A Neural Network Model of the Effect of Prior Experience with Regularities on Subsequent Category Learning

Casey L. Roark[1,2], David C. Plaut[1,2,3], and Lori L. Holt[1,2,3]

[1]Carnegie Mellon University Department of Psychology, [2]Neuroscience Institute, [3]Center for the Neural Basis of Cognition

Introduction

Category structures and neural mechanisms of learning

- A popular dual systems theory of category learning posits that the layout of categories in space drives neural mechanisms supporting learning.
- Rule-Based (RB) categories are thought to require selective attention to individual stimulus dimensions and Information-Integration (II) require pre-decisional integration across multiple dimensions (Ashby et al., 1998; Ashby & Maddox, 2011; Yi & Chandrasekaran, 2016).

Assumption of independence of perceptual dimensions

- Underlying psychological representations of simple visual input dimensions are well understood, but many dimensions are not as independent (Garner, 1974; Roark & Holt, 2019; Scharinger et al., 2013).
- In a simulation experiment, we investigate the effect of prior experience on subsequent learning of 'RB' and 'II' category structures to understand how psychological representations influence learning.

Neural Network Methods

Network Architecture

- Representation training: autoencoder with 20-unit sensory input, 10-unit hidden perceptual representation layer and 20-unit autoencoder output layers (Figure 1).
- Category learning: additional category decision output layer (2 units representing the two categories).

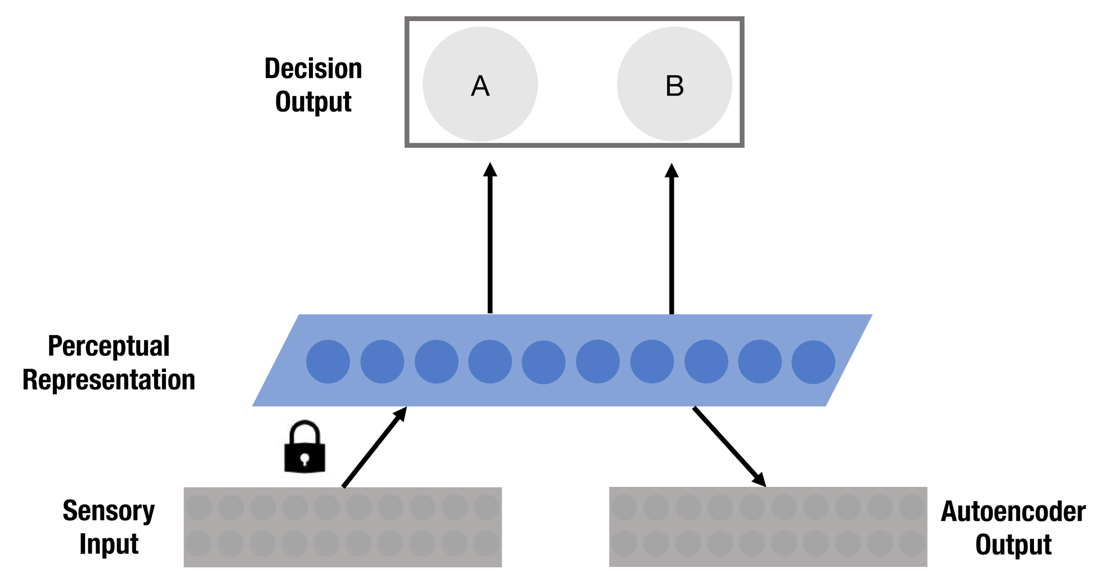

Figure 1. Layout of network architecture.

Representation Training

- Trained the network on five types of stimulus distributions representing five possible hypothetical relations between dimensions (Figure 2).

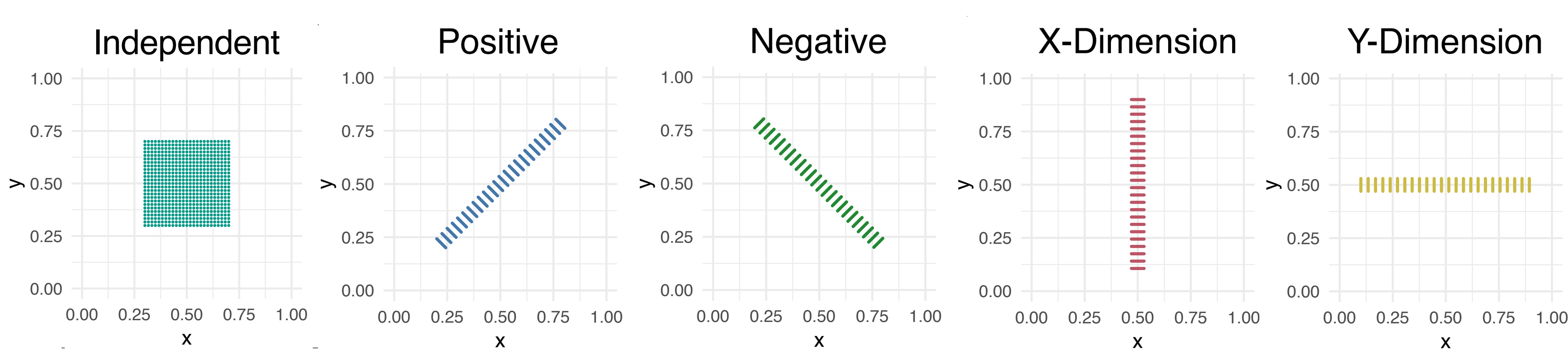

Figure 2. Stimulus distributions for the representation training phase.

