Learning to classify integral-dimension stimuli

ROBERT M. NOSOFSKY
Indiana University, Bloomington, Indiana

and

THOMAS J. PALMERI
Vanderbilt University, Nashville, Tennessee

The authors tested 288 participants in the classic category-learning tasks introduced by Shepard, Hovland, and Jenkins (1961). However, separable-dimension stimuli were used in previous tests, whereas integral-dimension stimuli were used in the present study. In contrast to previous results, which showed a superiority for Problem Type II over Problem Types III, IV, and V, the reverse pattern was observed in the present research. This result confirms a fundamental prediction made by modern exemplar-based models of classification learning. The results are interpreted in terms of the extent to which selective-attention learning mechanisms operate when separable-dimension versus integral-dimension stimuli are used.

In a classic investigation into the nature of category learning, Shepard, Hovland, and Jenkins (1961) tested people's ability to learn six types of classification problems. This landmark study was pivotal in developing constraints for models of category learning, and it remains highly influential. Numerous modern theorists continue to use the Shepard et al. tasks as benchmarks for models of classification (Anderson, 1991; Estes, 1994; Gluck & Bower, 1988; Kruschke, 1992; Nosofsky, 1984).

A central contribution of Shepard et al.'s (1961) study was the demonstration that models based solely on elementary principles of stimulus generalization were inadequate to explain the nature of category learning. Rather, some abstract process of selective attention to dimensions appeared to be critically involved. Furthermore, this study, together with others reported by Shepard and Chang (1963) and Shepard (1964), was instrumental in sparking the distinction between "integral" and "separable" dimensions that is fundamental in modern thinking about perception and cognition.

The structures of Shepard et al.'s (1961) problems are shown in Figure 1. In all cases, there were eight stimuli varying along three binary-valued dimensions. For purposes of illustration, in the example in Figure 1, the dimensions are color, shape, and size. In all the problems, four stimuli were assigned to Category A and the remaining four were assigned to Category B. Although there are 70 distinct ways of assigning four of eight stimuli to two categories, only six types of problems arise for stimuli varying along three binary-valued dimensions. All problems within a type have the same abstract structure, with

only the assignment of physical dimensions to the logical dimensions varying.

In Problem Type I, only a single dimension is relevant. The example in Figure 1 is to classify all squares into Category A and all triangles into Category B. In Problem Type II, exactly two dimensions are relevant. The example in Figure 1 is to classify black squares and white triangles into Category A and white squares and black triangles into Category B. In this example, the dimensions of color and shape are relevant, whereas size is irrelevant. In Problem Type VI, all three dimensions are equally relevant. Stating a rule for Type VI basically involves enumerating the stimuli in each of the categories. Finally, in terms of structural complexity, Types III, IV, and V are intermediate between Types II and VI. All three dimensions are relevant, but to differing extents. One way of thinking about these problems is as single-dimension-plus-exception structures. For instance, in the Type V example, squares belong to Category A and triangles to Category B, except that the small white triangle is switched with the small white square.

Shepard et al. (1961) found that upon initial exposure to each problem, people learned Type I most rapidly; followed by Type II; followed by Types III, IV, and V, which were approximately equal in difficulty; and finally by Type VI. This ordering of difficulty was also observed by Nosofsky, Gluck, Palmeri, McKinley, and Glauthier (1994), who conducted a replication and extension of the original study. Shepard et al.'s result was important, because a vast class of models based on elementary principles of stimulus generalization failed to predict this ordering. Most critically, these models predicted that Type II should be learned more slowly than Types III, IV, and V.

