To cite this chapter:

Vanpaemel, W., & Lee, M. (2007). A model of building representations for category learning. In D. McNamara & J. Trafton (Eds.), *Proceedings of the 29th annual conference of the cognitive science society* (pp. 1605-1610). Cognitive Science Society.

A Model of Building Representations for Category Learning

Wolf Vanpaemel (wolf.vanpaemel@psy.kuleuven.be)
Department of Psychology, K.U. Leuven
Tiensestraat 102, B-3000, Leuven, Belgium

Michael D. Lee (mdlee@uci.edu)
Department of Cognitive Sciences, University of California, Irvine
Irvine, CA, 92697-5100

Abstract

We develop a model of the interaction between representation building and category learning. Our model is a hierarchical extension of Nosofsky's (1986) Generalized Context Model of category learning, based on the extended set of stimulus representation possibilities developed by Vanpaemel, Storms and Ons (2005). Using Bayesian inference, the model provides an account of the representation people are using, and what process generated that representation. We apply the model to data sets from four category learning tasks, and demonstrate how the results inform the prototype vs exemplar representation debate, and the similarity- vs rule-based categorization debate.

Keywords: Category Representation; Hierarchical Bayesian Models; Generalized Context Model

Introduction

A fundamental challenge for cognitive science is to understand the interaction between building stimulus representations and learning category structures. When people learn categorical associations, they usually rely, in part, on internal representations of stimuli. Learning whether or not a newly encountered breed of dog is dangerous is facilitated by thinking about the known hostilities of familiar dogs. In this sense, stimulus representations are the building blocks for learning category structures.

Most representations, however, are themselves learned. And often category membership is a key source of the stimulus similarity that can guide the building of representations. The similarity between Las Vegas and Atlantic City derives, in large part, from their shared association to the gambling industry. In this sense, category structures are building blocks for learning stimulus representations.

Many successful models of category learning and stimulus representation make assumptions along these lines. Exemplar-based models of category learning, like the Generalized Context Model (GCM: Nosofsky, 1986), make categorization decisions using memory representations of related stimuli. Similarity-based models of representation, including dimensional and featural models (Goldstone, 1999), rely on similarity judgments sensitive to shared categorical associations between stimuli.

Rarely, though, is the tight coupling of representational and category learning processes modeled. Models of category learning usually assume a fixed stimulus representation. Often these representations are derived using similarity-based models, and sometimes the representations are manipulated during category learning by processes like selective attention, but more fundamental representational adaptation during category learning is not accommodated. Similarity-based representational modeling, on the other hand, has as its end goal the learning of the representations, and does not extend to accounting for category learning behavior.

We describe a hierarchical model that relates a representation building process to category learning behavior. It builds on the Varying Abstraction Model (VAM: Vanpaemel, Storms, & Ons, 2005). The VAM was designed to address the debate regarding the merits of exemplar- and prototype-based models of category learning. The VAM does this by specifying a class of representations for learning category structures that includes exemplar and prototype representations as special cases. Using the category learning processes of the GCM, the VAM makes inferences about representations, based on the decisions made in a category learning task. In effect, the VAM provides a model of *what* representations might be used for category learning. Our extension is to include a representational process describing *how* those representations might be generated. With this extension, we can make inferences about the representation building processes used to support their learning of categories.

In this paper, we test the hierarchical extension of the VAM by re-analyzing data from seminal category learning tasks (Nosofsky, 1986; Nosofsky, Clark, & Shin, 1989). These experiments involve single subjects or groups of subjects learning various two-category structures over a small number of stimuli, all of which have two-dimensional spatial representations. The collected data measure the way the trained stimuli, and a set of additional stimuli not seen in training, were categorized.

A Representation Building Process

The Class of Representations

The VAM considers all possible representations that can be obtained by merging the stimuli belonging to each category. The exemplar representation is the one where no stimuli are merged. The prototype representation is the one where all stimuli are merged. Intermediate possibilities involve some of the stimuli being merged.

The bottom row of Figure 1 provides a concrete example of VAM representations. Panel A shows the exemplar representation of four stimuli in two-dimensional space. Panel B shows the representation created when two of the stimuli are merged, with the original stimuli shown as smaller squares, joined by lines to their merged representation.

