Journal of Experimental Psychology:
Learning, Memory, and Cognition
1984, Vol. 10, No. 1, 83-94

Role of Feedback, Category Size, and Stimulus Distortion on the Acquisition and Utilization of Ill-Defined Categories

Donald Homa and Joan Cultice
Arizona State University

The magnitude, efficiency, and scope of learning ill-defined categories without feedback training were investigated for a wide range of stimulus-distortion conditions. In Experiment 1, subjects classified either low-, medium-, high-, or mixed-level distortions under feedback or no-feedback training and then received a common transfer test containing old, new, prototype, and unrelated stimuli. In Experiment 2, the training patterns were multidimensionally scaled, and a measure of categorical structure was derived for each training condition. The results indicated that feedback training was consistently superior to no-feedback conditions and that category size was important only for the feedback conditions. Minimal learning occurred in the no-feedback conditions, except when the training set was highly structured. A simple relationship was proposed, relating ease of learning to degree of categorical structure. The benefits of feedback were discussed in terms of choice reduction and schema formation, with the latter viewed as instrumental in preventing the formation of idiosyncratic groupings based on adventitious feature similarities in patterns from other categories.

The present study was concerned with the criticality of feedback on the acquisition of ill-defined categories. Specifically, do conditions exist that preclude learning unless feedback is provided? Furthermore, if such conditions exist, can they be identified in terms of the structural properties of the categories? In the present study, performance was contrasted for a variety of situations, and the utility of a recent measure of conceptual structure (Homa, Rhoads, & Chambliss, 1979) was assessed as a predictor of category learning in the presence and in the absence of feedback.

There is no dispute that some learning may occur in the absence of feedback. Numerous experiments involving the free sorting of patterns (Aiken & Brown, 1971; Bersted, Brown, & Evans, 1969) and the execution of same–different (Brown & Dansereau, 1970; Brown & Evans, 1969; Smallwood & Arnoult, 1974) and oddity judgments (Brown, Walker, & Evans, 1968; Evans & Edmonds, 1966) have provided ample evidence that significant learning is possible without feedback. However, the amount of learning is often slight and unlikely to asymptote at high levels of performance. For example, only 21% of the subjects performed better than chance in the free sorting task used by Evans & Arnoult (1967), and learning increments of less than 10% are often reported (Aiken & Brown, 1971; Brown & Evans, 1969; Smallwood & Arnoult, 1974).

Demonstrations of learning in the absence of feedback do not, however, address the broader issues concerned with the magnitude, efficiency, and scope of this phenomena. In numerous experiments, the number of acquisition trials were insufficient to allow estimates of asymptotic levels of performance. In addition, feedback controls were often lacking, and the obviousness of categorical structure was not manipulated. For example, only 15–20 presentations were used in the oddity task by Brown et al. (1968) and Evans and Edmonds (1966), and 40–60 stimulus presentations were provided in the same–different judgment tasks of Brown and Evans (1969) and Smallwood and Arnoult (1974).[1]

Portions of Experiment 1 were presented at the Rocky Mountain Psychological Association meeting in Tucson, Arizona, May 1980.

The authors wish to thank Julie Omohundro, C. Lock, and J. R. Whitmore for their comments on an earlier draft of this manuscript.

Requests for reprints should be sent to Donald Homa, Department of Psychology, Arizona State University, Tempe, Arizona 85287.

[1] A striking exception is the study by Bersted et al. (1969). Each subject was exposed to 19 trials of 30 patterns,

Failure to include a feedback control (Aiken & Brown, 1971; Bersted et al., 1969; Brown & Dansereau, 1970; Brown & Evans, 1969; Evans & Arnoult, 1967) creates two problems. First, the efficiency of learning in the absence of feedback cannot be determined without a feedback contrast. In addition, even the small amounts of learning often reported may be misleading, because general task difficulty (and not the difficulty occasioned by the unavailability of feedback) may prohibit sizable increments of learning. Failure to manipulate the obviousness of categorical structure (Bersted et al., 1969; Evans & Arnoult, 1967; Smallwood & Arnoult, 1974), such as by varying stimulus complexity or distortion, delimits an assessment of the range of conditions that can be effectively learned in the absence of feedback.

In spite of these methodological limitations, many researchers have claimed that feedback is either unnecessary (Gibson, 1969) or even potentially harmful (Smallwood & Arnoult, 1974) to category learning. For example, Gibson and Gibson (1955) proposed that perceptual learning occurs by extracting available stimulus information, with external feedback playing little role. This view is predicated on the implicit assumption that the environment contains structure that is readily extracted by the observer: "Perceptual learning is self-regulating, in the sense that modification occurs without the necessity of external feedback" (Gibson, 1969, p. 4). For Gibson, the selection of critical categorical information is guided by task constraints and cognitive motives.

Without further elaboration, however, it is unclear how task constraints and cognitive motives would operate to guide the selection of critical information, especially if categorical structure is less than obvious. The fact that subjects, when denied feedback, form far more categories than actually exist (Bersted et al., 1969) and substantially overestimate the number of categories represented by a collection of stimuli (Hartley & Homa, 1981) suggests that there are strong limitations to the successful extraction of categorical information when feedback is unavailable.

