Extending the ALCOVE Model of Category Learning to Featural Stimulus Domains

Michael D. Lee*
Communications Division
Defence Science and Technology Organisation

Daniel J. Navarro
Department of Psychology
University of Adelaide

Abstract

The ALCOVE model of category learning, despite its considerable success in accounting for human performance across a wide range of empirical tasks, is limited by its reliance on spatial stimulus representations. Some stimulus domains are better suited to featural representation, characterizing stimuli in terms of the presence or absence of discrete features, rather than as points in a multidimensional space. We report on empirical data measuring human categorization performance across a featural stimulus domain, and show that ALCOVE is unable to capture fundamental qualitative aspects of this performance. In response, a featural version of the ALCOVE model is developed, replacing the spatial stimulus representations that are usually generated by multidimensional scaling with featural representations generated by additive clustering. We demonstrate that this featural version of ALCOVE is able to capture human performance where the spatial model failed, and provide an explanation for the difference in terms of the different representational assumptions made by the two approaches. Finally, we discuss ways in which the ALCOVE categorization model might be extended further to use 'hybrid' representational structures combining spatial and featural components.

Introduction

The connectionist model of category learning known as ALCOVE (Kruschke 1992) is one of the most successful and widely used formal models in cognitive psychology. While there are some category learning effects that ALCOVE does not capture without modification or extension (e.g., Kruschke & Erikson 1995), its original formulation remains a simple and powerful account of a wide variety of categorization behavior. The most significant shortcoming of ALCOVE is that originally noted by Kruschke (1992, p. 40): "ALCOVE applies only to situations for which the stimuli can appropriately be represented as points in multidimensional psychological similarity space". Within cognitive psychology, it has often been argued (e.g., Tversky 1977) that many important stimulus domains are not amenable to spatial representation, but instead require a featural approach to representation. Motivated by a concrete example of a domain in which ALCOVE fails, apparently because of its spatial representation, the goal of this article is to

*This work was supported by a DSTO scholarship awarded to the second author. We wish to thank Simon Dennis, Robert Goldstone, John Kruschke, Robert Nosofsky, Douglas Vickers, Michael Webb, Chris Woodruff and several anonymous referees for their helpful comments on earlier versions of this article. Correspondence concerning this article should be addressed to: Michael D. Lee, Communications Division, Defence Science and Technology Organisation, PO Box 1500, Salisbury SA 5108 AUSTRALIA. Telephone: +61 8 8259 5821, Facsimile: +61 8 8259 7110, Electronic Mail: michael.d.lee@dsto.defence.gov.au

extend ALCOVE to accommodate stimulus domains that are represented in terms of the presence or absence of a set of discrete domain features.

The ALCOVE Model

In this section we summarize the way in which ALCOVE internally represents a stimulus set, categorizes a presented stimulus, and then learns from externally provided feedback that specifies whether or not the categorization decision was correct. A more detailed description of ALCOVE may be found in Kruschke (1992).

Categorization

Stimulus Representation The spatial representations used by ALCOVE locate each of the n stimuli at a point in an m dimensional space, as determined by multidimensional scaling or some other equivalent procedure. We denote the representative point for the ith stimulus by $\mathbf{f}_i = (f_{i1}, \ldots, f_{im})$.

Stimulus Comparison On each categorization trial, a stimulus is presented to ALCOVE, and its attention-weighted distance to each of the other stimuli is calculated. Although ALCOVE has been applied successfully to both integral and separable stimulus domains, we restrict ourselves to describing the separable case, since our concern is with the extension of ALCOVE to featural representations of stimuli. Any stimulus domain amenable to a featural characterization seems likely to contain dimensions that can be attended to individually, and may be regarded as separable within Garner's (1974) framework.

Accordingly, if the ith stimulus is presented, its distance to the jth stimulus, denoted d_{ij}, is given by the attention weighted city-block distance between their representative points:

$$d_{ij} = \sum_k a_k \left| f_{ik} - f_{jk} \right|, \tag{1}$$

where a_k is the attention weight applied to the kth dimension.

Generalization Gradient The distances between the presented stimulus, and the other stimuli are then transformed to similarities, denoted s_{ij}, using the exponential decay relationship advocated by Shepard (1987):

$$s_{ij} = \exp\left(-\sigma \sum_k a_k \left| f_{ik} - f_{jk} \right| \right), \tag{2}$$

where σ is a specificity or resolution parameter associated with the exponential function.

Response Probabilities After calculating these similarities, ALCOVE forms response strengths for each of the possible categories. These are calculated using associative weights maintained between each of the stimuli and the categories. The response strength for the xth category, r_x, is given by the similarity weighted sum of all of the associative weights to that category:

$$r_x = \sum_j w_{xj} s_{ij}, \tag{3}$$

where w_{xj} is the associative weight from the jth stimulus to the xth category.

From the response strengths, ALCOVE generates response probabilities using the choice rule (Luce 1963; Shepard 1957):

$$\Pr(X \mid i) = \frac{\exp(\phi r_X)}{\sum_x \exp(\phi r_x)} \tag{4}$$

where ϕ is a mapping parameter.

Learning

Having produced probabilities for each of the various possible categorization responses, ALCOVE is provided with feedback from an external source. This takes the form of a set of so-called 'humble teacher' values, one for each category, defined as:

$$t_x = \begin{cases} \max(+1, r_x) & \text{if stimulus } i \text{ is in category } x \\ \min(-1, r_x) & \text{otherwise.} \end{cases} \tag{5}$$

Two learning rules are then applied, both derived by seeking to minimize the error measure:

$$E = \frac{1}{2} \sum_x (t_x - r_x)^2 , \tag{6}$$

using a simple gradient descent approach to optimization.

Associative Learning The associative weights between the stimuli and response categories are adjusted using the learning rule:

$$w_{xj}^{new} = w_{xj}^{old} + \lambda_w (t_x - r_x) s_{ij}, \tag{7}$$

where λ_w is the associative learning rate parameter.

Attentional Learning Simultaneously, the attention weights for each dimension of the representational space are adjusted using the learning rule:

$$a_k^{new} = a_k^{old} - \lambda_a \sigma \sum_x (t_x - r_x) \sum_j w_{xj} s_{ij} |f_{ik} - f_{jk}| , \tag{8}$$

where λ_a is the attentional learning rate parameter.

Experiment

In developing a category learning experiment to explore ALCOVE's abilities with a featural stimulus domain, we were guided by a representational observation made by Choi, McDaniel and Busemeyer (1993). After examining the performance of ALCOVE on a set of stimuli varying along the inherently ordinal dimensions of size and number, represented using the spatial approach, they comment that "although this coding seems reasonable for size and number dimensions, it may not work well for color and shape dimensions. (Are triangles and hexagons psychologically twice distant from each other as they are from squares?)" (Note 4). Intuitively, Choi *et. al.* (1993) question the compatability of ALCOVE, because of its reliance on spatial representation, to deal with a domain built from discrete, nominal 'features' rather than continuous, ordered 'dimensions'.

Previous studies (Kruschke 1992; Nosofsky, Gluck, Palmeri, McKinley, & Glauthier 1994) have examined the ability of ALCOVE to model category learning data from the seminal experimental task introduced by Shepard, Hovland and Jenkins (1961), which involves what might be regarded as a 'featural' stimulus domain. This task measured human performance across a series of category structures that divided eight stimuli evenly between two categories. The stimuli were generated by exhaustively varying three binary dimensions such as {black, white}, {small, large} and {square, circle}. While a compelling case has been made (Kruschke 1992) that ALCOVE can capture human category learning on this task, it also the case that the binary featured domain happens to be readily amenable to spatial representation. By introducing an arbitrary ordering for each of the feature values, the stimulus domain can be represented as the vertices of a cube under a distance-based similarity model. This form of representation would not, however, have been possible if a third shape 'triangle' was introduced. This is not to say that a spatial representation would not be possible, but it would need to be a different sort of spatial representation, which may or may not be suited to modeling human categorization behavior. On the basis of these ideas, we chose to examine ALCOVE's performance using a featural stimulus domain

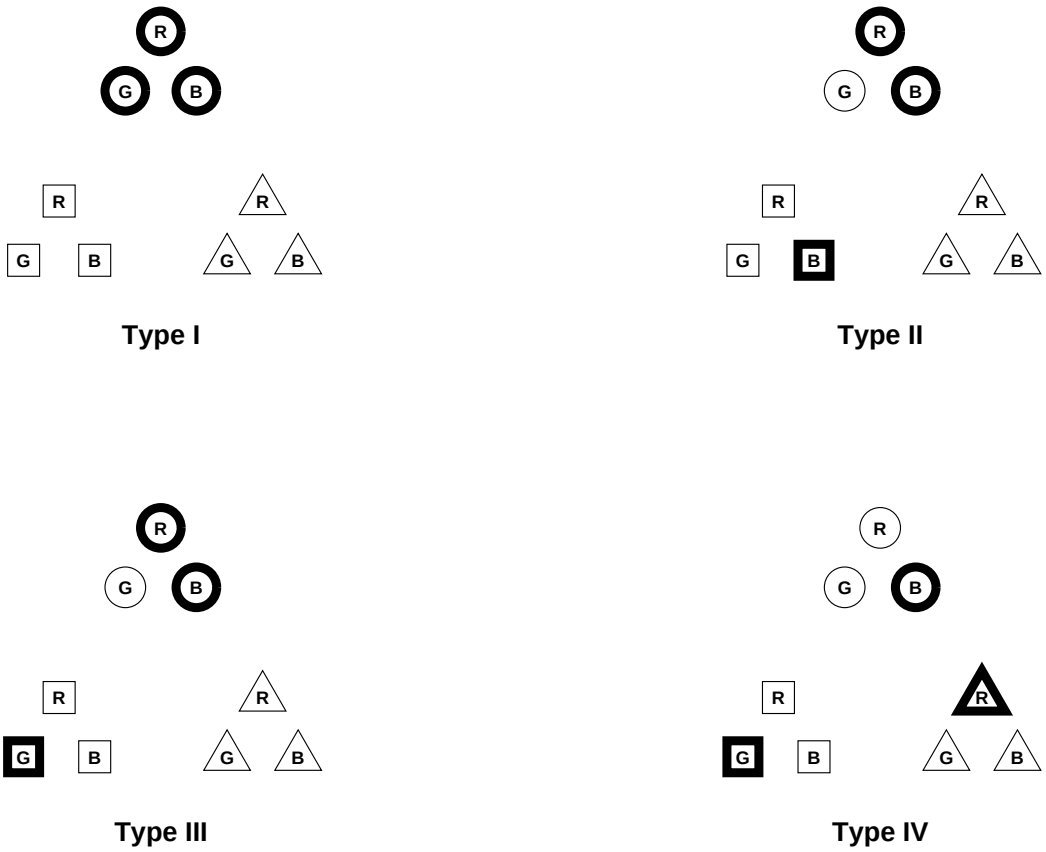


Figure 1: The four different category structures.

obtained by exhaustively combining the set of three colors {red, green, blue} with the set of three shapes {square, circle, triangle}, giving a total of nine stimuli.

