Mind the gap: comparing multiple models of scene representation in brain and behavior

Iris Groen, PhD New York University

Using goal-driven deep learning models to understand sensory cortex

Daniel L K Yamins[1,2] & James J DiCarlo[1,2]

Deep Neural Networks: A New Framework for Modeling Biological Vision and Brain Information Processing

Nikolaus Kriegeskorte

Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge, Cambridge CB2 7EF, United Kingdom; email: nikolaus.kriegeskorte@mrc-cbu.cam.ac.uk

Toward an Integration of Deep Learning and Neuroscience

Adam H. Marblestone[1], Greg Wayne[2] and Konrad P. Kording[3]*

Neural Networks and Neuroscience-Inspired Computer Vision

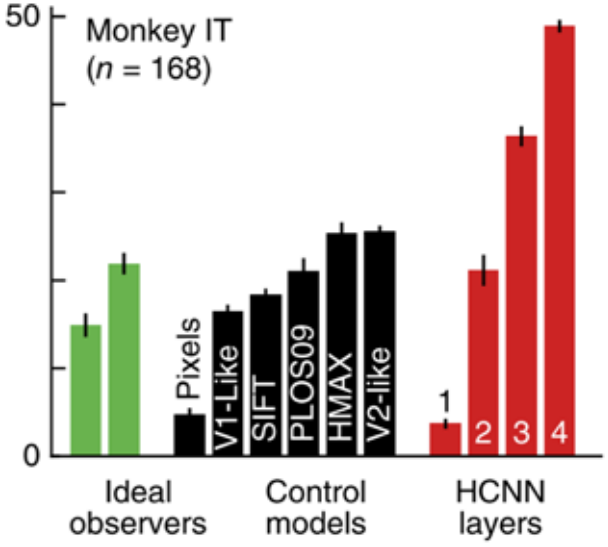
David Daniel Cox[1,2,3,*] and Thomas Dean[4,5]

Visual Object Recognition: Do We (Finally) Know More Now Than We Did?

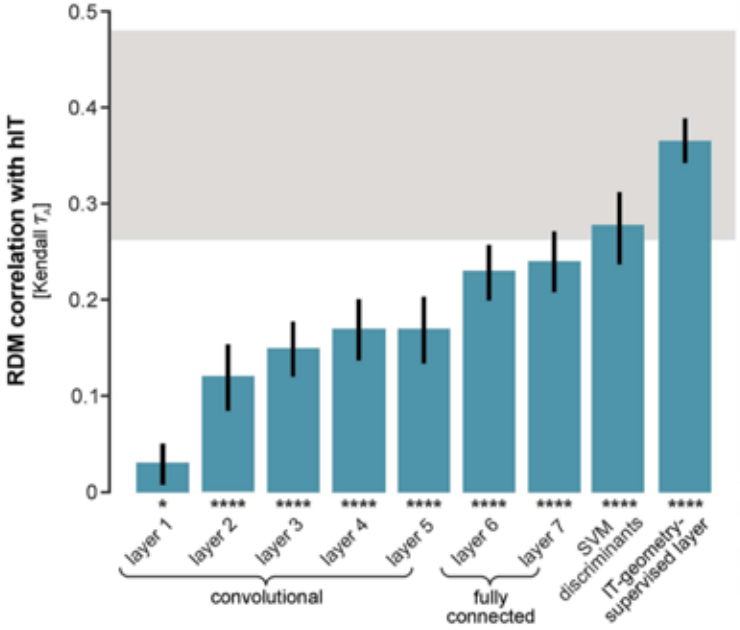
Isabel Gauthier[1] and Michael J. Tarr[2]

