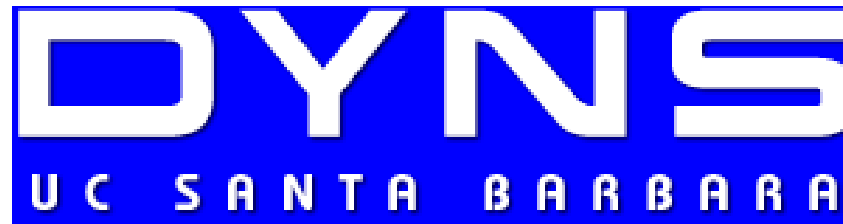


A Neural Interpretation of Exemplar Theory

Luke A. Rosedahl & F. Gregory Ashby

University of California, Santa Barbara, CA 93106



Exemplar theory has been an incredible success

Provides good fits to lots of data

- Google Scholar returns 2,480 hits
- Over 130 videos online
- Own Wikipedia page
- Has spread to many different fields
 - Many areas of cognitive psychology outside traditional categorization (e.g., recognition memory)
 - Social psychology
 - All areas of linguistics

Drawback: Lacking a Neural Interpretation

Previously has lacked a detailed neural interpretation

Limits its explanatory power for behaviors such as:

- 1: Button-Switch Interference (Ashby et al. 2003)
- 2: Feedback Delay Effects (Smith et al. 2014)
- 3: Retinal Specific Learning (Rosedahl, Eckstein, & Ashby, 2018)

Issues with Previous Neural Interpretation Attempts

Medial Temporal Lobe Explanations (e.g. Sakamoto & Love, 2004)

Predicts MTL damage should decrease performance

Not in general true (e.g. Janowsky et al. 1989; Bayley, et al. 2005; Filoteo et al. 2001)

Minimal task-related MTL activity during learning (Seger & Cincotta, 2005; Seger et al. 2010)

Overview

Exemplar theory is mathematically equivalent to a simplified version of the COVIS procedural system Ashby & Rosedahl (2017, Psychological Review)

I will then discuss using this result to predict categorization difficulty for tasks targeting the procedural system.

Exemplar Model

Nosofsky (1986, 2011)

The probability of responding A to stimulus k:

$$P(A|k) = \frac{\beta_A \sum_{i \in C_A} V_{iA} \eta_{ik}}{\beta_A \sum_{i \in C_A} V_{iA} \eta_{ik} + \beta_B \sum_{i \in C_B} V_{iB} \eta_{ik}}$$

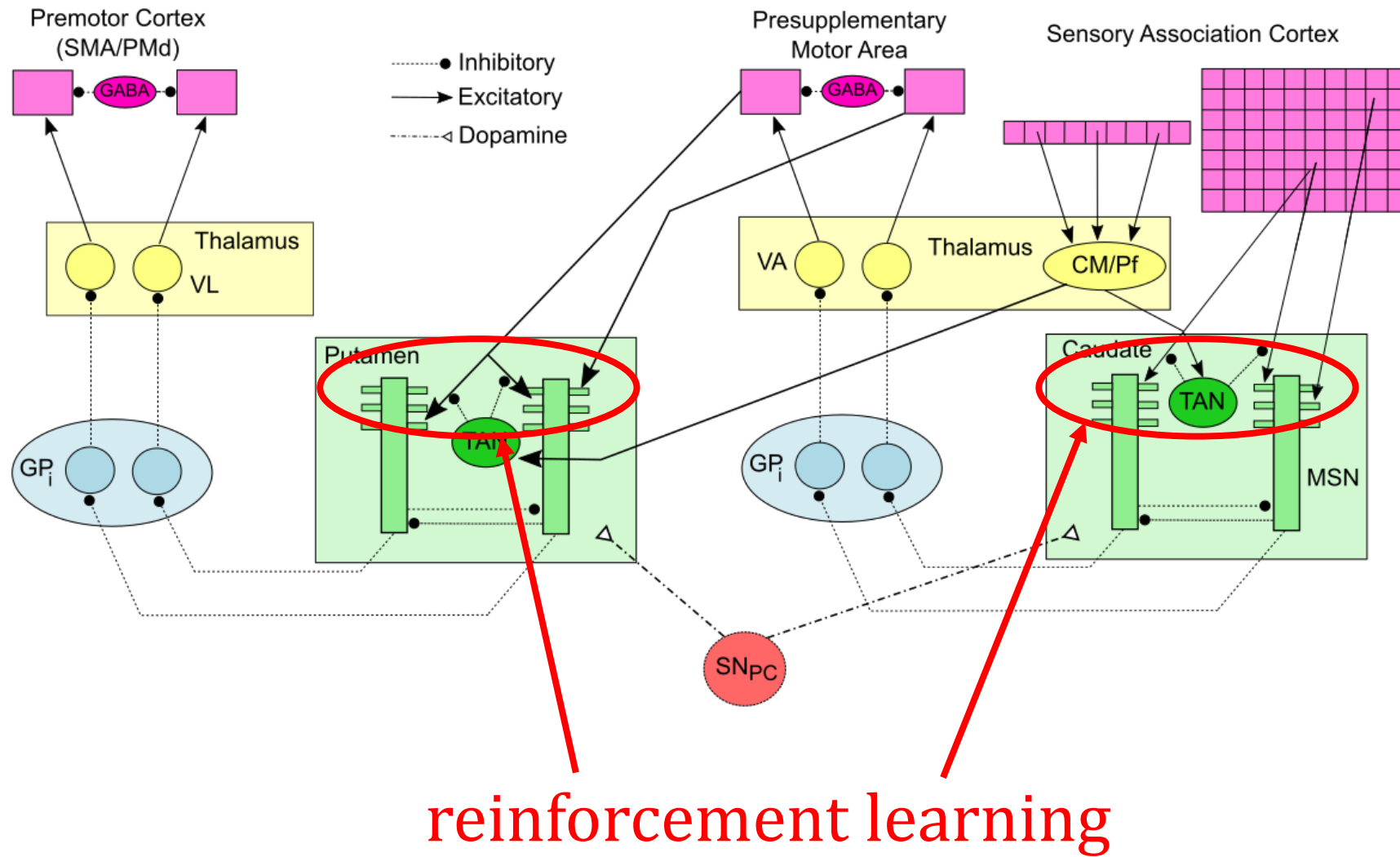
$C_A = \{\text{stimuli in category A}\}$