The particular category structures we used divided three stimuli into one category, and the remaining six into the other. The fact that a different number of stimuli is assigned to each category is inconvenient, because it potentially introduces issues concerning the base-rate of presentation for each category. Obviously, however, it is not possible to split nine stimuli into two categories evenly, and other experimental variations (such as introducing three possible category responses) seemed to constitute more radical departures from the successful methodology of Shepard *et. al.* (1961).

An analysis of the different category structures with three and six stimuli, allowing for isomorphisms arising from color or shape feature permutation, revealed that there are only four possible types. An example of each of these four category types is shown in Figure 1, which arranges the stimulus domain by forming an outer triangular grouping based on shape, and arranging the colors within these groupings. For each of the four category types, those stimuli belonging to the smaller category are indicated in bold.

Generating A Spatial Representation

Subjects Twenty volunteers served as subjects for collecting the similarity data. There were 19 males and one female, with ages ranging from 25 to 52 years.

Procedure Each subject rated the similarity of all $9 \times 8/2 = 36$ possible pairs of stimuli, presented in a random order, on a 5 point scale. For each presentation of a stimulus pair, the left/right display ordering

Table 1: The final similarity matrix for the stimulus domain.

	red circle	red square	red triangle	green circle	green square	green triangle	blue circle	blue square	blue triangle
red-circle	—								
red-square	0.613	—							
red-triangle	0.638	0.625	—						
green-circle	0.500	0.088	0.063	—					
green-square	0.050	0.550	0.050	0.613	—				
green-triangle	0.063	0.050	0.500	0.638	0.663	—			
blue-circle	0.525	0.063	0.050	0.500	0.125	0.100	—		
blue-square	0.100	0.525	0.088	0.075	0.563	0.088	0.600	—	
blue-triangle	0.088	0.050	0.488	0.088	0.038	0.538	0.588	0.650	—

was also randomly assigned. The final similarity matrix, shown in Table 1, was obtained by averaging across subjects, and made symmetric by transpose averaging.

Results A metric multidimensional scaling algorithm, using the Levenberg-Marquardt approach to non-linear least-squares optimization (More 1977), was used to generate the city-block spatial representation. A particular feature of this multidimensional scaling algorithm is that it automatically determines the appropriate dimensionality of the final solution. This is achieved using the Bayesian Information Criterion (Schwarz 1978) to balance improvements in data-fit with increased model complexity, as described by Lee (in press a). Figure 2 shows the pattern of change in data-fit and the Bayesian Information Criterion across representational spaces with different numbers of dimensions. What these results show is that a four-dimensional spatial representation, explaining 98.8% of the variance in the data, constitutes an appropriate balance between the number of dimensions used, and the level of data-fit achieved.

The coordinate locations of each stimulus for each dimension of this solution are detailed in Table 2, and an attempt to depict the representational space graphically is made in Figure 3. Plotting dimension 1 with dimension 2 shows the subspace of the representation that deals with the different colors of the stimuli. Effectively, each stimulus of the same color is located at the same point in this subspace, and the red, green and blue clusters are arranged in a triangle. This two-dimensional spatial configuration allows each of the three color types to be represented as (approximately) equally similar to the remaining two colors. Plotting dimension 3 with dimension 4 reveals the same representational strategy with respect to the shape component of the stimulus domain. In this subspace, all of the stimuli with the same shape are located at the same point, and the same triangle configuration is evident.

Category Learning

Subjects Twenty-two volunteers served as subjects. There were 14 males and 8 females, with ages ranging from 21 to 48 years.

Procedure Each subject was required to learn an instance of all four category structures, and the order in which the different structures were encountered was chosen randomly. At the beginning of the category learning task, the perceptual display features were also randomly assigned to the logical representational features, as were the two category labels, 'X' and 'Y', so that either could correspond to the smaller category. This meant, for example, that one category within the Type I structure learned by a particular

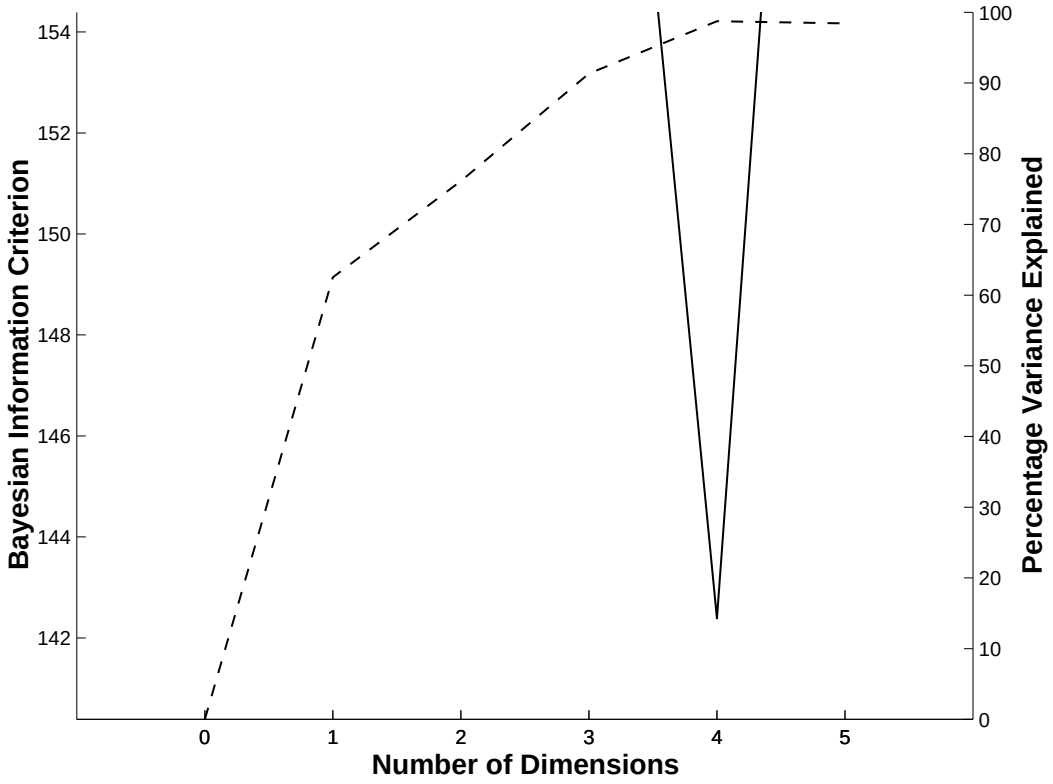


Figure 2: The pattern of change of the Bayesian Information Criterion (left hand scale, solid line), and Percentage Variance Explained (right hand scale, broken line) measures for spatial representations with different dimensionalities, obtained from the similarity data.

Table 2: The city-block multidimensional scaling representation of the stimulus domain.

Stimulus	Dimension 1	Dimension 2	Dimension 3	Dimension 4
red-circle	-0.261	-0.071	-0.205	-0.054
red-square	-0.259	-0.114	0.000	0.107
red-triangle	-0.261	-0.074	0.205	-0.054
green-circle	0.254	-0.080	-0.204	-0.050
green-square	0.254	-0.114	0.000	0.108
green-triangle	0.254	-0.079	0.205	-0.054
blue-circle	-0.013	0.169	-0.206	-0.054
blue-square	-0.001	0.182	0.000	0.108
blue-triangle	-0.004	0.181	0.205	-0.054

