Psychonomic Bulletin & Review
1999, 6 (3), 495-503

Learning categories at different hierarchical levels: A comparison of category learning models

THOMAS J. PALMERI
Vanderbilt University, Nashville, Tennesee

Three formal models of category learning, the rational model (Anderson, 1990), the configural-cue model (Gluck & Bower, 1988a), and ALCOVE (Kruschke, 1992), were evaluated on their ability to account for differential learning of hierarchically structured categories. An experiment using a theoretically challenging category structure developed by Lassaline, Wisniewski, and Medin (1992) is reported. Subjects learned one of two different category structures. For one structure, diagnostic information was present along a single dimension (1-D). For the other structure, diagnostic information was distributed across four dimensions (4-D). Subjects learned these categories at a general or at a specific level of abstraction. For the 1-D structure, specific-level categories were learned more rapidly than general-level categories. For the 4-D structure, the opposite result was observed. These results proved highly diagnostic for evaluating the models—although ALCOVE provided a good account of the observed results, the rational model and the configural-cue model did not.

In recent years, there has been tremendous growth in the development of formal models of classification. These include exemplar (e.g., Estes, 1994; Kruschke, 1992; Nosofsky, 1986; Nosofsky & Palmeri, 1997; Palmeri, 1997), connectionist (e.g., Gluck & Bower, 1988a, 1988b), Bayesian statistical (e.g., Anderson, 1990), decision bound (e.g., Ashby & Maddox, 1993; Maddox & Ashby, 1993), and rule-based models (e.g., Nosofsky & Palmeri, 1998; Nosofsky, Palmeri, & McKinley, 1994; Palmeri & Nosofsky, 1995). Although these models make different assumptions about category representations and processes, many of them make similar predictions of some elementary patterns of classification data. This has caused researchers to look at more detailed aspects of classification in order to evaluate models. In the present work, rather than simply asking whether models could account for transfer data following category learning, the models were instead evaluated on whether they could account for patterns of classification throughout the entire training sequence. This work follows in a line of recent studies that have focused on understanding the details of the category learning process (e.g., Estes, 1986; Estes, Campbell, Hatsopoulos, & Hurwitz, 1989; Kruschke, 1992; Nosofsky, Gluck, Palmeri, McKinley, & Glauthier, 1994; Nosofsky, Kruschke, & McKinley, 1992; Nosofsky & Palmeri, 1996).

Surprisingly, formal models of classification have largely neglected issues surrounding category learning at different hierarchical level (see, however, Estes, 1993). The seminal work of Rosch and colleagues (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) demonstrated that intermediate levels of a category hierarchy (the *basic level*) have a privileged status relative to superordinate or subordinate categories (see Lassaline, Wisniewski, & Medin, 1992, for one recent review). For example, objects are often classified most rapidly at the basic level (e.g., Murphy & Smith, 1982; Rosch et al., 1976), and categories are often learned most rapidly at the basic level (e.g., Lassaline et al., 1992). Although most models of supervised[1] category learning (in contrast to unsupervised learning; see Fisher & Langley, 1990, and Schyns, 1991) have not been formalized with hierarchical aspects of classification in mind, it seems reasonable to evaluate whether these models can account for differences in learning categories at different hierarchical levels.

For purposes of evaluating the category learning models, an intriguing category structure reported by Lassaline et al. (1992; Experiment 3 of Lassaline, 1990) was used. In their study, subjects learned one of two different category structures at either a specific or a general level of a category hierarchy. As shown in Table 1, the two different structures primarily differed in whether the defining features fell along just a single dimension (1-D structure) or fell along all four dimensions (4-D structure). Lassaline et al. observed an interesting, and theoretically challenging, pattern of results. For the 1-D structure, a "basic-level effect" was observed—fewer errors were made when learning the specific level categories than when learning the general level categories. For the 4-D structure, the opposite pattern of results was observed.

These results presented formidable challenges to three classification models Lassaline et al. (1992) examined: A category utility measure (Gluck & Corter, 1985), the

This work was supported by Vanderbilt URC Direct Research Support Grants and by Grant PHS R01 MH48494-06 from NIMH to Indiana University. Correspondence should be addressed to T. J. Palmeri, Department of Psychology, 301 Wilson Hall, Vanderbilt University, Nashville, TN 37240 (e-mail: thomas.j.palmeri@vanderbilt.edu).

Table 1
Category Structure Used in the Experiment (From Lassaline et al., 1992)

			Category Structure							
	Category Label		1-D				4-D			
Stimulus	General	Specific	D1	D2	D3	D4	D1	D2	D3	D4
1	A	C	*0*	0	2	3	*0*	1	3	2
2	A	C	*0*	1	3	0	*0*	2	1	3
3	A	C	*0*	2	0	1	*0*	3	2	1
4	A	D	*1*	3	1	0	1	*0*	2	1
5	A	D	*1*	0	2	1	2	*0*	3	1
6	A	D	*1*	1	3	2	3	*0*	1	2
7	B	E	*2*	2	0	2	3	2	*0*	1
8	B	E	*2*	3	1	3	1	3	*0*	2
9	B	E	*2*	0	2	0	2	1	*0*	3
10	B	F	*3*	1	3	1	2	3	1	*0*
11	B	F	*3*	2	1	2	3	1	2	*0*
12	B	F	*3*	3	0	3	1	2	3	*0*

Note—The italicized values highlight those feature values that are most diagnostic for category membership.

adaptive network model (Gluck & Bower, 1988b), and the context model (Medin & Schaffer, 1978) could not account for the interaction of level and structure. Category utility predicted a specific-level advantage for both structures, whereas the adaptive network model and the context model predicted a general-level advantage for both structures. Would these results also pose challenges to more recent, and potentially more sophisticated, category learning models? In particular, could the rational model (Anderson, 1990), the configural-cue model (Gluck & Bower, 1988a), and ALCOVE (Kruschke, 1992) account for an interaction of category level and structure?

