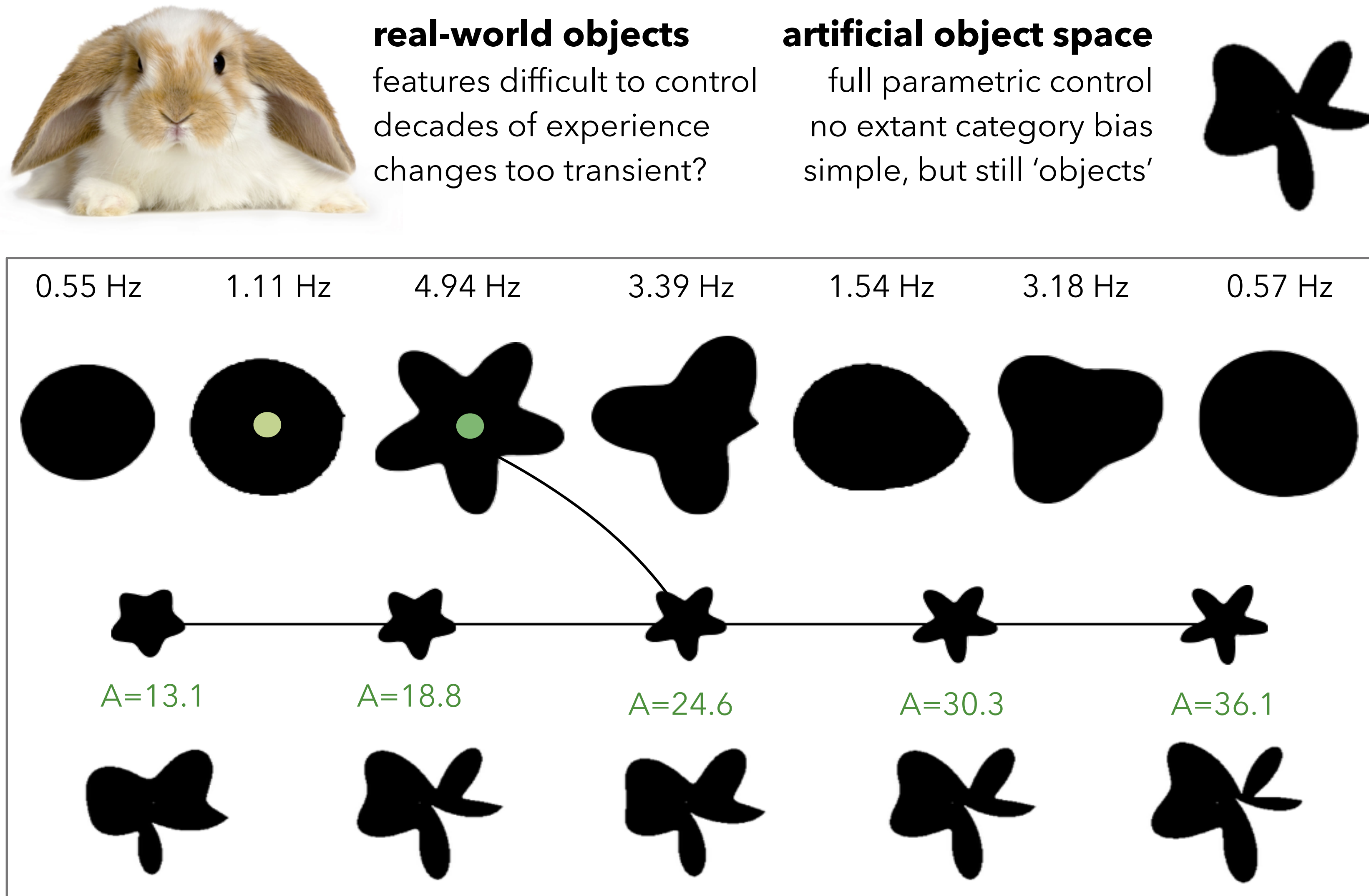# KL-Evidence: A Novel Multivariate Neurofeedback Method for Differentiating Representations