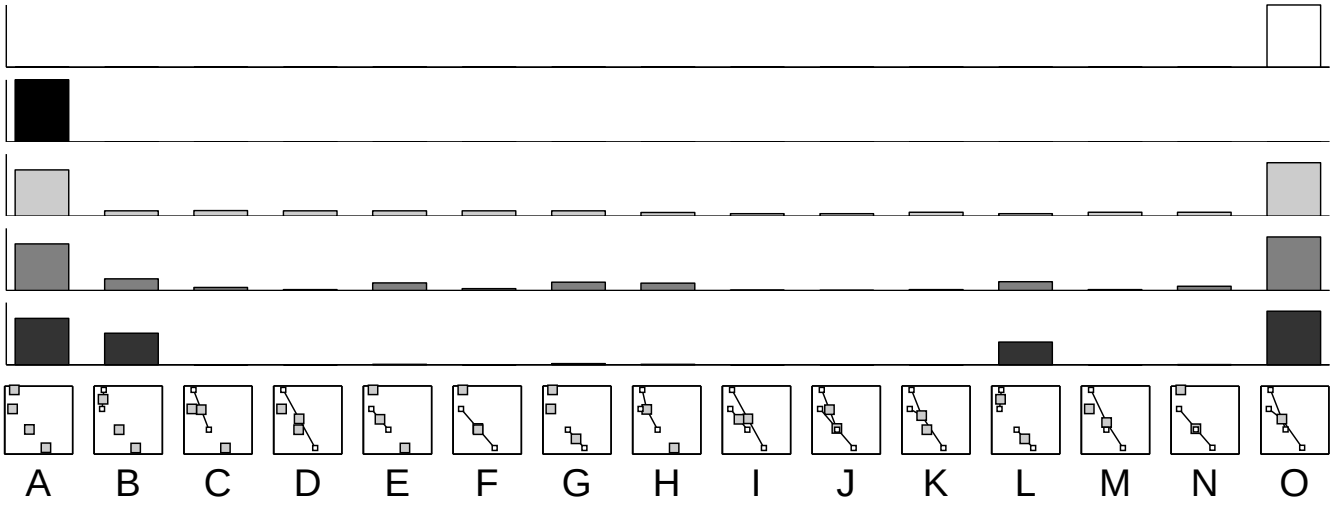

Figure 1: The bottom row shows the 15 possible VAM representations for a four-stimulus category structure. The top five rows give the probability distribution over these 15 representations for the generation process, corresponding to parameterizations $\theta = .99$, $\gamma = 1$ (white), $\theta = .01$, $\gamma = 1$ (black), $\theta = .7$, $\gamma = 0$ (light gray), $\theta = .7$, $\gamma = 1$ (medium gray), and $\theta = .7$, $\gamma = 10$ (dark gray).

The remaining panels in the bottom row of Figure 1 show the VAM representations resulting from averaging other stimuli. Panels B–G show the results of single merge, leaving three representations, while Panels H–N show the results of two merges, leaving two representations. The final VAM representation in Panel O shows the prototype representation in which all four stimuli are merged into a single representation.

The Generation of VAM Representations

We propose an iterative process for generating the class of VAM representations. The process has two parts; one controlling how many merges are made, and another deciding which stimuli are merged. Formally, $0 \leq \theta \leq 1$ is a parameter giving the probability that an additional merge will take place, and the iterative process will continue. This means, at any stage, there is a $1-\theta$ probability that the current representation will be maintained as the final one.

When an additional merge is undertaken, it is based on the similarities between all of the current representations (i.e., the original stimuli, or their merged replacement). The similarity between the ith and jth representations is, consistent with the GCM, modeled as an exponentially decaying function of the distance between their points, according to a Minkowski r-metric:

$$s_{ij} = \exp\left\{ -\left[\sum_k \left(\left| v_{ik} - v_{jk} \right|^r \right) \right]^{1/r} \right\}. \qquad (1)$$

Given these similarities, across all pairs in the current representation, the probability, m_{ij}, of choosing to merge the pair (i, j) is given by an exponentiated Luce-choice rule

$$m_{ij} = \frac{(\exp s_{ij})^\gamma}{\sum_x \sum_{y \geq x} (\exp s_{xy})^\gamma}. \qquad (2)$$

The parameter $\gamma \geq 0$ controls the level of emphasis given to similarity in determining the pair to be merged. As γ increases, the maximally similar pair dominates the others, and will be chosen as the pair to be merged with probability approaching one. At the other extreme, when $\gamma = 0$, similarity is not taken into account. All choices of pairs to merge then are equally likely, and the merge is essentially chosen at random. Values of γ between these two extremes result in intermediate behavior.

