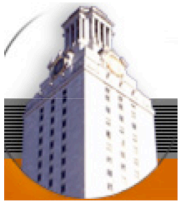


THE UNIVERSITY OF TEXAS AT AUSTIN

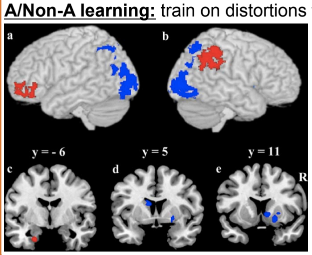
Modeling critical components of prototype learning: modulation of brain activity isolates regions associated with discriminability and strategy shifting

Marissa A. Gorlick, Brian D. Glass, Jeanette Mumford, Dagmar Zeithamova, W. Todd Maddox & David M. Schnyer; University of Texas at Austin

Introduction

Zeithamova and colleagues (2008)[1] identified neurally dissociable A/B and A/Non-A prototype learning systems:

A/B learning: train on distortions from two prototypes
A/Non-A learning: train on distortions from one prototype.



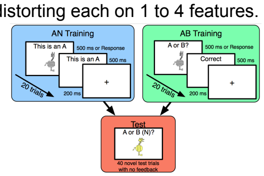
A/B Declarative Memory (red):
- (R) Inferior parietal
- (L) Orbitofrontal
- (L) Parahippocampal

A/Non-A Nondeclarative (blue):
- Inferior lateral occipital
- Superior parietal
- (R) Putamen
- Caudate

24 adults (age 18-30; 13 F) completed 2 runs of the A/B and A/Non-A task.

Prototype Learning Task

Stimuli: Ten binary features create 1024 exemplars. One exemplar is the A prototype and the B prototype has the opposite feature values. Categories are made by distorting each on 1 to 4 features.



Training (not scanned):
A/B: Categorize10 A and10 B stimuli with corrective feedback.
A/Non-A: Categorize 20 A stimuli passively.

Test (scanned): Identical for A/B and A/Non-A. On each of 42 trials, a stimulus was presented and the participant indicated the category membership of the stimulus. No feedback was provided.

Goals of Current Project

A/B and A/Non-A prototype models were applied to Zeithamova et al test data and were correlated with brain activation to examine:

1) how improved model fit correlates with activation in the task appropriate system. We predict some overlap with the active regions seen in Zeithamova.
2) how model parameters correlate with brain activation. We predict that discriminability represents different strategy selection systems for the A/B and A/Non-A task.

Computational Model

42 test trials were fit with a 2-proto model for the A/B data and a 1-proto model for the A/Non-A data using max likelihood[2]. Both models assumed attention-weighed distance between the current stimulus x_i and the A prototype is:

$$d_{xPA} = [\Sigma w_i (x_i - P_{Ai})^2]^{\frac{1}{2}}$$

where w_i represents the attention-weight of dimension i and the w_i sum to 1. The binary value of a dimension i is denoted by x_i and the value of the A prototype is P_{Ai}. The similarity between stimulus x_i, and the A prototype is:

$$\eta_{iA} = e^{-cd_{xPA}}$$

where c is the perceptual discriminability (c>0).

A/B model	**A/Non-A model**
The probability of responding A for each stimulus x is given by: $P(A\|x) = \frac{\eta_{iA}\beta_A}{\eta_{iA}\beta_A + \eta_{iB}\beta_B}$ Where 0<β<1 is the bias to respond A.	The probability of responding A for each stimulus x is given by: $P(A\|x) = \frac{\eta_{iA}}{\eta_{iA} + \beta}$ where β>0 is a bias to respond not-A.

- Maximum Likelihood Fit reflects the extent to which the participant was using the task appropriate strategy.
- Discriminability reflects how well the subject discriminates between exemplars.

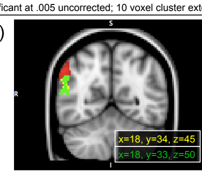
Maximum likelihood fit and Discriminability were entered into the model as regressors at the run level when comparing all trials to baseline. Fit and Discriminability predict accuracy in A/B and A/Non-A tasks.

Two-Prototype

Significant at .005 uncorrected; 10 voxel cluster extent

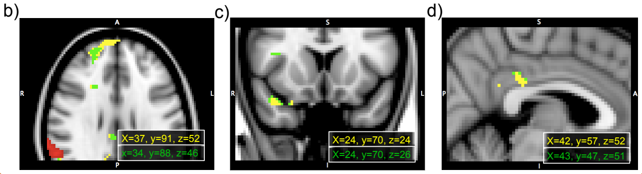
The **Fit (yellow)** and the **Discriminability parameter (green)** both correlate with a) right inferior parietal activity as seen by Zeithamova in the A/B task, however they did not correlate with orbitofrontal or parahippocampal activity.

a)



In addition to those regions identified in Zeithamova, Fit and Discriminability both correlated with b) right dorsolateral prefrontal cortex, c) right temporal pole, and d) posterior cingulate activation.

b) c) d)

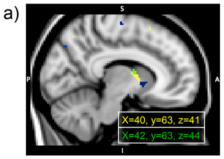


One-Prototype

Significant at .005 uncorrected; 10 voxel cluster extent

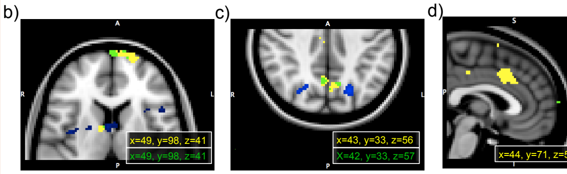
The **Fit (yellow)** and the **Discriminability parameter (green)** both correlate with a) right caudate activity as seen by Zeithamova in the A/Non-A task, however they did not correlate with occipital, putamen, left caudate, or superior parietal activation.

a)



In addition to regions identified in Zeithamova, Fit and Discriminability both correlated with b) orbitofrontal cortex and c) medial occipital activation.

b) c) d)



Fit also correlated with d) anterior cingulate activation.

Conclusions

Neural activity tracks with model parameters revealing regions not previously associated with accuracy. This may reflect new contributions compared to standard accuracy contrasts (correct > incorrect). Further studies are needed to define the mechanisms of these components.

A/B Learning

- Models suggest the parietal cortex (not active for correct trials in Zeithamova) may focus attention to perceptual features becoming more active as the category space is mapped.
- Activity in the DLPFC and anterior cingulate are consistent with attention and perceptual discriminability while testing verbal hypotheses.

A/Non-A Learning

- Models suggest the caudate (not active for correct trials in Zeithamova) indirectly affects accuracy, perhaps by processing feedback.
- Activity in the OFC is consistent with switching from the more salient episodic memory strategy to the optimal non-declarative strategy.

References

1. Zeithamova, D., Maddox, W. T., & Schnyer, D. M. (2008). Dissociable prototype learning systems: evidence from brain imaging and behavior. *The Journal of Neuroscience*, *28*(49), 13194.
2. 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*(6), 1411.

This work was supported by NIH Grant R01 MH077708 and Army Grant W911NF-07-2-0023, through The Center for Strategic and Innovative Technologies at The University of Texas at Austin. We thank Andrea Bozoki, Sean Maddox, Borami Lee, and Cristina Benavides for developing the rest of the stimulus sets. We also thank Tori Williams for valuable help with scanning.