

The genesis and use of exemplar vs. prototype knowledge in abstract category learning

DONALD ROBBINS

Rockefeller University, New York, New York 10021

JOHN BARRESI

Dalhousie University, Halifax, Nova Scotia, Canada

and

PHILIP COMPTON, ARTHUR FURST, MAY RUSSO, and MARGARET ANN SMITH

Emory University, Atlanta, Georgia 30322

Accurate classification of new exemplars in an abstraction paradigm may be due to their similarity to old exemplars rather than to abstract category (or prototype) knowledge. In the present study, subjects received initial training on a two-category problem before being transferred to a task in which half of the exemplar-response pairs had their responses reversed while the remaining half of the pairs were unchanged. When transfer occurred with no delay and involved old exemplars, more errors occurred for changed than for unchanged pairs. This result implies the use of exemplar-specific rather than abstract category knowledge. However, when transfer was delayed by 24 or 72 h, errors occurred equally often for changed and unchanged pairs. This result suggests that exemplar-specific knowledge is no longer used. Since subjects were still able to accurately classify exemplars prior to the transfer task at these delays, some form of abstract category knowledge is implicated.

Since naturally occurring categories such as trees, cars, handwritten letters, birds, and so on, rarely involve discrete dimensions and logically specifiable rules for inclusion, research in concept learning has shifted from the traditional concept-identification stimuli to stimuli more analogous to natural categories (Rosch, 1975). An example of this change is the procedure developed by Posner and his colleagues (Posner, 1969; Posner, Goldsmith, & Welton, 1967; Posner & Keele, 1968). Artificial categories are created by first placing nine nonoverlapping dots in a 30 by 30 implicit matrix to generate a prototype pattern. They then generate statistical distortions of these prototypes by moving each dot according to a statistical rule in approximately concentric circles around its original location. Each prototype and its distortions represent an artificial category.

A number of studies have been reported using artificial categories of this type by Posner and his colleagues as well as others (e.g., Barresi, Robbins, & Shain, 1975; Homa, Cross, Cornell, Goldman, & Schwartz, 1973). In most of these studies, subjects

receive training in discriminating distortions of a number of prototypes where the distortions of a prototype are assigned a common label. After this training, subjects are given a generalization test involving the classification of old and new distortions as well as the prototypes. The general findings are that subjects are very accurate in classifying old distortions and the prototype and fairly accurate in classifying new distortions.

The major interpretation of these results is that subjects abstract the central tendency of distortions received in training (i.e., the prototype), and that this prototype knowledge is used to "recognize" or classify instances in the transfer task (e.g., Posner, 1969; Reed, 1972). We feel that this interpretation is premature given current data. Since the classification of old distortions is generally better than that of new distortions and of the prototype, memory for these specific exemplars is indicated (e.g., Homa et al, 1973; Posner & Keele, 1968; Reitman & Bower, 1973). Given these data, it can be argued that abstraction has not occurred, but that new instances are classified accurately based on their similarity to the stored (old) exemplars (see Brooks, Note 1, for a stronger argument in favor of this position). Since the prototype is more similar to the old exemplars than are new distortions, its accuracy is also higher.

Posner and Keele (1968) attempted to provide evidence against this particular alternative by using some new distortions with the same average distance

This research was supported by a grant from the Graduate School of Arts and Sciences of Emory University. We wish to thank Ray Klein and Howard Rollins for their critical comments on an earlier version of this paper. Reprint requests should be sent to either John Barresi, Department of Psychology, Dalhousie University, Halifax, Nova Scotia, or Donald Robbins, Rockefeller University, New York 10021.

from the old distortions as the prototype. They found prototype classification higher than for the new distortions, which seemed to support the prototype abstraction view. However, a metric problem exists in this manipulation. They determined their "average distance" using a "city-block" metric, and there is reason to believe that a euclidean metric is more appropriate (Barresi et al., 1975; Reed, 1972; Barresi, Note 2). Using such a metric, the prototype would be closer to the old distortions than would the new distortions. Thus, their finding would be predicted using this metric, making the assumption of prototype learning unnecessary.

