

The multifaceted nature of unsupervised category learning

BRADLEY C. LOVE

University of Texas, Austin, Texas

A substantial portion of category-learning research has focused on one learning mode—namely, classification learning (a supervised learning mode). Subsequently, theories of category learning have focused on how the abstract structure of categories (i.e., the co-occurrence patterns of feature values) affects acquisition. Recent work in supervised learning has shown that a learner's interactions with the stimulus set also plays an important role in acquisition. The present study extends this work to unsupervised learning situations involving simple one-dimensional stimuli. The results suggest that categorization performance is a function of both learning mode (i.e., study conditions) and learning problem (i.e., category structure). Unsupervised learning, like supervised learning, appears to be multifaceted, with different learning modes best paired with certain learning problems.

A great deal of category-learning research focuses on how the structure of categories affects classification learning (Estes, 1994; Heit, 1992; Kruschke, 1992; Medin & Schaffer, 1978; Nosofsky, 1988). For example, Shepard, Hovland, and Jenkins (1961) compared the relative difficulty levels of six different assignments of eight stimuli to two contrasting categories. The focus of such studies is on characterizing the goodness of the mapping from inputs (i.e., the representation of a stimulus) to outputs (i.e., the category membership of a stimulus). Researchers vary the mapping of stimuli to categories in order to evaluate theories of learning.

Similarly, modeling and quantitative efforts focus on predicting the difficulty level associated with acquiring different category structures. For example, category utility (Corter & Gluck, 1992) explicitly measures the goodness of a category partition (i.e., the assignment of stimuli to categories), using an information theoretic measure. The higher the goodness of the mapping, the easier it should be for humans to acquire the category structure.

Recently, researchers have demonstrated that how a learner interacts with the stimulus set also plays a major role in determining how a category is acquired (Barsalou, 1991; Goldstone, 1996; Medin, Lynch, Coley, & Atran, 1997; Ross, 1997). This line of research examines variables other than the structure of the stimulus set. A related research program investigates how different learning modes interact with different category structures (Love, Markman, & Yamauchi, 2000; Yamauchi, Love, & Markman, 2002; Yamauchi & Markman, 1998).

For example, Yamauchi et al. (2002) compared classification and inference learning¹ and found that inference learning is better suited to linear category structures, whereas classification learning is a better match for non-linear category structures. Such studies do not focus simply on the role of category structure but, instead, explore which processes and aspects of the stimuli are stressed by different learning modes. This work is critical in developing more general theories of category learning. Focusing on any single induction task (e.g., classification learning) provides only one window into understanding categorization (Love, 2001; Schank, Collins, & Hunter, 1986).

Interactions observed between learning mode and category structure have largely involved complex stimuli (i.e., multidimensional stimuli) and supervised learning modes (i.e., subjects received corrective feedback on their category assignments). These factors may have made it more likely to observe the interactions. First, complex stimuli have multiple dimensions, which allow for selective attention, thus increasing the likelihood that different learning modes will stress different aspects of the stimulus set. Second, supervised learning is often characterized as an active learning process, whereas unsupervised learning is viewed as undirected, stimulus-driven, incremental accrual of information (Berry & Dienes, 1993; Cleermans, 1993; Hayes & Broadbent, 1988; Lewicki, 1986).² On this view, more variation should be seen across learning modes that are supervised, in comparison with learning modes that are unsupervised.

In order to evaluate the pervasiveness of learning mode and category structure interactions, in the present experiment, two category structures composed of simple one-dimensional stimuli were examined under four different unsupervised induction tasks. The four unsupervised learning modes (which will be discussed below) were intended to capture aspects of common real-world

This work was supported by AFOSR Grant F49620-01-1-0295 to B.C.L. I thank Art Markman, Doug Medin, Yasu Sakamoto, Todd Maddox, Todd Gureckis, and two anonymous reviewers for their helpful comments. Correspondence concerning this research should be addressed to B. C. Love, Department of Psychology, University of Texas at Austin, Austin, TX 78712 (e-mail: love@psy.utexas.edu).