Unfortunately, two aspects of Lassaline et al. (1992) make it impossible to use their results to evaluate these models. First, they only reported average accuracy throughout training. In order to rigorously evaluate the category learning models, the present study instead reported classification data throughout the course of training. Second, they used a category verification paradigm in which a stimulus and a category label were simultaneously displayed, and the subjects' task was to decide whether that category label was the correct one or not. The present study instead used a more typical category learning paradigm in which a stimulus was displayed, and the subjects' task was to decide which category label from a number of possible category labels to apply to the stimulus. This kind of task is preferable, largely because the various category learning models have been explicitly formulated to account for data from just such classification paradigms.

Note that in using this more typical kind of category learning task, the interesting interaction of category level and structure that Lassaline et al. (1992) reported is not guaranteed to be reproduced. In the category verification task, there are just two possible responses ("yes" or "no") at both the general level and the specific level. By contrast, in the classification task used in the present study, there were two possible responses at the general level, but there were four possible responses at the specific level. By simply guessing, subjects learning at the general level would be correct half of the time, whereas subjects learning at the specific level would be correct only one fourth of the time. In order to observe a specific-level advantage in this task, subjects would need to surmount this relatively large base rate difference—obtaining a crossover interaction between category level, and the amount of training is a real empirical challenge. Moreover, this pattern of results would provide an even more difficult test for the category learning models.

Because it has proven quite difficult to train individuals to classify objects at different hierarchical levels simultaneously, different groups of subjects in each of the four experimental conditions (1-D–specific, 1-D–general, 4-D–specific, and 4-D–general) were trained separately. In the studies in which subjects have been trained at multiple levels, they have typically been trained on one level at a time (e.g., Murphy, 1991; Murphy & Smith, 1982). In addition to posing difficulties for subjects, it is not obvious how to instantiate this segmented training regimen in the models that were evaluated. Although the hierarchical levels existed only virtually, in the sense that no subject learned more than one level throughout the experiment, any empirical differences between levels that are observed would still be challenging for the models to reproduce. If the models cannot account for differential learning of categories that exist in virtual hierarchies, it seems less likely that they will be able to account for naturally learned hierarchies either.

In the following experiment, a standard category learning paradigm was used. On every trial, a stimulus was presented, and the subject classified it into either one of two categories (subjects learning at the general level) or one of four categories (subjects learning at the specific level). Of particular interest was how classification accuracy would change with training as a function of which category structure subjects learned (1-D vs. 4-D) and at what level they learned to classify (specific vs. general).

METHOD

Subjects

The subjects were 100 undergraduates who voluntarily participated as part of an introductory psychology course.

Stimuli

The stimuli were computer-generated line drawings of rocketships varying in the shape of the wing, nose, porthole, and tail (closely modeled after stimuli originally used by Hoffman & Ziessler, 1983; see also Anderson, 1990). Each dimension had four possible values.

At the general level, rocketships were divided into two categories; at the specific level, rocketships were divided into four categories. As shown in Table 1, the two different category structures, 1-D and 4-D, differed in how the most relevant information was distributed across the four dimensions. For the 1-D structure, diagnostic features were contained along D1. For the 4-D structure, di-