What is needed, then, is some method for discriminating behavior that is under the control of old distortions (exemplar knowledge) from behavior that is under the control of abstract category information (prototype knowledge). An excellent method exists for addressing such a question. One of the classic methods in concept-identification studies used to discriminate between "conceptual learning" and independent subproblem learning is the extradimensional vs. reversal shift comparison (e.g., Kendler & Kendler, 1962).

Most recently, Medin (1973) and Tighe (1973) have analyzed the extradimensional shift in terms of the changed and unchanged problem. In a typical situation, the subjects are presented with a simultaneous discrimination task, where the stimuli vary on two discrete dimensions, for example, shape and color. On one trial type the choice is between a red triangle and a green circle, while on the other trial type the choice is between a green triangle and a red circle. If, during original learning, shape is the relevant dimension or concept and triangle is the positive value on the shape dimension, then the green and red triangles would be the correct stimuli on the two trial types, respectively. During an extradimensional shift, where color is the relevant dimension and green, for example, is the positive value, the green triangle and green circle are the correct stimuli on the two trial types. Note that during the shift phase the subjects need only change their choice behavior on green-circle/red-triangle trials. The green triangle, which was the correct choice during original learning on green-triangle/red-circle trials, is still the correct choice during the extradimensional shift phase. In this manner, the latter trial type is termed the "unchanged" subproblem, while the former is referred to as the "changed" subproblem.

Tighe (1973) presents a considerable amount of data revealing that lower organisms tend to make few or no errors on the unchanged problem, confining almost all of their errors to the changed problem. Medin (1973), however, points out that this is a weak assessment of independent learning in that one can conclude independent learning only if no errors are made on the unchanged problem. Tighe (1973) has also shown that college students show no differential behavior on the

changed in contrast to unchanged subproblem of an extradimensional shift, interpreting this result as evidence of "conceptual" in contrast to independent learning.

Although Tighe (1973) has applied the analysis to simultaneous discriminations, it is a small step to generalize to a successive discrimination task. Using the example cited above, if shape is the relevant dimension and triangles are categorized as "1s" and circles as "2s," then the original training phase may be viewed as a short paired associate list with the following pairs: red triangle, 1; green triangle, 1; red circle, 2; and green circle, 2. During an extradimensional shift to color, where red is 1 and green is 2, the subjects need only change their response to red circle and green triangle, defining the changed subproblems.

In the research reported here, we applied the subproblem analysis to a successive discrimination task involving distortions of two different prototypes. The general method used was to train subjects on a two-choice successive categorization task with feedback. During the initial phase, subjects learned to apply the labels 1 and 2, respectively, to four distortions each of two different prototypes. After a training criterion was reached, subjects were given a half-reversal shift (Bogartz, 1965; Slamecka, 1968; Trinder, Richman, & Gulkin, 1969); that is, half of the exemplar-response combinations were reversed. Thus, two exemplars from each prototype category did not change their response (unchanged subproblems), while the remaining two exemplars from each prototype category were shifted to the response of the other category (changed subproblems).

This manipulation differs somewhat from that previously used since, in the present research, prior to the shift the stimulus property that defines the conceptual dimension is a difference in the abstract relatedness of instances. Furthermore, the half-reversal shift used here does not yield an "extra" dimension that the subject can use in a conceptual manner to categorize stimuli. Thus, we must be careful in applying the logic of previous research involving this manipulation to the present study.

Let us first suppose that during initial training subjects do not develop detailed knowledge of each exemplar but, rather, abstract some of the features that provide distinctive cues for discriminating exemplars from the two categories (Barresi et al., 1975). They then use this abstract category knowledge to assign exemplars to the two response labels. If this were the case, then, when the half-reversal shift occurred, the conceptual knowledge would not aid the subjects in determining which particular exemplars had changed their response (changed subproblems) and which had not changed their response (unchanged subproblems). At this point they would be forced to learn all specific exemplar-response combinations anew, and thus (assuming no

response bias) they would be just as likely to make errors on unchanged as on changed subproblems. Thus, the ratio of changed-to-total errors should be approximately .5, which is also taken to indicate conceptual behavior in the extradimensional shift paradigm.