$\eta_{ik} = \text{similarity between stimuli } i \text{ and } k \quad \eta_{ik} = \exp(-c\delta_{ik}^\omega)$

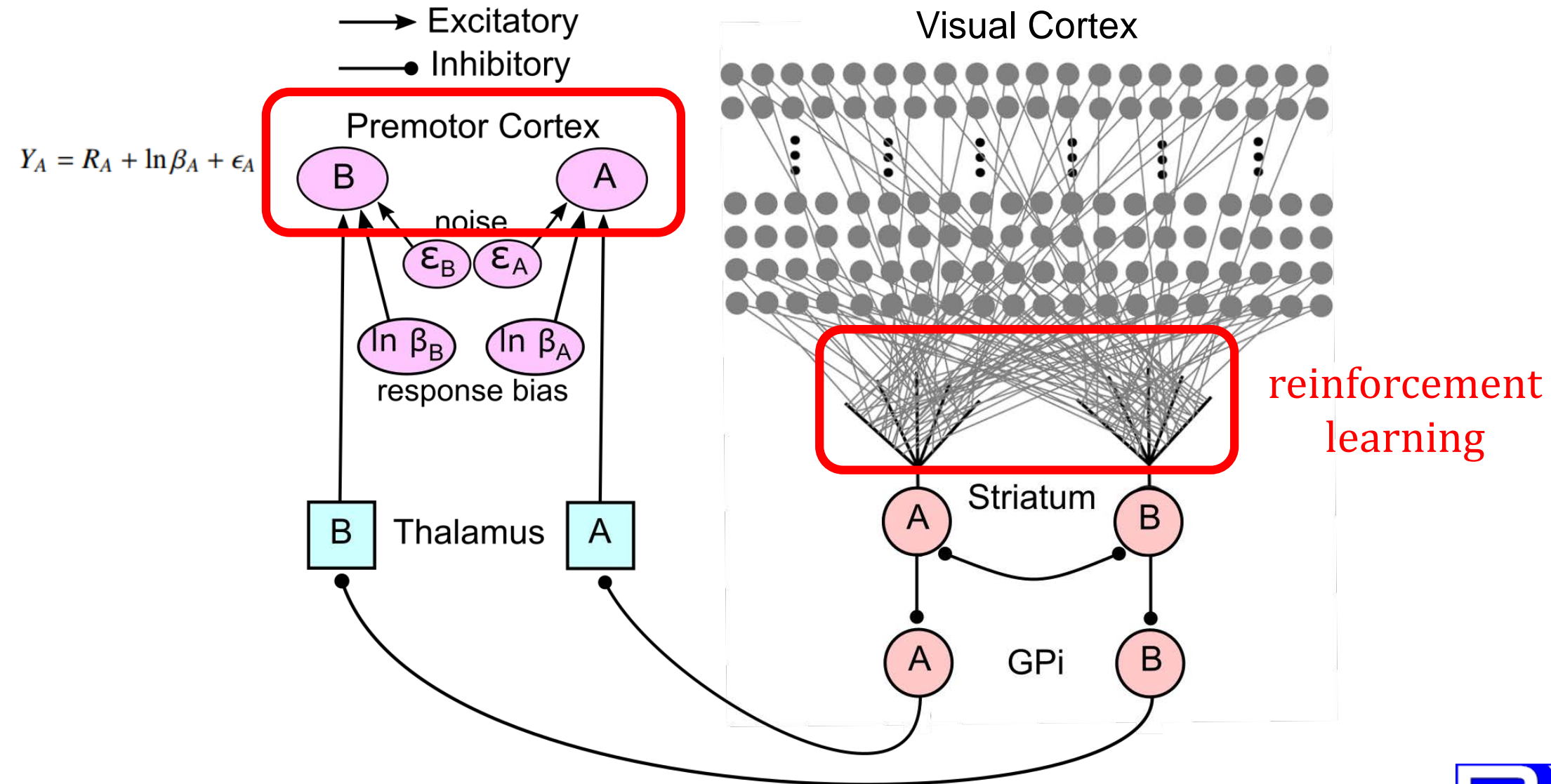
$\beta_A = \text{response bias toward A}$

