C36

Models of Visual Object Recognition in Humans

R.J. Peters (1), F. Gabbiani (1), J. Jovicich (1,2), L. Chang(2), T. Ernst(2) and C. Koch (1)
(1) Computation and Neural Systems, Caltech, Pasadena, CA 91125
(2) UCLA School of Medicine, Harbor-UCLA Medical Center, Torrance, CA 90209

1. Introduction

1.1 Background

Visual object recognition is about attaching a meaning or label to a visually perceived object. Recognition and categorization are close kin.

The labels depend on the question being asked:

What is it?	*What kind?*	*Which one?*
an apple	*a Fuji apple*	*the one on my plate*
Basic-level categorization	**Subordinate-level categorization**	**Individual exemplar recognition**

1.2 Goals

To investigate the representation(s) underlying visual object recognition:

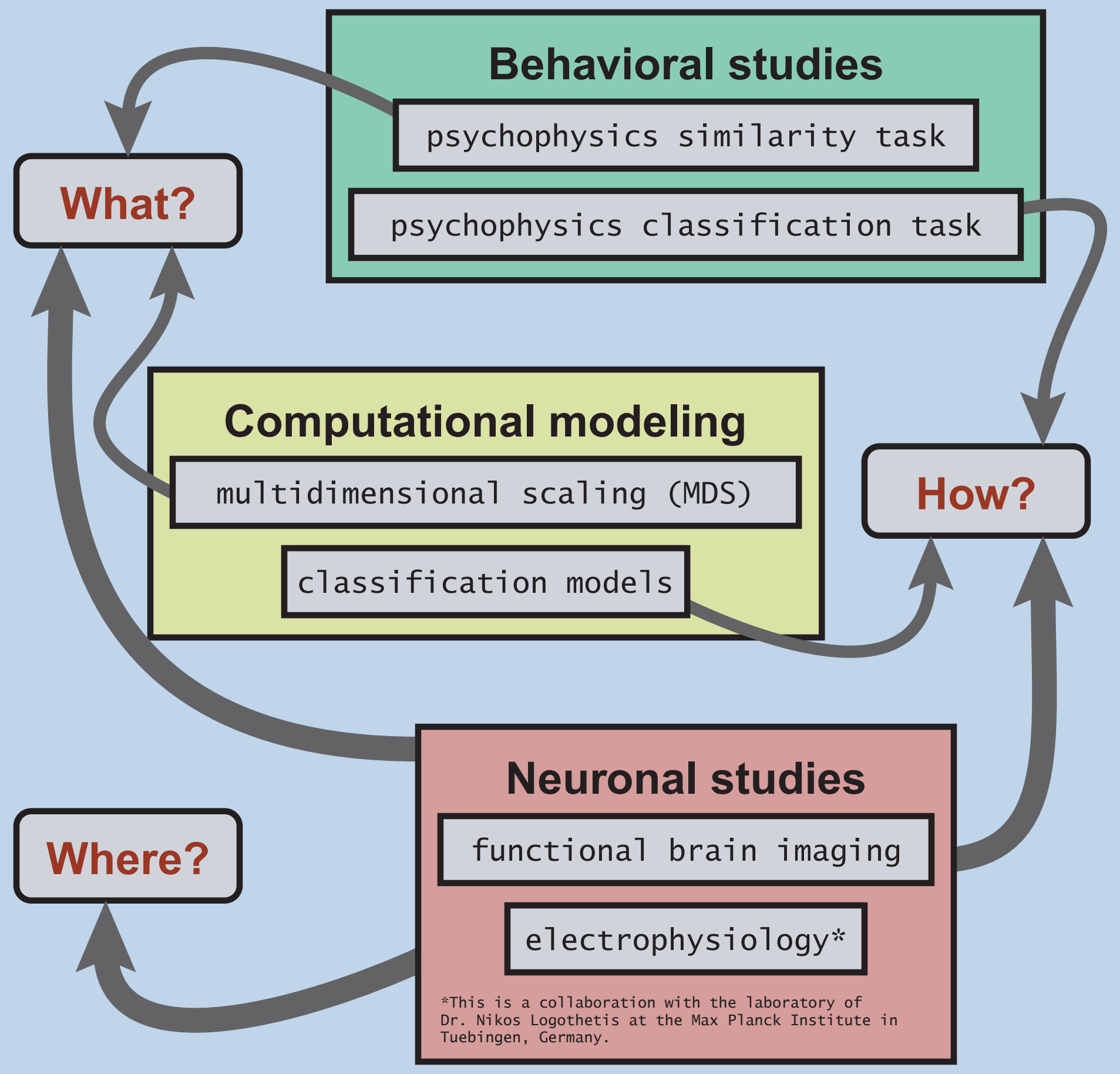
What is represented? (lines, features, distances?)
How is it represented? (what are single neurons doing?)
Where is it represented? (what brain regions are used?)

