Flexible Attention Weighting in Human Category Learning

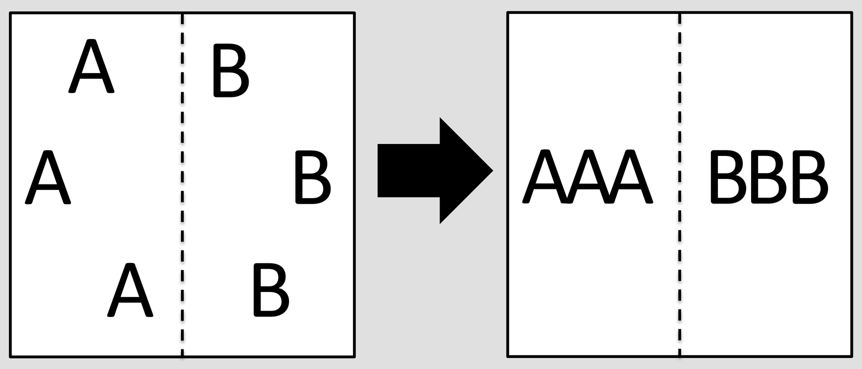
Nolan Conaway and Kenneth J. Kurtz
Department of Psychology, Binghamton University

Selective attention is an integral component of exemplar models. In conjunction with stimulus generalization theory, this design principle allows models to account for a wide variety of empirical phenomena. In exemplar models, selective attention is implemented as a set of weights that are used to differentiate the degree of contribution of each stimulus dimension in the computation of similarity between a target item and stored exemplars. A number of different reports suggest that human learners may not be limited to a single set of attention weights – instead, people appear to be able to apply different attentional sets to different areas of the stimulus space. This phenomenon can be explained using exemplar models if they are augmented with exemplar-specific attention. However, the psychological plausibility of exemplar-specific attention is questionable and has not been independently tested. Using the Divergent Autoencoder (DIVA) model, we propose a dynamic focusing mechanism that affords attentional flexibility without invoking exemplar-specific attention. We find that focusing enables DIVA to explain a wide range of empirical phenomena including the evidence suggesting a need for exemplar-specific attention.

Attention in exemplar models

Attention weights affect the calculation of similarity between exemplars and the presented cue.

Features relevant to a classification are heavily weighted. Irrelevant features are ignored.

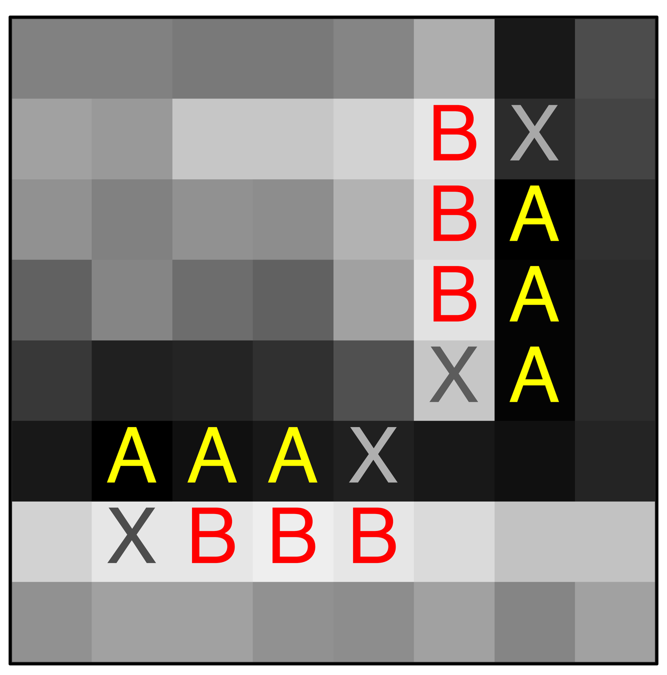


(Kruschke, 1992; Nosofsky, 1984)

Evidence for flexible feature-weighing

Aha & Goldstone (1992)

- Categories made out of two different simple rules
- Critical test items (X) are equally similar to exemplars from both categories, but classifiable by local rules
- Key Finding: human learners tend to generalize according to the local rules



Blair et al. (2009, Exp 2)

- 3D binary stimulus set divided into nested categories (4 classes)
- Eyetracking revealed different patterns of attention to features depending on whether the exemplar was A* or B*
- Learners only look at features that are relevant to the subclass

	F_1	F_2	F_3
A_1	0	0	0
	0	0	1
A_2	0	1	0
	0	1	1
B_1	1	0	0
	1	1	0
B_2	1	0	1
	1	1	1

Key point: The flexibility of feature weighing is not captured by feature-based selective attention. The best existing account of these phenomena is exemplar-specific attention, instead of attention specific to each feature (Aha & Goldstone, 1992).