$V_{iJ} = \text{the frequency that stimulus I is presented with category J feedback}$

The COVIS Procedural System



Architecture of the neural version of exemplar theory



Decision Depends on Premotor Activity

$$Y_J(n) = R_J(n) + \ln \beta_J + \epsilon_J$$

Assumption: Noise terms are independent random samples from identical double exponential distributions

$$P\left(u_k + \epsilon_k = \max_{i=1}^n \{u_i + \epsilon_i\}\right) = \frac{e^{u_k}}{\sum_{i=1}^n e^{u_i}} \quad \text{Yellott, 1977}$$

$$P(Y_J|k) = \frac{e^{Y_J}}{\sum_{i=1}^n e^{Y_i}}$$

Striatal Output

$$R_J(n) = F [A_J(n)]$$

F is a monotonically increasing function called an activation function in standard Firing Rate models.

Selecting the natural log for F gives:

$$R_J(n) = \ln [A_J(n)]$$

Striatal Activity

Striatal activity is equal to the weighed most active input

$$A_J(n) = w_{Jk}(n)I_{k|k}$$

Where I is normalized visual activity (i.e. ranges between 0 and 1)

Assuming one neuron will respond optimally for each stimulus:

$$\begin{aligned} A_J(n) &= w_{Jk}(n)I_{k|k} \\ &= w_{Jk}(n), \end{aligned}$$

Synaptic Weights

All weights start very small and are updated via reinforcement learning.

Assumption: error trials do not change synaptic strengths

if trial n was a correct category A response

$$w_{Ak}(n+1) = w_{Ak}(n) + \alpha_w A_{pre} A_{post} D^+$$

α_w = learning rate

A_{pre} = presynaptic activation

A_{post} = postsynaptic activation

D^+ = amount dopamine is above baseline

Standard
Three Factor
Learning

Reinforcement Learning

Solving for $w(n)$:

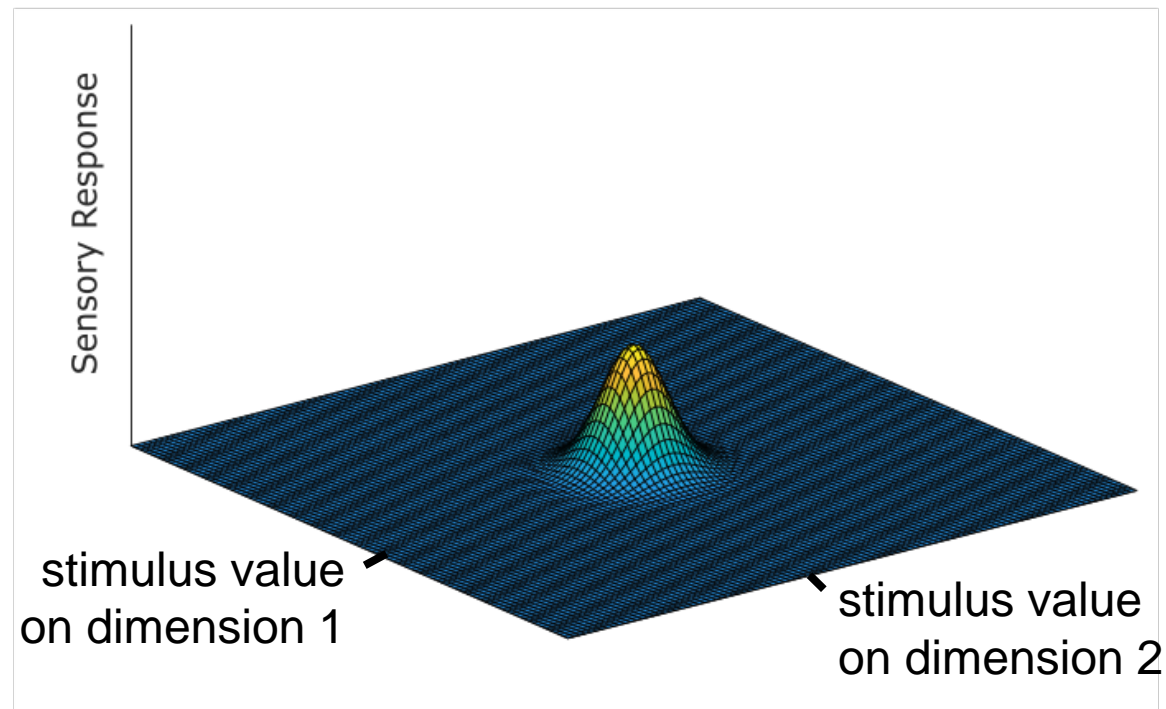
$$w_{Ak}(n + 1) = \alpha_w \sum_{i \in C_A} A_{\text{pre}}(i) A_{\text{post}}(i) D^+(i)$$