Deep nets and object recognition

Krizhevsky, Sutskever & Hinton, 2012

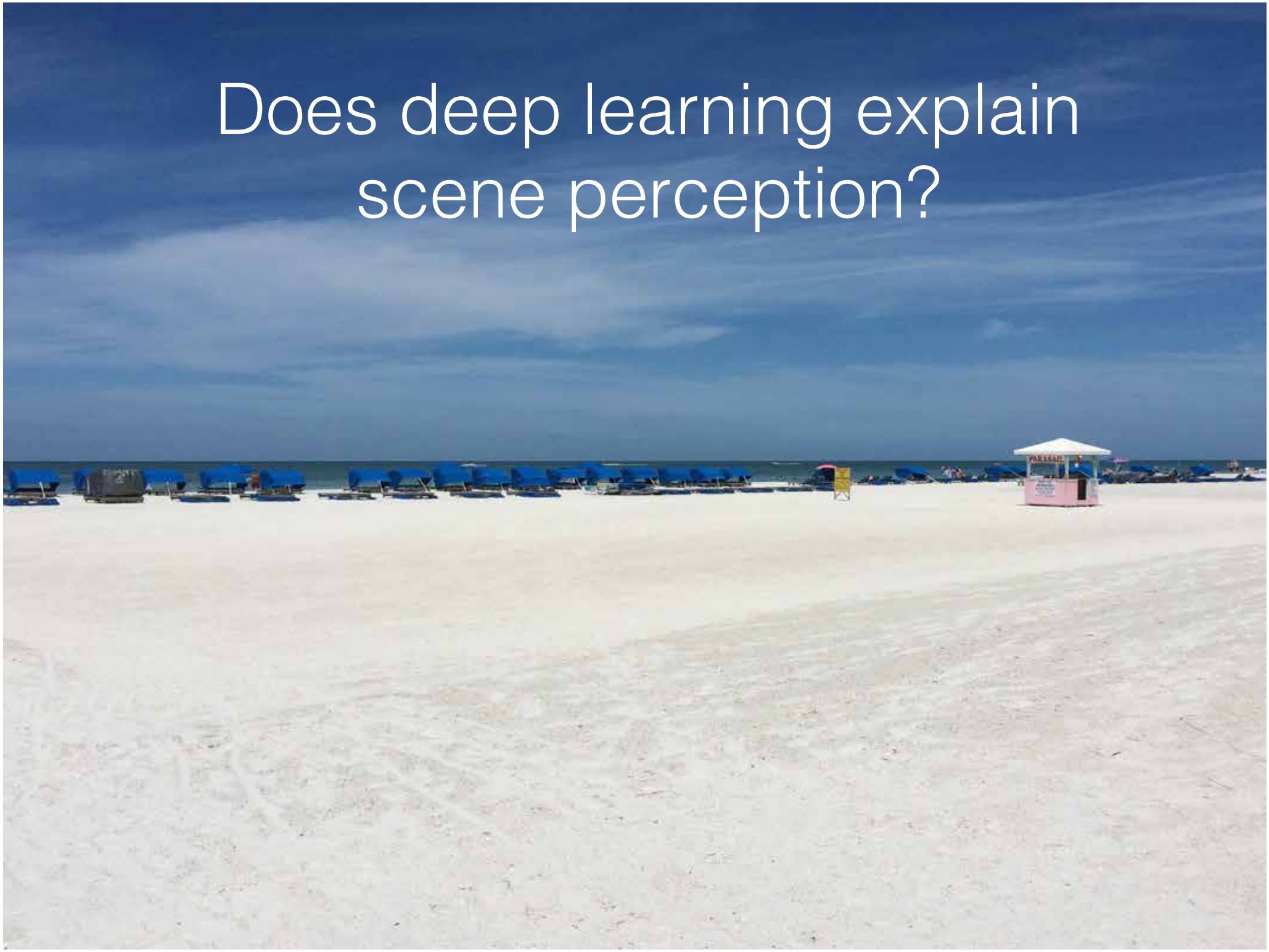


Yamins et al, 2014



Khaligh-Razavi & Kriegeskorte, 2014


beach hut
sand
beach chair
“beach”


Does deep learning explain
scene perception?

Outline

- Scene vs. object perception
- fMRI study 1: objects-in-context
- fMRI study 2: comparing multiple models
- How to move forward?

