Feature frequency and hypothesis testing in the acquisition of rule-governed concepts

RONALD T. KELLOGG
University of Colorado, Boulder, Colorado 80302

Alternative explanations of how people learn rule-governed concepts, frequency theory and hypothesis theory, were contrasted in a feature-identification task. On each trial of a problem, subjects received multiple instances of a biconditional concept. These were arranged so that some features occurred with high frequency and others with low frequency. Subjects provided a hypothesis about the pair of relevant features and, on some trials, attempted to recall the hypothesis given on the preceding trial. At the end of each problem, subjects described the prototype, the instance that best represented all exemplars. The results were as follows: (1) High-frequency features were sampled as hypotheses more often than would be expected on the basis of random sampling, (2) prototype descriptions included more high-frequency features than would be expected by chance, and (3) previous hypotheses were accurately recalled. Whereas the last finding supports hypothesis theory, the other results confirm predictions of frequency theory. It was concluded that a satisfactory account of concept learning may need to incorporate elements of both theories.

Traditionally, there have been two answers to the question of how concepts are acquired. Associative strength or frequency theorists have argued that people compile the frequency with which features occur among positive instances of a concept. Since relevant features occur consistently and irrelevant features occur inconsistently, the defining features of a concept may be identified by accumulating frequency information (Bourne & Restle, 1959; Hull, 1920; Restle, 1955). In contrast, hypothesis theorists have contended that we search for defining features by sampling, storing, and testing hypotheses (Levine, 1969; Restle, 1962). Although several versions of hypotheses theory have been proposed, they all characterize concept learning as a search for defining features, not as an accumulation of feature frequencies.

Many studies have supported hypothesis theory as an explanation of how well-defined rule-governed concepts are learned by showing that subjects sample and test hypotheses (Levine, 1969). Features of stimuli are hypothesized either on a random basis or on the basis of perceptual salience (Millward & Spoehr, 1973; Trabasso & Bower, 1968). According to hypothesis theory, the frequency with which features occur among positive instances does not provide a basis for sampling hypoth-

eses, an assumption that was supported by the reversal-shift experiments (Bower & Trabasso, 1963; Trabasso & Bower, 1966). These experiments indicated that subjects do not compile feature frequencies, since shifting the feature that occurred most often among exemplars failed to impair concept learning.

Such evidence led Levine (1975) to conclude that the central debate between frequency and hypothesis theories is settled: Only the details of hypothesis theory still require clarification. This conclusion may be premature, however. Three recent lines of evidence have rejuvenated interest in frequency theory.

First, Bourne, Ekstrand, Lovallo, Kellogg, Hiew, and Yaroush (1976) reported that the rate of identifying the relevant features of rule-governed concepts is linked to frequency information. To illustrate, consider a population of stimuli varying on four dimensions (e.g., number, size, color, and shape), each with three values. Suppose that color and shape are the relevant dimensions, with red and square designated as the relevant features. For a biconditional concept, the type of rule used in the present study, positive instances are defined as stimuli containing either both red and square or neither of these features. If equal numbers of these two types of positive instances are presented, red and square occur in 50% of the exemplars and the irrelevant colors and shapes occur in 25% of them. The irrelevant features on the number and size dimensions occur, by chance, in 33% of the positive instances. Thus, features occur equally often on irrelevant dimensions, whereas red and square occur more often than other features on the relevant dimensions. The magnitude of the frequency differential on the relevant dimensions can be increased by presenting positive instances containing both red and square more often than exemplars containing neither of

these. Bourne et al. found that as the magnitude of this differential increased, the difficulty of learning the concept substantially decreased. This finding suggests that subjects compile how often features occur among positive instances and identify features as relevant if they occur with high frequency.

Second, frequency theory offers one account of how prototypes of fuzzy, ill-defined concepts are acquired. Unlike a rule-governed concept, the boundary of a fuzzy concept is not sharply defined by criterial features. Moreover, exemplars lying within the ill-defined boundary differ with respect to how typical they are of the concept (Rosch, 1973; Rosch & Mervis, 1975). The prototype seems to be the stimulus consisting of features that have occurred most frequently among members of the category. Other explanations of prototype formation have been proposed, but studies comparing alternative models have yielded data favoring a frequency interpretation (Chumbley, Sala, & Bourne, 1978; Goldman & Homa, 1977; Hayes-Roth & Hayes-Roth, 1977; Neumann, 1977).