By contrast, let us suppose that during initial training the subjects learned each exemplar-response combination separately without taking note of their abstract category relation. With this exemplar-specific knowledge, we would expect the subjects to produce more errors on the changed than unchanged stimulus-response pairs after the shift had occurred. Thus the proportion of changed-to-total errors should produce a ratio close to 1.0, again as we would expect from the extradimensional shift paradigm assuming independent subproblem learning.

Although the comparison of changed to unchanged errors is sensitive to whether the subjects have only abstract category or only exemplar-specific knowledge, it is relatively insensitive to the occurrence of both forms of knowledge since it averages over all transfer trials. A measure that is more sensitive to the relative priority given to category vs. exemplar knowledge is the rate of "spontaneous reversals" (Tighe, 1973) on the first transfer trial. Spontaneous reversals are defined as follows: On the first transfer trial that a specific exemplar-response pair is reversed, all of the remaining exemplar-response pairs of the trial (where a trial contains one occurrence of each exemplar-response pair) are opportunities for the subject to reverse the previously given response; each instance in which this is done is considered a spontaneous reversal. If exemplar knowledge is secondary to category knowledge, then feedback that one exemplar in a category has changed responses could produce changed responses to other exemplars in the category. However, these spontaneous reversals should not occur if exemplar knowledge is given priority over category knowledge.

EXPERIMENT 1

In this study the subproblem analyses were made on groups given a successive discrimination task followed by an immediate half-reversal shift. The subjects received training on four distortions of two different prototypes. There were actually three different groups that received training on distortions. However, the variable distinguishing the three groups, namely, the relatedness of the prototypes and their resulting distortions, had no effect during transfer, which is the primary emphasis of this paper. The acquisition results are of interest because they represent a replication of a previous result with a different design and are presented here for this purpose. Those readers interested in this aspect are referred to Barresi et al. (1975) for arguments and details concerning the relatedness variable. As a

result, the transfer data are collapsed across this variable and the three groups combined are referred to as the experimental groups.

In addition, a control group was trained on eight different unrelated random dot patterns. The transfer behavior of this group should serve as a baseline of what to expect when subjects learn problems where only exemplar-specific knowledge is available. Previous research (Bogartz, 1965; Medin, 1972; Trinder et al., 1969) suggests that some conceptual behavior may be exhibited based purely on the response equivalence during training of unrelated stimuli. Thus this condition is essential to determine the baseline rate of spontaneous reversals and ratio of changed-to-total errors that can be attributed to response equivalence rather than abstract category structure. The groups were subdivided in transfer, with half of the subjects transferred to a problem involving the training exemplars and the remaining half of the subjects transferred to new exemplars. For the abstract category conditions, this involved transfer to a reversal shift of new exemplars from the same abstract categories. This provides a second important baseline control, since behavior exhibited by these subjects cannot represent exemplar-specific knowledge, although it may represent conceptual behavior based on abstract category knowledge. Thus, any differences that occur between subjects transferred to old in contrast to new distortions must be due to the fact that subjects transferred to old distortions have exemplar-specific knowledge that has differentially affected their behavior.

Method

Materials and Design. All patterns were constructed using standard 10-square/in. (3.94-square/cm) graph paper. The original patterns were constructed by placing filled dots in a 30 by 30 cell matrix, which in turn was centered in a 50 by 50 cell matrix. The prototypes were constructed in the following manner: A pattern was formed by randomly selecting nine cells within the 30 by 30 matrix. This pattern was then distorted by a rule used by Posner et al. (1967). Following Posner et al. (1967), a template of 400 cells was constructed about each dot, with the center cell (i.e., the dot) called zero. The surrounding cells were assigned the numbers 1-399 constructed in a spiral fashion. Five areas were then defined, consisting of Position Zero (the location of the dot), Positions 1-8, 9-24, 25-99, and 100-399 for Areas 1, 2, 3, 4, and 5, respectively). Distortions were constructed by assigning a probability to any given dot of moving to any of the five areas. Within an area, all of the cells were equally likely. The probabilities were assigned such that the total uncertainty of each dot could be described in terms of bits per dot. Alternatively, the distortions could be defined in terms of the average distance between paired dots, since its logarithm is a linear function of the level of distortion (Posner et al., 1967). It should be noted that the average distance per dot is calculated by measuring the hypotenuse of a right triangle. The right triangle is formed by connecting two related dots by the city-block method. The average of the hypotenuse for all of the dot pairs for a given pair of patterns is then computed.

