11-9-2011

The impact of category separation on unsupervised categorization

Shawn W. Ell
University of Maine, shawn.ell@maine.edu

Gregoryh F. Ashby
ashby@psych.ucsb.edu

Attention, Perception, & Psychophysics (2012), 74: 466-475. doi: 10.3758/s13414-011-0238-z

The impact of category separation on unsupervised categorization

Shawn W. Ell[1] & Gregory F. Ashby[2]

[1]University of Maine, Orono, ME 04469
[2]University of California, Santa Barbara, CA 93106

Abstract Most previous research on unsupervised categorization has used unconstrained tasks in which no instructions are provided about the underlying category structure or the stimuli are not clustered into categories. Few studies have investigated constrained tasks in which the goal is to learn pre-defined stimulus clusters in the absence of feedback. These studies have generally reported good performance when the stimulus clusters could be separated by a one-dimensional rule. The present study investigated the limits of this ability. Results suggest that even when two stimulus clusters are as widely separated as in previous studies, performance is poor if within-category variance on the relevant dimension is nonnegligible. In fact, under these conditions many participants failed even to identify the single relevant stimulus dimension. This poor performance is generally incompatible with all current models of unsupervised category learning.

Introduction

The vast majority of category learning theories have focused on supervised category learning (i.e., the ability to learn categories with the aid of corrective feedback). Several recent theories, however, have incorporated mechanisms for unsupervised category learning (i.e., the ability to learn categories without the aid of corrective feedback) (e.g., Love, Medin, & Gureckis, 2004; Pothos & Chater, 2002). Most empirical research on unsupervised categorization has used unconstrained tasks where participants are not explicitly informed that there is an underlying category structure. Furthermore, in most cases there is no underlying structure to discover in these experiments (i.e., there are no stimulus clusters). In constrained tasks, in contrast, the stimuli form separate clusters, participants are informed that there is an underlying category structure, and they are told that their goal is to attempt to learn the categories in the absence of trial-by-trial feedback. Unconstrained tasks tend to focus on the question of how participants prefer to construct categories whereas constrained tasks tend to focus on what types of category structures participants are capable of learning. Thus, these two approaches are complementary and a thorough understanding of the psychological processes involved in both is necessary in order to refine theories of unsupervised category learning.

In constrained unsupervised category-learning tasks, participants have had the most success when attempting to learn category structures where the optimal decision strategy requires selective attention to a single stimulus dimension (Ashby, Queller, & Berretty, 1999; Ell, Ashby, & Hutchinson, 2011; Zeithamova & Maddox, 2009). In addition, these data suggest that there may be a bias to use one-dimensional rules in constrained tasks. With unconstrained tasks, the evidence for such a one-dimensional bias is far less consistent. Some studies have reported a one-dimensional bias (e.g., Colreavy & Lewandowsky, 2008; Medin, Wattenmaker, & Hampson, 1987), while others have highlighted numerous methodological factors that mediate the bias to use one-dimensional strategies (Ahn & Medin, 1992; Milton, Longmore, & Wills, 2008; Milton & Wills, 2004; Pothos & Chater, 2005; Pothos & Close, 2008; Regehr & Brooks, 1995). For example, simply informing participants of the number of categories has

been argued to instill a one-dimensional bias (e.g., Murphy, 2002).

Studies demonstrating successful unsupervised learning of one-dimensional categorization rules have generally used highly separated categories – that is, category structures in which the within-category variances are low and/or the between-category distance is high. Consider, for instance, the Figure 1 category structures used by Ashby et al. (1999). The stimuli were lines varying continuously across trials in length and orientation and the optimal strategy (i.e., the strategy that maximized accuracy) was the one-dimensional rule "respond A if the line is short, otherwise respond B" (Figure 1). Thus the participant's task was to learn that length was relevant (and that orientation was irrelevant) and to learn a decision criterion on the length dimension. Participants were successful in learning the optimal rule regardless of whether they were trained under supervised or unsupervised conditions.

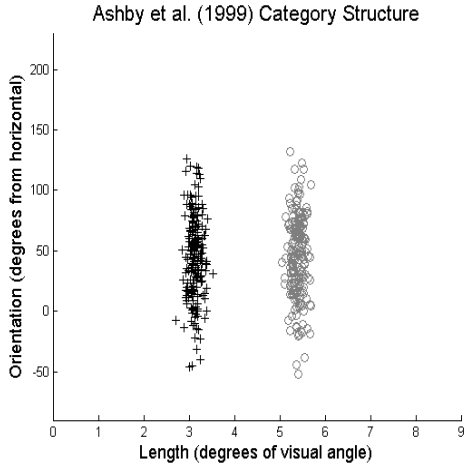

Figure 1. Scatterplot of the stimuli used in the Ashby et al. (1999) experiment. Each point represents a line of a particular length and orientation. Category A and B exemplars are depicted as black plus signs ('+') and gray circles ('o'), respectively. Perfect performance could be obtained by attending selectively to line length and learning the optimal position of a decision criterion that discriminates short and long lines.

within-category variance on the relevant dimension of the Ashby et al. (1999)