Scene vs. object perception

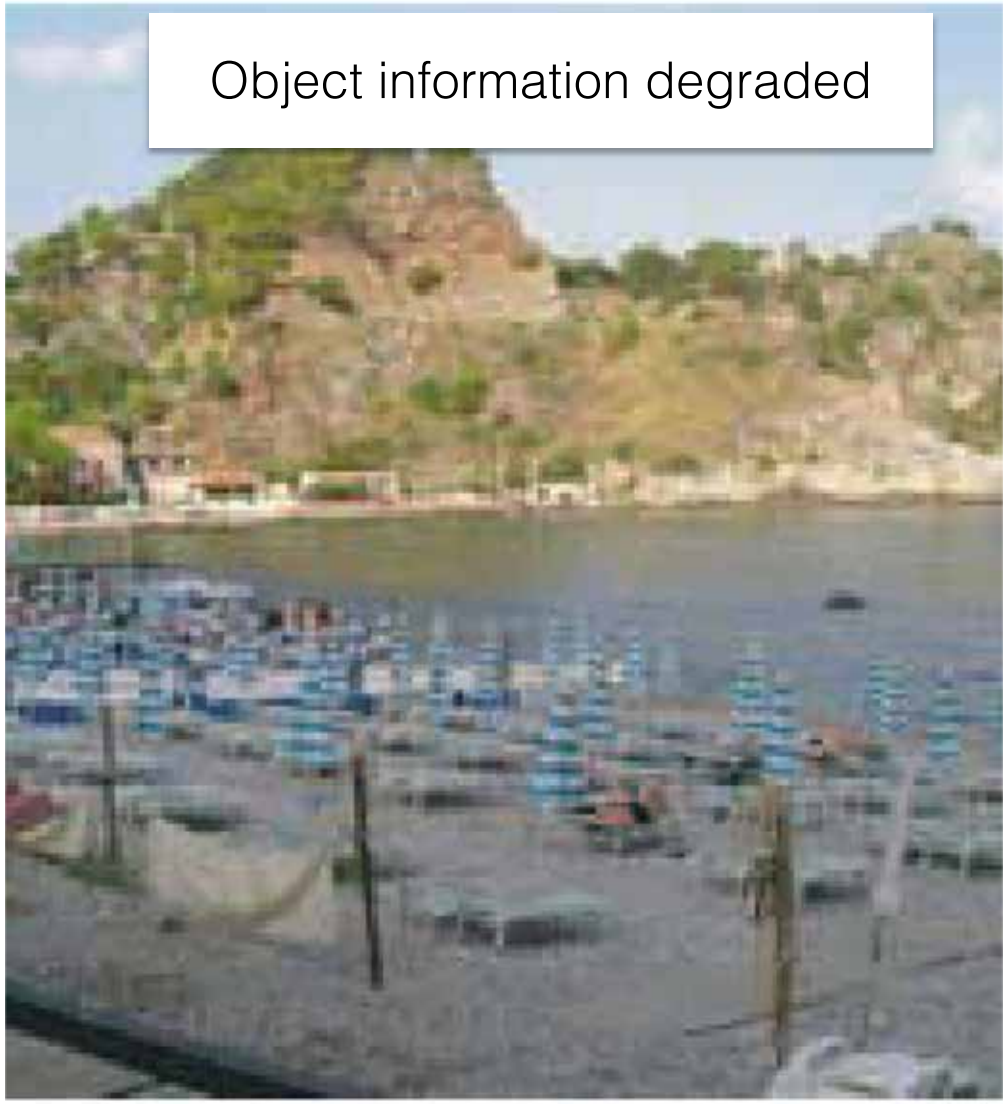
A scene is a semantically coherent (and often namable) view of a **real-world environment** comprising **background elements** and **multiple discrete objects** arranged in a **spatially licensed manner**.

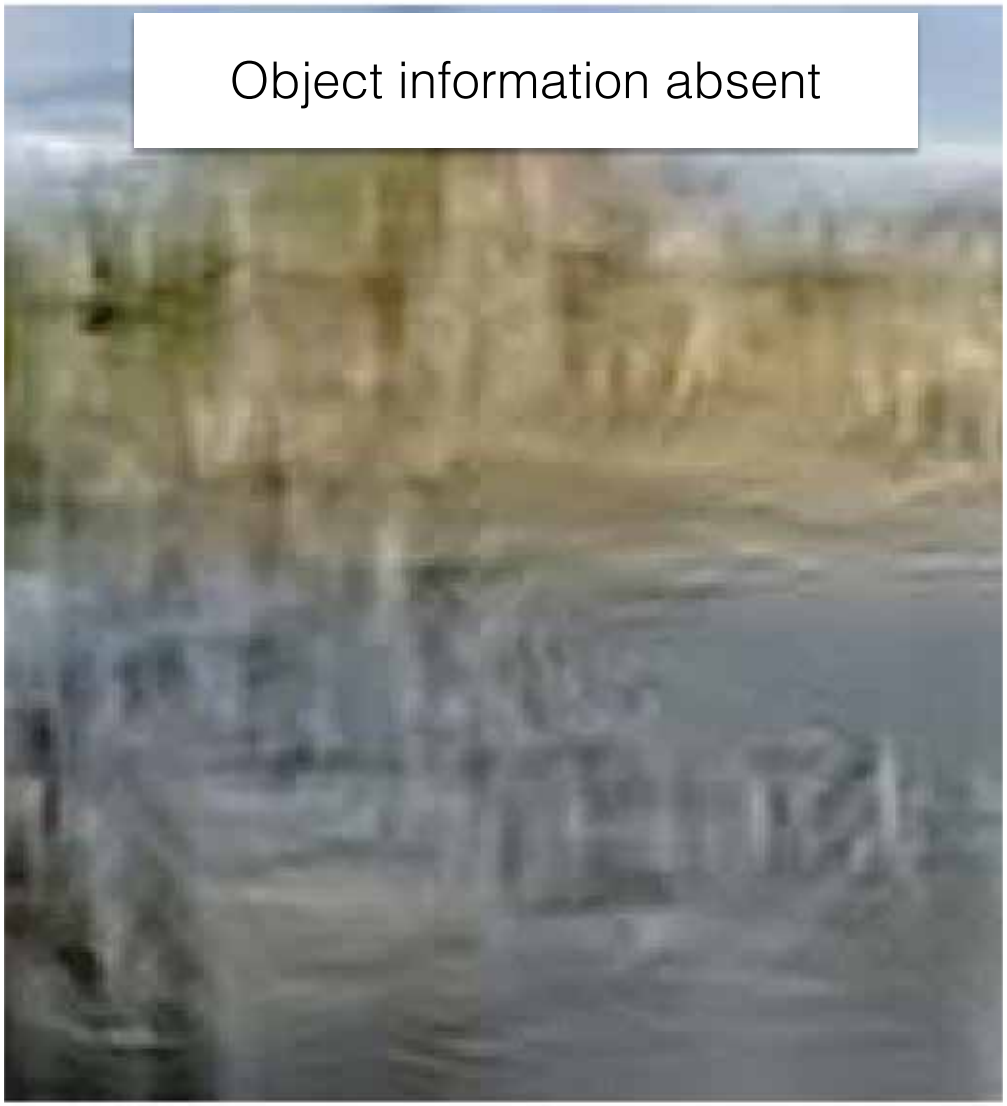
Henderson and Hollingworth (1999)

