Exemplar models can't see the forest for the trees

Nolan Conaway and Kenneth J. Kurtz
Department of Psychology, Binghamton University

We investigated human learning and generalization of three novel category structures based on eight exemplars in a continuous (9x9) stimulus space. Each category requires attention to both dimensions, but they differ in their organization. Critically, all three category types are matched on within- and between-category exemplar distances. The first category structure conforms to a condensation or information-integration type of problem with two classes separable by a diagonal bound. The other category structures cannot be solved with a linear decision boundary. We found that learners trained on the diagonal bound structure showed significantly better learning and generalization performance. In computational simulations, we found that an exemplar model (ALCOVE) could not account for the observed pattern. We posit that ALCOVE is constrained by the matched distances to learn these category structures at the same speed. Another similarity-based model with different basic design principles (DIVA) provided a good account of the behavioral data.

Introduction

Categorization in exemplar models is based on similarity (attention-weighted distance) between exemplars. Learning is faster to the extent that category membership is the same for similar exemplars (Kruschke, 1992; Nosofsky, 1984).

We provide a unique test of exemplar models using a novel set of category structures.

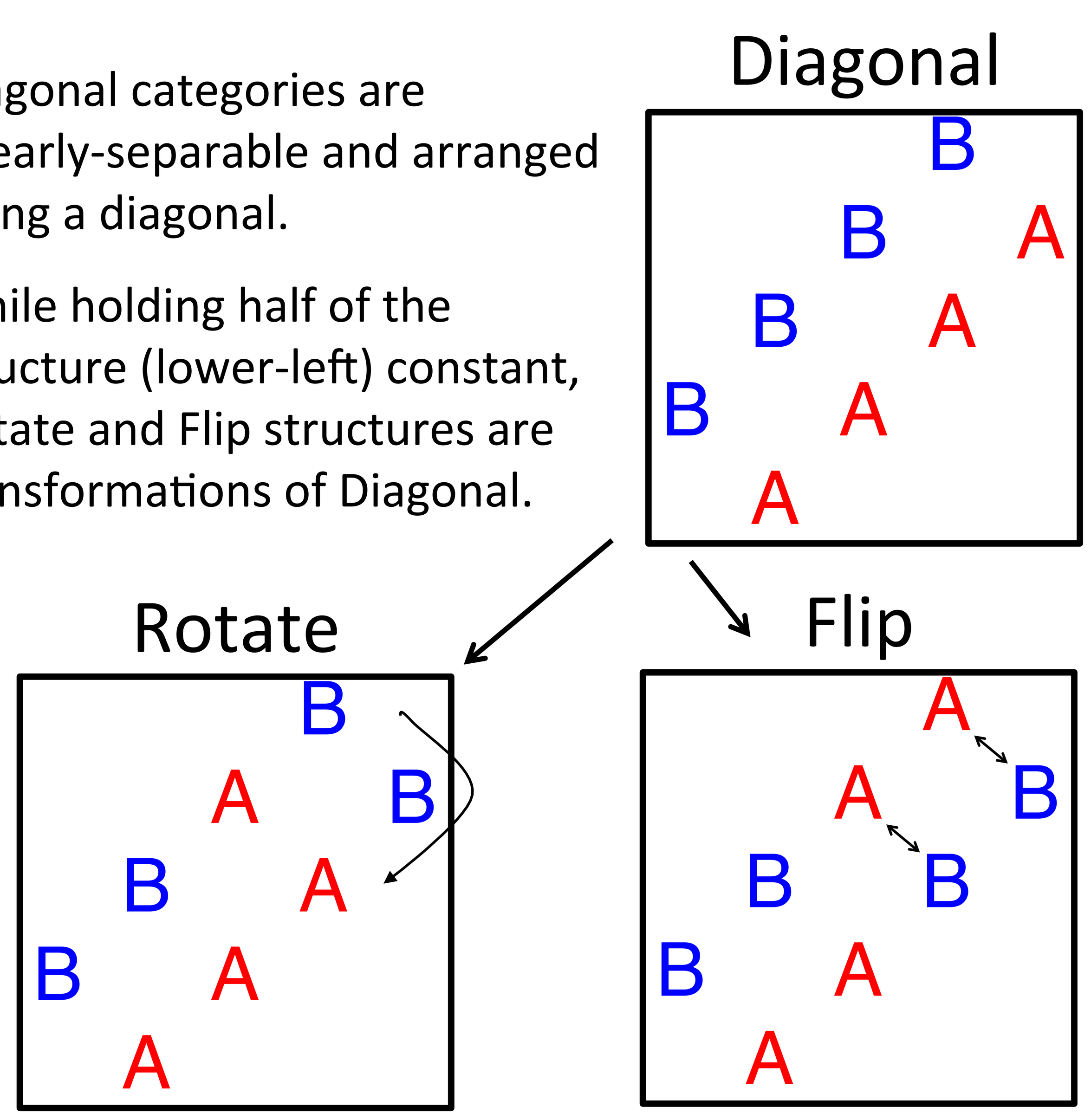
Our approach

Empirical tests of models typically involve a fixed set of stimuli presented to learners under different category assignments (e.g., Nosofsky et al., 1994; Shepard et al., 1961).

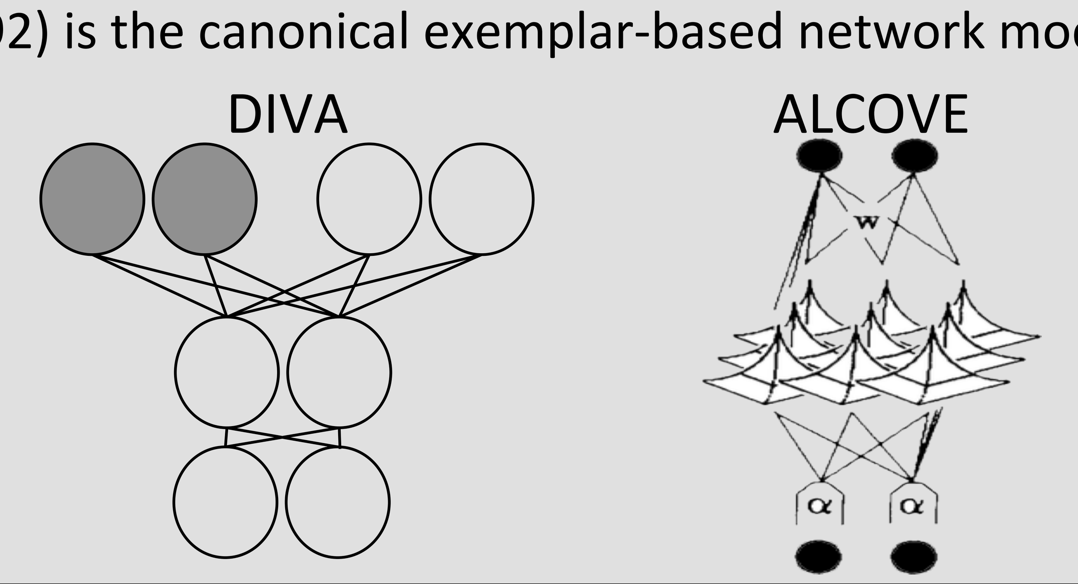
Models are compared for their ability to explain the relative learning difficulty of the resulting category types.

Diagonal categories are linearly-separable and arranged along a diagonal.