Marius Cătălin Iordan[1], Victoria J.H. Ritvo[1], Kenneth A. Norman[1]
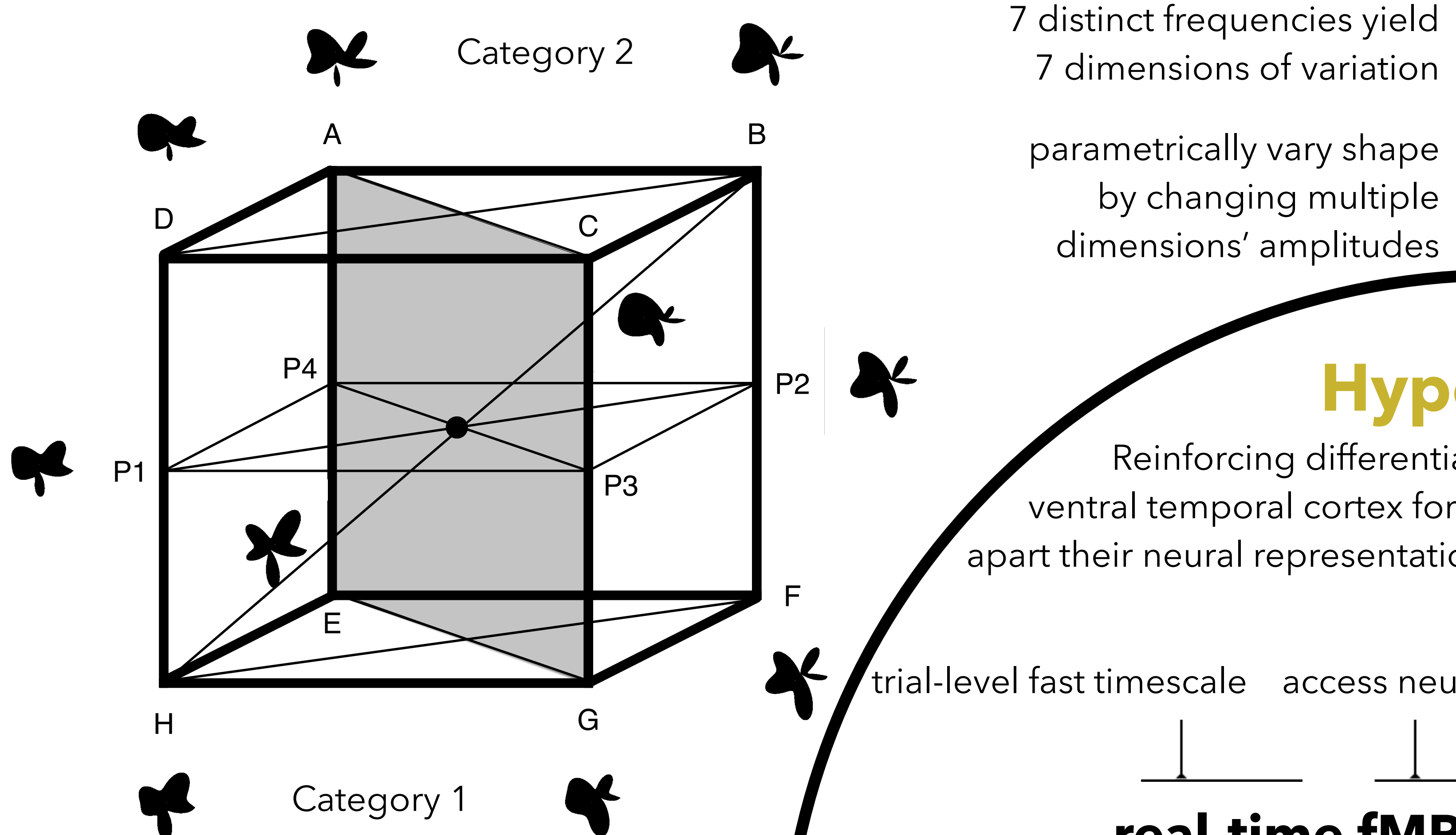