To study the neuronal and computational bases of subordinate-level category and individual exemplar representation.

To study how these representations depend on the familiarity of the observer with the visual stimulus.

1.3 General Methods

To answer *What*, *How*, and *Where*, we combine methods:



Behavioral studies
- psychophysics similarity task
- psychophysics classification task

Computational modeling
- multidimensional scaling (MDS)
- classification models

Neuronal studies
- functional brain imaging
- electrophysiology*

*This is a collaboration with the laboratory of Dr. Nikos Logothetis at the Max Planck Institute in Tuebingen, Germany.

Acknowledgements

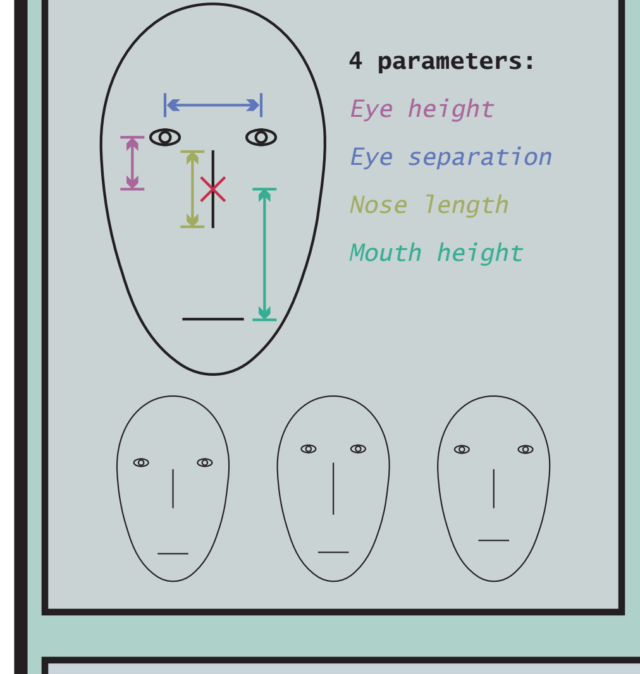
This work was supported by grants from NSF, NIMH, Keck and HHMI.

Thanks to N. Kanwisher for the use of the face/house photographs.

2. Psychophysics

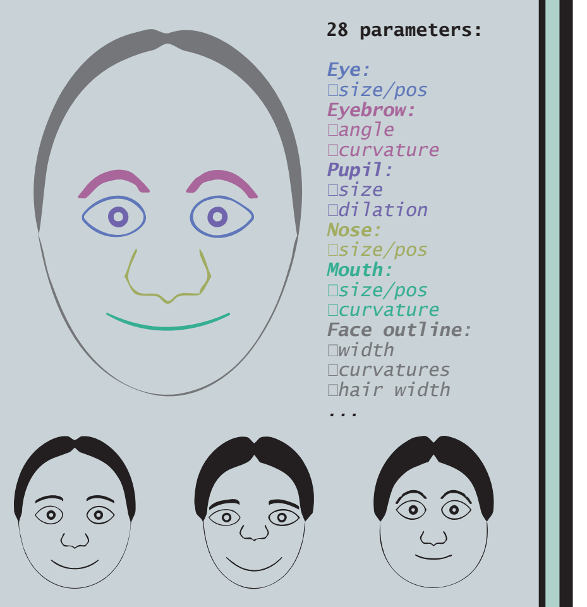
2.1 Stimuli

Brunswik faces



4 parameters:
- *Eye height*
- *Eye separation*
- *Nose length*
- *Mouth height*

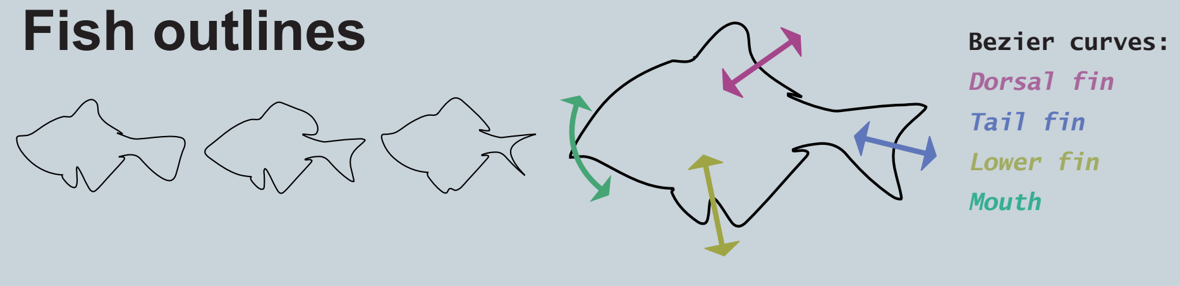
Cartoon faces



28 parameters:

Eye:
□*size/pos*
Eyebrow:
□*angle*
□*curvature*
Pupil:
□*size*
□*dilation*
Nose:
□*size/pos*
Mouth:
□*size/pos*
□*curvature*
Face outline:
□*width*
□*curvatures*
□*hair width*
...

Fish outlines

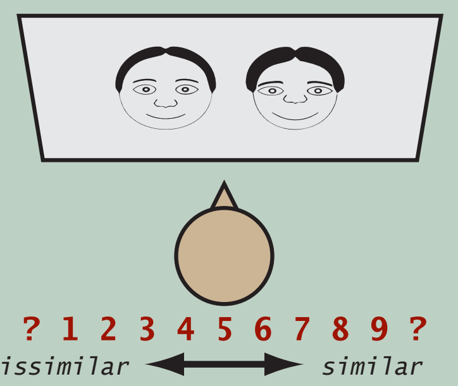


Bezier curves:
- *Dorsal fin*
- *Tail fin*
- *Lower fin*
- *Mouth*

The objects' parameters were unknown to subjects.

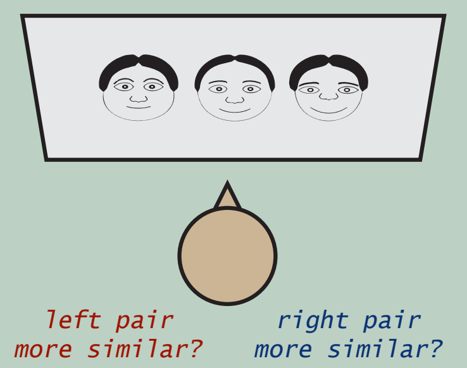
2.2 Similarity tasks

Pairs task



Subjects viewed pairs of objects for up to 2 seconds, and pressed a key between 1-9 to indicate how similar the two objects appeared. For each set of 20 objects, 1200 trials were run (*20 objects* x *20 objects* x *3 repeats*).

Triads task



Subjects viewed triplets of objects for up to 2 seconds, and pressed a key to indicate whether the left or right pair appeared more similar. For the 20 Brunswik faces, 1721 trials were run (selected to optimize the MDS fit).

Other experiments have typically use either the pairs or triads task, but not both. However, we needed to validate comparisons between monkey and human data obtained with the different tasks, because 1) monkeys can only learn the triads task, and 2) human subjects prefer the pairs task (it requires fewer trials). Thus we compared the two tasks directly in our initial human studies with the Brunswik faces. When used in MDS analysis and classification models, the two tasks gave similar results, so in later experiments only the pairs task was used.

2.3 Classification task

For each set of objects, two categories were defined, each consisting of 5 exemplar objects. Each set also contained 10 test exemplars different from the training exemplars.

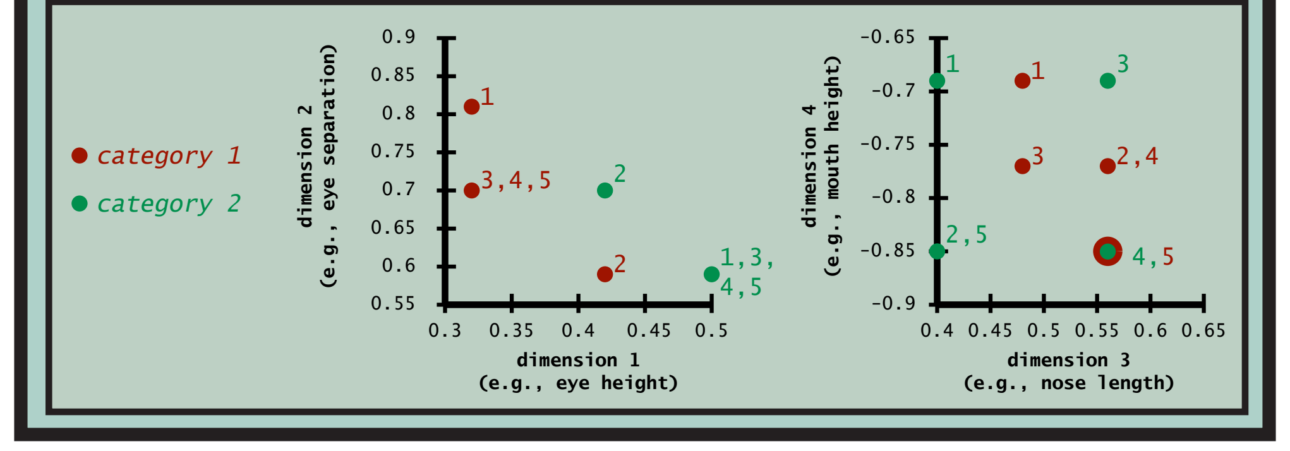
In a 2-AFC task with auditory feedback, subjects were trained to categorize the training exemplars at 85% accuracy. This typically required 2-5 blocks of 100 trials (1 block = 10 training exemplars x 10 repeats).

