

01 Mar 1981

## Feature Frequency In Concept Learning: What Is Counted?

Ronald Thomas Kellogg  
*Missouri University of Science and Technology*

Follow this and additional works at: [https://scholarsmine.mst.edu/psysci\\_facwork](https://scholarsmine.mst.edu/psysci_facwork)



Part of the [Psychology Commons](#)

---

### Recommended Citation

Kellogg, R. T. (1981). Feature Frequency In Concept Learning: What Is Counted?. *Memory & Cognition*, 9(2), pp. 157-163. Springer; Psychonomic Society.

The definitive version is available at <https://doi.org/10.3758/BF03202330>

This Article - Journal is brought to you for free and open access by Scholars' Mine. It has been accepted for inclusion in Psychological Science Faculty Research & Creative Works by an authorized administrator of Scholars' Mine. This work is protected by U. S. Copyright Law. Unauthorized use including reproduction for redistribution requires the permission of the copyright holder. For more information, please contact [scholarsmine@mst.edu](mailto:scholarsmine@mst.edu).

# Feature frequency in concept learning: What is counted?

RONALD T. KELLOGG

*University of Missouri, Rolla, Missouri 65401*

Frequency theories of concept learning assume that people count how often features occur among instances of a concept, but different versions make various assumptions about what features they count. According to the basic feature model, only basic features are counted, whereas according to the configural model, basic features and configural features (all combinations of basic features) are counted. Two experiments assessed the predictions of both versions of frequency theory. Subjects viewed schematic human faces, which included both positive and negative instances of the concept to be learned, and then provided typicality ratings, classification responses, and frequency estimates of configural features, basic features, and whole exemplars. Because both models assume that basic features are counted, they make the same predictions in many situations. Here, the basic feature estimation and whole exemplar tests were designed such that both models make the same predictions, whereas the typicality rating, classification, and configural feature estimation tests were designed to distinguish between the models. The pattern of results clearly supported the basic feature version of frequency theory.

Many theorists have proposed feature frequency models of concept learning. These models assume that people compile the frequency with which the values or features of a stimulus dimension occur among instances of a concept. In support of frequency models, there is evidence that subjects can correctly estimate the relative frequencies of features (Kellogg, 1980a) and that they use this information in hypothesizing the relevant features of a concept by sampling features that occur with high relative frequency (Kellogg, 1980b). In addition, each instance of the concept can be described in terms of an overall frequency score, which is calculated by summing the frequencies of its constituent features. This frequency score seems to provide the basis for classifying stimuli (Bourne, Ekstrand, Lovallo, Kellogg, Hiew, & Yaroush, 1976; Hayes-Roth & Hayes-Roth, 1977) and for forming the prototype of a category, as indexed by typicality ratings (Chumbley, Sala, & Bourne, 1978; Kellogg, Bourne, & Ekstrand, 1978; Rosch & Mervis, 1975), prototype preferences (Goldman & Homa, 1977), frequency estimates of whole exemplars (Kellogg, 1980a), and recognition confidence ratings (Hayes-Roth & Hayes-Roth, 1977; Neumann, 1974, 1977; Reitman & Bower, 1973).

Granted, some evidence supports prototype-distance models, the major alternative to frequency models

This research was conducted at the Institute for the Study of Intellectual Behavior at the University of Colorado. The work was supported by a research grant and a postdoctoral fellowship from the National Science Foundation. The author thanks Lyle E. Bourne, Jr., and the reviewers for their helpful comments on earlier drafts of this article. Reprint requests should be addressed to Ronald T. Kellogg, Department of Psychology, University of Missouri, Rolla, Missouri 65401.

(Franks & Bransford, 1971; Reed, 1972). But, except for a study by Lasky and Kallio (1978), studies that have directly compared the two types have concluded that frequency models best account for the data (Chumbley et al., 1978; Goldman & Homa, 1977; Hayes-Roth & Hayes-Roth, 1977; Neumann, 1977).

The strong empirical support for frequency models should not blind us to a still unsolved problem with them, however. To date, it is uncertain what to count. Unless we know what features people count, frequency models are of limited value. Concerning this fundamental issue, there are two versions of frequency theory to consider: basic feature and configural feature models.