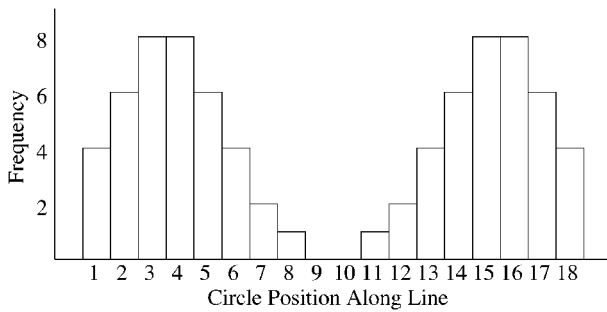


Figure 1. Two quasi-normal distributions (i.e., clusters) illustrating the stimuli seen by the subjects in the study phase of the two-cluster learning problem. In the one-cluster problem, the subjects were exposed only to stimuli drawn from the left distribution.

learning situations. Predictions for each learning mode were derived on the basis of which aspects of the stimulus set were stressed. Observing strong interactions between learning mode and category structure in the experiment (with simple stimuli and under unsupervised learning) would demonstrate the pervasiveness of such interactions. Furthermore, such a pattern of results would indicate that unsupervised learning is multifaceted (as supervised learning has been shown to be).

EXPERIMENT

In the experiment, learning problem (i.e., the category structure) and learning mode (i.e., the unsupervised induction task) were varied factorially. The subjects first completed a study phase (the nature of which was determined by the learning mode). After completing the study phase, the subjects completed the test phase. The test phase was identical for all the subjects. In the test phase, the subjects judged whether novel items were members of the original category. The design allowed for a comparison of category learning performance across different learning problems and modes.

Each stimulus consisted of a circle centered on a horizontal line. The stimuli differed from one another in the circle's position along the line. The two category structures (one cluster and two clusters) examined are illustrated in Figure 1. In the one-cluster problem, the study phase consisted of stimuli with circles toward the left end of line. In the two-cluster problem, the study phase consisted of stimuli with circles toward both the left and the right ends of the line.³ Although these two learning problems are simple, important differences are predicted to occur across the four learning modes. Below, the four learning modes are described, along with predictions for the one-cluster and the two-cluster learning problems.

The Learning Modes

The explicit learning mode. Learning is intentional. Subjects are instructed that they will view a series of

stimuli that are all members of the same category. Subjects are encouraged to discover the pattern or rule that characterizes the category members. Although this learning mode is unsupervised, because corrective feedback is not provided to help the learner discover the underlying concept (i.e., the learner is simply exposed to positive examples of the concept), subjects should still engage in hypothesis testing and discover the pattern that characterizes the study items. Given that the stimuli vary on one dimension and the category structure for both learning problems is relatively simple (i.e., the appropriate rule is verbalizable; see Ashby, Queller, & Berretty, 1999, for a discussion of similar unsupervised learning tasks under intentional conditions), learning under intentional conditions should lead to excellent test phase performance for both learning problems. In all conditions, test phase performance is measured as the correlation between each subject's category endorsement judgments and the actual frequency distribution of training items (see Figure 1). For example, if a subject endorsed items toward the left end of the line in the one-cluster condition, the subject's correlation would be large and positive.

The explicit learning mode is similar to classification learning in that subjects are likely to perform hypothesis testing and construct rules. The explicit learning mode differs from classification learning in that no feedback is provided (learning is unsupervised) and only positive examples of the category are shown (i.e., there is no contrasting category). Thus, the explicit learning mode is properly characterized as unsupervised, because the learner must discover the category structure without guidance (i.e., a learner must be self-supervising). Many real-world learning problems may involve positive examples only.

The pleasantness learning mode. Learning is incidental. Subjects are told that the purpose of the experiment is to norm a set of stimuli for a future experiment and that the experimenter needs the subject to rate the pleasantness of each item. Performance should be excellent for the two learning problems because, in order to make a pleasantness rating, subjects must encode the spatial position of the circle within each stimulus, because position is the only dimension that varies and is, therefore, the only basis on which to make a pleasantness rating. The information emphasized in the study phase is exactly the information required for excellent test phase performance. In many real-world learning situations, we are often exposed to information whose utility is apparent only in a subsequent task.

The same/different learning mode. Learning is incidental. Subjects are instructed that they will see two letters appearing inside a circle and that they should decide as quickly as possible whether the two letters are the same or different. Subjects are told that the position of the circle will vary so as to make the task more difficult. Subjects are also told that the purpose of the experiment is to determine how similar letters are to one another, so that a future experiment can have precisely controlled

stimuli (a cover story similar to that for the pleasantness learning mode). The letters appear in lowercase white type and are completely inscribed within the red circle.

As compared with both the explicit and the pleasantness learning modes, test performance should be lower, because the same/different letter judgments do not emphasize the encoding of spatial position. Although subjects may note the position of the circle while making a same/different judgment, the spatial position of the circle (which defines the category) is irrelevant to the same/different judgments. Position is a nuisance variable. This contrasts with the pleasantness rating task, in which the only basis for a rating is spatial position. Under the same/different study conditions, one would expect that the encoding of spatial position would be degraded, thus reducing performance levels for both learning problems.