Shepard et al. (1961) suggested that a process of selective attention was involved when people learned to solve the problems. To solve Type II, people need attend to only two of the three dimensions; the third dimension is
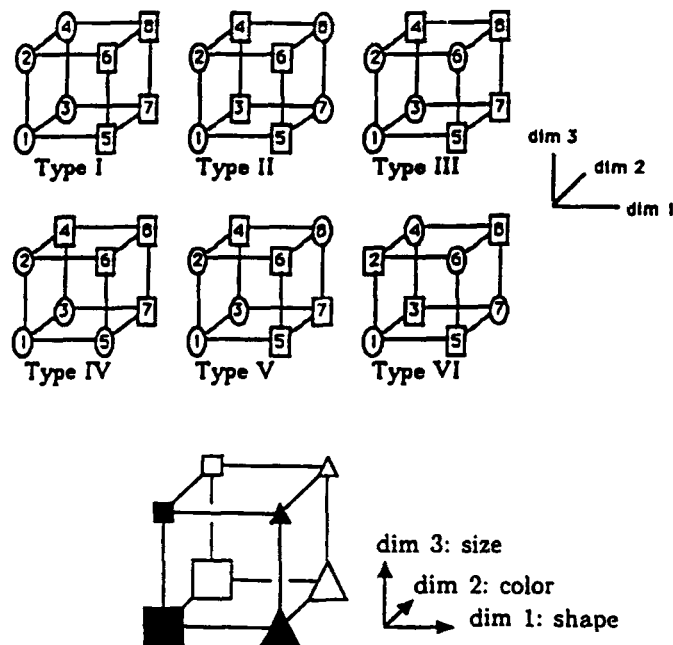
222

Figure 1. (Top) The six types of categorization problems tested by Shepard, Hovland, and Jenkins (1961). The eight stimuli are denoted by the corners of the cubes. Assignments to categories are denoted by the ovals or rectangles that enclose the stimulus numbers. (Bottom) Example in which the stimuli vary along the dimensions of shape, color, and size.

irrelevant. But to solve Types III, IV, and V, one needs to spread attention across all three dimensions.

This interpretation about the influence of selective attention was corroborated in theoretical analyses of Shepard et al.'s (1961) tasks conducted by Nosofsky (1984) and Kruschke (1992). These investigators demonstrated that, without allowing for selective attention processes, modern exemplar models of category learning, which formalize in quantitative fashion the principles of stimulus generalization assumed by Shepard et al., predict that Type II should be learned more slowly than Types III, IV, and V. However, with principles of selective attention incorporated, these same models predict perfectly the ordering of difficulty of the six problem types.

The stimuli used in Shepard et al.'s (1961) original tasks and in Nosofsky et al.'s (1994) replication varied along highly "separable" dimensions. Separable dimensions remain psychologically distinct when in combination; an example is forms varying in shape and color. A vast amount of converging evidence suggests that people are highly efficient at selectively attending to separable dimensions. By contrast, "integral" dimensions combine into relatively unanalyzable unitary wholes; an example is colors varying in hue, brightness, and saturation. Although people can selectively attend to integral dimensions to some degree, the process is far less efficient than what occurs for separable-dimension stimuli.

Shepard and Chang (1963) reasoned that, if people learned to classify integral-dimension stimuli, models

based solely on principles of stimulus generalization might indeed capture the results, because the selective-attention process that operates for separable-dimension stimuli would be largely precluded. Shepard and Chang confirmed this prediction by demonstrating that models of stimulus generalization without the incorporation of selective attention provided reasonably good fits to data from six new category-learning problems in which participants classified integral-dimension color stimuli.

A shortcoming of the seminal investigations of Shepard et al. (1961) and Shepard and Chang (1963), however, is that in addition to varying whether the dimensions were separable or integral, different category structures were tested in the two studies. Indeed, the stimuli used by Shepard et al. varied along three binary-valued dimensions, whereas in Shepard and Chang's studies the stimuli varied along two continuous dimensions. Furthermore, the critical qualitative contrasts in problem difficulty that were present in Shepard et al.'s (1961) studies (Type II vs. Types III–V) did not arise for the structures tested by Shepard and Chang.

Thus, after all these years, there is still a missing part of the picture. It is critically important to replicate the Shepard et al. (1961) tasks, except using integral-dimension stimuli instead of separable-dimension ones. The purpose of the present research was to conduct such an experiment. The prediction stemming from current exemplar models is that when integral-dimension stimuli are used, Problem Type II should be learned more slowly than Problem Types III, IV, and V.