According to the basic model, people count the so-called basic features of a stimulus (Chumbley et al., 1978; Goldman & Homa, 1977; Kellogg et al., 1978). For instance, the basic features of a face might include the values exhibited by the face on the nose, eyes, and hair dimensions. According to a configural model (Hayes-Roth & Hayes-Roth, 1977; Reitman & Bower, 1973), the units of frequency analysis include the basic features plus all combinations of these features. In our example, these combinational or configural features include all possible combinations of values on the nose-eyes, nose-hair, eyes-hair, and nose-eyes-hair dimensional combinations. In short, the basic feature model asserts that the stimulus dimensions are independent, whereas the configural model claims that the dimensions interact.<sup>1</sup>

Whether or not stimulus dimensions are independent is an issue that cuts across several areas of research, ranging from sentence recall in humans (Anderson, 1976), for example, to classical conditioning in rabbits (Saavedra, 1975). In these areas the results have been

mixed, with some experimental situations producing results in accordance with basic models and others yielding support for configural models. In concept learning, previous research has also failed to provide a consistent answer to the question of dimensional independence.

For instance, Hayes-Roth and Hayes-Roth (1977) employed linguistic descriptions of fictitious people as stimuli in a concept learning task. After learning to classify a series of training stimuli, subjects were required to recognize old and new exemplars. The observed pattern of recognition confidence ratings was best accounted for by configural models (Hayes-Roth & Hayes-Roth, 1977, Property Set Models R1 and R2). It should be noted, however, that some basic models (R21 and R22) fit the data nearly as well. Reitman and Bower (1973), using letter strings as stimuli, also reported recognition confidence ratings that were consistent with a configural model, but since they did not evaluate a basic model, the implications of their data are uncertain.

In clear support of a basic model, Chumbley et al. (1978), who also used linguistic stimuli, noted that most subjects treated stimulus dimensions independently in concept formation, as assessed by typicality ratings. Also, Neumann (1974) reported that in many, but not all, situations a basic model can account for recognition confidence ratings. In his study, a stimulus consisted of four geometric designs, arranged in the corners of a square. When the two designs on top were separated from the designs on the bottom by a horizontal line on both training and test stimuli, subjects were sensitive to configural features. But even then, subjects counted only horizontal combinations of features, not vertical or diagonal combinations. And when the horizontal line was omitted, from either the training or the test stimuli, subjects ignored configural features altogether.

Although the evidence, on balance, seems to favor a basic feature version of frequency theory in concept learning, existing data do not decisively rule out configural models. Neumann's (1974) observations suggest that the particular nature of the stimuli may influence whether configural features are counted. Thus, if special stimuli are selected to enhance the use of configural features, then clear evidence in support of a configural model should be obtained. Human faces seem to be such a special type of stimuli, judging from evidence that configural features are important in the perception and recognition of faces (Anderson & Paulson, 1978; Carey & Diamond, 1977; Rock, 1974; Smith & Nielson, 1970). Here, two experiments were conducted using facial stimuli, in an effort to provide clear support for a configural model. If the results still favor the basic feature model, then the appropriateness of a configural model of concept learning, at least, must be seriously questioned.

As noted by Hayes-Roth and Hayes-Roth (1977), different models of concept learning often make identical predictions in the standard paradigms reported in the literature. This is particularly true of basic and configural

models, since both types assume that people at least compile the frequency of basic features. Here, five measures of concept learning were employed; the two models make different predictions for three of these, but all frequency models make the same predictions for the other two.

For all predictions, the basic logic was the same. A particular frequency distribution was established during the study trials for the basic and configural features of certain critical facial dimensions. Subjects examined faces that the experimenter labeled as positive or negative instances of the concept to be learned. Then, on the trials of five different types of tests, half of the stimuli contained a feature that occurred with high frequency among positive instances during study trials and half contained a low-frequency feature. The two types of test items were identical with respect to all other features.

On typicality, classification, and configural feature estimation of frequency tests, a configural feature distinguished the high- and low-frequency items. Thus, the configural model predicts that subjects should give higher typicality ratings, more correct classification responses, and higher configural feature estimates to high-frequency items than they give to low-frequency items. The basic model, in contrast, predicts that subjects should respond the same to high- and low-frequency items on these tests, because it assumes that configural features are not counted.

On basic feature and whole exemplar estimation of frequency tests, a basic feature distinguished the high- and low-frequency items. Because any frequency model presumes that basic features are counted, both the basic and configural models predict that subjects should give higher estimates to high-frequency items than they do to low-frequency items on both of these tests.