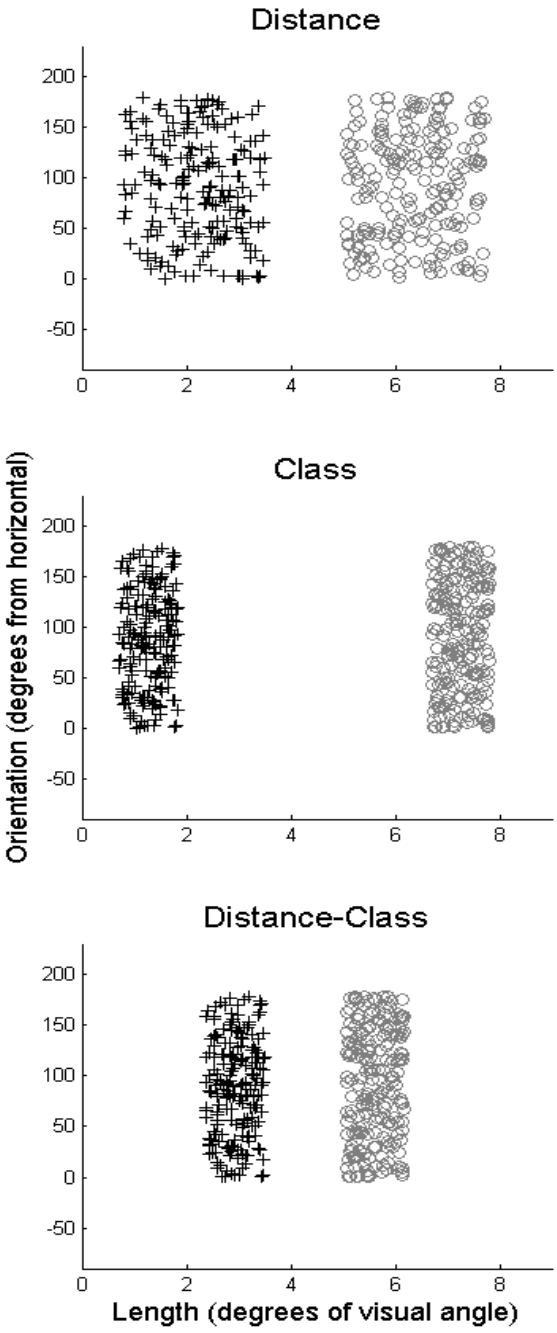

Figure 2. Scatterplots of the stimuli used in the present experiments. Each point represents a line of a particular length and orientation. Category A and B exemplars are depicted as black plus signs ('+') and gray circles ('o'), respectively.

categories was so small that many participants may have perceived this dimension as binary, with one level for category A and another for category B. This feature of the experiment could have been critical because evidence suggests that within-category variance strongly influences unsupervised category learning (Kloos & Sloutsky, 2008). One of the goals of the present study is to determine whether variation in within-category variance along the relevant dimension affects the ability to learn in constrained tasks as well as the bias to use one-dimensional rules.

Any increase in within-category variance of the Ashby et al. categories, in isolation, would also decrease category separation. Thus, in order to permit a comparison to Ashby et al. (1999), it was necessary to manipulate within-category variability while controlling for category separation. We used two different measures of separation. One equates the distance between the nearest exemplars from the contrasting categories (i.e., the between-category distance). This is the Distance condition in Figure 2. A second method equates class separation by equating the standardized distance between the category means using a multivariate analog of the signal detection measure d′ (Fukunaga, 1990). This is the Class condition in Figure 2.

A comparison of the Distance and Class conditions also provides a test of the importance of within-category variability. This comparison, however, is confounded by a difference in the between-category distance. To address this confound, we also included a condition with the same between-category distance as in the Distance condition and the same within-category variance as in the Class condition. This is the Distance-Class condition in Figure 2.

If within-category variance is critical, accuracy should be higher in the Class and Distance-Class conditions than the Distance condition. If category separation is also important, then one might expect the following ordering by accuracy: Class, Distance-Class, Distance. A qualitative comparison to the one-dimensional categories of Ashby et al. (1999) will provide a further test of the importance of within-category variability as the Distance and Class conditions increase within-category variability while controlling for category separation.

Method

Participants and Design.

Sixty participants were recruited from the University of California, Santa Barbara and University of Maine communities and received partial course credit for participation. Twenty participants were randomly assigned to each of three experimental conditions: Distance, Class, and Distance-Class. No participant completed more than one experimental condition. All participants had normal (20/20) or corrected to normal vision. Each participant completed one session of approximately 45 minutes duration.

Stimuli and Apparatus. The stimuli in all experiments were lines that varied continuously along the dimensions of length and orientation[1]. The complete set of stimuli used in the three experimental conditions is

[1] We focused on categories defined by variation in length for two reasons. First, Ashby et al. (1999) observed no differences between length-relevant and orientation-relevant categories. Second, categories where orientation is the only relevant dimension pose serious difficulties when studying unsupervised category learning. Orientation (unlike length) has anchor points that can influence categorization decisions (e.g., Zeithamova & Maddox, 2007). More specifically, people are drawn to highly salient rules that place the criterion on horizontal, vertical, or 45 degree orientations. This is especially problematic with unsupervised studies because such initial biases can dominate performance making it difficult to determine whether the participant's behavior is a result of learning or bias.

shown in Figure 2. The experiment used a variation of the randomization technique introduced by Ashby and Gott (1988) in which each category was defined as a bivariate uniform distribution. Each category distribution was specified by the minimum and maximum on each dimension (see Table 1 for category parameters and class separation and Appendix A for more detail on the calculation of class separation).