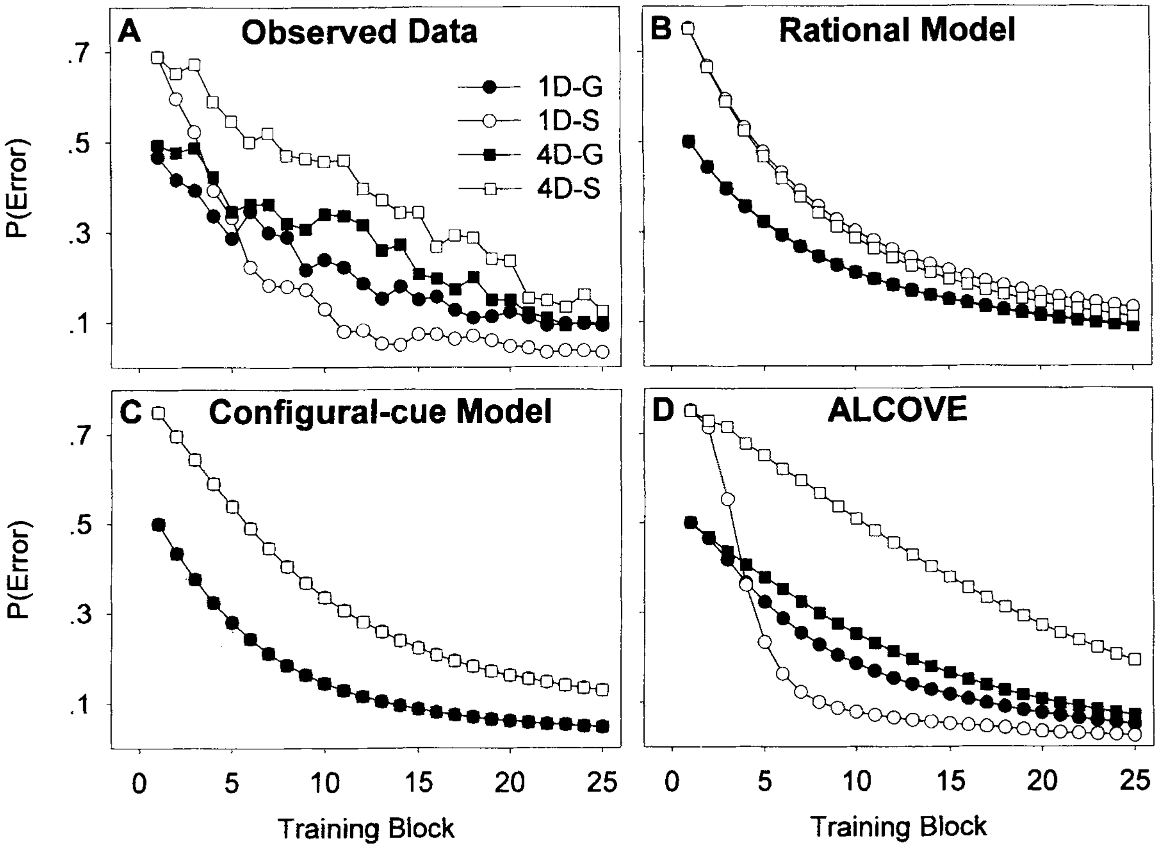

Figure 1. Average probability of error as a function of the number of training blocks and condition. Specific level of classification (S) is indicated by open symbols, and general level of classification (G) is indicated by filled symbols. The 1-D category structure is indicated by circles, and the 4-D category structure is indicated by squares. Panel A displays the observed data, panel B displays the predictions by the rational model, panel C displays the predictions by the configural-cue model, and panel D displays the predictions by ALCOVE.

agnostic features were distributed across all four dimensions. In particular, at the specific level, for the 1-D structure, Values 0, 1, 2, and 3 along D1 signaled Categories C, D, E, and F, respectively; for the 4-D structure, at the specific level, Value 0 along Dimensions D1, D2, D3, and D4, signaled Categories C, D, E, and F, respectively. Both category structure (1-D vs. 4-D) and category level (specific vs. general) were manipulated between subjects.

The assignments of physical dimensions and features to abstract dimensions and features were randomized for every subject. For example, a given subject learning the 1-D–specific categories might need to learn that a particular shape of the wing was associated with each category; a given subject learning the 1-D–general categories might need to learn that two different shapes of the wing were associated with each category. A given subject learning the 4-D–specific categories might need to learn that a particular shape of the wing was associated with the first category, that a particular shape of the nose was associated with the second category, that a particular shape of the porthole was associated with the third category, and that a particular shape of the tail was associated with the fourth category; a given subject learning the 4-D–general categories might need to learn that a particular shape of the wing or a particular shape of the nose was associated with the first category and that a particular shape of the porthole or a particular shape of the tail was associated with the second category.

Procedure

A standard supervised category learning procedure was used in which the subjects were supplied with corrective feedback after every response. Half of the subjects learned category structure 1-D, whereas the other half learned category structure 4-D. For the subjects learning each category structure, half of them learned to classify each of the 12 stimuli into one of two categories (general level), whereas the other half learned to classify each of the 12 stimuli into one of four categories (specific level). Each stimulus was presented once per block for a total of 25 training blocks. On every trial, a randomly chosen stimulus was presented, the subject classified that stimulus into either one of two possible categories (those learning at the general level) or one of four possible categories (those learning at the specific level), and then corrective feedback was supplied for 1 sec. The learning trials were terminated when the subject completed two error-free training blocks.

EMPIRICAL RESULTS AND DISCUSSION

Figure 1A displays classification error probabilities as a function of training. First, error rates decreased as a function of training. Second, more errors were made by subjects given the 4-D structure (square symbols) than those given the 1-D structure (circle symbols). Third, category level interacted with category structure: For the 1-D structure, the specific level (open circles) was learned more rapidly than the general level (filled circles); for the 4-D structure, the general level (filled squares) was learned more rapidly than the specific level (open squares). Using a very different learning procedure from the one used by Lassaline et al. (1992), a very similar interaction between category level and category structure was observed.

A 2 (1-D vs. 4-D) × 2 (general vs. specific) × 25 (training block) analysis of variance was conducted on the data, with category structure and category level as between-subjects factors and training block as a within-subjects factor.[2] An alpha level of .05 was set for all statistical tests. Corroborating the above impressions, significant main effects of category structure and training block were found [$F(1,89) = 15.30$, $MS_e = 0.54$, and $F(24,2136) = 98.69$, $MS_e = 0.02$, respectively]. A significant two-way category structure × category level interaction reflected the more rapid learning of the specific level than the general level for the 1-D structure and the more rapid learning of the general level than the specific level for the 4-D structure [$F(1,89) = 4.43$, $MS_e = 0.54$]. A significant two-way category structure × training block interaction reflected the quicker learning of the 1-D structure than the 4-D structure [$F(24,2136) = 3.51$, $MS_e = 0.02$]. Finally, a significant two-way category level × training block interaction partially reflected the initial advantage of general level over specific level due to the different number of response categories [$F(24,2136) = 5.63$, $MS_e = 0.02$].

