# Description of ALCOVE

Catlearn Research Group*

August 17, 2016

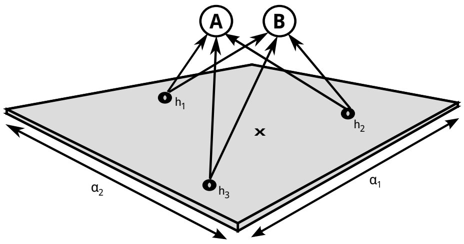


Figure 1: Architecture of the ALCOVE model (Kruschke, 1992). The gray quadrilateral is a three-dimensional depiction of a two-dimensional plane, representing a psychological stimulus space. Points $h_1$, $h_2$, and $h_3$ represent the location of radial-basis ("exemplar") units within that space. Point $\mathbf{x}$ represents the presented stimulus. The lettered circles are category representations, and the arrows connecting to them are variable-strength connection weights from the radial-basis units. $\alpha_1$ and $\alpha_2$ represent the attention allocated to each of the two dimensions of the space; the arrows beside these $\alpha$ illustrate that dimensional allocation acts to stretch and squash psychological space in ALCOVE. Image author: Andy J. Wills. CC BY 4.0.

Figure 1 summarizes the architecture of the ALCOVE model (Kruschke, 1992). ALCOVE is a connectionist model that assumes stimuli are represented as points in a multidimensional psychological stimulus space (Figure 2A). Thus, each stimulus is represented by a vector, which we will denote here as $\mathbf{x}$. For example, for stimuli varying in size and angle, one might write $\mathbf{x} = \begin{pmatrix} 0.4 & 0.5 \end{pmatrix}$, where the two values represent the psychological size and angle of the presented stimulus.

Presentation of a stimulus leads to the activation of radial-basis nodes (Cheney, 1966). Radial-basis nodes, like stimulus representations, can be considered as points in stimulus space (Figure 1). In virtually all applications of ALCOVE there is exactly one radial-basis node for each unique training stimulus. Although these radial-basis nodes are often called "exemplar" nodes, this description is something of a misnomer as, in most applications, all the nodes exist before training begins. It is perhaps better to think of these radial-basis nodes as a simplification of the abstract concept behind ALCOVE, which is that there are radial-basis nodes randomly scattered across stimulus space (the "COVEring map" of ALCOVE).

However one prefers to think about it, the architecture of the radial-basis layer of ALCOVE is fully specified by the matrix $\mathbf{h}$, which has the same number

*Email: andy@willslab.co.uk

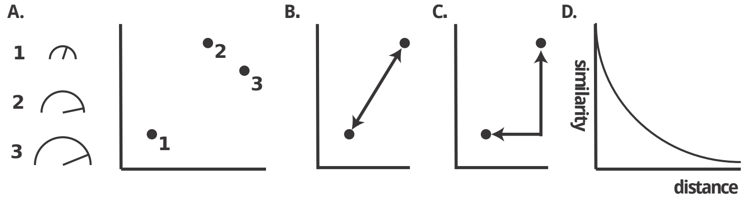


Figure 2: A. Representing the similarity structure of stimuli 1, 2, and 3 in a two-dimensional geometric space; in this example, the dimensions of this space are readily interpretable as size and angle. B. Euclidean distance ($distance^2 = x^2 + y^2$). C. City-block distance ($distance = x + y$). D. An exponential decay relationship between similarity and distance in psychological space. Reproduced from "On the Adequacy of Current Empirical Evaluations of Formal Models of Categorization" by A.J. Wills & E.M. Pothos, 2012, Psychological Bulletin, 138, 102–125. Copyright 2012 by American Psychological Association.

of columns as there are radial-basis units ($j$ columns), and the same number of rows ($i$ rows) as there are psychological stimulus dimensions. For example, in a simple size-angle experiment with four stimuli, the architecture of the radial-basis layer might be described as

$$\mathbf{h} = \begin{pmatrix} 0.4 & 0.8 & 0.8 & 0.4 \\ 0.4 & 0.4 & 0.8 & 0.8 \end{pmatrix}$$

where each column represents the location of one training exemplar in stimulus space.

ALCOVE computes the activations of each of the radial basis nodes with the following equation:

$$a_j^h = exp[-c(\sum_i \alpha_i |h_{ji} - x_i|^r)^{q/r}] \quad (1)$$

This equation specifies that the activation of each radial-basis node is a decreasing function of its distance from the presented stimulus. Where $r = 2$, that distance is Euclidean (Figure 2B); where $r = 1$, the distance is city-block (Figure 2C). Euclidean distance is typically used for integral stimuli, city-block for separable stimuli (see Garner, 1976). Where $q = 1$, the decreasing function is exponential (Figure 2D); where $q = 2$, it is Gaussian. Exponential decay is typically used (Shepard, 1987); occasionally Gaussian decay is used where stimuli are highly confusable (Ennis, 1988).

Stimulus space can be uniformly contracted or expanded using $c$ (see Figure 3C). $c$ is largely treated as an arbitrarily variable parameter (Wills & Pothos, 2012), although psychologically it is intended to represent cognitive discriminability or memorability of stimuli, so if information about this could be derived independently for a set of stimuli then it would constrain model fitting somewhat (minimally, $c$ would need to be a non-decreasing function of discriminability/memorability). In related models, application to amnesic data takes this form (e.g. Nosofsky & Zaki, 1998).