EXPERIMENT

We tested participants on the Shepard et al. (1961) problems, except that, instead of using separable-dimension stimuli, we used integral-dimension colors. Pilot work was conducted to find a set of computer-generated colors that approximated the Munsell configuration illustrated in Figure 2A. In this configuration, the colors vary in hue (purple-blue 7.5 vs. purple 7.5), brightness (value 3 vs. value 6), and saturation (chroma 4 vs. chroma 10). Similarity-scaling work was performed to verify the dimensional structure of the stimuli.

In the classification experiment, each participant solved one of the category-learning problems illustrated in Figure 1. The assignment of the dimensions of hue, brightness, and saturation to the abstract structure of each problem was balanced over participants. We then measured the average rate at which each of the problems was learned.

Method

Participants. The participants in the classification experiment were 288 undergraduates from Indiana University who received credit for a course requirement. Forty-eight participants were randomly assigned to each of the six problems. An additional 34 undergraduates from the same population were tested in a similarity-scaling study.

Stimuli. The stimuli were colors presented on computer screens. Pilot work was conducted to create a set of colors that approximately matched the Munsell configuration illustrated in Figure 2A. Similarity-scaling work (described in the Results section) was then performed to verify that the colors had the requisite psychological structure. The colors were generated on CompuAdd 14-in. monitors by adjusting the red, green, and blue (RGB) color channels on CompuAdd 486 machines. The RGB values for the colors were: 1 [84,68,96], 2 [100,64,80], 3 [136,108,140], 4 [160,108,132], 5 [104,0,132], 6 [128,0,88], 7 [164,0,208], and 8 [204,0,180]. Each color occupied a 9 × 7 cm rectangle surrounded by a gray background. Participants entered responses by pressing appropriate keys on the computer keyboard.

Procedure. Assignment of the dimensions of hue, brightness, and saturation to the logical structure of each problem was balanced over participants. The procedure for the learning of each problem was similar to the one used by Shepard et al. (1961). In the first and second block of 8 trials, each color appeared once in random order. In each subsequent block of 16 trials, each color appeared twice in random order. On each trial, a color was presented, the participant classified it, and feedback was provided. Learning continued until a participant reached a criterion of 4 consecutive subblocks of 8 trials with no errors or for a maximum of 400 trials (25 blocks of 16 trials).
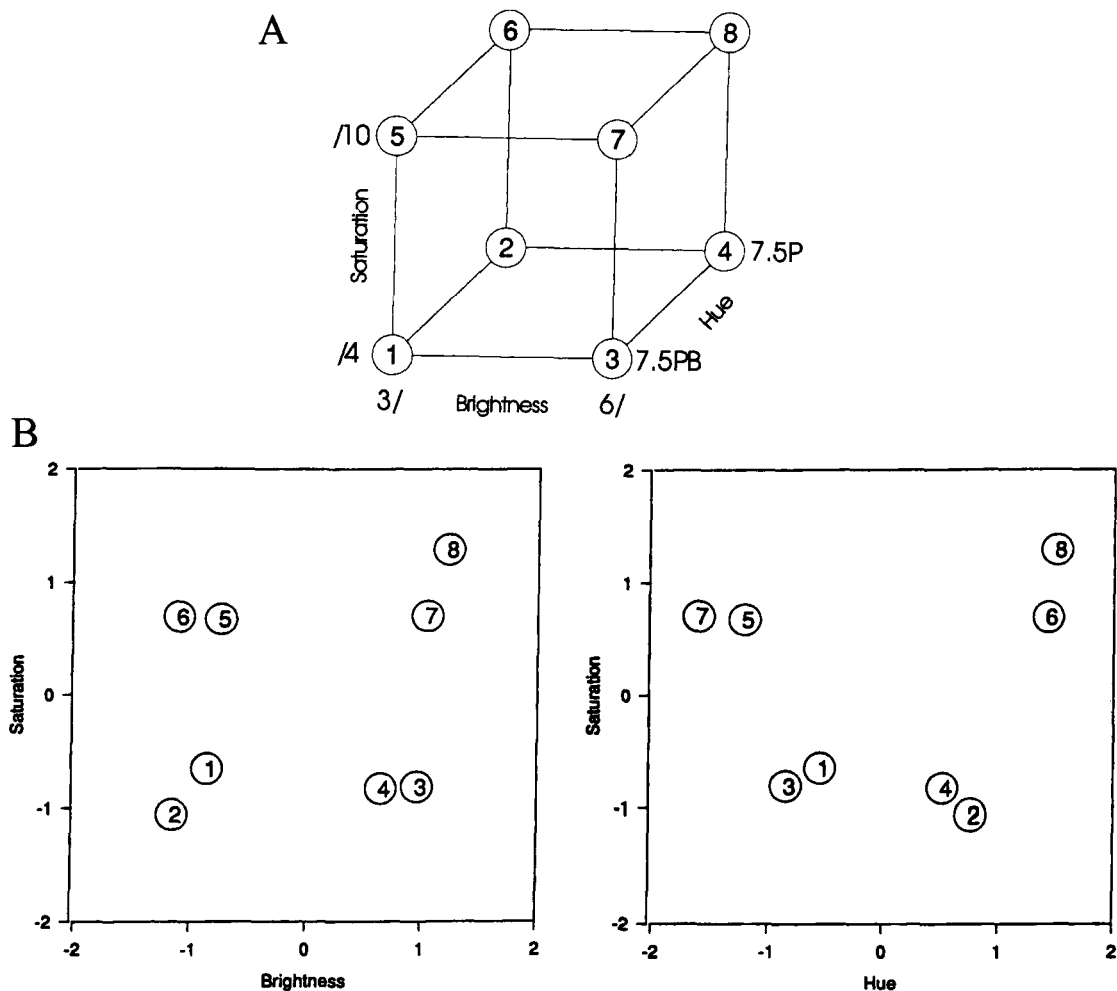