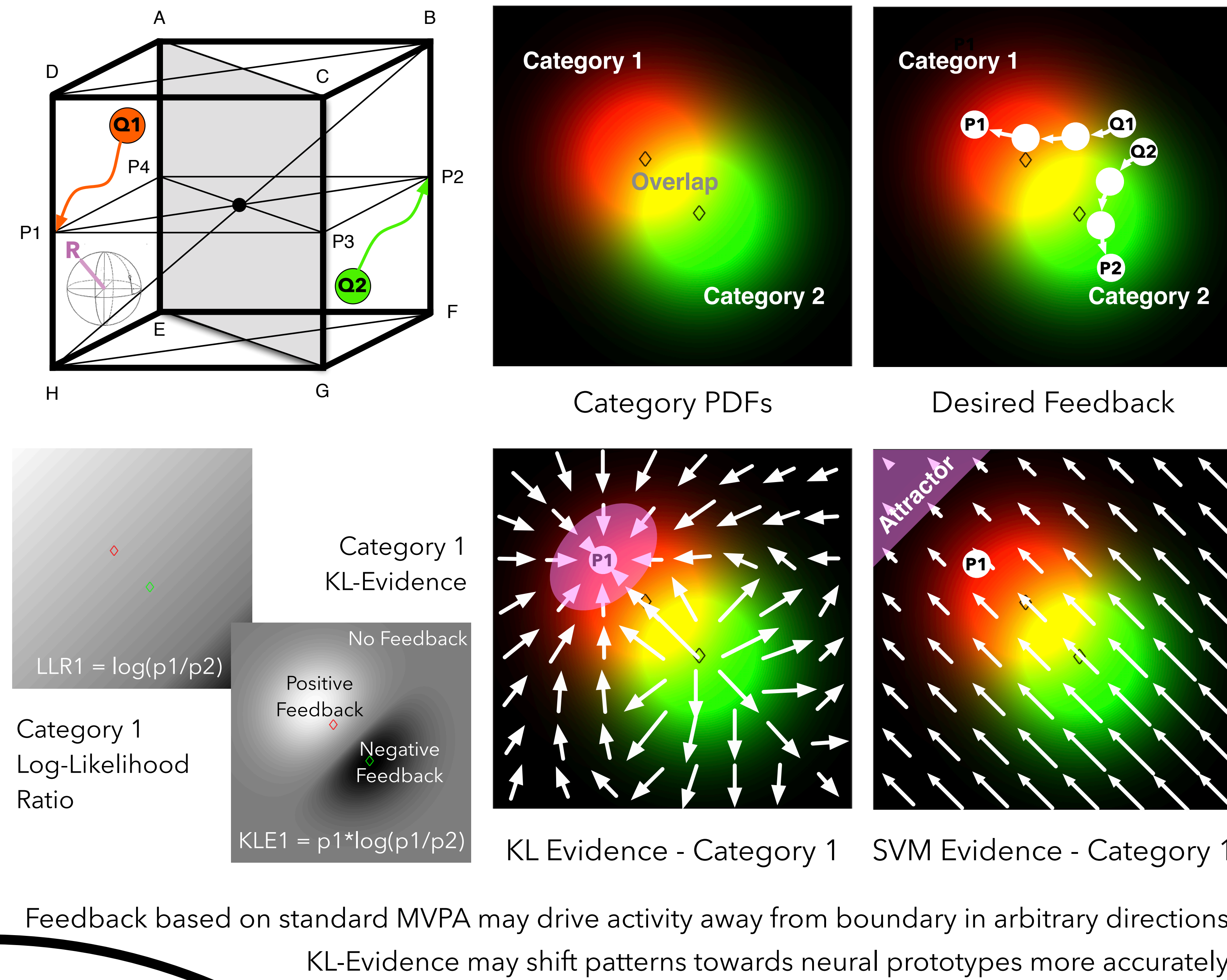Nicholas B. Turk-Browne[1,2], Jonathan D. Cohen[1]