On each trial, a random sample (x, y) was drawn from the category A or B distribution and these values were used to construct a line of x pixels in length (ranging from .7 to 7.8 degrees of visual angle) and y degrees of orientation (counterclockwise from horizontal). A total of 400 stimuli (200 from each category) were generated. All stimuli were generated offline and a linear transformation was applied to ensure that the sample statistics matched the population parameters. The experiment was run using the Psychophysics toolbox (Brainard, 1997; Pelli, 1997) in the Matlab computing environment. Each line was presented in white on a black background and was displayed on a 15-inch CRT with 832 x 624 pixel resolution at a viewing distance of 58 inches in a dimly lit room.

Procedure. Each participant was run individually. Participants were told that lines varying in length and orientation would be presented one at a time on a monitor and their task was to learn to categorize the stimuli into two categories. Following Ashby et al. (1999), five observation-only blocks (blocks 1, 3, 5, 7, and 9) alternated with five response blocks (blocks 2, 4, 6, 8, and 10). The same 400 stimuli were presented during the observation and response blocks with presentation order randomized. During the observation-only blocks, participants were instructed to look at 80 sequentially presented stimuli and to try and learn about the categories. The stimuli in the observation-only blocks were presented for 1 s with an inter-stimulus interval of 0.5 s. The observation-only blocks were included in an effort to increase the number of stimuli that the participants were exposed to during an experimental session. The observation-only blocks do not require a response and, thus, take less time to complete than the response blocks (Ashby, et al., 1999). During the response blocks participants were instructed to select a category for each stimulus and to press a button labeled “A” or a button labeled “B” to show which category had been selected. The participants were told that the category labels were arbitrary, but were instructed to be consistent with what they called a member of category A and what they called a member of category B. Given that the category labels were arbitrary, it was assumed that participants assigned the stimuli to the two categories in a manner that resulted in the highest accuracy (percent correct) for each block. Therefore, it was impossible for participants to achieve accuracy below 50% correct in any given block. The participants were told that perfect accuracy was possible, but were never given any feedback about their performance. The stimulus display was response terminated (with 5 s maximum exposure duration) in the response blocks and the response-to-stimulus interval was 0.5 s. The break between blocks was participant paced.

Results

Accuracy-based analyses

Preliminary inspection indicated that the data from all conditions and response blocks were highly bimodal with one mode near chance accuracy and one mode near optimal accuracy (Figure 3). Given these data, we opted to use a series of nonparametric analyses. First, an analysis of the change in accuracy across response blocks (Friedman’s test) indicated that accuracy did not generally improve with training in any condition

Table 1. Parameters of the uniform distributions used to generate the category structures for the three conditions as well as measures of category separation.

	Length (pixels)		Orientation (degrees)		Class Separation	d′
	Min	Max	Min	Max		
Distance					7.5	5.5
Category A	55	245	0	180		
Category B	355	545	0	180		
Class					85.2	18.5
Category A	50	129	0	180		
Category B	471	550	0	180		
Distance-Class					17.2	8.3
Category A	166	245	0	180		
Category B	355	434	0	180		

Note. See Appendix A for details on the calculation of class separation.

[Distance: $\chi^2(4) = .23$, $p = .99$; Class: $\chi^2(4) = 6.31$, $p = .18$; Distance-Class: $\chi^2(4) = 2.55$, $p = .64$]. These data suggest that participants who responded optimally either learned the category structures very early in training, or else guessed the optimal categorization rule at the outset of the experiment.

Next, we computed the proportion of successful participants, with success being defined as above chance accuracy (i.e., 59%)[2] during the majority of response blocks. These data, plotted in Figure 4A, suggest an ordering by condition across the Class, Distance-Class, and Distance conditions. Although the proportion of successful participants was higher in the Class condition relative to the Distance condition [$\chi^2(1) = 6.67$, $p = .03$], the proportion of successful participants in the Distance-Class condition did not differ significantly from either the Class [$\chi^2(1) = 1.6$, $p = .6$] or Distance conditions [$\chi^2(1) =$

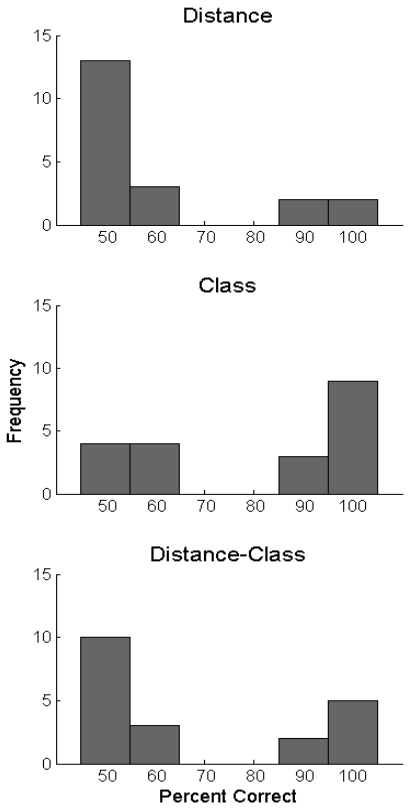

Figure 3. Frequency distributions of the accuracy rates during the final response block for all conditions (bin width = 10%). These data are representative of the frequency distributions for all response blocks.

[2] The criterion for chance performance, 59% correct, was estimated using a binomial distribution ($n = 80$, $p = .5$) at $\alpha = .05$ (one-tailed).