Because typicality differences are obtained with ill-defined concepts, one might argue that such evidence is not damaging to hypothesis theory, which was designed to explain the acquisition of rule-governed concepts. Indeed, Rosch (1973) argued that members of rule-governed categories are equally typical. Perhaps the structures of rule-governed and ill-defined concepts are fundamentally different, with each type requiring a different learning theory. On the other hand, the two types may differ only with respect to the sharpness of the category boundary, not with respect to the typicality of instances lying within the boundary. So if subjects learn a frequency-based prototype even for rule-governed concepts, an issue not answered by the extant literature, one could not argue that frequency theory applies only to ill-defined concepts.

Third, Kellogg, Robbins, and Bourne (1978) discovered that subjects may learn a concept yet fail to store current hypotheses in short-term memory. Different versions of hypothesis theory assume different memorial abilities, but all versions posit that a current hypothesis is retained when it leads to correct classification of stimuli. Counter to this prediction, recognition of the hypothesis given on the immediately preceding trial was only about 75% correct on a binary choice test, even on postsolution trials when classification of stimuli was flawless. Although subjects were able to generate hypotheses that matched the relevant features, they did not accurately store these in short-term memory.

Such forgetting is paradoxical from the perspective of hypothesis theory, since it assumes that subjects learn the relevant features by storing and testing hypotheses. But according to frequency theory, remembering hypotheses is unnecessary. Because relevant features occur more often among positive instances than do irrelevant features, one could always generate the correct hypothesis

by compiling and using knowledge of relative frequencies, without ever storing the hypothesis in short-term memory. Given the theoretical importance of hypothesis forgetting, it is essential to explore the generality of this phenomenon.

The present study addressed the following questions: (1) Are hypotheses generated on the basis of frequency information? (2) Are frequency-based prototypes formed for rule-governed concepts? And (3) is hypothesis forgetting a widespread phenomenon? Biconditional concepts were employed in a feature-identification task; subjects knew beforehand the nature of the rule, and their task was to identify the pair of relevant features. On each trial a series of positive instances was presented, and then subjects stated their current hypothesis concerning the relevant features. On 40% of the trials, they were asked to recall the hypothesis given on the preceding trial. Finally, after multiple trials, subjects were asked to describe the prototype, the stimulus that best represented all positive instances. The exemplars were arranged so that one of the three features on each of the four dimensions occurred 50% of the time; the other two features each occurred in 25% of the instances. For some problems, the relevant features were high-frequency values, whereas for others the relevant features occurred only 25% of the time. To offset possible differences in the perceptual salience of features, the high-frequency and relevant features were counterbalanced across problems and subjects. In short, the frequency distributions of biconditional concepts were manipulated, creating a situation in which frequency theory and hypothesis theory offer contrasting predictions.

According to frequency theory, hypotheses should include more high-frequency features than would be expected on the basis of random sampling. Also, problems should be solved more often if the relevant features occur with high frequency than if they occur with low frequency. Confirmation of these predictions would constitute direct evidence that subjects generate hypotheses on the basis of their knowledge of how often features occur among positive instances. Conversely, if hypotheses include high-frequency features at a chance rate, or if problems are solved equally well regardless of how often the relevant features occur, the results would support hypothesis theory.

According to hypothesis theory, prototype descriptions, at least for unsolved problems, should include a random selection of features, because subjects presumably do not accumulate frequency information. In contrast, frequency theory predicts that these descriptions should include high-frequency features more often than would be dictated by chance for both solved and unsolved problems. Both theories contend, of course, that for solved problems the relevant features should be included in the prototype descriptions. The defining features should be part of the prototype regardless of how they are identified.

With regard to memory for hypotheses, accurate recall would support hypothesis theory, whereas poor recall would support frequency theory. Hypothesis theory predicts that subjects should always remember their previous hypothesis if it is the pair of relevant features. Frequency theory, on the other hand, permits subjects to forget their hypothesis on any trial, even once they have solved the problem. By this account, knowledge of relative frequencies allows subjects to generate the relevant features without ever storing hypotheses in short-term memory.