Scene perception can be usefully contrasted to object perception: whereas objects are **spatially compact entities** that **one acts upon**, **scenes are spatially distributed entities** that **one acts within**.

Epstein (2005)

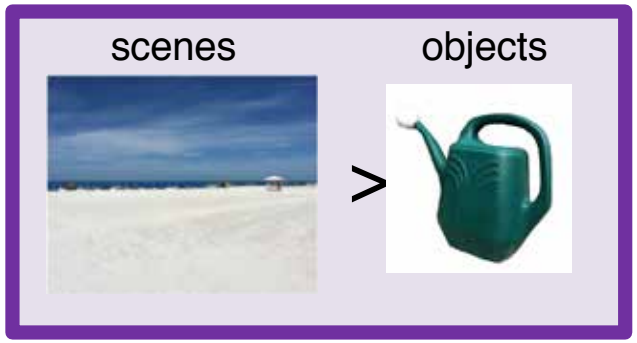
Scene recognition without objects

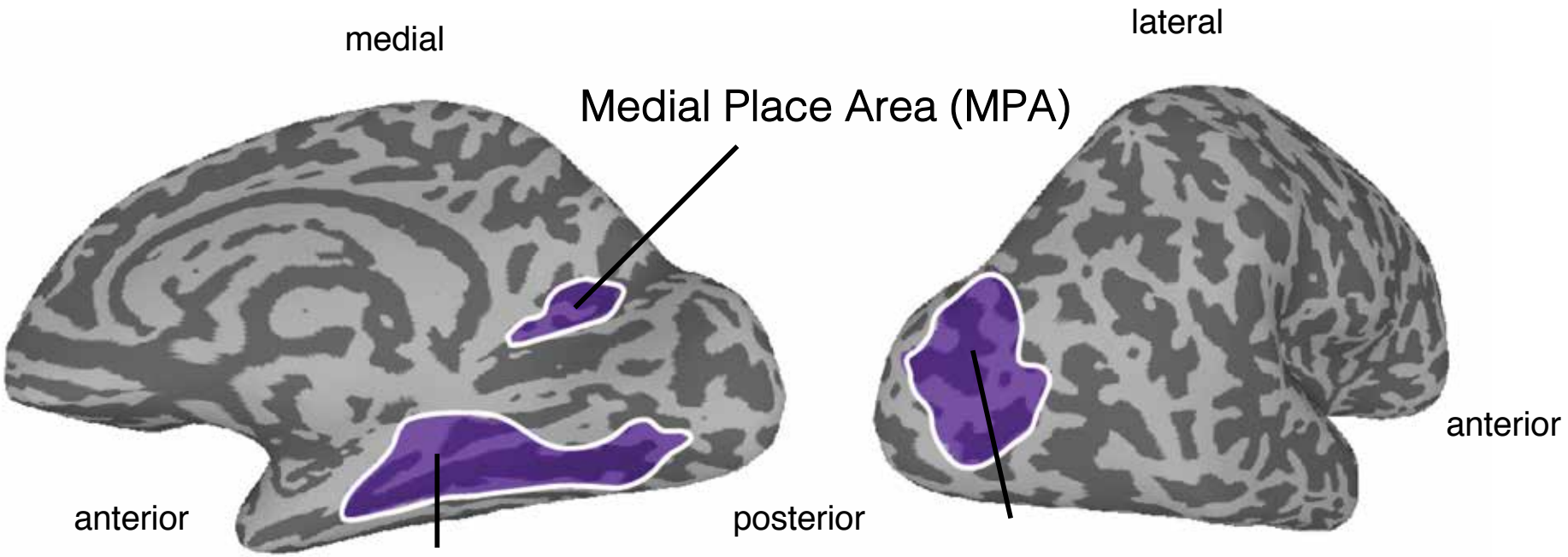




Oliva, 2005