Given a value for the θ and γ parameters, every VAM representation has some probability of being generated by the process just described. The top five rows in Figure 1 show the probability of the VAM representations being generated for different parameters. In the top row $\theta = 0.99$, so merging is very likely, and hence the prototype representation almost always results. In the second row $\theta = 0.01$, so merging is very unlikely, and hence the exemplar representation is almost always retained.

The third, fourth and fifth rows show, for a fixed $\theta = 0.7$, the effect of the γ parameter. When $\gamma = 0$ in the third row, the exemplar and prototype representations are most likely, but all others are possible. In particular, any representation arising from a single merge is equally likely, and any representation arising from two merges is equally likely, because the pair of stimuli to be merged is chosen at random. In the fourth row, when $\gamma = 1$, representations like B and L that involve merging similar stimuli become much more likely, although some other possibilities remain. Once $\gamma = 10$ in the fifth row, only the most similar stimuli are merged, and B and L are the only intermediate possibilities between exemplar and prototype representation with non-negligible probability.

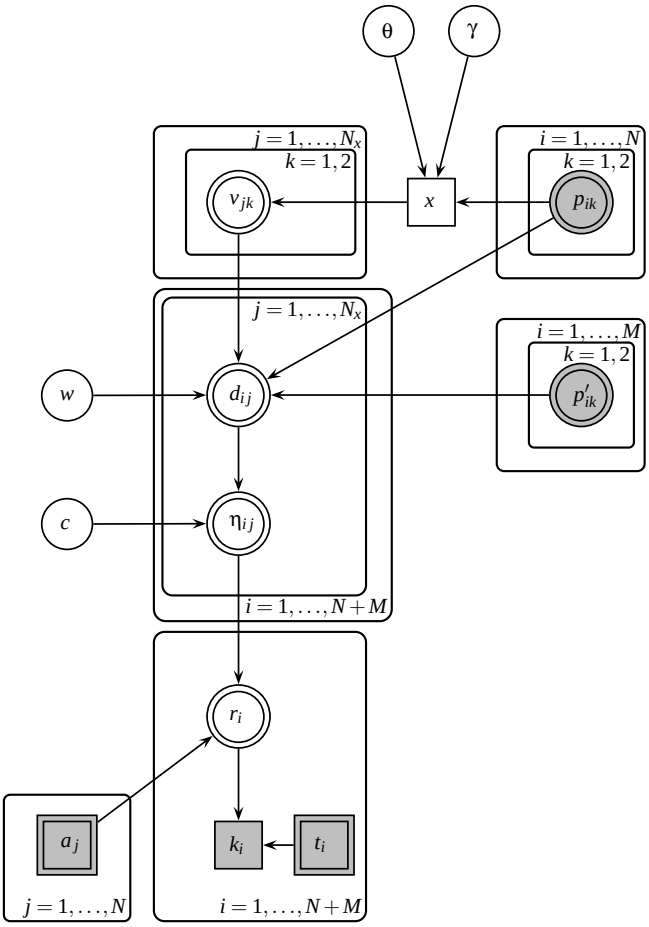

Figure 2: Graphical model representation of the hierarchical VAM used in the applications.

The Hierarchical VAM

Graphical Model Notation Figure 2 shows as a graphical model the specific adaptation of the hierarchical VAM we applied to re-analyze the Nosofsky (1986) and Nosofsky et al. (1989) data. Figure 2 uses a standard graphical model representation (Jordan, 2004). Nodes represent the labeled variables. The directed graph structure indicates dependencies between the variables, with children depending on their parents. Stochastic variables have single-borders and deterministic variables have double-borders. Observed variables have shading and unobserved variables have no shading. Continuous variables have circular nodes and discrete variables have square nodes. Independent replications in the model are represented by enclosing parts of the graph in square boundaries called plates, and are labeled by the indexing of the replications.

Representation At the top of Figure 2 are the coordinate locations p_{ik} and p'_{ik} for the N training and M additional stimuli, respectively, in $k = 1,2$ dimensions, as found by previous multidimensional scaling analysis. The training stimuli are the ones assigned to categories, and so are the basis for the representation building process.