METHOD

Subjects

Twenty undergraduates enrolled in introductory psychology participated in the experiment to fulfill a course requirement. Subjects were tested individually. They were randomly assigned to one of four conditions, designed for counterbalancing purposes, on the basis of their order of appearance at the laboratory.

Materials

Stimuli were color slides of geometric objects varying along four dimensions, each with three values. The colors were red, yellow, and black; the sizes were large, medium, and small; the shapes were square, triangle, and hexagon; the numbers were one, two, and three objects. A Kodak slide projector displayed the stimuli on a white wall, and a Hunter timer controlled the presentation rate.

Design

All subjects received four feature-identification problems. The biconditional rule governed all problems, but the relevant features were different in each case. A problem consisted of five trials with four positive instances presented on each trial. There was one high-frequency feature, exhibited by two of the four instances per trial, on each dimension. The two low-frequency features on each dimension each appeared in one of the four instances. On two problems, the relevant features were also high frequency (50% problems). On the other two problems, the relevant features were low frequency (25% problems).

Written examples of the pictorial stimuli, presented on a single trial of a 50% and a 25% problem, are shown in Table 1.

Table 1
Stimuli Presented per Trial for 50% and 25% Problems

Number	Size	Color	Shape
		Stimulus Dimension	
		50% Problem	
one	large	yellow	triangle
two	small	red	square
one	large	yellow	hexagon
three	medium	black	triangle
		25% Problem	
one	large	yellow	triangle
two	small	red	square
one	medium	black	hexagon
three	small	black	triangle

Note—Size and color are the relevant dimensions, with large and yellow serving as the relevant features. Each of the four positive instances of the biconditional concept contains both or neither of the relevant features.

The relevant features in both cases are large and yellow. Since the problem is a biconditional, these either appear as a pair or do not appear at all in each of the positive instances. For the 50% problem, the two high-frequency features on the relevant dimensions are, of course, large and yellow, and the two high-frequency features on the irrelevant dimensions are one and triangle. For the 25% problem, the two high-frequency features on the relevant dimensions are small and black, and the two high-frequency features on the irrelevant dimensions are one and triangle. Note that both types of problems contained four high-frequency features.

A different solution was assigned to each of the four problems (two-square, small-square, one-red, and large-yellow). The two 50% problems differed in terms of the particular values that occurred with high frequency on the irrelevant dimensions; the two 25% problems differed in the same way. Since each problem had a different solution, the high-frequency features on the two relevant dimensions also differed across the two 50% and the two 25% problems. To counterbalance the order in which subjects received the various solutions and problem types, a Greco-Latin square was employed. Solution type (the four pairs of relevant features) and problem type (the two 50% and the two 25% problems) were the Greek and Latin letter variables. Problem order (the order in which the solution type/problem type combinations were presented) and problem number (first, second, third, and fourth problems) were the row and column variables. Five subjects were assigned to each of the four rows.

The important characteristics of the design were that across subjects and problems (1) each of the 12 possible dimensional values served as a high-frequency feature two or three times in both 50% and 25% problems, and (2) each of the four solutions occurred twice in both 50% and 25% problems. By collapsing data over subjects and problems, then, frequency effects were not confounded with possible differences in feature salience.

Two of the five trials on each problem included hypothesis-memory probes. Subjects were asked to recall the hypothesis that had been assigned the highest confidence rating on the previous trial (see Procedure section). The memory probe occurred after either one or four instances of the next trial had been presented. The particular trials on which the lag-one and lag-four probes occurred were varied across problems, so subjects could not anticipate when they would be asked to retrieve a previous hypothesis.

Procedure

The instructions began with an explanation of the feature-identification task. The biconditional rule was used to illustrate the problem. The subject was told to learn which 2 of the 12 possible features defined the concept according to the rule that positive instances exhibited either both or none of the relevant features. A sorting task was performed with cards, which displayed geometric stimuli, to insure that the subject understood the biconditional rule. In addition, the subject worked on one practice problem until it was solved under conditions identical to the actual experiment.