Then, in a similar 2-AFC task, but without audio feedback, subjects categorized both the training exemplars and the test exemplars. These blocks also consisted of 100 trials (1 block = (10 training exemplars + 10 test exemplars) x 5 repeats).

The training/testing cycle was repeated in 5 separate sessions for each subject.

2.4 Categories

For each object type, the categories formed the same logical configuration of training exemplars, by substituting the objects' parameters for dimensions 1-4 (shown below). Note that the categories were linearly separable. The parameters for each dimension were quantized to three values, so that the entire set of objects occupied a 3x3x3x3 lattice.



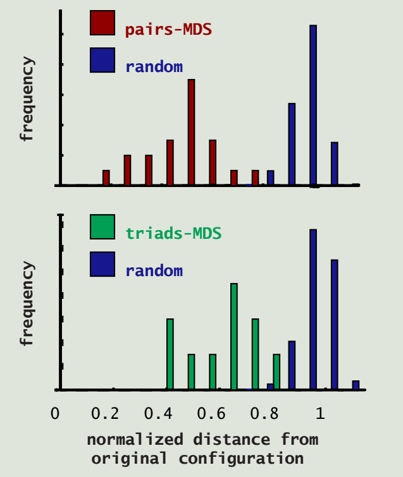
3. Models

3.1 MDS (Multidimensional scaling)

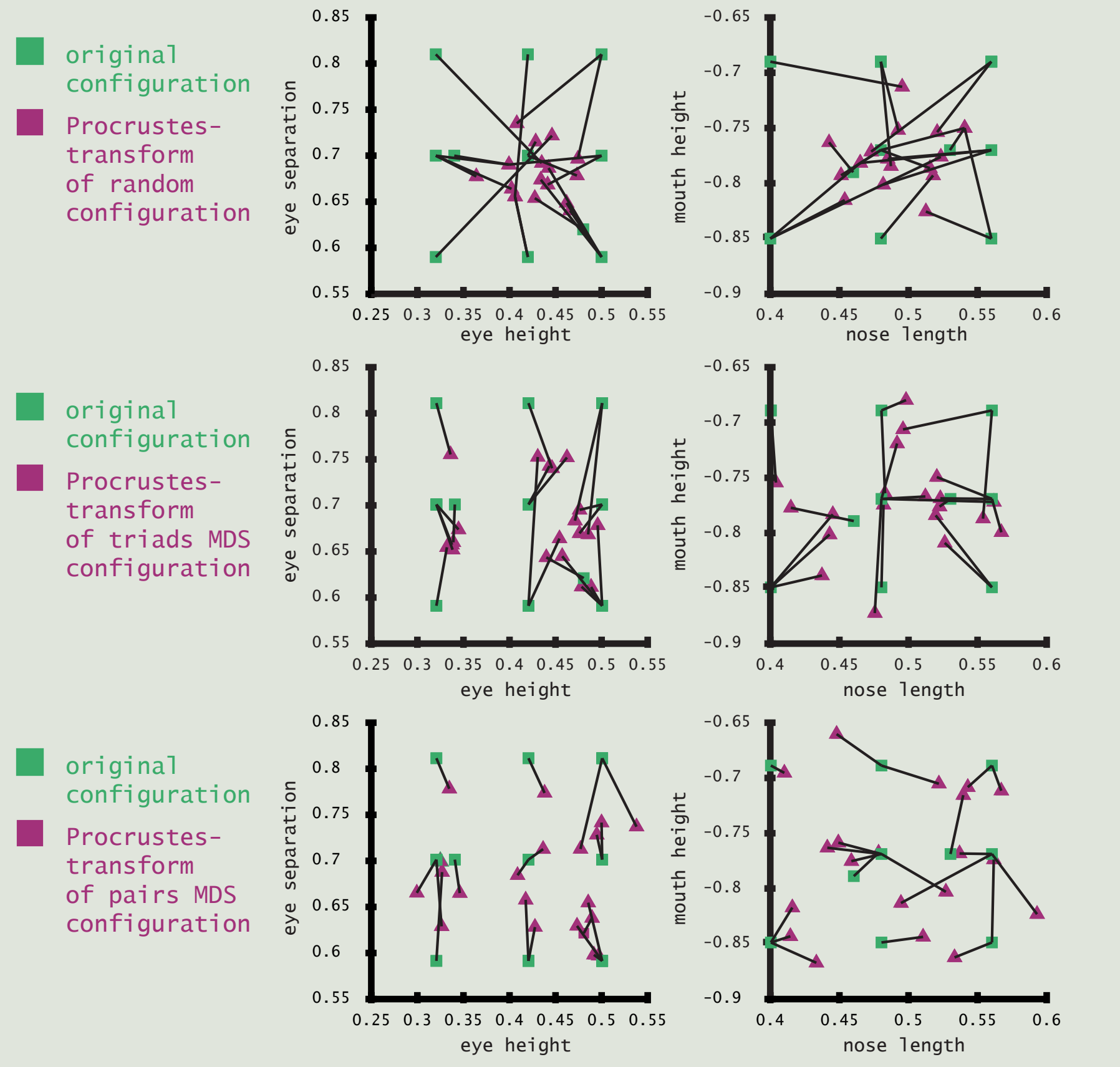
Using data from the similarity task as a distance metric, MDS was used to generate 4-dimensional configurations of the object sets. A Procrustes transform (uniform scaling, reflection, and rotation) was applied to align this configuration with the original object configuration.