In Equation 1, $\alpha_i$ represents dimensional attention on dimension $i$. Dimensional attention acts as a multiplier to distance, stretching psychological space uniformly across one axis (see Figure 3B). The model is typically initialised

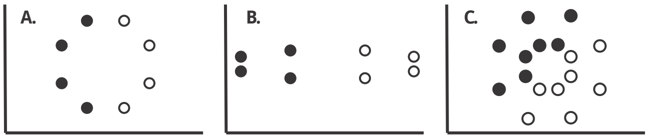


Figure 3: A. Geometric representation of two categories, each of four stimuli (category membership denoted by color of dot). B. Stretching along the $x$ axis and compression along the $y$ axis, thereby increasing within-category similarity and decreasing between-category similarity. C. Overall expansion of psychological similarity space. Reproduced from "On the Adequacy of Current Empirical Evaluations of Formal Models of Categorization" by A.J. Wills & E.M. Pothos, 2012, Psychological Bulletin, 138, 102–125. Copyright 2012 by American Psychological Association.

with equal attention to all dimensions, conventionally summing to unity, e.g. $\alpha = \begin{pmatrix} 0.5 & 0.5 \end{pmatrix}$.

The activation process represented by Equation 1 results in a vector of radial-basis node activations, e.g. $\mathbf{a^h} = \begin{pmatrix} 0.4 & 0.8 & 0.4 & 0.8 \end{pmatrix}$. Radial-basis node activation propagates forward to a set of output (category) nodes, which then have activation $\mathbf{a^o}$. There is one output node for each category, and each radial-basis node has one variable weight connection to each output node. The weight-state of the model is thus a matrix of the following form:

$$\mathbf{w} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

which has $k$ rows (one for each ouput node) and $j$ columns (one for each hidden node). Activation of the output nodes is calculated with the standard connectionist equation

$$a_k^o = \sum_j w_{kj} a_j^h \tag{2}$$

The forward propagation of activation ends with a standard exponential ratio rule to convert activation to response probability

$$P(K) = exp(\phi a_K^o) / \sum_k exp(\phi_k^o) \tag{3}$$

where $\phi$ is a non-negative response-scaling parameter. Low values of $\phi$ lead to approximately probability-matching behavior (category selection probability is proportional to the ratio of output node activations). High values of $\phi$ lead to approximately winner-take-all behavior (the category with the highest activation is always selected). We note in passing that this exponential ratio rule is probably a poor model of categorical decisions (Wills et al., 2000).

In ALCOVE, learning is driven by "teacher" ($t$) values. The presence of a category label is represented by a teacher signal of 1; absence of a category label is typically represented by -1. The teacher is typically considered to be "humble". This means that if the output activation is more extreme than the +1/-1 teaching value, then the ouput activation is used as the teaching signal.

Learning of connection weights from radial-basis nodes to output nodes uses a standard summed-error term equation[1]

[1] See Le Pelley (2004) for a discussion of summed- and separate- error term equations.

$$\Delta w_{kj} = \lambda_w (t_k - a_k^o) a_j^h \tag{4}$$

where $\lambda_w$ is the associative learning-rate parameter, which can range from 0 to 1. This equation acts to change connection weights in the direction that most rapidly reduces error.

Attentional weights are also learned. This is achieved by the backpropagation of error (Rumelhart et al., 1986; Werbos, 1974) to the radial-basis nodes in the standard manner:

$$b_j = a_j^h \sum_k (t_k - a_k^o) w_{kj} \tag{5}$$

This back-propagated error is then used to change the attentional weight for each dimension:

$$\Delta \alpha_i = -\lambda_\alpha \sum_j b_j c |h_{ji} - x_i| \tag{6}$$

where $\lambda_\alpha$ is the attention learning-rate parameter, which again can range from zero to one. Implementations of ALCOVE constrain attentional weights to be non-negative. ALCOVE's attentional learning system acts to stretch and squash psychological stimulus space (Figure 3B) in the directions that most rapidly reduce error.

# References


Cheney, E. (1966). *Introduction to approximation theory.* New York: McGraw-Hill.

Ennis, D. M. (1988). Confusable and discriminable stimuli: Comment on Nosofsky (1986) and Shepard (1986). *Journal of Experimental Psychology: General*, *117*, 408–411.

Garner, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology*, *123*, 98–123.

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

Le Pelley, M. E. (2004). The role of associative history in models of associative learning: A selective review and a hybrid model. *Quarterly Journal of Experimental Psychology*, *57B*, 193–243.

Nosofsky, R. M., & Zaki, S. R. (1998). Dissociation between categorization and recognition in amnesic and normal individuals: An exemplar-based interpretation. *Psychological Science*, *9*, 247–255.

Rumelhart, D. E., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In D. E. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1* (pp. 318–362). Cambridge, MA: MIT Press.

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

Werbos, P. J. (1974). *Beyond regression: New tools for prediction and analysis in the behavioral sciences* (Unpublished doctoral dissertation). Harvard University.

Wills, A. J., & Pothos, E. M. (2012). On the adequacy of current empirical evaluations of formal models of categorization. *Psychological Bulletin*, *138*, 102–125.

Wills, A. J., Reimers, S., Stewart, N., Suret, M. B., & McLaren, I. P. L. (2000). Tests of the ratio rule in categorization. *Quarterly Journal of Experimental Psychology*, *53A*, 983–1011.



**ATTRIBUTION** Please attribute the text of this document to the *Catlearn Research Group.* Current and previous contributors to the CRG: Andy J. Wills, Garret O'Connell, C.E.R. Edmunds, and Angus B. Inkster.