Each problem proceeded as follows. Four positive instances were presented sequentially for 5 sec each. After the instances were presented, the subject wrote on an answer sheet all pairs of features that might be relevant. Spaces were provided for eight pairs of features; however, the subject was instructed to list only pairs currently under consideration. A confidence rating was then given to each pair to identify the subject's working hypothesis, the one the subject thought was most likely to be correct. A number ranging from one to eight was written next to each pair, with eight representing maximum confidence in the correctness of the hypothesis. After responding on each trial, the subject gave the current answer sheet to the experimenter to prevent looking back at previous hypotheses. After the fifth trial, the prototype description was obtained.

Prototype descriptions were not explained during the initial instructions. Instead, this aspect of the task was introduced after the first problem had been completed. At that point, the subject was told: "Before going on to the next problem, I would like you to write down a description of a positive instance of the problem you just worked which best describes all of the positive instances you saw: that is, what set of four features, one from each dimension, represents the best example of the positive category?"

The subject was warned during the initial instructions that on certain trials memory for previous hypotheses would be probed. Typically, the four positive instances were presented without interruption. Occasionally, however, the sequence was interrupted by a blank slide, which served as a cue for the subject to recall the feature pair that had been given the highest confidence rating on the preceding trial and to record it in a box provided on all answer sheets.

RESULTS

Subjects generally gave only a single pair of features on each trial instead of multiple hypotheses. The mean proportions of multiple hypotheses for Trials 1-5 were .65, .36, .45, .29, and .14. Because secondary hypotheses were given relatively infrequently, only working hypotheses were statistically analyzed. The level of significance for all statistical tests was p < .05.

Each hypothesis was scored for the number of high-frequency features it included (0-2). The overall maximum number of high-frequency features that could have been given was 800 (2 features × 5 trials × 4 problems × 20 subjects), and the actual number given was 468. Each feature was treated as an independent observation in a test employing the normal approximation to the binomial distribution: The observed proportion of high-frequency features (.59) was significantly greater than would be expected by chance (.33) (z = 15.13). Thus, high-frequency features were hypothesized more often than would be expected on the basis of random sampling. This effect was reasonably consistent across subjects; 75% of the subjects hypothesized high-frequency features more than half of the time.

The proportion of high-frequency features was significantly greater than chance for 50% problems (.72; z = 16.56) and for 25% problems (.45; z = 4.86). The difference between the mean proportions of high-frequency features for the two types of problems was also significant [t(19) = 4.5, SE = .06]. This difference was probably due to some subjects' managing to discover the low-frequency relevant features of 25% problems. Solving 25% problems limited the number of high-frequency features that could be hypothesized, whereas solving 50% problems insured that some high-frequency features would be hypothesized. Note, though, that high-frequency features were given significantly more often than predicted by random sampling, even for the 25% problems.

According to hypothesis theory, features are randomly sampled until the relevant features are discovered. It was important, therefore, to examine the type of

features sampled only on presolution trials, those on which the relevant features either were not given or were given but then were abandoned on the next trial. The proportion of high-frequency features on presolution trials was significantly greater than chance for both 50% problems (.47; z = 4.22) and 25% problems (.52; z = 7.38). Thus, like the analysis based on all trials, the presolution analysis indicated that features were not randomly sampled.

Each hypothesis was scored for the number of relevant features it included (0-2). Based on all problems and trials, the mean proportions of relevant features for 50% and 25% problems were .53 and .31, respectively. These means differed significantly [t(19) = 3.14, SE = .07]. In addition, the proportion of solved problems, those for which the working hypothesis on Trial 5 was the relevant pair, was significantly greater for 50% problems (mean = .68) than for 25% problems (mean = .28) [t(19) = 4.0, SE = .10]. The results indicated, then, that relevant features were identified most easily if they occurred with high frequency.