Formally, the parameters θ and γ determine the index x of the VAM representational class, which we write

$$x \sim \text{Merge}(\theta,\gamma). \quad (3)$$

This index defines the N_x points in the VAM representation, with v_i denoting the ith of these representations. We used Monte Carlo estimates of $p(x \mid \theta,\gamma)$ to define the Merge distribution, found by simulating the iterative process over the stimuli and category structures used in the applications across the grid $\theta = (0.025, 0.05, \ldots, 0.975)$ and $\gamma = (0, 0.1, \ldots, 10)$.

For this representation component of the model, we use priors

$$\begin{aligned} \theta &\sim \text{Uniform}(0,1), \\ \gamma &\sim \text{Erlang}(1). \end{aligned} \quad (4)$$

The uniform prior for the rate θ is an obvious choice. The Erlang prior for γ gives support to all positive values, but has most density around the modal value one, corresponding to our prior expectations.

Categorization Having generated the VAM representation, the remainder of the model deals with the categorization process, and follows closely the GCM. The only difference is that the category similarity of the $N+M$ stimuli presented to participants is formed by summing over their similarities to the N_x representations constituting the VAM representation.

First, the attention-weighted distances between the stimuli and representations are calculated, according the Minkowski r-metric, so that

$$d_{ij} = \left(\left[w\,(p_{i1} - v_{j1}) \right]^r + \left[(1-w)\,(p_{i2} - v_{j2}) \right]^r \right)^{\frac{1}{r}}, \quad (5)$$

for the training stimuli, and analogously for the additional stimuli, where w is the attention weight parameter measuring the relative emphasis given to the first stimulus dimension over the second.

From the distances, the generalization gradient with scale parameter c and shape α determines the similarities,

$$\eta_{ij} = \exp\left\{ -c d_{ij}^{\alpha} \right\}. \quad (6)$$

The assignment of the representations to the two categories is defined by the category structure, and is given by indicator variables, so that $a_j = 1$ means the jth representation belongs to Category A, with $a_j = 0$ otherwise. For the current reanalyses we ignore the possibility of response bias, and so the probability of the ith stimulus being chosen as a member of the Category A is determined by the sum of similarities between the ith stimulus to the N_x representations in each category, according to a Luce choice rule,

$$r_i = \frac{\sum_j a_j \eta_{ij}}{\sum_j a_j \eta_{ij} + \sum_j (1 - a_j)\,\eta_{ij}}. \quad (7)$$

Finally, the response probabilities are used to account for the observed data, D, which are the counts, k_i of the number

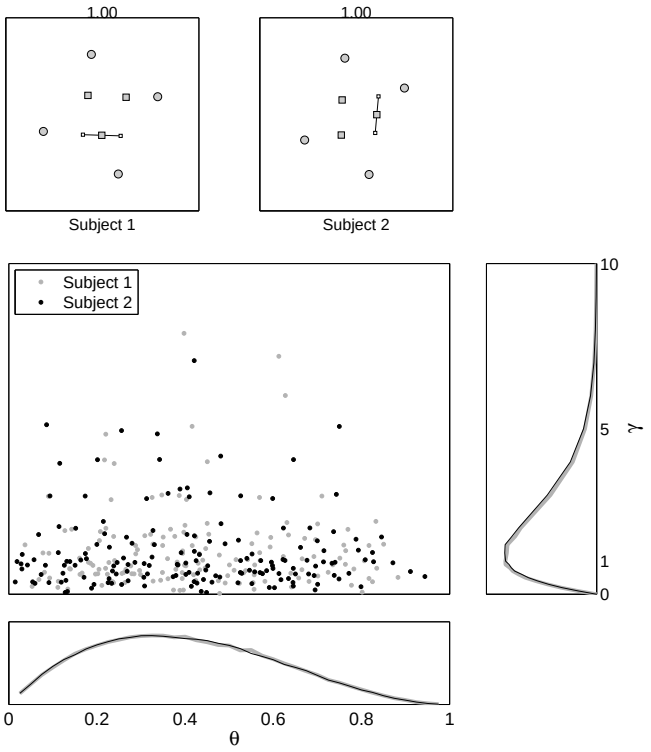

Figure 3: Hierarchical VAM account of the Nosofsky (1986) Interior-Exterior category data.