In particular, performance in the two-cluster learning problem should suffer more than that in the one-cluster problem. The one-cluster problem does not require that position information should be as precise as in the two-cluster problem. Thus, a subject's performance in the one-cluster problem should not be as strongly affected by imprecise spatial position information. Beyond intuitions, Monte Carlo simulations confirm this conclusion. A set of stimuli was created (see Figure 1), and the position of the circle in each stimulus was perturbed randomly to the left or the right. The perturbation was a random variable drawn from a normal distribution (i.e., normally distributed noise). The noise represented the effect of degraded spatial encoding. As the noise increased (i.e., the standard deviation of the normal distribution generating the noise is increased), the correlation between the target frequency distribution (see Figure 1) and the distribution resulting after the perturbation decreased. As was predicted, the correlation fell much faster for the two-cluster problem than for the one-cluster problem. In fact, at high noise levels, a moderate correlation was observed for the one-cluster problem, whereas the correlation for the two-cluster problem was not significantly different from zero.

Outside the laboratory, there are numerous situations in which category-relevant information is irrelevant in the induction task. One example is a father washing his child's clothing. The father quickly sorts the clothing into piles on the basis of color. Later, the daughter asks the father if he washed her clothing or her older brother's. Even though the father was exposed to the size and style of the clothing, the father may have difficulty answering this question because of his focus on color.

The relative position learning mode. Learning is incidental. Subjects are told that the experiment investigates how the human perceptual system functions. Subjects are instructed to judge whether the circle in the current stimulus is to the left or to the right of the circle in the previous stimulus. Test performance should suffer because the relative position judgment does not emphasize a stimulus's spatial position; rather, it emphasizes the spatial position of the circle *relative* to the position of the circle in the previous stimulus. This focus should

especially hurt performance in the one-cluster problem. In the one-cluster problem, the distribution of relative distances of circle positions is uninformative (all distances are relatively small because all stimuli are drawn from one quasi-normal distribution). Relative distance information is not useful for localizing the position of the cluster, which is critical for test performance. In the two-cluster problem, the distributions of relative distances will be bimodal (roughly half the relative position judgments will be for stimuli drawn from the same quasi-normal distribution, and half will span the two quasi-normal distributions). Only a two-cluster category structure satisfies this distribution of relative distances (i.e., the intended category structure). Thus, subjects should not be greatly impaired in the two-cluster problem.

Instead of focusing subjects on the encoding of exemplars, the construction of rules, or the creation of category abstractions, the relative position learning mode focuses subjects on pairwise comparisons of stimuli. This can have negative consequences for abstraction. For example, a market researcher examining a list of products from some category may not be able to discover the structure of the category by engaging in pairwise comparisons of products when all the products are perturbations of an underlying prototype (i.e., differing slightly from one another on value, quality of advertising, packaging, etc.). The differences between the products do not specify the underlying prototype. On the other hand, if the products are naturally partitioned into two or more centroids, structure discovery might occur from pairwise comparisons. In such a case, the market researcher may notice (without supervision) that one cluster of products tends to be low cost, sparsely advertised, and address basic needs, whereas another cluster of products tends to be high cost, highly advertised, and frivolous.

Method

Subjects. One hundred thirty-three University of Texas undergraduates participated for course credit.

Design. The two independent variables, learning mode (explicit, pleasantness, same/different, or relative position) and learning problem (one cluster or two clusters), were factorially combined in a between-subjects design. After completing the study phase, the subjects in all the conditions completed test trials that accessed what category information was acquired during the study phase.

Apparatus. The experiment was run on Pentium III computers operating in DOS. Data were collected using an in-house real-time data collection system. The monitors had 15-in. CRT color displays and a refresh rate of 16.67 msec.

Stimuli. Each stimulus (displayed on the monitor) depicted a red circle with its center located on a white horizontal line. The background was black. The white horizontal line was approximately 10 cm long and subtended 9.5° of visual arc. In pixel values, the horizontal line was 200 pixels long. The red circle had a radius of 10 pixels. The stimuli differed in circle position along the line. The circle was never centered on the outer 10 pixels of either end of the line (i.e., the circle never crossed either end of the line). The circles could be centered at each of the 180 legal pixel positions. Thus, there were 180 distinct stimuli.

Study phase procedure. Instructions and procedure varied depending on which learning mode a subject was assigned. In all the conditions, instructions were displayed on the computer monitor.

