

INVITED ADDRESS

**MODELING INDIVIDUAL DIFFERENCES IN LEARNING
HIERARCHICALLY ORGANISED CATEGORIES**

Steven VERHEYEN & Gert STORMS

Katholieke Universiteit Leuven

It is commonly believed that exemplar models have difficulty accounting for more accurate classification of items at the lower level of a nested hierarchy than classification of these items at a more general level. We identified groups of similarly performing participants in an experiment where such a hierarchically organised category structure had to be mastered and fitted a differently parameterised version of the ALCOVE exemplar model to each of these groups. The exemplar model had no difficulty predicting better performance at the lower level of the hierarchy than at a more general level for those groups of participants that actually displayed the supposedly challenging behavioural pattern.

Ever since similarity has taken front stage in the study of classification (Rosch & Mervis, 1975) arguments on the nature of category representations have been going back and forth between proponents of two main theories. Proponents of prototype theory have argued for a central tendency representation (e.g., Hampton, 1979, 1993), while researchers adhering to exemplar theory believe that classification is dependent on the activation of memory traces of previously encountered category members; not on an abstract summary of these exemplars (e.g., Medin & Schaffer, 1978; Nosofsky, 1984, 1986). For several decades prototype and exemplar theory have been compared using a considerable variety of classification experiments, providing for the development of formal instantiations of both theories, but not for a stop to the controversy (Smith & Minda, 1998). New input to the ongoing discussion comes from category learning experiments in which participants have to master categories at multiple levels of abstraction.

Steven Verheyen is the recipient of the 2007 BAPS Best Thesis Award and was therefore invited to write the present article. He is a research assistant at Research Foundation-Flanders at the Department of Psychology, Katholieke Universiteit Leuven; Gert Storms is Professor at the Department of Psychology, Katholieke Universiteit Leuven.

This research was supported in part by grant OT/05/27 awarded to Gert Storms by the Research Council of the Katholieke Universiteit Leuven.

We would like to thank Eef Ameel, Bieke Bollen, Matthew J. Dry, and Wouter Voorspoels for their helpful comments on earlier versions of this manuscript.

Correspondence concerning this article should be addressed to Steven Verheyen, Department of Psychology, Katholieke Universiteit Leuven, Tiensestraat 102, 3000 Leuven. E-mail: steven.verheyen@psy.kuleuven.be

Although it has long been established that many of the entities we encounter in everyday life are referred to with multiple names (e.g., bulldog, dog, mammal, animal) indicating a hierarchical organisation of categories of different abstraction (Rosch, 1978), it has taken researchers dealing with classification a long time to incorporate this finding into their experimental paradigms. Until recently it was common to have participants only discriminate among categories at a *single* level of abstraction. The structure among categories of *various* abstraction was for the most part left untouched in traditional classification experiments. Studies by Lassaline, Wisniewski, and Medin (1992), Murphy (1991), and Murphy and Smith (1982) constitute notable exceptions for they show that people participating in a classification experiment which involves multiple levels of abstraction can make sense out of the task put before them. However successful prototype and exemplar models might be at dealing with classification at a single level of abstraction, for them to remain valid accounts of human classification performance, they should capture the behavioural patterns arising in these experiments with multiple levels of abstraction as well (Estes, 1993; Palmeri, 1999).

Doubts about formal classification models' ability to do so have primarily arisen against the class of exemplar models (Murphy, 2002). Unlike rival accounts of classification, exemplar models are suspected to have great difficulty accounting for more accurate classification of an item at the lower level of a nested hierarchy than classification of the item at a more general level (Palmeri, 1999). If these speculations would prove to be true, it could invalidate exemplar theory. Superior performance at the lower level of a nested hierarchy is after all reminiscent of the well-established faster and more accurate classification of items at the so-called basic level of abstraction (e.g., car) than at the superordinate level (e.g., vehicle) in natural language taxonomies (Rogers & Patterson, 2007; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976).