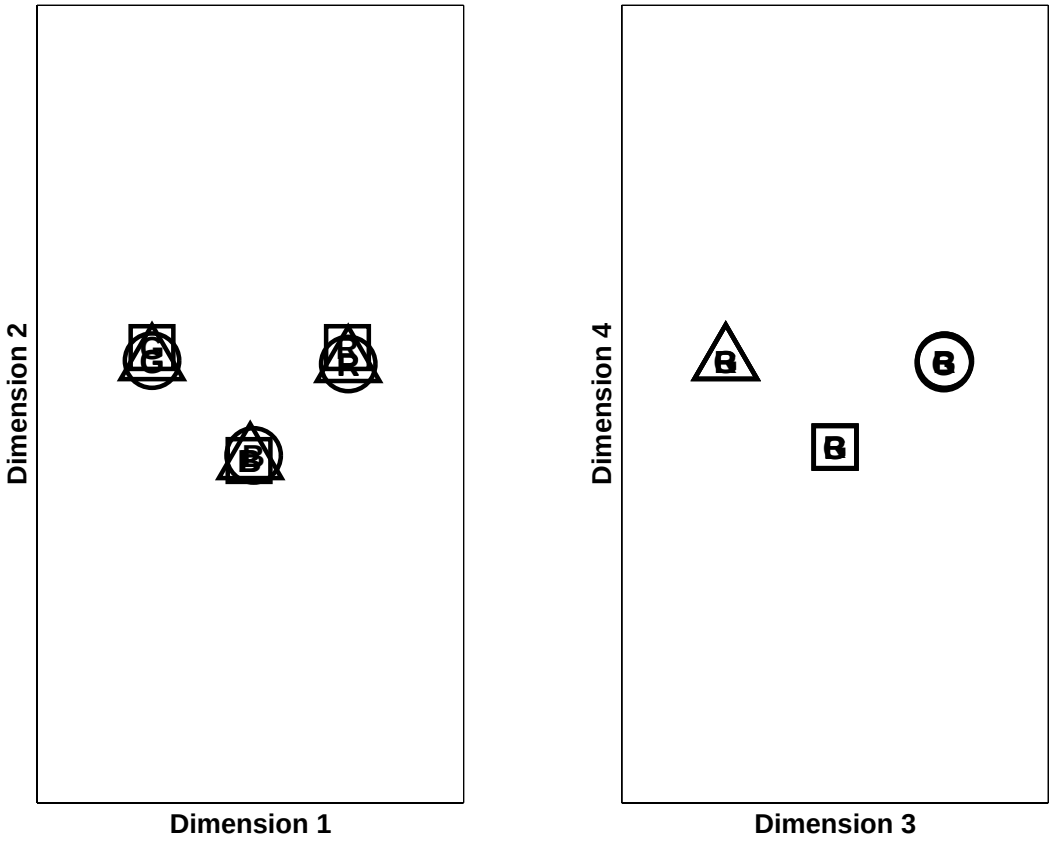


Figure 3: The four-dimensional spatial representation of the stimulus domain, shown in terms of two subspaces. The left panel plots dimension 1 and dimension 2, which capture the variation relating to color. The right panel plots dimension 3 and dimension 4, which capture the variation relating to shape.

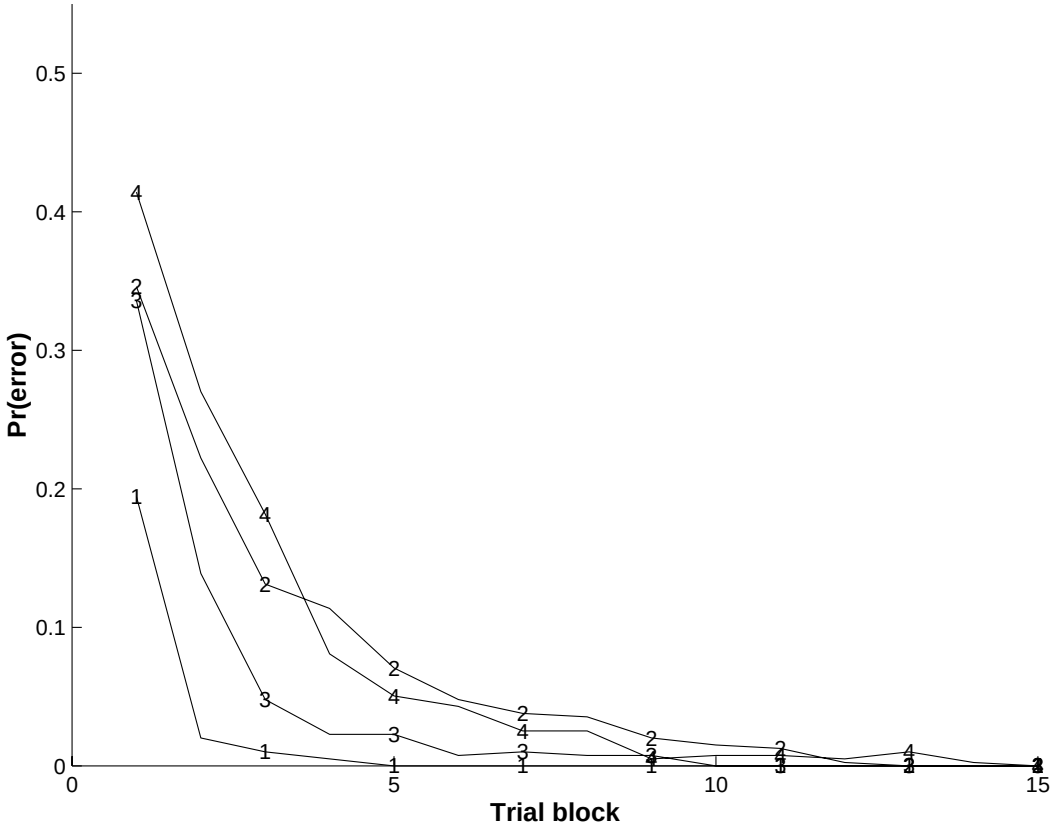


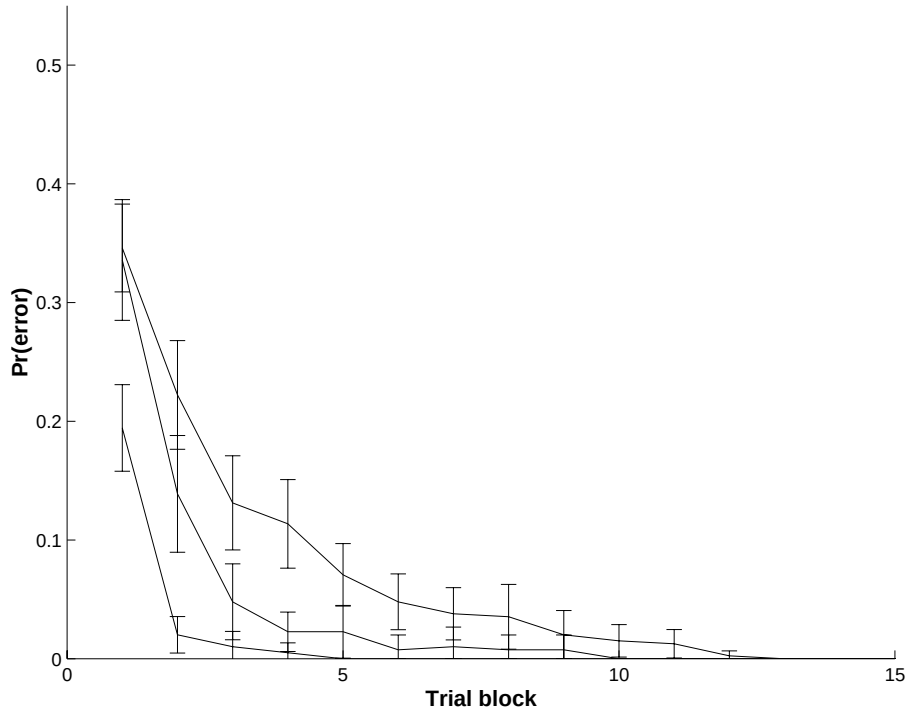
Figure 4: Averaged human performance on the four categorization tasks.

subject could be {red circle, red square, red triangle}, {red circle, blue circle, green circle}, or any of the four other possibilities, and this category could be labeled 'X' or 'Y'.

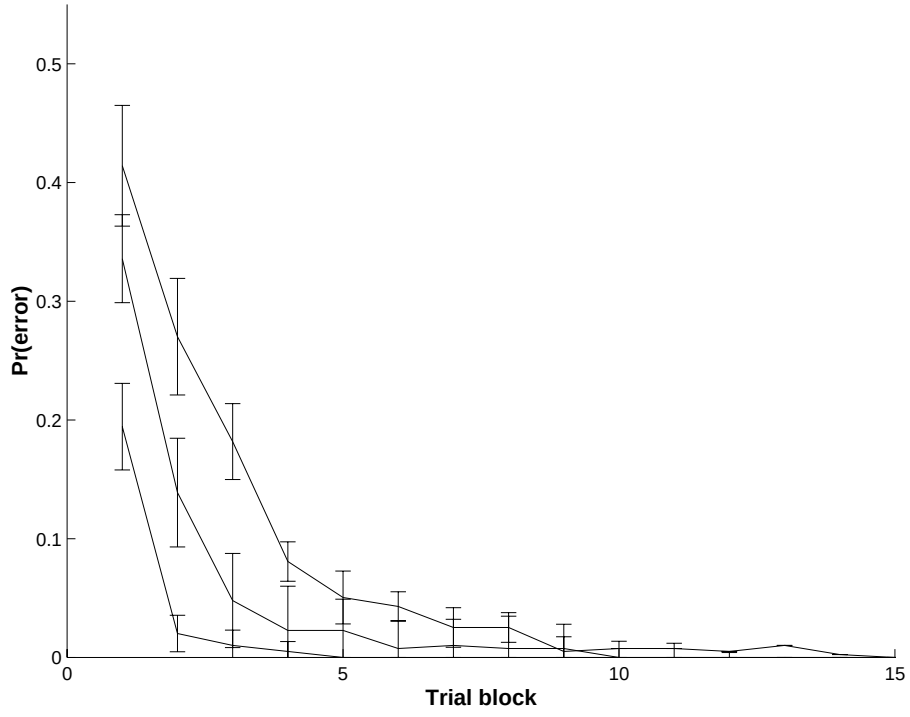
The stimuli were presented in a series of blocks, each of which involved the presentation of nine stimuli. Successive pairs of these blocks were constrained to contain exactly two presentations of each stimulus, but the ordering of their presentation within these two blocks was random. Upon presentation, subjects were required to provide a category response using the mouse within approximately 5s. Feedback was then provided for approximately 3s by showing the correct category label before the next stimulus was presented. This process continued until subjects reached a criterion of 36 consecutive correct responses, or a total of 50 presentations of each stimulus had been made. Following Nosofsky *et. al.* (1994), subjects who reached criterion were deemed to have learned the category structure, and error free performance for the remaining blocks was assumed.

Results The way in which humans learned the four category structures, summarized by averaging the error probabilities across subjects, is shown Figure 4. The averaged data suggest that Type I was learned most quickly, and with the fewest errors, Type III was the next most easily learned, and Types II and IV were the most difficult to learn.