The reader is referred to Barresi et al. (1975) for the details of generation of related prototypes, that is, the high- and low-variance conditions. In general, the high-variance prototypes

had some dot locations that were highly similar and some that were very different. The low-variance prototypes had dot locations that were all a moderate distance from each other. The unrelated prototypes had no relationship to each other and were generated as by Posner et al. (1967).

Having generated the prototype patterns, eight 5-bit/dot distortions for each of the prototype patterns were then constructed (Posner et al., 1967). The movement probabilities were .2, .3, .4, .05, and .05 to Areas 1, 2, 3, 4, and 5, respectively, with an average distance per dot moved of 1.91 units. The 5-bit/dot distortions with their respective prototypes represented the artificial categories used during learning and transfer.

The subjects in the experimental (exemplar-trained) groups were trained on four distortions of each of the two prototypes. The prototypes from which the distortions were generated defined each of the two categories, which were randomly assigned a response number, 1 or 2. Each distortion was reproduced on 8.5 x 11 in. (21.6 x 27.9 cm) paper. A test booklet consisted of one of each of the eight distortions. The correct category response number for a pattern was on the top of the center of the next page, which also contained another pattern. Paste-on labels were used to prevent the subject from observing the pattern and response on the next page. Each subject was provided with a separate answer sheet. There were 10 different test booklets for the distortions of each specific prototype pair. The order of the distortions within a booklet was randomly determined. During transfer, 10 test booklets were constructed in a similar manner for the distortions of each specific prototype pair.

For the control group, 16 unrelated nine-dot patterns were generated, half were randomly assigned to one category while the remaining half were assigned to the other category. During training these subjects were exposed to four of each category for a total of eight patterns during training. Test booklets containing one of each of the eight patterns were constructed in a manner similar to those of the other groups. Transfer booklets were similarly constructed. For all groups, the patterns that served as training and transfer stimuli were counterbalanced across subjects.

During training there were three groups ($n = 20$ each) that received training on distortions of prototypes that were either unrelated, or had a low- or high-variance relation. The control group ($n = 20$) received training on eight unrelated patterns. All four groups were evenly subdivided in transfer in terms of new or old exemplars. During transfer half of the subjects in the unrelated, low-variance, and high-variance groups were given a half-reversal shift of the original distortions, so that half the original exemplar-response pairs were given a response reversal, while the remaining half of the original exemplar-response pairs were unchanged. The remaining subjects received a half-reversal shift on new distortions from the same prototype categories they were trained on. For the control group, half of the subjects were given a half-reversal shift, so that half the training pattern-response pairs had the responses reversed and

the other half of the pattern-response pairs were unchanged. The other half of the subjects were given eight new unrelated nine-dot patterns for a new two-choice successive discrimination problem.

Subjects. Eighty subjects from the Emory University introductory psychology subject pool partially fulfilled a course requirement by participating in a 45- to 50-min session.

Procedure. The subjects were run in small groups of up to four people. The subjects were randomly assigned to groups in order of appearance. The subject was informed that the task was to correctly assign each pattern to the categories "1" and "2." The trials were self-paced. After studying a pattern, the subject recorded the response on an answer sheet provided for this purpose and turned the page. The subject was told that the correct response was at the top of the next page. When a subject finished a booklet, it was turned in and another booklet was given. This was continued until two booklets were completed without an error (yielding a criterion of at least 16 correct responses in a row). A transfer test booklet was then given. The same procedure as in training was followed, with the exception that a maximum of 15 trials (booklets) were given during transfer.

Results and Discussion

For all statistical tests, the $p \leq .05$ level was used. The training data, shown in Table 1, reveal that the groups were ordered in terms of trials to criterion and total number of errors during training (not including the criterion trials): unrelated < high variance < low variance < control. An analysis of variance revealed a significant group effect for trials to criterion [$F(3,72) = 14.99$, $MSe = 10.99$] and total errors [$F(3,72) = 11.30$, $MSe = 111.98$]. Orthogonal partitions revealed that the control group differed significantly from the other three groups combined on both measures [$F_s(1,72) \geq 3.99$], and the high- and low-variance groups differed from each other [$F_s(1,72) \geq 5.24$]. These data replicate and extend previous findings by Barresi et al. (1975), who, using a within-subjects design, found faster learning in the high- in contrast to the low-variance condition.