Monte Carlo simulations showed that both pairs- and triads-based MDS configurations are always closer to the original configuration than would be expected of a configuration drawn by chance. This suggests a similarity between the features encoded at the neuronal level and the variable features in the visual objects.

Random configurations and *pairs-MDS* and *triads-MDS* configurations were Procrustes-transformed to the original object space. The distance between two configurations was defined as the average distance between the corresponding exemplar representations in the two configurations. The distances between configurations were normalized to the mean distance from the target configuration of random configurations. The histograms demonstrate that the pairs-based MDS configurations are somewhat closer to the original space than are the triads-based MDS configurations.

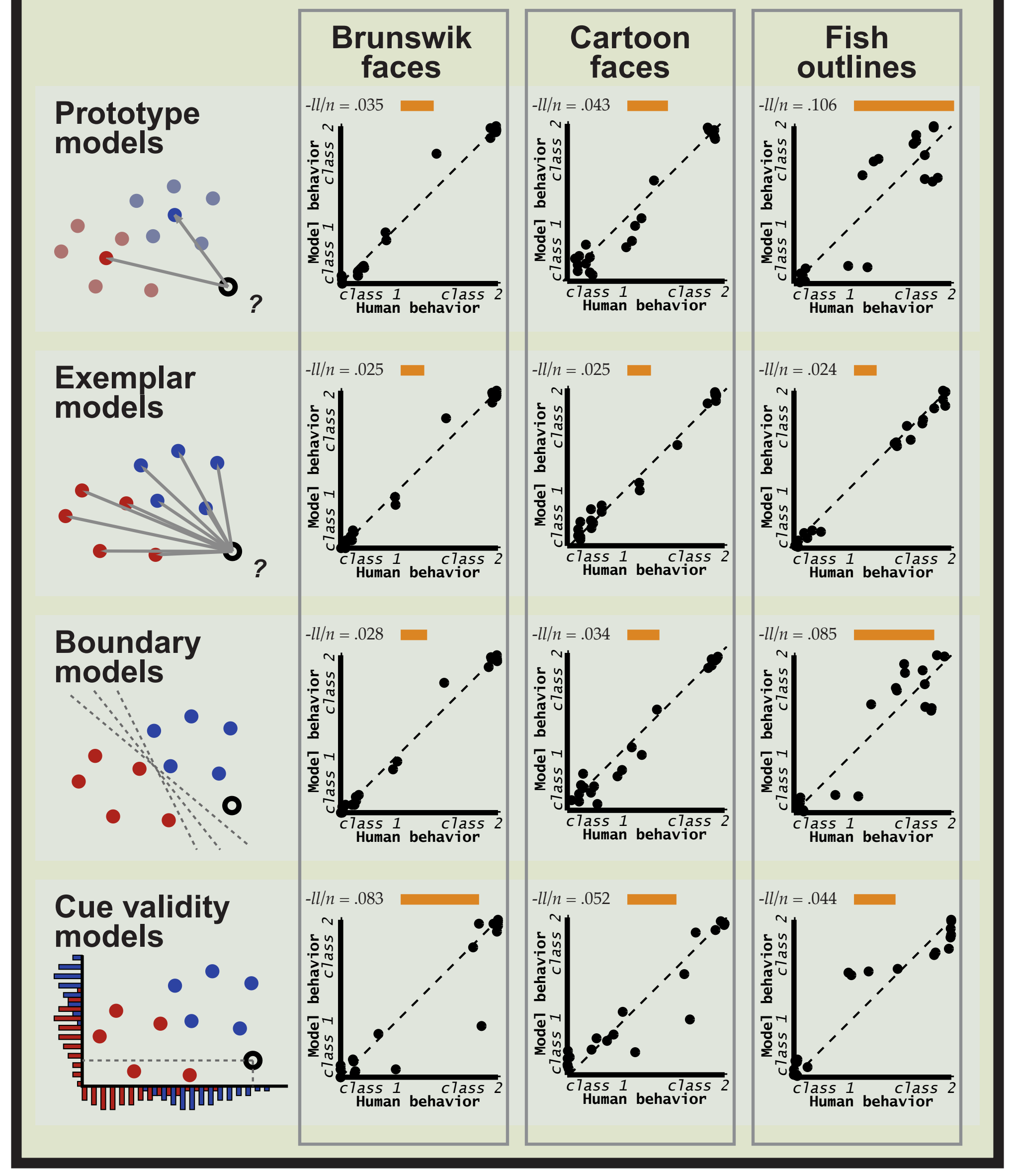