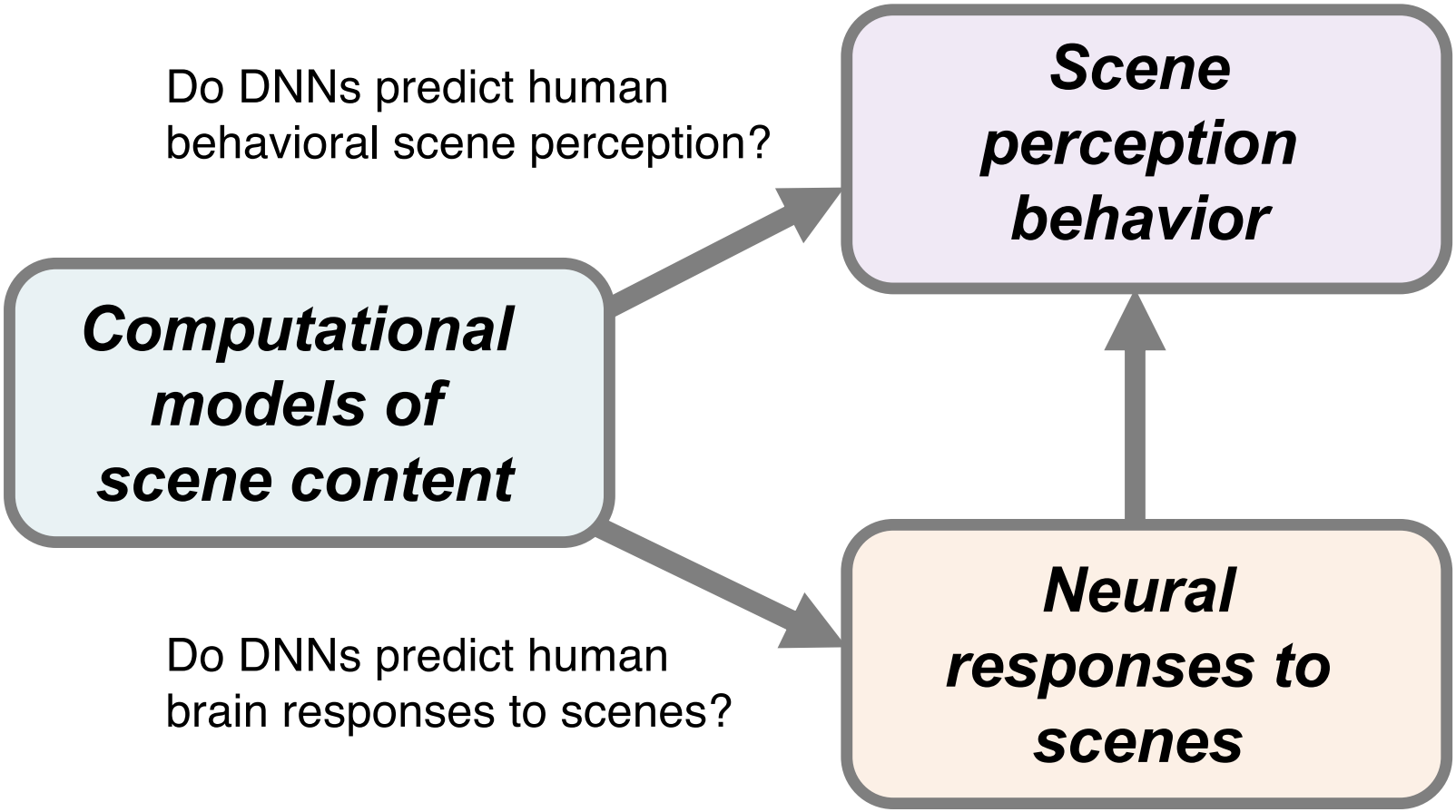
Scene-selective brain regions





See e.g. Epstein (2014) for a review

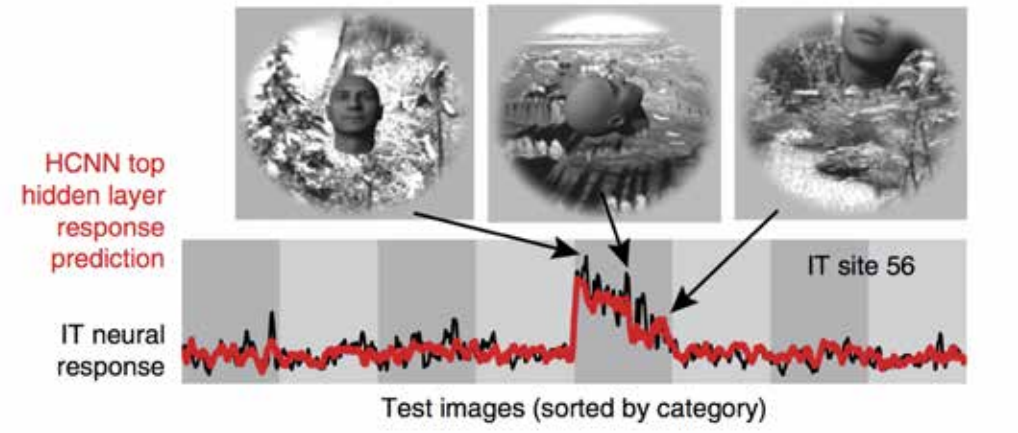
Understanding scene perception



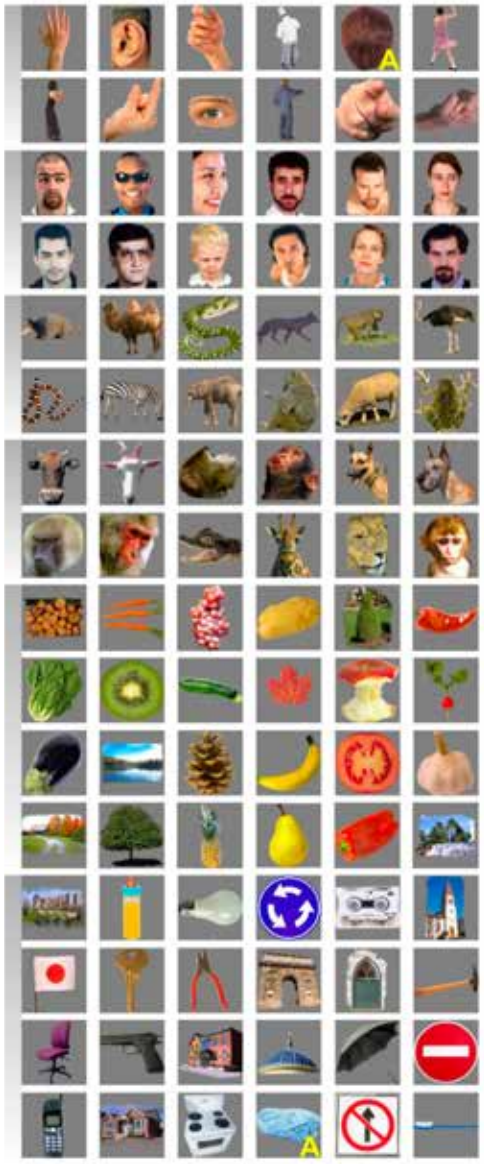
Outline

- Scene vs. object perception
- fMRI study 1: objects-in-context
- fMRI study 2: comparing multiple models
- How to move forward?