The transfer data are also presented in Tables 1 and 2. The error analyses separated Trial 1 from the other transfer trials. Many subjects failed to reach criterion on the transfer task, as shown in Table 2. As a result, these subjects were assigned a score of 15 (the maximum number of transfer trials) for all relevant analyses. On the transfer task the control group took significantly fewer trials to learn the transfer task and made fewer

Table 1
Mean Trials to Criterion and Mean Total Errors for the Various Groups in Experiment 1

	Training		Transfer					
	Trials to Criterion	Total Errors	Trials to Criterion			Total Errors*		
			Old	New	Mean	Old	New	Mean
High Variance	4.2	10.6						
Low Variance	6.4	18.4	11.1	12.3	11.7	32.8	33.3	33.1
Unrelated	2.6	5.8						
Control	8.9	23.6	8.5	6.2	7.4	20.6	10.9	15.8

*Excluding Trial 1.

Table 2
Transfer Data From Experiment 1

Group	Proportion of Subjects Failing to Learn		Mean Ratio of Changed Problem Errors to Total Errors (Excluding Trial 1)		Mean Proportion of Spontaneous Reversals on Trial 1	
	Old	New	Old	New	Old	New
Experimentals	.33	.40	.67	.57	.25	.45
Control	.40	.00	.71	.52*	.13	.50**

*For the control group given eight new unrelated patterns during transfer, half of the patterns were randomly categorized as "changed problems," while the other half were categorized as "unchanged problems."

**For this condition there were eight new unrelated patterns during transfer. Since these patterns were both unrelated to each other and unrelated to the training patterns (in contrast to new exemplars for the three experimental groups), no a priori definition of a spontaneous reversal results. However, given any randomly determined category assignment, the subject's responses would yield a proportion of approximately .50.

errors than the other three groups [$F(1,72) = 6.45$, $MSe = 15.7$, for trials to criterion; $F(1,72) = 7.11$, $MSe = 227.0$, for total errors]. For both dependent variables, the Group by Transfer Stimuli (old or new) interaction was not statistically significant nor was the transfer stimuli effect ($F_s < 1$).

Table 2 also presents the ratio of changed-to-total errors, as well as the proportion of spontaneous reversals. The proportion of spontaneous reversals was calculated from the first transfer trial, as described previously. Specifically, for each subject, the proportion of responses that were the reverse of the training response after the subject first experienced reversal in transfer was calculated. The range of total possible spontaneous reversals was 4-7. In addition, the proportion of changed-to-total errors (excluding Trial 1) was calculated for each subject. No differences were found between groups on the mean ratio of changed-to-total errors, nor was the Group by Transfer Stimuli interaction significant ($F < 1$). However, lower error ratios were found overall on the new distortions in contrast to the old distortions [$F(1,72) = 21.33$, $MSe = .028$]. Further, a greater proportion of spontaneous reversals occurred for the unrelated, the high-variance, and the low-variance groups when transfer involved new rather than old stimuli [$F(1,54) = 12.46$, $MSe = .106$]. In addition, the control group showed a lower proportion of spontaneous reversals when transferred to old stimuli than did the experimental groups [$F(1,36) = 6.17$, $MSe = .12$].

Generally, these results suggest that exemplar-specific behavior was exhibited when transfer involved old stimuli. Since the ratios of changed-to-total errors did not differ among the four groups, specific knowledge of exemplars is implicated. This is further suggested by the lower ratios found when transfer involved new distortions, since a bias for responding with the original category labels is not indicated. However, the proportions of spontaneous reversal for the experimental groups transferred to old exemplars was greater than the comparable proportion for the control group. This more sensitive measure suggests that abstract category knowledge may also be available and may

initially guide at least some subjects' behavior. However, the effect is quite small and the proportion is far smaller than found when transfer involved new distortions. Thus, even the spontaneous reversal data are most congruent with the inference that the subject's behavior was predominantly under the control of exemplar-specific rather than abstract category knowledge.