[1] Princeton Neuroscience Institute & Psychology Department, Princeton University [2] Psychology Department, Yale University

mci@princeton.edu   vej@princeton.edu   knorman@princeton.edu   ntb@princeton.edu   jdc@princeton.edu

Abstract Multidimensional STIMULUS SPACE

real-world objects
features difficult to control
decades of experience
changes too transient?

artificial object space
full parametric control
no extant category bias
simple, but still 'objects'

0.55 Hz   1.11 Hz   4.94 Hz   3.39 Hz   1.54 Hz   3.18 Hz   0.57 Hz

A=13.1   A=18.8   A=24.6   A=30.3   A=36.1

7 distinct frequencies yield
7 dimensions of variation

parametrically vary shape
by changing multiple
dimensions' amplitudes

Category 2

Category 1

NEUROFEEDBACK: KL-Evidence Model & Training

if two shapes become more similar neurally, they may be perceived similarly
Drive neural activity for shapes near category boundary towards category prototypes

Category PDFs

Desired Feedback

Category 1
KL-Evidence

$LLR1 = \log(p1/p2)$

Category 1
Log-Likelihood
Ratio

No Feedback
Positive Feedback
Negative Feedback

$KLE1 = p1*\log(p1/p2)$

KL Evidence - Category 1

SVM Evidence - Category 1

Feedback based on standard MVPA may drive activity away from boundary in arbitrary directions
KL-Evidence may shift patterns towards neural prototypes more accurately

Shape oscillates with variable radius, centered randomly **Task**
Push neural pattern of shape towards prototype **Goal**
Make the shape stop oscillating! **Instructions**

Radius R **Neurofeedback Manipulation**
Positive feedback: decrease R
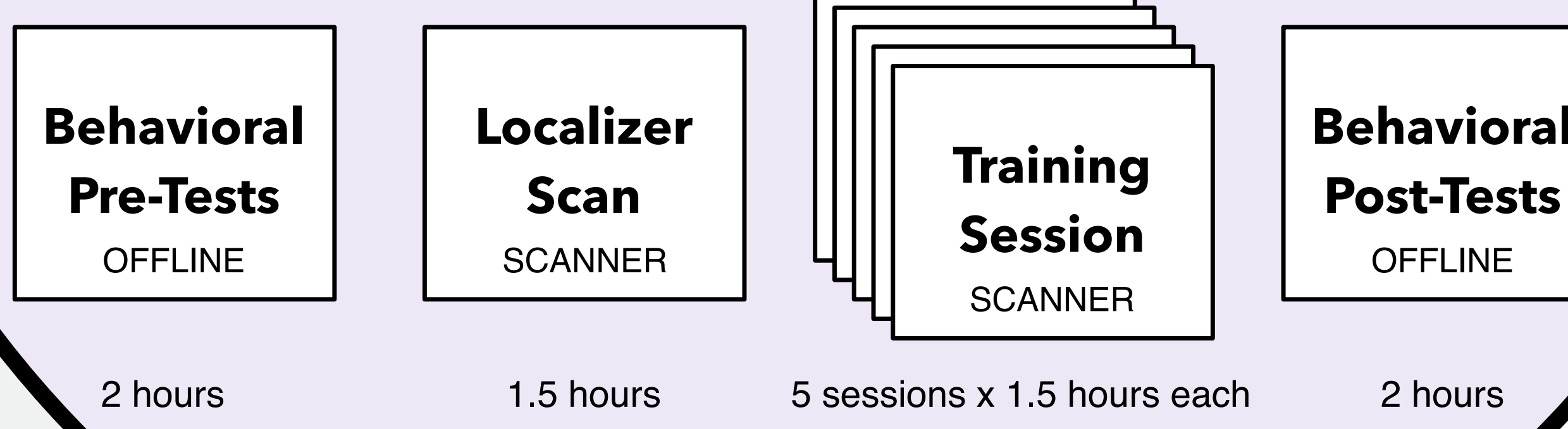Negative feedback: increase R

Hypothesis

Reinforcing differential neural activity patterns in
ventral temporal cortex for visually similar shapes will drive
apart their neural representations and reduce perceptual similarity

trial-level fast timescale   access neural pattern   change neural representation