1.91, p = .5][3]. The distribution of successful participants by condition was virtually identical when defining success as above chance accuracy during the final response block (Figure 4B).

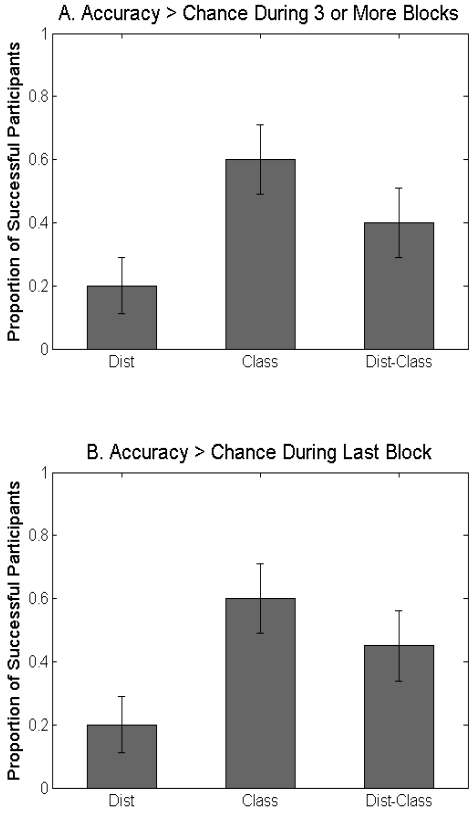

Figure 4. Proportion of successful participants (+/- the standard error of proportion) by condition using two definitions of success. **A.** A successful participant is defined as a participant performing greater than chance during the majority of response blocks. **B.** A successful participant is defined as a participant performing above chance during the last response block.

Chance performance was used as the criterion for success because it is an objective standard against which to judge performance that would not be influenced by idiosyncrasies of the particular sample. That being said, chance represents only a minimal criterion against which to judge performance. Thus, for descriptive purposes, we also investigated the impact of varying the accuracy cutoff on the proportion of successful participants. In Figure 5, the data corresponding to the two definitions of success used in Figure 4 are plotted. Importantly, the numerical ordering of the three conditions was robust across the range of accuracy cutoffs. In sum, the numerical ordering of the three conditions and the superior performance of the Class condition relative to the Distance condition suggest that both within-category variance and between-category

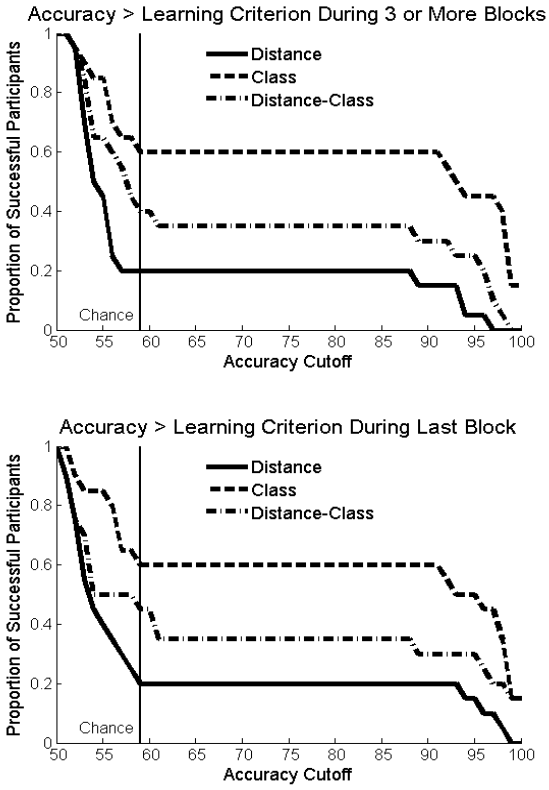

Figure 5. Proportion of successful participants as a function of the accuracy cutoff used to define a success. **A.** A successful participant is defined as a participant performing greater than the cutoff during the majority of response blocks. **B.** A successful participant is defined as a participant performing above the cutoff during the last response block. The vertical line in both plots denotes the criterion used to define success in Figure 4 (i.e., chance). Note that the large range of accuracy cutoffs for which the proportion of successful participants changes very little (i.e., from a cutoff of approximately 60% to a cutoff of approximately 90%) is consistent with the bimodal nature of the accuracy distributions described in Figure 3.

[3] A Sidak correction for multiple comparisons was applied here, and throughout the manuscript.