These 2x2D plots show the set of Brunswik faces in their original configuration, and in Procrustes-transformed configurations based on either a random configuration or a triads or pairs MDS configuration from a single subject.



3.2 Classification models

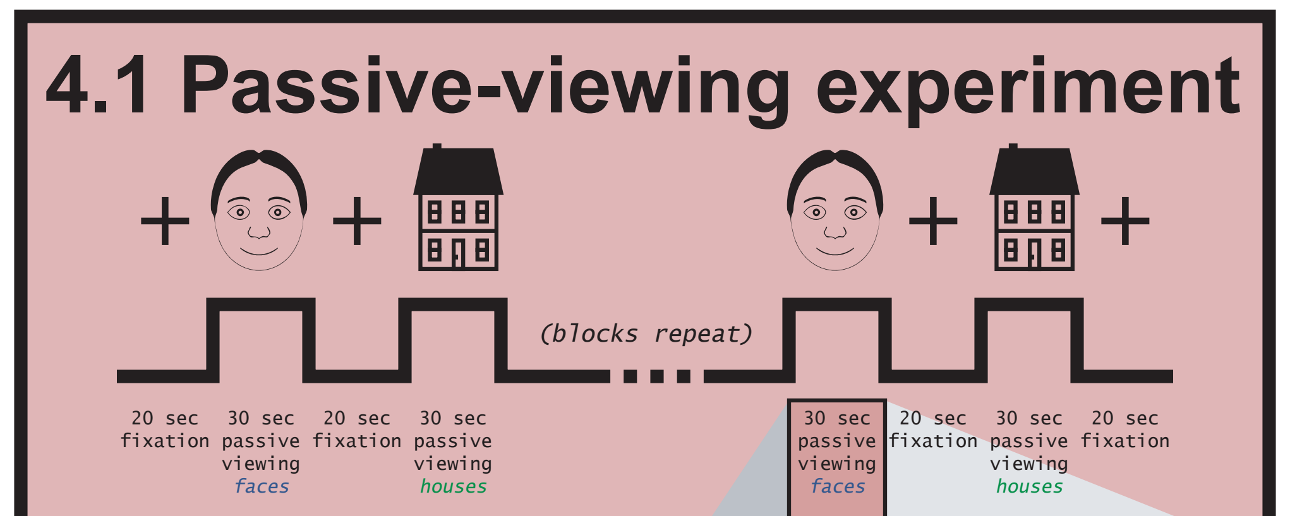
For each object class, we tested four classification models using both the original and MDS configurations. The models were designed to test human classification strategies, rather than to maximize overall classification accuracy. The graphs below summarize the performance of the models based on the original configurations (which typically performed at least as well as the MDS-based models). The human classification data are pooled across subjects (n=5). For each model, *minus the loglikelihood per trial* ($-ll/n$) is shown along with a bar of proportional length. A smaller value represents a better fit. The pattern of model performances depends significantly on the object class, but in general, *exemplar models perform best, while a prototype alone does not account for human performance*.