In all the conditions, the stimuli were constructed as described above and were presented in a different random order for each subject. At the start of a trial, the stimulus was shown, and a tone sounded after the subject responded. The stimulus remained displayed for 834 msec (50 screen refreshes), and then a blank screen (all black) was shown for 834 msec. The learning modes differed primarily in the task the subjects were instructed to perform. Aspects of the instructions discussed in The Learning Modes section will not be repeated here.

In the explicit learning conditions, the subjects were instructed to press the space bar whenever they were ready to advance to the next trial. After pressing the space bar, a positive tone sounded.

In the pleasantness conditions, the subjects rated pleasantness on a 9-point scale (from 1 to 9, with 9 signifying a *very pleasant stimulus*) by pressing the corresponding numeric key. After pressing one of the keys, a positive tone sounded.

In the same/different conditions, the subjects judged whether two letters were the same or different. The two letters were randomly generated from the letters of the alphabet, with the constraint that, on half the trials, the two letters were the same (*same* trials) and, on half the trials, the two letters were different (*different* trials). In the one-cluster problem, there was either one more *same* trial or one more *different* trial, because the total number of trials was odd. The ordering of *same* and *different* trials was random for each subject. The subjects pressed the S key when they thought the two letters were the same and pressed the D key when they thought the two letters were different. After the subjects responded, a positive tone sounded when they were correct, and a negative tone sounded when they were incorrect.

In the relative position conditions, the subjects pressed the Q key (on the left side of the keyboard) when they thought the current stimulus's circle was to the left of the previous stimulus's circle. The subjects pressed the P key (on the right side of the keyboard) when they thought the current stimulus's circle was to the right of the previous stimulus's circle. After they responded, a positive tone sounded when they were correct, and a negative tone sounded when they were incorrect.

Learning problem stimulus distributions. Depending on the learning problem (one cluster or two clusters), the stimuli tended to contain a circle located toward the left end of the line or both ends of the line. The line was divided into 18 qualitative regions in which the circle could appear. Figure 1 displays two quasi-normal distributions discretized into the 18 qualitative regions. The normal distribution was approximated with a discretized distribution that was sampled without replacement, because an actual random sample may not be representative of the underlying distribution when the sample size is small (e.g., $n = 39$).

The two distributions shown in Figure 1 characterize the study phase items shown in the two-cluster problem (i.e., the stimuli contained circles appearing on both the left and the right ends of the line). The stimulus frequencies are given by the bar heights in Figure 1. The two distributions of circle positions are mirror images of one another. The centers of the two quasi-normal distributions are 6.0 standard deviations apart (a standard deviation is 20 pixels). The center of each distribution is 2.0 standard deviations from the edge of the line (again, circles were not centered on the 10 pixels closest to each end). Within each quasi-normal distribution, 39 examples appeared during the study phase (randomly sampled without replacement). Each stimulus was constructed by randomly centering the circle on 1 of 10 pixel locations within the region (the uniform distribution was sampled).

In the one-cluster problem, the stimuli for the study trials were drawn (without replacement) from only the left distributions in Figure 1. Thus, there were only 39 study trials for the one-cluster problem but 78 study trials for the two-cluster problem. In order to evaluate interactions between learning mode and learning problem, it was preferable to confound learning problem and number of study trials than to confound learning problem and number of study trials per cluster.

Test phase instructions and procedure. The instructions and procedure for the test phase were identical for all the conditions. The subjects were told that the items viewed in Part 1 (the study phase) were all members of the same category. The subjects were told that they would now see a series of items, some of which were members of the previous category, and some of which were not members of the category. The subjects were instructed to judge the membership of the new items the best they could, even if they were not sure what the category was. The subjects rated the likelihood of a stimulus's category membership on a 9-point scale (from 1 to 9, with 9 signifying *likely membership*) by pressing the corresponding numeric key. After a response, a positive tone sounded. The stimulus then was displayed for 843 msec (50 screen refreshes), and a blank screen (all black) was displayed for 843 msec.

Test phase distributions. In the test phase, the subjects in all the conditions experienced the same distribution of examples—namely, a uniform distribution spanning all 18 qualitative regions. As in the study phase, each stimulus was constructed by centering the circle at 1 of the 10 pixel locations (chosen randomly) within a qualitative region. One stimulus was randomly sampled (without replacement) from each of 18 qualitative regions. This procedure was repeated twice, for a total of 54 test trials.

Results

Mean category endorsement judgments (averaged over subjects) are shown in Figure 2. The subjects' test phase category judgments were analyzed. For each subject, the correlation between the subject's mean category endorsement judgment for each of the 18 qualitative regions and the actual frequency distribution from the study phase (see Figure 1) was calculated.⁴ The mean of these correlations for each of the eight conditions is shown in Table 1. The proportion of subjects displaying statistically significant positive correlations at the .05 alpha level (two tailed) is displayed in Table 2. The count data shown in Table 2 are analyzed here. These nonparametric analyses are not subject to concerns (e.g., skew, restricted range) that could arise in analyzing the correlations shown in Table 1 (although such analyses lead to the same conclusions).