It is important to underline that the reservations towards exemplar theory are prompted by educated speculations regarding exemplar models' functioning; not by explicit simulation of empirical results. The argument made against exemplar theory generally starts with the assumption that – just as humans participating in an experiment – exemplar models have to rely upon the available evidence of membership in order to make correct classifications. Within exemplar models the evidence E_X for classifying a stimulus in a particular category X is accumulated by summing the similarity of each of the category exemplars to the stimulus. The probability $P(X)$ of assigning the stimulus to category X is generally taken to be the ratio of the evidence E_X to the evidence that the stimulus belongs to any of the categories at the abstraction level of the target category. If a hierarchically organised category structure is assumed, comprising at the one hand two specific categories C and D ,

which are subordinate to a more general category A, and on the other hand two additional specific categories E and F, constituting a general category B, the probability of classifying a particular stimulus in category A is then said to be

$$P(A) = E_A / (E_A + E_B),$$

while the probability of appointing the same stimulus to category C becomes

$$P(C) = E_C / (E_C + E_D + E_E + E_F).$$

Since in a true hierarchy the evidence for classifying a stimulus in a category can be considered the summed evidence for classifying the stimulus in the constituting categories, it becomes difficult to see how classification of the stimulus in one of those constituting categories can ever be superior to classification of the stimulus in the more general one. When

$$E_A = E_C + E_D \text{ and } E_B = E_E + E_F,$$

the denominators in the classification probability formulas will be identical at both levels of abstraction, while the numerator for classification of the stimulus in a general category is at least as high as the one for classification of the stimulus in one of the constituting specific categories. Exemplar models that rely on this classification strategy would therefore be incapable of predicting better category performance at the specific level of abstraction than at the more general level, it is argued.

Current exemplar models, however, have moved beyond the rather simple learning and response mechanisms discussed in the above argument. A widely used instantiation of exemplar theory like the ALCOVE model (Kruschke, 1992) for instance, can not only rely on attention learning to acquire the correct structure among categories, but also incorporates nonlinearities in its response rule. Each of these properties separately, or both of them combined, might exempt the ALCOVE model from the problem that was put forth in the above argument (Palmeri, 1999). Whether it actually does, was put to test by Verheyen, Ameel, Rogers, and Storms (2008) who designed an experiment which held the conditions said to prove difficult for exemplar models to master.

In their experiment Verheyen and colleagues (2008) had a group of undergraduate students perform a supervised category learning task on 12 schematic drawings of spaceships. These stimuli were modelled after the ones introduced by Hoffman and Ziesler (1983). They differed along four dimensions (shape of the nose, tail, wings, and porthole) and along each of these dimensions every spaceship took one of four possible shapes. The stimulus structure in Table 1 exemplifies how dimensions and shapes were com-

bined to form the 12 stimuli.

Table 1 also holds the category structure Verheyen et al. (2008) employed. It shows how the classification experiment held categories of various abstraction. Each of the 12 stimuli *at the same time* belonged to one of four specific categories (C, D, E, or F) and to one of two more general categories (A or B). The italicised values in the table highlight how participants had to pay attention to the first stimulus dimension in order to make correct classification decisions. Each of the four values (shapes) along the first dimension was associated with a specific category. By having values 1 and 2 along this dimension signify Category A membership and having values 2 and 3 signify membership of Category B, a hierarchical category structure was obtained. Categories C and D were subordinate to A, while E and F were subordinate to B.

Classification at the two levels of abstraction occurred in separate stages. First, participants were asked to discriminate between the two general categories. Upon presentation of a stimulus participants indicated their decision and feedback on the correctness of this response was provided. Each one of the 12 stimuli was presented once per block for a total of 25 blocks. After completing the first learning stage, participants were required to classify the 12 stimuli into the four specific categories. It was stressed that the 12 stimuli were identical to those in the first stage, but no further information on the relationship between both learning tasks was provided. In the second learning stage, participants again completed 25 blocks of trials.

Verheyen and colleagues (2008) found that participants' performance at the specific level of abstraction was superior to that at the general level. On average, the probability of committing an erroneous classification was lower when stimuli were to be placed in one of the lower level categories than when they were to be put in one of the general ones. This is exactly the kind of result exemplar models should find difficult to capture according to the argument introduced above. However, when Verheyen and colleagues (2008) put the ALCOVE exemplar model to test on the results, they found that the model was quite adept to account for the behavioural pattern arising in the experiment. ALCOVE demonstrated higher accuracy in classifying at the specific level than in classifying at the general level.¹

¹The experiment of Verheyen and colleagues (2008) also held a condition in which a different group of participants had to learn the hierarchical category structure in the reverse order. First they learned to distinguish between the specific level categories, and then the distinction between general categories had to be mastered. We do not elaborate on this condition's results since they do not pertain to the discussion at hand (e.g., no specific level advantage was found). When in the remainder of the text the Verheyen et al. experiment (2008) is referred to or the term hierarchical classification experiment is employed, the condition of the experiment in which general categories were differentiated before specific ones is intended.