	Brunswik faces	Cartoon faces	Fish outlines
Prototype models	$-ll/n = .035$	$-ll/n = .043$	$-ll/n = .106$
Exemplar models	$-ll/n = .025$	$-ll/n = .025$	$-ll/n = .024$
Boundary models	$-ll/n = .028$	$-ll/n = .034$	$-ll/n = .085$
Cue validity models	$-ll/n = .083$	$-ll/n = .052$	$-ll/n = .044$

4. fMRI

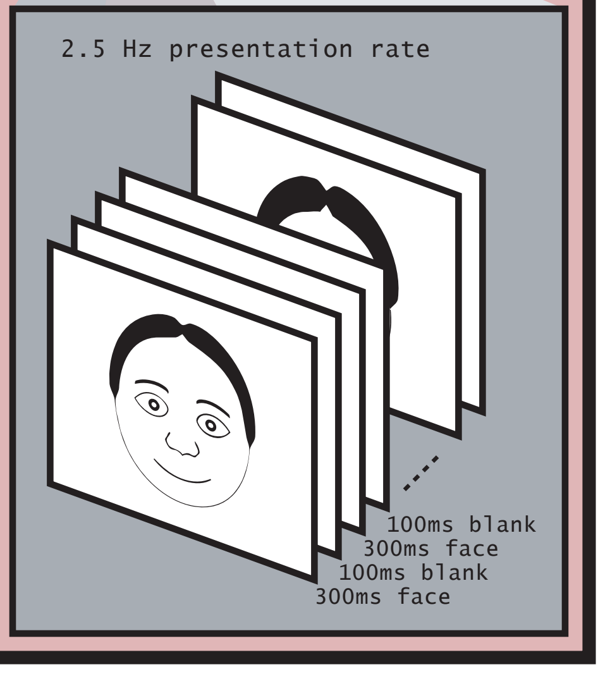
4.1 Passive-viewing experiment


(blocks repeat)
20 sec fixation
30 sec passive viewing faces
20 sec fixation
30 sec passive viewing houses
30 sec passive viewing faces
20 sec fixation
30 sec passive viewing houses
20 sec fixation

Scanner: Whole brain single shot T_2*-weighted spiral functional images were acquired using the manufacturer's head birdcage coil on a 1.5-Tesla scanner (General Electric Signa).

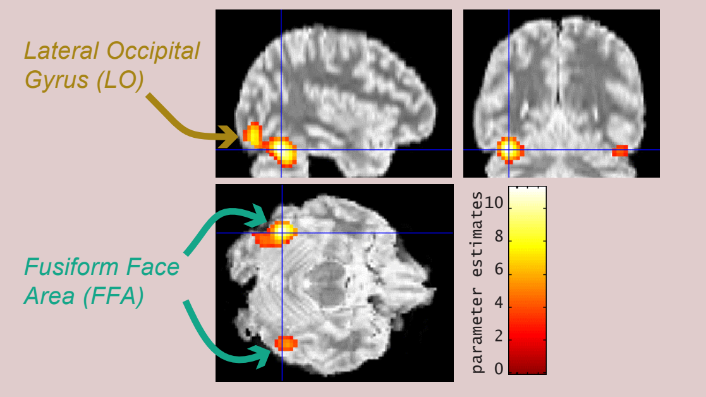
Imaging parameters: TE=50ms, TR=2500ms, 3.125x3.125mm in-plane resolution, 4mm-thick axial slices, 1mm slice gap.

Data analysis: Standard procedures were used to process the data in SPM99b. Data were motion corrected, Tailarach normalized, and spatially smoothed prior to statistical analysis. In each subject, regions of interest were identified that responded preferentially to faces. The peak activities in these areas were averaged