Potential Limitations of Learning Without Feedback

When feedback is unavailable, the likelihood of successful category learning is complicated by the number of potential category assignments that must be considered. For example, assigning 12 stimuli to three categories of four members each involves 5,775 potential combinations of assignment, one of which is correct. If the number of categories represented by the 12 stimuli is unknown, but it can be assumed that each category contains an equal number of members, then the number of potential assignments to categories containing two, three, four, and six members each is 10,395, 15,400, 5,775, and 462, respectively. If the subject cannot assume the equal-membership property, then the number of potential assignments is increased even more. Finally, if task difficulty is further increased, such as by increasing the number of stimuli to be classified and/or by introducing memory demands by presenting the stimuli sequentially, then the potential benefits of feedback are clear: Feedback reduces choice and permits the observer to analyze those members that are known to belong to the same category for common categorical information.

When feedback is unavailable, the only factor that can override the number of potential categorical choices is the obviousness of categorical structure. If the structure is, in some sense, prominent, then the observer's task may be no more than "merely a process of finding clusters of points in a multidimensional space" (Evans, 1967, p. 88). The minimal amounts of learning demonstrated in previous studies suggests, however, that the task of learning categorical structure in the absence of feedback is far from trivial.

Quantification of Structure

A useful definition of structure should be applicable to a wide range of stimulus materials and capture intuitive properties of categories, such as the range of permissible distortion

spread over 10 days, where the patterns were never repeated and feedback was absent. Only four of the seven subjects demonstrated consistency in sorting the patterns into appropriate categories, and all of the subjects used six to eight categories throughout, until instructed to reduce this number. Feedback controls were not used, however, and level of stimulus distortion was not manipulated.

within a category. Ideally, this measure should also predict ease of learning in the presence and in the absence of feedback.

The measure employed in the present study was provided by Homa et al. (1979), who defined degree of categorical structure as the ratio of within-category to between-category distances in a multidimensional space. Here, distances were derived from the multidimensional scaling (Kruskal, 1964; Shepard, 1962) of category members. This measure is applicable to any material that is suitable for multidimensional scaling and has been shown to be a moderate predictor of classificatory performance (Hartley & Homa, 1981).

Experiment 1

The purpose of Experiment 1 was to evaluate systematically the effectiveness of corrective feedback on ill-defined category learning, where both the degree of category experience and exemplar experience were varied across a wide range. From a methodological viewpoint, Experiment 1 differed from previous studies in that a wider range of distortion levels was explored, feedback controls were included, and a sufficient number of learning trials were provided to obtain reasonable estimates of asymptotic performance. In addition, a variable of known importance in shaping ill-defined categories under feedback training (category size; e.g., Homa, Cross, Cornell, Goldman, & Shwartz, 1973) was evaluated for its importance under no-feedback learning. Categories were defined by three, six, or nine different stimuli, where the stimuli were either exclusively low level, medium level, high level, or a mixture of distortion levels. Half of the subjects attempted to classify these patterns in the absence of feedback, whereas the remaining subjects received corrective feedback during learning. Afterward, all subjects were transferred to a common set of stimuli, including old, new, prototype, and unrelated patterns, and the subject had the option of using a "junk" category on the transfer test. The new patterns were drawn from three levels of distortion, so that a determination of the categorical breadth (Homa & Vosburgh, 1976) could be made for the various conditions. The transfer test also provided an additional check on the amount of category learning, separate from the acquisition results.

Method

Subjects. The subjects were 192 Arizona State University undergraduates in an introductory psychology course, who participated to partially fulfill a class requirement. A total of 24 subjects were randomly assigned to eight conditions (two levels of feedback × four levels of training distortion).

Materials and apparatus. The stimuli were statistically distorted forms that belonged to five different prototype classes. Construction of these forms has been described previously (Homa, 1978). In brief, a form category is created by randomly assigning nine dots within a 50 × 50 grid and then connecting the dots with lines, forming a closed figure. This pattern is arbitrarily designated as a prototype. Members of this category are then generated by moving each of the dots according to a statistical decision rule and connecting the dots in the same order as the prototype. Patterns that are low-level distortions of the category prototype have each dot displaced, on the average, about 1.10 Euclidean units from each corresponding dot of the prototype. For patterns that are medium- and high-level distortions, each dot is displaced, on the average, about 2.90 and 4.80 units, respectively. Generally, patterns that are low-level distortions appear to be similar to each other and to the prototype, whereas high-level distortions share little obvious physical similarity to either the prototype or to each other (Homa, 1978).

All patterns were drawn in black ink by a Cal-Comp plotter, photographed onto 35-mm film, and mounted in 2 × 2 inch (5.08 × 5.08 cm) slides. During the learning and transfer phases, the stimuli were presented by a 650 Kodak Carousel projector.

Procedure. Subjects were tested in small groups of 3 to 8. They were told that a series of patterns would be shown and that their task was to determine which patterns belonged to Group A, Group B, and Group C. Subjects were further informed that they should not expect an equal number of patterns in each group.

With a single exception, the procedure in the learning and transfer phases was identical for feedback and for no-feedback subjects. In the learning phase, each pattern was shown for 8 s. For subjects receiving corrective verbal feedback, the experimenter said "respond" after 6 s, which signaled the subjects to record their response (A, B, or C) if they had not already done so. One second later, the experimenter stated the correct category response for that stimulus (e.g., "C"), paused 1 s, and then presented the next stimulus. For subjects in the no-feedback conditions, the pattern was simply shown for 2 s following the respond signal, after which the next stimulus was shown.

A total of 18 different stimuli were presented on each of eight learning trials, where the three categories were represented by 3, 6, and 9 different stimuli, respectively. All patterns assigned to the same category were generated from the same prototype. The same 18 stimuli were presented in three random orders, with 8 s separating each learning trial. Subjects recorded their answers on prepared forms; after each learning trial, they were instructed to fold their response form so that the previous responses could not be seen.