To examine the extent to which the ordering of the averaged learning curves is supported by the underlying individual subject data, standard errors for each of the averaged error probabilities at each trial block were calculated, and used to generate 90% confidence intervals. Figure 5 shows these confidence intervals as error bars on the averaged data for the two cases of interest. In Figure 5(a) the curves for Type I, Type II and Type III are shown, and in Figure 5(b) the curves for Type I, Type III and Type IV are shown. In both cases, over the learning trials spanned by blocks 2, 3, 4 and 5, where the bulk of the learning takes place, there is strong separation between the learning curves.



(a)



(b)

Figure 5: Averaged human performance, with 90% confidence intervals, for the category structures (a) Type I (bottom), Type III (middle), and Type II (top), and (b) Type I (bottom), Type III (middle), Type (IV) top. Note that the data in these figures is the same as that displayed in Figure 4.

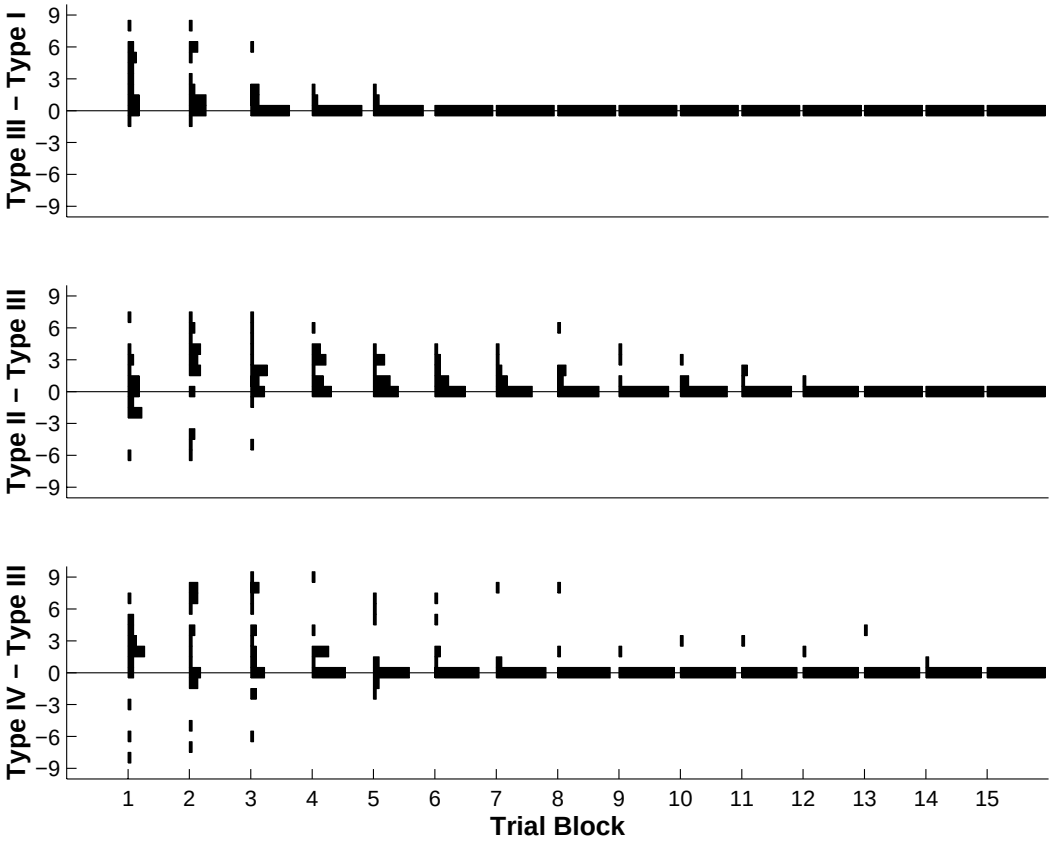


Figure 6: Frequency distribution of within-subject error differences across trial blocks, for three category type comparisons.

Since each subject learned each of the four category structures, it is also possible to conduct a within-subject analysis, comparing the differences in the number of errors each subject made at each point in the learning curve. In terms of the evident ordering in Figures 4 and 5, the important comparisons are between Type III and Type I, Type II and Type III, and Type IV and Type III. For an individual subject to display the same learning order as the aggregated data, the first category type in each of these three comparisons should involve more errors, and hence the difference should be positive. Figure 6 summarizes the difference scores calculated for these three comparisons, showing frequency histograms for the within-subject difference scores across each trial block. In each case, it can be seen that the vast majority of error differences are positive. Coupled with the analysis across-subjects averaged learning curves, this within-subjects analysis provides strong evidence for asserting that the subjects learned the Type I category structure most easily, then Type III, and then Types II and IV.

Fitting Spatial ALCOVE

To examine the ability of ALCOVE to model human category learning, we performed multivariable optimization across the four free parameters λ_w, λ_a, σ and ϕ, using the sum-squared deviation from the human block error probabilities as the objective function. The optimization approach we used combined a global grid search with local tuning based on sequential quadratic programming (see, for example, Gill, Murray, & Wright 1981), and returned parameter values of $\lambda_w = 0.21$, $\lambda_a = 0.01$, $\sigma = 14.0$ and $\phi = 2.84$, with an associated sum-squared deviation of 0.048. The learning curves produced by ALCOVE with these parameter values are shown in Figure 7. Note that the evident ordering of the learning curves is different from that shown by the human subjects in Figure 4. The final attention weights for each of the

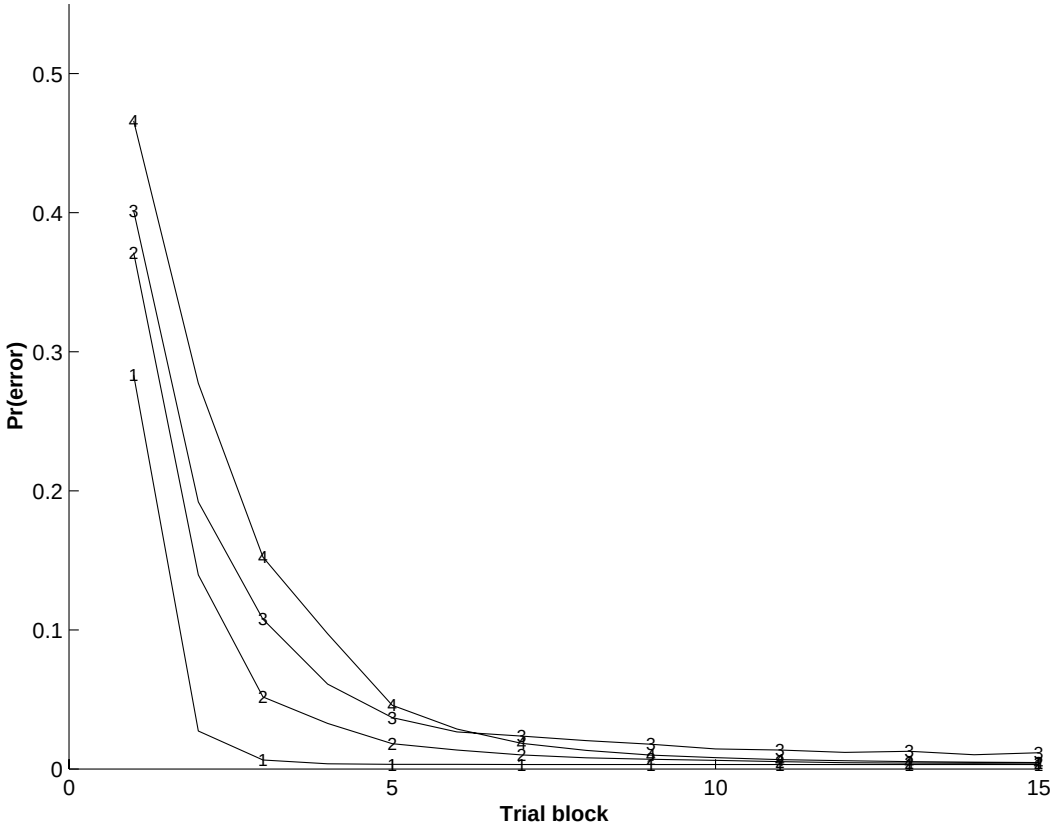


Figure 7: The performance of ALCOVE on the four categorization tasks, using the best-fitting parameter values $\lambda_w = 0.21$, $\lambda_a = 0.01$, $\sigma = 14.0$ and $\phi = 2.84$.

four category types are listed in Table 3.

This sort of analysis, which examines the degree to which ALCOVE is able to produce learning curves that are 'close' to the human curves provides one measure of its ability to capture human performance. There are a number of difficulties, however, with this approach, relating to issues of model complexity. For example, there is nothing in our optimization approach that guarantees best-fitting parameter values will not lie in unstable regions of the parameter space. That is, it is possible that small changes to the parameter values used to generate Figure 7 may result in large differences in the learning curves produced by ALCOVE. From a general model theoretic standpoint (e.g., Kass & Raftery 1995; Myung & Pitt 1997), models that require a precise tuning of parameter values to explain data are complicated models, and should be rejected in favor of simpler accounts. Indeed, many quantitative measures of model complexity, such as the Laplacian approximation (see Kass & Raftery 1995, p. 777), explicitly measure the robustness of a model's fit to the data across the region of the parameter space surrounding the best-fitting parameter values. Accordingly, one way to address the model complexity issue would be to evaluate ALCOVE against the human data using a measure that incorporated both data-fit and model complexity components, such as those described by Kass and Raftery (1995), or Myung, Balasubramanian and Pitt (2000)[1].

An alternative approach, that effectively sidesteps the detailed consideration of data-fit and complexity, is to evaluate a model in terms of its ability to capture fundamental qualitative features of the constraining data. Without wishing to make the point too strongly, there is some merit in Rutherford's

[1]It is worth noting that some of the more easily applied measures in this class, such as the Bayesian Information Criterion (Schwarz 1978), would not be suitable, since they are insensitive to the complexity effects caused by the functional form of parametric interaction (Myung & Pitt 1997).

Table 3: The final attention weights applied to the stimulus dimensions, for each of the four category structures.