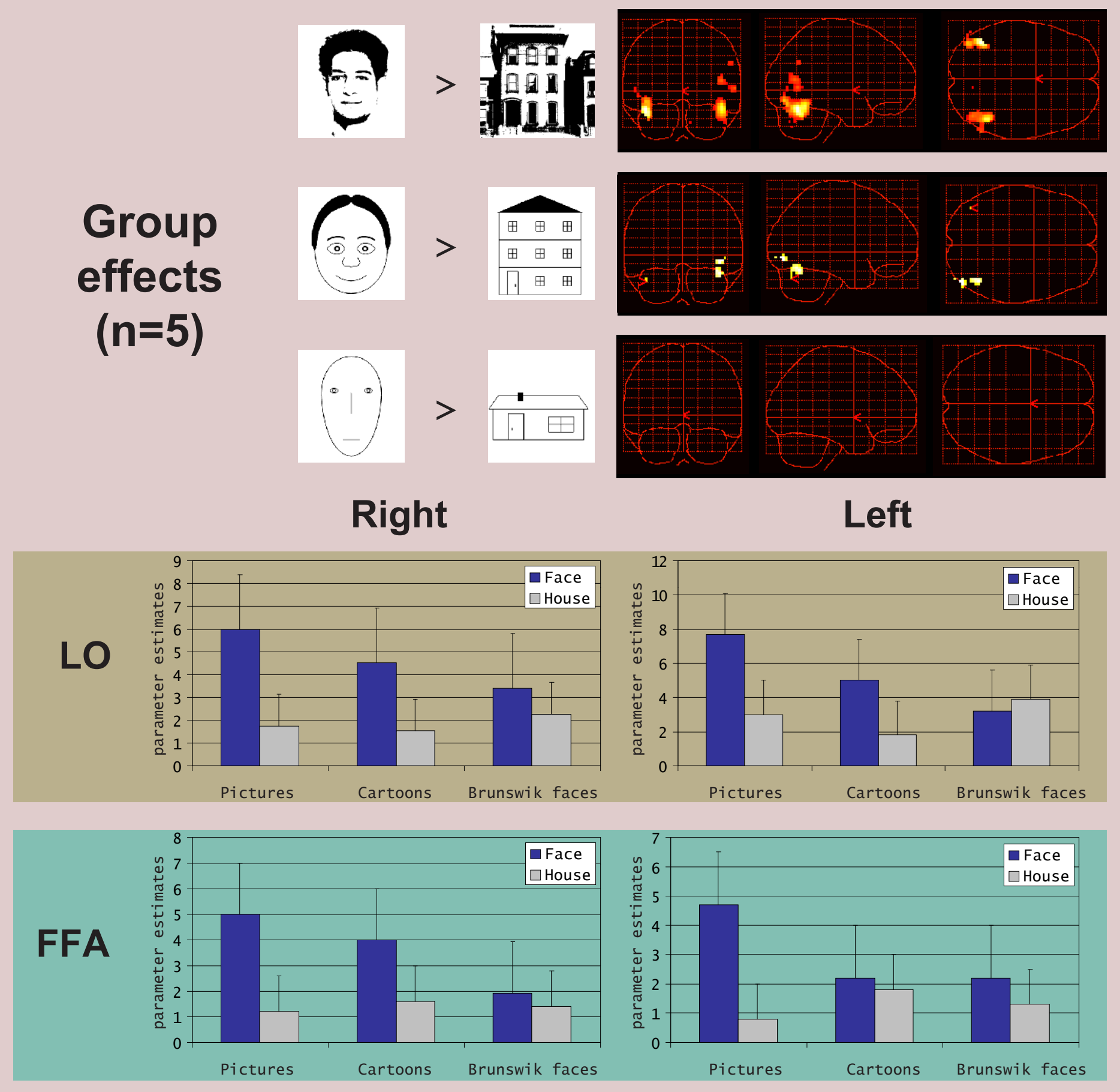


4.2 Results

Passive viewing of face pictures activates areas in the middle fusiform gyrus (FFA) and in the lateral occipital gyrus (LO).



Face-related activity in LO and FFA decreases as the objects become more schematic.



5. Summary

The experiments reported here were designed to investigate the representations underlying visual object recognition.

What is represented?

The similarity of subjects' MDS spaces to the objects' original parameter spaces suggests (1) there is flexibility in what is represented, since subjects learned the features during the experiment, and (2) this flexibility allows internal representations to faithfully reflect natural external parameter spaces.

How is it represented?

The results of the classification models suggest that representations of subordinate-level categories must retain information about individual exemplars. The differences between classification models for face and fish images may reflect different representations used for familiar and novel stimuli.

Where is it represented?

Cartoon faces with sufficient detail give rise to activity in FFA and LO. The lack of activity for simpler faces may be due to a lack of familiarity, or to lower variability within the dataset. Future human fMRI and monkey electrophysiology studies will consider how the representations in areas such as FFA and LO may be used to accomplish classification and recognition tasks.