One key prediction of the experiment is that different learning modes should stress different aspects of the stimulus set. This prediction should manifest itself in differences in performance across learning modes, as well as in an interaction between learning mode and learning problem. All predictions were confirmed.

Both main effects were significant. A greater proportion of subjects (43 of 64 vs. 31 of 69) achieved significant (and positive) correlations in the one-cluster learning problem, as compared with the two-cluster learning problem [$\chi^2(1, N = 133) = 6.67, p < .01$]. A main effect was also found for learning mode [$\chi^2(1, N = 133) = 11.86, p < .001$]. To evaluate the interaction between learning problem and learning mode, a log linear model was fit to the data. In the present context, fitting the log linear model is analogous to performing a $2 \times 2 \times 4 \chi^2$ test (see Haberman, 1972, for more information). As was predicted, the interaction between learning problem and learning mode was significant [$\chi^2(3, N = 133) = 11.35, p < .01$]. The interaction was driven principally by the same/different and relative position learning modes. When only these two learning modes are considered, the

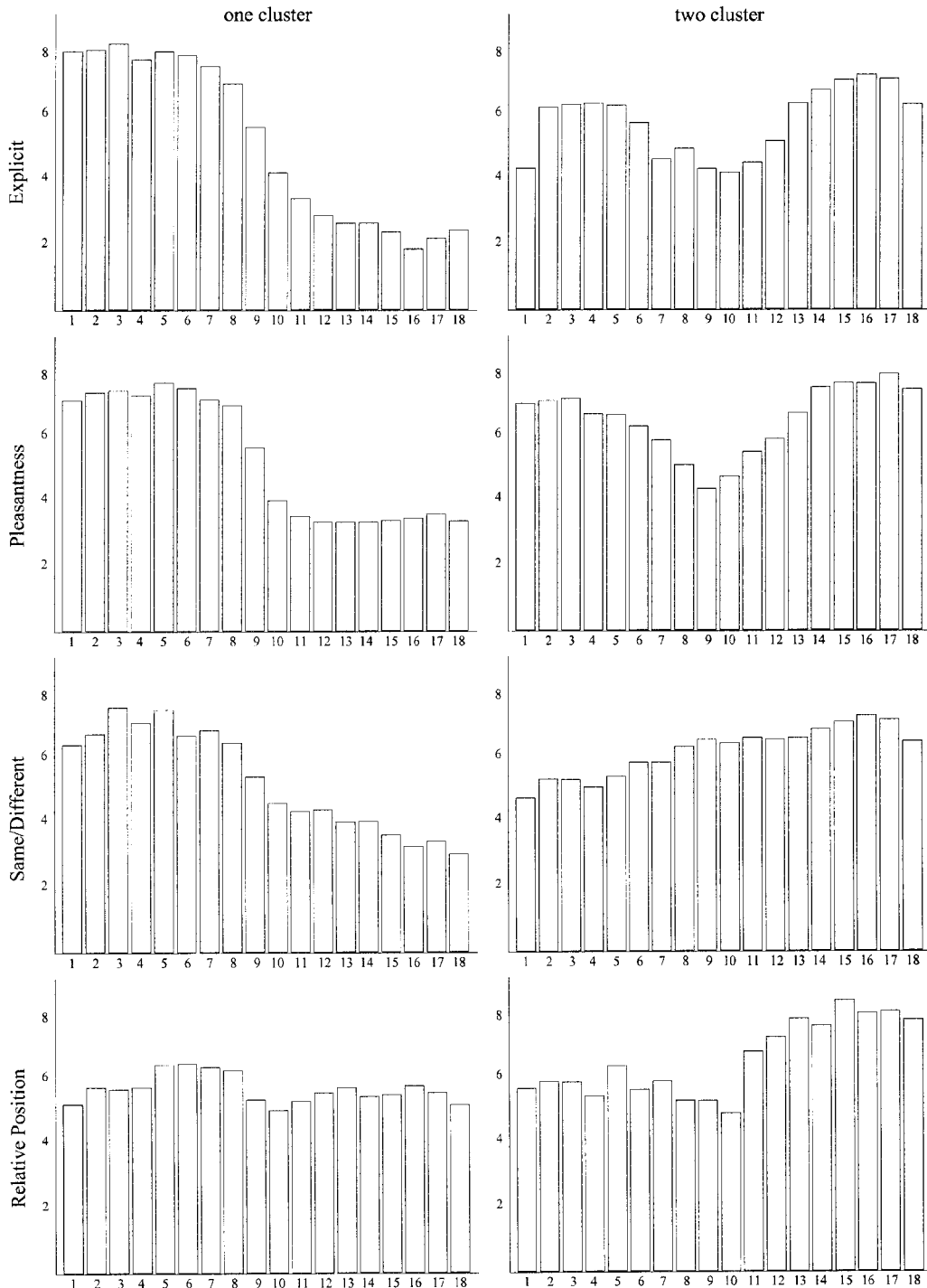


Figure 2. The mean of the subjects' category endorsement ratings for the 18 regions for each of the eight conditions. Care should be taken in interpreting averaged data. For example, the relative position two-cluster condition plot results from a mix of two patterns of endorsement judgments. One pattern correlates highly with the studied distribution. The other pattern involves endorsing stimulus items with circles falling toward the right end of the line (subjects' default strategy).

interaction remains significant [$\chi^2(1, N = 68) = 8.73, p < .01$].

The subjects' performance in the study phase is not formally analyzed here. It suffices to say that the perfor-

mance levels were high in the same/different and the relative position judgment tasks. The subjects' pleasantness ratings (for both the one- and the two-cluster problems) were higher for circle position toward the right end of the

Table 1
The Means of Subjects' Individual Correlations for Each Condition (With Standard Errors of the Means)

Learning Mode	One Cluster		Two Clusters	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Explicit	.67	.11	.36	.110
Pleasantness	.42	.17	.49	.081
Same/different	.45	.14	-.01	.085
Relative position	.07	.16	.29	.097

line; thus, the study phase responses themselves were not likely to serve as the basis for test phase category membership ratings.

Discussion

Even in the domain of unsupervised learning with one-dimensional stimuli, how a learner approaches the stimulus set plays a large role in determining what is acquired. The present findings suggest that interactions between learning mode and learning problem are not the exception, but the rule. In the present study, performance was maximal when learning mode and learning problem were well paired.