Immediately following the completion of the learning phase, the transfer instructions were given. All of the subjects were told that, as before, they would be shown a series of patterns, some of which belonged to the same groups represented in the initial phase (A, B, C) and some of which belonged to none of these groups. If they thought a pattern belonged to one of the original groups, then they should record that response; if, however, they felt the pattern belonged to none of the original groups, then they should record "none." In all, 60 of the 80 transfer patterns belonged to the original categories, and 20 belonged to the none category. Each transfer pattern was shown for 8 s, and the experimenter again provided a "record" prompt after 6 s. The 80 patterns were presented in random order with no temporal break in the total sequence. The composition of the 80 transfer patterns was such that 25% each belonged to categories A, B, C, and none. Of the 60 patterns that belonged to the original categories, 9 were old patterns (3 from each category), 45 were new (15 from each category), and 6 were the category prototypes (two copies of each of the 3 prototypes). The 45 new patterns were composed of 5 low-, 5 medium-, and 5 high-level distortions from each category. The 20 unrelated patterns were members of two random prototypes (10 each), in which 6 were low level, 6 were medium level, and 8 were high-level distortions.

The eight treatment conditions were defined by two levels of feedback (corrective feedback, no feedback) × four levels of stimulus distortion (low, medium, high, mixed). For subjects in the low-level distortion conditions, all 18 learning stimuli were low-level distortions of the category prototypes. In the medium-level and high-level conditions, all 18 patterns were medium-level or high-level distortions, respectively, of the category prototypes. In the mixed condition, the 18 patterns were equally drawn from three levels of distortion: The three-instance category was represented in learning by one low-, one medium-, and one high-level distortion; the six-instance category was represented by two low-, two medium-, and two high-level distortions, and so forth. The only difference in composition of the transfer set for the eight conditions occurred for the old patterns. For subjects who trained on low-, medium-, and high-level patterns, the 9 old patterns on the transfer set were drawn equally from the three categories at that level of distortion; for subjects who trained on a mixture of distortion levels, the old patterns were equally represented by low-, medium-, and high-level distortions.

Design. A mixed design was used, with feedback (corrective feedback, no feedback) and pattern distortion in learning (low, medium, high, mixed) functioning as between-subject variables, and category size (three, six, nine) and pattern type (old, new, prototype, unrelated) in transfer functioning as within-subject variables. The three form prototypes (P3, P4, P6) used in the learning phase were balanced across category size (three, six, nine) in a 3 × 3 Latin square, with 8 subjects assigned to each row of the resulting square. Subjects assigned to the same row of the Latin square received the same sequence of patterns on the transfer test, with the random order changing for subjects in different rows.

Scoring of data. There are a number of ways to score the learning and transfer performance in the no-feedback conditions. Objectively, each learning trial contained 18 different stimuli that belonged to three different categories. The scoring procedure finally adopted was to essentially maximize performance, working backward from Trial 8 to Trial 1. The only requirement was that each trial be maximally scored for three identifiable categories. For example, if a subject grouped the three patterns from one objective category and the six patterns from another category into one category, and the remaining patterns into a second category, the subject was given credit for 15 (out of 18) correct responses—0 (out of 3) in the three-instance category, 6 (out of 6) in the six-instance category, and 9 (out of 9) in the nine-instance category. If no discernible consistency of category labels was evident from trial to trial, the subject's score was simply maximized for each trial. A similar procedure was adopted for the transfer trials, with one exception. If the subject appeared to switch use of category labels from learning to transfer, transfer performance was scored to maximize the number correct, given that three identifiable categories were required.

Results

Learning. Errors were tabulated for each trial, as a function of feedback condition (F = feedback, NF = no feedback) and level of exemplar distortion in learning (low = low level, med = medium level, high = high level, mix = mixed level). Figure 1 shows the mean percentage of correct performance, averaged across subjects, for the four learning conditions on each of the eight learning trials. The left panel shows performance for the feedback conditions; the right panel shows the corresponding conditions when feedback was omitted.

For subjects receiving feedback, performance in all learning conditions systematically improved across trials. It is not surprising that performance in the low learning condition was superior to the performance obtained in the other conditions and that low was the only condition to reach 100% classification performance (by the fourth trial). Performance in the med and mix learning conditions were similar, with the mix condition enjoying a slight advantage on the early learning trials but reaching a terminal level of performance equal to performance in the med condition (about 93%). Performance in the high condition began at near-chance levels on Trial 1 (.405, with chance = .333), but rapidly improved, reaching a terminal level of accuracy of about 89%. In summary, learning rapidly proceeded under feedback in all four conditions, with subjects in the low condition quickly reaching errorless performance; in the remaining conditions, the magnitude of improvement across learning trials ranged from 39% to 48%.