Category Learning

- Trained the network with feedback—using the learned hidden perceptual representations—on four types of category structures that require different usage of the two dimensions (Figure 3):
  - Information-Integration (II) Positive & Negative
  - Rule-Based (RB) X Dimension and Y Dimension

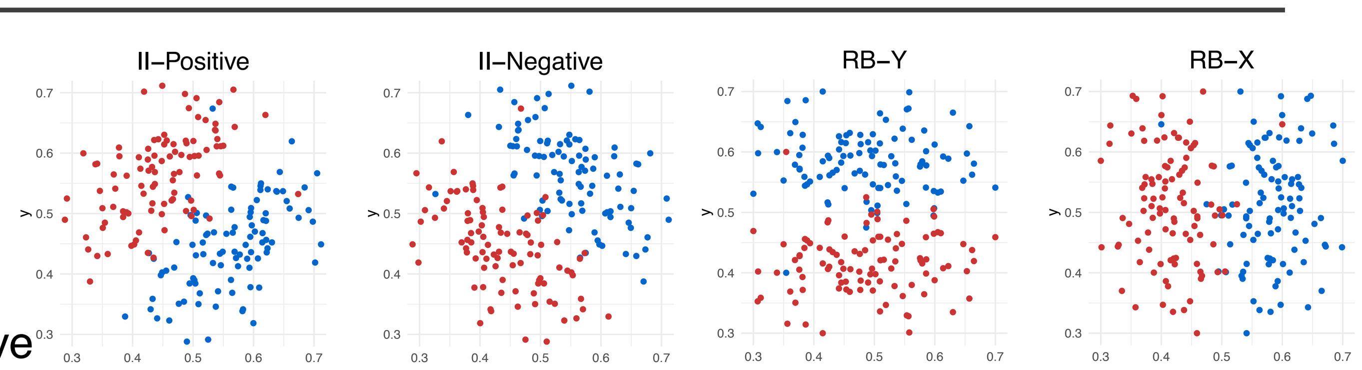

Figure 3. Category distributions for the category training phase.

Acknowledgements

This work was supported by a training grant from the National Institute General Medical Sciences to Casey L. Roark (T32GM081760). Note: CLR is now at the University of Pittsburgh, Department of Communication Science & Disorders

Simulation Results

- We assessed the impact of representation training (Independent, Positive, Negative, Y-Dim, X-Dim) on the four types of category learning structures (IIPos, IINeg, RBY, RBX). We trained and tested on all 200 stimuli from each category.
- Accuracy was the percent of category exemplars for which the model met a target activation criterion of 0.5 (Figure 4).
- **Performance of the model on these category structures greatly depended on the nature of the pre-trained representations.**
- Existing perceptual representations can impact the outcomes of category learning, especially when the physical dimensions or experimenter-defined dimensions do not align with the dimensions of representations.