DNNs and object recognition

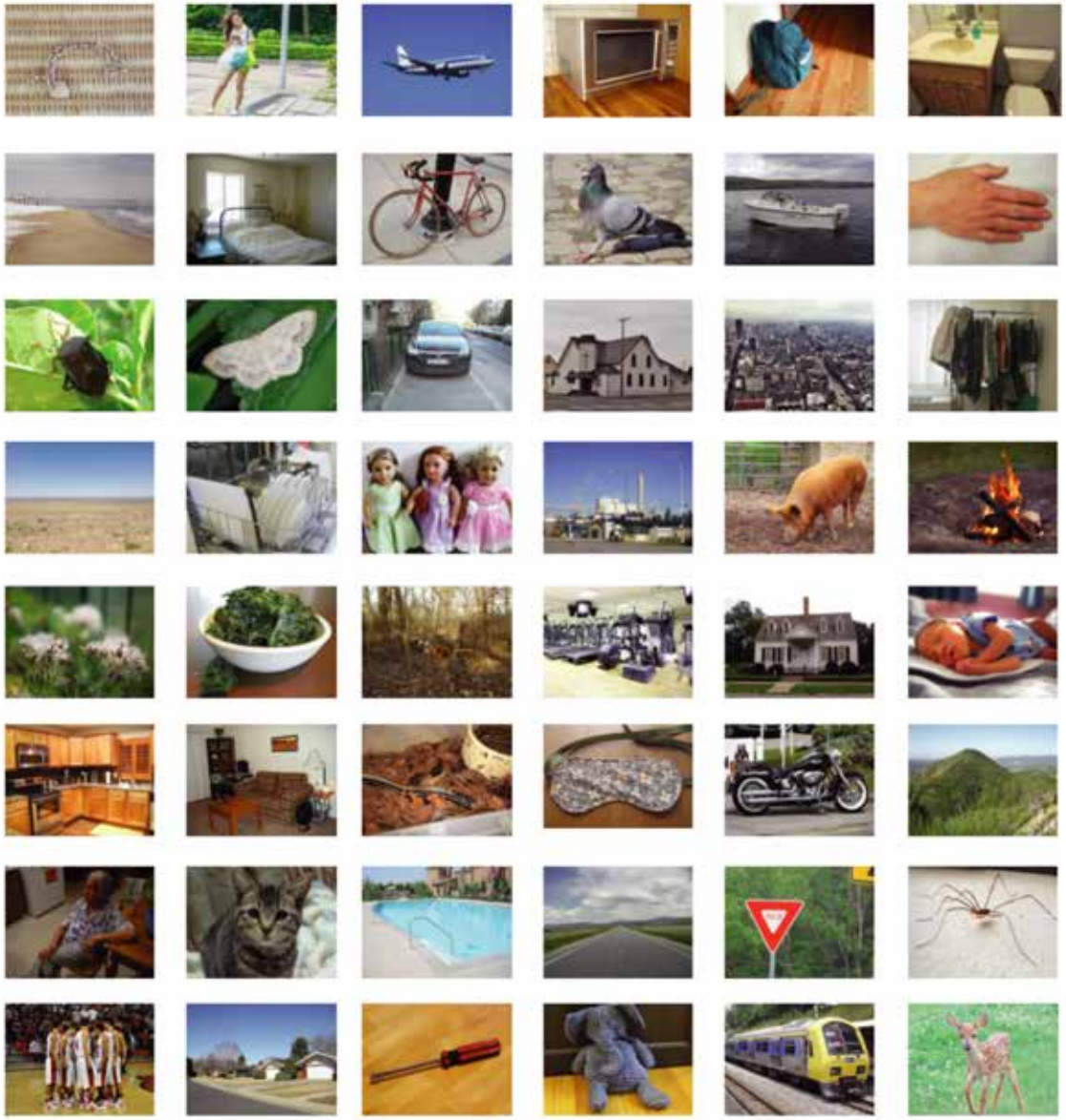


Yamins et al, 2014



Khaligh-Razavi & Kriegeskorte, 2014