An alternative to exemplar-specific attention

DIVA (Kurtz, 2007, 2015) is a *Div*ergent *A*utoencoder that models category learning in terms of autoassociative, error-driven learning of internal representations.

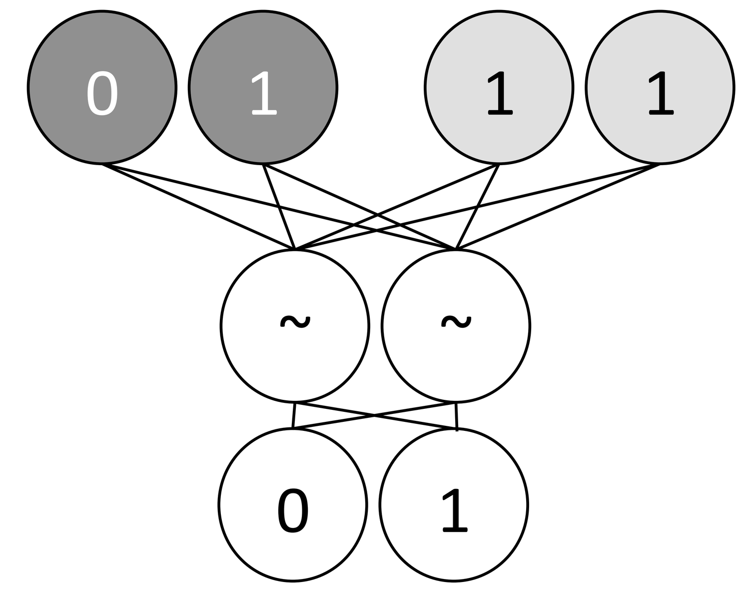
DIVA's focusing mechanism dynamically weighs features (i) based on the diversity (D) across categories (A, B)

$$D_i = |A_i - B_i|$$

$$W_i = \frac{e^{\beta D_i}}{\sum_i e^{\beta D_i}}$$

Dimensions that differ more across the category channel reconstructions are more heavily weighted in response rule.

A focusing parameter (β) controls the degree of focusing.