$A_{\text{post}}(i) D^+(i) = \text{a constant } K$ then

$$w_{Ak}(n + 1) = K \sum_{i \in C_A} A_{\text{pre}}(i)$$

Presynaptic activation from visual cortex determined by receptive fields

Commonly modeled as Gaussian Radial Basis
Functions (RBFs)



Modeling a Receptive Field

Consider a trial when stimulus k is presented,
then activation in visual unit i equals:

$$A_{\text{pre}}(i) = \exp(-\delta_{ik}^{\omega}/\gamma)$$

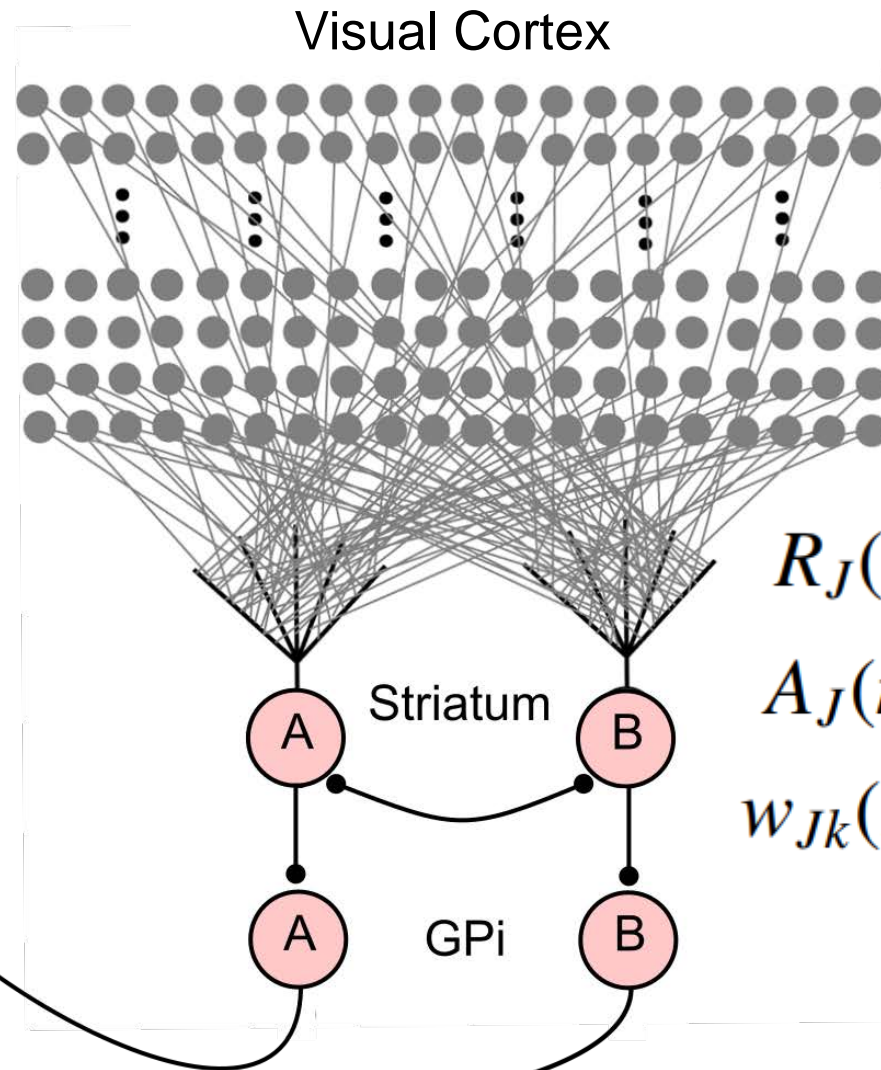
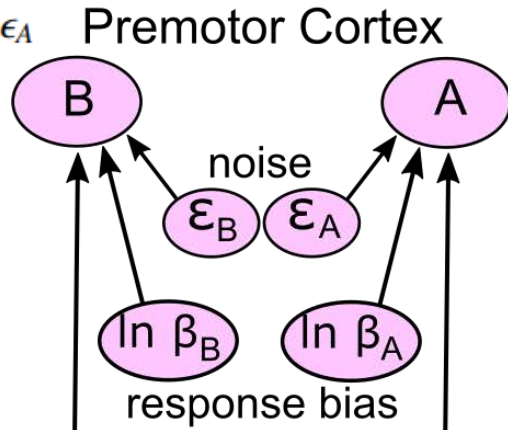
where δ_{ik} = distance in perceptual space
between the representations of stimuli i and k