In contrast, performance for the no-feedback conditions (right panel, Figure 1) was radically different. Substantial improvement

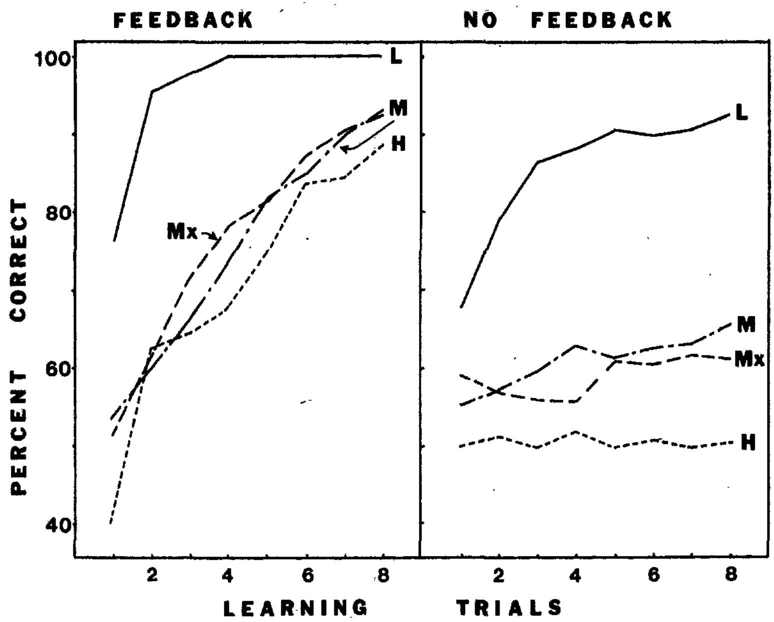

Figure 1. Mean percentage correct performance across learning trials, as a function of feedback (feedback, no feedback) and level of training distortion (L = low distortion, M = medium disortion, H = high distortion, Mx = mixed distortion).

across trials was obtained only in the low condition, with performance ranging from 68% on Trial 1 and terminating at 92% on Trial 8. For subjects in the remaining conditions, the magnitude of improvement across trials was either slight or totally absent; in the med, mix, and high conditions, the difference in classification accuracy between the eighth and initial trial was +.104, +.020, and +.005, respectively. Of these latter three conditions, only the med condition demonstrated a significant improvement across trials, $F(7, 161) = 5.86$ ($MS_e = 1.585$), $p < .05$. Even for this condition, however, it is doubtful that errorless performance is attainable, because the learning function appears to asymptote at a value considerably below 100%. The minimal improvement obtained for the mix and high conditions was not statistically significant (both *p*s > .20) and suggests that learning, in the absence of feedback, may be impossible for these conditions.[2]

An analysis of individual subject performance was also done. Table 1 shows the percentage of subjects in the two feedback conditions × the four learning conditions who (a) exhibited less than a 12% improvement across learning trials, (b) had a terminal performance less than 80% accuracy, and (c) were both below an 80% terminal level of learning and who exhibited less than a 12% improvement across learning trials. If this latter measure is taken as a rough indicator of the percentage of subjects who would never reach errorless learning, regardless of the number of trials provided, then over 90% of the subjects in the high condition and over 50% of the subjects in the med and mix conditions would fall into this category when feedback is omitted during training. In contrast, only 5% or less of all subjects receiving feedback during training would fall into this category.

In summary, the acquisition of categorical information seems highly unlikely unless feedback is provided or training occurs with minimal-level distortions (low).

Transfer performance. Mean percentage correct transfer performance for the various learning conditions is shown in Figure 2 as a

[2] Data for the NF conditions were also scored separately for each category size. The patterning of results across trials was unchanged by this analysis. In particular, learning curves were either flat (high) or showed a minimal increase across trials (med and mix). Within each NF condition, category size was a weak variable.

Table 1
Proportion of Subjects at Various Learning Criteria (Shown Separately for Feedback × Training Distortion Conditions)

	Training condition							
	Feedback				No feedback			
Learning criteria	Low	Med	High	Mix	Low	Med	High	Mix
Less than 12% improvement across trials	.250	.083	.000	.000	.250	.583	.917	.708
Less than 80% accuracy on terminal trial	.000	.167	.208	.042	.208	.833	1.000	.792
Both less than 12% improvement and below 80% on terminal trial	.000	.042	.000	.042	.125	.542	.912	.625

function of (a) category size (three, six, nine), (b) stimulus type (old, new, prototype), and (c) distortion level of the new patterns (low, medium, high level); performance under feedback training is shown in the top panel of Figure 2, and the corresponding performance under no-feedback training is shown in the bottom panels.

Overall, feedback subjects outperformed their no-feedback contrasts by amounts ranging from 6.9% in the low condition ($p < .10$) to 18.7% in the med condition, 13.4% in the high condition, and 20.8% in the mix condition (each $p < .01$). Furthermore, the advantage of feedback increased with increases in category size: For the three-, six-, and nine-instance categories, F − NF = 10.0%, 12.9%, and 22.0%, respectively ($ps < .01$). Generally, the performance of feedback subjects mirrored that obtained in previous studies (e.g., Homa & Vosburgh, 1976), in that the effect of category size was facilitative and performance systematically declined across distortion levels. For the no-feedback subjects, both variables were haphazard determinants of performance.[3]

Classification of unrelated patterns. On the transfer test, 25% of all of the patterns belonged to none of the three categories represented in the learning phase. Table 2 shows the percentage of unrelated patterns correctly identified as belonging to the none category as well as the percentage of unrelated patterns erroneously incorporated into the learned categories.

For the feedback subjects, correct classification of the unrelated patterns mirrored overall classification accuracy, with over 76% of the unrelated patterns identified as such in the low condition and only 40% in the high condition. In the high condition, a greater percentage of the unrelated patterns were erroneously classified into the six- and nine-instance categories, relative to the three-instance category, a result that had been obtained previously (e.g., Homa, et al., 1973; Homa, Sterling, & Trepel, 1981). Little bias was evident in the remaining conditions, in that near-equal percentages of unrelated patterns were incorporated into the various category sizes.