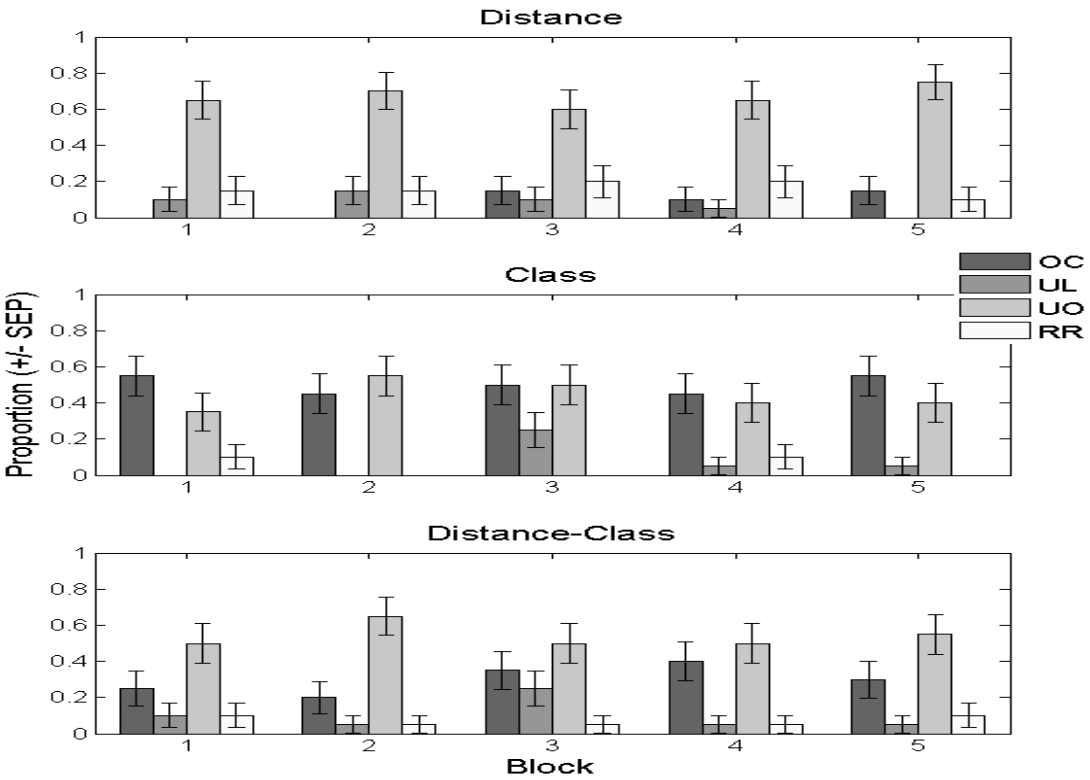

Figure 6. Proportion of participants (+/- the standard error of proportion) in the Distance, Class, and Distance-Class conditions whose data were best fit by the optimal classifier (OC), the suboptimal one-dimensional classifier on length (UL), the suboptimal one-dimensional classifier on orientation (UO), or a model assuming that participants were responding randomly (RR). One block from one participant in the Distance condition and one block from three participants in the Distance-Class condition were best fit by the linear classifier. These data have been excluded from the figure for brevity.

separation influence unsupervised categorization on constrained tasks.

Model-Based Analyses

Analysis of the accuracy data does not directly address the question of what decision strategies were used to perform the Figure 2 tasks. For instance, does near chance performance reflect guessing or that participants adopted a highly suboptimal decision strategy (e.g., a strategy based on orientation)? The following analyses represent a quantitative approach to investigating these questions.

Three different types of models were evaluated, each based on a different assumption concerning the participant's strategy. First, the one-dimensional classifiers assume that the participant attends selectively to one dimension (e.g., if the line is long, respond B; otherwise respond A). There were three versions of the one-dimensional classifier, one assuming participants used the optimal decision strategy on length, one assuming participants used a one-dimensional classifier with a suboptimal intercept on length, and one assuming participants used a one-dimensional classifier on orientation. Second, the general linear classifier assumes that participants integrate the stimulus information from both dimensions prior to making a categorization decision. This model predicts that participants will use a linear decision bound that can have any slope and

intercept. Finally, the random responder models assume that participants guessed. Each of these models was fit separately to the data from every response block for all participants using a standard maximum likelihood procedure for parameter estimation (Ashby, 1992b; Wickens, 1982) and the Bayes information criterion for goodness-of-fit (Schwarz, 1978) (see Appendix B for a more detailed description of the models and fitting procedure).

The proportion of participants best fit by each model type is plotted in Figure 6. In the Distance condition, there was a strong and consistent bias to use a one-dimensional rule on the irrelevant dimension suggesting that the relatively low accuracy was driven by the use of an inappropriate rule rather than by guessing. In the Class condition, a similar proportion of participants used one-dimensional rules on the relevant and irrelevant dimensions. Mirroring the accuracy data, the distribution of best-fitting models in the Distance-Class condition was intermediate between the Distance and Class conditions. Consistent with this descriptive analysis, analyzing the proportion of participants best fit by the optimal classifier across conditions (focusing on block 5 for simplicity) indicated that although the optimal classifier was more frequently used in the Class condition than in the Distance condition [$\chi^2(1) = 6.21$, $p = .04$], the Distance-Class condition did not differ significantly from either the Class [$\chi^2(1) = 2.06$, $p = .45$] or Distance conditions [$\chi^2(1) = 1.29$, $p = .77$]. In sum, the accuracy advantage for participants in the Class condition was driven by more frequent use of the optimal classifier and, in general, there was a strong and consistent bias to use one-dimensional rules across all three conditions.

General Discussion

The ability to categorize in the absence of feedback has been an area of ongoing interest in the categorization literature with the majority of work focusing on categorization preferences in unconstrained tasks where often there is no underlying category structure to discover. Clearly, the question of categorization preference is important, but knowledge of the limitations of unsupervised category learning is also critical for a thorough understanding of real-world cognition. Constrained tasks, such as those investigated in the present study, contribute to this issue by investigating the limits on unsupervised category learning that result from manipulating category separation (i.e., within-category variance and between-category distance). Our results suggest that even when the categories are as widely separated as in Ashby et al. (1999), performance is poor if within-category variance on the relevant dimension is nonnegligible. In fact, under these conditions many participants failed even to identify the single relevant stimulus dimension.