## EXPERIMENT 1

### Method

**Stimuli.** IDENTIKIT<sup>2</sup> faces were used as stimuli. Three basic features (a, b, and c), which, based on pilot data, were easy to tell apart, were employed on each of six dimensions, lips (L), eyebrows (B), mustache (M), nose (N), eyes (E), and hair (H). The facial perimeter and the ears were identical for all stimuli.

There were 20 different positive and 20 different negative instances of the concept. The L dimension defined the boundary of the category: Half of the positive instances exhibited La and half Lb, whereas all negative instances exhibited Lc. The B dimension was irrelevant in that the three features occurred about equally often among both positive and negative instances.

The M, N, E, and H dimensions were the critical dimensions that were used to differentiate between the basic feature and configural models. The frequency distributions of the critical basic and configural features for positive instances are shown in Table 1.

First, note that the frequencies of the basic features Ma and Na occurred more often than did those of Mb and Nb among the positive instances presented during the study phase. This manipulation set the stage for the basic feature estimation and whole face estimation tests. Second, the configural features EaHa and EbHb each occurred among 50% of the positive

**Table 1**  
Frequency Distributions of the Critical Basic and Configural Features for Positive Instances in Experiment 1

Feature	Dimension					
	M	N	E	H	MN	EH
Basic Features						
a	60	60	50	50		
b	20	20	50	50		
c	20	20	0	0		
Configural Features*						
aa					30	50
bb					0	50
ab					15	0
ba					15	0

*Note*—Frequencies are expressed as the percent of study instances that exhibited each feature. \*All other configural features occurred infrequently and about equally often, given the constraints imposed by the above frequency distributions. These are not shown here; of course, the total for the MN configural features, including those not shown, equaled 100%.

study instances, and EaHb and EbHa never occurred, setting the stage for the typicality and classification tests. Finally, the configural features MaNa occurred among 30% of the positive study instances, whereas MaNb occurred among 15%, setting the stage for the configural feature estimation test.

The basic and configural features shown in Table 1 occurred about equally often among negative instances. All other configural features occurred infrequently and about equally often among both positive and negative instances (for simplicity, these are not shown in Table 1). The predictions of the frequency models, then, were relevant only to responses given to positive test items, and these predictions were determined solely by the frequency distributions shown in Table 1.

None of the faces shown during the study phase was presented during the test phase. Some of these new test items were presented more than once across the five types of tests, in order to meet the constraints of the design.

**Design.** The typicality rating test included 12 positive instances. Half were high-frequency test items, with three exhibiting EaHa and three exhibiting EbHb. Half were low-frequency items, with three exhibiting EaHb and three exhibiting EbHa. Except for these critical dimensions, each high-frequency item had an exact duplicate among the low-frequency test items with regard to the other features. This insured a valid comparison between the ratings given to high- and low-frequency items. Two random presentation orders were employed, and half of the subjects were assigned to each order.

The classification test included eight positive and eight negative instances. Half of the positive instances were high-frequency items, with four exhibiting EaHa and four exhibiting EbHb. The others were low-frequency items, with four exhibiting EaHb and four exhibiting EbHa. Again, except for these critical dimensions, each high-frequency item was identical to a

particular low-frequency item. Frequency of test items was a pseudovariable for the negative instances, with an equal number randomly assigned to the high- and low-frequency levels. Two random presentation orders were employed, and half of the subjects were assigned to each order.

The configural feature estimation test included two positive instances: a high-frequency item (MaNa) and a low-frequency item (MaNb). The other features of these items were identical. Half of the subjects received the high-frequency item first, and half received the low-frequency item first.

The basic feature estimation test included two positive instances: a high-frequency item (Ma and Na) and a low-frequency item (Mb and Nb). The other features of these items were identical. Half of the subjects received the high-frequency item first and estimated the frequency of Ma; then these subjects received the low-frequency item and estimated the frequency of Nb. The other subjects received the low-frequency item first and estimated the frequency of Mb; then these subjects received the high-frequency item and estimated the frequency of Na.

The whole exemplar estimation test involved only a single item: either a high-frequency item (LaBaMaNaEaHa) or a low-frequency item (LbBbMbNbEbHb). The other features of these items were identical. Half of the subjects received the high-frequency item and half, the low-frequency item.

