

# The abstraction and long-term retention of ill-defined categories by children

DONALD HOMA and JAMES LITTLE  
*Arizona State University, Tempe, Arizona*

The acquisition, utilization, and long-term retention of ill-defined categories was assessed in 10.8-year-old children. Children initially classified 18 distorted form stimuli into three categories containing 3, 6, and 9 members, and then received transfer tests immediately following classification, one week later, and one month later. On the transfer tests, old, new, prototype, and unrelated forms were presented, and the subjects had available an optional junk category. The results showed that classification of new instances at all levels of distortion was unaffected by a month's delay, with classification of the prototype and new instances strongly facilitated by increases in category size. The acquired breadth of these concepts was stable across the month's delay. A comparison with adult performance in a highly similar task revealed no qualitative differences.

We investigated the acquisition, utilization, and long-term retention of ill-defined form categories in young children. Most studies on children's concept learning have utilized familiar or well-defined categories, with emphasis either on the development of formal operations in cognition (e.g., Farnham-Diggory & Gregg, 1975; Markman, 1978) or on contrasting performance of children with performance of adults (e.g., Lasky, 1974). Furthermore, when conceptual knowledge is assessed, investigators usually have employed an immediate or brief retention interval. In contrast, we investigated the acquisition of ill-defined form categories, and the stability of these concepts was evaluated for time periods up to one month following learning. The forms used in the present study are considered ill-defined (Neisser, 1967), because the patterns belonging to a category are infinitely variable, the dimensions underlying the category are obscure, and a simple rule cannot be used to classify the patterns; rather, these patterns are statistical distortions of a category prototype. For Neisser, most naturally occurring categories are ill-defined.

The present study mirrored the procedure used in adult category abstraction tasks, in that both learning and transfer phases were used. With the exception of a slight relaxation of the learning criterion, little attempt was made to tailor the task for children; transfer occurred only following near-errorless acquisition, and pattern distortions were used that spanned a wide range, including high-level distortions that result in poor performance by adults (Homa, 1978). There was reason to believe that children should

perform capably in a category abstraction task. First, children are able to detect slight changes in related form patterns (Gaines, 1973). Second, children, like adults, employ multiple features in the free sorting of variable forms (Aiken & Williams, 1975).

Four major questions were of interest in the present study: (1) Are the concepts of children enhanced by manipulations of category size in the learning phase? (2) Are these concepts preserved after lengthy retention intervals (1 month), or do these concepts undergo rapid deterioration after initial learning? (3) Do children tend to form overly-narrow concepts (Saltz & Sigel, 1967), and is the breadth of the acquired concepts (Homa & Vosburgh, 1976) affected by time delays? (4) Are children more instance-bound than adults (Tighe, Tighe, & Schechter, 1975)? The rationale for varying category size (3, 6, 9) in the learning phase was based on findings with adult subjects: Adults typically show improved transfer to novel patterns belonging to categories defined by more instances in the learning phase (Homa, Cross, Cornell, Goldman, & Shwartz, 1973; Homa, 1984).

## METHOD

### Subjects

The subjects were 12 fifth-grade pupils from Fuller Elementary School in Tempe, Arizona. The mean age of the subjects was 10.8 years. Each subject took part in three sessions. The first session lasted about one hour; the latter two sessions involved transfer tests only and lasted about 20 min each. These two transfer sessions occurred one week and one month after criterion learning. One subject who failed to reach learning criterion was replaced.

### Materials and Apparatus

The stimuli were nine-sided forms, the construction of which has been described previously (Homa, 1978). Briefly, a prototype form is created by the random placement of nine dots in a 50 × 50 grid, and the dots are then connected with straight lines in an arbitrary order, forming a closed figure. Distorted exemplars of the prototype are determined by displacing each of the dots according to a statistical decision rule, and

We would like to thank Julie Omohundro and Ward Mirz for their helpful comments. We also would like to express our appreciation to Doris Cornett, the school principal at Fuller Elementary, and Robert Curry of the Tempe Research Committee. Requests for reprints should be sent to D. Homa, Department of Psychology, Arizona State University, Tempe, AZ 85287.

the points are connected in the same order as the prototype. The magnitude of the distortion level for a given pattern is determined by the average Euclidean distance moved per dot from the corresponding points in the prototype. Patterns defined as low-level distortions of the prototype are patterns whose points are displaced, on the average, a small distance (1.20 units/dot), and which, consequently, share a good deal of physical similarity to other low-level distortions and to the prototype. The mean dot movement for medium-level (2.80 units/dot) and high-level (4.80 units/dot) distortions is considerably greater, with the high-level distortions having little obvious physical similarity to either the prototype or other high-level distortions.