Study 1: objects-in-context + scenes

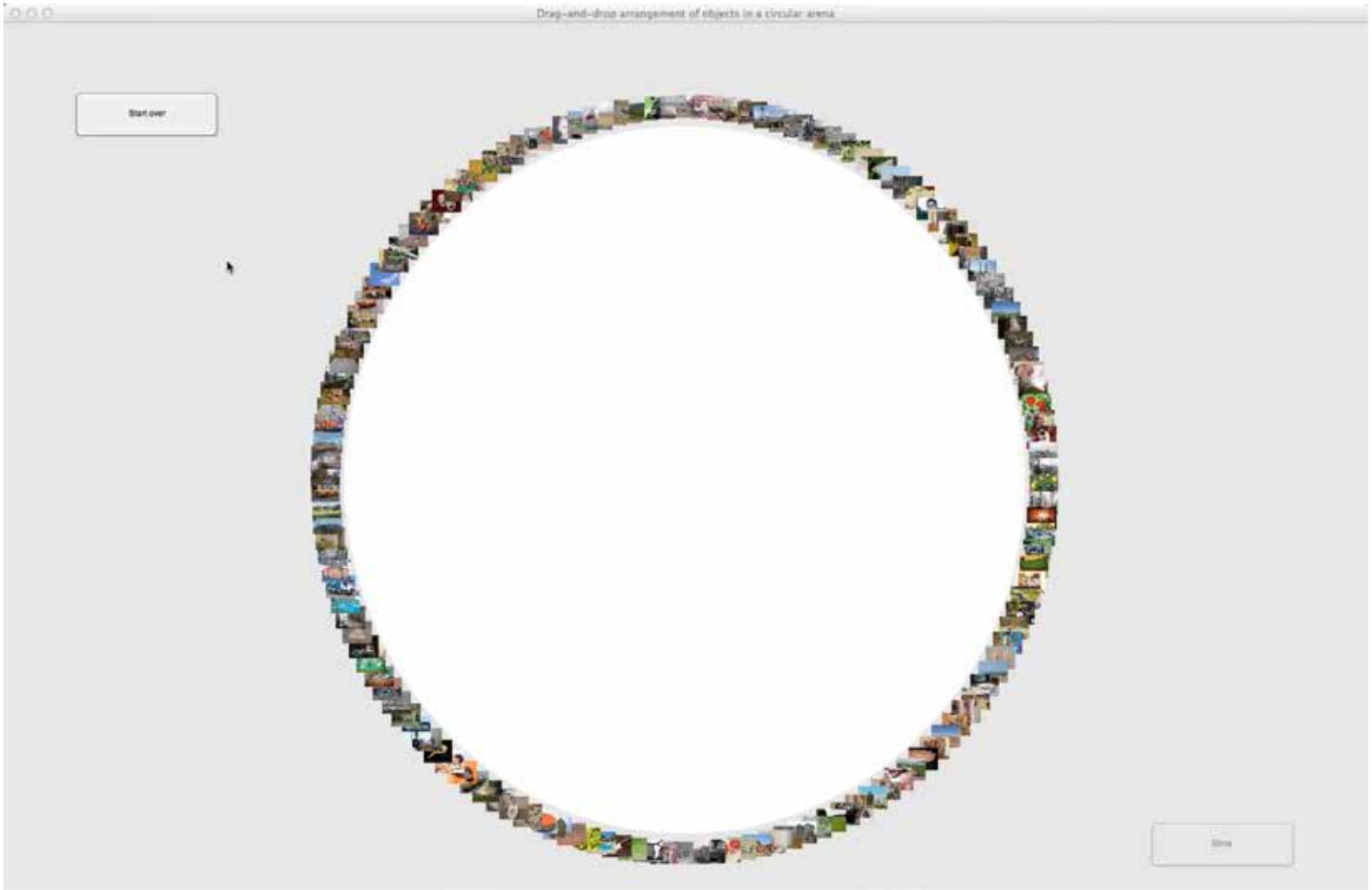


48 naturalistic image categories
(3 exemplars per category, 2 sets)