of times the ith stimulus was chosen in Category A out of the t_i trials it was presented. The counts k_i follow a Binomial distribution

$$k_i \sim \text{Binomial}(t_i, r_i). \quad (8)$$

For this categorization component of the model, we use priors

$$\begin{aligned} w &\sim \text{Uniform}(0,1), \\ c^2 &\sim \text{Gamma}(\varepsilon,\varepsilon). \end{aligned} \quad (9)$$

The uniform distributions for w is again an obvious choice. The c parameter functions as an inverse scale (i.e., $1/c$ scales the distances), implying c^2 functions as a precision, and so is given the standard near non-informative Gamma prior with $\varepsilon = .001$ set near zero.

Applications

This section presents four applications of our model to the seminal category learning data reported and analyzed by the GCM in Nosofsky (1986) and Nosofsky et al. (1989). We consider seven of the eight available data sets, leaving the 'Dimensional' category structure, because its heavy reliance on selective attention manipulations is not well accommodated by our current model. We follow the original analyses in using $r = \alpha = 2$ for Nosofsky (1986), and $r = \alpha = 1$ for Nosofsky et al. (1989).

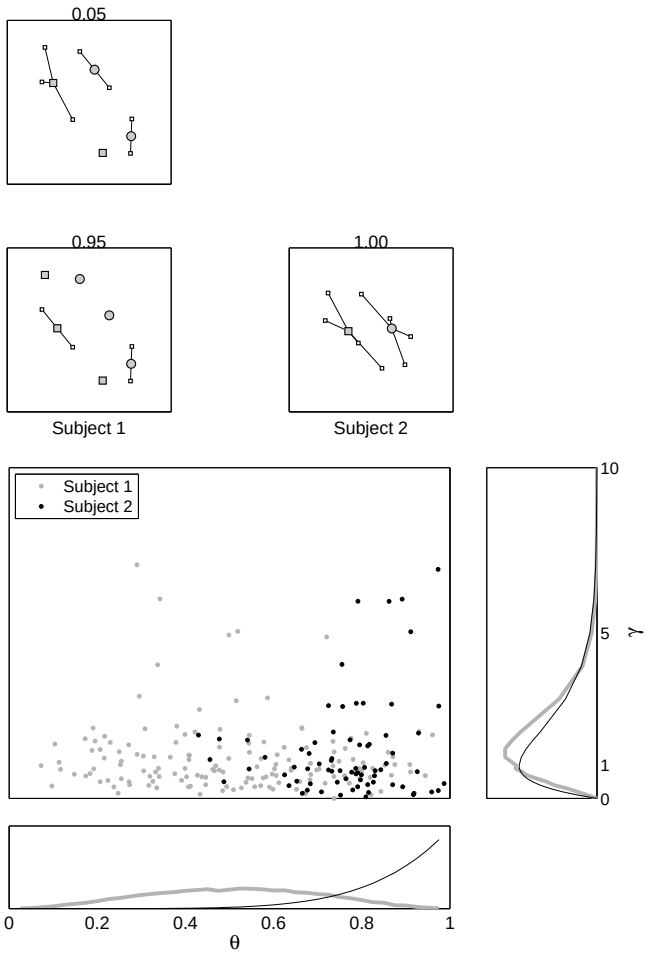

Figure 4: Hierarchical VAM account of the Nosofsky (1986) Diagonal category data.

Our primary interest is on two posterior distributions: $x \mid D$, which describes the inferences made by the model about what VAM representation is being used; and $(\theta, \gamma) \mid D$, which describes inferences about what process people used to generate that representation.

Interior-Exterior from Nosofsky (1986) Figure 3 summarizes the results from the Interior-Exterior category structure. The upper panels show the VAM representations inferred from the categorization decisions made by each subject. The stimuli for the two categories are shown by circles and squares, together with any merged representations. For both subjects, there is a single VAM representation that captured virtually all of the posterior probability. These representations are extremely similar, and involve a single merge of two of the stimuli in the interior category.