Accounting for individual differences

By carrying out explicit simulations of an experiment which held hierarchically organised categories, Verheyen et al. (2008) demonstrated how speculations regarding exemplar models' adequacy in dealing with classification at various levels of abstraction were unwarranted. At least one formal instantiation of exemplar theory, the connectionist ALCOVE model (Kruschke, 1992), had no problem accounting for higher accuracy of classification at the specific level of a hierarchy than at a more general level. However, even with these speculations regarding the model's functioning being dealt with, one might still question the demonstration of Verheyen et al. (2008) by pointing towards the fact they obtained model predictions for averaged learning curves.

It has long been recognised that averaging across participants carries the danger of obscuring important individual differences inherent to the data (Estes, 1956) and might even lead researchers in believing that a behavioural pattern not observed in any of the individual participants' data is characteristic of the population's behaviour (Ashby, Maddox, & Lee, 1994; Martin & Caramazza, 1980). Smith, Murray, and Minda (1997) have shown how in classification tasks an exemplar model might benefit from the averaging process. Since participants in their experiments could not be regarded invariants, averaging across them did not result in the desired reduction of measurement error, but in the distortion of the structure of the data. Amongst a subgroup of the participants, Smith et al. (1997) identified a pattern of performance that proved difficult for the employed exemplar model to capture. When the model was to predict the performance across *all* participants however, it encountered no difficulties since the challenging pattern was done away with by the averaging.

One way to tackle the finding by Smith et al. (1997) that aggregating across all participants is not always appropriate when evaluating an exemplar model, would be to attain model predictions for individual participants' classification behaviour (e.g., Ashby & Lee, 1991; Cohen & Nosofsky, 2003; McKinley & Nosofsky, 1996). Although it guarantees that the structure of the data is honoured, this approach is not without disadvantages either (Lee & Webb, 2005; Webb & Lee, 2004). By looking at how individuals are different instead of how they are similar, one runs the risk of providing an overly complicated account of the phenomenon under study. Each time a formal model is fitted to an individual's data, an extra set of parameters is required, while this is not necessarily very informative. The extra estimated parameters might point towards processes already found in other individuals, or might just reflect the fitting of noise in the individual's data.

Another means of dealing with the issue would constitute identifying sub-

groups of similarly performing participants (e.g., Palmeri & Nosofsky, 1995; Smith et al., 1997; Vanpaemel & Navarro, 2007). In a way, this approach achieves the best of modeling endeavours at the aggregated and at the individual level. By allowing for subgroups one can get at the differences that might exist among participants, while averaging across the participants in each one of the subgroups reduces the measurement error. A model can then be fit to each subgroup's averaged data. Providing these subgroups were put together in a sensible way, the differences in parameterisation of the estimated model should testify to important ways in which the groups differ.

Bayesian model selection criteria can be applied to determine the appropriate number of subgroups to consider (Lee & Webb, 2005; Webb & Lee, 2004). The more subgroups are withheld, the more complicated the account of the phenomenon under study becomes, for it requires extra models to be fitted. Despite the improvement in quantitative measures of fit this might invoke, the extra models might not add much to our understanding of the data (Myung, 2000). Therefore, it is now commonly held that the goal of obtaining good quantitative fits to behavioural data should be offset by moderation in the complexity of the formal account given (Myung & Pitt, 1997). The Bayesian Information Criterion (BIC; Schwarz, 1978) is an index that achieves this: it balances the goodness of the fit with the complexity of the models employed.

If differences in classification performance exist among participants discriminating among categories at a single level of abstraction, individual differences can also be expected to arise when the structure among categories of various levels of abstraction has to be inferred. In order to investigate whether (i) the behavioural pattern observed in the Verheyen et al. (2008) hierarchical classification experiment is not just an artefact of the averaging process and (ii) the exemplar model ALCOVE is still able to capture the fundamental behavioural patterns in hierarchical classification when individual differences are allowed for, the modeling performed by Verheyen and colleagues (2008) at the aggregated level was undertaken at the level of groups of similarly performing participants. Identification of these subgroups of participants was achieved through the procedure introduced by Webb and Lee (2004). Participants were first grouped according to the similarity of their learning curves. Since at the level of a subgroup participants can then be said to be invariants, we averaged across the participants in a group and fitted ALCOVE to the resulting classification pattern. We employed the BIC to decide upon the appropriate number of groups to divide the participants in or the equivalent amount of ALCOVE parameterisations to tackle the empirical data with. In doing so, we showed that even when the aggregated level of analysis is left, superior classification performance at the lower level of a nested hierarchy is displayed and can be accounted for by the exemplar model ALCOVE.