EXPERIMENT 2

The results of Experiment 1 indicate that transfer involving old distortions produced behavior primarily under the control of exemplar-specific knowledge, although there was some suggestion that abstract category knowledge might also be available. However, specific knowledge of exemplars may decay rapidly with time, while abstract category or prototype knowledge may have a slower rate of decay. Previous research (Homa et al., 1973; Posner & Keele, 1970) has shown that accuracy in classifying old exemplars decays significantly with time, while classification of prototypes and new exemplars shows relatively little forgetting. In order to determine whether the changes in classification behavior are indeed due to a relative decrease in exemplar knowledge in comparison to prototype knowledge, Experiment 2 was conducted. Subjects were trained on distortions of two unrelated prototypes. They were then given a classification test containing prototypes, old and new distortions, and new patterns, either immediately or after a delay. This was followed by a half-reversal shift involving old distortions. If exemplar knowledge decays more rapidly than prototype knowledge, we would expect more subjects to show "conceptual" behavior, that is, more spontaneous reversals on transfer Trial 1 and a similarity of error rates on changed and unchanged problems, as the interval between training and transfer increases.

Method

Materials and Design. Only the patterns and booklets for the unrelated group of Experiment 1 were used. The classification test booklet administered after training consisted of two copies of each of the two prototypes, the eight old (training) distortions, eight new distortions, and four totally new nine-dot

patterns that were not related to each other or to the training stimuli. Three groups were given training identical to the unrelated group of Experiment 1. They were given the classification test and transfer trials either immediately, or after 24 or 72 h. In addition, a control group that received only the test and transfer trials was included to assess the difficulty of the task and as a gross assessment of negative transfer of training for the other three groups. All transfer trials involved "old" stimuli.

Subjects. Forty subjects from the Emory University introductory psychology pool partially fulfilled a course requirement by participating in the study. There were 10 subjects in each of four groups.

Procedure. The procedure was identical to that of Experiment 1, with the exception of the inclusion of the classification test booklet consisting of the prototypes, old and new distortions, and new patterns. For the test booklet, the control subjects (who had no prior training) were told to guess 1 or 2 for each pattern in the test booklet. No feedback was given on these trials. Subjects failing to learn during transfer were assigned a score of 15, the maximum number of precriterion trials. However, as in Experiment 1, many subjects failed to learn the transfer task within 15 trials (see Table 3). As a result, post hoc analyses were performed using a lenient criterion of 6/8 correct on any one trial.

Results and Discussion

Table 3 shows the training, test, and transfer data. No group differences were found in training or transfer in trials to criterion or mean total errors ($F_s < 1$). The results of the classification test trials given immediately prior to transfer indicated that subjects in all three of the prototype-distortion groups correctly recognized old and new distortions, as well as the prototypes, between 91% and 100% of the time. This indicates a failure to find a decrease in recognition of old distortions over the delay intervals. Furthermore,

there was little evidence of a strong response bias on these test trials, since new random patterns were categorized as a "1" response (in contrast to a "2") approximately 50% of the time.

During transfer, the immediate test group had a significantly lower mean proportion of spontaneous reversals than the 24- or 72-h delay groups combined [$F(1,27) = 6.01$, $MSe = .033$], which did not differ from each other ($F < 1$). Comparison of the ratio of changed-to-total errors revealed a significant overall difference [$F(3,36) = 3.68$, $MSe = 1.93$], while subsequent orthogonal analyses revealed a significantly higher ratio for the immediate group compared to the 24-h, 72-h, and control groups combined [$F(1,36) = 9.17$], with no differences between the other three groups ($F_s \leq 1.27$). Both of these results indicate that transfer behavior is affected by exemplar-specific knowledge only when transfer is immediate. This suggests that exemplar-specific information has decayed prior to the delayed transfer task.