For the no-feedback subjects, the classification of unrelated patterns as members of the none category was accurately achieved only in the low condition; in the remaining conditions, 62%–83% of the unrelated patterns were erroneously incorporated into the training categories, an outcome that suggests that categorical discriminability was nearly absent. Generally, the rate of erroneous classification of unrelated patterns was much higher in the no-feedback conditions. No systematic trend across category size was obtained for the no-feedback conditions.

Discussion

The results of Experiment 1 are reasonably straightforward: With increasing levels of stimulus distortion in the learning phase, the likelihood of acquiring categorical information

[3] The complete set of analyses, including the full data matrix, for the transfer task is available on request.

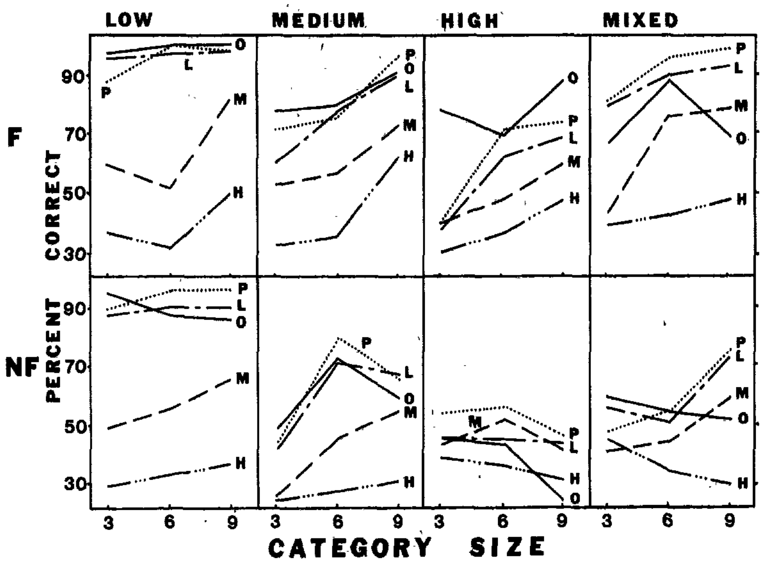

Figure 2. Mean percentage correct transfer performance, as a function of category size, stimulus type (P = prototype, O = old, L = new low distortion, M = new medium distortion, H = new high distortion), and training condition (low, medium, high, mixed). (Feedback [F] is shown in the upper panels, and no-feedback [NF] training is shown in the bottom panels.)

in the absence of feedback becomes increasingly remote. It is especially noteworthy that even moderate levels of distortion were unlearnable. The failure to observe substantial improvement across learning trials in the med, mix, and high conditions implies that some external aid must be available when even moderate levels of preexisting structure are to be learned. Without external aids, such as corrective feedback, the subject is likely to flounder at levels only slightly better than chance.

The patterning of results on the transfer test were similarly quite striking: Subjects who trained in the presence of feedback enjoyed a substantial advantage over no-feedback subjects, regardless of training condition. In addition, the effects of category size on transfer were significant for the feedback conditions, whereas little systematic benefit of category size was apparent in the no-feedback conditions. Finally, subjects in the no-feedback conditions exhibited great difficulty in discriminating category from noncategory (unrelated) stimuli, whereas feedback subjects were able to make this discrimination with reasonable accuracy.

Additional training trials would undoubtably affect transfer performance. However, given the minimal learning exhibited by the no-feedback subjects in the med, mix, and high conditions, the obtained benefits of feedback training would likely be increased by this

Table 2
Classification of Unrelated Patterns

Training distortion condition	Response assignment: Feedback 3	Feedback 6	Feedback 9	Feedback None	No feedback 3	No feedback 6	No feedback 9	No feedback None
Low	.058	.090	.090	.762	.079	.142	.085	.694
Medium	.152	.206	.188	.454	.190	.183	.235	.392
High	.104	.250	.248	.398	.350	.306	.181	.162
Mixed	.140	.179	.162	.519	.196	.279	.148	.377

manipulation. Among the feedback conditions, additional learning trials would be most advantageous to the high condition, because terminal level of learning (lowest for high) is correlated with later transfer performance.

Experiment 2

The purpose of Experiment 2 was twofold: to provide a more general measure of categorical structure than stimulus distortion and to relate an "ease-of-learning" measure with categorical structure. Categorical structure was defined here in a manner identical to that employed previously (Homa et al., 1979): $S = d_w/d_b$, where S indexes the overall degree of structure of a space containing members from multiple categories, d_w is the average within-category scaled distance of stimuli to members of the same category, and d_b is the average between-category scaled distance for stimuli belonging to different categories. The distances are obtained from a multidimensional scaling (Kruskal, 1964; Shepard, 1962) of category members. Defined in this way, S is continuous over the interval .00–1.00+, with $S = 1.00$ for a totally unstructured space, and S decreasing to .00 with increasing amounts of categorical structure.[4]

In Experiment 2, the patterns used in the low, med, mix, and high conditions were rated for similarity and subjected to a multidimensional scaling analysis. The derived measures of categorical structure, S, were then related to an ease-of-learning measure that incorporates initial and asymptotic performance for these conditions and the rate that asymptote is achieved.

Method

Subjects. The subjects were 80 Arizona State University undergraduates from an introductory psychology course. None of these subjects had participated in Experiment 1. The 80 subjects were run in four groups of 20 each, with each group receiving the stimuli appropriate to one of the four learning conditions in Experiment 1 (low, med, mix, high).

Materials and apparatus. The stimuli used in the learning phase of Experiment 1 were used in Experiment 2, the only difference being that one set of 18 stimuli from a learning condition was used rather than all three sets. Two 650 Kodak Carousel projectors were used to project the stimulus pairs onto a blank wall.