Method

Partitioning participants

Eighteen undergraduate students participated in the Verheyen et al. experiment (2008). In two separate stages they acquired the category structure detailed in Table 1. For 25 blocks they classified the 12 artificial stimuli into two general categories. For another 25 blocks they classified the same 12 stimuli into four subordinate categories. This provided us with two separate learning curves for every participant; one at each level of abstraction. In order to divide the participants into sensible groups, we chose to investigate the similarity of their learning curves at the lower level of abstraction. By employing the curves from the second stage of the classification experiment, we meant to capture similarities (and differences) regarding the inference of a relationship between both stages. Since the challenge particular to the experiment was in the inference of the hierarchical nature of this relationship, we speculated that this would constitute an important source of individual differences. Participants who are fast at establishing the hierarchical relationship between both stages of the classification experiment should demonstrate a learning advantage over participants who are not. Since the former participants can bring the knowledge they acquired during the first learning stage to bear on the second one, they should exhibit more rapidly decreasing learning curves than the latter participants.

Table 1.
Stimulus and category structure of the hierarchical classification experiment.

| Stimulus | Stimulus structure | | | | Category structure | |
|----------|--------------------|----|----|----|--------------------|----------|
| | D1 | D2 | D3 | D4 | General | Specific |
| 1 | <i>1</i> | 1 | 3 | 4 | A | C |
| 2 | <i>1</i> | 2 | 4 | 1 | A | C |
| 3 | <i>1</i> | 3 | 1 | 2 | A | C |
| 4 | 2 | 4 | 2 | 1 | A | D |
| 5 | 2 | 1 | 3 | 2 | A | D |
| 6 | 2 | 2 | 4 | 3 | A | D |
| 7 | 3 | 3 | 1 | 3 | B | E |
| 8 | 3 | 4 | 2 | 4 | B | E |
| 9 | 3 | 1 | 3 | 1 | B | E |
| 10 | 4 | 2 | 4 | 2 | B | F |
| 11 | 4 | 3 | 2 | 3 | B | F |
| 12 | 4 | 4 | 1 | 4 | B | F |

Note. Italicised values indicate the dimension values defining category membership.

The 18 learning curves taken from the second stage of the classification experiment were correlated to obtain a measure of their similarity. These similarities were then subjected to a singular value decomposition. From this decomposition, only the singular vectors with a value greater than 1.0 were retained to form a representation suitable for the application of k-means clustering. Using the clustering algorithm, 7 partitions of the 18 participants were obtained. Each partition held a different number of participant groups, ranging from one group holding all 18 participants in the first partition, to seven groups, with differing numbers of participants in each group, in the seventh partition.

Model fitting

In all simulations we used the ALCOVE variant that employs discrete values along the input dimensions (Palmeri, 1999). When a stimulus is presented along this model's four input dimensions (corresponding to the four stimulus dimensions displayed in Table 1) it activates a set of 12 exemplars through the similarity of its representation to theirs. The degree to which a particular exemplar j is activated is given by

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

where the variable d_{ji} takes the value 0 when input stimulus and exemplar match along dimension i . It takes the value 1 when they do not. The extent of the activation is further determined by the learned strength of attention α_i , which is allocated to each input dimension, and the value of specificity parameter c , which acts as a scaling factor.

The exemplars' activations in turn elicit the activation a_X^{out} of every category node X through learned associations weights w_{jX} . The category nodes' activations

$$a_X^{out} = \sum_j w_{jX} a_j^{hid}$$

are finally converted to classification probabilities $P(X)$ according to a softmax function

$$P(X) = \exp(\phi a_X^{out}) / \sum_C \exp(\phi a_C^{out}),$$

where the sum in the denominator is taken across all categories at the abstraction level of the target category and ϕ is a response-mapping param-

ter. At this point feedback on the correctness of the model's response is provided and attention strength values α_i and association weights w_{jX} are adjusted accordingly through backpropagation. The rate with which these adjustments are made is determined by free parameters λ_α and λ_w , respectively. With c and ϕ to be estimated as well, this brings the total amount of free model parameters at four.