Based on the stringent criterion, no differences were found between the four groups on trials to criterion or total errors ($F_s < 1$). However, since 18 of the 40 subjects failed to achieve this stringent criterion, a post hoc analysis was performed using the lenient criterion that only one subject failed to achieve. These analyses comparing all four groups revealed a significant difference in trials to criterion [$F(3,36) = 4.46$, $MSe = 9.6$] and in total errors [$F(3,36) = 3.52$, $MSe = 9.91$]. Orthogonal analyses on both variables revealed no differences between the immediate and control groups or between the 24- and 72-h groups. However, the latter two groups combined, in contrast to

Table 3
Training, Test, and Transfer Data From Experiment 3

Criterion†	Experimental Group				
	Immediate Test	24-h Test	72-h Test	Control Group	
Training					
Mean Trials to Criterion	4.9	4.8	4.9		
Mean Total Errors	10.1	9.3	8.9		
Test					
Proportion Correct—Prototype	1.00	.97	1.00	.48	
Proportion Correct—Old Distortions	.98	.97	.95	.50*	
Proportion Correct—New Distortions	.95	.95	.91		
Random Patterns**	.55	.50	.48	.53	
Transfer					
Number of Subjects Failing to Learn	Lenient	0	0	1	0
	Stringent	3	5	7	3
Mean Trials to Criterion	Lenient	6.9	9.7	11.5	7.7
	Stringent	12.3	12.0	12.3	11.5
Mean Total Errors (Excluding Trial 1)	Lenient	28.5	35.9	42.8	24.8
	Stringent	36.5	40.9	43.3	31.7
Mean Ratio of Changed Problem Errors to Total Errors (Excluding Trial 1)	Lenient	.72	.60	.52	.56**
	Stringent	.68	.59	.55	.56**
Mean Proportion of Spontaneous Reversals on Trial 1		.30	.55	.52	

*Old and new distortions were combined since the distinction is irrelevant for this group. **Proportion of "1" responses.

†Values are based on the stringent criterion of two successive trials without error (16/16) or the lenient criterion of 6/8 correct.

the former two groups combined, showed significantly more trials to criterion [$F(1,36) = 11.34$] and total errors [$F(1,36) = 8.88$]. This evidence of negative transfer in the delayed conditions, combined with the apparent decay of exemplar-specific knowledge, suggests that the negative transfer effect is due to interference based on abstract category knowledge that may have developed during the interval between initial training and transfer.

GENERAL DISCUSSION

The present study attempted to determine the genesis and use of exemplar and abstract category knowledge for artificial categories generated by distorting prototype patterns. Previous research using a postlearning classification task, while seeming to provide evidence of the use of abstract category knowledge, could just as well be interpreted in terms of exemplar knowledge. Thus the experimental separation of their possible effects motivated the present transfer task.

Based on the previous research with the subproblem analysis, we have assumed that, if subjects have specific knowledge of the exemplars and if they treat the transfer task in a nonconceptual manner, they should show a high ratio of changed-to-total errors and a low rate of spontaneous reversals. Further, if subjects have abstract category knowledge rather than specific exemplar knowledge, they should show a ratio of changed-to-total errors of about .5 and also show a high rate of spontaneous reversals. The combined results of the two experiments indicated that both exemplar and abstract category knowledge may be available at the time of transfer. When transfer was immediate and involved old exemplars, strong evidence was found to indicate that behavior was under the control of exemplar knowledge. The delayed transfer data indicate that this exemplar-specific information was no longer being used and was probably no longer available. Subjects who received delayed transfer performed poorly on the transfer task. However, they did correctly classify category instances (old and new) before the transfer task and also classified the prototype almost without error. Since reliance on a conceptual solution would retard learning the transfer task, this suggests that delayed classification during the test trials and transfer may be based on a more abstract form of knowledge. The long interval before transfer may have led to the development of relatively impenetrable concepts of the two categories, which inhibited learning during transfer because these abstract representations must again be broken down to instance knowledge.

Taken together, the two experiments indicate that immediately after training, behavior is partly conceptual but primarily exemplar specific. Since no strong

evidence of conceptual behavior based on abstract category knowledge occurs in the immediate transfer task, it seems highly likely that classification behavior tested immediately after training is accomplished through the similarity of instances to old exemplars. The fact that delayed transfer data no longer exhibit exemplar-specific behavior indicates the rapid decay of the use of exemplar knowledge. The present study and previous research (Homa et al., 1973; Posner & Keele, 1970) indicate that classification behavior at long delays is still high so that abstract category knowledge is implicated. Although available research does not indicate exactly when this abstract knowledge evolves, there is no strong evidence that conceptual behavior based on abstract knowledge occurs immediately, that is, during training. As a result it is possible that abstract category knowledge comes about primarily through the decay of detailed knowledge of specific exemplars. Furthermore, this interpretation is most congruent with the delayed transfer results of the second experiment, which show no evidence of exemplar-specific knowledge, while at the same time providing evidence of the development of relatively impenetrable abstract concepts.