The proportions of high-frequency features hypothesized for solved and unsolved problems in the 50% and 25% conditions are shown in Table 2. Unlike previous analyses, not all subjects contributed a data point to all four cells of the analysis shown in Table 2. Because of the possibility of subject-selection problems, these proportions should be interpreted cautiously. Nonetheless, the results are consistent with those based on all subjects. In only one case did subjects fail to hypothesize more high-frequency features than would be expected by chance. For solved 25% problems, the relevant features, which were consistently hypothesized, were low frequency; therefore, it is not surprising to find high-frequency features selected at a chance rate in this case. When subjects failed to solve problems of either type, they hypothesized high-frequency features about twice as often as chance. For solved 50% problems, over 9 out of 10 features hypothesized were high frequency. This inflated proportion was due in part to the fact that the relevant features were also high frequency.

Each prototype description was scored for the number of high-frequency features it included (0-4). The overall maximum number of high-frequency features that subjects could have given was 320 (4 features × 4 problems × 20 subjects), and the actual number given was 183. This proportion of high-frequency features (.57) was significantly greater than the proportion

Table 2
Proportions (P) of High-Frequency Features Hypothesized for Various Types of Problems

Solution	50% Condition		25% Condition	
	P	z	P	z
Solved	.91	19.57	.25	−1.88
Unsolved	.68	8.60	.67	12.15

expected by chance (.33) (z = 9.07). Individual differences in this case were the same as in the case of hypothesis features; 75% of the subjects gave high-frequency features more than half of the time in describing prototypes. Interestingly, the 75% of the subjects showing the frequency effect in the prototype data were not exactly the same set of individuals showing the effect in the hypothesis data. The descriptions suggested that subjects abstracted a prototype consisting of high-frequency features.

High-frequency features were listed as part of the prototype more often than would be expected by chance for both 50% problems (.66; z = 8.86) and 25% problems (.48; z = 3.98). Such features were given significantly more often for 50% problems than for 25% problems [t(19) = 3.6, SE = .05]. Also, prototype descriptions of 50% problems contained more relevant features than those of 25% problems. Each prototype description was scored for the number of relevant features it included (0-2). For each subject, the maximum number of relevant features that could have been listed for either type of problem was 4 (2 problems X 2 features). The mean proportions of relevant features included in the prototypes of 50% and 25% problems were .81 and .34, respectively; the difference was significant [t(19) = 5.88, SE = .08].

Such differences between problem types were probably a consequence of higher solution rate for 50% problems than for 25% problems. As any theory would predict, the relevant features were included in the prototype description of 92% of the solved problems. Because the relevant features were generally part of the prototype, solving 25% problems limited the number of high-frequency features that could be listed, whereas solving 50% problems insured that some high-frequency features would be listed. Again, note that even for 25% problems the prototype descriptions contained high-frequency features significantly more often than would be expected by chance.

Regardless of whether problems were solved, a frequency-based prototype seemed to be abstracted. The proportion of high-frequency features included in the prototype was significantly greater than the chance proportion (.33) for both solved (.64; z = 8.0) and unsolved (.51; z = 4.92) problems.

An analysis in terms of solved-unsolved and 50%-25% problems is shown in Table 3. The prototype descriptions presented here, like the hypothesis-selection data shown

Table 3
Proportions (P) of High-Frequency Prototype Features
for Various Types of Problems

Solution	50% Condition		25% Condition	
	P	z	P	z
Solved	.66	7.24	.45	1.76
Unsolved	.67	5.26	.49	3.69

in Table 2, may be tainted by subject-selection problems. But once again, the data reinforce the conclusions based on summary analyses. The proportions of high-frequency features were significantly greater than chance (two-tailed test) in all but one case: the solved 25% cell. The value of z for this cell reached significance only in a one-tailed test. The higher proportions observed for 50% problems probably reflect the tendency for the relevant features, which were often discovered in these problems, to be included in the prototypes.

It could be argued that subjects learned a prototype only because they knew the descriptions of a best instance would be collected. The relevant evidence is somewhat equivocal. Recall that subjects were told about the prototype task after the exemplars of the first problem had been presented; hence, the description given on the first problem should be free of expectancy or learning effects. On the one hand, the data argue strongly against the expectancy hypothesis. Averaged across different types of problems, the proportion of high-frequency features given on Problem 1 (.56) was still significantly greater than would be dictated by chance (.33) (z = 4.35). On the other hand, this effect was limited to solved 50% (.80; z = 4.47) and unsolved 25% (.65; z = 3.04) problems. For unsolved 50% (.40; z = .67) and for solved 25% (.35; z = .76), high-frequency features were included in the prototype of the first problem no more often than chance. Since these cells showed significant proportions of high-frequency features averaged across all problems, it is likely that subjects may have learned both to expect the prototype task and to give high-frequency features in these situations. In interpreting these data, though, it should be noted that only five subjects contributed to each proportion.