Maximum likelihood estimates for these parameters were obtained through the use of a Nelder-Mead simplex algorithm. A six (two general and four specific) category node version of the model outlined above was initialised by setting attention strengths and association weights to zero and awarding random starting values to the four free parameters. The model was then trained on a sequence of 25 x 12 randomly ordered stimuli. During this sequence attention values and association weights leading up to the two general category nodes were adjusted through backpropagation. When simulation of the first stage of the classification experiment was completed, the model was further trained on a second sequence of 25 x 12 randomly ordered stimuli. During this sequence the attention values continued to be updated through backpropagation after each trial. Association weights leading up to the four specific category nodes were adjusted in this manner as well, but those connected to the general category nodes were now left unchanged, for in the second stage of the classification experiment feedback was only provided on the correctness of responses at the specific level of abstraction.

For each of the different groups in a partition, this procedure was repeated 500 times, awarding different starting values to the free parameters with each repetition. Out of these 500 attempts to fit the averaged data of a group, the predictions by the model with the highest likelihood were withheld. This likelihood $p(D | \theta)$ of obtaining the averaged data of a group D , given the model's parameterisation θ , entered in the computation of the BIC for the partition *as a whole*

$$\text{BIC} = -2 \ln p(D | \theta^*) + P \ln N,$$

where $p(D | \theta^*)$ is the maximum likelihood across all groups in a partition (e.g., the multiplied likelihoods of the models that were fit to the separate groups in a certain partition), P is the number of parameters used to model a partition (e.g., the sum of all the parameters used by the models fit to the groups in a certain partition), and N is the total number of datapoints. Seven BIC values were computed in this manner, corresponding to the seven partitions that were obtained through k-means clustering.

Results and model predictions

Figure 1 shows the BIC values for the seven partitions obtained using the k-means clustering algorithm. The lowest BIC value is found for the partition involving five subgroups, indicating that the most adequate account of the behaviour demonstrated in the hierarchical classification experiment is provided by five different parameterisations of the ALCOVE model.

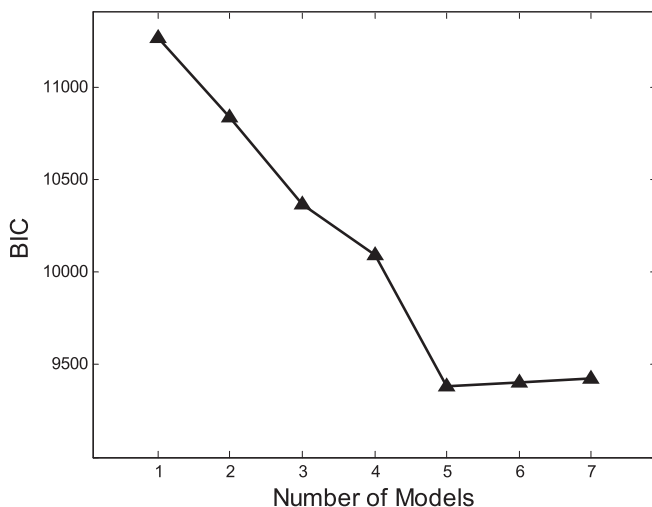


Figure 1.

The pattern of change in BIC values for seven partitions of the hierarchical classification data.

One of the five subgroups retained only held a single participant. This particular participant differed from all others in that she did not display learning – neither at the general level of abstraction, nor at the specific was her performance above chance level. Figure 2 displays in solid lines the averaged probability of committing a classification error, with regard to level of abstraction and learning block, for the four other retained groups. Panels A, B, and C show learning curves averaged across four, six, and five participants, respectively. They make up the majority of the 18 participants that entered in the clustering. Common to the classification behaviour displayed by the participants in these groups is that they show rapid learning of the specific categories in the beginning of the second learning stage. As a result, they display higher classification accuracy at the specific level of abstraction than at the general level. In other words, even when individual differences are

allowed for, the most compelling pattern arising from the analysis is similar to that found at the aggregated level of analysis.