The logic of the experimental design is illustrated in Table 2. Consider first the tests that distinguish between basic and configural models, beginning with the typicality test. During the study phase, the configural feature EaHa, for example, occurred more often than EaHb among positive instances. Also, it must be noted that the frequency of the basic features Ea, Eb, Ha, and Hb were equal among the positive study instances. On the typicality test, subjects were presented test items that were identical in all respects, except for the configural features on these critical dimensions. If subjects count all configural features, then they should give higher typicality ratings to high-frequency test items, exhibiting EaHa, than they give to low-frequency test items, exhibiting EaHb. The basic model, on the other hand, predicts that subjects should give equivalent ratings to both types of positive instances, given that the study frequency of the basic features Ha and Hb were equal.

The same logic applies to the classification test. According to the configural model, subjects should make fewer errors in classifying high-frequency relative to low-frequency test items. But according to the basic model, the same number of errors should be made in classifying both types of positive instances.

The logic behind the configural feature estimation test was similar; however, different dimensions were involved. The configural feature MaNa occurred more often than MaNb among the positive study instances. On the frequency estimation test, subjects were shown either the high-frequency test item (MaNa) or the low-frequency test item (MaNb) and were asked to estimate how often that particular combination of M and N features occurred among positive instances. According to the configural model, subjects should estimate that MaNa occurred more often than MaNb. In contrast, if subjects count only basic features, then they should have no accurate knowledge of the frequency of these configural features.

Both models assume that subjects count basic features. The

**Table 2**  
Logic of the Experimental Design

Type of Test	Frequency of Test Item		Prediction of Model	
	High	Low	Configural	Basic
Typicality and Classification	EaHa	EaHb	High > Low	High = Low
Configural Feature Estimation	MaNa	MaNb	High > Low	High = Low
Basic Feature Estimation	Ma	Mb	High > Low	High > Low
Whole Exemplar Estimation	Ma	Mb	High > Low	High > Low

*Note*—Other basic and configural features were exhibited by high- and low-frequency test items (see the Design section of Experiment 1). For the sake of clarity, only the features listed above are used to describe the predictions.

basic feature Ma, for example, occurred more often than Mb among the study positive instances. Subjects were shown either the high-frequency test item (Ma) or the low-frequency test item (Mb) and were asked to estimate the frequency of that particular M feature. Any frequency model predicts that the Ma feature should receive higher estimates than the Mb feature.

On the whole exemplar estimation test, subjects were asked to estimate how often a particular face occurred among positive instances. The high-frequency test item exhibited features that had occurred more often among positive study instances (e.g., Ma) than the features exhibited by the low-frequency test item (e.g., Mb). It should be noted that the frequencies of the other features of these test items were equivalent. Subjects were expected to estimate that the high-frequency item occurred more often than the low-frequency item, a prediction that is common to both property set and basic feature models.

**Procedure.** The instructions named the facial dimensions that varied and asked subjects to attend to every dimension. Subjects were told that they would be shown both positive and negative instances of a concept. They were asked to learn what defined the concept, that is, how positive and negative instances differed. Subjects were told nothing about the test phase until after the study phase was completed.

The study phase consisted of two presentations of the negative instances, followed by two presentations of the positive instances. Thus, a series of 40 negative instances was presented, followed by a series of 40 positive instances; the experimenter identified the type of instance before each presentation series began. Each face was projected onto a white wall for 5 sec, using a Kodak slide projector and a Hunter timer.

Next, the test phase began. It consisted of the typicality rating, basic feature estimation, configural feature estimation, whole face estimation, and classification tests, in that order for all subjects.

On the typicality test, subjects rated faces according to how well they fit their concept of positive instances, by circling a number on a 5-point scale (1 = "not very well"; 5 = "very well"). On the basic feature estimation test, subjects were first told to "look at the nose dimension," for example, and then were shown a face exhibiting the basic feature to be estimated. Subjects were informed that they had been shown 40 positive instances during the study phase and 12 more on the typicality test. Their task was to estimate how often they had seen that particular feature among all positive instances shown to them, writing a number between 0 and 52. Similarly, on the configural feature test, subjects were first told to "look at the mustache and nose dimensions" and then were shown a face exhibiting the configural feature to be estimated. They wrote a number (0-52) reflecting how often they had seen that particular combination of features. On the whole face estimation test, subjects estimated how often (0-52) they had seen that particular face. Finally, on the classification test, subjects were asked to classify each test item as a positive or negative instance by circling a "P" or an "N."