Patterns were drawn by a Cal-Comp plotter and affixed to 6 × 9-in. cards. Each stimulus was centered within a 6 × 6-in. square, with most patterns having a vertical and horizontal extent of about 3 in. All sessions were run in a small (10 × 15 ft), sound-proofed room at the school. During the experiment, the subject and experimenter faced each other across a table, which had a small divider located in the middle. The divider prevented the subject's viewing of the score sheets and stimuli. The experimenter hand-held the stimuli over the divider.

**Procedure**

At the initial session, a category was informally defined as a group of things that belonged together because they shared something in common. Both verbal and visual examples of categories were given. First, the subject was told that poodles, Great Danes, and German shepherds belonged to the dog category rather than the cat category. Then, the subject was shown drawings of different kinds of triangles, circles, and lines, and the subject was encouraged to determine which shapes went together. If the subject appeared to fully appreciate what was meant by a category, the experiment proper was begun.

The subject was told that a series of patterns would be shown which belonged to the A, B, or C category, and that the task was to determine which patterns belonged to which category. The subject was further informed that each category need not contain an equal number of patterns. On the initial trial, each of the learning patterns was briefly presented, and the experimenter verbally stated the correct category label for each pattern. On all subsequent presentations, the subject was required to provide the label, guessing if necessary, and the experimenter provided corrective feedback after each response ("no, that is a B pattern"). Learning was terminated once the subject reached a criterion of two or fewer errors on a trial. Each trial contained 18 different patterns, 3 of which belonged to one category, 6 to a second category, and 9 to a third category; the patterns classified together were generated from the same prototype. These same 18 patterns were presented in three different random orders, and learning trials were rotated across these random orders until criterion was reached. Each category was represented by a mixture of distortion levels; the 3-instance category contained 1 low-level, 1 medium-level, and 1 high-level pattern; the 6-instance category contained 2 low-level, 2 medium-level, and 2 high-level distortions; etc.

After reaching learning criterion, an immediate transfer test was given. The subjects were told that they would be shown a number of patterns, some of which belonged to the categories they had just learned (A, B, or C), and some of which belonged to none of these categories ("junk"). If they thought a pattern belonged to a learned category, then they should indicate which one; if, however, they thought the pattern belonged to

none of the learned categories, then they should assign that pattern to the junk category. The subjects were told also that they would not be told if their choices were correct, and that, unlike the learning trials, each category would contain approximately the same number of patterns.

The transfer test contained 76 different stimuli, 19 in each of the three learned categories, and 19 in the junk category. For each learned category, the 19 patterns included 3 old, 15 new, and the category prototype. The 3 old transfer patterns for each category contained 1 low-level, 1 medium-level, and 1 high-level distortion; the 15 new patterns similarly contained an equal proportion of low-, medium-, and high-level distortions. The 19 unrelated patterns (junk) were high-level distortions from three unrelated prototype categories.

The 76 transfer stimuli were presented in random order, with groups of three subjects receiving the same random order. After the immediate test, the subject was dismissed and brought back to the test room about one week and one month later. At these later sessions, the subject was reminded of the earlier task and informed that another transfer test would be given as before. The only procedural difference on the delayed transfer tests was that the 76 transfer stimuli were presented in a random order different from before.

**Design**

A within-subjects design was used, with number of patterns per category during learning (3, 6, 9), distortion level (low-, medium-, high-level), stimulus type (old, new, prototype, unrelated), and time of test (immediate, 1 week, 1 month) functioning as major variables. A Greco-Latin square was used to assign the three prototypes (P2, P3, P4) and the three category names (A, B, C) to each category size (3, 6, 9). With 12 subjects, it was possible to assign 4 subjects to each of the three rows of the resulting square. As a consequence, each of the three prototypes was used equally often to represent the categories containing 3, 6, and 9 instances, and each category size was called an A, B, or C category equally often.

**RESULTS**

**Learning**

The mean number of trials to criterion was 9.50, with an individual subject range of 6-15 trials. The mean percent correct classification rate on the last trial for the 3-, 6-, and 9-instance categories was 91.7%, 87.5%, and 93.5%, respectively, *p* > .20. Thus, categories differing in size were learned to a similar criterion prior to transfer.