Of course, the division of the 15 participants who showed better performance at the lower level of the nested hierarchy into three groups also testifies to important differences between them. Visual inspection of panels A, B, and C of Figure 2 suggests the main difference lies in the speed with which participants pick up on the category structure. The four participants of Group 1 steadily learn to distinguish between general categories A and B, but on average never achieve perfect performance for this level of abstraction. After six blocks of learning at the specific level of abstraction they do attain perfect scores. Classification performance of participants in Group 2 is highly similar to that of Group 1 participants at the specific level of abstraction. After six blocks of learning they too achieve perfect performance on this level of abstraction. At the general level of abstraction however, they show somewhat more rapid learning than the Group 1 participants, although they do not achieve perfect classification either. The five Group 3 participants are by far the fastest to pick up on the available category structure. They attain perfect performance at both levels of abstraction, and at the specific level of abstraction require fewer learning blocks than other participants to do so.

The fourth group constitutes two participants who display a classification pattern that differs from the large majority of people who participated in the hierarchical classification experiment. For 15 out of the 18 participants, we saw higher classification accuracy at the specific level of abstraction than at the general level from early on in the second stage of the experiment. In the two Group 4 participants this classification pattern only becomes apparent in the last four blocks of learning. Until then their performance at the specific level of abstraction had been much worse than that of the participants in Groups 1, 2, and 3, providing for a lesser classification accuracy overall.

In order to determine whether ALCOVE is able to account for human performance in the hierarchical classification experiment when individual differences are allowed for, we apply the commonly used approach of verifying whether the model captures the fundamental qualitative patterns of the constraining data (Lee & Navarro, 2002) to each of the identified participant groups.

To account for the performance of the one participant who did not display learning, the ALCOVE model fitted to her data predicted a .50 probability of committing a classification error at the general level of abstraction, and a .75 probability of committing such an error at the specific level. Since at the general level two category responses were available, and at the specific level four categories were to be chosen from, this corresponds to performance at chance level.

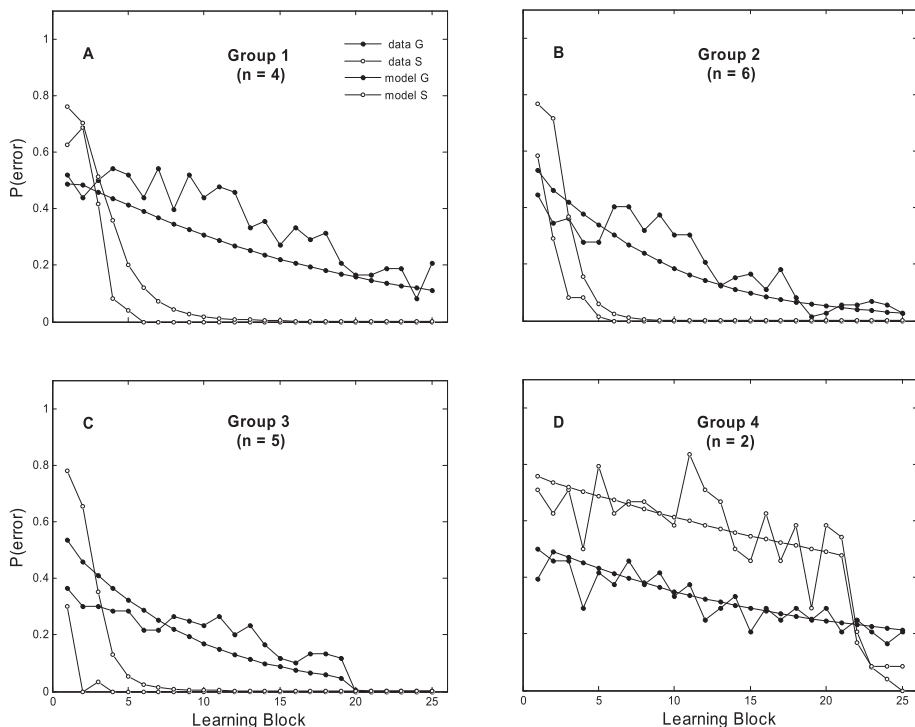


Figure 2.

Probability of error with regard to level of abstraction and learning block.

The general level of abstraction (G) is indicated by filled symbols, while the specific level (S) is indicated by open symbols. Solid lines represent averaged empirical data, while dotted lines represent the predictions by ALCOVE.

Panels A, B, C, and D each display the results for a subgroup of the participants.