**real-time fMRI neurofeedback
below threshold of awareness**

no explicit top-down learning signal

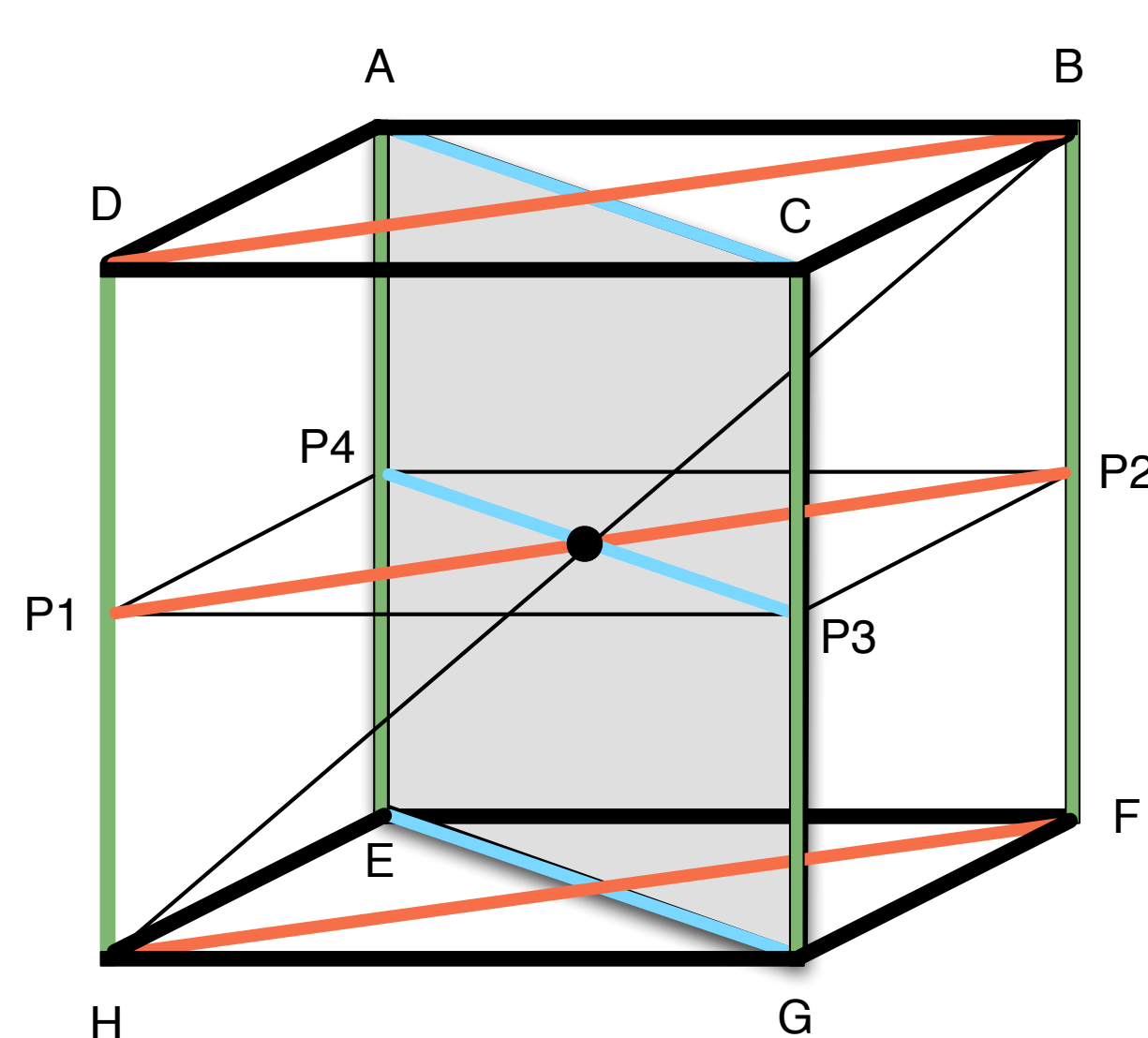| Behavioral Pre-Tests OFFLINE	Localizer Scan SCANNER	Training Session SCANNER	Behavioral Post-Tests OFFLINE
2 hours	1.5 hours	5 sessions x 1.5 hours each	2 hours