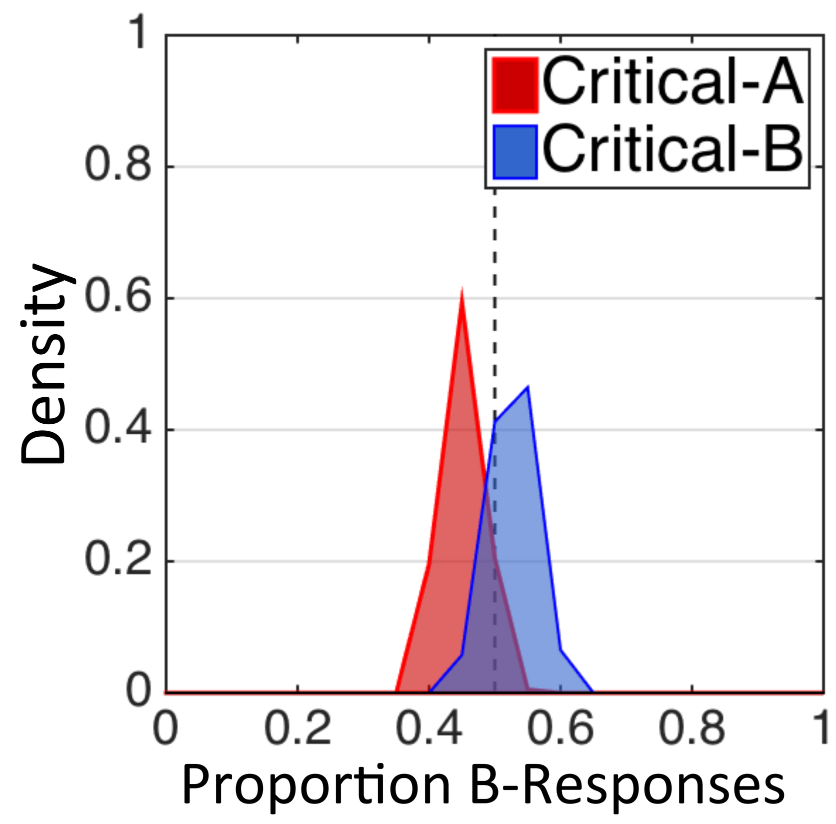
Where β=2

$D_1 = |0-1| = 1$

$D_2 = |1-1| = 0$

$W_1 = e^{2*1} / (e^{2*1} + e^{2*0}) = \mathbf{0.88}$

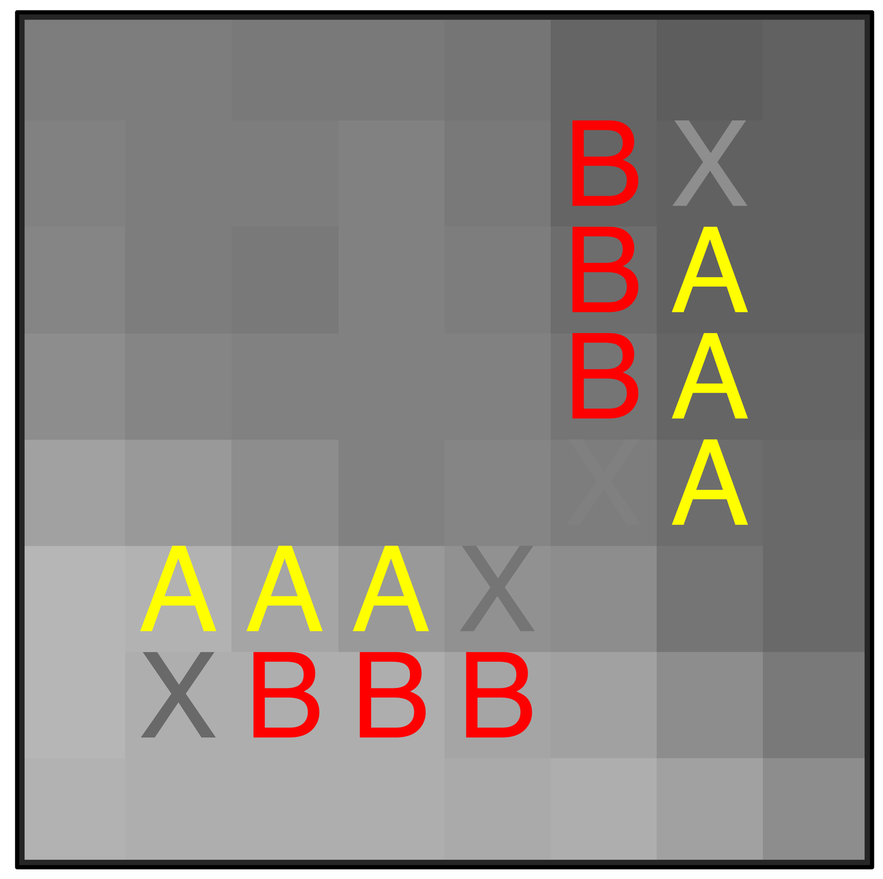
DIVA's account of Aha and Goldstone (1992)



Grid-search over large number of DIVA parameterizations reveals:

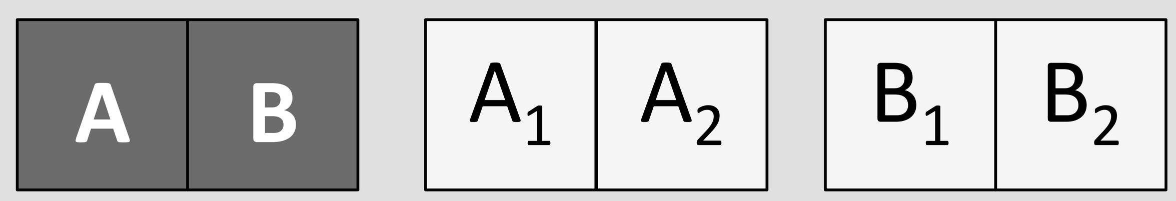
Fig. (right): Sample weights for Y-dim (note: darker = less attention). Greater focusing on Y-dim for items in the region (lower left) where Y-dim is diagnostic.

Fig. (left): Histogram of responses to critical items. *DIVA systematically generalizes according to local unidimensional boundaries.*



DIVA's account of Blair et al. (2009)