On the first four tests, subjects provided their responses while the test item was projected for 10 sec. But on the classification test, each test item was projected for 3 sec and then removed during a 3-sec response interval.

**Subjects.** Twenty introductory psychology students participated as part of a course requirement. Subjects were tested in small groups ( $n = 1-5$ ) and were randomly assigned in equal numbers to the two test order conditions, which differed with respect to the order in which items were presented on each test.

## Results and Discussion

Data on all tests were collapsed across the test order variable. The .05 level of statistical significance was employed.

High-frequency (mean = 3.33) and low-frequency (mean = 3.22) items received nearly identical typicality ratings, as predicted by the basic feature model. An analysis of variance (ANOVA) confirmed that the frequency variable had no influence on typicality ratings ( $F < 1.0$ ).

Overall, subjects correctly classified a mean of 96.25% of the test items, indicating that they learned the difference between positive and negative instances. For positive items, subjects correctly classified high-frequency (mean = 97.5%) and low-frequency (mean = 96.25%) items about equally well; however, for negative items, they correctly classified fewer high-frequency items (mean = 92.5%) than low-frequency items (mean = 98.75%). The pattern for negative items was unexpected, because frequency was a pseudovisible for these items. For positive instances, though, high-frequency items were predicted by the basic feature model to be classified just as often as low-frequency items. Because performance was so close to the ceiling, however, this evidence is weak. A 2 by 2 ANOVA indicated that the main effects of frequency and instance type were nonsignificant ( $F_s \cong 1.0$ ) and that the interaction was only marginally significant [ $F(1,19) = 4.17$ ,  $MSe = 67.43$ ].

The configural feature estimates for the high-frequency item (mean = 17.35) were slightly greater than those for the low-frequency item (mean = 15.98). The configural feature exhibited by the high-frequency item actually occurred 16 times, and the one exhibited by the low-frequency item actually occurred 8 times. The small difference in the mean estimates was in the direction predicted by the configural model; however, the estimates were highly variable, and an ANOVA indicated that the difference was statistically negligible ( $F < 1.0$ ). Thus, the configural features estimates, like the typicality and classification data, were best accounted for by the basic model.

The basic feature estimates for the high-frequency item (mean = 24.0) were significantly greater than those for the low-frequency item (mean = 16.65) [ $F(1,19) = 10.59$ ,  $MSe = 51.03$ ]. The actual frequencies of the basic features exhibited by the high- and low-frequency items were 31 and 11, respectively. Thus, subjects underestimated the high-frequency feature and overestimated the low-frequency feature, but they nonetheless showed that they knew which feature had occurred most often. Such a pattern of results, characterized by relative but not absolute accuracy, is commonly obtained when subjects estimate the frequency with which a word has occurred among a series of unrelated words in a memory task (Begg, 1974) and in feature estimates following concept learning (Kellogg, 1980a). The basic feature estimates, then, confirmed the predictions of both the configural and the basic models.

The whole exemplar estimates for the high-frequency item (mean = 21.6) were much greater than those for

the low-frequency item (mean = 1.0), as predicted by any frequency model. The difference in means was significant [ $F(1,18) = 21.72$ ,  $MSe = 97.69$ ]. Both test items actually occurred only once (during the typicality test). Thus, the high-frequency item was strikingly overestimated, whereas the low-frequency item was accurately estimated.

Such a pattern of frequency estimates diverges from the usual pattern of overestimated low-frequency events and underestimated high-frequency events. But the marked overestimation of the high-frequency item, which frequency models predict to be the prototype of the positive instances of the concept, has been observed before (Kellogg, 1980a). When subjects judge the frequency with which a face has occurred among a series of related faces in a concept learning task, their estimates are related to the typicality of the face. Frequency estimates of whole exemplars reveal that people believe they have seen the prototype of a category many times more often than they actually have, just as recognition confidence ratings show that people are convinced they recognize the prototype when in fact it is a new item (Franks & Bransford, 1971; Neumann, 1974; Reitman & Bower, 1973).

In summary, the results of Experiment 1 support the basic model. The only confirmed predictions of the configural model were those that follow from any frequency model.

