Representational Dissimilarity Analysis as a Tool for Neural Network Model Search

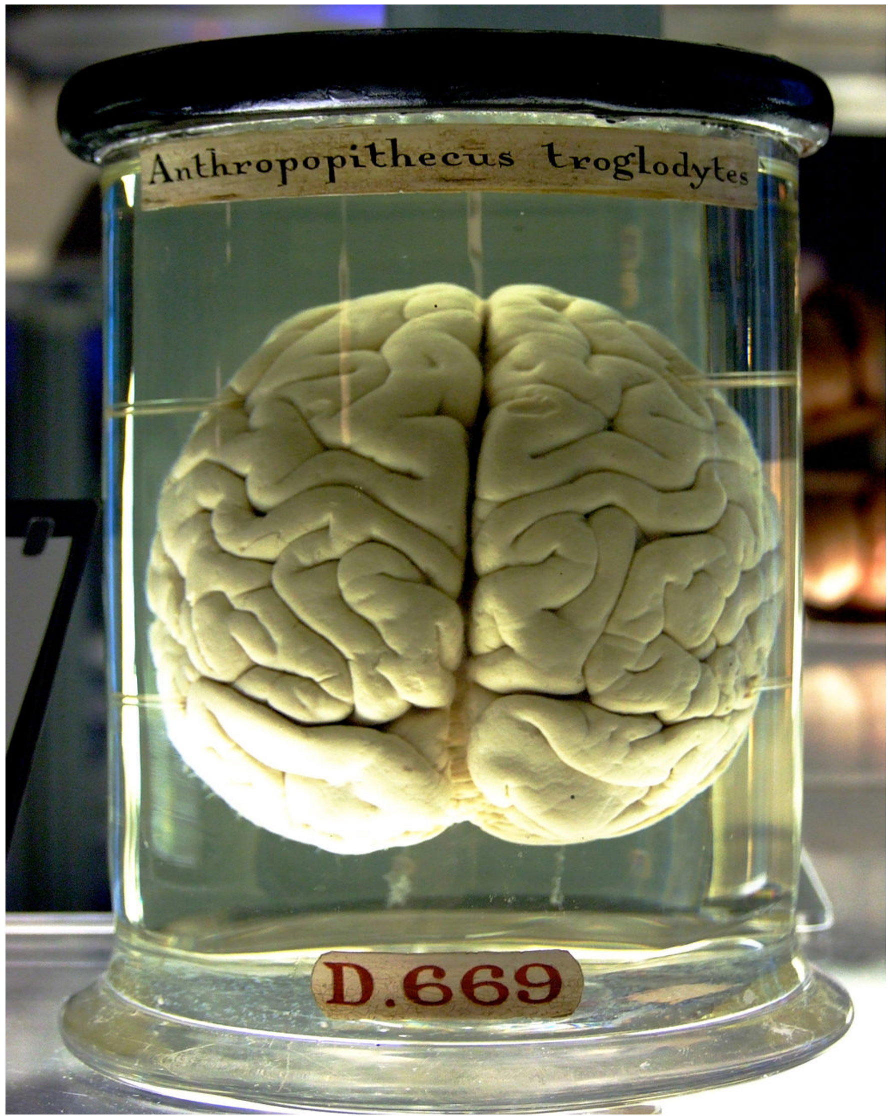
Walter J. Scheirer

Computer Vision Research Laboratory
Department of Computer Science and Engineering

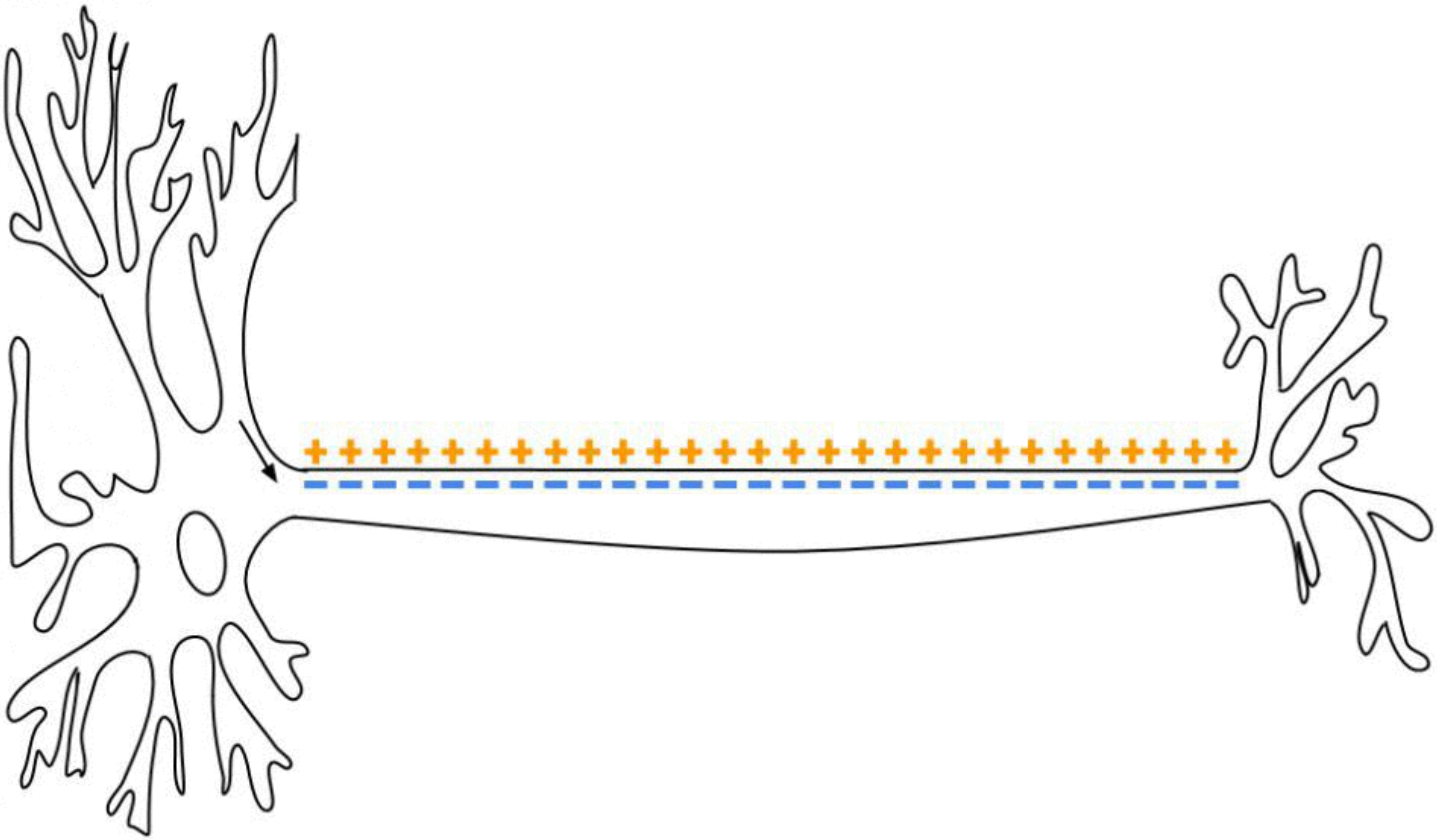
UNIVERSITY OF NOTRE DAME

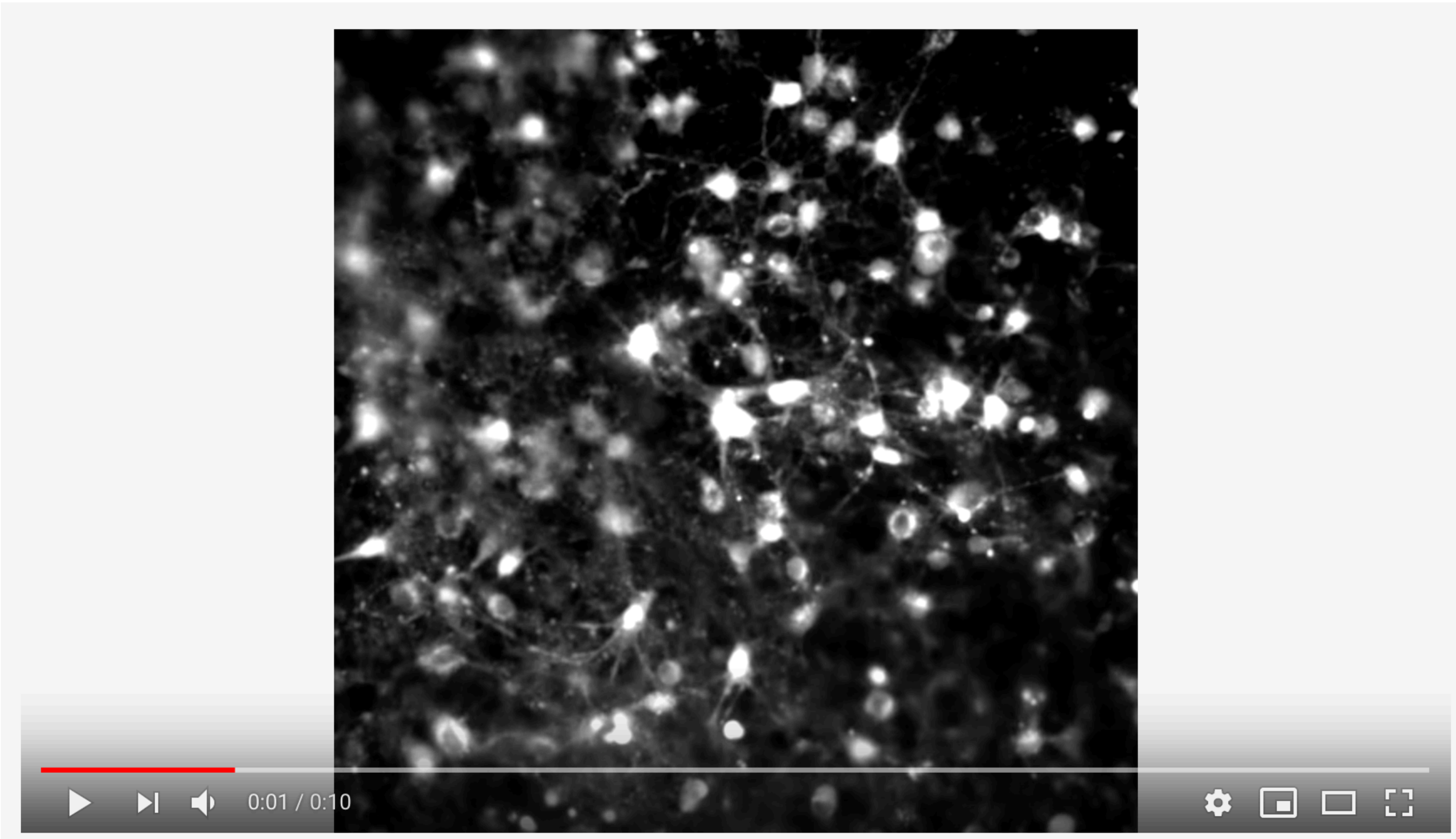
COMPUTER VISION @ND

A model of vision that works:



