Do we develop visual representations based on pair-wise statistics of the visual scene?

József Fiser

Brandeis University

Gergő Orbán
Brandeis University

Máté Lengyel
Gatsby Unit, UCL

Richard Aslin
University of Rochester

Conceptual framework

- What to compute — Independent underlying components (functional model)
- How to compute — Probabilistic normative approach (full inference)
- How to validate — Compare it to physiological measures (spontaneous activity)

How to extend this framework to visual learning?

- What to compute — Learning representations of the underlying independent components
- How to compute — Different types of representational learning schemes
- How to validate — Compare the performance of learning schemes to human performance

Question

How do humans develop internal representations of the hierarchically structured sensory information they perceive?

The problem of chunking

Some possible answers

By remembering everything holistically

Keeping simple statistics of appearances

Applying recursively pairwise Hebbian associative learning

......

Alternative learning models

1-2 Lookup table for occurrence, co-occurrence frequencies

$P(\blacksquare)$ or $P(\blacksquare, \maltese)$ – averaged sum of episodic memory traces
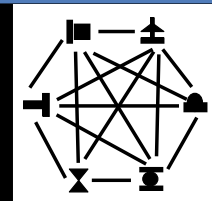
3 Lookup table for conditional probabilities

$P(\blacksquare | \maltese)$ – statistical learning   *Naïve statistical learners*
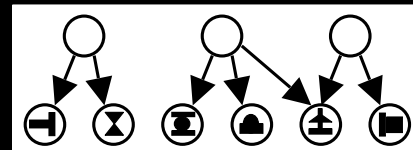
4 Second order correlations

any (recursive) pairwise associative learning

5 Bayesian ideal learner

explicit chunking

Probabilistic learners

An "ideal" Bayesian learner

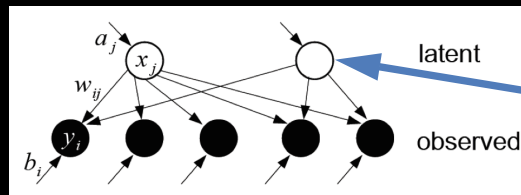Optimal chunking $\Rightarrow$ a model-selection problem

Choose the right inventory of chunks that both captures the previous scenes and generalizes well

Bayesian model comparison

$$P(scene_1, scene_2, \ldots scene_n \mid inventory)$$

$\Rightarrow$ Inventory = A complete model of the visual world
a.k.a. learned memory representation

Sigmoid Belief Network

Independent causes

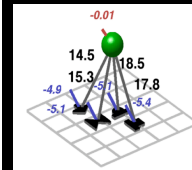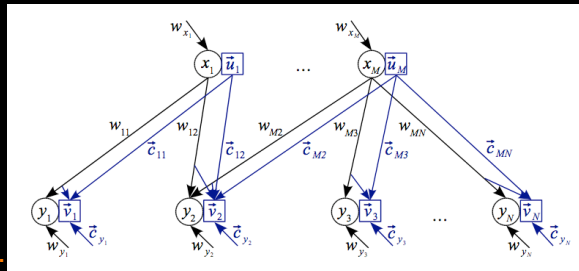# Product of Gaussians Sigmoid Belief Network

x - latent cause

u - latent position

$$P(x, u \mid \Theta_I, I) = \prod_i P(x_i, u_i \mid \Theta_I, I)$$
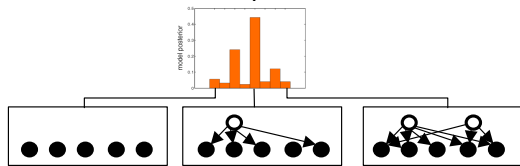
y - observed feature

v - observed relative pos.

$$P(y, v \mid x, u, \Theta_I, I) = \prod_j P(y_j, v_j \mid x, u, \Theta_I, I)$$
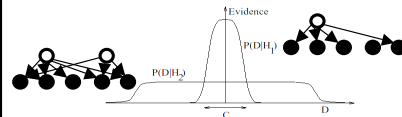

Model selection and automatic Occam's razor

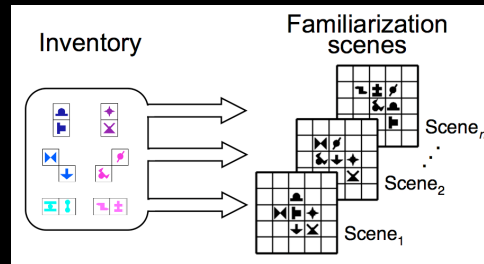Learn a distribution of possible model structures

Automatic Occam's Razor

$$\frac{P(\mathcal{H}_1 \mid D)}{P(\mathcal{H}_2 \mid D)} = \frac{P(\mathcal{H}_1)}{P(\mathcal{H}_2)} \frac{P(D \mid \mathcal{H}_1)}{P(D \mid \mathcal{H}_2)}$$