## EXPERIMENT 2

This experiment was designed to replicate and extend the findings of Experiment 1. Although the typicality and classification results of Experiment 1 failed to support the configural model, the configural feature estimates were in the predicted direction. One might argue that the difference in the actual frequencies of the configural features MaNa and MaNb was simply not great enough. After all, in Experiment 1, this difference was less than half of the difference in the actual frequencies of the basic features Ma and Mb, for example. Consequently, the spread between high- and low-frequency test items may have been large enough to yield a significant frequency effect on the basic feature estimation test but not on the configural feature estimation test. Before concluding that subjects counted only basic features, it is important that the spread for the configural feature test be at least as great as the spread for the basic feature test.

### Method

The materials, subject pool, procedure, test items, and test order manipulations were the same as in Experiment 1 for both the replication and extension conditions. In the extension condition, the configural feature MaNa occurred among 60% of the positive study instances, whereas MaNb never occurred. This was achieved by perfectly correlating the features of the M and N dimensions: MaNa, MbNb, and McNc were the only

combinations that appeared. Among typicality test items, MaNb occurred twice, however. Thus, the frequency spread between these features, across the 52 positive instances that had been seen prior to the beginning of the estimation tests, was 60% vs. 4%. The frequency distributions of the basic features (see Table 1) and of all other configural features were about the same as in Experiment 1. A total of 40 subjects were assigned to the conditions in equal numbers.

### Results and Discussion

In the replication condition, the high-frequency (mean = 3.53) and low-frequency (mean = 3.44) items received about the same typicality ratings. Similarly, in the extension condition, the high-frequency items (mean = 3.36) and low-frequency items (mean = 3.4) received roughly equal ratings. A 2 by 2 ANOVA showed that all four means were statistically equivalent: The frequency, condition, and Frequency by Condition effects were nonsignificant ( $F_s \cong 1.0$ ). Thus, the typicality ratings once again supported the basic model.

The overall mean percent of correct classifications was 94.69%, a high level of performance that was similar to the outcome of Experiment 1. For the replication condition, the mean percents of correct classification responses were 93.75%, 95.0%, 96.25%, and 97.5% for the positive high-frequency, positive low-frequency, negative high-frequency, and negative low-frequency test items, respectively; these figures for the extension condition were 92.5%, 91.25%, 95.0%, and 96.25%.

Neither in the replication condition nor in the extension condition was there any evidence, though, of the Frequency by Instance Type interaction predicted by the configural model. In fact, averaged across the two conditions, low-frequency positive items (mean = 93.13%) were classified just as well as high-frequency positive items (mean = 93.13%), as predicted by the basic feature model. Again, these results must be interpreted cautiously due to the possibility of a ceiling effect. Low-frequency negative items (mean = 96.88%) were classified only slightly better than high-frequency negative items (mean = 95.63%). Although this unexpected difference was in the same direction as in Experiment 1, its magnitude was much smaller. Overall, more negative items (mean = 96.25%) than positive items (mean = 93.13%) were correctly classified. A 2 by 2 ANOVA revealed a significant main effect of instance type [ $F(1,38) = 5.46$ ,  $MSe = 71.54$ ]. But the Frequency by Instance Type interaction and all other sources of variance were nonsignificant ( $F_s \cong 1.0$ ).

In the replication condition, the configural feature estimates for the high-frequency item (mean = 16.0) were slightly greater than those for the low-frequency items (mean = 15.6). The configural features exhibited by the high- and low-frequency items in this condition actually occurred 16 and 8 times, respectively. In the extension condition, on the other hand, the configural features exhibited by the high- and low-frequency items actually occurred 31 and 2 times, respectively. Yet, in

the extension condition, the estimates were about equal to those of the replication condition: High-frequency and low-frequency mean estimates were 17.35 and 15.98, respectively. A 2 by 2 ANOVA revealed that the four means were statistically equivalent; none of the sources of variance was significant.

The spread of the mean estimates, then, bore no relation to the spread of the actual frequencies. If configural features were counted, one would expect that, certainly, in the extension condition, the high-frequency item would have received higher estimates than the low-frequency item. The configural features estimates, therefore, failed to confirm a key prediction of the configural model. These results, like the typicality and classification data, supported the basic model.