The panels in Figure 2 hold in dotted lines, for each of the other participant groups, the predictions made by the best-fitting ALCOVE model. Among participants of Groups 1, 2, and 3 the most fundamental pattern that emerged was that of higher classification accuracy at the specific level of abstraction than that at the general level. In all three groups, the ALCOVE architecture proves appropriate for displaying this critical pattern of performance. It is able to do so by allowing for a rapid decrease in classification error at the specific level of abstraction during the first blocks of learning. Participants in the three groups differed with respect to their performance on the general level of abstraction, with Group 3 participants achieving higher accuracy at this level of abstraction than Group 2 participants, who outperformed participants from Group 1. The ALCOVE models' predictions dis-

play the same ordering of performance. Admittedly, the same is not entirely true for the models' predictions of classification errors at the specific level of abstraction. The stage at which participants in the three groups achieved perfect performance is not accurately reflected by the models' learning curves. This should not come as a surprise since the ALCOVE model was originally intended to capture discrimination of categories at a *single* level of abstraction. In this study, it is applied to a more complex task which requires the differentiation of categories at *multiple* levels of abstraction. In light of the little task-specific tailoring it was permitted, the model's difficulty capturing the subtleties of the presented empirical results should be regarded as insignificant compared to the more fundamental qualitative patterns it does produce.

The ALCOVE model also proved adequate when put to test on the data from two participants who did not display a specific level advantage until the latest blocks of learning. As was the case in the Group 4 participants' data, a steep decrease of the classification error at the specific level of abstraction was not observed in the model predictions until the 22nd block of learning.

Summary

At the onset of this manuscript we discussed arguments which were raised against exemplar models' adequacy when it comes to dealing with performance differences in a classification experiment involving multiple levels of abstraction. Superior performance at the lower level of a nested hierarchy, reminiscent of the basic level effect in natural language taxonomies, was alleged not to arise in exemplar models' predictions. Verheyen et al. (2008) demonstrated that at least one particular instantiation of exemplar theory, the connectionist ALCOVE model (Kruschke, 1992), did display this difference in performance when the constraining data asked for it. With its ability to allocate different weights to dimensions of varying informativeness and its nonlinear response rule, ALCOVE is not confined to predicting superior classification accuracy at the highest level of a hierarchical organisation of categories.

We continued to note that people willing to object against this demonstration, might refer to the work by Smith et al. (1997). They argued that the level of aggregation exemplar models should be fitted to is that of subgroups of similarly performing participants; not that of the entire group of participants as was done by Verheyen and colleagues (2008). In order to dismiss this argument against exemplar models' predictions of hierarchical classification behaviour, we identified groups of similarly performing participants in the Verheyen et al. experiment (2008), following a procedure developed by

Webb and Lee (2004). A differently parameterised version of the ALCOVE exemplar model was fit to each of these groups and Bayesian model selection criteria indicated that five different parameterisations of the model were required to capture the fundamental behavioural patterns arising in the experiment.

The Webb and Lee partitioning procedure identified three groups of participants that demonstrated the behavioural pattern previously established at the aggregate level of analysis: better performance when distinguishing categories at the specific level of abstraction than when differentiating between the more general categories. The ALCOVE model did not face difficulties mimicking the supposedly challenging behavioural pattern of these groups. Furthermore, the partitioning procedure set apart a group of only two participants who did not show the specific level advantage until the final learning blocks, and in addition singled out the one participant who did not display learning. The behavioural patterns of these two groups too were accounted for by the respective ALCOVE models' parameterisations.

Even with individual differences allowed for in the shape of groups of similarly performing participants, the most compelling finding in the hierarchical classification experiment discussed, was that of higher classification accuracy at the lower level of the hierarchically organised category structure. It was observed among 15 out of a total of 18 participants, constituting three of five identified subgroups. Despite speculations to the contrary, this behavioural pattern was adequately modelled by a formal instantiation of exemplar theory. With exemplar theory shown to be adequate in dealing with classification at a single level of abstraction as well as classification at multiple levels of generality, it is now up to proponents of rival theories to establish their models' validity in regard to both these phenomena as well.

References

- Ashby, F.G., & Lee, W.W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General*, 120, 150-172.
- Ashby, F.G., Maddox, W.T., & Lee, W.W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychological Science*, 5, 144-151.
- Cohen, A.L., & Nosofsky, R.M. (2003). An extension of the exemplar-based random-walk model to separable-dimension stimuli. *Journal of Mathematical Psychology*, 47, 150-165.
- Estes, W.K. (1956). The problem of inference from curves based on group data. *Psychological Bulletin*, 53, 134-140.
- Estes, W.K. (1993). Models of categorization and category learning. *Psychology of Learning & Motivation*, 29, 15-56.