Subjects showed a high level of recall of previous hypotheses. The mean proportion of errors for lag-one and lag-four probes were .08 and .16; the increase across delay intervals was nonsignificant. All of the errors occurred on presolution trials. Thus, subjects rarely forgot hypotheses and never forgot them once the relevant features were discovered.

DISCUSSION

The present study yielded three important findings. First, subjects hypothesized high-frequency features nearly twice as often as would be expected on the basis of random sampling. Consequently, relevant features were identified more readily when they occurred with high frequency among positive instances of a biconditional concept than when they occurred with low frequency. Second, descriptions of the best example of the concept did not correspond to a random sampling of features; rather, for both solved and unsolved problems, subjects included high-frequency features in prototype descriptions nearly twice as often as would be expected by chance. Third, in contrast to a recently reported case

of hypothesis forgetting (Kellogg, Robbins, & Bourne, 1978), the present results showed few errors in recalling the hypothesis given on the immediately preceding trial.

The first finding supports the view that in a feature-identification task people compile the frequency with which features occur among positive instances of a rule-governed concept (Bourne et al., 1976). The feature that occurs most frequently relative to other dimensional values is most likely to be hypothesized as a relevant feature. It should be recalled, though, that the outcome of the reversal-shift experiments argued against the use of strength or frequency information (Bower & Trabasso, 1963; Trabasso & Bower, 1966). Since these studies employed a rule-governed concept in a feature-identification task, the reason for the discrepancy is unclear. It is clear, however, that feature frequencies are accumulated in learning the prototype of an ill-defined concept (Chumbley et al., 1978; Goldman & Homa, 1977; Hayes-Roth & Hayes-Roth, 1977; Kellogg, Bourne, & Ekstrand, 1978; Neumann, 1977; Rosch & Mervis, 1975). So one might expect frequency processing to operate in the acquisition of a rule-governed concept as well. In agreement with this expectation, the present study provided strong evidence that we generate hypotheses about the relevant features of rule-governed concepts by selecting high-frequency features.

The relevance of the literature on ill-defined concepts is implied by the second finding: A prototype consisting of high-frequency features seems to be abstracted for rule-governed and ill-defined concepts alike. These two types of concepts differ, of course, in that the boundaries of rule-governed categories are unambiguous, whereas those of ill-defined categories are not. Nevertheless, both types seem to be characterized by variations in the typicality of exemplars. Additional research is needed to see whether only complex rule-governed concepts, like the biconditional concepts used here, are characterized by prototypes. Some recent evidence on this point, though, suggests that a frequency-based prototype is abstracted for a simple affirmational concept, defined by a single criterial feature (Kellogg, Note 1). Hence, differences in the typicality of rule-governed exemplars may be a ubiquitous phenomenon. If so, ill-defined and rule-governed concepts must be regarded as more similar than they were previously thought to be (Rosch, 1973).

The third finding indicates that hypothesis forgetting is not found under all experimental conditions. The situation in which subjects forgot the hypothesis they had given on the immediately preceding trial differed from the present one in several ways. For instance, in the experiments reported by Kellogg, Robbins, and Bourne (1978), subjects provided their current hypothesis on only 50% of the trials; they answered memory probes about their previous hypothesis on only 12.5% of the trials; they answered memory probes about stimuli, classification responses, and feedback as well as

about hypotheses; and they classified a stimulus as a positive or negative instance on every trial. In contrast, here, all responses dealt with hypotheses: Subjects provided their current hypothesis on every trial, answered memory probes on 40% of the trials, and never classified stimuli. Thus, it may be that subjects remember hypotheses when strong emphasis is placed on them. But since the two studies differed in ways other than the emphasis placed on hypotheses (e.g., the stimulus materials and the nature of the rule governing the concepts), such a conclusion should be interpreted with caution.