OVERVIEW OF THE CATEGORY LEARNING MODELS AND THEIR PREDICTIONS

I will summarize the key aspects of the three category learning models, discuss some possible expectations for how well the models might account for the observed data, and then summarize the actual fits of the models to the observed data. More details of the model fitting are provided in the Appendix.

Rational Model

According to the rational model (Anderson, 1990, 1992), classification involves a Bayesian statistical analysis of the environment. Internal representations of subcategories, or partitions, are created to the extent that objects in the world are divided up into disjoint sets whose members probabilistically share certain features (cf. Rosch et al., 1976). Partitions are similar to prototypes in that they may be abstractions of a number of specific instances; a single category in the world may be represented as one or more internally defined partitions. The probability of classifying an object as a member of some particular category is essentially a function of the similarity of that object to the central tendency of each partition weighted by how likely that particular category label is associated with objects contained within the partition. Anderson (1990) applied the rational model to experiments examining basic-level effects that were similar to the present one (e.g., Hoffman & Ziessler, 1983; Murphy & Smith, 1982). Although the model was not actually fitted to experimental data, the rational model did show a strong preference for creating partitions at the experimentally determined basic level—that is, each basic-level category was represented by a single partition. Thus, it may seem reasonable to expect the model to predict a specific-level advantage in the present task. In addition, although the rational model does not assume any mechanism of selective attention[3] to psychological dimensions,[4] Anderson (1990, 1992) applied the model to classification tasks in which certain dimensions were highly diagnostic for determining category membership (e.g., Medin & Schaffer, 1978; Shepard, Hovland, & Jenkins, 1961). Thus, it may seem reasonable to expect the model to predict a 1-D advantage as well.

The best-fitting predicted learning curves of the rational model are shown in Figure 1B. Contrary to the intuitive predictions generated above, the qualitative fit of the model was quite poor. The model was unable to predict any specific-level advantage and was essentially unable to predict any difference between the 1-D and 4-D conditions. It should be noted, however, that although the model-fitting routine settled on parameters that maximized quantitative fit, the best-fitting parameter values were fairly extreme (see the Appendix). In essence, fit was maximized when the rational model created individual partitions for every stimulus, effectively reducing to a pure exemplar-based model (see Nosofsky, 1991).

Because these shortcomings of the rational model may be partially a product of the demands of maximizing the quantitative fit of model, it seemed important to explore the qualitative predictions of the model using more reasonable parameters similar to what was used in previous research (Anderson, 1990; Nosofsky, 1991).[5] In these new simulations, the model still predicted essentially no difference between the 1-D and 4-D conditions and did not predict a crossover in learning the specific- and general-level categories. Examination of the partitions that were formed was particularly revealing. When learning categories at a general level, partition formation was fairly idiosyncratic: Depending on the particular sequence of training stimuli, between three and six partitions were formed for each category. For example, in one particular sequence, when learning the 1-D structure, Stimuli 1, 3, and 5 formed one partition, Stimuli 2 and 6 formed another partition, and Stimulus 4 formed its own partition; in one particular sequence, when learning the 4-D structure, Stimuli 3, 4, and 5 formed one partition, while Stimuli 1, 2, and 6 formed their own separate partitions. By contrast, when learning categories at a specific level, all stimuli within a given category were grouped together into a single partition; however, the formation of such partitions was insensitive to whether the diagnostic features were present along a single dimension (1-D) or present along various dimensions (4-D). So the rational model did indeed create consistent partitions when learning categories at the specific level. Unfortunately, although such partitioning may suggest some sort of an advantage for "basic-level" categories, the model did not effectively utilize this advantageous partitioning to predict these categories to be more quickly learned than categories at higher levels of the hierarchy.

To assess whether the model could possibly account for a specific-level advantage when learning only the 1-D structure, the model was fitted to the 1-D data alone—it was not possible to find parameters that permitted the model to predict the observed crossover between specific- and general-level categories.

Configural-Cue Model

The configural-cue model (Gluck & Bower, 1988a) is a two-layer connectionist model of category learning. The input layer contains a single node for every individual cue (i.e., particular values along psychological dimensions) and combination of cues (configural cues) that compose an item. The output layer contains a single node for every category. Association weights are learned between cues and categories via gradient descent on error. Gluck, Corter, and Bower (1996) applied the configural-cue model to existing data (Hoffman & Ziessler, 1983; Murphy & Smith, 1982) and new experiments examining basic-level effects in artificial categories. Thus, it may seem reasonable to expect the model to predict a specific-level advantage in the present results. Also, although the configural-cue model does not incorporate selective attention to psychological dimensions (see note 3), Gluck and Bower (1988a) provided simulations demonstrating the effectiveness of the model in accounting for classic attentional phenomena in classification. Thus, it may seem reasonable to expect the model to predict a 1-D advantage as well.

The best-fitting predictions of the configural-cue model are shown in Figure 1C. The fit was quite poor—the model failed to predict a specific-level advantage, and the predicted learning curves for the 1-D and 4-D conditions overlapped completely. But could the model account for the observed specific-level advantage if it was fitted only to the data from the 1-D condition? In contrast to what was reported by Gluck et al. (1996), using different category structures, it was not possible to find parameter values that allowed the model to predict the observed crossover of the specific- and general-level conditions.