Figure 2. (A) Munsell configuration that served as a model for the computer-generated colors. (B) Scaling solution for the colors used in the experiment.

In the scaling study, there were four blocks of similarity judgments. On each block, all 28 unique pairs of the eight colors were presented, 1 pair per trial, in a random order. These four blocks were preceded by 20 practice trials. The color rectangles were presented simultaneously in the middle of the screen, separated by 3 cm. Participants made similarity ratings by using a 10-point scale (1 = *very dissimilar*, 10 = *very similar*).

Results

Similarity scaling. A multidimensional scaling (MDS) solution for the colors was derived by fitting the standard Euclidean model to the averaged similarity ratings. The three-dimensional solution yielded a stress of .003 and accounted for 99.99% of the variance in the data. The solution is shown in Figure 2B. The derived dimensions correspond reasonably well to the desired configuration, although some issues arise that require discussion. As anticipated, colors 1, 2, 3, and 4 are relatively unsaturated, whereas colors 5, 6, 7, and 8 are saturated; colors 1, 2, 5, and 6 are dark, whereas colors 3, 4, 7, and 8 are light; and color group 1, 3, 5, and 7 is separated from color group 2, 4, 6, and 8 in hue. However, the hues of the colors that are low in saturation are less discriminable than are the hues of the colors that are high in saturation. This same property holds for the original Munsell colors that served as models for our computer-generated ones. A basic property of the Munsell system is that as saturation decreases, all hues converge to a constant gray. We address concerns about the possible influence of this change in discriminability in our analyses of the classification-learning data.

Classification learning. The average probabilities of errors for each problem in each block of 16 trials are shown in Figure 3A. The means on late blocks reflect

zero values for participants who had already reached criterion. Our assumption is that the participants who had reached criterion, and who thereby had already achieved between 32 and 40 consecutive correct responses, would have continued to respond without error if they maintained the same level of motivation.

The learning data confirm our critical prediction: although Problem Type II was learned more quickly than Problem Types III–V in previous studies in which separable-dimension stimuli were used (Nosofsky et al., 1994; Shepard et al., 1961), the reverse is observed for the present integral-dimension stimuli. The overall ordering of difficulty for the six problems in terms of average error probabilities is I, IV, III, V, II, and VI. Using the average error probability for each individual problem as the unit of analysis, pairwise t tests indicate that IV and III were learned with significantly fewer errors than was II, although the difference between V and II was not statistically significant.