- Hampton, J.A. (1979). Polymorphous concepts in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 18, 441-461.
- Hampton, J.A. (1993). Prototype models of concept representations. In I. Van Mechelen, J.A. Hampton, R.S. Michalski, & P. Theuns (Eds.), *Categories and concepts: Theoretical views and inductive data analysis* (pp. 67-95). London: Academic Press.
- Hoffman, J., & Ziessler, C. (1983). Objektidentifikation in künstlichen Begriffshierarchien [Object identification in artificial concept hierarchies]. *Zeitschrift für Psychologie*, 16, 243-275.
- Kruschke, J.K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99, 22-44.
- Lassaline, M.E., Wisniewski, E.J., & Medin, D.L. (1992). Basic levels in artificial and natural categories: Are all basic levels created equal? In B. Burns (Ed.), *Percepts, concepts, and categories: The representation and processing of information* (pp. 327-378). Amsterdam: North-Holland.
- Lee, M.D., & Navarro, D.J. (2002). Extending the ALCOVE model of category learning to featural stimulus domains. *Psychonomic Bulletin & Review*, 9, 43-58.
- Lee, M.D., & Webb, M.R. (2005). Modeling individual differences in cognition. *Psychonomic Bulletin & Review*, 12, 605-621.
- Martin, R.C., & Caramazza, A. (1980). Classification in well-defined and ill-defined categories: Evidence for common processing strategies. *Journal of Experimental Psychology: General*, 109, 320-353.
- McKinley, S.C., & Nosofsky, R.M. (1996). Selective attention and the formation of linear decision boundaries. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 294-317.
- Medin, D.L., & Schaffer, M.M. (1978). Context theory of classification learning. *Psychological Review*, 85, 207-238.
- Murphy, G.L. (1991). Parts in object concepts: Experiments with artificial categories. *Memory & Cognition*, 19, 423-438.
- Murphy, G.L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.
- Murphy, G.L., & Smith, E.E. (1982). Basic-level superiority in picture categorization. *Journal of Verbal Learning and Verbal Behavior*, 21, 1-20.
- Myung, I.J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology*, 44, 190-204.
- Myung, I.J., & Pitt, M.A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review*, 4, 79-95.
- Nosofsky, R.M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 104-114.
- Nosofsky, R.M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, 115, 39-57.
- Palmeri, T.J. (1999). Learning categories at different hierarchical levels: A comparison of category learning models. *Psychonomic Bulletin & Review*, 6, 495-503.
- Palmeri, T.J., & Nosofsky, R.M. (1995). Recognition memory for exceptions to the category rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 548-568.
- Rogers, T.T., & Patterson, K. (2007). Object categorization: Reversals and explana-

- tions of the basic-level advantage. *Journal of Experimental Psychology: General*, 136, 451-469.
- Rosch, E. (1978). Principles of categorization. In E. Rosch & B.B. Lloyd (Eds.), *Cognition and categorization* (pp. 27-48). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rosch, E., & Mervis, C.B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573-605.
- Rosch, E., Mervis, C.B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382-439.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, 6, 461-464.
- Smith, J.D., & Minda, J.P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24, 1411-1436.
- Smith, J.D., Murray, M.J., & Minda, J.P. (1997). Straight talk about linear separability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23, 659-680.
- Vanpaemel, W., & Navarro, D.J. (2007). Representational shifts during category learning. In D. McNamara & G. Trafton (Eds.), *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 1599-1604). Mahwah, NJ: Erlbaum.
- Verheyen, S., Ameel, E., Rogers, T.T., & Storms, G. (2008). Learning a hierarchical organization of categories. In B.C. Love, K. McRae, & V.M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 751-756). Austin, TX: Cognitive Science Society.
- Webb, M.R., & Lee, M.D. (2004). Modeling individual differences in category learning. In K. Forbus, D. Gentner, & T. Regier, (Eds.), *Proceedings of the 26th Annual Conference of the Cognitive Science Society* (pp. 1440-1445). Mahwah, NJ: Erlbaum.

Received October 1, 2007

Revision received April 30, 2008

Accepted June 11, 2008