Chimp Brain in a jar

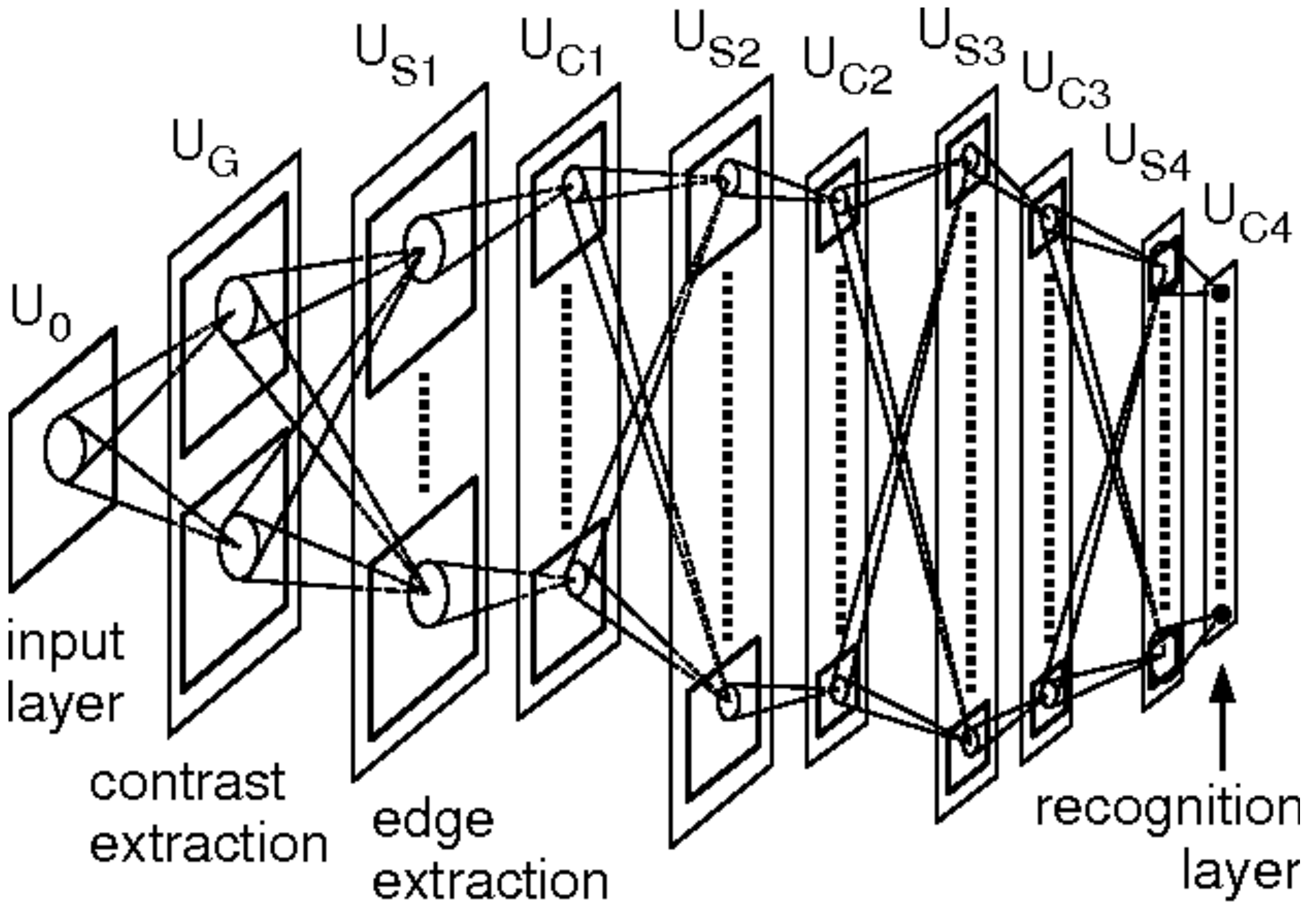




https://www.youtube.com/watch?v=yy994HpFudc

Hypothesis: networks exhibiting brain-like activation behavior will demonstrate brain-like characteristics, e.g., stronger generalization capabilities.

Fukushima 1979: Neocognitron



Is there any correspondence between activity measured in the brain and activity measured in artificial neural networks?

Monkey performing an object recognition task

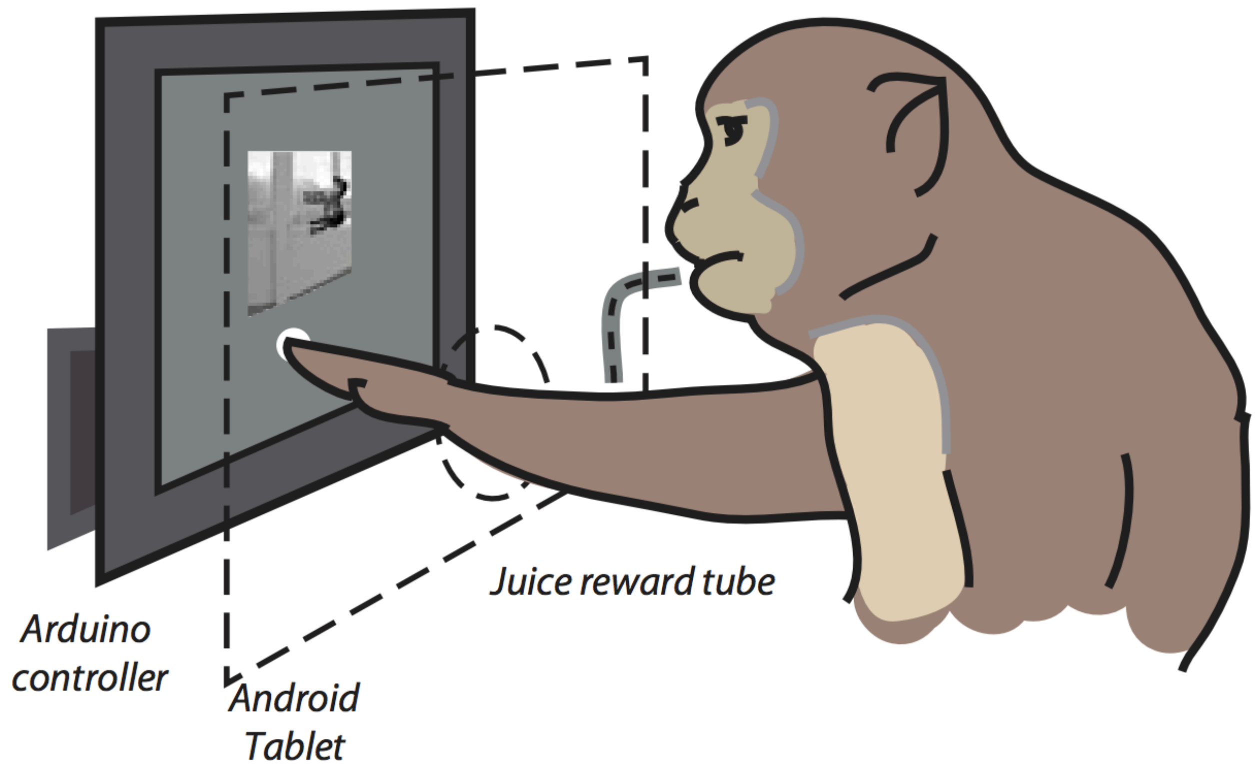
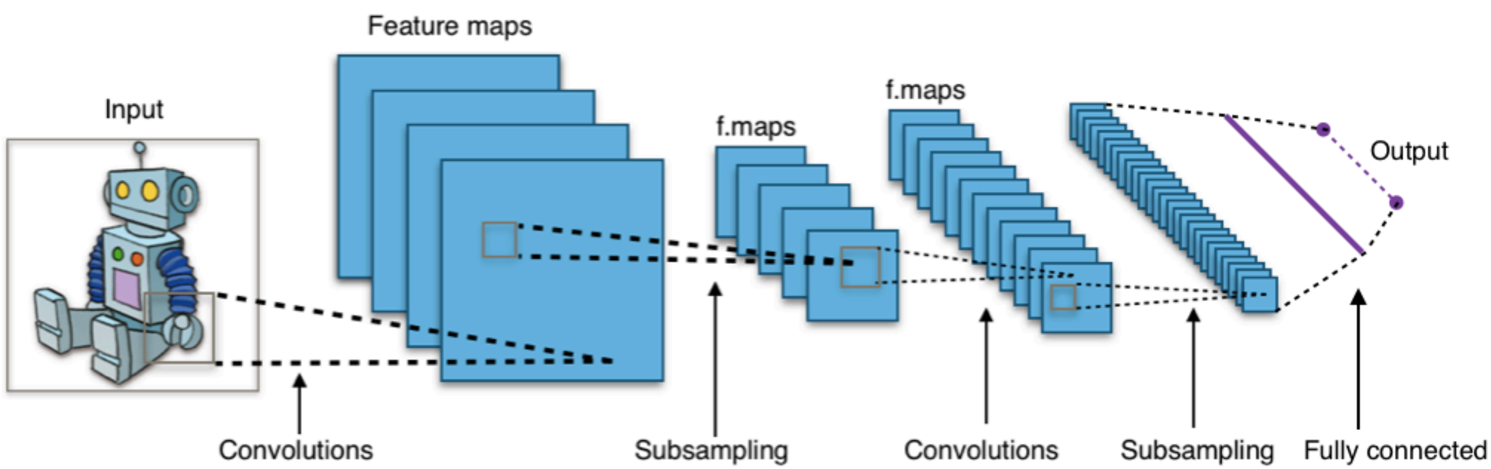