In the replication condition, the basic feature estimates for the high-frequency item (mean = 22.9) were greater than those for the low-frequency item (mean = 19.35). Similarly, in the extension condition, the high-frequency item (mean = 22.55) received greater estimates than the low-frequency item (mean = 16.2). Thus, averaged across both conditions, subjects estimated that the basic feature exhibited by the high-frequency item (mean = 22.73) occurred more often than the basic feature exhibited by the low-frequency item (mean = 17.78). As in Experiment 1, they underestimated the actual frequency of the high-frequency feature, which occurred 31 times, and overestimated the low-frequency feature, which occurred 11 times. Yet, as predicted by any frequency model, subjects knew the relative frequencies of these features. A 2 by 2 ANOVA revealed a significant main effect of frequency [ $F(1,38) = 6.77$ ,  $MSe = 72.39$ ]; the other sources of variance were non-significant ( $F_s < 1.0$ ).

In the replication condition, the whole exemplar estimates given to the high-frequency item (mean = 11.9) were greater than those given to the low-frequency item (mean = 1.8). The extension condition provided the same pattern of results, with the high-frequency item (mean = 16.6) receiving greater estimates than the low-frequency item (mean = .9). Both items actually occurred only once. Thus, as in Experiment 1, subjects in general markedly overestimated the frequency of the high-frequency item (mean = 14.25), the prototype of the concept. In contrast, they accurately estimated the frequency of the low-frequency item (mean = 1.35). A 2 by 2 ANOVA yielded a significant main effect of frequency [ $F(1,36) = 23.95$ ,  $MSe = 69.49$ ], which is predicted by any frequency model. The main effect of condition and the Frequency by Condition interaction were nonsignificant ( $F_s \cong 1.0$ ).

One might argue that the large difference in whole exemplar estimates was in part due to the difference in study frequency of the MaNa vs. MbNb configural features. Because the high-frequency test item exhibited the a features on all dimensions, whereas the low-frequency item exhibited the b features, such an argument could be made.

This possibility is not well supported by the data, however. If the argument were correct, then one would expect that subjects in the extension condition, who received study faces exhibiting MaNa 60% of the time, would give greater estimates to the high-frequency item than would subjects in the replication condition, who saw MaNa in only 30% of the study faces. Although there was a trend in this direction, the Frequency by Condition interaction failed even to approach significance. Moreover, the overall average of the estimates given by subjects in Experiment 1 combined with those in the replication condition (mean = 16.75) was essentially identical to the estimates given by subjects in the extension condition (mean = 16.6) to the high-frequency item. Finally, if the configural features played a part in the frequency effect obtained on the whole exemplar test, then one would expect the typicality, classification, and configural feature estimation tests to have provided evidence supporting the configural model. Because they did not, it seems unlikely that configural features played any role in the present research.

To summarize, the results of Experiment 2 corroborated those of Experiment 1. The replication condition matched the observations of Experiment 1. Moreover, the extension condition demonstrated that the failure to obtain a significant frequency effect on the configural feature estimation test was not because of the relatively small frequency spread between the high- and low-frequency items. Rather, the null result seemed to be a consequence of subjects not counting these configural features, regardless of the size of the frequency spread.

## GENERAL DISCUSSION

In answer to the question of what stimulus features are counted in concept learning, the present experiments provide a consistent response: basic but not configural features. Typicality rating, classification, and configural feature estimation data indicated that people do not count how often configural features occur among instances of a concept. On these tests, high- and low-frequency items received identical mean scores, as predicted by the basic feature but not the configural model (see Table 2). By contrast, high-frequency items received greater mean scores than did low-frequency items on the basic feature estimation and whole exemplar estimation tests. Both versions of frequency theory predict these effects, since both assume that basic features are counted. Such a pattern of frequency effects across the five measures of concept learning implies that only basic features are counted.

Because schematic human faces were used as stimuli in the present experiments, the failure to support the assumption that people count configural features is particularly damaging to configural models of concept learning. Configural effects obtained in several experiments with facial stimuli suggest that configural features are important in the perception of faces (Anderson &

Paulson, 1978; Carey & Diamond, 1977; Rock, 1974; Smith & Nielson, 1970). Thus, the use of faces should have biased the results in favor of the configural model. Although configural features of faces may be important in some experimental situations, they had no measurable impact in the concept learning task investigated here.

It might be premature to infer that the basic feature version of frequency theory is always applicable. There may be a mix of stimulus materials, instructions, and task requirements that would yield support for the configural model. For instance, if the concept to be learned were defined by a rule involving two or more dimensions, such as an inclusive disjunction, rather than by the single-dimension rule employed here, one might succeed in forcing subjects to count configural features. In this regard, however, it should be noted that Neumann (1974) used a complex bidimensional rule and yet failed to find that subjects consistently heeded configural features.