Experimental Design

Inducing and Measuring PERCEPTUAL CHANGES

Categorical Perception of Space
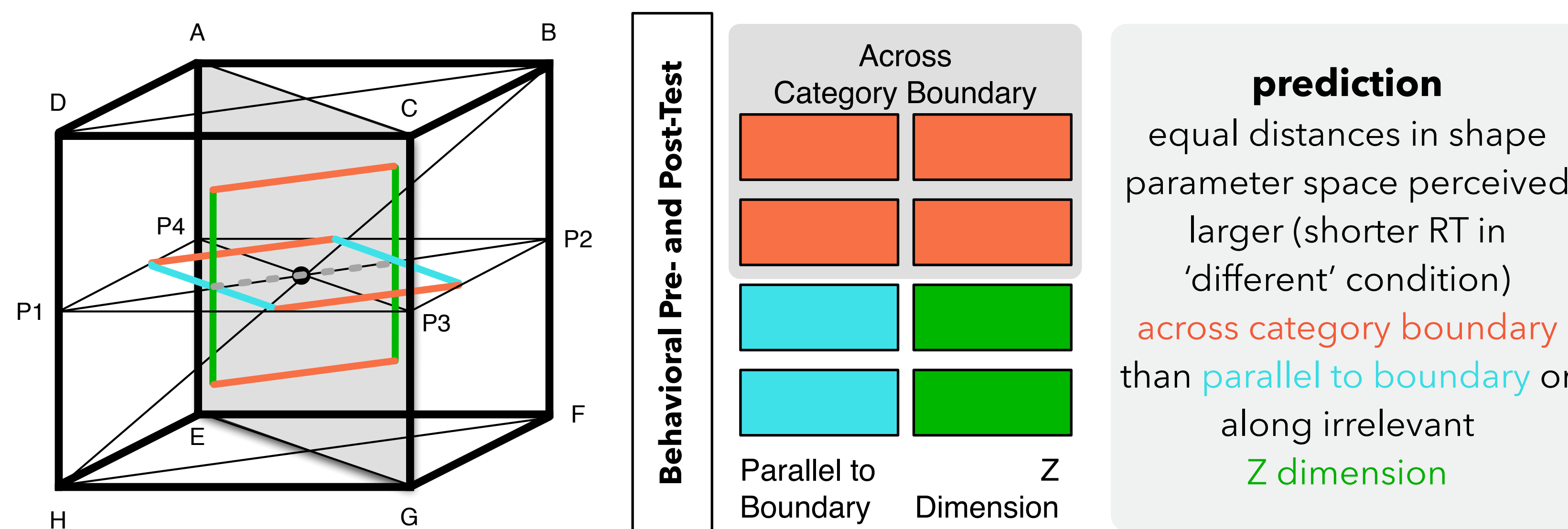2AFC between line endpoints | 10 lines

prediction
sharper category
boundaries (steeper
psychometric function)
for feedback dimension
versus untrained dimensions

center of space perceived equidistant
from endpoints across all dimensions

Amazon Mechanical Turk, n = 24-30 per line

Proportion Chosen as Category 'C'
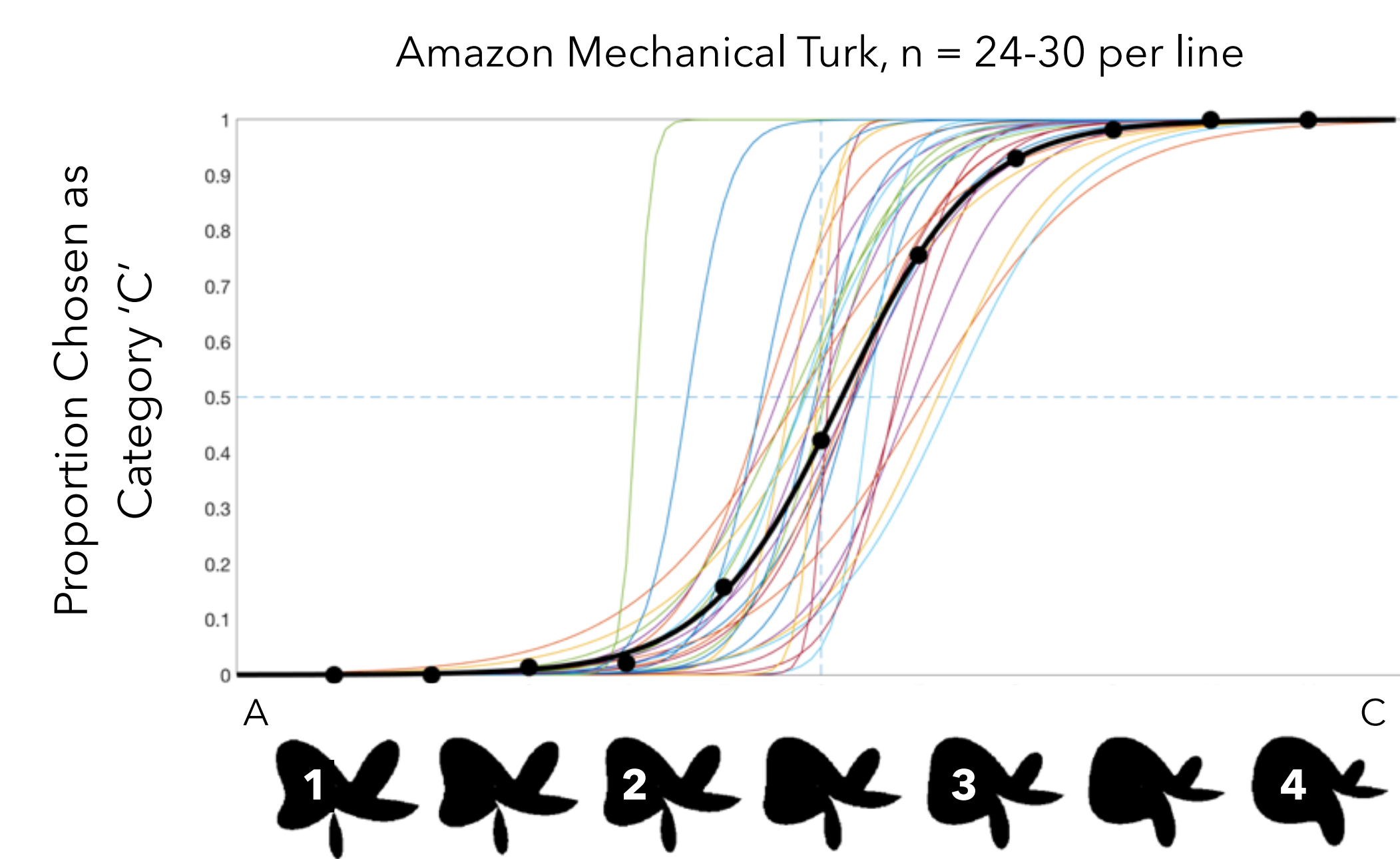
Across Boundary	Along Boundary
DB	AC
P1P2	P4P3
HF	EG
AE	BF
CG	DH