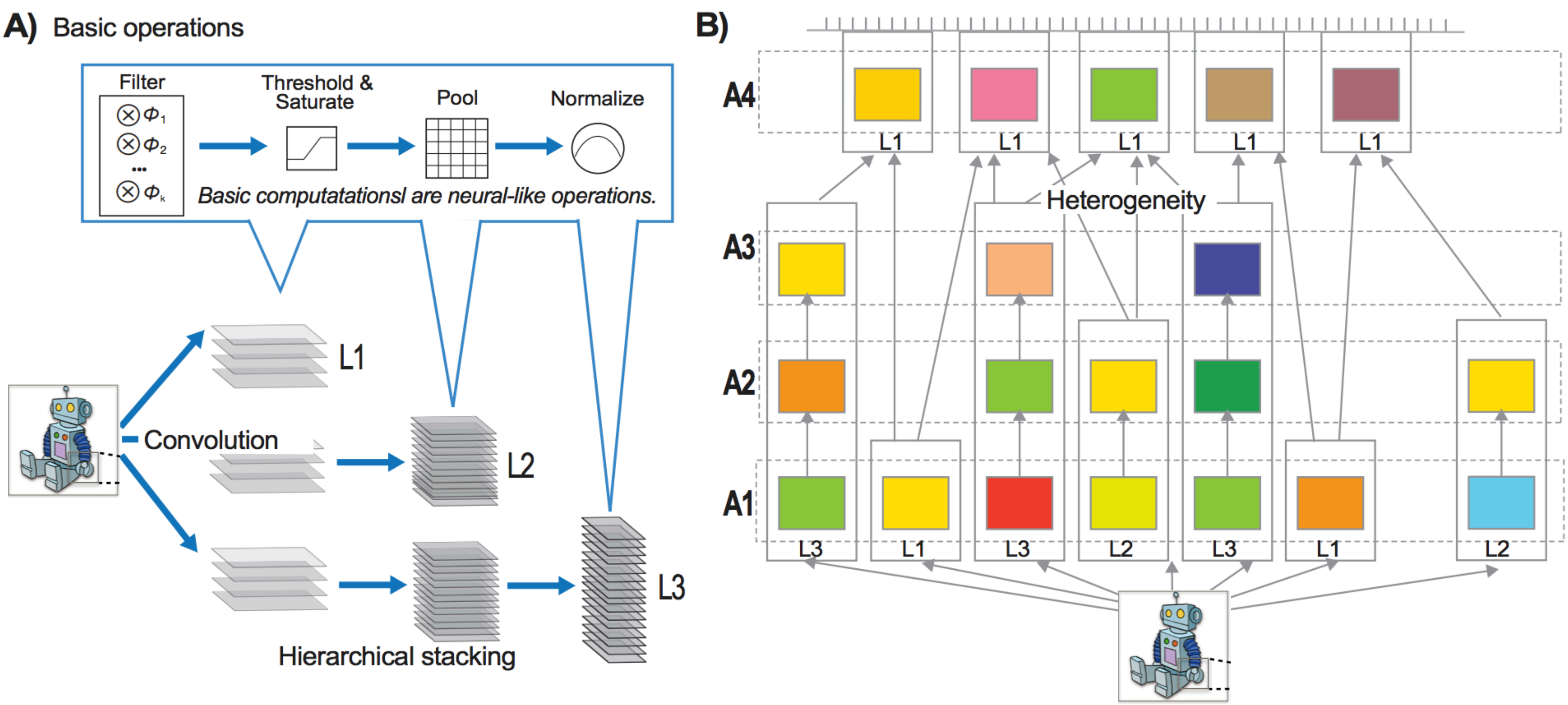
Image adapted from: Rajalingham et al. JNeurosci 2018

CNN for Object Recognition



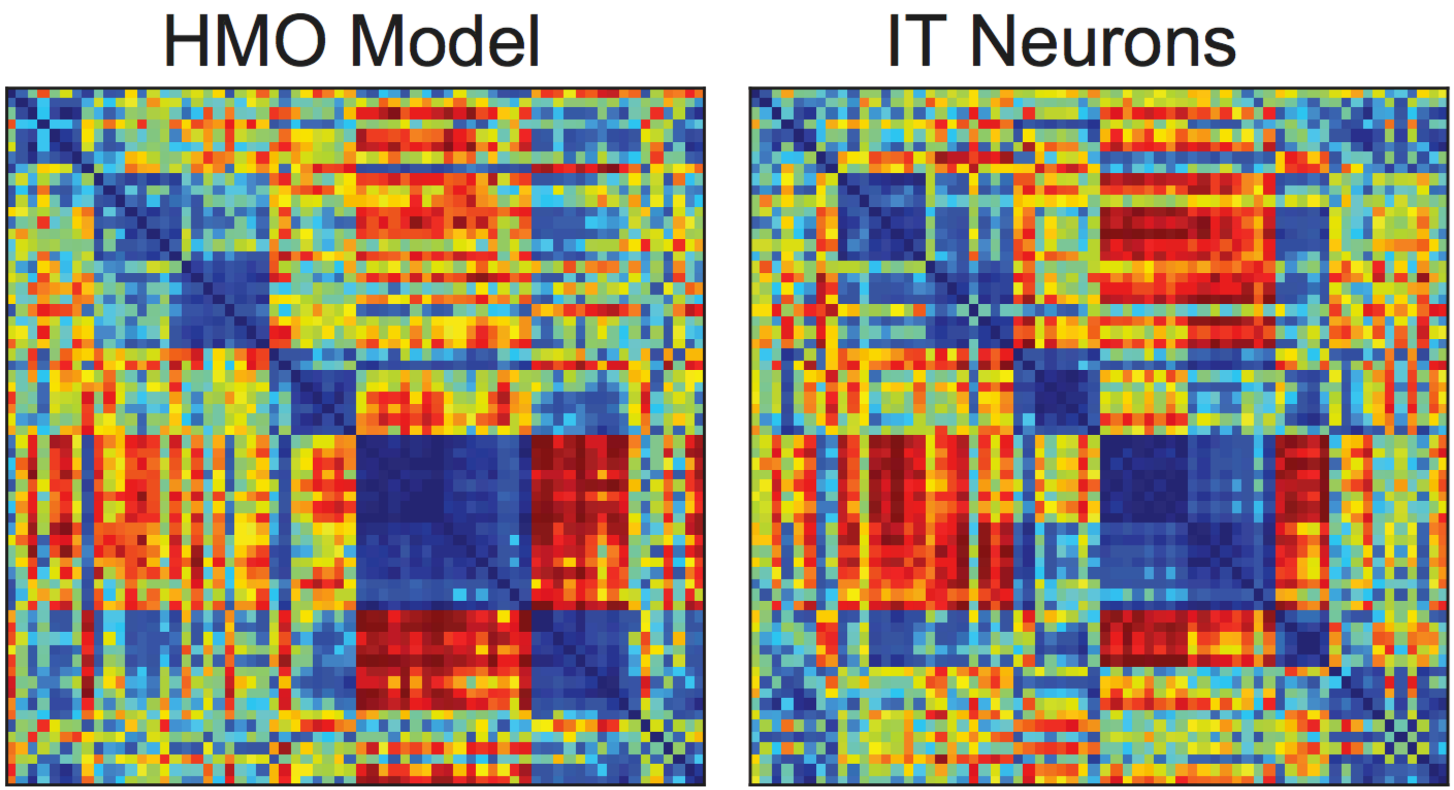
Typical CNN architecture

Heterogeneous Hierarchical CNN



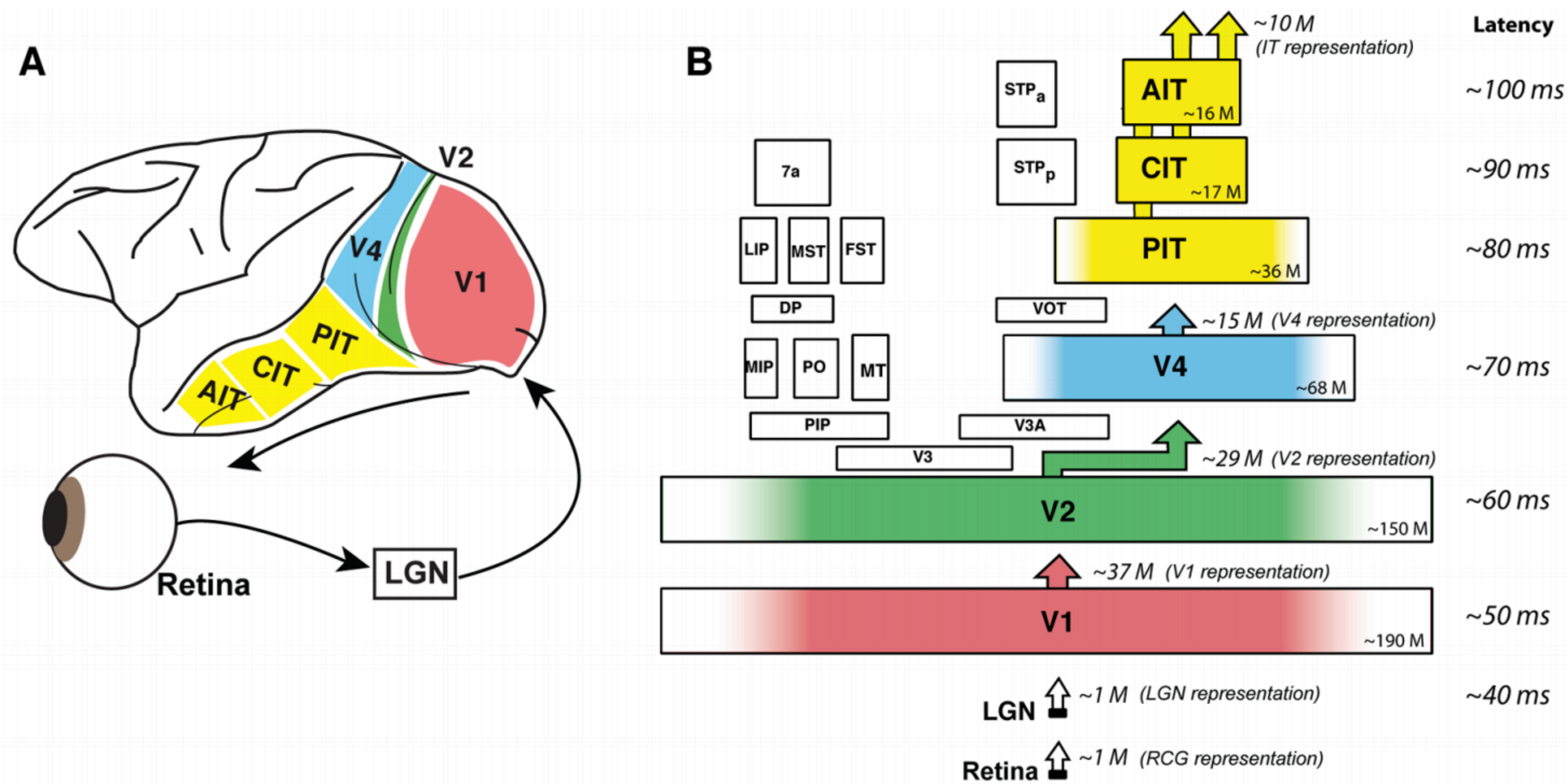
Yamins et al. NeurIPS 2013

Population Responses: Model vs. Brain



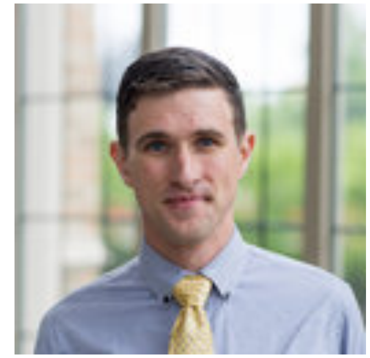
Yamins et al. NeurIPS 2013

Where in the brain is area IT?