$$K \sum_{j \in C_A} n_{a_j} \exp(-\delta_{kj}^{\omega}/\gamma)$$

Architecture of the neural version of exemplar theory

$$P(A|k) = \frac{e^{Y_A}}{e^{Y_A} + e^{Y_B}}$$

$$Y_A = R_A + \ln \beta_A + \epsilon_A$$



$$R_J(n) = \ln [A_J(n)]$$

$$A_J(n) = w_{Jk}(n)$$

$$w_{Jk}(n) = K \sum_{j \in C_A} n_{a_j} \exp(-\delta_{kj}^\omega / \gamma)$$

Equivalence

Ashby & Rosedahl (2017, Psychological Review)

$$P(A|k) = \frac{\beta_A \sum_{j \in C_A} n_{a_j} \exp(-\delta_{kj}^\omega / \gamma)}{\beta_A \sum_{j \in C_A} n_{a_j} \exp(-\delta_{kj}^\omega / \gamma) + \beta_B \sum_{j \in C_B} n_{b_j} \exp(-\delta_{kj}^\omega / \gamma)}$$

$$P(A|k) = \frac{\beta_A \sum_{i \in C_A} V_{iA} \eta_{ik}}{\beta_A \sum_{i \in C_A} V_{iA} \eta_{ik} + \beta_B \sum_{i \in C_B} V_{iB} \eta_{ik}}$$

$$\eta_{ik} = \exp(-c\delta_{ik}^\omega)$$

Relaxing Assumptions

We ran the model for the Dimensional, Criss-Cross, Interior-Exterior, and Diagonal Conditions of Nosofsky 1986.

The average r^2 between the model with and without assumptions was .983, so removing the assumptions has little impact on the model's performance

The average r^2 between the GCM and the model without assumptions was .992 (lowest was .990), so relaxing the assumptions does not lead to behaviors that fall outside of GCM

Question

So if the assumptions are required for mathematical equivalence, why are the models still equivalent without them?

Big Picture

Summed Similarity is Equivalent to Kernel Density Estimation with RBFs
(Ashby & Alfonso-Reese 1995)

KDE is the sum of the Radial Basis Functions

Big Picture

Summed Similarity is Equivalent to Kernel Density Estimation with RBFs

(Ashby & Alfonso-Reese 1995)

KDE is the sum of the Radial Basis Functions

Reinforcement Learning (RL) with RBFs is also the sum of RBFs

Big Picture

Summed Similarity is Equivalent to Kernel Density Estimation with RBFs

(Ashby & Alfonso-Reese 1995)

KDE is the sum of the Radial Basis Functions

Reinforcement Learning (RL) with RBFs is also the sum of RBFs

Therefore:

Summed Similarity = Summed RBFs = RL with RBFs

Benefits of the Neural Interpretation

Expands Exemplar Theory to explain results such as:

1: Button-Switch Interference (Ashby et al. 2003)

2: Feedback Delay Effects (Smith et al. 2014)

3: Retinal Specific Learning (Rosedahl, Eckstein, & Ashby, 2018)

4: Task-Related Striatal Activity (Seger et al. 2010)

Predicting Categorization Difficulty

We can build off of the results here to derive a difficulty measure for procedural categorization tasks for a human performer.