Why did the configural-cue model fail to account for the observed difficulty of learning the 4-D category structure? According to the model, a cue is any specific feature value, such as the particular shape of the tail or the particular shape of the porthole of a rocketship stimulus. The model assumes that people learn the relevance of such cues, and combinations of cues, for making category decisions. However, no distinction is made between learning the relevance of two cues along the same psychological dimension (e.g., a circular vs. a rectangular porthole) versus two cues along different psychological dimensions (e.g., a circular porthole vs. a triangular wing). Psychological dimensions, per se, do not exist in the model, only featural cues and cue combinations. By contrast, it is likely that people might learn the relevance of psychological dimensions as well as the relevance of particular values along those dimensions.

ALCOVE

ALCOVE is a connectionist instantiation of an exemplar model of classification, the generalized context model (GCM; Medin & Schaffer, 1978; Nosofsky, 1986). According to exemplar models, categories are represented in terms of the individual remembered exemplars. ALCOVE assumes that classification decisions are determined by the similarity of a target item to each remembered exemplar and by the learned association strength between each exemplar and each category. A fundamental property of ALCOVE is that psychological dimensions can be learned to be selectively attended to according to their diagnosticity. Selective attention acts to "stretch" the psychological space along relevant dimensions and "shrink" it along irrelevant dimensions. For example, if stimuli varied in shape, size, and color, but shape was particularly diagnostic for deciding which category a stimulus belonged in, then differences along the shape dimension would be accentuated, whereas differences along the size and color dimensions would be attenuated. Because ALCOVE learns to attend to dimensions based on their diagnosticity, it may seem reasonable to expect the model to predict a 1-D advantage in the present results.

However, one generally acknowledged shortcoming of many exemplar models is that, in a typical category learning paradigm, they cannot predict classification of objects at lower levels of a nested category hierarchy to ever be superior to classification of those same objects at higher levels of a hierarchy (e.g., Lassaline et al., 1992)—they fail to predict a "basic-level" advantage. To illustrate, suppose that at the general level there are two categories, A and B, and that these two categories can be divided up at the specific level into two categories each, C and D for Category A, and E and F for Category B. According to the context model (Medin & Schaffer, 1978; Nosofsky, 1986), the evidence that some item belongs to a category is found by summing up the similarities of that item to all exemplars of the category. The probability of classifying an item as a member of a category is given by the ratio of the evidence for that category to the total evidence for all categories. For example, when classifying at the general level $P(\mathrm{A}) = E_\mathrm{A}/(E_\mathrm{A} + E_\mathrm{B})$, and when classifying at the specific level $P(\mathrm{C}) = E_\mathrm{C}/(E_\mathrm{C} + E_\mathrm{D} + E_\mathrm{E} + E_\mathrm{F})$, where E_X is the evidence (summed similarity) for Category X (where X can be Category A, B, C, D, E, or F). If categories are represented in terms of stored exemplars, then categories at higher levels of a hierarchy are simply the union of the exemplars of categories at lower levels of the hierarchy. This means that the summed similarities to categories at higher levels of a hierarchy are simply equal to the sum of the summed similarities to categories at lower levels; for example, $E_\mathrm{A} = E_\mathrm{C} + E_\mathrm{D}$. In computing classification response probabilities, the denominator in the ratios are identical at the general level and the specific level, and the numerator is greater at the general level than at the specific level, so classification probabilities are constrained to be less accurate at the specific level than the general level—the context model fails to predict a "basic-level" advantage.

ALCOVE is not necessarily constrained to predict a general-level advantage for several reasons. First, because ALCOVE learns to attend to dimensions based on their diagnosticity, the similarity between an item and an exemplar in memory can depend on the category level that was learned. Therefore, evidence for classification at a higher level of a hierarchy is not a simple linear combination of evidences at a lower level. Second, ALCOVE learns to associate exemplars with categories via a connectionist error-driven learning algorithm. It is possible for association weights to be larger for categories learned at the specific level than at the general level. By contrast, the context model just tallies the number of times each exemplar has been associated with a given category. Finally, the response rule used by ALCOVE to map category activations to response probabilities is highly nonlinear. Unlike the context model, evidence for classification at higher levels of a hierarchy is not simply the additive combination of evidences at lower levels. In summary, ALCOVE is not subject to the same constraints as the context model. Yet, explicit simulations of the model are necessary to assess whether the model can predict the observed specific level.

The best-fitting predictions of ALCOVE are shown in Figure 1D. The quantitative fit of the model was quite good (see Appendix). More importantly, the model was able to account for all of the qualitative results: ALCOVE predicted more rapid learning of the 1-D structure over the 4-D structure and was able to predict the observed crossover interaction of category structure with category level.

Not surprisingly, dimensionalized selective attention learning is critical for allowing ALCOVE to predict the observed 1-D over 4-D category structure advantage. To demonstrate this, a restricted version of ALCOVE without allowing for learned selective attention was fitted to the observed data. As was the case for the configural-cue model, this restricted version of ALCOVE predicted absolutely no difference in learning the 1-D and 4-D category structures.[6] This finding adds additional support for the theoretical claim that dimensionalized selective attention is a critical component of category learning (see Kruschke, 1992; Nosofsky, 1986; Nosofsky, Gluck, et al., 1994; Nosofsky & Palmeri, 1996). Essentially, ALCOVE captures the notion that people generally find it is easier to learn to pay attention to differences along a single psychological dimension than differences along multiple psychological dimensions.