More detailed analyses revealed that, regardless of which dimension was irrelevant (hue, saturation, or brightness), performance on Problem Type II was always worse than that on Problem Type IV. This result militates against concerns that poor performance on II relative to IV arose solely from the discriminability differences that existed on the hue dimension. Even when hue was irrelevant, performance on II was worse than performance on IV.

THEORETICAL ANALYSIS

In previous work, Nosofsky et al. (1994) demonstrated that the ALCOVE model (Kruschke, 1992) provided an
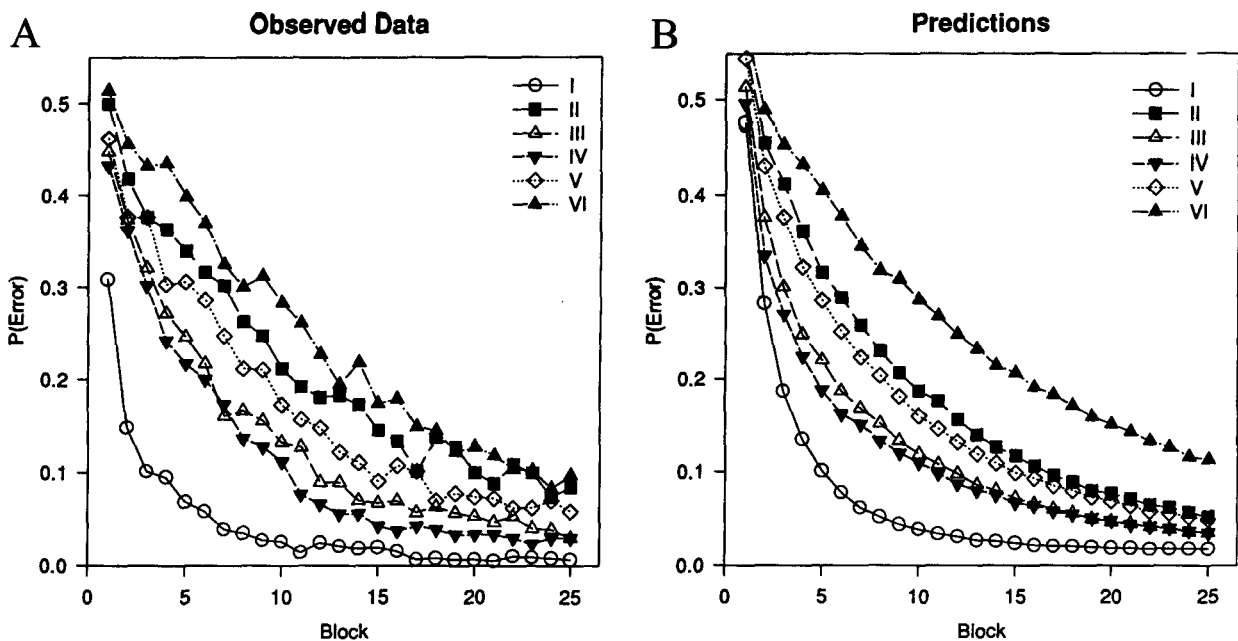

Figure 3. (A) Average probabilities of errors for each problem in each block of 16 trials. (B) Predicted probabilities of errors from the ALCOVE model.

excellent quantitative account of the averaged learning data for the Shepard et al. (1961) tasks in a situation involving separable-dimension stimuli. According to ALCOVE, people learn categories by storing individual exemplars in memory. Classification decisions are based on how similar an item is to these stored exemplars. Selective-attention processes, represented by a set of attention-weight parameters, modify the structure of the space in which the exemplars are embedded. ALCOVE provides an explicit mechanism for how these attention weights are learned. If the mechanism is active, then ALCOVE basically learns to distribute attention over the dimensions so as to optimize performance (Nosofsky, 1984). So, for example, it learns to attend selectively to the single dimension that is relevant in Problem Type I and to the two relevant dimensions in Problem Type II.

The version of ALCOVE fitted by Nosofsky et al. (1994) had four free parameters: an overall sensitivity parameter (κ), a background-noise constant (b), an association-weight learning rate (λ_w), and an attention-weight learning rate (λ_α) (see Kruschke, 1992, and Nosofsky et al., 1994, for more detailed discussion). The same model is now used to fit the current data. Because of the integral nature of the stimulus dimensions, however, the expectation is that the best fitting value of the attention-learning parameter will be near zero.