Accurate recall of hypotheses is consistent with some indirect evidence that we remember hypotheses (Levine, 1969) and with the tenets of hypothesis theory. Despite the results favoring this theory (Levine, 1975), however, the assumptions of frequency theory cannot be ignored. Feature frequencies are used in generating hypotheses and in abstracting prototypes; in addition, under some circumstances, concepts are learned even though hypotheses are forgotten (Kellogg, Robbins, & Bourne, 1978). Before a new detailed theory is proposed, additional research is needed to determine whether, in learning natural concepts, relative frequencies are always used to form hypotheses and prototypes and whether hypotheses are reliably stored in short-term memory. Still, the available data clearly suggest that a satisfactory account of concept learning should combine elements of traditionally opposing viewpoints: frequency theory and hypothesis theory.

REFERENCE NOTE

1. Kellogg, R. T. *Feature frequency in concept learning: when, what, and how to count.* Manuscript submitted for publication, 1979.

REFERENCES

Bower, G. H., & Trabasso, T. Reversals prior to solution in concept identification. *Journal of Experimental Psychology,* 1963, **66**, 409-418.

Bourne, L. E., Jr., Ekstrand, B. R., Lovallo, W. R., Kellogg, R. T., Hiew, C. C., & Yaroush, R. A. Frequency analysis of attribute identification. *Journal of Experimental Psychology: General,* 1976, **105**, 294-312.

Bourne, L. E., Jr., & Restle, F. A mathematical theory of concept identification. *Psychological Review,* 1959, **66**, 278-296.

Chumbley, J. I., Sala, L. S., & Bourne, L. E., Jr. Bases of acceptability ratings in quasinaturalistic concept tasks. *Memory & Cognition,* 1978, **6**, 217-226.

Goldman, D., & Homa, D. Integrative and metric properties of abstracted information as a function of category discriminability, instance variability, and experience. *Journal of Experimental Psychology: Human Learning and Memory,* 1977, **3**, 375-385.

Hayes-Roth, B., & Hayes-Roth, F. Concept learning and the recognition and classification of exemplars. *Journal of Verbal Learning and Verbal Behavior,* 1977, **16**, 321-338.

Hull, C. L. Quantitative aspects of the evaluation of concepts: An experimental study. *Psychological Monographs,* 1920, **28** (1, Whole No. 123).

KELLOGG, R. T., BOURNE, L. E., JR., & EKSTRAND, B. R. Feature frequency and the acquisition of natural concepts. *American Journal of Psychology*, 1978, **91**, 211-222.

KELLOGG, R. T., ROBBINS, D. W., & BOURNE, L. E., JR. Memory for intratrial events in feature identification. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, **4**, 256-265.

LEVINE, M. Neo-noncontinuity theory. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 3). New York: Academic Press, 1969.

LEVINE, M. *A cognitive theory of learning*. New York: Erlbaum, 1975.

MILLWARD, R. B., & SPOEHR, K. T. The direct measurement of hypothesis-sampling strategies. *Cognitive Psychology*, 1973, **4**, 1-38.

NEUMANN, P. G. Visual prototype formation with discontinuous representation of dimensions of variability. *Memory & Cognition*, 1977, **5**, 187-197.

RESTLE, F. A theory of discrimination learning. *Psychological Review*, 1955, **62**, 11-19.

RESTLE, F. The selection of strategies in cue learning. *Psychological Review*, 1962, **69**, 329-343.

ROSCH, E. H. Natural categories. *Cognitive Psychology*, 1973, **4**, 328-350.

ROSCH, E. H., & MERVIS, C. B. Family resemblances: Studies in the structure of categories. *Cognitive Psychology*, 1975, **7**, 573-605.

TRABASSO, T., & BOWER, G. H. Presolution dimensional shifts in concept identification: A test of the sampling with replacement axiom in all-or-none models. *Journal of Mathematical Psychology*, 1966, **3**, 163-173.

TRABASSO, T., & BOWER, G. H. *Attention in learning*. New York: Wiley, 1968.