DiCarlo et al. Neuron 2012

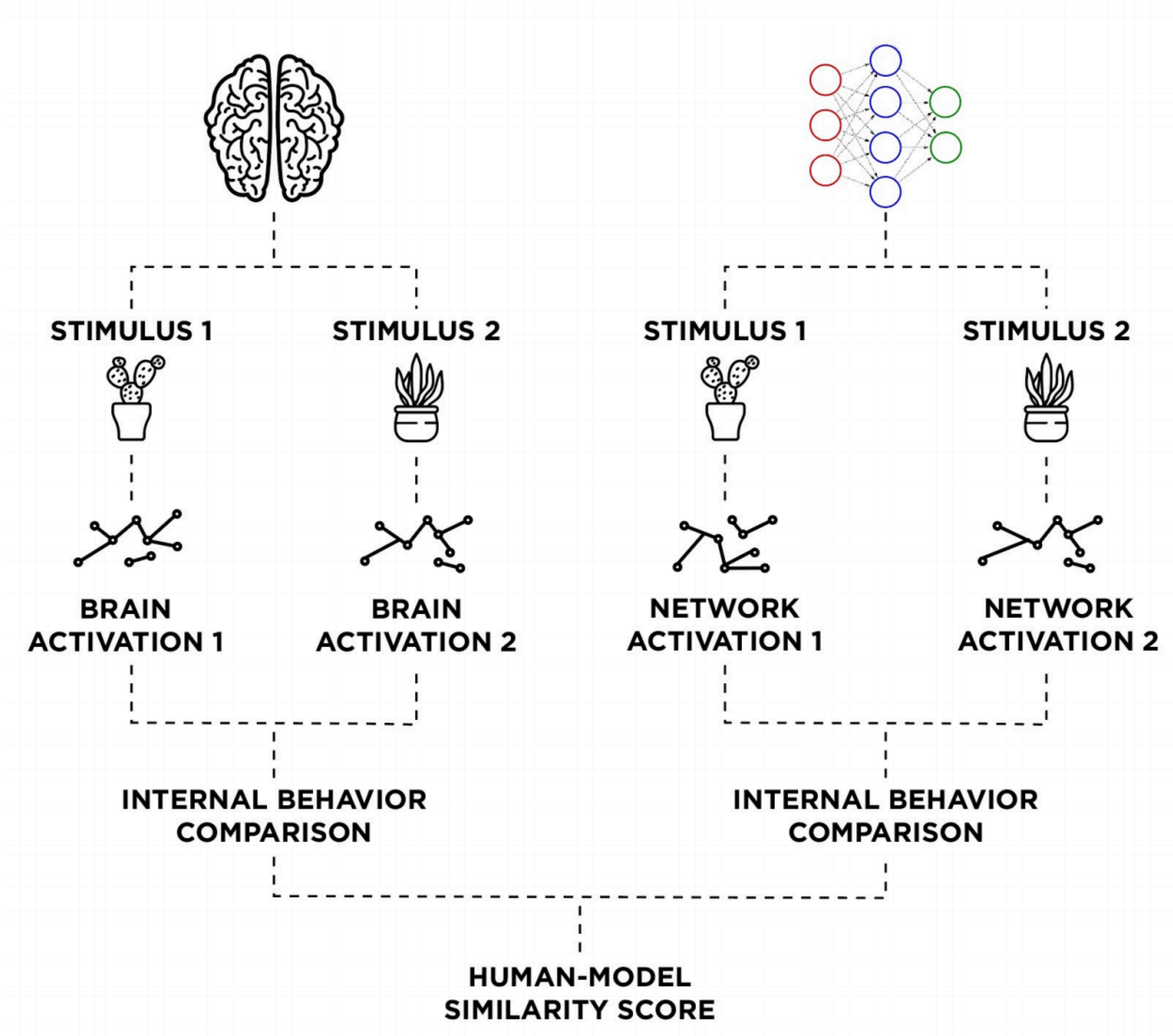
How do we compare the activity in brains with the activity in artificial neural networks?