Thus, unambiguous support for the configural version of frequency theory is obviously hard to obtain. Previous results tend to favor the basic feature model, and the present findings add weight to this conclusion.

In one sense, such a conclusion bolsters the foundation of frequency theory in general. The frequency approach to concept learning has been criticized on the grounds that it demands the use of highly complex encoding processes (Medin & Schaffer, 1978). Although the configural model is vulnerable to this criticism, because it presumes that people encode basic features and all possible configural features, the basic feature model is not. If people count only the basic features of most types of stimuli, then arguments against frequency theory based on the complexity of its encoding processes lose force.

## REFERENCES

- ANDERSON, J. R. *Language, memory, and thought*. Hillsdale, N.J.: Erlbaum, 1976.
- ANDERSON, J. R., & PAULSON, R. Interference in memory for pictorial information. *Cognitive Psychology*, 1978, 10, 178-202.
- BEGG, I. Estimation of word frequency in continuous and discrete tasks. *Journal of Experimental Psychology*, 1974, 102, 1046-1052.
- 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.
- CAREY, S., & DIAMOND, R. From piecemeal to configurational representation of faces. *Science*, 1977, 195, 312-314.
- CHUMBLEY, J. I., SALA, L. S., & BOURNE, L. E., JR. Bases of acceptability ratings in quasinnaturalistic concept tasks. *Memory & Cognition*, 1978, 6, 217-226.
- FRANKS, J. J., & BRANSFORD, J. D. Abstraction of visual patterns. *Journal of Experimental Psychology*, 1971, 90, 65-74.
- GARNER, W. R. *The processing of information and structure*. Potomac, Md: Erlbaum, 1974.
- GARNER, W. R. Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology*, 1976, 8, 98-123.
- 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, 321-328.
- KELLOGG, R. T. Simple feature frequency versus feature validity models of prototype formation. *Perceptual and Motor Skills*, 1980, 51, 295-306. (a)
- KELLOGG, R. T. Feature frequency and hypothesis testing in the acquisition of rule-governed concepts. *Memory & Cognition*, 1980, 8, 297-303. (b)
- 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.
- LASKY, R. E., & KALLIO, K. D. Transformational rules in concept learning. *Memory & Cognition*, 1978, 6, 491-495.
- MEDIN, D. L., & SCHAFFER, M. M. Context theory of classification learning. *Psychological Review*, 1978, 85, 207-238.
- NEUMANN, P. G. An attribute frequency model for the abstraction of prototypes. *Memory & Cognition*, 1974, 2, 241-248.
- NEUMANN, P. G. Visual prototype formation with discontinuous representation of dimensions of variability. *Memory & Cognition*, 1977, 5, 187-197.
- REED, S. K. Pattern recognition and categorization. *Cognitive Psychology*, 1972, 3, 382-407.
- REITMAN, J. S., & BOWER, G. H. Storage and later recognition of exemplars of concepts. *Cognitive Psychology*, 1973, 4, 194-207.
- ROCK, I. The perception of disoriented faces. *Scientific American*, 1974, 230, 78-86.
- ROSCH, E., & MERVIS, C. B. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 1975, 7, 573-605.
- SAAVEDRA, M. A. Pavlovian compound conditioning in the rabbit. *Learning and Motivation*, 1975, 6, 314-326.
- SMITH, E. E., & NIELSON, G. D. Representations and retrieval processes in short-term memory: Recognition and recall of faces. *Journal of Experimental Psychology*, 1970, 85, 397-405.

## NOTES

1. This discussion assumes that the stimulus dimensions are separable, not integral, in the sense defined by Garner (1974, 1976). If integral dimensions were employed, then it would be difficult to conceive of the values of the dimensions as basic features. I assume that the basic features of a stimulus should permit one to selectively attend to the values of each dimension. If not, the basic features might be best equated with the dimensionally inseparable points defined within the euclidean space of the integral dimensions. Combinations of these points, then, would serve as the configural features.

2. The facial features were selected from a collection produced by the IDENTIKIT Company (1230 East Warner Avenue, Santa Ana, California 92705).

(Received for publication January 9, 1980;  
revision accepted September 17, 1980.)