The lower panels of Figure 3 show the joint posterior distribution over θ and γ inferred from the data. The main central panel shows samples from this joint posterior for each subject. The side panels show the marginal distributions for θ and γ for each subject. These distributions are also extremely

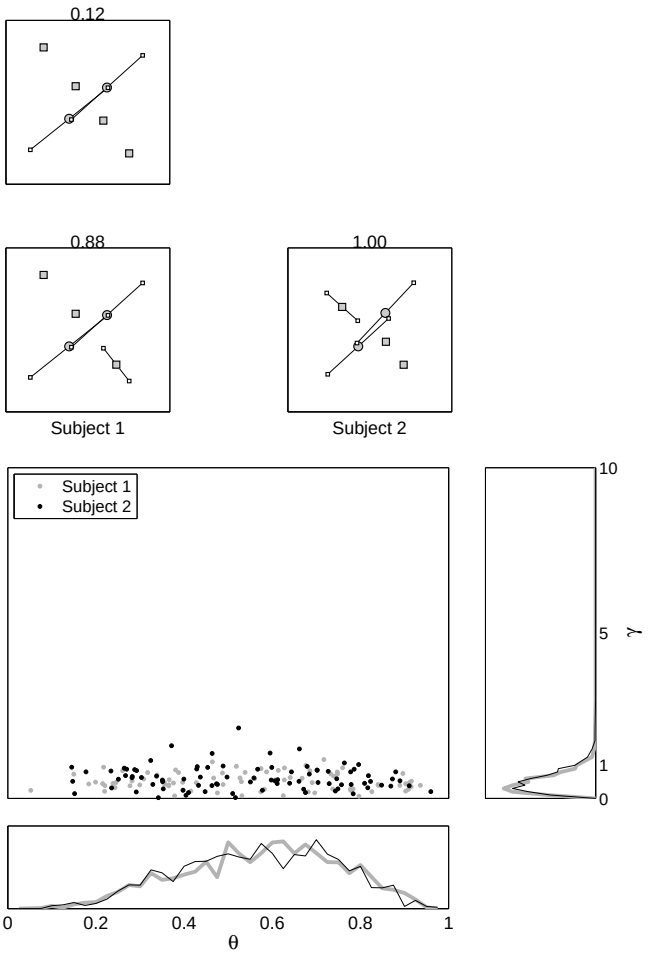

Figure 5: Hierarchical VAM account of the Nosofsky (1986) Criss-Cross category data..

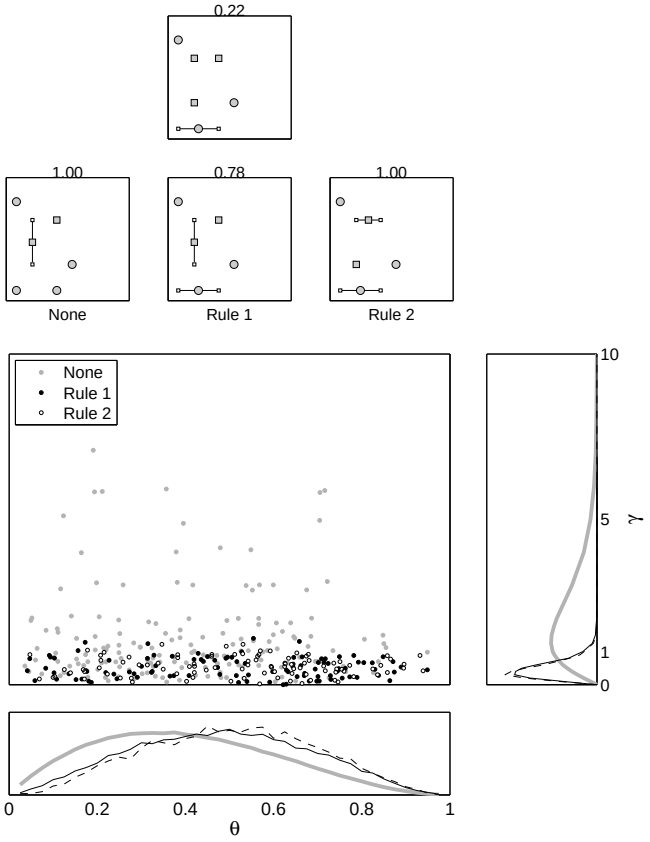

Figure 6: Hierarchical VAM account of the Nosofsky et al. (1989) category data.

similar for both subjects. In both cases, the value of θ is likely to be relatively low, indicating the use of a near-exemplar representation.