Category Structure	Dimension 1 (Color)	Dimension 2 (Color)	Dimension 3 (Shape)	Dimension 4 (Shape)
Type I	0.213	0.198	0.364	0.225
Type II	0.159	0.073	0.340	0.429
Type III	0.111	0.010	0.310	0.569
Type IV	0.304	0.076	0.280	0.339

assertion that "If your experiment needs statistics, you ought to have done a better experiment". In particular, if there is a strong qualitative trend that characterizes human performance in a cognitive task, then models of that cognitive task should exhibit the same behavior. A good example of evaluating ALCOVE using this sort of approach is provided by Kruschke (1992) in relation to the Shepard *et. al.* (1961) task, where it is shown that the attention learning mechanism allows ALCOVE to capture the ordering of the learning curves for the six category types.

A similar constraint is supplied by the ordering of the human learning curves for the current task, as shown in Figure 4, and examined more closely in Figures 5 and 6. For ALCOVE to capture human performance, it must be able to display the ordering Type I, then Type III, and then Types II and IV. As shown in Figure 7, ALCOVE does not do this when using the best-fitting parameter values. In fact, as part of a more general survey of the parameter space, we were unable to find any combination of parameter values that allowed ALCOVE to learn Type III more easily than Types II and IV.

Discussion

An examination of the concrete examples of the four category structures shown in Figure 1 suggests that this deficiency may be caused by the spatial representation it uses. The Type I category structure is easily learned because it allows the 9 stimuli to be 'collapsed' into three groups of three, collecting together the circles, squares and triangles. The Type III category structure would benefit from a more complicated form of representational collapse, which brought together the features red and blue, effectively reducing the problem to six nodes, rather than nine. Meanwhile, neither the Type II nor the Type IV encourage any form of representational collapse.

For the Type III category structure, however, there is no way for ALCOVE to manipulate the spatial representation to bring together the features red and blue. As is clear from Figure 3, the only way to align the red and blue stimuli is to reduce the attention weights for dimension 1 to zero, but this manipulation has the unwanted side-effect of aligning the green stimuli, and makes it impossible to learn the category structure. The way in which ALCOVE attempts to overcome this fundamental difficulty is made clear by the best-fitting specificity parameter value. The value $\sigma = 14.0$ corresponds to an extremely sharp generalization gradient, meaning that ALCOVE is effectively using local, rather than distributed, stimulus representation. Intuitively, this means that each of the different category types is being learned by establishing appropriate associative weights to apply to local regions of the spatial representation. The small best-fitting attention learning rate of $\lambda_a = 0.01$ shows that ALCOVE does not use selective attention to provide more significant levels of generalization. In other words, because the dimensional structure of the spatial representation is not well suited to learning the category structures through processes of selective attention and generalization, the best-fitting parameter values indicate ALCOVE uses a less compelling learning strategy based on establishing associative weights.

For this reason, the final attention weights shown in Table 3 are not very informative. In particular, they do not reflect the outcome of an attention-based learning strategy. For each category type, the extent to which the final attention weights differ from the starting point of equality tends to reflect the number of learning trials involved. Since ALCOVE only modifies its attention weights when it makes an

incorrect categorisation, greater change is evident for the more difficult category types.

More importantly, the ordering of the learning curves shown in Figure 7 are readily explained in terms of the 'local learning' process. For Type I, those stimuli that belong to the smaller category are similar stimuli within the original spatial representation. In other words, the appropriate category structure is largely already captured by the stimulus representation, meaning that the categorization task is reasonable easy to accomplish even without attentional learning. There is less consistency, however, between the spatial representation with the Type II category structure, because the stimuli in the smaller category are less similar to each other. This similarity decreases further for Type III, because there is one less feature in common across the stimuli in the smaller category. Finally, the Type IV category structure is least well captured by the spatial representation. This pattern of correspondences, under the learning approach used by ALCOVE with the best-fitting parameter values, leads to the ordering Type I , then Type II, then Type III, and finally Type IV, as evident in Figure 7.

A complication for this analysis is that the featural permutation of Type III that required 'red' and 'green' to be collapsed could be accomplished using the spatial representation, while the 'green' and 'blue' permutation suffers the same difficulty as 'red' and 'blue'. In this sense, the model fitting results presented in Figure 7 may be justified as representing the dominant behavior of the model. More fundamentally, it seems theoretically implausible that different featural permutations lead to different category learning performance, and there is no evidence in the collected data to support such an assertion. In particular, as Figure 5 shows, the human learning of Type III is not significantly more variable than Types II or IV, despite the fact that all category structures were tested across all of the permissible perceptual display permutations.

The general conclusion, therefore, is that it seems the inability of ALCOVE to learn the four category structures in the same order as humans may arise not because of a fault of ALCOVE *per se*, but through its reliance on spatial representation. The fact that the spatial representation is an accurate and intuitively reasonable description of the stimulus domain, explaining 98.8% of the variance in the data using an interpretable structure, gives some suggestion that the difficulty lies in a fundamental incompatibility between the representational assumptions embodied by the spatial approach, and those used by humans. For this reason, it is worth examining the ability of an ALCOVE-like model, using stimulus representations generated according to the alternative featural approach, to model the category learning data. While it remains entirely plausible that modifications to the processed used by ALCOVE might be able to account for the learning order (e.g., Erickson & Kruschke 1998; Kruschke & Blair in press; Kruschke & Johansen 1999), there is a sense in which a simple representational change would constitute a more direct and elegant solution.

Featural Stimulus Representations

The distinction between spatial and featural approaches to mental representational modeling has been a classic one in cognitive psychology. The spatial approach adopted by ALCOVE represents stimuli as points in a multidimensional space, while the featural approach represents stimuli in terms of the presence or absence of a number of discrete (often binary) features. It has frequently been observed (e.g., Carroll 1976, p. 440, Tenenbaum 1996, p. 3, Tversky 1977, p. 328) that the nature of spatial representation means that it is better suited to domains where stimuli vary continuously along a relatively small number of dimensions, while the discrete nature of the featural approach makes it more appropriate for modeling domains where stimuli are defined in terms of a set of properties or features.

The Contrast Model

For stimulus domains where discrete featural representations are deemed to be appropriate, it is necessary to develop an analogue of the distance-based approach to measuring stimulus similarity used with spatial representations. This analogue is provided by Tversky's (1977) Contrast Model, which assumes that the similarity between two stimuli is a function of their common and distinctive features. Formally, the similarity takes the form:

$$s_{ij} = \theta F\left(\mathbf{f}_i \cap \mathbf{f}_j\right) - \alpha F\left(\mathbf{f}_i - \mathbf{f}_j\right) - \beta F\left(\mathbf{f}_j - \mathbf{f}_i\right), \tag{9}$$

where $\mathbf{f}_i \cap \mathbf{f}_j$ denotes the features common to the ith and jth stimuli, $\mathbf{f}_i - \mathbf{f}_j$ denotes the features present in the ith, but not the jth, stimulus, and $F(\cdot)$ is some monotonically increasing function. By manipulating the positive weighting parameters θ, α and β, different degrees of importance may given to the common and distinctive components in assessing stimulus similarity. In particular, Tversky (1977) and others (e.g., Carroll & Corter 1995; Gati & Tversky 1984; Restle 1961; Sattath & Tversky 1987) have placed some emphasis on the two extreme alternatives of the contrast model obtained by setting $\theta = 1, \alpha = \beta = 0$, which results in a purely common features model of similarity, or setting $\theta = 0, \alpha = \beta = 1$, which results in a purely distinctive features model.

In terms of developing a featural extension to ALCOVE, it is natural to ask whether a common or distinctive features approach to similarity (or some balance between the two) should be used in place of the distance measures used for spatial representations. In answering this question, it is important to distinguish between the two different roles distance measures play in measuring similarity in ALCOVE. One role is to underpin the generation of stimulus representations, since the primary aim of techniques such as multidimensional scaling is to model the distance relations specified by similarity data. The second role is to serve in the generation of stimulus similarities during the categorization of a presented stimulus. Within the spatial representational approach of ALCOVE, the same distance metric is used for both types of similarity.

It is, however, widely recognized (e.g., Goodman 1972; Nosofsky 1986; Rips 1989; see Goldstone 1994 for an overview) that similarity is not a unitary phenomenon, and the way in which it is measured may change according to different cognitive demands. As Goldstone, Medin and Halberstadt (1997) argue: "the aggregate of evidence suggests that similarity is not just simply a relation between two objects; rather, it is a relation between two objects and a context" (p. 238). In particular, there is considerable empirical evidence for context dependency when featural similarities are generated according to the Contrast Model (e.g., Gati & Tversky 1984; Ritov, Gati, & Tversky 1990; Sattath & Tversky 1987), with the general conclusion being that "the weighting of common and distinctive features is context dependent, but these variations are systematic rather than random" (Ritov *et. al.* 1990, p. 40).

Of specific concern here is the suggestion that the two contexts involved in ALCOVE—the generation of similarity judgments, and the generation of category responses—involve different processes when dealing with featural stimulus representations. Gati and Tversky (1984) argue that different task demands can induce significant changes on the relative weighting of common and distinctive features. In particular, they propose that "judgments of similarity focus on common features whereas judgments of dissimilarity focus on distinctive features" (Gati & Tversky 1984, p. 367; see also Markman 1996). On this basis, it would seem likely that a common features approach to similarity should be used to extract a domain representation from similarity data, while a distinctive features approach should be used when categorizing a presented stimulus. It is worth examining each of these claims in more detail.

In terms of feature extraction from similarity data, it is known that the distinctive features approach is formally equivalent to the common features approach when complementary features are present (Sattath & Tversky 1987). This means that, when a feature belonging to a subset of stimuli is identified, another feature belonging to all of the other stimuli is implied, and all of the stimuli that do not have the feature are consequently made relatively more similar. As previously argued by Lee (1998), this is sensible in the (relatively rare) case of 'global' domain features, but prevents the extraction of 'local' domain features. For example, consider the featural modeling of the abstract conceptual properties of the numbers $0, 1, \ldots, 9$ (see Shepard, Kilpatrick, & Cunningham 1975; Tenenbaum 1996). It would be possible, under the distinctive approach, to find features corresponding to 'even numbers' and 'odd numbers', because they are complementary. The feature corresponding to 'multiples of three', however, is unlikely to be found, since its complement (the numbers 0, 1, 2, 4, 5, 7 and 8) does not correspond to any feature. As Lee (1998) goes on to argue, a common features model of stimulus similarity is needed to extract these sorts of features from similarity data.