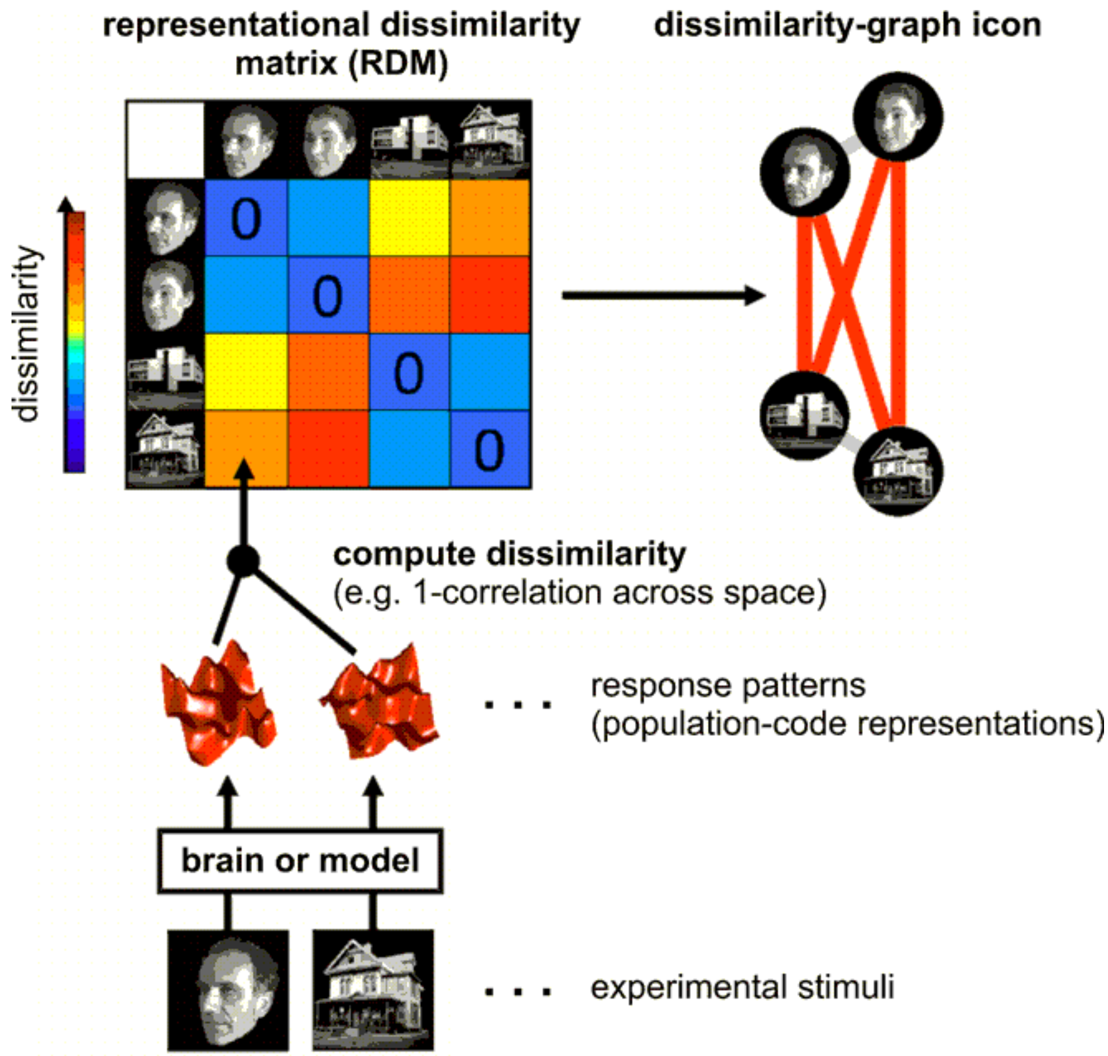
Nathaniel Blanchard

A Neurobiological Evaluation Metric for Neural Network Model Search

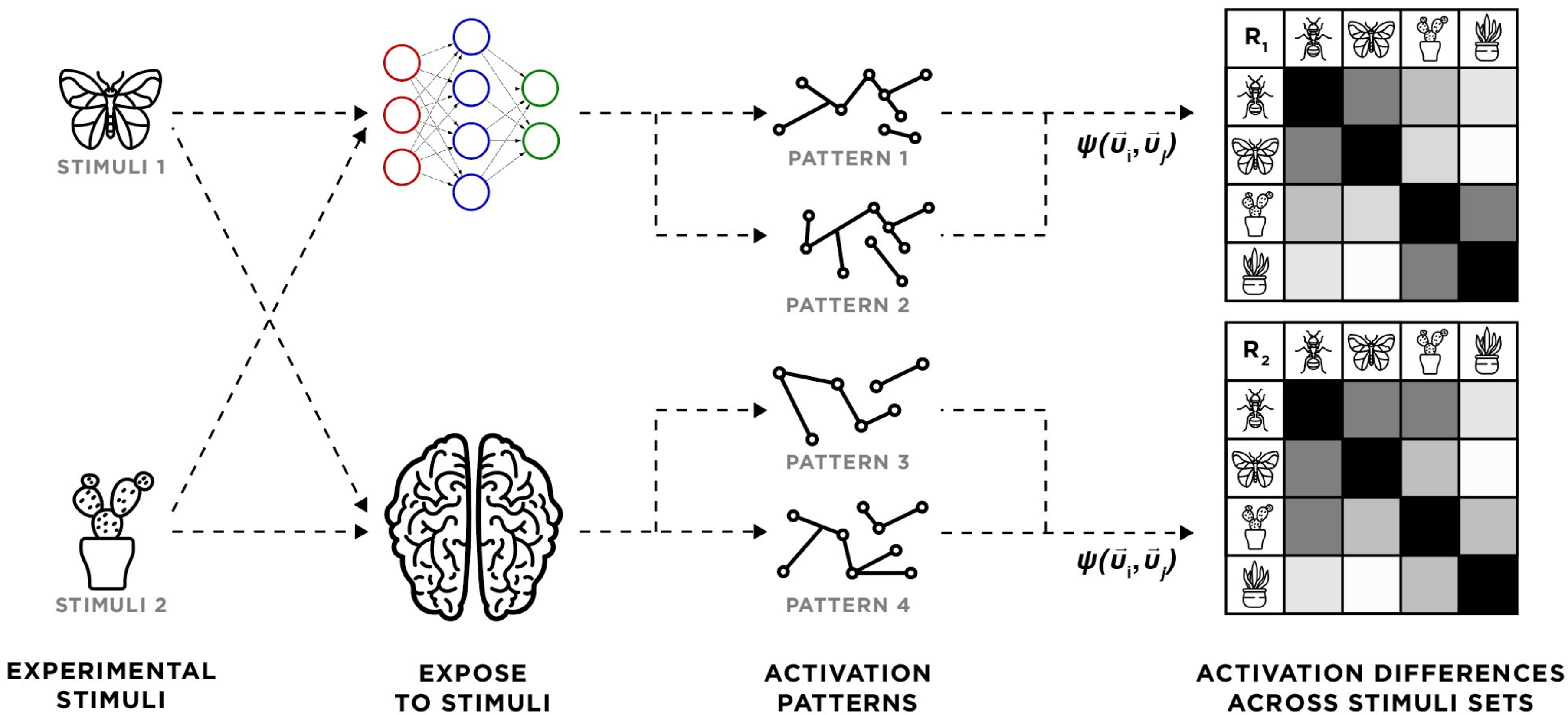
IEEE/CVF CVPR, 2019


STIMULUS 1
STIMULUS 2
STIMULUS 1
STIMULUS 2
BRAIN ACTIVATION 1
BRAIN ACTIVATION 2
NETWORK ACTIVATION 1
NETWORK ACTIVATION 2
INTERNAL BEHAVIOR COMPARISON
INTERNAL BEHAVIOR COMPARISON
HUMAN-MODEL SIMILARITY SCORE

Kriegeskorte et al.: Representational Similarity Analysis



N. Kriegeskorte, M. Mur and P. A. Bandettini, “Representational Similarity Analysis – Connecting the Branches of Systems Neuroscience,” Frontiers in Systems Neuroscience, 2008



STIMULI 1
STIMULI 2
PATTERN 1
PATTERN 2
PATTERN 3
PATTERN 4
$\psi(\vec{u}_i, \vec{u}_j)$
$\psi(\vec{u}_i, \vec{u}_j)$
R_1
R_2
EXPERIMENTAL STIMULI
EXPOSE TO STIMULI
ACTIVATION PATTERNS
ACTIVATION DIFFERENCES ACROSS STIMULI SETS

RDM Step 1: Data Representation

Given a single feature f and a single stimulus s, $v = f(s)$, where v is the value of feature f in response to s. Likewise, the vector

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}^T = \begin{bmatrix} f_1(s) \\ f_2(s) \\ \vdots \\ f_n(s) \end{bmatrix}^T$$

can represent the feature values of a collection of n features, $f_1, f_2, \ldots, f_n$, in response to s.

RDM Step 1: Data Representation

If one expands the representation of s to a set of m stimuli $S = s_1, s_2, \ldots, s_m$, the natural extension of $\overrightarrow{v}$ is the set of feature value collections $V = \overrightarrow{v_1}, \overrightarrow{v_2}, \ldots, \overrightarrow{v_m}$, in which $s_i \in S$ is paired with $\overrightarrow{v_i} \in V$ for each $i = 1, 2, ..., m$.

RDM Step 2: Dissimilarity

Define the dissimilarity score between any two $\overrightarrow{v_i} \in V$ and $\overrightarrow{v_j} \in V$:

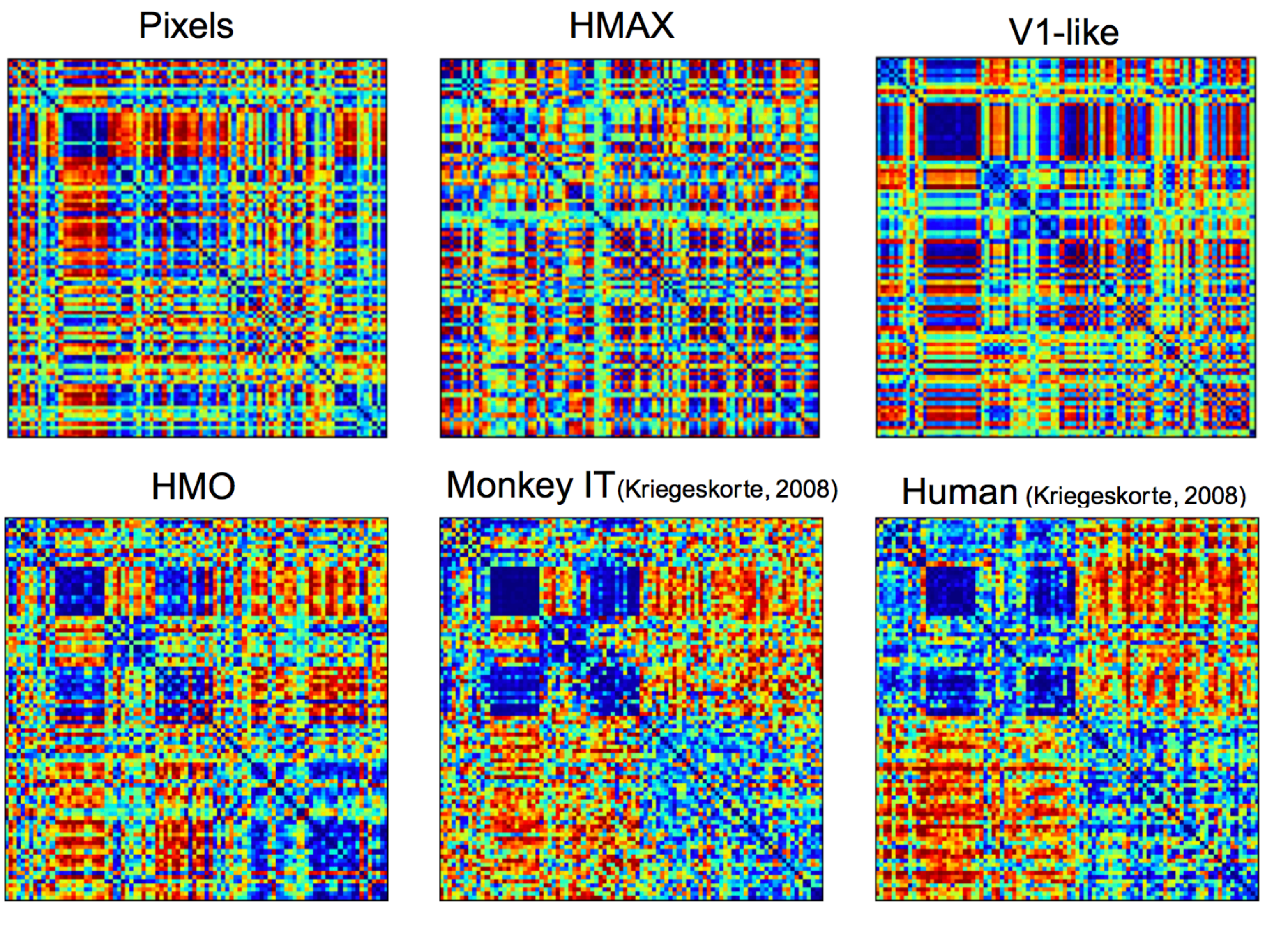
$$\psi(\vec{v}_i, \vec{v}_j) := 1 - \frac{(\vec{v}_i - \bar{v}_i) \cdot (\vec{v}_j - \bar{v}_j)}{\|\vec{v}_i - \bar{v}_i\|_2 \|\vec{v}_j - \bar{v}_j\|_2}$$

RDM Step 3: Construct Matrix

An RDM R may then be constructed from S, V , and ψ as:

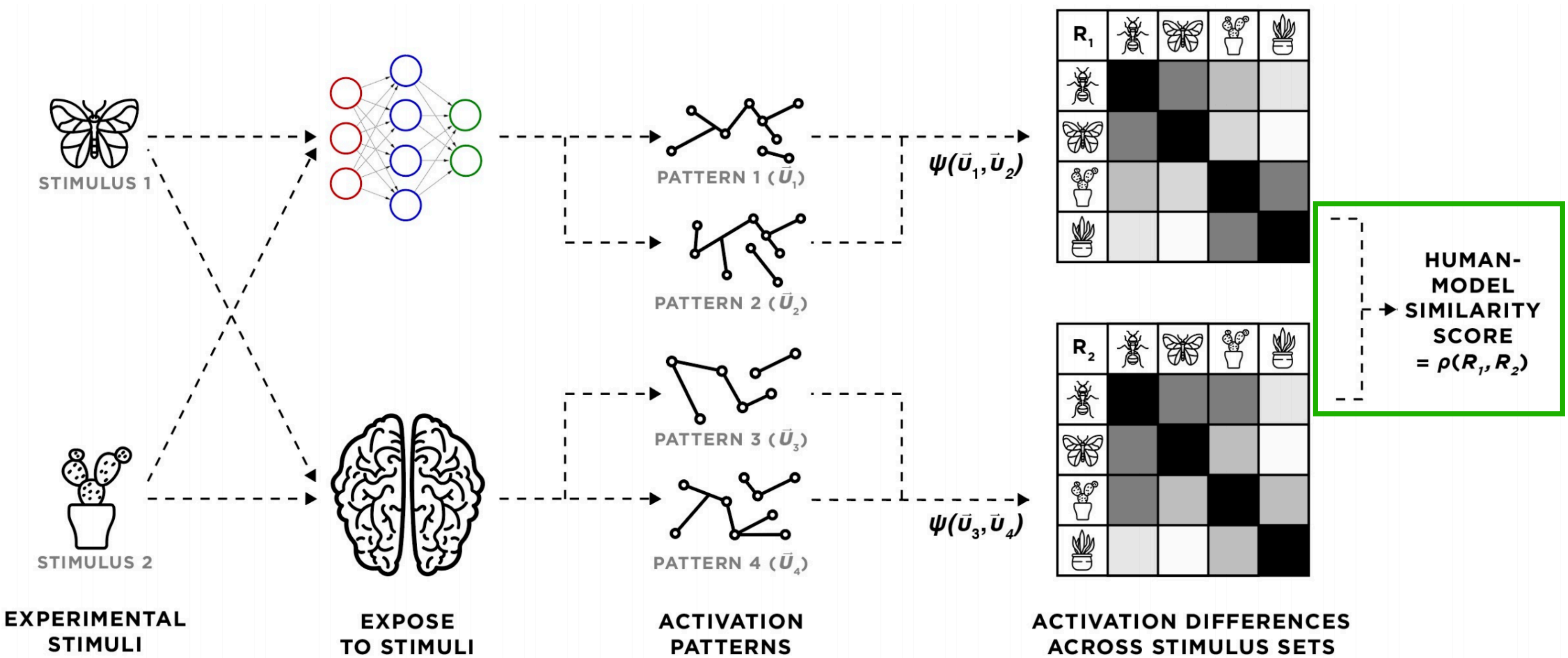
$$R = \begin{bmatrix} \psi(\vec{v}_1, \vec{v}_2) & \psi(\vec{v}_1, \vec{v}_3) & \cdots & \psi(\vec{v}_1, \vec{v}_m) \\ & \psi(\vec{v}_2, \vec{v}_3) & \cdots & \psi(\vec{v}_2, \vec{v}_m) \\ & & \ddots & \vdots \\ & & & \psi(\vec{v}_{m-1}, \vec{v}_m) \\ & & & \end{bmatrix}$$