Why can ALCOVE predict a specific-level advantage for the 1-D structure when the context model failed? As discussed above, multiple factors relieve ALCOVE from the constraint of assuming evidence for classification at the higher level of a hierarchy to be a simple linear summation of evidences at lower levels. In breaking this constraint, dimensionalized selective attention is clearly important—a restricted version of ALCOVE without attention could not predict a specific-level advantage. For the 1-D structure, when trained on the specific-level categories, ALCOVE learned to attend solely to Dimension 1 because Dimensions 2–4 were completely nondiagnostic. By contrast, when trained on the general-level categories, ALCOVE learned to attend to Dimensions 2–4 because they were somewhat diagnostic. These differences in selective attention cause the relative evidence for specific-level classification to be greater than the relative evidence for general-level classification. Error-driven learning and nonlinear response rules are also very important—restricted versions of ALCOVE with Hebbian (correlational) learning and a linear response mapping rule could not predict a specific-level advantage either. Accordingly, when provided the 1-D structure, ALCOVE developed stronger association weights from exemplars to categories when trained at the specific level than when trained at the general level.

SUMMARY

The present article reported an extension of an experiment by Lassaline et al. (1992). For one category structure in which diagnostic information was present along a single dimension, specific-level categories were learned more rapidly than general-level categories; for another category structure in which diagnostic information was spread across dimensions, the reverse pattern of results was found.

Predicting this interaction of category structure and category level proved quite challenging. Neither the rational model (Anderson, 1990) nor the configural-cue model (Gluck & Bower, 1988a) was able to predict a difference in learning the two category structures, largely because neither model incorporates dimensionalized selective attention (see Nosofsky, Gluck, et al., 1994; Nosofsky & Palmeri, 1996). Also, neither model was able to predict a specific-level advantage for the 1-D category structure. Although the rational model can indeed produce partitions at the specific level when learning specific-level categories (Anderson, 1990), the model was un-able to take advantage of this partitioning to predict a specific-level advantage in category learning. Although the configural-cue model showed some promise in predicting basic-level effects in previous work (Gluck et al., 1996), the model was unable to account for the present results. By contrast, ALCOVE (Kruschke, 1992) was able to account for the interaction of category structure and category level. The presence of dimensionalized selective attention allowed the model to predict faster learning of the 1-D structure than of the 4-D structure. A nonlinear response mapping rule, differences in association weights when learning at different category levels, and differences in learned selective attention weights when learning at different category levels all contributed to ALCOVE predicting a specific-level advantage when learning the 1-D structure. The surprising failure of the rational model and the configural-cue model and the surprising success of ALCOVE highlight the importance of carrying out explicit simulations. Hintzman (1990) noted the difficulties in predicting the behavior of complex psychological models based on some a prior understanding of the models: "Surprises are likely when the model has properties that are inherently difficult to understand, such as variability, parallelism, and nonlinearity—all, undoubtedly, properties of the brain" (p. 111).

It should again be emphasized that the present empirical and theoretical work is limited in that the category hierarchies existed only virtually, in the sense that individual subjects never learned categories at more than one level. However, the model-based analyses are still relevant in pointing out important limitations of the rational model and the configural-cue model in accounting for observed patterns of categorization behavior, irrespective of whether these particular empirical results say anything about learning natural categories at multiple levels of a hierarchy. In extending these results to the more general issue of how people might actually learn natural category hierarchies, the present work highlights the possibility that learning categories at different hierarchical levels may require attending to the psychological dimensions of stimuli in very different ways. That these patterns of selective attention to dimensions seem to vary as a function of category level may help explain why it has proved so difficult to train people to classify stimuli at multiple levels of a category hierarchy at the same time (Lassaline et al., 1992; Murphy, 1991; Murphy & Smith, 1982). But these varying patterns of selective attention to dimensions also point out an important limitation of ALCOVE. If ALCOVE is to account for natural situations in which the same individual learns categories at multiple levels, then there must be inclusion of some mechanisms for remembering specific patterns of selective attention weights and setting them to their appropriate values depending on the given category context. Future research will be needed to extend ALCOVE to account for learning multiple category levels at the same time and to then test whether ALCOVE can explain basic-level effects in other classification situations.

REFERENCES

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1992). The adaptive nature of human categorization. *Psychological Review*, **98**, 409-429.

Ashby, F. G., & Maddox, W. T. (1993). Relations between prototype, exemplar, and decision bound models of categorization. *Journal of Mathematical Psychology*, **37**, 372-400.

Estes, W. K. (1986). Memory storage and retrieval processes in category learning. *Journal of Experimental Psychology: General*, **115**, 155-174.

Estes, W. K. (1993). Models of categorization and category learning. *Psychology of Learning & Motivation*, **29**, 15-56.

Estes, W. K. (1994). *Classification and cognition.* Oxford: Oxford University Press.

Estes, W. K., Campbell, J. A., Hatsopoulos, N., & Hurwitz, J. B. (1989). Base-rate effects in category learning: A comparison of parallel network and memory storage-retrieval models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 556-571.

Fisher, D., & Langley, P. (1990). The structure and formation of natural categories. *Psychology of Learning & Motivation*, **26**, 241-284.

Garner, W. R. (1974). *The processing of information and structure.* New York: Wiley.

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

Gluck, M. A., & Bower, G. H. (1988b). From conditioning to category learning: An adaptive network model. *Journal of Experimental Psychology: General*, **117**, 225-244.