Precise encoding of spatial position was emphasized in both the explicit learning mode (intentional learning) and the pleasantness learning mode (incidental learning). Subsequently, the subjects performed at high levels. The subjects in the same/different learning mode, which did not stress the precise encoding of spatial position, performed at lower levels, particularly on the two-cluster problem, which required more precise encoding of spatial position (as evidenced through intuition and Monte Carlo simulations; see The Learning Modes section above). The subjects in the relative position learning mode, which stressed encoding of relative distances between stimuli, and not the absolute spatial positions, also performed at lower levels. Interestingly, this deficit was not as pronounced for the two-cluster problem. In the relative position two-cluster condition, relative distance information can support categorization judgments—the distribution of relative distances is bimodal, and this distribution uniquely indicates a two-cluster category structure (see The Learning Modes section).

In contrast to the present study, the majority of category-learning studies do not explore interactions between learning mode and learning problem. Instead, the majority of studies focus on one mode of acquisition—namely, classification learning (e.g., Shepard et al., 1961). Examining a single learning mode allows researchers to focus on the role the abstract structure of the stimulus set plays in learning. However, the present results demonstrate that general statements about problem difficulty cannot be made without regard to learning mode. The present results suggest that theories of learning need to address the interaction of learning mode and learning problem in order to explain human learning across a range of typical learning environments.

Interestingly, the present results are not easily addressed by positing multiple memory or learning systems (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Cohen & Eichenbaum, 1993; Squire, 1992). For instance, positing implicit and explicit memory systems does not explain the results, because there is no possible way to label the four learning modes as either implicit or explicit and make general statements about what is learnable implicitly versus explicitly. For example, if the three incidental conditions are labeled as implicit learning tasks, there is a case in which difficulties arise in learning the two-cluster problem (the same/different learning mode), a case in which the one-cluster problem presents difficulties relative to the other learning modes (the relative position learning mode), and a case in which both problems are mastered (the pleasantness learning mode). However, the present results do suggest that learning (and possibly memory) is multifaceted (even in the context of unsupervised learning). Perhaps the answer lies in a synthesis of multiple memory systems theories and transfer-appropriate processing theories (e.g., Jacoby, 1983; Roediger, Weldon, & Challis, 1989) in which memory systems are defined by the processes they subservise.