Procedure. A procedure identical to that employed previously (Homa et al., 1979) was used. In brief, a similarity scale ranging from *maximal similarity* (0) to *maximal dissimilarity* (10) was used. Subjects were encouraged to use the entire range of similarity values but were not required to do so. On a given trial, two stimuli were projected side by side for about 10 s. With 18 stimuli per condition, each subject rated 153 stimulus pairs. For subjects in the low condition, the 18 patterns were low-level distortions from three prototypes, 3 belonging to one category, 6 to a second, and 9 to a third; for subjects in the med, mix, and high conditions, the stimuli appropriate to these conditions were used. For each condition, each of the 18 stimuli appeared about equally often as the left or right member of a presented pair. At no time were subjects informed of the number of categories contained in the 18 stimuli.

Results

Measurement of categorical structure. The three-dimensional scaling solution was used for each of the four levels of training distortion because these solutions provided reasonable fits to the data (stress values = .069, .121, .106, and .141 for low, med, mix, and high, respectively). The obtained structural values ($S = d_w/d_b$) for the three-dimensional solutions were lawfully related to the level of stimulus distortion used in original learning: S = .176, .441, .535, and 0.801 for low, med, mix, and high, respectively. One noteworthy result was that the high condition, which resulted in no learning in the absence of feedback (Experiment 1), was substantially more structured than was a random configuration (S = 0.801 vs. S = 1.00).

Learning and structure. Intuitively, measures of learning performance include a number of components: (a) the initial level of performance, such as obtained on Trial 1; (b) the

---

[4] A variant of the structural ratio measure can be applied to categories having discrete features. Average within- and between-category similarity is estimated by the ratio of nonshared features within and between categories. This modified measure may explain why Medin and Schwanenflugel (1981) failed to obtain better learning for their linearly separable (LS) categories, relative to their nonlinearly separable (NLS) categories. In their Experiments 1–4, the ratio of nonshared features within category to between category was, for LS and NLS categories, 54/72 vs. 54/72, 54/68 vs. 54/72, 32/40 vs. 32/40, and 24/30 vs. 24/30, respectively. Based only on these ratios, one might expect no differences in learning between LS and NLS categories in Experiments 1, 3, and 4; an advantage favoring the NLS categories exists in Experiment 2. Precisely these outcomes occurred.

terminal (asymptotic) level of performance; and (c) the rate that learning occurs. Trial 1 performance provides an estimate of the readily available and usable structure, in the relative absence of learning. As such, Trial 1 performance should be highly correlated with S. Similarly, the level of asymptotic performance and the rate that asymptote is reached both define, in part, the ultimate difficulty of the material to be learned.

Estimates of each of these components were obtained by fitting a general exponential function across the learning trials (Experiment 1), separately for each Feedback × Training Distortion condition.[5] The resulting parameter estimates for initial level (I), asymptotic level (A), and rate (λ) for each learning condition are shown in Table 3, along with the corresponding structural ratios (S).

Each of the learning components (I, A, λ) corresponds to the structural ratios (S) in lawful ways. Still, the parameters, taken separately, are deficient in some way. For example, the asymptotic values (A) correspond to the structural ratios for the no-feedback conditions but not for the feedback conditions, where A = 1.00 in each case. Similarly, rate of learning (λ) is reasonably lawful for feedback conditions (larger values of λ correspond to smaller values of S) but not for the no-feedback conditions. In the latter case, these values are deceptively large and are produced by the fact that asymptote is reached rapidly when minimal learning occurs.

The ease-of-learning (EOL) measure proposed here covaries with degree of categorical structure and incorporates the three learning components into a single value: EOL = $\lambda(A - I)/(1 - I)$. The value $(A - I)$ estimates the amount of material that is learnable, the value $(1 - I)$ estimates the amount of material that is available to be learned, and the ratio $(A - I)/(1 - I)$ defines the percentage of available material that is, in fact, learnable given the constraints of learning. This percentage is then indexed by the learning-rate parameter (λ), and EOL becomes an estimate of how rapidly the percentage of available material can be acquired. The EOL parameters are shown in the fourth column of Table 3. Overall, the EOL parameters are larger for the feedback conditions and, for both conditions, inversely related to degree of preexisting structure (S).

Table 3
Learning Parameters and Structural Ratios for Training Distortion × Feedback Conditions

Training distortion condition	Learning parameters I	A	λ	$\lambda\frac{(A-I)}{(1-I)}$	S
No feedback					
Low	.48	.87	0.80	0.601	.176
Medium	.30	.44	1.11	0.226	.441
Mixed	.40	.42	3.00	0.105	.535
High	.25	.26	3.44	0.055	.801
Feedback					
Low	.67	1.00	1.12	1.123	.176
Medium	.22	1.00	0.25	0.254	.441
Mixed	.24	1.00	0.28	0.276	.535
High	.19	1.00	0.21	0.210	.801

Note. I = initial level; A = asymptotic level; λ = rate; S = structural ratio.

General Discussion

Two major results emerge from the present study. First, learning of categorical information, in the absence of external feedback, was strongly modulated by the degree of preexisting categorical structure; without feedback, minimal learning occurred when moderate or higher levels of distortion were used. Second, degree of categorical structure and ease of learning were found to covary.[6] That two constructs, such as S and EOL, were found to covary in a single study is unremarkable. Still, S and EOL capture some of the intuitive properties of categorical structure and learning ease. We suggest that a general relationship may exist between learning and categorical structure, independently of the exact form of the classified material. Highly structured material should be rapidly learned, whether the stimuli are random dot patterns, paintings of various artists, or even different species of

[5] A best-fitting exponential function of the form: $y = A(1 - e^{-\lambda t})$ was used, where t = trial number, and A = asymptote. Accuracy was first corrected for guessing, with g = .33. I would like to thank Peter Killeen for this analysis.