In terms of the categorization process requiring the distinctive features approach, some considerable insight is provided by considering the category learning task studied by Shepard *et. al.* (1961). As was noted earlier, the key observation is that this stimulus domain is equally well represented using both the spatial and featural approaches. A small, black square, for example, is just as well conceived as a stimulus with the features 'small', 'black' and 'square', as it is a point in a three dimensional space (the vertex of a cube) that corresponds to the extremes values of 'small', 'black' and 'square' along stimulus

dimensions of 'size', 'color' and 'shape'. Since ALCOVE is able to capture the learning differences between the six category structures found empirically, the implication is that a featural extension of ALCOVE should reduce to the standard spatial version for this stimulus domain. In looking to achieve this equivalence, an examination of the learning rule for the attention weights (Eq. 8) shows that their attention weight learning is entirely driven by those stimuli that are different from the presented stimulus on each dimension, which provides strong evidence in favor of using the distinctive feature model of stimulus similarity.

Taken together, these arguments suggest that the requirements of extracting features from similarity data, and adapting attention weights during category learning, are fundamentally different. By treating the common and distinctive feature measures of stimulus similarity as specializations of the overarching Contrast Model, it is possible to satisfy these different demands. Under the established framework provided by the Contrast Model, it is natural to use the common features measure when it is needed for generating stimulus representations, and a distinctive features measure when it is needed for category learning.

Additive Clustering

The obvious means of extracting featural representations from similarity data, using a common features approach, is by applying additive clustering techniques (Shepard & Arabie 1979). These techniques find a set of domain features, and assign a saliency weight to each, so that the observed similarity between a pair of stimuli is approximated by the sum of the weights of the clusters common to both stimuli. Formally, if the presence or absence of the kth feature in relation to the ith stimulus is defined as:

$$f_{ik} = \begin{cases} 1 & \text{if stimulus } i \text{ has feature } k \\ 0 & \text{otherwise,} \end{cases} \tag{10}$$

and the kth feature is assigned a saliency weight a_k, then the similarity between the ith and jth stimuli is given by:

$$s_{ij} = \sum_k a_k f_{ik} f_{jk} + c \tag{11}$$

where c is an additive constant, corresponding to a 'universal' feature that is shared by every stimulus.

As a concrete example of an additive clustering representation, Table 4 presents the results of analyzing Rosenberg and Kim's (1975) similarity data for kinship terms. This representation was generated using a modified version of the algorithm described by Lee (in press c), based on a stochastic hill-climbing approach to combinational optimization. As with the multidimensional scaling algorithm described earlier, the particular strength of this algorithm is that it uses an additive clustering version of the Bayesian Information Criterion (Lee in press b) to balance the competing demands of maximizing data-fit while minimizing model complexity.

The representation itself explains 92.8% of the variance in the data using ten features, and the universal feature containing all stimuli. There are features relating to which generation each kinship term belongs, whether or not that they are once-removed, and their gender. The important point is that each of these three perspectives 'cuts across' the other two, and demands a clustering model that allows arbitrary patterns of overlap between clusters. Only with overlapping clusters, for example, can the kinship term 'brother' belong to the clusters that correspond to the features 'sibling', 'nuclear family' and 'male'. Because it allows the necessary flexibility, additive clustering is able to generate an accurate representation of the kinship stimulus domain using a relatively small number of features.

This representation of the kinship stimulus domain, together with a wide range of others generated by additive clustering (e.g., Shepard & Arabie 1979; Lee 1999), would appear to be suitable featural counterparts to the multidimensionally scaled spatial representations used by ALCOVE. Given that the ALCOVE model was developed specifically for use with spatial representations, however, it is necessary to make some modifications before it is able to accommodate featural representation. In the next section, we develop a featural version of ALCOVE, identifying the changes that need to be made within the framework we used to describe the original ALCOVE.

Table 4: 10-cluster representation of kinship data.

Stimuli In Cluster	Weight
brother sister	0.391
father mother	0.372
daughter son	0.370
granddaughter grandfather grandmother grandson	0.366
aunt uncle	0.330
nephew niece	0.326
aunt cousin nephew niece uncle	0.277
aunt daughter granddaughter grandmother mother niece sister	0.269
brother father grandfather grandson nephew son uncle	0.268
brother daughter father mother sister son	0.208
additive constant	0.062
Variance Explained	92.8%

A Featural Version Of ALCOVE

Stimulus Representation

The featural representations, as generated by additive clustering, take the form of binary membership variables f_{ik}, denoting whether or not the ith stimulus has the kth feature (refer Eq. 10), and a set of saliency weights $(w_1, \ldots, w_K)$ for the K features.

Stimulus Comparison The original ALCOVE model calculates the distance between each stimulus and the presented stimuli, using the known locations of the representative points, and the metric structure of the space. While featural representations have neither spatial locations nor metric structure, generalizing the notion of distance from spatial to featural representation is relatively straightforward. All that is required is the selection of an appropriate functional form, $F(\cdot)$, in the Contrast Model (Eq. 9) under a distinctive features parameterization $\theta = 0, \alpha = \beta = 1$. Following the lead taken by additive clustering under the common features approach, a simple additive functional form seems reasonable.

In this way, featural distance may be defined as the sum of the weights of the features that differ between two stimuli, as follows:

$$d_{ij} = \sum_k a_k f_{ik} (1 - f_{jk}) + \sum_k a_k (1 - f_{ik}) f_{jk} \tag{12}$$

where a_k now denotes the saliency of the kth feature.

The notion of saliency for featural representation corresponds to the notion of dimensional attention for spatial representation. Accordingly, it is appropriate for each feature initially to have the attention weight prescribed by the additive clustering solution, rather than simply assuming all featural saliencies to be equivalent at the beginning of category learning[2]. During the course of category learning, these attention weights are then modified according to the category structure being presented, with features

[2]The use of equal initial attention weights in the spatial ALCOVE model is justified, however, since the representations generated by multidimensional scaling implicitly encode dimensional saliencies by using different degrees of extension along the various spatial dimensions.

that distinguish between categories becoming highly weighted, and irrelevant features receiving little or no attention. The initial attention weightings, therefore, reflect only the *a priori* expectation regarding the salience of each feature, based on the evidence provided by the similarity data.

Generalization Gradient The original ALCOVE model converted stimulus distances into stimulus similarities, using an exponential decay function. As presented in Shepard (1987), however, the theoretical basis for this relationship relies on probabilistic geometry, and is inherently spatial. This means that the use of the exponential decay function for featural representations cannot be based on Shepard's (1987) results.

Fortunately, however, Russell (1986, see also Gluck 1991) provides theoretical analysis of generalization gradients across featural representations, which uses the same approach as Shepard (1987), and finds that stimulus similarity still decays exponentially with respect to featural distance. As Shepard (1994) summarizes, the change to featural representations "still yields an exponential type of falloff of generalization with distance, where distance is now defined in terms of the sum of the weights of the features that differ between the two objects" (p. 25). This means that stimulus similarity may be calculated as:

$$s_{ij} = \exp\left(-\sigma\left[\sum_k a_k f_{ik}\left(1-f_{jk}\right) + \sum_k a_k\left(1-f_{ik}\right) f_{jk}\right]\right) \tag{13}$$

Response Probabilities Once these similarities have been found, the use of featural representation does not require any change to the way ALCOVE generates response strengths:

$$r_x = \sum_j w_{xj} s_{ij}, \tag{14}$$

or response probabilities:

$$\Pr\left(X \mid i\right) = \frac{\exp\left(\phi r_X\right)}{\sum_x \exp\left(\phi r_x\right)}. \tag{15}$$

Learning

Once the featural version of ALCOVE has generated category response probabilities for a presented stimulus, the same 'humble teacher' values are used:

$$t_x = \begin{cases} \max\left(+1, r_x\right) & \text{if stimulus } i \text{ is in category } x \\ \min\left(-1, r_x\right) & \text{otherwise,} \end{cases}, \tag{16}$$

and the same error measure is defined:

$$E = \frac{1}{2}\sum_x \left(t_x - r_x\right)^2. \tag{17}$$

Associative Learning Because the method of response generation was not affected by the use of featural representations, there is no need to alter the associative learning rule:

$$w_{xj}^{new} = w_{xj}^{old} + \lambda_w\left(t_x - r_x\right) s_{ij}. \tag{18}$$

Attentional Learning The change to the way stimulus similarity is expressed for featural stimuli (Eq. 13) does, however, warrant a change to the attention learning rule. It now becomes:

$$a_k^{new} = a_k^{old} - \lambda_a \sum_x \left(t_x - r_x\right) \sum_j w_{xj} s_{ij}\left(f_{ik} + f_{jk} - 2 f_{ik} f_{jk}\right). \tag{19}$$

It is important to understand that, in purely computational terms, this learning rule does not differ from the spatial version (Eq. 8). This is a consequence of the fact that, as noted by Nosofsky (1991, pp. 103–105) the feautral distance measure (Eq. 12) is identical to the spatial distance measure (Eq. 1) for binary variables, and hence the featural similarity measure (Eq. 13) reduces to the spatial similarity measure (Eq. 2). Conceptually, however, it is often useful to distinguish the representational interpretations demanded by the spatial and featural approaches. For example, conceiving of featural representations as the vertices of a hypercube can be counter-productive, since the intuitive notion of spatial distance does not correspond to stimulus dissimilarity under any model of stimulus similarity that uses common features.