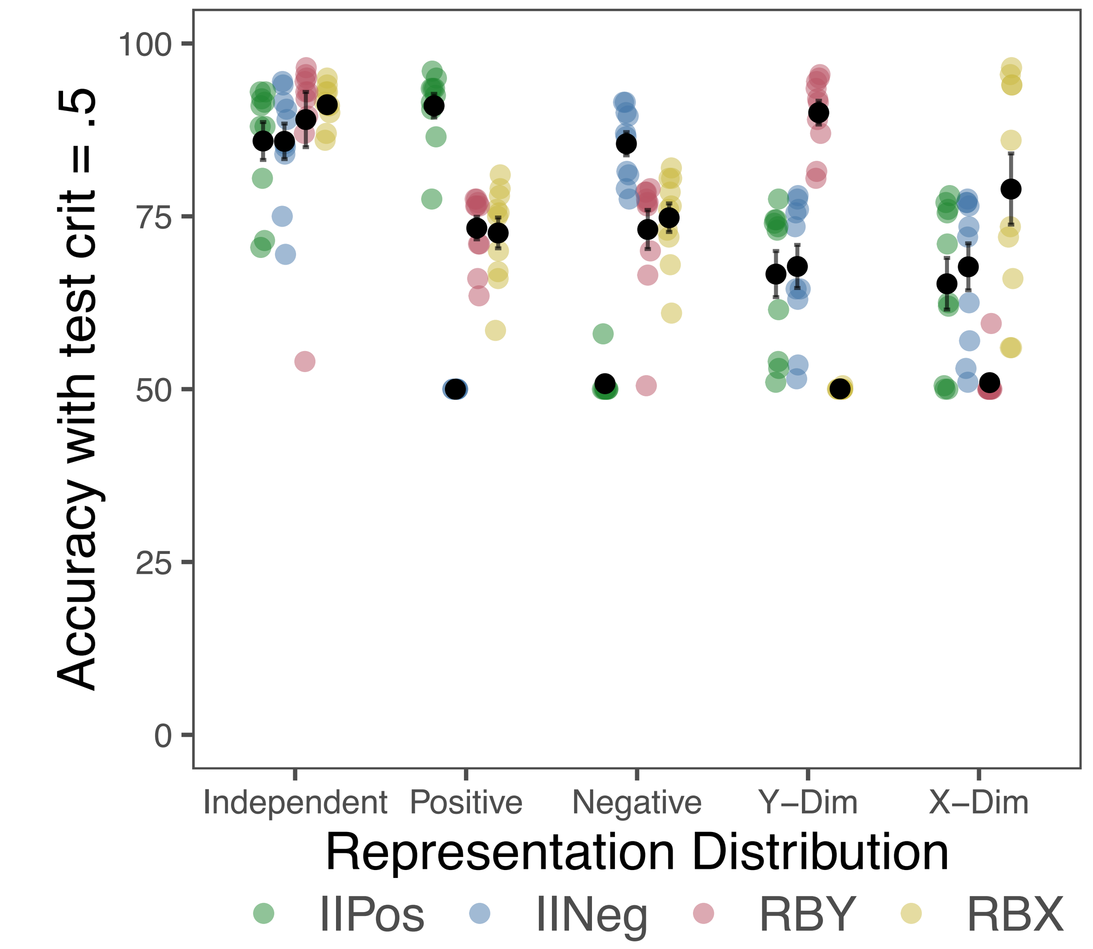

Figure 4. Model's performance accuracy in the category learning phase.

Comparison to Human Behavior

- For sake of brevity, we present a comparison of human behavior in one auditory experiment. Please see the full conference paper for other comparisons.
- Humans in Roark & Holt (2019) learned auditory categories defined on the dimensions of center frequency (Hz) and modulation frequency (Hz), which are difficult to selectively attend to.
- The pattern of the human behavioral accuracy (Figure 5) most closely aligns with the model's performance after having been pre-trained on the *Positive* training distribution.
- **These acoustic dimensions may be represented in a manner that reflects a long-term positive relationship between the dimensions, providing insight into the nature of perceptual representations.**

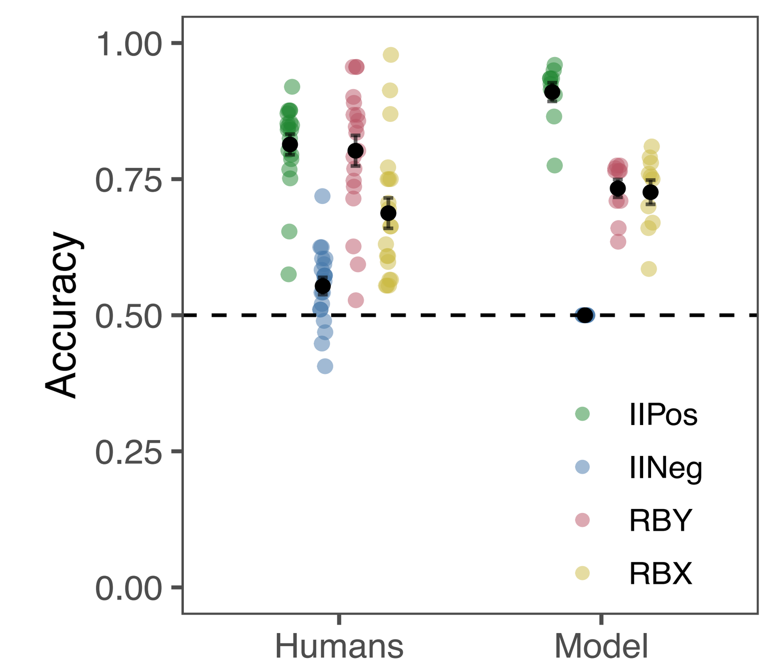

Figure 5. Human behavior from Roark & Holt (2019) and model accuracy after *Positive* training.

Conclusions

- The *perceptual* side of perceptual category learning has drifted out of focus of theories of learning.
- These simulations demonstrate that psychological representations of sensory information, shaped by experience, can strongly influence learning.

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review, 105*(3), 442–481.

Ashby, F. G., & Maddox, W. T. (2011). Human category learning 2.0. *Annals of the New York Academy of Sciences, 1224*, 147–161.

Garner, W. R. (1974). *The Processing of Information and Structure.* Hillsdale, NJ: Erlbaum.

Roark, C. L., & Holt, L. L. (2019). Perceptual dimensions influence auditory category learning. *Attention, Perception, and Psychophysics, 81*(4), 912–926.

Scharinger, M., Henry, M. J., & Obleser, J. (2013). Prior experience with negative spectral correlations promotes information integration during auditory category learning. *Memory & Cognition, 41*(5), 752–768.

Yi, H.-G., & Chandrasekaran, B. (2016). Auditory categories with separable decision boundaries are learned faster with full feedback than with minimal feedback. *The Journal of the Acoustical Society of America, 140*(2), 1332–1335.