Diagonal from Nosofsky (1986) Figure 4 summarizes the results from the Diagonal category structure. Subject 1 has two VAM representations with posterior probabilities of 0.95 and 0.05, while the second subject has a single representation. In this case, there are significant individual differences, with Subject 1 using some merging, but maintaining many stimuli, while Subject 2 uses a prototype representation. The θ parameter captures these individual differences, taking values close to one for Subject 2, but lower values for Subject 1.

This model analysis provides a useful explanation of some striking features of the raw data in Nosofsky (1986, Table 3). In particular, the use of prototype representations results in a loss of sensitivity to the diagonal structure of the category boundary, and so explains the observation that the additional stimuli labeled '7' and '10', located just below and right, and just above and left of center respectively (see Nosofsky, 1986, Figure 6), will be categorized quite differently by the two subjects.

Criss-Cross from Nosofsky (1986) Figure 5 summarizes the results from the Criss-Cross category structure. The representations are similar for both subjects, and involve merging dissimilar stimuli within each category. The joint posteriors are also similar for both subjects, with the posterior for γ taking small values (i.e., less than one) indicating the merging of dissimilar stimuli.

These results suggest a deficiency in our representation building process. A more natural generative account of the representations in Figure 5 would involve the deletion of stimuli, rather than relying solely on merging for building representations.

Nosofsky et al. (1989) Nosofsky et al. (1989) considered a single category learning task, similar in structure to the Nosofsky (1986) Interior-Exterior task, but using different stimuli, and giving different instructions to three groups of subjects. The first group was given no special instructions, and so was expected to learn the category structure using the similarity-based principles that underly the GCM. The remaining two groups were instructed to use one of two simple rules accurately describing the category structure.

Figure 6 summarizes the results from the hierarchical VAM analysis of each group. The group with no special instructions are accounted for by a VAM representation that does follow stimulus similarity, by collapsing the similar stimuli in the interior category to a prototype, and largely preserving the less similar stimuli as exemplars in the exterior category.

The groups given the rule instructions, however, do not follow stimulus similarity closely, especially through their merging of the same two dissimilar exterior stimuli. An examination of the rules reveals that both had in common a logical proposition that directly corresponds to these two dissimilar stimuli, and so encouraged this merging.

As in the previous example, the posterior for γ neatly distinguishes whether or not representations were similarity-based, taking large values for the group given no special instructions, and values less than one for the two rule groups.

Discussion

The extension to the VAM developed here provides a process model for the class of representations it previously just assumed. The obvious benefit of this extension is that it permits inferences about the process of representation building. The applications demonstrated the ability of the model to make inferences about two theoretically interesting parameters: θ, measuring the extent of compression, which is relevant to the exemplar vs prototype debate; and γ, measuring to what extent compression is based on stimulus similarity, which is relevant to the similarity- vs rule-based categorization debate.

A further, perhaps less obvious, contribution of the representation building process is that it naturally overlays a sensible inductive bias on the VAM class of representations. A strength of the VAM is that it considers a wide range of representational possibilities, but a reasonable criticism is that it gives each of these equal prior status. Intuitively, some the VAM representations seem more reasonable than others, and the prior predictions made by the hierarchical extension, as shown in Figure 1 seem intuitively satisfying. In particular, there is a strong inductive bias towards the exemplar and prototype representations, as well as a bias towards similarity-based representational compression. Our last two applications show that these biases can be overcome by data, and so we believe the hierarchical VAM strikes the right balance of having theoretically-based expectations, without losing flexibility by simply assuming the basic tenets of those theories.

Against these strengths, the applications presented here suggest a deficiency in the particular representation building process we proposed. It seems, at least for some categorization tasks, people ignore a subset of the stimuli to learn the category structure. Future work intends to refine the current model with a more general representation building process that allows for this possibility.

Acknowledgments

WV was supported by the Research Council of the University of Leuven (IDO/02/004).

References

Goldstone, R. L. (1999). Similarity. In R. A. Wilson & F. C. Keil (Eds.), *MIT Encyclopedia of the Cognitive Sciences* (pp. 763–765). Cambridge, MA: MIT Press.

Jordan, M. I. (2004). Graphical models. *Statistical Science*, *19*, 140–155.

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

Nosofsky, R. M., Clark, S. E., & Shin, H. J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(2), 282–304.

Vanpaemel, W., Storms, G., & Ons, B. (2005). A varying abstraction model for categorization. In B. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 2277–2282). Mahwah, NJ: Erlbaum.