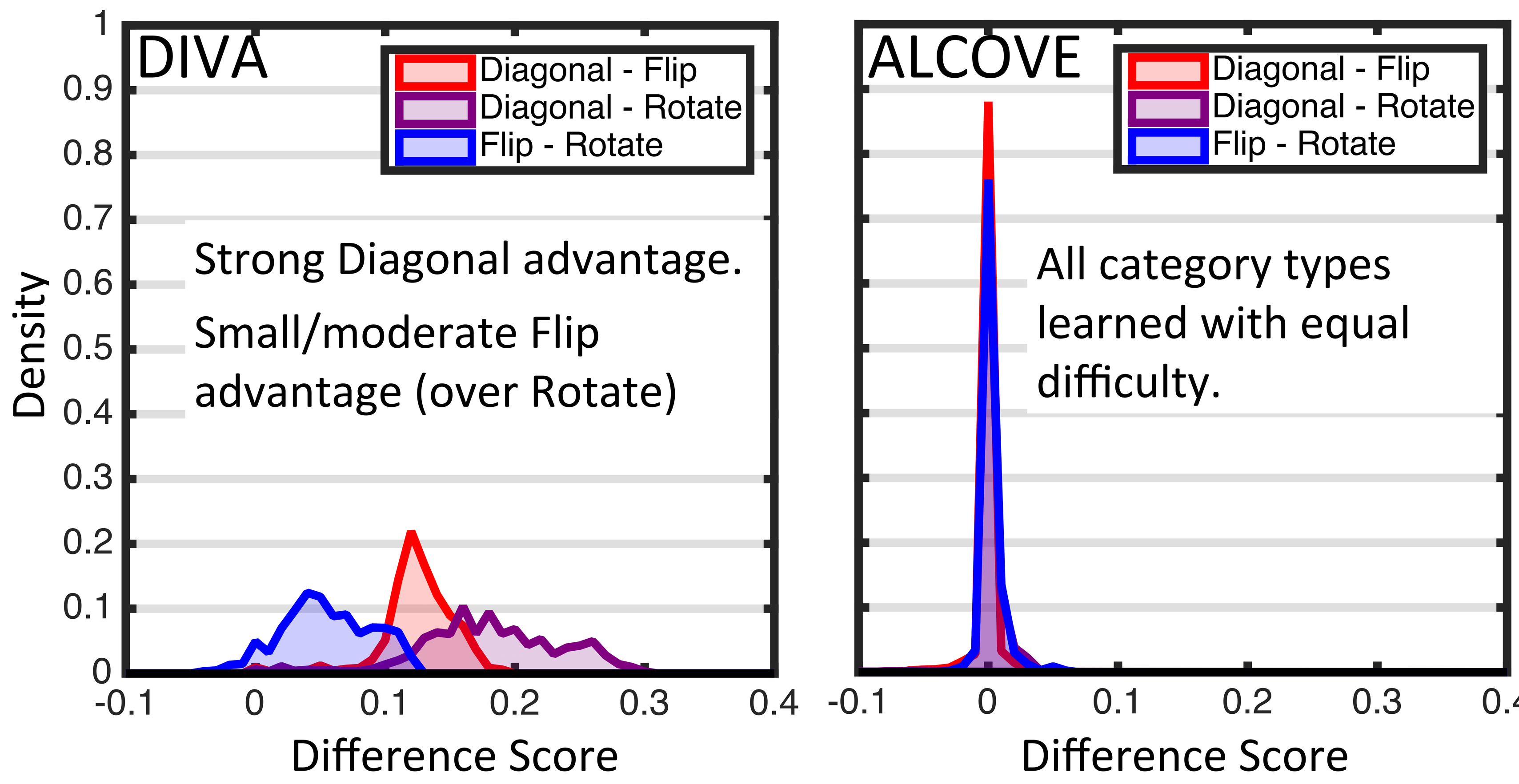
While holding half of the structure (lower-left) constant, Rotate and Flip structures are transformations of Diagonal.



DIVA (Kurtz, 2007) is a divergent autoencoder that models category learning in terms of autoassociative, error-driven learning of internal representations. ALCOVE (Kruschke, 1992) is the canonical exemplar-based network model.



What do the models predict? A priori simulations using DIVA and ALCOVE



Goal: Identify model predictions independent of behavioral data.

- 'Grid search' over model parameters.
- Compute average difference in training accuracy for each pair of category types over many parameterizations.
- Distribution of difference scores reveals each model's predictions:

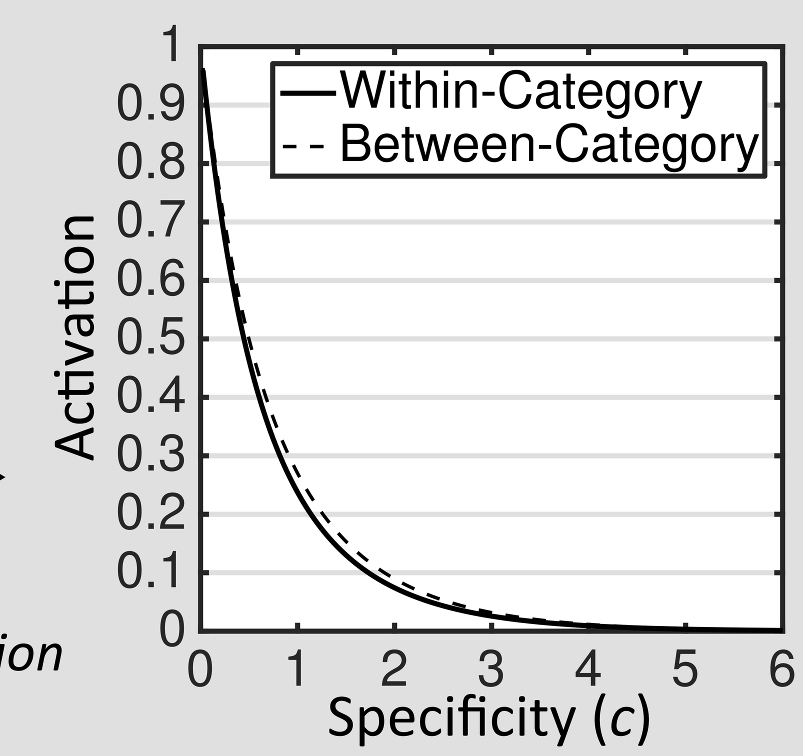
DIVA	*Diagonal > Flip >= Rotate*
ALCOVE	*Diagonal = Flip = Rotate*