Gluck, M. A., & Corter, J. E. (1985). Information, uncertainty, and the utility of categories. In *Proceedings of the Seventh Annual Conference of the Cognitive Science Society* (pp. 283-288). Hillsdale, NJ: Erlbaum.

Gluck, M. A., Corter, J. E., & Bower, G. H. (1996). *Basic levels in learning category hierarchies: An adaptive network model.* Unpublished manuscript.

Hintzman, D. L. (1990). Human learning and memory: Connections and dissociations. *Annual Review of Psychology*, **41**, 109-139.

Hoffman, J., & Ziessler, C. (1983). Objectidentifikation in kunstlichen Begriffshierarchien [Object identification in artificial concept hierarchies]. *Zeitscrift für Psychologie*, **16**, 243-275.

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

Lassaline, M. E. (1990). *The basic level in hierarchical classification.* Unpublished master's thesis, University of Illinois, Champaign.

Lassaline, M. E., Wisniewski, E. J., & Medin, D. L. (1992). Basic levels in artificial and natural categories: Are all basic levels created equal? In B. Burns (Ed.), *Percepts, concepts, and categories: The representation and processing of information* (pp. 327-378). Amsterdam: North-Holland.

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

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review*, **85**, 207-238.

Murphy, G. L. (1991). Parts in objects concepts: Experiments with artificial categories. *Memory & Cognition*, **19**, 423-438.

Murphy, G. L., & Smith, E. E. (1982). Basic level superiority in picture categorization. *Journal of Verbal Learning & Verbal Behavior*, **21**, 1-20.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General*, **115**, 39-57.

Nosofsky, R. M. (1991). Relations between the rational model and the context model of categorization. *Psychological Science*, **2**, 416-421.

Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition*, **22**, 352-369.

Nosofsky, R. M., Kruschke, J. K., & McKinley, S. C. (1992). Combining exemplar-based category representations and connectionist learning rules. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 211-233.

Nosofsky, R. M., & Palmeri, T. J. (1996). Learning to classify integral-dimension stimuli. *Psychonomic Bulletin & Review*, **3**, 222-226.

Nosofsky, R. M., & Palmeri, T. J. (1997). An exemplar-based random walk model of speeded classification. *Psychological Review*, **104**, 266-300.

Nosofsky, R. M., & Palmeri, T. J. (1998). A rule-plus-exception model for classifying objects in continuous-dimension spaces. *Psychonomic Bulletin & Review*, **5**, 345-369.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review*, **101**, 53-79.

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 324-354.

Palmeri, T. J., & Nosofsky, R. M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experiment Psychology: Learning, Memory, & Cognition*, **21**, 548-568.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, **8**, 382-439.

Schyns, P. G. (1991). A modular neural network model of concept acquisition. *Cognitive Science*, **15**, 461-508.

Shepard, R. N., Hovland, C. L., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, **75**(13, Whole No. 517).

NOTES

1. Two main forms of category learning have been investigated. In *supervised learning* tasks, explicit trial-by-trial feedback is supplied about whether particular category responses are correct or incorrect. In *unsupervised learning* tasks, no feedback is supplied, and subjects form their own categories on the basis of some internal criteria for category cohesiveness.

2. Unfortunately, due to a procedural error, 7 of the 25 subjects in the 4-D–general condition did not finish all 25 training blocks (none of these subjects completed fewer than 20 training blocks). Although their available data were included in the data plotted in Figure 1, which was used to assess the model predictions, our statistical package could not accommodate partial observations from individual subjects.

3. A number of categorization studies have found that the diagnosticity of certain stimulus dimensions has a profound effect on the speed of learning particular categories (e.g., Nosofsky, Gluck, et al., 1994; Shepard, Hovland, & Jenkins, 1961) and on transfer of category knowledge to new stimuli (e.g., Medin & Schaffer, 1978; Nosofsky, 1986). These results have been taken as evidence for some form of a dimensionalized selective attention mechanism in categorization (in other words, particular dimensions of stimuli are weighted more heavily because they are relatively more diagnostic for determining category membership). However, the rational model and the configural-cue model, which do not incorporate any explicit form of selective attention to dimensions, have been able to account for some of these empirical results (see, however, Nosofsky, Gluck, et al., 1994).

4. Note that *psychological dimensions*, such as shape, color, or size, are contrasted with particular values (or features) along those dimensions, such as circular, red, or large (see Garner, 1974).

5. Typically, the coupling parameter is set to some intermediate value, dimensional salience is significantly larger than label salience, and the response-mapping parameter is not used (Anderson, 1990; Nosofsky, 1991). These simulations were conducted with $c = 0.3$, $s_D = 1.0$, $s_L = 0.1$, and $r = 1.0$.

6. The fact that this restricted version of ALCOVE and the best-fitting version of the configural-cue model accounted for the observed data equally poorly is not simply a matter of coincidence. Rather, a version of ALCOVE without selective attention is formally identical to a version of the configural-cue model with only complete exemplar cue combinations present; the best-fitting parameters of the configural-cue model had nonzero learning rates only for complete exemplars (λ_4).