nested classification task → *six* output channels

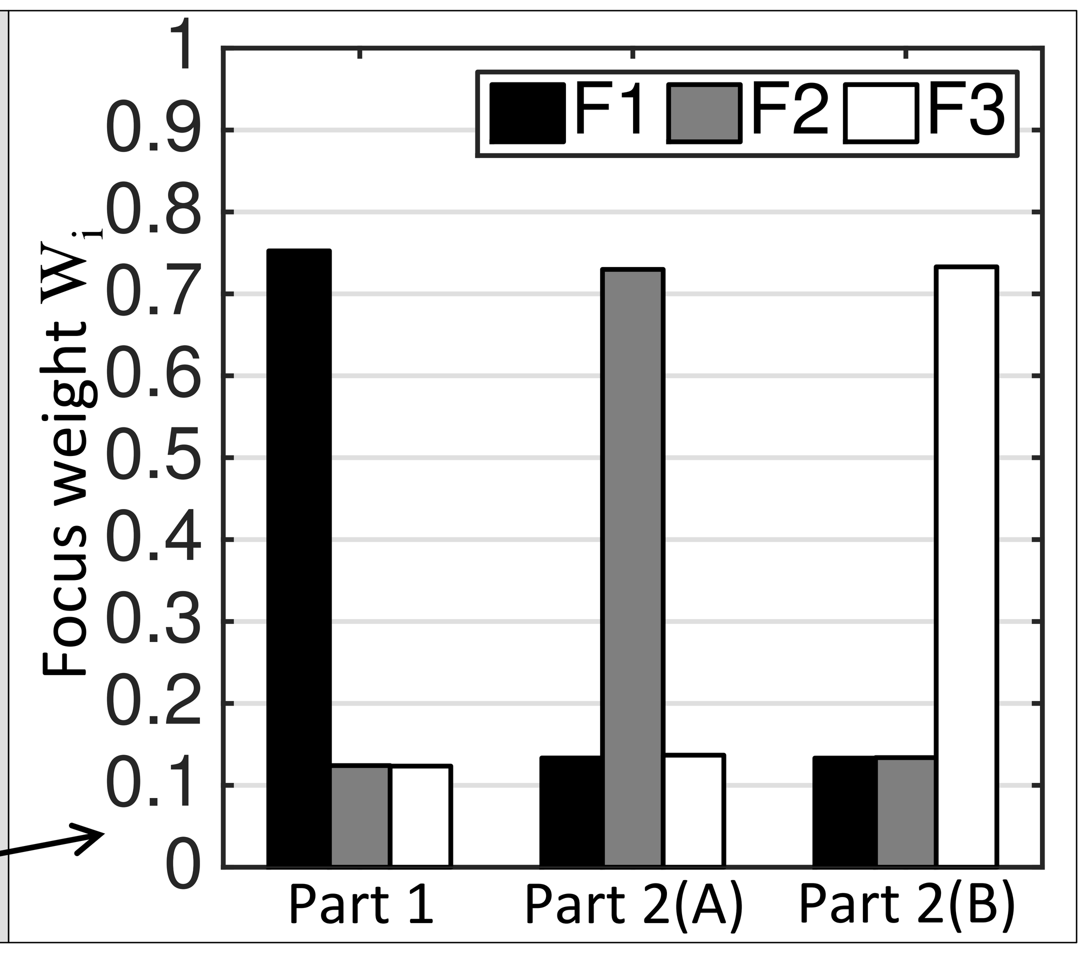


Classification is a two-part process:

1. Choose between A and B categories (A_1/A_2 vs. B_1/B_2)
2. Choose between subcategories (A_1 vs. A_2 or B_1 vs. B_2)

Network is trained on correct global and subclass channels.

DIVA captures the observed attentional patterns



DIVA explains qualitative patterns found by Aha & Goldstone (1992):

- Critical test items classified in accord with local unidimensional boundary.
- Focus weights systematically reflect diagnostic value of local boundary.

Focus weights mirror eyetracking findings from Blair et al. (2009)

- Nested process allows us to simulate feature weighting over time.
- Features are weighted only if they are relevant for a given sub-classification.

Discussion

- We report a focusing mechanism that can explain the evidence suggesting a need for flexible feature weighting.
- While our simulations used DIVA, focusing may also be applicable in exemplar models. Future work will address this possibility.

Other evidence?

- Exemplar-specific attention can also explain generalization of rule + exception problems (Rodrigues & Murre, 2007), and memory for exception items (Sakamoto et al., 2004)
- Future work will address whether focusing enables DIVA to explain these phenomena.

References

Aha, D. W., & Goldstone, R. L. (1992). Concept learning and flexible weighting. Proc Cog Sci

Blair et al., (2009). Extremely Selective Attention: Eye-Tracking Studies of Dynamic Attention Allocation to Stimulus Features in Categorization. JEPLMC

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. PR.

Kruschke, J. K. (1993). Human category learning: Implications for backpropagation models. Conn Sci.

Kurtz, K. J. (2007). The divergent autoencoder (DIVA) model of category learning. PB&R.

Kurtz, K. J. (2015). Human Category Learning: Toward a Broader Explanatory Account. PL&M.

Nosofsky, R. (1984). Choice, similarity, and the context theory of classification. JEPLMC

Rodrigues, P. & Murre, J. (2007). Rules-plus-exception tasks: A problem for exemplar models? PB&R.

Sakamoto, Y., Matsuka, T., & Love, B. (2004). Dimension-wide vs. exemplar-specific attention in category learning and recognition. 6th International Conference on Cognitive Modeling.

Special thanks to Sarah Laszlo