Behavioral Pre- and Post-Test

Z Dimension

Perceived Distances Across Dimensions
RTs for same-different task

prediction
equal distances in shape
parameter space perceived
larger (shorter RT in
'different' condition)
across category boundary
than parallel to boundary or
along irrelevant
Z dimension

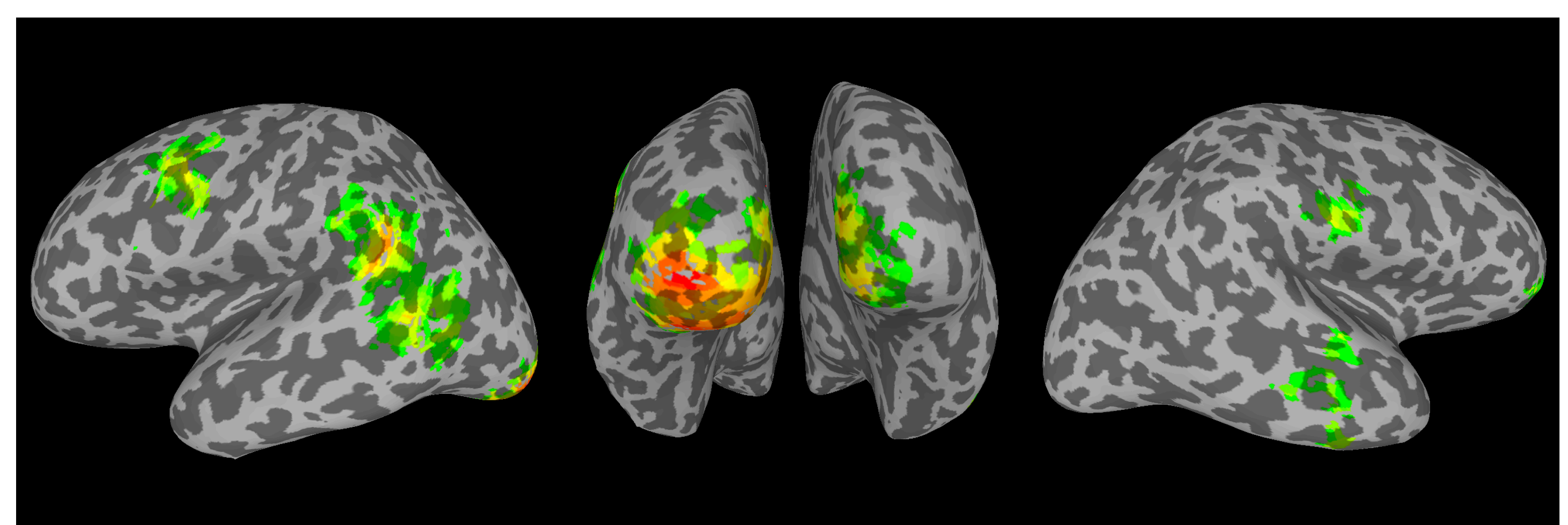Across Category Boundary

Parallel to Boundary

Z Dimension

Behavioral Pre- and Post-Test