[6] Both EOL and S are derived measures. Given these, and given the fact that additional data need to be collected with different types of materials, it is premature to curve fit the relationship between EOL and S. Still, it is interesting to note that an exponential function also describes this relationship quite nicely.

storks. Similarly, ease of learning should be poor whenever the material is weakly structured. Feedback becomes critical only when the degree of categorical structure falls below some critical value. In the ideal case, estimates of initial, asymptotic, and learning-rate parameters could be obtained from a single measure—the degree of preexisting categorical structure. Obviously, this tentative proposal requires research with different kinds of materials.

The results of the present study seem inconsistent with those few studies that have purported to show that verbal corrective feedback is ineffective (Brown et al., 1968; Smallwood & Arnoult, 1974): "Extrinsic feedback in the form of correction, in conjunction with the less than perfectly reliable intrinsic information, may confuse the subject and make it difficult for him to determine accurately why he is correct or incorrect" (Smallwood & Arnoult, 1974, p. 581). The results of the present study could hardly be less in accord with this view: Regardless of the degree of preexisting structure, feedback was markedly facilitative on learning. When feedback was omitted, only the most structured set of categorical information was readily acquired, and at a rate still inferior to its feedback counterpart. A number of procedural differences may account for the discrepant conclusions:

1. Few learning trials were provided in these earlier studies. Had more study trials been used, the benefits of corrective feedback may have been obtained.

2. A slower rate of presentation was used in these studies, which may have precluded adequate schema formation. For example, 30 s elapsed between successive trials in the study by Brown et al., and 13 s separated presentations in the Smallwood and Arnoult study. Although presentation rate has not been systematically manipulated in category abstraction studies, substantial delays between successive presentations may prohibit integration mechanisms from operating, especially if category labels are forgotten from trial to trial.

3. In these studies, the subject did not classify patterns per se but provided, in some sense, a similarity judgment. Regardless, the results of the present study were unequivocal: The acquisition of categorical information is enhanced by feedback, regardless of the level of categorical structure.

What Is Learned With Feedback?

The benefits of feedback probably result from reduced choice. By knowing the category of each member, commonalities within categories and contrasts among categories can be readily identified. With moderately or weakly constrained categories, the intrinsic structure is, apparently, too diffuse to be extracted without feedback.

A related factor may operate to create an additional advantage under feedback training. When subjects are asked to describe their learning strategies and to specify what information they use to identify the various categories, they often respond that *pattern segments* initially dominate their concepts. Across trials, additional information within each pattern is used. At the conclusion of learning, subjects often sketch *overall shapes,* accompanied by statements such as "It usually points downward." Not infrequently, subjects indicate how a pattern can be stretched and shaped to conform to its category. A similar progression of learning, in which concepts are initially represented by parts and, later, by more schematic information, is contained in the extensive records of Fisher's (1916) monograph.

Therefore, the reduction of categorical choice resulting from feedback may then promote the use of increasingly greater amounts of information within each pattern. Being able to use additional pattern information is a distinct advantage in classification, because two patterns from the same category may vary considerably for a particular component, yet match reasonably well on other pattern information. In contrast, two patterns from different categories may have a similar pattern segment. If other pattern information is not considered and/or feedback is unavailable, the development of highly idiosyncratic categories seems likely. An interesting parallel was observed by Lesgold, Feltovich, Glaser, and Wang (1981): Less experienced residents often pursued erroneous paths in their diagnosis of X rays because of a prominent but misleading feature; experienced radiologists, however, were able to ignore the same feature, because the remaining information in the X ray failed to fit their schema for this disease category.

An illustration of the chaotic representation of idiosyncratic concepts is shown in Table 4, which shows trial-by-trial performance of a

Table 4
Classification Into Categories A, B, and C for Subject 10 as a Function of Trials and Objective Prototype

	Trial							
	1	2	3	4	5	6	7	8
Prototype	ABC	ABC	ABC	ABC	ABC	ABC	ABC	ABC
1 (3)	030	030	120	120	120	120	120	120
2 (6)	141	042	033	042	042	033	042	024
3 (9)	711	414	504	612	513	504	513	405

typical subject in a no-feedback condition (high distortion). Two aspects are noteworthy. First, classification is not random; some consistency is evident across trials, and patterns belonging to the same category are often grouped together. For example, this subject often grouped two of the three patterns generated from one prototype into one category, four of the six patterns from another prototype into a second category, and six of nine patterns from the remaining prototype into a third category. Second, the tendency to cross objective categorical boundaries is quite strong. On the last trial, for example, this subject sorted nine patterns into his Category C; four (of the six available) were derived from one prototype and five (of nine) belonged to a second prototype. Categories A and B similarly contained patterns from two prototypes. Across trials, there is little evidence that between-category discriminability improved.

It is not that less information is potentially available when feedback is absent. Rather, the lack of feedback, combined with less than prominent structure, is likely to result in adventitious groupings, perhaps based on pattern segments. In previous research, experience with category members has been shown to increase the degree of categorical structure, especially for those categories previously defined by the greatest number of different members (Homa et al., 1979). Similarly, increases in category size have been shown to enhance subsequent transfer to novel members (e.g., Homa, 1978) while minimizing the contribution of similarity to stored exemplars (Homa et al., 1981). We have argued elsewhere that the variable of category size promotes the formation of prototypical information. A possible explanation of the benefits of feedback, therefore, is that feedback allows more of the available pattern information to be used, and the use of this pattern information is increasingly constrained by an evolving schema or prototype.