A number of philosophers and psychologists (e.g., Cassierer, 1953; Goodman, 1965; Nelson, 1974) have argued against the notion that abstract concepts evolve through a process of forgetting differences among particulars. Rather, they argue for a purposeful process of concept formation through the selection of some rule of behavior equivalence. Our data suggest the possibility of both forms of concept formation. For unrelated stimuli, such as in the control condition in the first experiment, conceptual behavior may be based purely on response equivalence. However, when natural categories (or in our case "artificial natural categories") are involved, where a set of stimuli are highly related to each other and unrelated to other stimuli, a concept-formation process based on forgetting differences and retaining similarities is probably involved.

REFERENCE NOTES

1. Brooks, L. *Nonanalytic concept formation and memory for instances*. Paper presented at the S.S.R.C. Conference on Human Categorization, Berkeley, 1976.
2. Barresi, J. *The modification of perceptual space in the discrimination of abstract categories*. Unpublished manuscript, 1977.

REFERENCES

- BARRESI, J., ROBBINS, D., & SHAIN, K. Role of distinctive features in the abstraction of related concepts. *Journal of Experimental Psychology: Human Learning and Memory*, 1975, **104**, 360-368.
- BOGARTZ, W. Effects of reversal and nonreversal shifts with CVC stimuli. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 484-488.
- CASSIERER, B. *Substance and function*. New York: Dover 1953.

- GOODMAN, N. *Fact, fiction and forecast*. Indianapolis: Bobbs-Merrill, 1965.
- HOMA, D., CROSS, J., CORNELL, D., GOLDMAN, D., & SCHWARTZ, S. Prototype abstraction and classification of new instances as a function of number of instances defining the prototype. *Journal of Experimental Psychology*, 1973, **101**, 116-122.
- KENDLER, H. H., & KENDLER, T. S. Vertical and horizontal processes in problem solving. *Psychological Review*, 1962, **69**, 1-16.
- MEDIN, D. L. Response factors in verbal learning and transfer. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 234-238.
- MEDIN, D. L. Subproblem analysis of discrimination shift learning. *Behavior Research Methods & Instrumentation*, 1973, **5**, 332-336.
- NELSON, K. Concept, word, and sentence: Interrelations in acquisition and development. *Psychological Review*, 1974, **81**, 267-285.
- POSNER, M. I. Abstraction and the process of recognition. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 3). New York: Academic Press, 1969.
- POSNER, M. I., GOLDSMITH, R., & WELTON, R. E., JR. Perceived distance and the classification of distorted patterns. *Journal of Experimental Psychology*, 1967, **73**, 28-38.
- POSNER, M. I., & KEELE, S. W. On the genesis of abstract ideas. *Journal of Experimental Psychology*, 1968, **77**, 353-363.
- POSNER, M. I., & KEELE, S. W. Retention of abstract ideas. *Journal of Experimental Psychology*, 1970, **83**, 304-308.
- REED, S. K. Pattern recognition and categorization. *Cognitive Psychology*, 1972, **3**, 382-407.
- REITMAN, J. S., & BOWER, G. H. Storage and later recognition of examples of concepts. *Cognitive Psychology*, 1973, **4**, 194-206.
- ROSCH, E. Universals and cultural specifics in human categorization. In R. Brislin, S. Bochner, & W. Lonner (Eds.), *Cross-cultural perspectives on learning*. New York: Sage/Halsted Press, 1975.
- SLAMECKA, N. A methodological analysis of shift paradigms in human discrimination learning. *Psychological Bulletin*, 1968, **69**, 423-438.
- TIGHE, T. Subproblem analysis of discrimination learning. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 7). New York: Academic Press, 1973.
- TRINDER, J., RICHMAN, C. L., & GULKIN, R. Effects of a Slamecka shift on reversal and half-reversed shift tasks. *Proceedings, 77th Annual Convention, American Psychological Association*, 1969.

(Received for publication November 1, 1977;
accepted March 18, 1978.)