Increasing within-category variance and/or decreasing between-category distance did not reduce the tendency of participants to use one-dimensional rules, but did greatly reduce their ability to find the one relevant stimulus dimension. Participants in the condition with high within-category variance and low between-category distance (i.e., the Distance condition) were less likely to use the optimal decision strategy than participants in the condition with low within-category variance and high between-category distance (i.e., the Class condition). Somewhat surprisingly, however, one-dimensional strategies on the irrelevant stimulus dimension were prevalent in all conditions and their use did not differ in frequency across conditions [$\chi^2(2) = 5.02$, $p = .08$].

An open, but critically important question is whether our participants learned anything in this experiment. Evidence favoring learning can be found in the large number of participants who responded optimally, but evidence against learning comes from the

statistical analyses that failed to find any evidence that accuracy improved across blocks in any experimental condition. If there was no learning, then why did so many participants respond optimally? One possibility is that participants have a strong preference to use one-dimensional rules, and that each stimulus dimension was equally salient. This hypothesis provides a good account of our results. First, it correctly predicts no improvement in accuracy with training (because there was no learning). Second, it predicts that by chance, roughly half the participants will select the optimal rule and half will select a rule on the irrelevant dimension. This pattern roughly matches the results in each condition. On the other hand, note that this hypothesis predicts no difference across conditions. Thus, the slightly better performance we observed in the Class condition is evidence that at least in this condition, some category learning occurred.

Recall that in the Ashby et al. (1999) experiments, the distance between categories was the same as in our Distance condition and the class separation was the same as in our Class condition. Yet virtually all participants in the Ashby et al. one-dimensional conditions were responding with near perfect accuracy by the end of their unsupervised training and the responses of all of those participants were best fit by the optimal one-dimensional classifier during their final response block. In contrast, many participants in our Distance and Class conditions were responding with near chance accuracy at the end of their training and roughly half of these participants were basing their categorization responses on the value of the stimulus on the irrelevant dimension. Our data therefore strongly suggest that the excellent performance of the Ashby et al. (1999) participants was not due only to the distance between the categories or to their class separation.

Why were the participants in the one-dimensional conditions of Ashby et al. (1999) so much better than our participants? One obvious hypothesis is that the within-category variance on the relevant dimension was much smaller for the Ashby et al. categories (i.e., 75) than for any of our conditions (e.g., 520 in our Class condition). In fact, as mentioned earlier, there was so little variance along the relevant dimension in the Ashby et al. categories that participants in those (one-dimensional) conditions may have noticed only two discrete values – and associated one of them with each category. If so, then their optimal behavior might not be unexpected. This hypothesis seems to predict that successful unsupervised category learning is likely quite rare – in effect, limited to categories that can be separated on a single stimulus dimension and for which all category exemplars share (or nearly share) a common value on that dimension. It is important to note, however, that the within-category variance hypothesis does not provide a complete account of the data as there was no significant difference between the proportion of participants performing above chance in the Class and Distance-Class conditions.

A second, less obvious possibility is that the variance along the irrelevant dimension is also important. More specifically, the ratio of the within-category variance along the irrelevant dimension to the variance along the relevant dimension may be an important factor. The idea is that learning should be easier the greater this ratio because large ratios may draw more attention to the relevant dimension. Indeed, similar category complexity measures have been shown to be predictive of supervised (Alfonso-Reese, Ashby, & Brainard, 2002) and unsupervised (e.g., Kloos & Sloutsky, 2008) category-learning performance. This variance ratio correctly predicts the ordering by task difficulty across the Ashby et al. (125), Class (5.2) and Distance (0.9) category structures. Note though that the variance ratio is not influenced by between-category distance and,

therefore, it incorrectly predicts no difference between the Distance and Distance-Class conditions. In this sense, class separation (or other d′ like statistics) provides a better account of our data because it correctly predicts the difficulty ordering of all three conditions. The problem, of course, is that class separation incorrectly predicts no difficulty difference between our Class condition and the one-dimensional conditions of Ashby et al. (1999). Thus, none of the common metrics discussed here provide a complete explanation of the performance differences across the Distance, Class, and Distance-Class conditions and the one-dimensional categories of Ashby et al. (1999).

Implications for Models of Category Learning

The finding that unsupervised categorization performance is improved if within-category variance is reduced and/or if between-category distance is increased is consistent with many current computational models of unsupervised categorization (e.g., Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Fried & Holyoak, 1984; Love, et al., 2004). Even so, this fact alone does not guarantee that a model will be able to predict our results. For example, Pothos and Chater's (2002) simplicity model predicts that the larger the within-category similarity and the smaller the between-category similarity, the more intuitive the categories (Pothos & Bailey, 2009). Assuming that higher intuitiveness implies higher accuracy, the simplicity model correctly predicts that the categories from the Class condition are more intuitive than the categories from the Distance condition, but it also incorrectly predicts that the categories from the Class condition are more intuitive than the Ashby et al. (1999) categories[4]. It is likely, however, that more recent extensions of the simplicity model will be able to account for these data upon further development (e.g., Pothos & Close, 2008).

