# References

- Brooks, L. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch & B. Lloyd (Eds.), Cognition and categorization (pp. 170–207). New York: Wiley.
- Fried, L. S. (1979). Perceptual learning and classification with ill-defined categories (Tech. Rep. No. 79–6). Ann Arbor: University of Michigan, Michigan Mathematical Psychology Program.
- Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psy*chology: Learning, Memory, and Cognition, 10, 234–257.
- Gibson, E. J. (1953). Improvement in perceptual judgments as a function of controlled practice and training. *Psychological Bulletin*, 50, 401– 431.
- Hayes-Roth, B., & Hayes-Roth, F. (1977). Concept learning and the recognition and classification of exemplars. *Journal of Verbal Learning* and Verbal Behavior, 16, 321-338.
- Jacoby, L. L., & Brooks, L. R. (1984). Nonanalytic cognition: Memory, perception, and concept learning. In G. H. Bower (Ed.), *The psychology* of learning and motivation: Advances in research and theory, Vol. 18. New York: Academic Press.
- Malmi, R. A., & Samson, D. J. (1983). Intuitive averaging of categorized numerical stimuli. Journal of Verbal Learning and Verbal Behavior, 22, 547-559.
- Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207-238.
- Neumann, P. G. (1977). Visual prototype formation with discontinuous representation of dimensions of variability. *Memory & Cognition*, 5, 187-197.
- Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. Journal of Experimental Psychology, 77, 353-363.
- Reed, S. K. (1972). Pattern recognition and categorization. Cognitive Psychology, 3, 382-407.
- Reitman, J. S., & Bower, G. H. (1973). Storage and later recognition of exemplars of concepts. *Cognitive Psychology*, 4, 194-206.
- 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.
- Thomas, D. R., Freeman, F., Svinicki, J. G., Burr, D. E. S., & Lyons, J. (1970). Effects of extradimensional training on stimulus generalization. *Journal of Experimental Psychology Monographs*, 83(1, Pt. 2).

Received August 26, 1984 Revision received July 10, 1985