Why can't ALCOVE see the forest for the trees?

We calculated the exemplar node activation upon presentation of all 8 exemplars and averaged the activations for the same- and opposite-category exemplars.

Each of the three category types produced the *exact* same profile. →

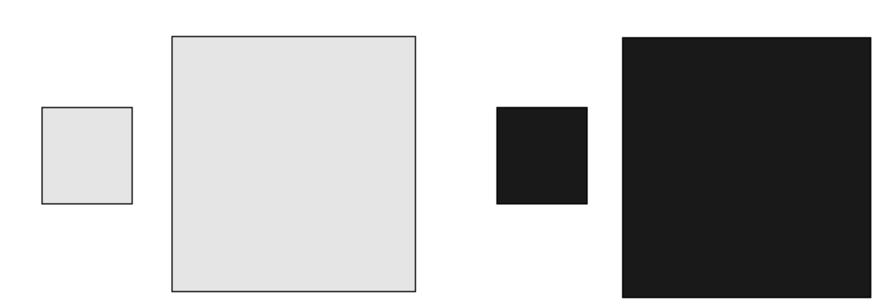
ALCOVE is sensitive to inter-exemplar distances (trees), but not the organization of each category type (forest).

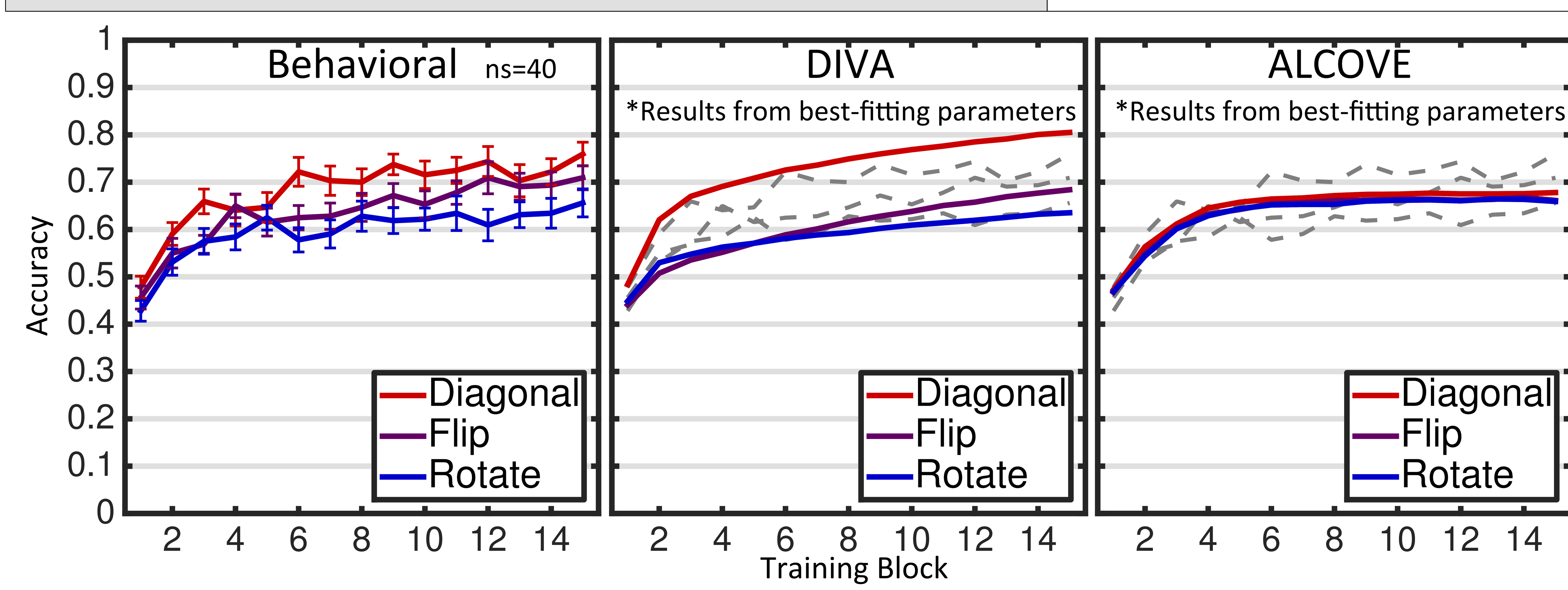


Experiment: *Do people learn the Diagonal structure more easily than Flip and Rotate?*

– Classification with feedback
– 15 training blocks (120 trials)

Sample stimuli:





Diagonal was learned the most quickly, and Flip was learned marginally faster than Rotate ($p = 0.053$). Behavioral data match DIVA's predictions, while ALCOVE fails to capture the observed learning order.

What are participants learning?

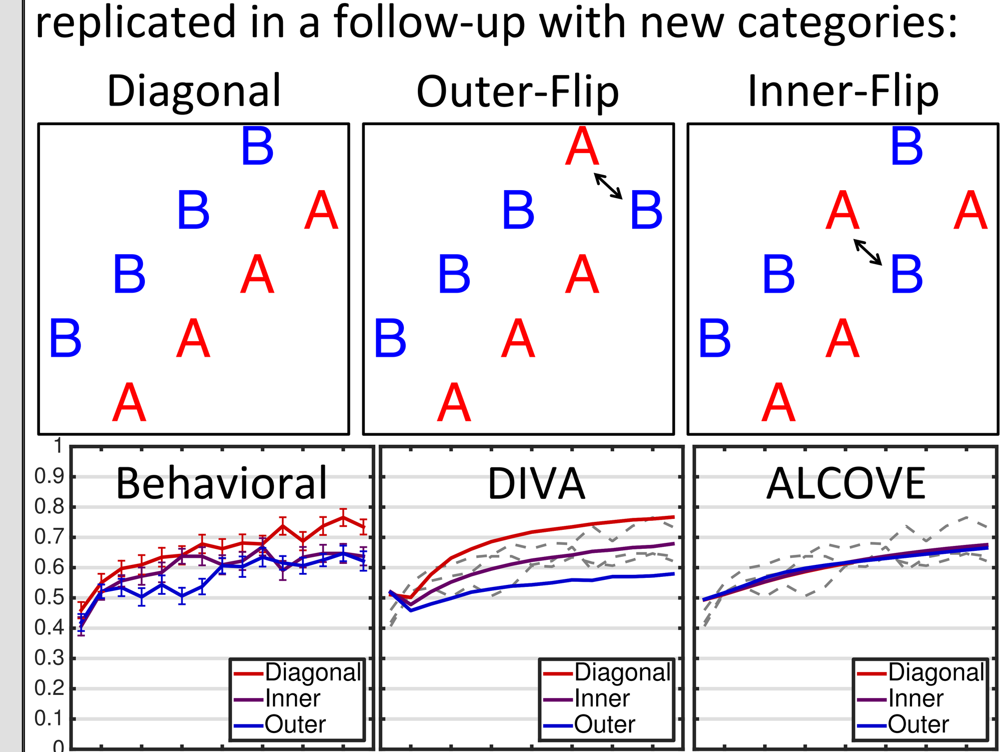
– Diagonal categories contain internal regularities (linear separability, central tendency) not shared with Flip and Rotate.
– Learners likely use these regularities to speed acquisition.

Implications for classification models

– Strong evidence against exemplar account when selective attention is not a factor.
– DIVA has abstractive capability and shows sensitivity to overall structure of categories.
– Other models that use abstraction (clusters, prototypes) may also succeed where pure exemplar accounts fail.

Replication and Extension

Results of the behavioral experiment were fully replicated in a follow-up with new categories:



Learners acquired Diagonal most quickly. ALCOVE unable to explain the observed order.

References

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. Psychological Review, 99(1), 22-44 .
Kurtz, K. J. (2007). The divergent autoencoder (DIVA) model of category learning. Psychonomic Bulletin & Review, 14, 560 –576
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., Gluck, M. A., Palmeri, T. J., & McKinley, S. C. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). Memory & Cognition, 22(3), 352-369.
Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. Psychological Monographs: General and Applied, 75, 1–42.