Validation: Human visual statistical learning
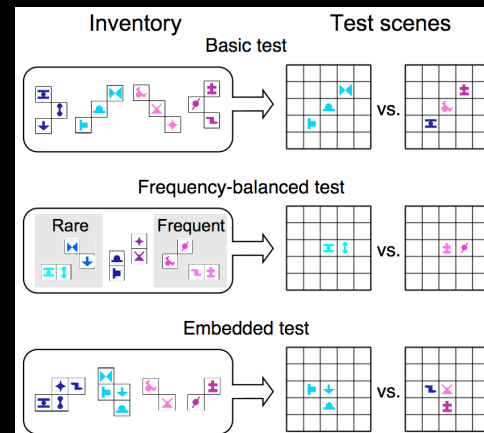### *(Fiser & Aslin 2001, 2002, 2005)*

Basic paradigm

- Unsupervised learning

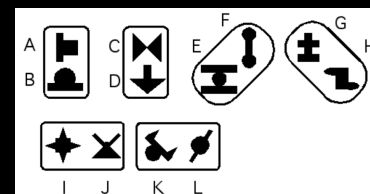- Familiarization and test

Test types
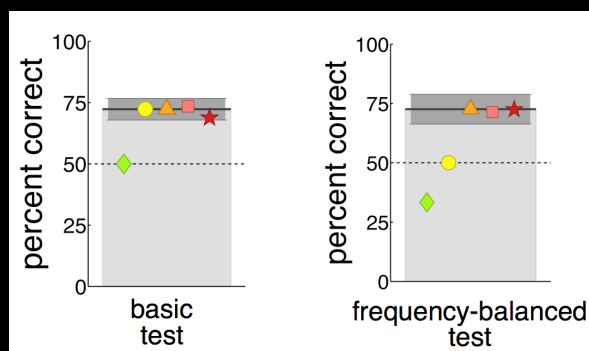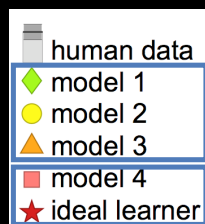
- Basic

- Frequency balanced

- Embedded

Validation results I

Basic experiments with pairs

Learning models learns from the same practice trials as humans do and chooses during test according to the log Prob ratio:
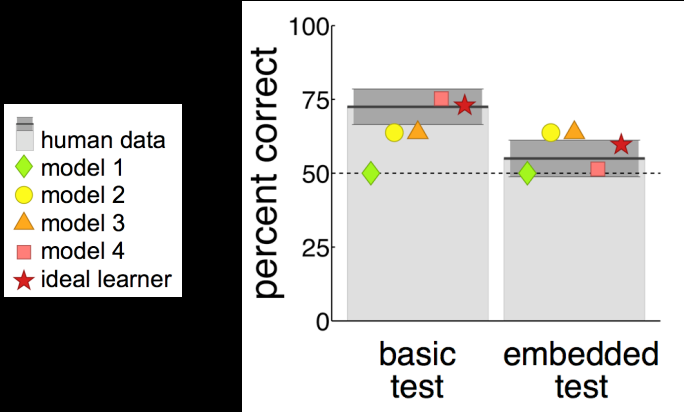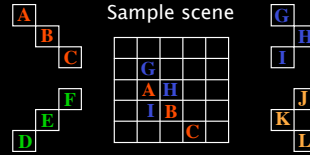
$$P(T) = \sum_I \int d\Theta_I P\left(y^{(T)}, v^{(T)} \mid \Theta_I, I\right) P(\Theta_I \mid I, D) P(I \mid D)$$

Legend:
- human data
- model 1
- model 2
- model 3
- model 4
- ideal learner

- Frequency-counting models fail

Validation results II

Embedded experiment with triplets

Sample scene

Legend:
- human data
- model 1 (green diamond)
- model 2 (yellow circle)
- model 3 (orange triangle)
- model 4 (red square)
- ideal learner (red star)
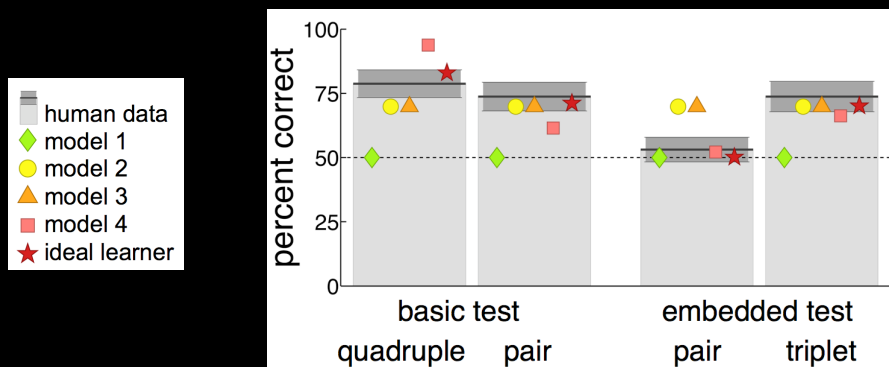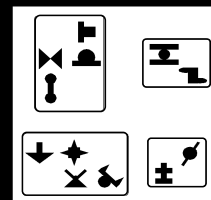
y-axis: percent correct (0, 25, 50, 75, 100)
x-axis: basic test, embedded test