Generality of the Present Results

Without some index of preexisting structure, it is impossible to determine whether functional and naturally occurring categories possess sufficient intrinsic structure to be learned in the absence of feedback. Some esthetic categories (Hartley & Homa, 1981) and the drawing styles of children (Hartley, Sommerville, Jensen, & Eliefja, 1982) are weakly structured and probably require feedback for learning. Possibly, many categories possess sufficient structure to permit global partitioning into superordinate or basic level categories but not finer grained subordinate ones.

Multidimensional scaling is increasingly employed by taxonomists (Sneath & Sokal, 1973) and numerical biologists (Wood, in press) as one way to display structure. If the conclusions of the present study are general, then natural categories, once indexed for their degree of structure, should follow the same learning relationship as obtained with artificial categories.

References

Aiken, L. S., & Brown, D. R. (1971). A feature utilization of the perception of pattern class structure. *Perception & Psychophysics, 9,* 279–283.

Bersted, C. T., Brown, B. R., & Evans, S. H. (1969). Free sorting with stimuli clustered in a multidimensional attribute space. *Perception & Psychophysics, 6,* 409–413.

Brown, B. R., & Dansereau, D. F. (1970). Functional equivalence between same–different classifications and judged similarity of Markov patterns. *Perception & Psychophysics, 7,* 307–310.

Brown, B. R., & Evans, S. H. (1969). Perceptual learning in pattern discrimination tasks with two and three schema categories. *Psychonomic Science, 15,* 101–103.

Brown, B. R., Walker, D. W., & Evans, S. H. (1968). Schematic concept formation as a function of constraint redundancy and knowledge of results. *Psychonomic Science, 11,* 75–76.
Evans, S. H. (1967). A brief statement of schema theory. *Psychonomic Science, 8,* 87–88.
Evans, S. H., & Arnoult, M. D. (1967). Schematic concept formation: Demonstration in a free sorting task. *Psychonomic Science, 9,* 221–222.
Evans, S. H., & Edmonds, E. M. (1966). Schema discrimination as a function of training. *Psychonomic Science, 5,* 303–304.
Fisher, S. C. (1916). The process of generalizing abstraction; and its product, the general concept. *Psychological Monographs, 21*(Whole No. 90).
Gibson, E. J. (1969). *Principles of perceptual learning and development.* New York: Appleton-Century-Crofts.
Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review, 62,* 32–41.
Hartley, J., & Homa, D. (1981). Abstraction of stylistic concepts. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 33–46.
Hartley, J. L., Somerville, S. C., Jensen, D. V. C., & Eliefja, C. C. (1982). Abstraction of individual styles from the drawings of five-year-old children. *Child Development, 53,* 1193–1214.
Homa, D. Abstraction of ill-defined form. (1978). *Journal of Experimental Psychology: Human Learning and Memory, 4,* 407–416.
Homa, D., Cross, J., Cornell, D., Goldman, D., & Shwartz, S. (1973). Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. *Journal of Experimental Psychology, 101,* 116–122.
Homa, D., Rhoads, D., & Chambliss, D. (1979). The evolution of conceptual structure. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 11–23.
Homa, D., Sterling, S., & Trepel, L. (1981). Limitations of exemplar-based generalization and the abstraction of categorical information. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 418–439.
Homa, D., & Vosburgh, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 322–330.
Kruskal, J. B. (1964). Nonmetric multidimensional scaling: A numerical method. *Psychometrika, 29,* 28–42.
Lesgold, A. M., Feltovich, P. J., Glaser, R., & Wang, Y. (1981). *The acquisition of perceptual diagnostic skill in radiology* (Report No. PDS-1). University of Pittsburgh: Learning Research and Development Center.
Medin, D. L., & Schwanenflugel, P. J. (1981). Linear separability in classification learning. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 355–368.
Shepard, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function: I. *Psychometrika, 27,* 125–140.
Smallwood, R. A., & Arnoult, M. D. (1974). A comparison of simple correction and functional feedback in schema learning. *Perception & Psychophysics, 15,* 581–585.
Sneath, P. H. A., & Sokal, R. R. (1973). *Numerical taxonomy.* San Francisco: Freeman.
Wood, D. S. (in press). A phenetic assessment of the ciconiidae (aves) using skeletal morphology. *Annals of the Carnegie Museum.*

Received November 22, 1982
Revision received July 15, 1983 ■

Roediger Appointed Editor, 1985–1990

The Publications and Communications Board of the American Psychological Association announces the appointment of Henry L. Roediger III, Purdue University, as Editor of the *Journal of Experimental Psychology: Learning, Memory, and Cognition* for a 6-year term beginning in 1985. As of February 1, 1984, manuscripts should be directed to

Henry L. Roediger III
Department of Psychology
Purdue University
West Lafayette, Indiana 47907

Manuscript submission patterns for the *Journal of Experimental Psychology: Learning, Memory, and Cognition* make the precise date of completion of the 1984 volume uncertain. Therefore, authors should note that although the current editor, Richard M. Shiffrin, will receive and consider manuscripts until January 31, 1984, should the 1984 volume be completed before that date, manuscripts may be redirected to Roediger for consideration in the 1985 volume.