Comparing Spatial And Featural ALCOVE

The most striking property of the featural ALCOVE model is how little it differs from the established spatial ALCOVE model. Stimulus similarities are generated across featural representations in a way that is conceptually different, but computationally equivalent, to the spatial approach. The same applies to the learning rules, which can be thought to have differences in form, but not in substance. Indeed, the only real difference, in terms of the way ALCOVE learns to categorize stimuli, is that attention weights are maintained for each stimulus feature, rather than each stimulus dimension.

The fundamental difference between the two models, however, is the representational difference. Using additive clustering to generate a featural representation of the stimulus domain, rather than multidimensional scaling to generate a spatial representation, leads ALCOVE to understand the structure of the stimulus domain in an entirely new way. Given the plausible argument that ALCOVE's failure to capture human performance on the categorization task presented earlier may have been due to limitations in the spatial representation of the domain, it is clearly worth examining the capability of the featural version.

The Experiment Revisited

Featural Stimulus Representation

To generate a featural representation of the color and shape domain the additive clustering algorithm previously used for the kinship domain was applied to the averaged similarity data given in Table 1. Figure 8 shows the pattern of change of data-fit and the Bayesian Information Criterion as extra clusters are added to the featural representation. A clear minimum in the Bayesian Information Criterion is evident at the point where 6 clusters are used, indicating that this representation constitutes the appropriate balance between accuracy and simplicity.

The structure of this representation, which explains 99.3% of the variance in the data, is given in Table 5. Each of the clusters is readily interpreted in terms of its defining feature, and these are the specific colors and shape from which the stimuli were constructed. Interestingly, the saliency weights of the features suggest that subjects assigned relatively greater emphasis to common color, as compared to common shape, when judging the similarity between stimuli.

Fitting Featural ALCOVE

We fit the featural ALCOVE model using the same multivariable optimization approach previously applied to the spatial version. The parameter values returned weref $\lambda_w = 0.12$, $\lambda_a = 0.09$, $\sigma = 6.90$ and $\phi = 2.85$, and had an associated sum-squared deviation of 0.022. The learning curves produces by the featural version of ALCOVE, using the best-fitting parameter values, are shown in Figure 9, and the final attention weights for the features in each of the category structures are given in Table 6. The important aspect of the learning curves is, of course, that they exhibit the same ordering as the human data. In particular, The Type III category structure is learned more quickly than Types II and IV.

Once again, there is no guarantee that the parameter values found to generate Figure 9 lie in a stable region of the parameter space. As with the spatial ALCOVE model, however, extensive simulation showed that the learning orders are largely insensitive to parametric variation. Across a broad range of parameter values, the featural ALCOVE model learned Type III more easily than Types II or IV, which were approximately equally difficult.

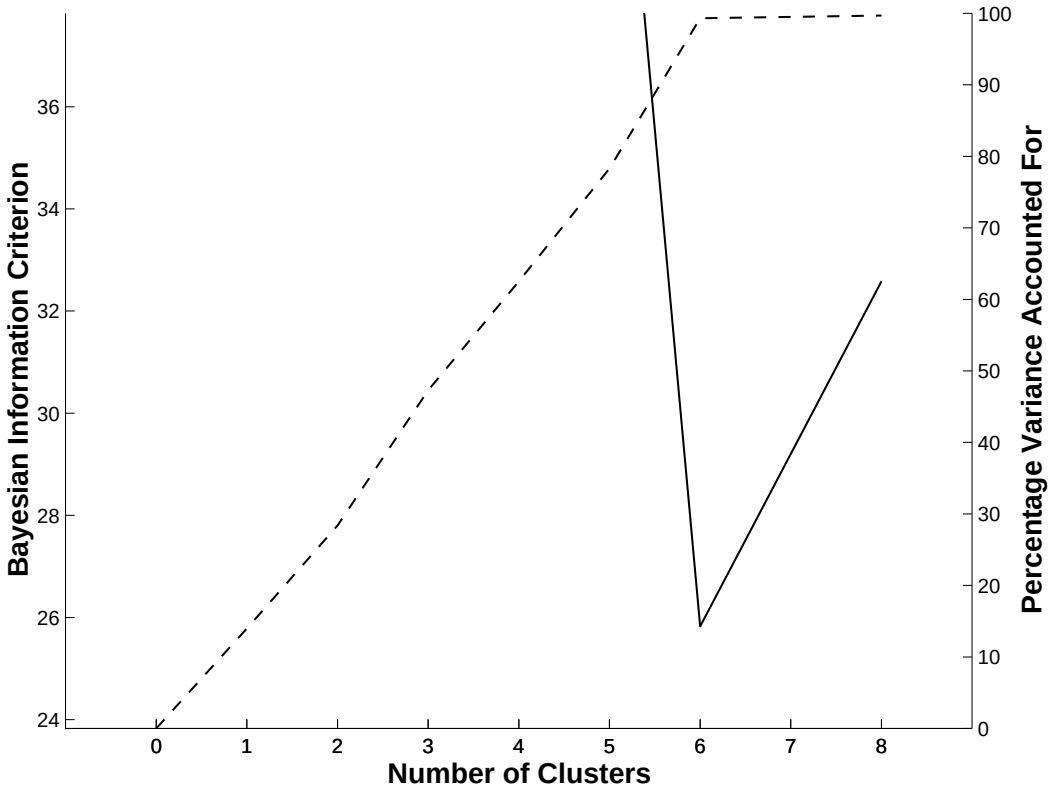


Figure 8: The results of applying the additive clustering algorithm to the similarity data, showing the pattern of change of the Bayesian Information Criterion (left hand scale, solid line), and Percentage Variance Explained (right hand scale, broken line) measures for featural representations with different numbers of clusters.

Table 5: The additive clustering representation of the color and shape stimulus domain.

Stimuli In Cluster	Interpretation	Weight
`green-circle green-square green-triangle`	green	0.602
`red-circle red-square red-triangle`	red	0.590
`blue-circle blue-square blue-triangle`	blue	0.577
`red-square green-square blue-square`	square	0.510
`red-triangle green-triangle blue-triangle`	triangle	0.473
`red-circle green-circle blue-circle`	circle	0.473
additive constant		0.073
Variance Explained		99.3%

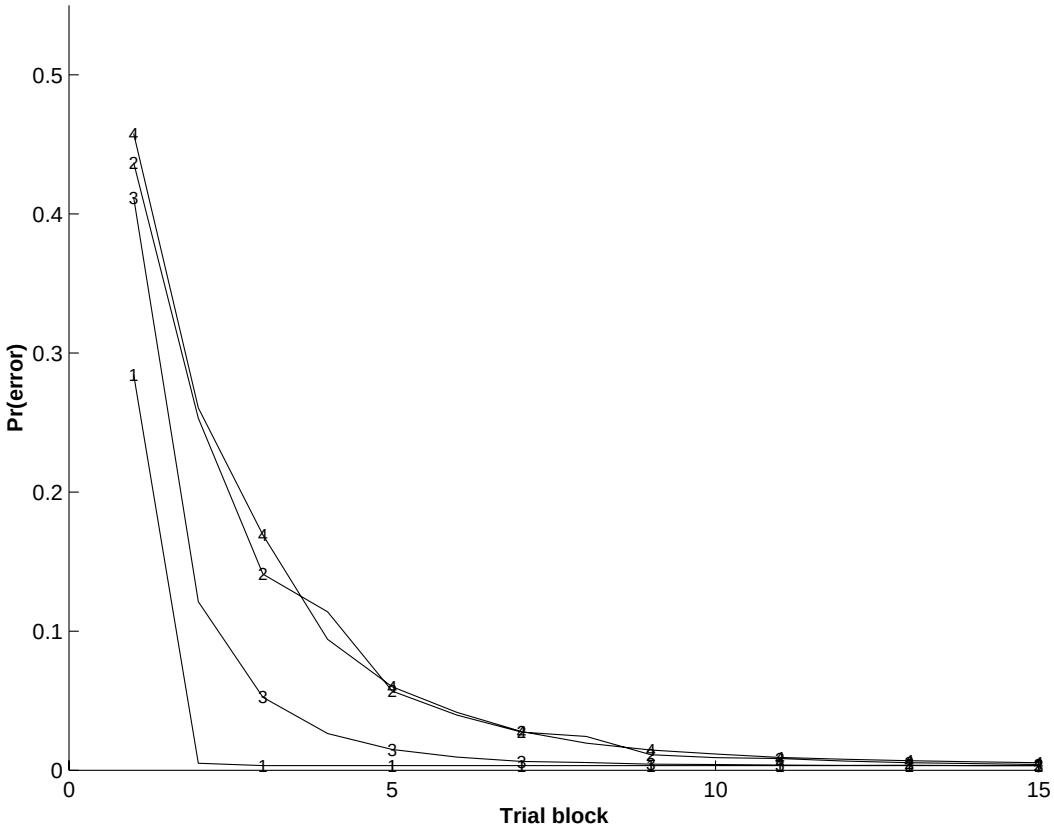


Figure 9: The performance of the featural ALCOVE model on the four categorization tasks, using the best-fitting parameter values $\lambda_w = 0.12$, $\lambda_a = 0.09$, $\sigma = 6.90$ and $\phi = 2.85$.

Table 6: The final attention weights applied to the stimulus features, for each of the four category structures.

Category Structure	Feature 1 (Green)	Feature 2 (Red)	Feature 3 (Blue)	Feature 4 (Square)	Feature 5 (Circle)	Feature 6 (Triangle)
Type I	0.000	0.000	0.000	0.225	0.552	0.224
Type II	0.177	0.146	0.179	0.145	0.175	0.178
Type III	0.401	0.000	0.000	0.353	0.148	0.098
Type IV	0.169	0.167	0.164	0.169	0.167	0.169

Discussion

The best-fitting specificity parameter value of $\sigma = 6.90$ for the featural ALCOVE model is less than half of the corresponding value for the spatial ALCOVE model, and the attention learning rate of $\lambda_a = 0.09$ is much higher. This indicates that the use of featural representations allowed the featural ALCOVE model to use selective attention and generalization processes to learn the category structures. In particular, the earlier analysis of the poor performance of the spatial ALCOVE model presented earlier suggested that, to learn Type III more easily than Types II and IV, the ability to group the features 'red' and 'blue' may play and important role. The final attention weights shown in Table 6 demonstrate how featural ALCOVE achieves this representational manipulation. Of the color features, only 'green' maintains a non-zero attention weight, meaning that the color of each stimulus is effectively reduced to the distinction 'green' or 'not green'. In this way, the colors red and blue are treated as one, and the Type III category structure is able to be learned more easily than Types II and IV, both of which require attention to all six features.