- All three of the naïve statistical models fail

Validation results III

Embedded experiment with quadruples and pairs

Legend:
- human data
- model 1 (green diamond)
- model 2 (yellow circle)
- model 3 (orange triangle)
- model 4 (red square)
- ideal learner (red star)

y-axis: percent correct (0, 25, 50, 75, 100)
x-axis: basic test (quadruple, pair), embedded test (pair, triplet)
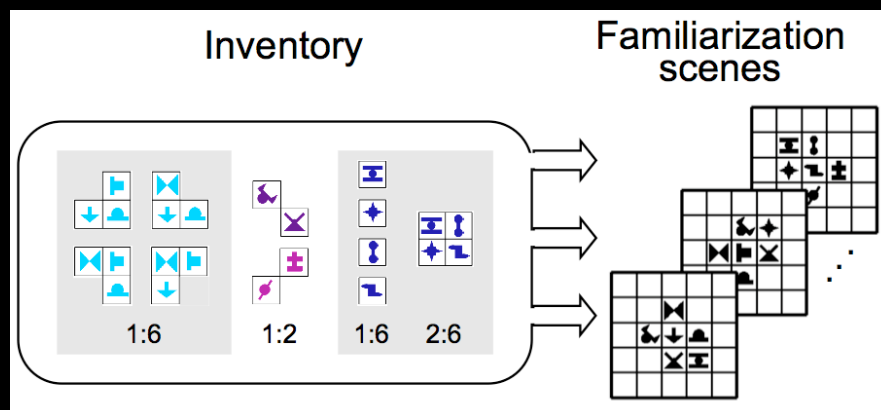
- All models but the Bayesian ideal learner fail

A wish list

- An experiment with different predicted outcome by the associative and the ideal Bayesian learning model

- A quantitavie prediction by the ideal learner

- … without any change in the model parameters

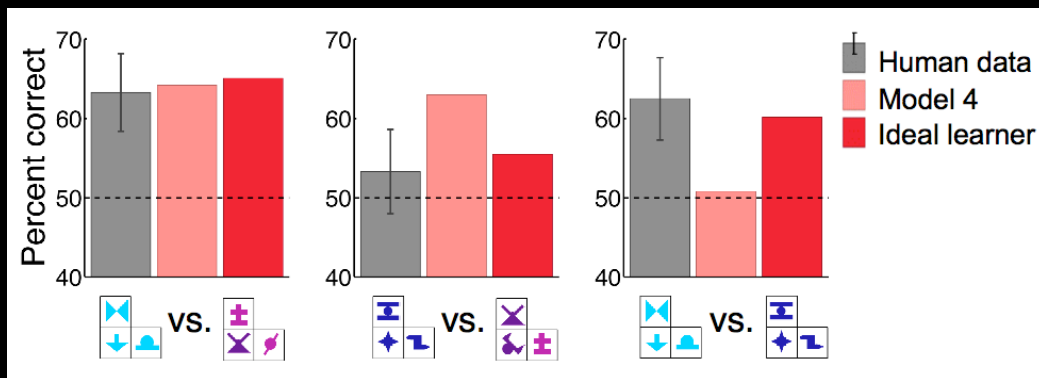# Chunks versus second-order correlations

Should humans learn chunks when first and second-order statistics contain no relevant information?
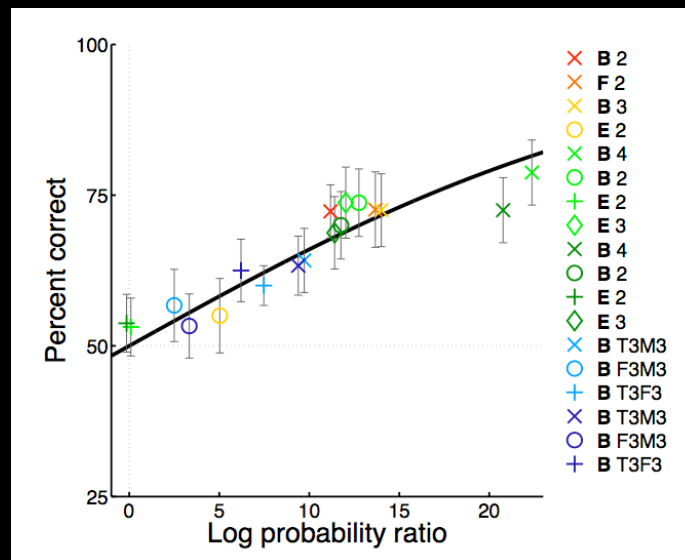
Experiment:

Test: true and false triplets against random triplets

Results

- The Bayesian ideal learner follows human performance whereas the associative learner (Model 4) does not

Summary

- Correct prediction without any parameter tweaking

Conclusions

- Human visual chunk learning performance is remarkably close to Bayes-optimal and cannot be captured by naïve statistical learning or by iterative pair-wise associative learning

- Humans seem to learn new complex information by generating the simplest sufficient representation of the input based on previous experience and not by encoding its full correlational structure

- The present results open the possibility to treat instant perception and learning behavior within the same probabilistic framework