The present results are also particularly challenging for current learning models. The machine-learning community views unsupervised learning as a density estimation problem akin to estimating the joint probability distribution of all features (i.e., all possible parameters are estimated), whereas supervised learning is viewed as more focused, with only the conditional probabilities associated with the category label requiring estimation (Dayan & Hinton, 1996). Similar views permeate the psychology literature and have, in turn, influenced models of human learning. Unsupervised learning is viewed as engaging fairly automatic processes that estimate joint frequency distributions (Hock, Malcus, & Hasher, 1986; Kellogg, 1982) and these views are reflected in learning and memory models (J. A. Anderson, Silverstein, Ritz, & Jones, 1977; McClelland & Rumelhart, 1985). The results presented here suggest that these approaches do not address all of the key phenomena.

Alternative views of unsupervised learning posit that human learners actively discover categories embedded within the stimulus set (J. R. Anderson, 1991; Clapper & Bower, 1991; Love, Medin, & Gureckis, in press). Category discovery models fare no better with respect to the

Table 2
The Proportion of Subjects With Individual Correlations Above .47 (the Minimum Positive Correlation Necessary for Significance at the .05 Alpha Level, Two-Tailed) for Each Condition

Learning Mode	One Cluster	Two Clusters
Explicit	.93 (14/15)	.53 (9/17)
Pleasantness	.73 (11/15)	.67 (12/18)
Same/different	.71 (12/17)	.13 (2/16)
Relative position	.35 (6/17)	.44 (8/18)

present data. These models induce a new category (or cluster) when the current stimulus is not sufficiently similar to any existing category. Like existing models of supervised learning (e.g., Kruschke, 1992; Nosofsky, Palmeri, & McKinley, 1994), these unsupervised learning models are sensitive to the structure of the stimulus set (i.e., the learning problem) but are not sensitive to the learning mode. Even SUSTAIN (Love et al., 2000; Love & Medin, 1998), which can account for data from multiple learning modes (e.g., unsupervised correlation learning, unsupervised sorting, classification learning, and inference learning) because of its sensitivity to both the structure of the stimulus set and the overall pattern of the feedback, cannot address the present data. SUSTAIN fails because the pattern of feedback does not vary in the present study (there is no category-relevant feedback).

One way to model the present data is to fit the data from each learning mode with a unique set of parameters (cf. Nosofsky & Zaki, 1998). For example, parameters related to encoding sensitivity could be reduced in the same/different conditions. Such an approach would probably meet with modest success, but ideally models would serve a greater explanatory role and their behavior would be tailored to the current learning situation without intervention from the modeler. Another potential problem is that parameter differences are unlikely to be informative in all cases.

What is really needed are models that take issues of processing and representation to heart. How is the stimulus represented? How does the current task affect the stimulus's representation? These basic psychological issues are not addressed by models that consider only the abstract structure of the stimulus set.