Meanwhile, featural ALCOVE achieves the representational collapse required to learn the Type I category structure by reducing attention weights for the color features to zero, effectively attending only to stimulus shape. It is interesting to note that the 'circle' feature that defines the smaller category (refer Figure 1) is given relatively greater attention that the other two shape features.

General Discussion

Human performance on the color and shape task, as characterized by the order in which the different category structures are learned, provides a strong constraint for any model of category learning. The ALCOVE model, when relying on a spatial stimulus representation, is unable to produce the same learning order. A slightly modified version of ALCOVE, however, designed to accommodate featural representations, reliably produces the correct ordering. This finding provides strong evidence in favor of the need to represent the color and shape domain using discrete features, rather than continuous dimensions, and demonstrates the utility of generalizing ALCOVE to consider both types of stimulus representation.

Perhaps the fact that change to featural stimulus representation demanded few changes to the category learning processes used by ALCOVE should not be surprising. ALCOVE evolved from successful models of category learning (Medin & Schaffer 1978; Nosofsky 1984), and has been demonstrated to be empirically successful in its own right. In addition, the processes used by ALCOVE are both simple and directly interpretable in terms of basic principles of category learning (Kruschke 1993). For this reason, one might expect that ALCOVE, with minor modifications, would be capable of dealing with any reasonable form of stimulus representation. The fact that ALCOVE was originally cast in spatial terms need not imply that it is better suited to spatial or featural representations.

An interesting future application of the featural ALCOVE model involves transfer effects, particularly those involving positive transfer to novel values along a previously relevant dimension. These effects would seem to require a featural representation that captured the 'higher-order' relationships between features, recognizing, for example, that 'red' and 'blue' are both colors, but that 'square' is not. The representational freedom afforded by additive clustering models makes it well suited to generating these sort of hierarchical feature structures, while still maintaining the possibility of overlapping features. There is even the possibility of using an extended form of additive clustering that is built on the full contrast model of similarity, rather than just relying on its common features special case. Whether the featural ALCOVE model developed here displays the appropriate transfer effects using more sophisticated featural representations is a worthwhile topic for future investigation.

Thinking in a similar vein, we suspect ALCOVE could be modified to deal with richer stimulus representations than are allowed by either the spatial or featural approaches. These two representational formalisms can be viewed as being at the extremes of a representational continuum, and many stimulus domains would probably benefit from a representation that combined aspects of both. As Carroll (1976) argues: "Since what is going on inside the head is likely to be complex, and is equally likely to have both discrete and continuous aspects, I believe the models we pursue must also be complex, and have both discrete and continuous components" (p.462). There does not seem to be any barrier preventing a

modified ALCOVE model from using stimulus representation structured in terms of this hybrid spatial-featural approach. Indeed, we would suggest that the 'distance' between stimuli represented as points in a multidimensional space, and having a number of saliency-weighted features, is simply the sum of the metric spatial distance between them and the weights of their distinctive features. Using this measure, stimulus similarities could be calculated, and appropriate learning rules derived for a very general model of category learning. The main difficulty would seem to be the development of a technique to fulfill the role of multidimensional scaling and additive clustering by generating these hybrid representations. While there are techniques for combining spatial representation with the partitioning clusterings they reveal (e.g., DeSarbo, Howard, & Jedidi 1991), we know of no general hybrid technique that affords the full flexibility of both multidimensional scaling and additive clustering.

In the meantime, however, allowing the ALCOVE model of category learning to use featural representations significantly extends the type of stimulus domain to which it can be applied. While many stimuli are appropriately represented in continuous coordinate spaces, many others are better described in terms of the presence or absence of discrete features. Being able to apply the ALCOVE model to both types of representations enhances its generality, and may offer fresh insights into the fundamental cognitive process of categorization.

References

Carroll, J. D. (1976). Spatial, non-spatial and hybrid models for scaling. *Psychometrika 41*, 439–463.

Carroll, J. D. & Corter, J. E. (1995). A graph-theoretic method for organizing overlapping clusters into trees, multiple trees, or extended trees. *Journal of Classification 12*, 283–313.

Choi, S., McDaniel, M. A. & Busemeyer, J. R. (1993). Incorporating prior biases in network models of conceptual rule learning. *Memory & Cognition 21*(4), 413–423.

DeSarbo, W. S., Howard, D. J. & Jedidi, K. (1991). MULTICLUS: A new method for simultaneously performing multidimensional scaling and cluster analysis. *Psychometrika 56*(1), 121–136.

Erickson, M. A. & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General 127*(2), 107–140.

Garner, W. R. (1974). *The processing of information and structure*. Potomac, MD: Erlbaum.

Gati, I. & Tversky, A. (1984). Weighting common and distinctive features in perceptual and conceptual judgments. *Cognitive Psychology 16*, 341–370.

Gill, P. E., Murray, W. & Wright, M. H. (1981). *Practical Optimization*. London, UK: Academic Press.

Gluck, M. A. (1991). Stimulus generalization and representation in adaptive network models of category learning. *Psychological Science 2*, 50–55.

Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition 52*, 125–157.

Goldstone, R. L., Medin, D. L. & Halberstadt, J. (1997). Similarity in context. *Memory & Cognition 25*(2), 237–255.

Goodman, N. (1972). Seven strictures on similarity. In N. Goodman (Ed.), *Problems and Projects*, pp. 437–446. New York, NY: Bobbs-Merrill.

Kass, R. E. & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association 90*(430), 773–795.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review 99*(1), 22–44.

Kruschke, J. K. (1993). Three principles for models of category learning. *The Psychology of Learning and Motivation 29*, 57–90.

Kruschke, J. K. & Blair, N. J. (in press). Blocking and backward blocking involve learned inattention. *Psychonomic Bulletin & Review*.

Kruschke, J. K. & Erikson, M. A. (1995). Six principles for models of category learning. Talk presented at the 36th Annual Meeting of the Psychonomic Society, 10 November 1995, Los Angeles, CA.

Kruschke, J. K. & Johansen, M. K. (1999). A model of probabilistic category learning. *Journal of Experimental Psychology: Learning, Memory and Cognition 25*(5), 1083–1119.

Lee, M. D. (1998). Neural feature abstraction from judgments of similarity. *Neural Computation 10*(7), 1815–1830.

Lee, M. D. (1999). An extraction and regularization approach to additive clustering. *Journal of Classification 16*(2), 255–281.

Lee, M. D. (in press a). Determining the dimensionality of multidimensional scaling representations for cognitive modeling. *Journal of Mathematical Psychology*.

Lee, M. D. (in press b). On the complexity of additive clustering models. *Journal of Mathematical Psychology*.

Lee, M. D. (in press c). A simple method for generating additive clustering models with limited complexity. *Machine Learning*.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush & E. Galanter (Eds.), *Handbook of Mathematical Psychology*, pp. 103–189. New York, NY: Wiley.

Markman, A. B. (1996). Structural alignment in similarity and difference judgments. *Psychonomic Bulletin & Review 3*(2), 227–230.

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

More, J. J. (1977). The Levenberg-Marquardt algorithm: Implementation and theory. In G. A. Watson (Ed.), *Lecture Notes in Mathematics, 630*, pp. 105–116. Springer-Verlag.

Myung, I. J., Balasubramanian, V. & Pitt, M. A. (2000). Counting probability distributions: Differential geometry and model selection. *Proceedings of the National Academy of Sciences 97*, 11170–11175.

Myung, I. J. & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review 4*(1), 79–95.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition 10*(1), 104–114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General 115*(1), 39–57.

Nosofsky, R. M. (1991). Stimulus bias, asymmetric similarity, and classification. *Cognitive Psychology 23*, 94–140.

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

Restle, F. (1961). *Psychology of Judgment and Choice*. New York, NY: Wiley.

Rips, L. J. (1989). Similarity, typicality, and categorization. In S. Vosniadou & A. Ortony (Eds.), *Similarity and Analogical Reasoning*, pp. 21–59. New York, NY: Cambridge Univeristy Press.

Ritov, I., Gati, I. & Tversky, A. (1990). Differential weighting of common and distinctive components. *Journal of Experimental Psychology: General 119*(1), 30–41.

Rosenberg, S. & Kim, M. P. (1975). The method of sorting as a data-generating procedure in multivariate research. *Multivariate Behavioral Research 10*, 489–502.

Russell, S. J. (1986). A quantitative analysis of analogy by similarity. In *Proceedings of the National Conference on Artificial Intelligence*, Philadelphia, PA, pp. 284–288. AAAI.

Sattath, S. & Tversky, A. (1987). On the relation between common and distinctive feature models. *Psychological Review 94*(1), 16–22.

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

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika 22*(4), 325–345.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science 237*, 1317–1323.

Shepard, R. N. (1994). Perceptual-cognitive universals as reflections of the world. *Psychonomic Bulletin & Review 1*(1), 2–28.

Shepard, R. N. & Arabie, P. (1979). Additive clustering representations of similarities as combinations of discrete overlapping properties. *Psychological Review 86*(2), 87–123.

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

Shepard, R. N., Kilpatrick, D. W. & Cunningham, J. P. (1975). The internal representation of numbers. *Cognitive Psychology 7*, 82–138.

Tenenbaum, J. B. (1996). Learning the structure of similarity. In D. S. Touretzky, M. C. Mozer, & M. E. Hasselmo (Eds.), *Advances in neural information processing systems*, Volume 8, pp. 3–9. Cambridge, MA: MIT Press.

Tversky, A. (1977). Features of similarity. *Psychological Review 84*(4), 327–352.