Works well for assessing biological fidelity:



Yamins et al. NeurIPS 2013

Computing Human-Model Similarity



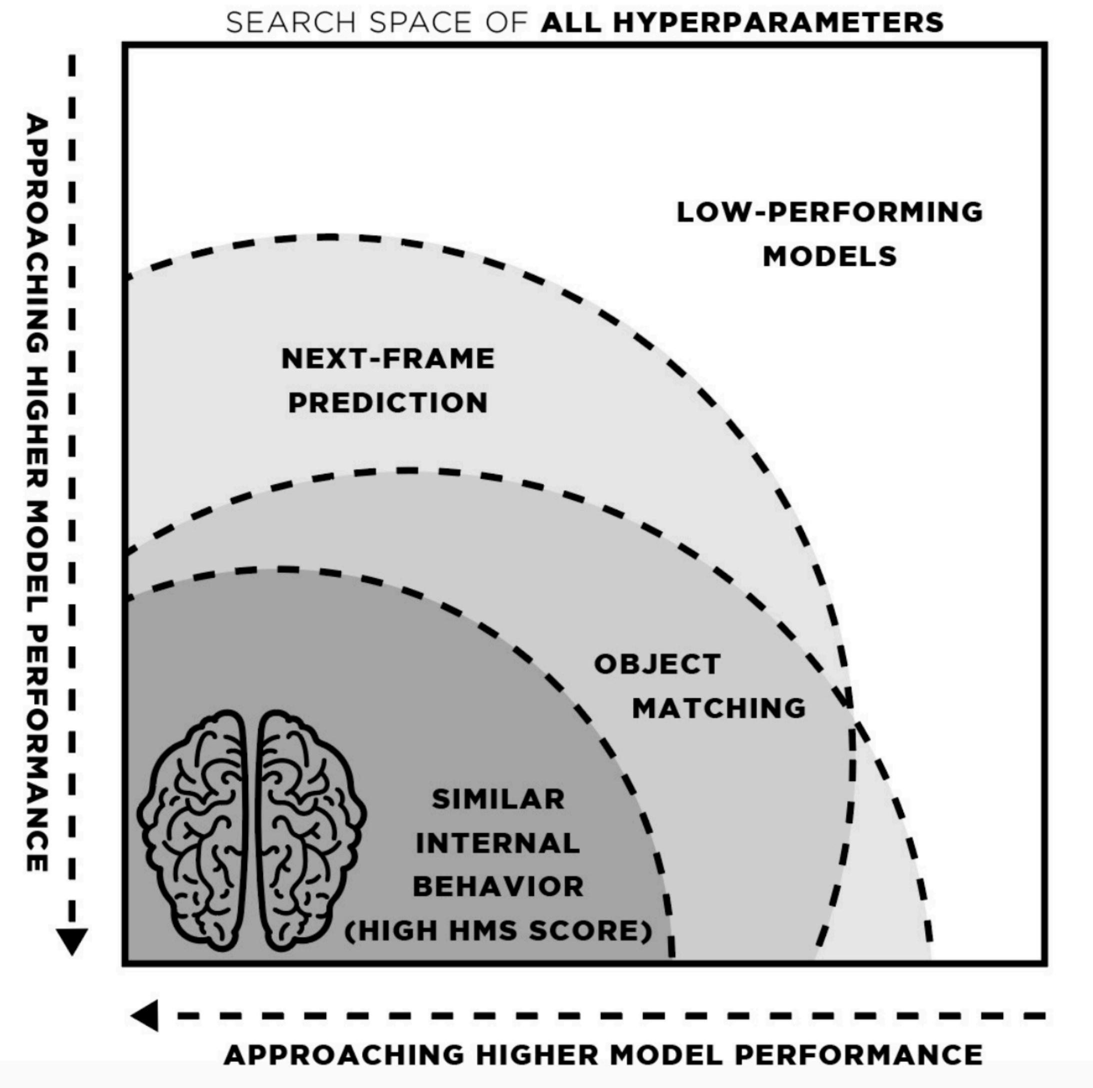
Human Model Similarity Score

RDM 1 RDM 2

$$HMS = \rho(\hat{R}_1, \hat{R}_2)$$

Spearman's Rank Correlation

AI as a search problem



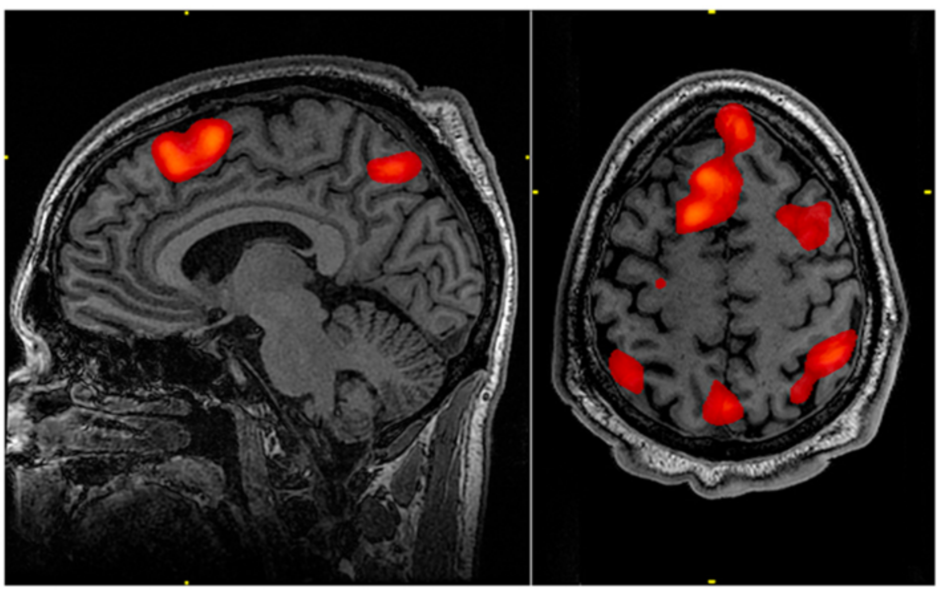
Experiments

fMRI

A direct way to measure human brain activity

Non-invasive experimentation with humans

Uses blood flow as a proxy for neuronal activations



Spatial resolution good enough to identify Brodmann areas

fMRI Data

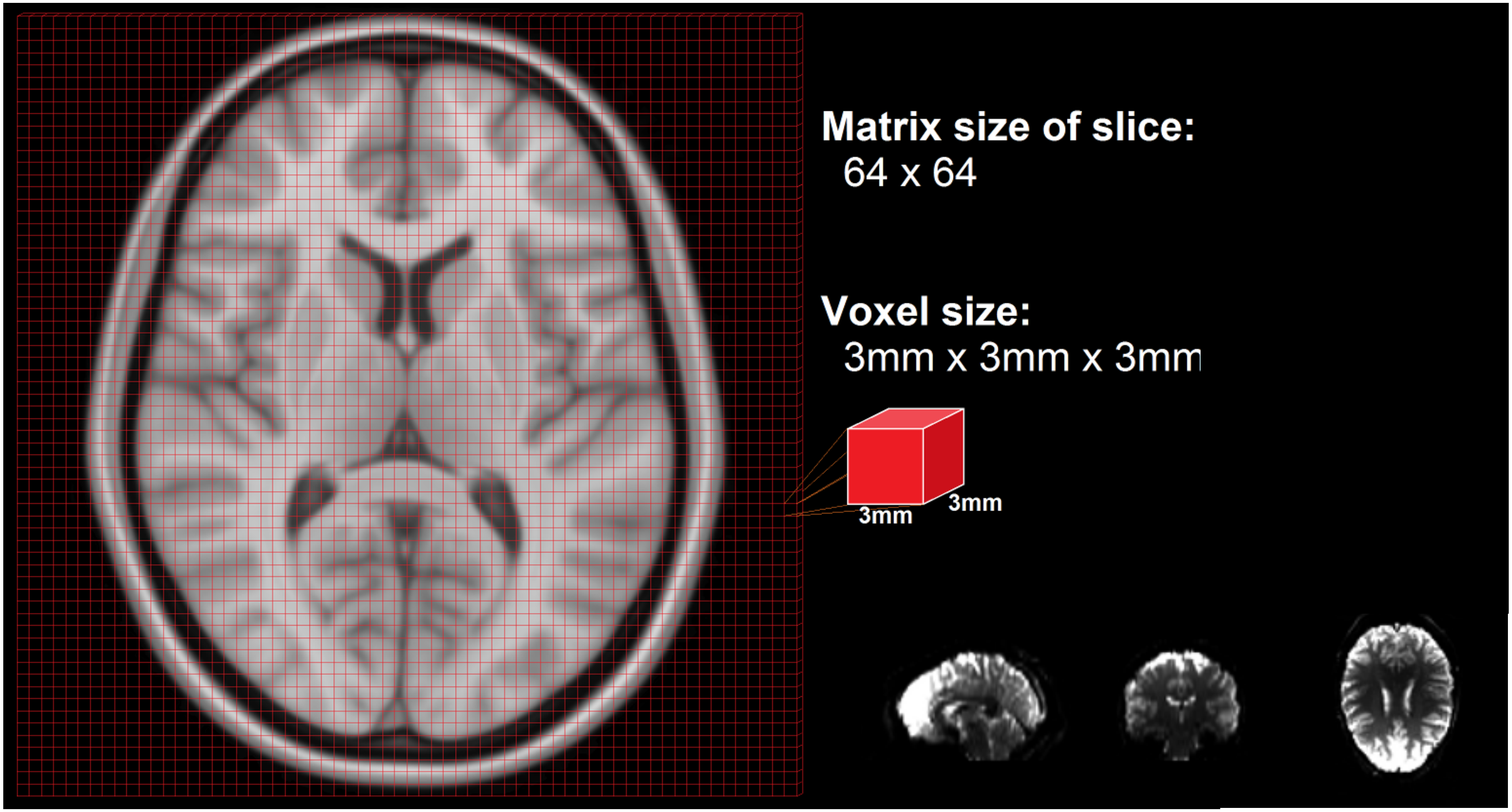


Image adapted from: https://miykael.github.io/nipype-beginner-s-guide/neuroimaging.html