NEURAL REPRESENTATION of Shape Space

Parametric Shape Localizer
average 6 lines | n=8 | anatomical ROI

Lateral Occipital RSA   **r = 0.96**   Ideal Similarity Matrix
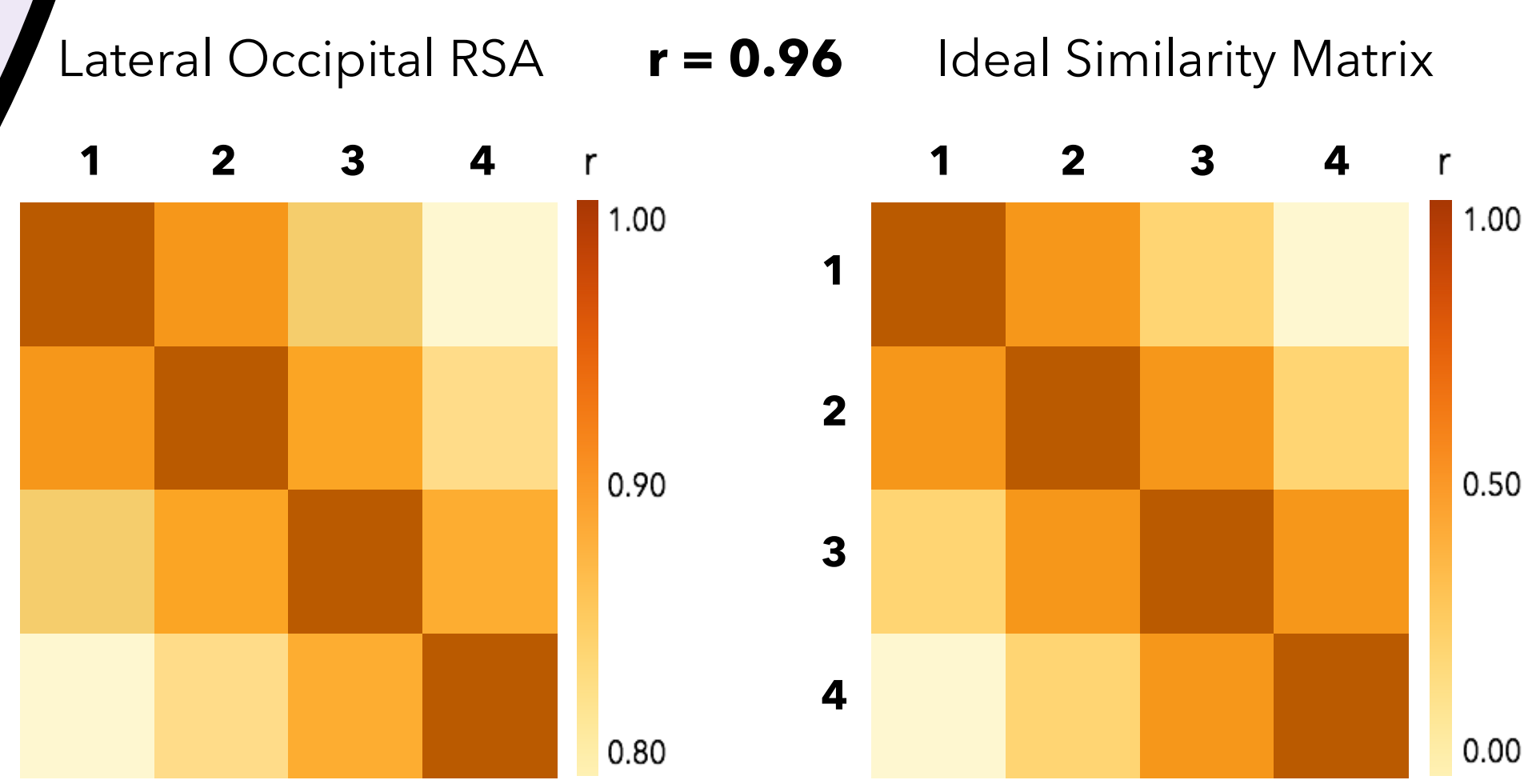
Feedback Training ROI Searchlight for parametric regions: r > 0.50

Category Prototype Classification: P1 vs. P2   n=1 | 10 runs

Early Visual Cortex   Lateral Occipital ROI   Searchlight ROI

Decoding Accuracy

SVM   KLE