Single Stimulus Difficulty

- $P(A | k) = \frac{S_{A|k}}{S_{A|k} + S_{B|k}}$

$$S(A|k) = \sum_{i \in C_A} \eta_{ki}, \quad \eta_{ki} = \exp(-D_{ki}^\omega / \gamma),$$

- Difficulty for stimulus $k \propto \frac{S_{B|k}}{S_{A|k}}$

Category Difficulty

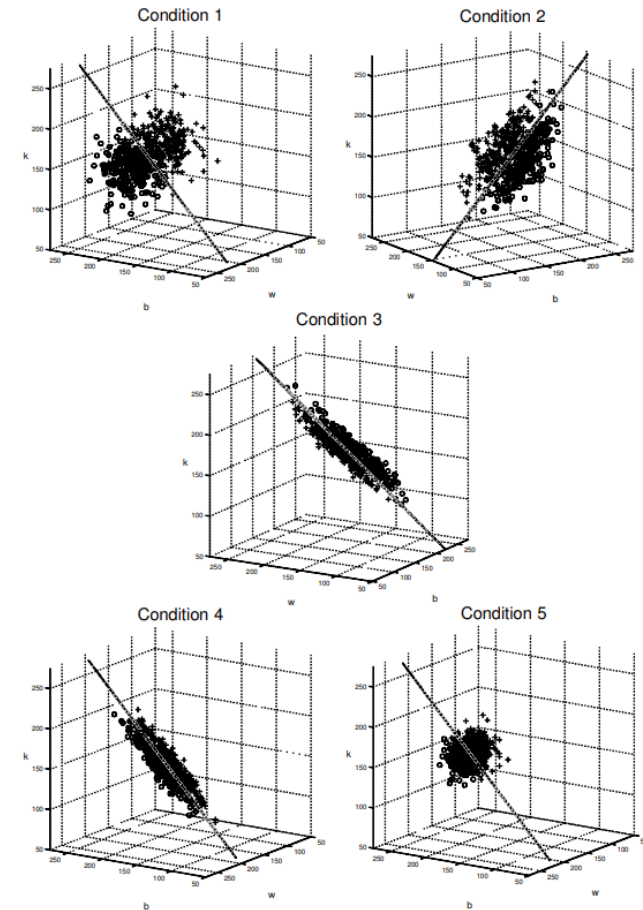
- Difficulty for category A $\propto \frac{S_{B|A}}{S_{A|A}}$

$$S(A|A) = \sum_{k \in C_A} \sum_{i \in C_A, i \neq k} \eta_{ki},$$

- Difficulty for A & B $\propto \frac{S_{B|A} + S_{A|B}}{S_{A|A} + S_{B|B}}$
 - Infinitely Separated Difficulty = 0
 - Totally Overlapping Difficulty = 1

Previous Attempts

- Covariance Complexity
 - $C3=C4>C2>C1=C5$
- Ideal Observer
 - $C5 \gg C1 > C2 > C3=C4$
- Our Measure
 - $C3 \approx C5 \approx C4 > C2 > C1$