APPENDIX
Details of the Category Learning Models and Model Fitting

In this Appendix, I provide additional details about the three category learning models and how they were fitted to the observed learning curves. Best-fitting predictions from each of the three category learning models were generated by adjusting free parameters of the models using a hill-climbing routine that minimized sum of squared deviations (*SSD*) between observations and predictions. To guard against local minima emerging from the hill-climbing routine, a number of different starting parameter values were used to initialize the search. To fit each of the models, 100 random stimulus sequences were generated, predictions were obtained for each of these sequences, and these predictions were then averaged. These average predicted category learning curves constituted the predictions that were fitted to the observed data. The best-fitting parameters and fit values for the three models are given in Table A1.

Rational Model

According to the rational model, categories are learned by grouping objects into partitions. The probability that some object joins an existing partition is a function of both the similarity of that exemplar to the partition's central tendency and the prior probability of the partition. The prior probability of a partition is jointly determined by the size of the partition and by the value of a coupling parameter, c, which is a free parameter; large values of c produce large partitions, and small values of c produce small partitions. In most applications, the coupling parameter has an intermediate value (Anderson, 1990, used $c = 0.3$). The similarity of an exemplar to a partition's central tendency is found using a multiplicative similarity rule somewhat analogous to the similarity rule used in ALCOVE and the context model (Nosofsky, 1991). This similarity is jointly determined by whether the dimensions of the object sufficiently match those stored in the partitions and priors specified by salience terms for stimulus dimensions and category labels, s_D and s_L, respectively, which are free parameters; unlike other classification models, the category label is treated as just another stimulus dimension in the stored representation. In most applications, dimensional salience is significantly larger than label

Table A1
Best-Fitting Model Parameters

Model	Parameters	Fit
Rational	$c = 0.001$, $s_D = 0.125$, $s_L = 0.000$, $r = 1.486$	*SSD* = 0.958, *RMSD* = 0.098, Var = 67.2
Configural-cue	$\lambda_1 = 0.000$, $\lambda_2 = 0.000$, $\lambda_3 = 0.000$, $\lambda_4 = 0.074$, $\phi = 1.788$	*SSD* = 1.318, *RMSD* = 0.115, Var = 54.9
ALCOVE	$c = 7.440$, $\lambda_w = 0.017$, $\lambda_\alpha = 0.712$, $\phi = 3.746$	*SSD* = 0.228, *RMSD* = 0.054, Var = 90.1

Note—*SSD*, sum of squared deviations; *RMSD* = root mean squared deviations; Var, percentage of variance accounted for.

APPENDIX (Continued)

salience (Anderson, 1990; Nosofsky, 1991). The probability, p_A, that a given category label is assigned to an object is found by summing the probability that the object belongs to each partition multiplied by the probability that each partition signals that category label. A response-mapping parameter, r, transforms the internal category label probabilities, p_A, into actual response probabilities, $P(A)$. Given a category label probability, p_A, the actual probability of a Category A response is given by

$$P(A) = \frac{p_A^r}{\sum_C p_C^r},$$

where the subscript C ranges over all possible categories. This mapping function allows probabilities more or less extreme than those ordinarily predicted by the rational model (see Nosofsky et al., 1994). In the present application, there are four estimated parameters: the coupling parameter, c, the dimensional salience, s_D, the label salience, s_L, and the response-mapping parameter, r.

Configural-Cue Model

Inputs to the configural-cue model are individual cues and configural cues that comprise a given object. With four dimensions, each having four possible values, there are 16 single nodes (4 dimension × 4 features), 96 possible double nodes, 256 possible triple nodes, and 256 quadruple nodes (one for each possible exemplar). The activation of each input node, a_i, is set equal to one if the relevant configuration is present within a stimulus, otherwise it is set equal to zero. The activation of a category output node is given by

$$O_A = \sum_i w_{ia} a_i,$$

where w_{iA} is the learned association weight between cue i and Category A. Output activations are converted into response probabilities by

$$P(A) = \frac{\exp(\phi O_A)}{\sum_C \exp(\phi O_C)},$$

where the subscript C ranges over all possible categories and ϕ is a response mapping constant. In the present application, there are five estimated parameters: the response mapping constant (ϕ), and learning rates (λ_1, λ_2, λ_3, and λ_4), for updating the association weights between singles, doubles, triples, and quadruples and the category output nodes, respectively.

ALCOVE

Formally, ALCOVE is a three-layer feedforward network. The input layer consists of a single node for every psychological dimension of an object. The hidden layer consists of a single node for every stored exemplar. The activation of each hidden node is a function of the similarity between the current input representation and the exemplar representation of that node and is given by

$$a_j^{hid} = \exp\left(-c \sum_i \alpha_i d_{ji}\right),$$

where α_i is the learned selective attention to dimension i, and d_{ji} is an indicator variable equal to zero if the input stimulus and the exemplar match along dimension i and equal to 1 if they mismatch. The positive constant c, called the *specificity*, acts as a scaling factor. Every hidden node is connected to every category output node via a learned association weight. The activation of Category node A is given by

$$O_A = \sum_j w_{jA} a_j^{hid},$$

where w_{jA} is the learned association weight. Output activations are converted into response probabilities by

$$P(A) = \frac{\exp(\phi O_A)}{\sum_C \exp(\phi O_C)},$$

where the subscript C ranges over all possible categories, and ϕ is a response-mapping constant. In the present applications, there were four estimated parameters: the sensitivity parameter (c), the response-mapping constant (ϕ), and learning rates (λ_w and λ_α), for updating the exemplar association weights and selective attention weights (see Kruschke, 1992, for details).

(Manuscript received March 19, 1998;
revision accepted for publication December 9, 1998.)