We fitted ALCOVE by searching for the free parameters that minimized the sum of squared deviations between predicted and observed error probabilities. First, we constructed 120 random stimulus sequences. The characteristics of each sequence matched the constraints in our experimental design. For any given set of parameters, the model was used to generate predictions on the basis of each random sequence. These 120 sets of predicted values were then averaged, and the averaged values constitute the predictions that were fitted to the observed data. A hill-climbing parameter-search routine was used to find the best fitting parameters. Finally, in fitting ALCOVE, the MDS solution derived for the colors (Figure 2B) was used for computing Euclidean distances among exemplars, and predictions of the model were averaged over balanced assignments of the dimensions of hue, brightness, and saturation to the logical structure of each problem.

The predicted learning curves are shown in Figure 3B. The four-parameter model provides a reasonably good quantitative fit to the data, accounting for 94.2% of the variance in the 150 error probabilities. The best fitting parameters were $\kappa = .608$, $b = .018$, $\lambda_w = .338$, and $\lambda_\alpha = 0$. Of greatest importance, the best fitting attention-weight learning rate turned out to be zero. ALCOVE predicts perfectly the ordering of difficulty of the six problems for these integral-dimension stimuli. Most critically, it predicts that Problem Type II is learned more slowly than Problem Types IV, III, and V.

The model also characterizes the subtle differences among Problem Types III–V. Although each of these problems can be characterized as single-dimension-plus-

exception structures, the nature of the exception varies. For instance, similarity relations among exemplars are more favorable in the Type IV structure than in the Type V structure. Type IV is linearly separable, in the sense that the categories can be partitioned by drawing a single oblique plane through the cube. The exemplars in each category form a similarity cluster on each side of the plane. By contrast, in Type V the exception in each category is isolated from the remaining three exemplars, which leads to relatively inefficient learning.

DISCUSSION

The present research verified a fundamental prediction of modern exemplar models incorporating selective-attention learning mechanisms. We replicated the classic Shepard et al. (1961) category problem-solving tasks, except that we used integral-dimension stimuli instead of separable-dimension ones. While previous results involving separable-dimension stimuli had revealed that Problem Type II was solved more quickly than Problem Types III, IV, and V, the reverse was observed for the present integral-dimension stimuli. Both sets of results are explained by selective-attention exemplar models by assuming a high attentional learning rate in a situation involving separable-dimension stimuli but a low attentional learning rate in the present situation involving integral-dimension stimuli.

No single study can be considered definitive, and caution is needed in making cross-experiment comparisons. Conceivably, other differences between the experiments could also account for the dramatically different patterns of results. For example, the rate of learning on all problems was slower in the present study than in the previous studies that had used separable-dimension stimuli. Perhaps differences in overall discriminability among members of the stimulus sets affected selective-attention processes. Although further research is needed, the present study nevertheless goes a long way toward completing the picture conceived long ago by Shepard et al. (1961) and Shepard and Chang (1963) regarding the role of integral and separable dimensions in classification learning.

REFERENCES

ANDERSON, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, **98**, 409-429.

ESTES, W. K. (1994). *Classification and cognition*. New York: Oxford University Press.

GLUCK, M. A., & BOWER, G. H. (1988). Evaluating an adaptive network model of human learning. *Journal of Memory & Language*, **27**, 166-195.

KRUSCHKE, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, **99**, 22-44.

NOSOFSKY, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 104-114.

NOSOFSKY, R. M., GLUCK, M. A., PALMERI, T. J., McKINLEY, S. C., & GLAUTHIER, P. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition*, **22**, 352-369.

SHEPARD, R. N. (1964). Attention and the metric structure of the stimulus space. *Journal of Mathematical Psychology*, **1**, 54-87.

SHEPARD, R. N., & CHANG, J. J. (1963). Stimulus generalization in the learning of classifications. *Journal of Experimental Psychology*, **65**, 94-102.

SHEPARD, R. N., HOVLAND, C. I., & JENKINS, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, **75**(13, Whole No. 517).