ALFONSO-REESE, ASHBY, AND BRAINARD, 2002

Previous Attempts

- Covariance Complexity

- $C3=C4>C2>C1=C5$

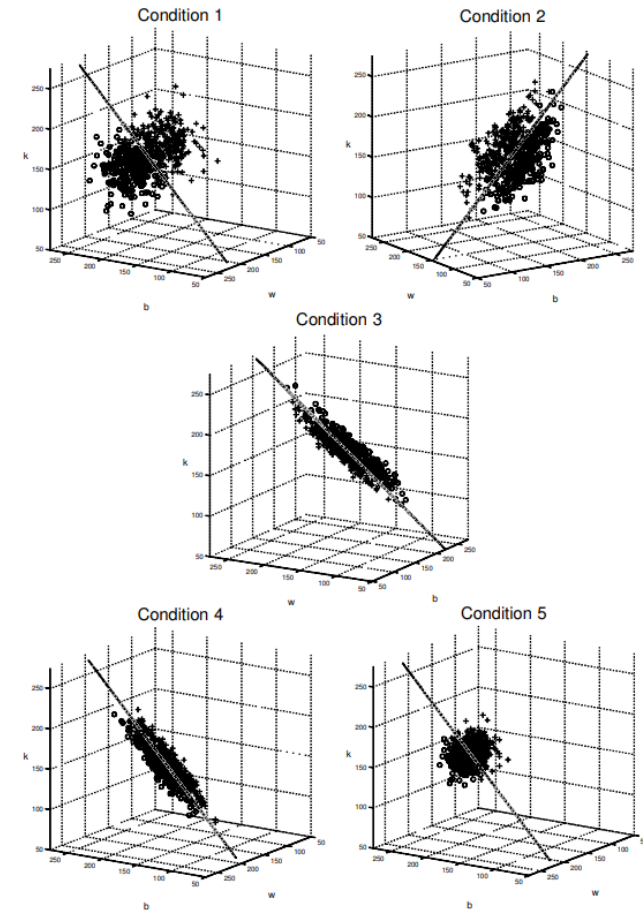
- Ideal Observer

- $C5 \gg C1 > C2 > C3 = C4$

- Our Measure

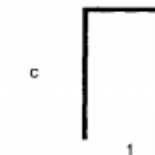
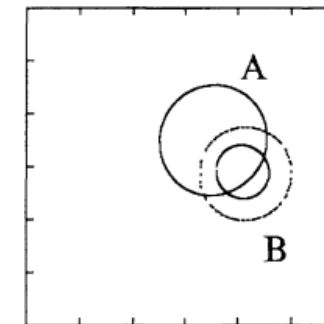
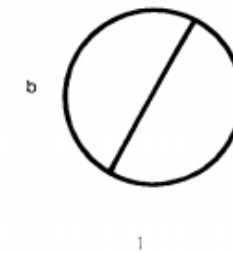
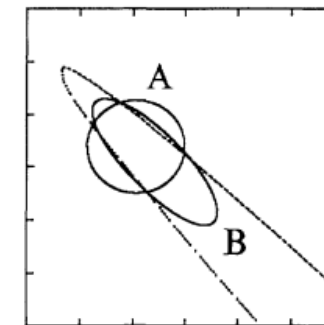
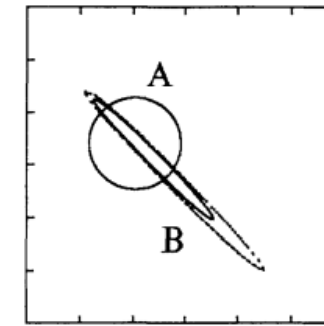
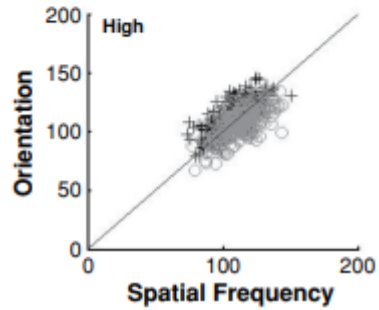
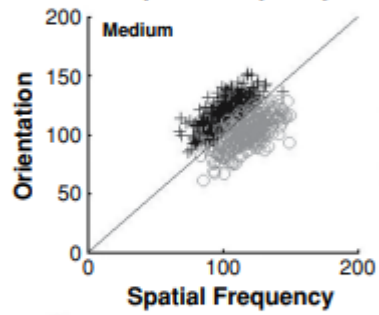
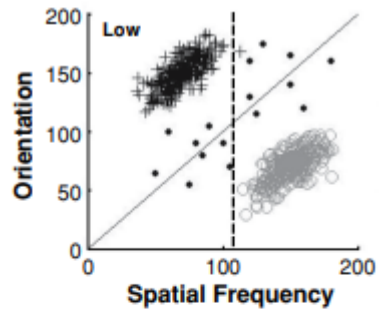
- $C3 \approx C5 \approx C4 > C2 > C1$

Correct!