REFERENCES

- ANDERSON, J. A., SILVERSTEIN, J. W., RITZ, S. A., & JONES, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, **84**, 413-451.
- ANDERSON, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, **98**, 409-429.
- ASHBY, F. G., ALFONSO-REESE, L., TURKEN, A., & WALDRON, E. (1998). A neuropsychological theory of multiple-systems in category learning. *Psychological Review*, **105**, 442-481.
- ASHBY, F. G., QUELLER, S., & BERRETTY, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics*, **61**, 1178-1199.
- BARSALOU, L. W. (1991). Deriving categories to achieve goals. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 27, pp. 1-64). San Diego: Academic Press.
- BERRY, D. C., & DIENES, Z. (1993). *Implicit learning: Theoretical and empirical issues*. Hillsdale, NJ: Erlbaum.
- CLAPPER, J. P., & BOWER, G. H. (1991). Learning and applying category knowledge in unsupervised domains. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 27, pp. 65-108). San Diego: Academic Press.
- CLEERMANS, A. (1993). *Mechanisms of implicit learning: Connectionist models of sequence processing*. Cambridge, MA: MIT Press.
- COHEN, N. J., & EICHENBAUM, H. (1993). *Memory, amnesia, and the hippocampal system*. Cambridge, MA: MIT Press.
- CORTER, J. E., & GLUCK, M. A. (1992). Explaining basic categories: Feature predictability and information. *Psychological Bulletin*, **111**, 291-303.
- DAYAN, P., & HINTON, G. E. (1996). Varieties of Helmholtz machine. *Neural Networks*, **9**, 1385-1403.
- ESTES, W. K. (1994). *Classification and cognition*. New York: Oxford University Press.
- GOLDSTONE, R. L. (1996). Isolated and interrelated concepts. *Memory & Cognition*, **24**, 608-628.
- HABERMAN, S. J. (1972). Log-linear fit for contingency table—Algorithm AS51. *Applied Statistics*, **21**, 218-225.
- HAYES, N., & BROADBENT, D. E. (1988). Two modes of learning for interactive tasks. *Cognition*, **28**, 249-276.
- HEIT, E. (1992). Categorization using chains of examples. *Cognitive Psychology*, **24**, 341-380.
- HOCK, H. S., MALCUS, L., & HASHER, L. (1986). Frequency discrimination: Assessing global and elemental letter units in memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 232-240.
- JACOBY, L. L. (1983). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning & Verbal Behavior*, **22**, 484-508.
- KELLOGG, R. T. (1982). When can we introspect accurately about mental processes? *Memory & Cognition*, **10**, 141-144.
- KRUSCHKE, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22-44.
- LEWICKI, P. (1986). *Nonconscious social information processing*. New York: Academic Press.
- LOVE, B. C. (2001). Three deadly sins of category learning modelers. *Behavioral & Brain Sciences*, **24**, 687-688.
- LOVE, B. C., MARKMAN, A. B., & YAMAUCHI, T. (2000). Modeling classification and inference learning. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence* (pp. 671-676). Cambridge, MA: MIT Press.
- LOVE, B. C., & MEDIN, D. L. (1998). SUSTAIN: A model of human category learning. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence* (pp. 671-676). Cambridge, MA: MIT Press.
- LOVE, B. C., MEDIN, D. L., & GURECKIS, T. (in press). SUSTAIN: A network model of human category learning. *Psychological Review*.
- MCCLELLAND, J. L., & RUMELHART, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, **114**, 159-188.
- MEDIN, D. L., LYNCH, E. B., COLEY, J. D., & AFRAN, S. (1997). Categorization and reasoning among tree experts: Do all roads lead to Rome? *Cognitive Psychology*, **32**, 49-96.
- MEDIN, D. L., & SCHAFFER, M. M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207-238.
- NOSOFSKY, R. M. (1988). Similarity, frequency, and category representations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 54-65.
- NOSOFSKY, R. M., PALMERI, T. J., & MCKINLEY, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, **101**, 53-79.
- NOSOFSKY, R. M., & ZAKI, S. F. (1998). Dissociations between categorization and recognition in amnesic and normal individuals. *Psychological Science*, **9**, 247-255.
- ROEDIGER, H. L., III, WELDON, M. S., & CHALLIS, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 3-41). Hillsdale, NJ: Erlbaum.
- ROSS, B. H. (1997). The use of categories affects classification. *Journal of Memory & Language*, **37**, 240-267.
- SCHANK, R. C., COLLINS, G. C., & HUNTER, L. E. (1986). Transcending inductive category formation in learning. *Behavioral & Brain Sciences*, **9**, 639-686.
- SHEPARD, R. N., HOVLAND, C. L., & JENKINS, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, **75**(13, Whole No. 517).
- SQUIRE, L. R. (1992). Memory and the hippocampus: A synthesis from findings with rats, monkeys, and humans. *Psychological Review*, **99**, 195-231.
- WHITTLESEA, B. W. A., & DORKEN, M. D. (1993). Incidentally, things in general are particularly determined: An episodic-processing ac-

- count of implicit learning. *Journal of Experimental Psychology: General*, **23**, 227-248.
- YAMAUCHI, T., LOVE, B. C., & MARKMAN, A. B. (2002). Learning non-linearly separable categories by inference and classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **28**, 585-593.
- YAMAUCHI, T., & MARKMAN, A. B. (1998). Category learning by inference and classification. *Journal of Memory & Language*, **39**, 124-149.

NOTES

1. Inference learning is a supervised learning mode (like classification learning). Unlike classification learning, learners are given the category membership of a stimulus and are asked to infer the value of an unknown dimension (e.g., "This is a bird, do you think it has wings?"). On different trials, different dimensions are unknown and queried. After corrective feedback is received, the entire stimulus representation is

known (as in classification learning). In fact, after feedback, inference and classification learning are informationally equivalent.

2. See Whittlesea and Dorken (1993) for an opposing view.

3. In addition to these two learning problems, pilot subjects completed an alternative form of the one-cluster problem, in which the one cluster was centered toward the right of the line. This learning problem was not pursued further because it was discovered that subjects are biased to endorse stimuli that lie toward the right end of the line (even in the absence of training). The learning problems included in the experiment provide a more conservative measure of category-learning performance.

4. Seven subjects gave the same rating for each of the 18 regions. For these subjects, the correlation was defined to be zero. Excluding these subjects does not change the pattern of data.

(Manuscript received May 30, 2001;
revision accepted for publication March 1, 2002.)