fMRI Experimental Setup

Data collected by the Kriegeskorte lab at the University of Cambridge*

Eight RDMs were constructed from fMRI recordings of four subjects over two sessions in response to 92 random stimuli

Recordings were from measurements of $1.95 \times 1.95 \times 2\text{mm}^3$ within an occipitotemporal measurement slab (5cm thick).

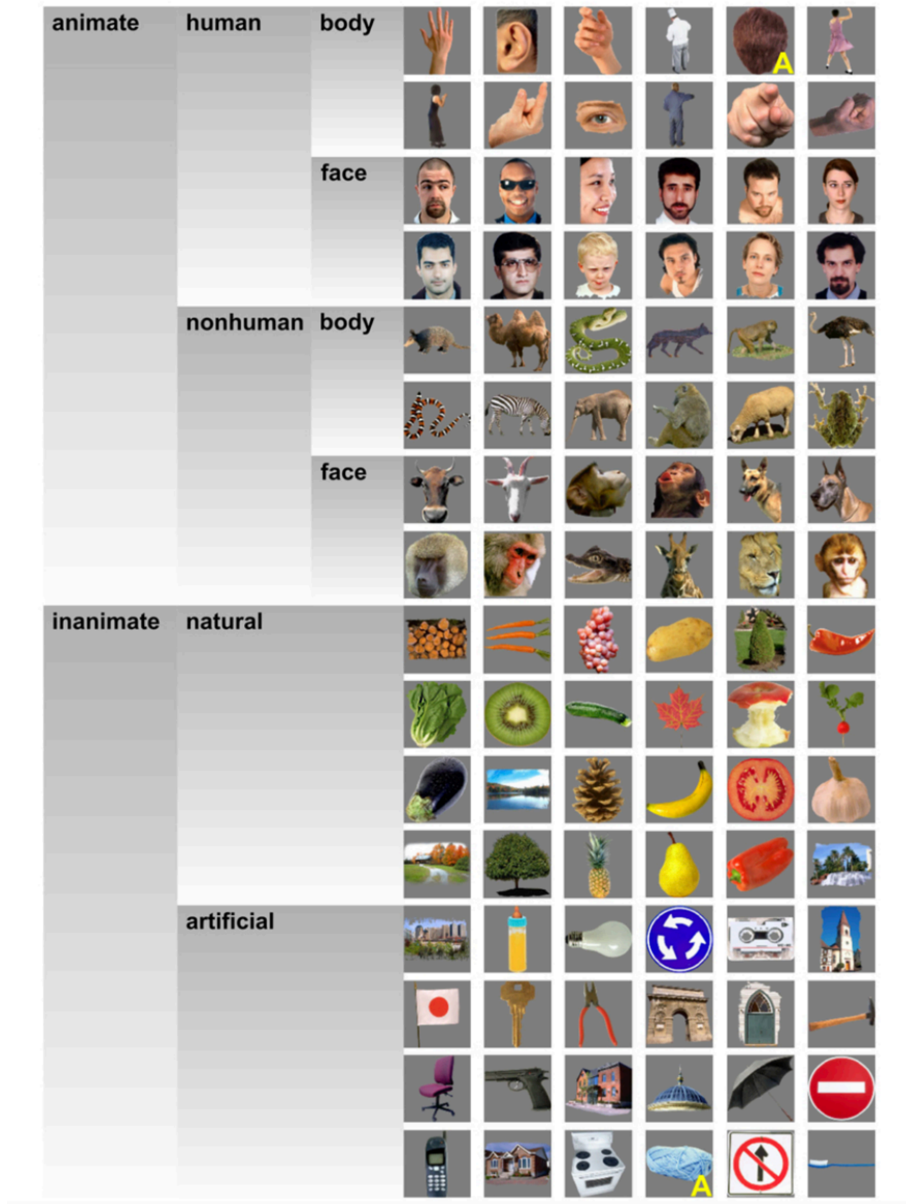
Each stimulus was displayed for 300 milliseconds, every 3700 milliseconds, with four seconds between stimuli.

Subject RDMs were averaged together into a mean human brain RDM, which reduced noise.

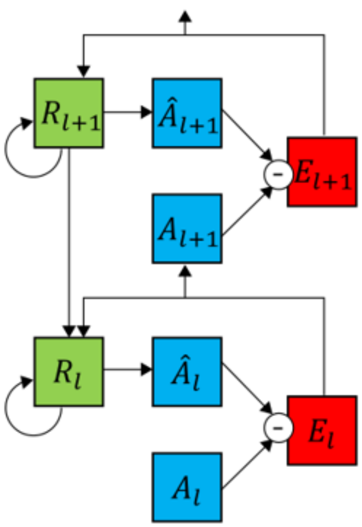
01 Siemens MAGNETOM Trio (cc) BY-SA 2.0 Image Editor

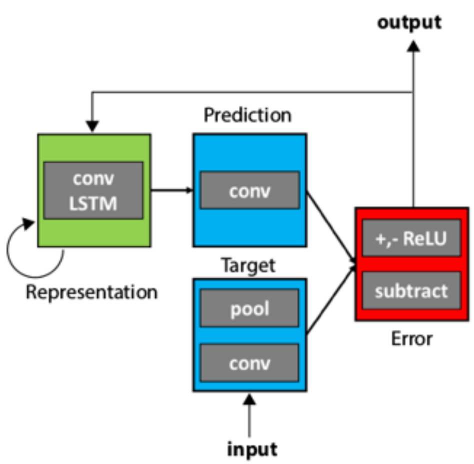
H. Nili, C. Wingfield, A. Walther, L. Su, W. Marslen-Wilson, and N. Kriegeskorte. A toolbox for representational similarity analysis. PLoS Computational Biology, 10(4):e1003553, 2014.

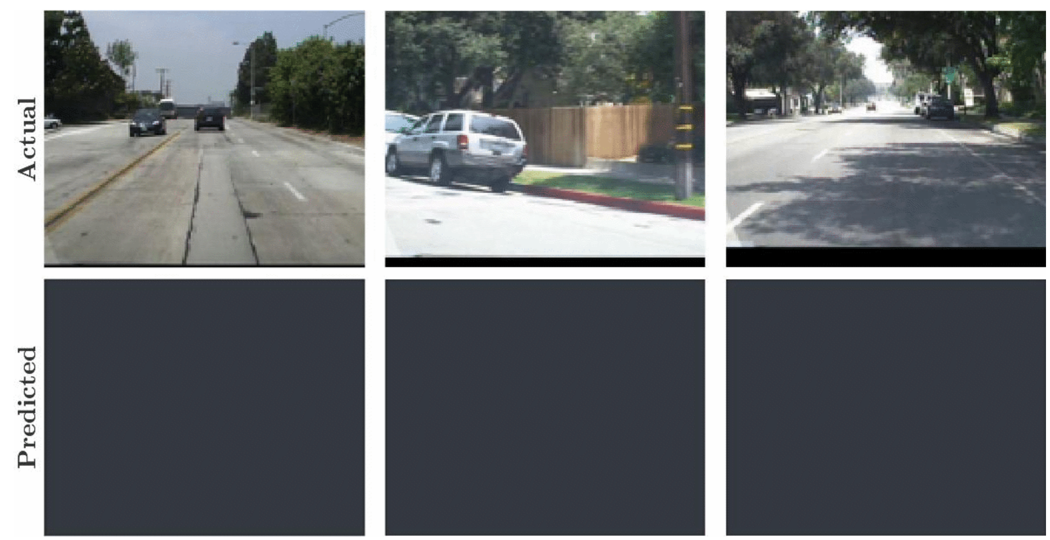
fMRI Stimuli Set



Architecture: PredNet

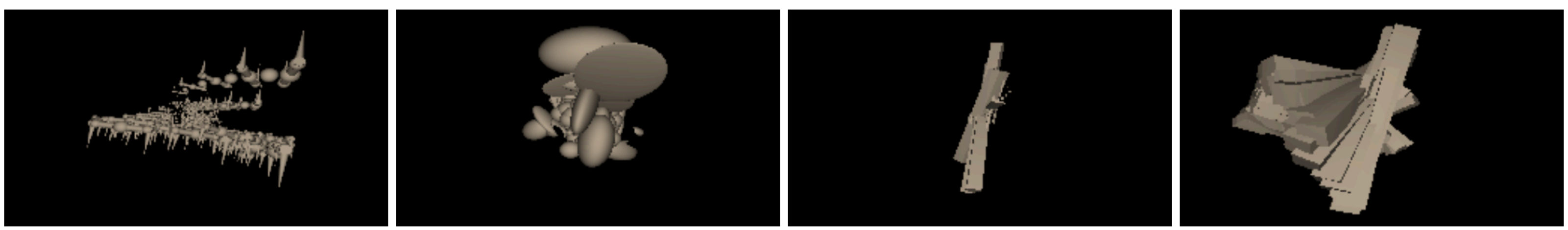






W. Lotter, G. Kreiman, and D. D. Cox, "Deep Predictive Coding Networks for Video Prediction and Unsupervised Learning," ICLR, 2017.