**Transfer**

The mean classification accuracy for each stimulus type (old, new, prototype, unrelated) at each category size (3, 6, 9) is shown in Table 1, as a function of retention interval (immediate, one week, one month). In addition,

**Table 1**  
Mean Classification Accuracy for Each Stimulus Type  
as a Function of Category Size and Time of Transfer Test

Stimuli	Immediate				One Week				One Month				Sorted into Junk		
	Category Size				Category Size				Category Size				Immed	1 Wk	1 Mo
	3	6	9	Mean	3	6	9	Mean	3	6	9	Mean			
Old	.667	.667	.861	.731	.472	.639	.722	.611	.667	.417	.722	.602	.037	.074	.074
New-Low	.517	.717	.867	.700	.483	.733	.917	.711	.433	.650	.850	.644	.106	.067	.067
New-Medium	.233	.400	.700	.444	.283	.350	.700	.444	.367	.433	.733	.511	.250	.261	.156
New-High	.300	.400	.333	.344	.167	.350	.467	.328	.233	.400	.467	.367	.372	.344	.278
Prototype	.583	.917	.999	.833	.333	.750	.917	.667	.250	.667	.917	.611	.056	.083	.083
Mean	.460	.620	.752	.611	.348	.564	.745	.552	.390	.513	.738	.547	.164	.166	.132

performance on new patterns at each level of distortion (low-, medium-, high-level) is shown.

Overall, classification accuracy was enhanced by about 35% by increases in category size and decreased by about 34% by increases in distortion level. An analysis of variance on only the new patterns revealed that the effect of category size,  $F(2,22) = 14.00$ , and distortion level,  $F(2,22) = 36.93$ , were both highly significant, both  $ps < .001$ . The effect of delayed testing, which resulted in a slight enhancement of 1.1%, was not significant,  $F < 1$ . The only other significant source of variance was the Category Size  $\times$  Distortion Level interaction,  $F(4,44) = 2.97$ ,  $p < .05$ . This interaction was due to the fact that low- and medium-level distortions were enhanced by a greater amount with increases in category size than were the high-level distortions (40% vs. 19% facilitation).

A second analysis was performed with the variables of stimulus type (old, new, prototype), category size (3, 6, 9), and time of test (immediate, one week, one month) as main variables. As before, category size was a significant source of variance,  $p < .001$ , and time of test was not,  $p > .20$ . The effect of stimulus type was highly significant,  $F(2,22) = 14.42$ ,  $p < .001$ , and reflected the fact that the prototype stimuli (70.4%) were classified better than either old (64.8%) or new (49.9%) stimuli. The Category Size  $\times$  Stimulus Type interaction was significant,  $F(4,44) = 5.01$ ,  $p < .01$ , and was due to the fact that new and prototype stimuli were enhanced more by increases in category size than were the old stimuli (34% and 56% vs. 17%). Finally, both the Stimulus Type  $\times$  Delay and Stimulus Type  $\times$  Category Size interactions were significant,  $F(4, 44) = 3.41$  and  $F(8, 88) = 2.25$ , respectively, both  $ps < .05$ . These interactions were produced by sizable retention losses across the month's delay for the old (-13%) and prototype (-22%) stimuli, relative to the new stimuli (+1%). The decrement across the month's delay for the prototype stimuli was confined largely to the 3-instance category (-33%), with minimal loss for the 9-instance category (-8%).

In summary, increases in category size during learning was strongly facilitative on subsequent transfer, especially for the new and prototype patterns. The effect of a month's delay was insignificant for the new patterns and for the prototype belonging to the 9-instance category. The effect of the delay was deleterious for the old instances and for the prototype belonging to the 3-instance category.

### Classification of Unrelated Patterns

On the transfer test, 25% of all patterns were statistically unrelated to the learning categories. Table 2 shows the likelihood that unrelated patterns were correctly identified as junk patterns and erroneously classified into the learning categories. Overall, 42% of the unrelated patterns were correctly classified into the junk category, with the 9-instance category receiving a greater percentage of the unrelated patterns (24.1%) than the 3- (18.0%) or 6- (16.1%) instance categories. These tendencies were un-

**Table 2**  
Classification Rate of Unrelated Forms into the 3, 6, 9, and Junk Categories, as a Function of Time of Transfer Test.

Transfer	Category Size			Junk
	3	6	9	
Immediate	.180	.188	.238	.395
Week Delay	.184	.149	.219	.447
Month Delay	.175	.145	.268	.412
Mean	.180	.161	.241	.418

affected by the delayed tests. To assess whether subjects had a bias to classify more novel stimuli into the larger categories, the number of hits and false alarms were determined for each subject as a function of category size and time of test. Of the resulting nine correlations, only two approached significance,  $r = +.52$  (3-instance category, week delay) and  $r = +.51$  (9-instance category, month delay), both  $.05 < p < .10$ . None of the remaining seven correlations approached significance. Thus, subjects who were good classifiers on the transfer tests were as likely to erroneously sort few unrelated patterns into the learned categories as were the poor classifiers, an outcome contrary to a simple bias effect.

## DISCUSSION

The results demonstrated that 10- to 11-year-old children readily acquire ill-defined categorical information, and that this information is relatively stable across a month's retention interval. Old stimuli and the prototype for the 3-instance category were most likely to be forgotten. Neither the new stimuli, nor the prototype for the 9-instance category suffered any decrement in performance across a month's delay. The variable of category size, which is crucial to the shaping of concepts in adults, was also a potent factor for children, enhancing classification of new and prototype stimuli by 30-50%. Classification of the category prototype for the 9-instance category was superior to any other type of category information, exceeding 90% even after a month's delay.