Even though some unsupervised models may be able to account for the ordering by task difficulty that we observed across our three conditions, they would all have difficulty accounting for the high prevalence of one-dimensional strategies on the irrelevant dimension. At first glance, the explicit (i.e., rule-based) subsystem of the COVIS model of category learning (Ashby, et al., 1998) might be in the best position to predict these data. COVIS was designed as a model of supervised category learning, but because it assumes that there is a bias to use one-dimensional rules (a bias that cannot be overcome in the absence of feedback), it could have some success predicting these data. In COVIS, however, the stimulus dimension that is selected is determined by the relative salience. If length and orientation were equally salient, COVIS would predict that the one-dimensional rules on length and orientation would be used approximately equivalently. Although such a prediction is generally consistent with the data from the Class and Distance-Class conditions, it is inconsistent with the Distance condition (and the data of Ashby et al., 1999). Similarly, relatively greater salience on either length or orientation would result in a misprediction for some subset of the available data. As is the case with many models of category learning (e.g., Erickson & Kruschke, 1998; Kruschke, 1992), COVIS assumes that salience can change as a result of learning. Learning-

[4] We verified this in the following way. First, we generated samples of 400 stimuli (200 from each category) from the Table 1 distributions or from the one-dimensional condition of Ashby et al. (1999). For computational ease we used between- and within-category dissimilarity. To determine between-category dissimilarity we computed the sum of all pairwise Euclidean distances for stimuli from contrasting categories from a sample of 400 stimuli (200 from each category) generated. To determine the within-category dissimilarity, we computed the sum of all pairwise distances for stimuli from the same category. Intuitiveness was computed as between-category dissimilarity minus within-category dissimilarity. Thus, larger values imply greater intuitiveness.

related changes in salience would improve the ability of COVIS to account for these data, but this learning mechanism is driven by external feedback and, therefore, would not be predicted to contribute on unsupervised categorization tasks.

Summary

In sum, our results suggest that people are surprisingly poor at unsupervised category learning on constrained tasks. Roughly half of our participants performed at chance, even on widely separated categories that differed on only one relevant dimension. These results present a challenge to extant models of unsupervised category learning. We argue that these data suggest a need for a more thorough investigation of the properties of category structures that bias selective attention processes toward different stimulus dimensions. More specifically, models of unsupervised category learning should include a more detailed mechanism by which category separation can influence predictions regarding how the task is learned.

References

Ahn, W., & Medin, D. L. (1992). A two-stage model of category construction. *Cognitive Science, 16*, 81-121.

Alfonso-Reese, L. A., Ashby, F. G., & Brainard, D. H. (2002). What makes a categorization task difficult. *Perception & Psychophysics, 64*, 570-583.

Ashby, F. G. (1992a). Multidimensional models of categorization. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition*. Hillsdale, NJ: Erlbaum.

Ashby, F. G. (1992b). Multivariate probability distributions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 1-34). Hillsdale: Lawrence Erlbaum Associates, Inc.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*, 442-481.

Ashby, F. G., & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 33-53.

Ashby, F. G., & Lee, W. W. (1993). Perceptual variability as a fundamental axiom of perceptual science. In S. C. Masin (Ed.), *Foundations of percpetual theory* (pp. 369-399). Amsterdam: Elsevier.

Ashby, F. G., Queller, S., & Berretty, P. M. (1999). On the dominance of unidimensional rules in unsupervised categorization. *Perception & Psychophysics, 61*, 1178-1199.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93*, 154-179.

Brainard, D. H. (1997). Psychophysics software for use with MATLAB. *Spatial Vision, 10*, 433-436.

Colreavy, E., & Lewandowsky, S. (2008). Strategy development and learning differences in supervised and unsupervised categorization. *Mem Cognit, 36*(4), 762-775.

Ell, S. W., Ashby, F. G., & Hutchinson, S. (2011). Unsupervised category learning with integral-dimension stimuli. *submitted for publication*.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General, 127*, 107-140.

Fried, L. S., & Holyoak, K. J. (1984). Induction of category distributions: A framework for classification learning. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 10*, 234-257.

Fukunaga, K. (1990). *Introduction to statistical pattern classification*. USA: Academic Press.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Kloos, H., & Sloutsky, V. M. (2008). What's behind different kinds of kinds: effects of statistical density on learning and representation of categories. *J Exp Psychol Gen, 137*(1), 52-72.

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

Love, B. C., Medin, D. L., & Gureckis, T. M. (2004). SUSTAIN: A network model of category learning. *Psychological Review, 111*, 309-332.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics, 53*, 49-70.

Medin, D. L., Wattenmaker, W. D., & Hampson, S. E. (1987). Family resemblance, conceptual cohesiveness, and category construction. *Cognitive Psychology, 19*, 242-279.

Milton, F., Longmore, C. A., & Wills, A. J. (2008). Processes of overall similarity sorting in free classification. *J Exp Psychol Hum Percept Perform, 34*(3), 676-692.

Milton, F., & Wills, A. J. (2004). The influence of stimulus properties on category construction. *J Exp Psychol Learn Mem Cogn, 30*(2), 407-415.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge: MIT Press.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437-442.

Pothos, E. M., & Bailey, T. M. (2009). Predicting category intuitiveness with the rational model, the simplicity model, and the generalized context model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1062-1080.

Pothos, E. M., & Chater, N. (2002). A simplicity principle in unsupervised human categorization. *Cognitive Science, 26*, 303-343.

Pothos, E. M., & Chater, N. (2005). Unsupervised categorization and category learning. *Q J Exp Psychol A, 58*(4), 733-752.

Pothos, E. M., & Close, J. (2008). One or two dimensions in spontaneous classification: a simplicity approach. *Cognition, 107*(2), 581-602.