Predictive Coding Network Performance: 95 Nets



Tenenbaum et al. Science 2011

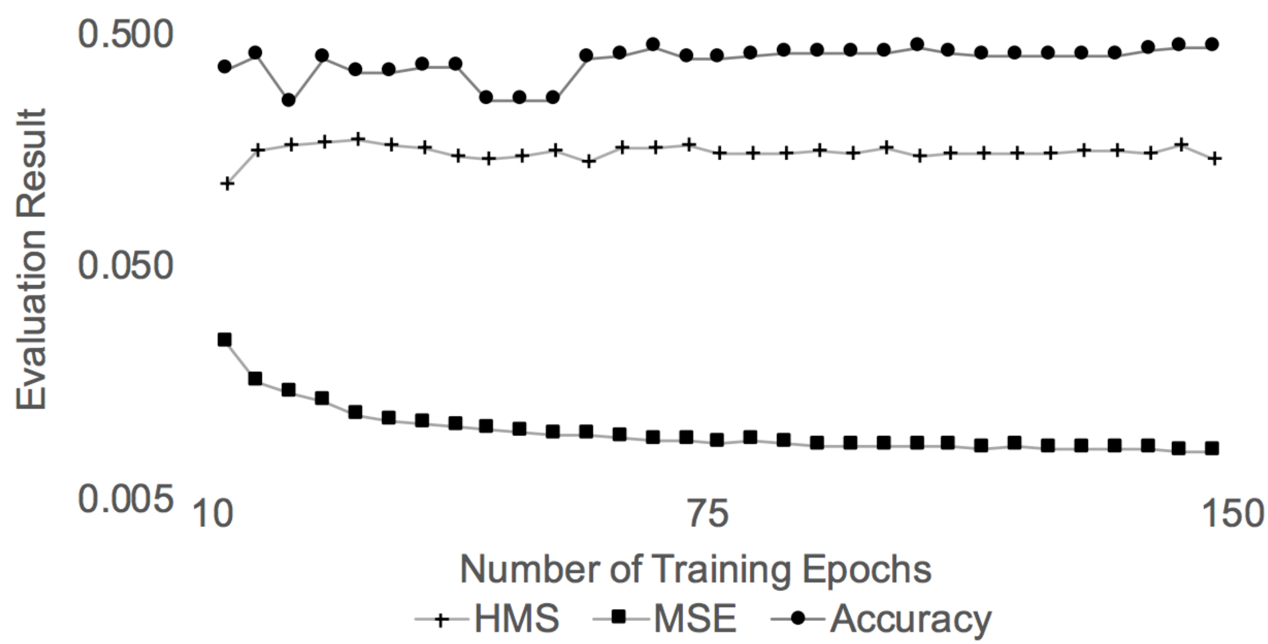
Evaluation Task	Metric	Mean (SD)	Top Ten HMS Mean (SD)
Next Frame Prediction Error	Pixel MSE	0.092 (0.148)	0.009 (0.003)
Object Matching	Accuracy	0.367 (0.134)	0.459 (0.049)
Human-Model Similarity	RDM Correlation	0.106 (0.055)	0.178 (0.011)

HMS is predictive of network performance on other metrics

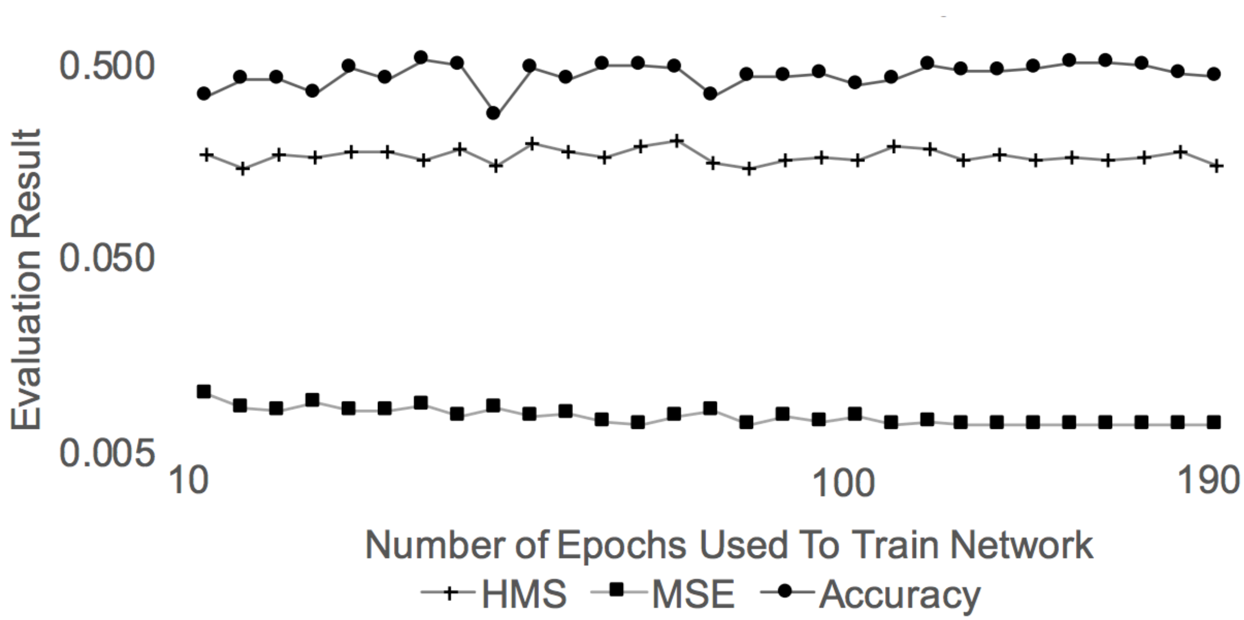
Variable	Accuracy	HMS	Learning Rate
Next Frame Prediction Error	-0.791^{**}	**-0.646^{**}**	0.635^{**}
Object Matching Accuracy	.	**0.575^{**}**	-0.517^{**}
Human-Model Similarity	.	.	-0.452^{**}

$^{**}p < 0.001$

Within-Network Stability

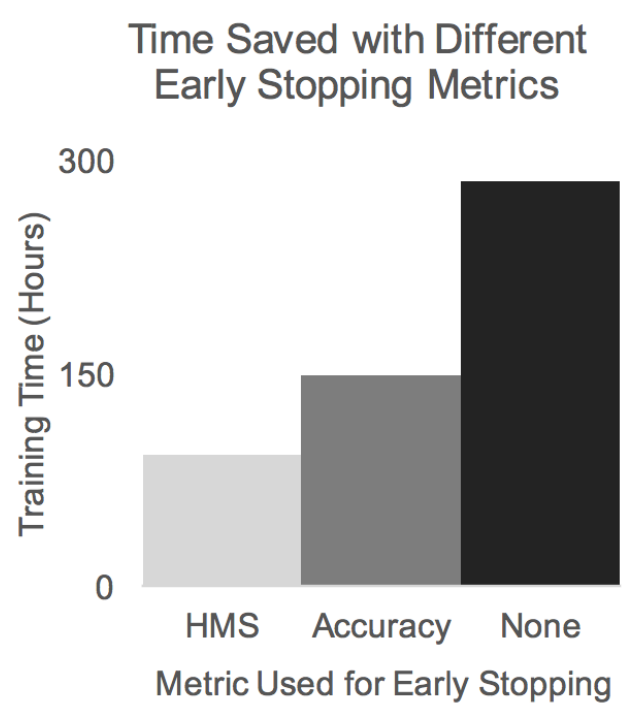


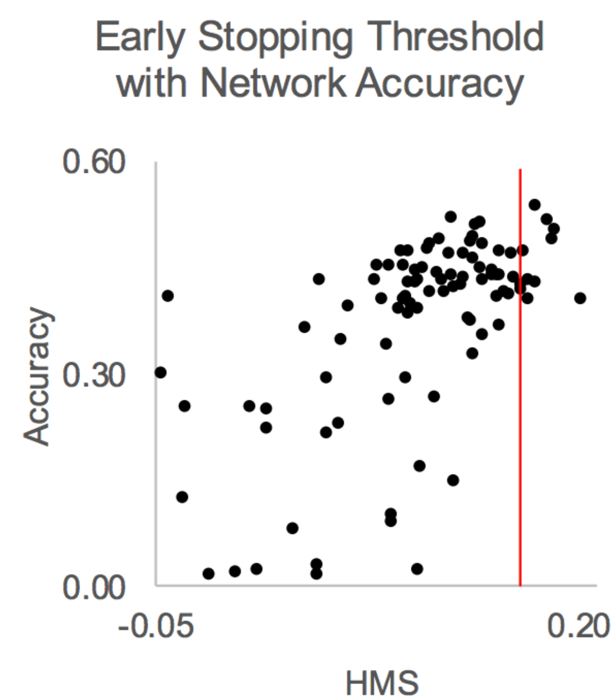
Across-Network Stability



66 Models, Mean

HMS-Driven Early Stopping



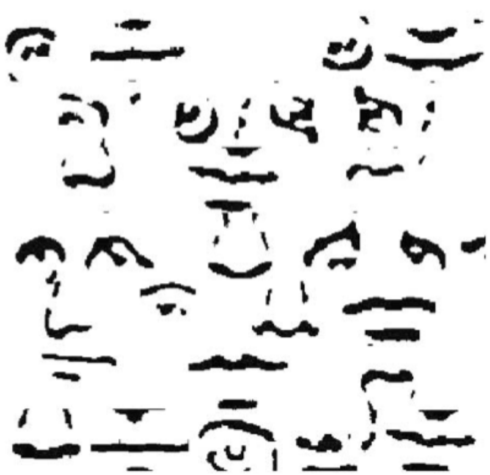


Is fMRI the best reference for this?

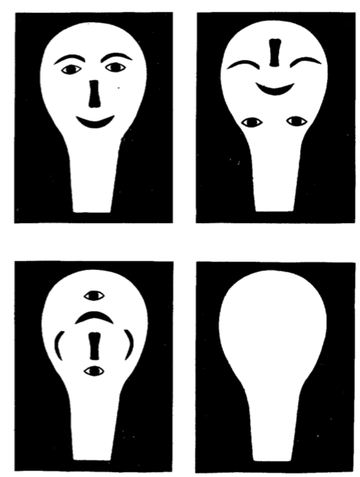
Easier: Human Behavior

Visual Psychophysics: probe psychological and perceptual thresholds through controlled manipulation of stimuli.

Careful management of stimulus construction, ordering and presentation allows for precise determination of perceptual thresholds.



Garrido et al 2011



Goren et al 1975

Thank you!