### Comparison with Adult Performance

The results of a similar study using adult subjects (Homa & Cultice, 1984, mixed condition) suggest that children perform somewhat worse than adults, but qualitative differences are absent. In the adult study, only an immediate test was used, and a somewhat different learning procedure was used. However, the learning and transfer manipulations exactly matched those used in the present study, and the terminal level of acquisition prior to transfer was virtually identical in the two studies.

Overall, adults classified the novel transfer stimuli somewhat better (a 15% advantage). However, the patterning of results was quite similar: Both children and adults showed an orderly progression of transfer performance as a function of stimulus distortion, and both groups exhibited enhanced performance with increases in category size. In fact, the children benefited more by increases in category size (35%) than did the adult subjects (18%). If category size can be interpreted as a major experiential component of ill-defined category learning, then children may be more sensitive to this manipulation than adults are.

### Breadth of Ill-Defined Concepts in Children and Adults

Accuracy of classification was markedly diminished by increasing degrees of stimulus distortion. This gradient of classification was somewhat sharper for the children than for adults in an identical condition (Homa & Cultice, 1984), suggesting that children formed concepts having a narrower breadth than the adult subjects did. However, performance

by children in the present study did not drop precipitously with increases in distortion level, in the sense that most low-level stimuli were classified accurately and most moderate- and high-level distortions were not. Because no evidence of an interaction between stimulus distortion and delay of test was found, the acquired breadth of the concepts formed by children was relatively stable across the month's delay.

#### Instance versus Categorical Representation

An index of instance-memory can be gauged by performance on the old stimuli, because these stimuli can be classified correctly by recognition mechanisms rather than by category generalization. In the present study, classification of the old stimuli decreased across the month delay (73% immediately following, 60% one month later). In spite of this drop in performance on the old patterns, no reduction in performance on the new patterns was obtained. If children were more instance-bound than adults (Tighe et al., 1975), then it might be expected that the old information would remain relatively stable across the various delayed tests, and that the classification of new and prototypical information would progressively degenerate. In fact, the opposite result was obtained.

#### Conclusion

The results of the present study provide little evidence of a qualitative shift in ill-defined category learning and subsequent generalization abilities between children of age 10 and college-age students. Perhaps the most remarkable result was the stability of these concepts across a month's delay. An obvious issue is whether the results of the present study are unique to ill-defined categories, which place a premium on abstraction abilities and minimize rule-learning strategies. The ability to integrate information that is discriminably different but conceptually based may represent a cognitive skill that is prior to and more fundamental than the learning of concepts based on well-defined rules.

#### REFERENCES

- AIKEN, L. S., & WILLIAMS, T. M. (1975). Development of multiple dimension use in form classification. *Child Development*, **46**, 123-132.
- FARNHAM-DIGGORY, S., & GREGG, L. W. (1975). Color, form, and function as dimensions of natural classification: Developmental changes in eye movements, reaction time, and response strategies. *Child Development*, **46**, 101-114.
- GAINES, R. (1973). Matrices and pattern detection by young children. *Developmental Psychology*, **9**, 143-150.
- HOMA, D. (1978). Abstraction of ill-defined form. *Journal of Experimental Psychology: Human Learning & Memory*, **4**, 407-416.
- HOMA, D. (1984). On the nature of categories. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 18). New York: Academic Press.
- 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., & CULTICE, J. (1984). Role of feedback, category size, and stimulus distortion on the acquisition and utilization of ill-defined categories. *Journal of Experimental Psychology: Learning, Memory & Cognition*, **10**, 83-94.
- HOMA, D., & VOSBURGH, R. (1976). Category breadth and the abstraction of prototypical information. *Journal of Experimental Psychology: Human Learning & Memory*, **3**, 322-330.
- LASKY, R. E. (1974). The ability of six-year-olds, eight-year-olds, and adults to abstract visual patterns. *Child Development*, **45**, 626-632.
- MARKMAN, E. M. (1978). Empirical versus logical solutions to part-whole comparison problems concerning classes and collections. *Child Development*, **49**, 168-177.
- NEISSER, U. (1967). *Cognitive psychology*. New York: Appleton-Century-Crofts.
- SALTZ, E., & SIGEL, I. E. (1967). Concept overdiscrimination in children. *Journal of Experimental Psychology*, **73**, 1-8.
- TIGHE, T. J., TIGHE, L. S., & SCHECHTER, J. (1975). Memory for instances and categories in children and adults. *Journal of Experimental Child Psychology*, **20**, 22-37.

(Manuscript received for publication March 18, 1985.)