Regehr, G., & Brooks, L. R. (1995). Category organization in free classification: The organizing effect of an array of stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(347-363).

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461-464.

Wickens, T. D. (1982). *Models for behavior: Stochastic processes in psychology*. San Francisco: W. H. Freeman.

Zeithamova, D., & Maddox, W. T. (2007). The role of visuospatial and verbal working memory in perceptual category learning. *Memory & Cognition, 35*, 1380-1398.

Zeithamova, D., & Maddox, W. T. (2009). Learning mode and exemplar sequencing in unsupervised category learning. *Journal of Experimental Psycholgy: Learning, Memory, & Cognition, 35*(3), 731-741.

Author Notes

This research was supported in part by National Science Foundation Grant BCS99-75037 and Public Health Service Grants R01 MH3760-2 and P01 NS044393. We would like to thank Jared Williams for his assistance with data collection and Emmanuel Pothos and two anonymous reviewers for their comments on a previous draft of the manuscript. Correspondence concerning this article should be addressed to Shawn W. Ell, Psychology Department, University of Maine, 5742 Little Hall, Room 301, Orono, ME 04469-5742 (email: shawn.ell@umit.maine.edu) or F. G. Ashby, Department of Psychological & Brain Sciences, University of California, Santa Barbara, CA 93106 (e-mail: ashby@psych.ucsb.edu).

Appendix A

Class Separation

Class separation is a multivariate analog of the signal detection measure from the statistical pattern recognition literature. Class separation is based on a measure of the variability between category means, denoted by S_b, and a measure of the variability within

each category, denoted by S_w (Fukunaga, 1990). The between category variability matrix S_b is defined as

$$S_b = \frac{1}{2}\left[\left(\underline{\mu}_A - \underline{m}\right)\left(\underline{\mu}_A - \underline{m}\right)^T\right] + \frac{1}{2}\left[\left(\underline{\mu}_B - \underline{m}\right)\left(\underline{\mu}_B - \underline{m}\right)^T\right], \text{ and } \underline{m} = \frac{1}{2}\left(\underline{\mu}_A - \underline{\mu}_B\right),$$

where $\underline{\mu}_A$ and $\underline{\mu}_B$ are the means of categories A and B, respectively. When the two categories have the same variance-covariance matrix (as in the present experiments), the within-category variability matrix S_w equals the common variance-covariance matrix of each category (i.e., Σ). Given these definitions, class separation is defined as

$$J = \text{trace}(S_w^{-1} S_b),$$

where the trace of a matrix equals the sum of all elements on the main diagonal.

Appendix B

Model-Based Analyses

To get a more detailed description of how participants categorized the stimuli, a number of different decision bound models (Ashby, 1992a; Maddox & Ashby, 1993) were fit separately to the data for each participant from every block. Decision bound models are derived from general recognition theory (Ashby & Townsend, 1986), a multivariate generalization of signal detection theory (Green & Swets, 1966). It is assumed that, on each trial, the percept can be represented as a point in a multidimensional psychological space and that each participant constructs a decision bound to partition the perceptual space into response regions. The participant determines which region the percept is in, and then makes the corresponding response. While this decision strategy is deterministic, decision bound models predict probabilistic responding because of trial-by-trial perceptual and criterial noise (Ashby & Lee, 1993).

The appendix briefly describes the decision bound models. For more details, see Ashby (1992a) or Maddox and Ashby (1993).

One-dimensional Classifier

This model assumes that the stimulus space is partitioned into two regions by setting a criterion on one of the stimulus dimensions. Three versions of the one-dimensional classifier were fit to these data: one assumed that participants attended selectively to length (UL) and another assumed participants attended selectively to orientation (UO). The one-dimensional classifier has two free parameters: a decision criterion on the relevant perceptual dimension and the variance of internal (perceptual and criterial) noise (i.e., σ^2). A third version is a special case of the UL, the optimal one-dimensional classifier (UC), that assumes that participants use the one-dimensional decision bound that maximizes accuracy (Figure 2). This special case has one free parameter (σ^2)

General Linear Classifier

This model assumes that a linear decision bound partitions the stimulus space into two regions and integrates the perceived values on the stimulus dimensions prior to producing a categorization response. The general linear classifier (LC) has three parameters, slope and intercept of the linear bound, and σ^2.

Random Responder Models

Equal Response Frequency (ERF). This model assumes that participants randomly assign stimuli to the two response frequencies in a manner that preserves the category base rates (i.e., 50% of the stimuli in each category). This model has no free parameters.

Biased Response Frequency (BRF). This model assumes that participants randomly assign stimuli to the two response frequencies

in a manner that matches the participant's categorization response frequencies (i.e., the percentage of stimuli in each category is computed from the observed response frequencies). This model has no free parameters.

Model Fitting

The model parameters were estimated using maximum likelihood which entails finding the parameters that maximize the likelihood of the data (or, equivalently, minimizing the negative natural log of the likelihood) (Ashby, 1992b; Wickens, 1982). The goodness-of-fit statistic was

$$BIC = r \ln N - 2 \ln L,$$

where N is the sample size, r is the number of free parameters, and L is the likelihood of the model given the data (Schwarz, 1978). The BIC statistic penalizes a model for poor fit and for extra free parameters. To find the best model among a set of competitors, one simply computes a BIC value for each model, and then chooses the model with the smallest BIC.