ALFONSO-REESE, ASHBY, AND BRAINARD, 2002

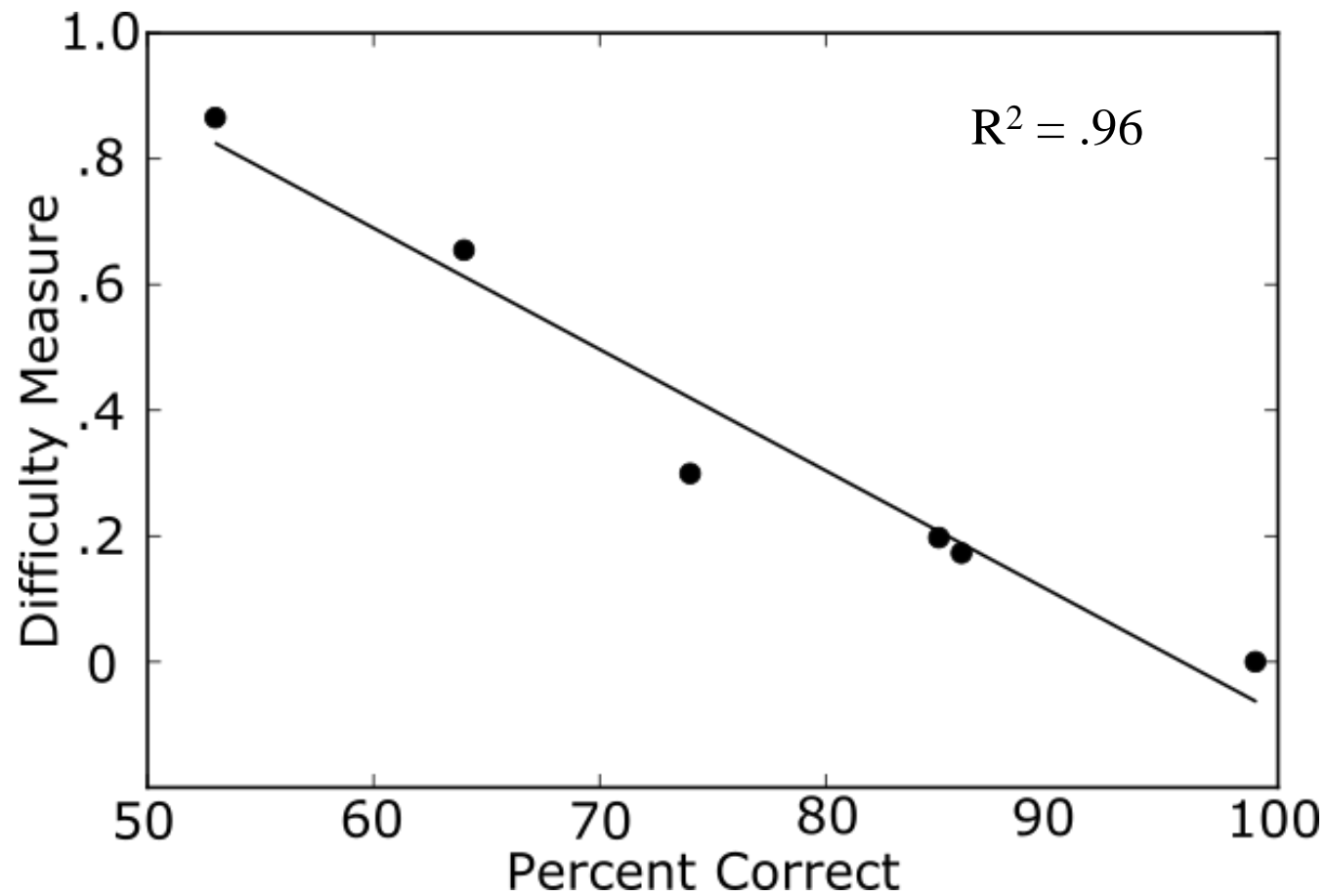
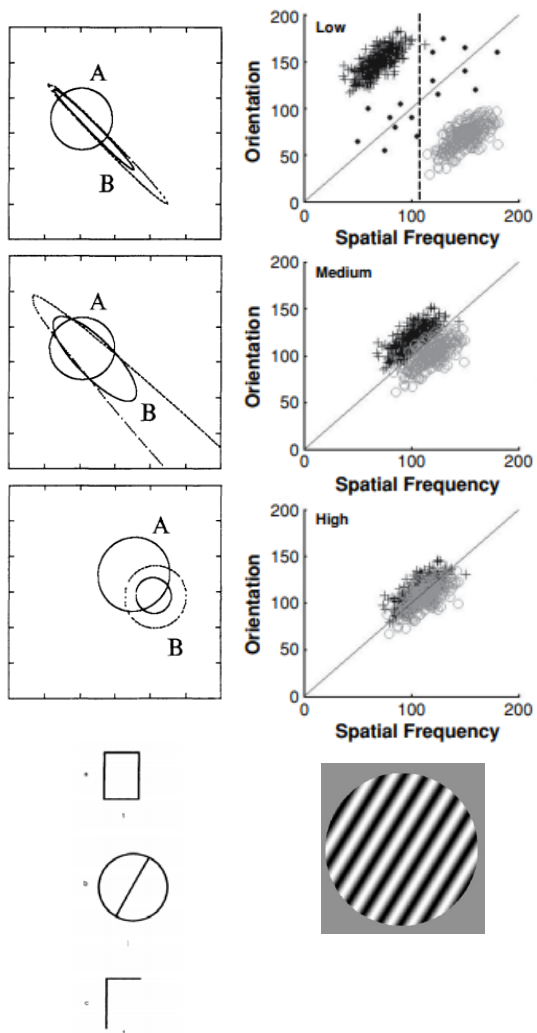
Comparing Across Stimuli Types



Eli & Ashby, 2006

Ashby & Maddox, 1992

Results



Conclusions

The neural version of exemplar theory:

- expands exemplar theory to be compatible with almost all patient and neuroimaging categorization data
- can be used to develop a difficult measure applicable to a wide range of category structures and stimulus types
- Potentially useful when providing computer assistance during categorization tasks

Acknowledgements



Laboratory for Computational
Cognitive Neuroscience



F. Gregory Ashby

Funding: National Science Foundation Graduate Research
Fellowship under Grant No. 1650114, NIMH Grant
MH3760-2