accessories
adults
airplanes
appliances
bags
bathrooms
beaches
beds
bikes
birds
boats
body parts
bugs
butterflies
cars
churches
cityscapes
clothes
deserts
dishes
dolls
factories
farm animals
fire
flowers
food
forests
gyms
houses
kids
kitchens
living rooms
lizards/snakes
masks
motorcycles
mountains
older adults
pets
pools
roads
signs
spiders
sports
suburbs
tools
toys
trains
wild animals

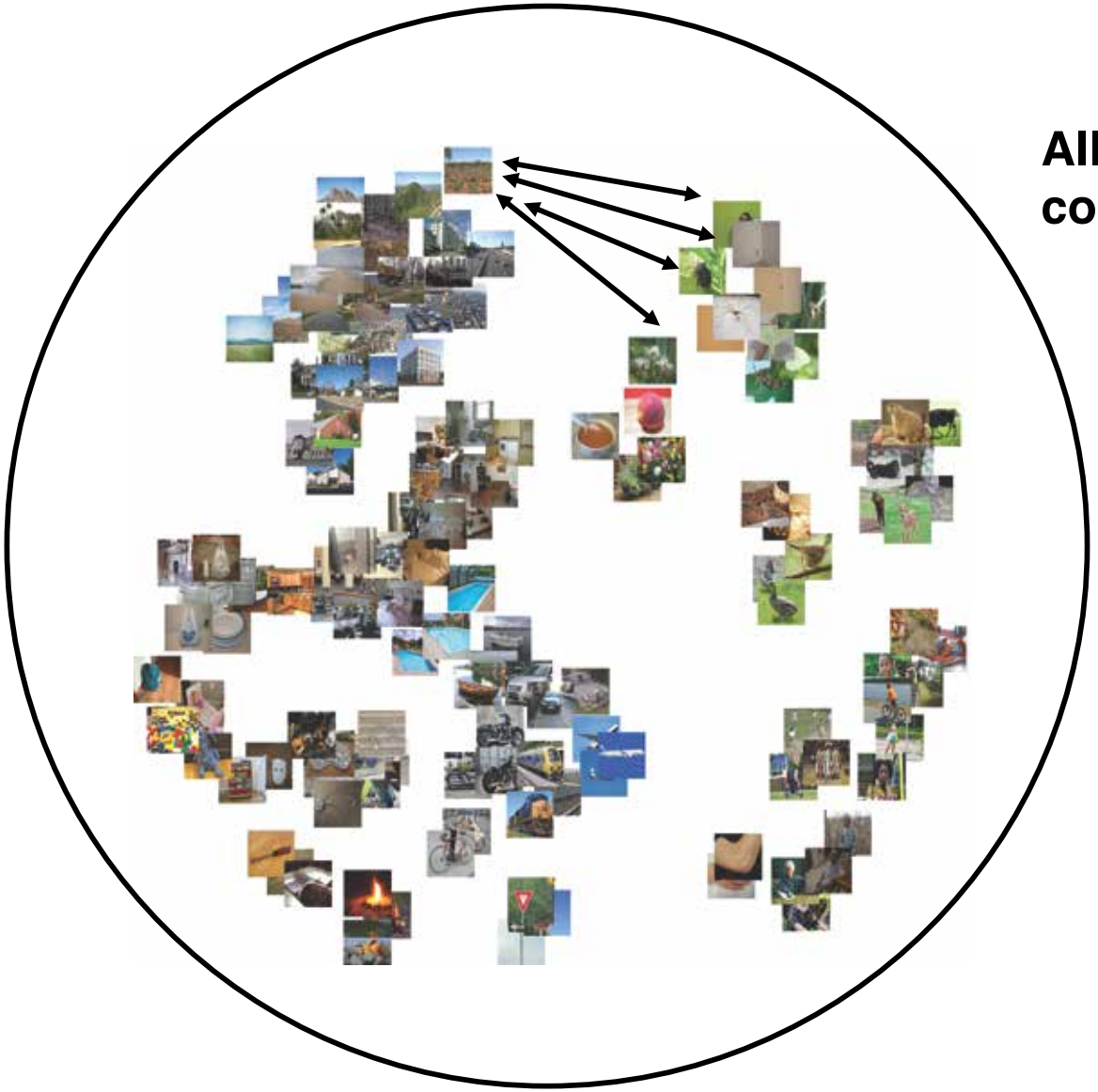
Free arrangement behavioral task

"Group these images according to their similarity"



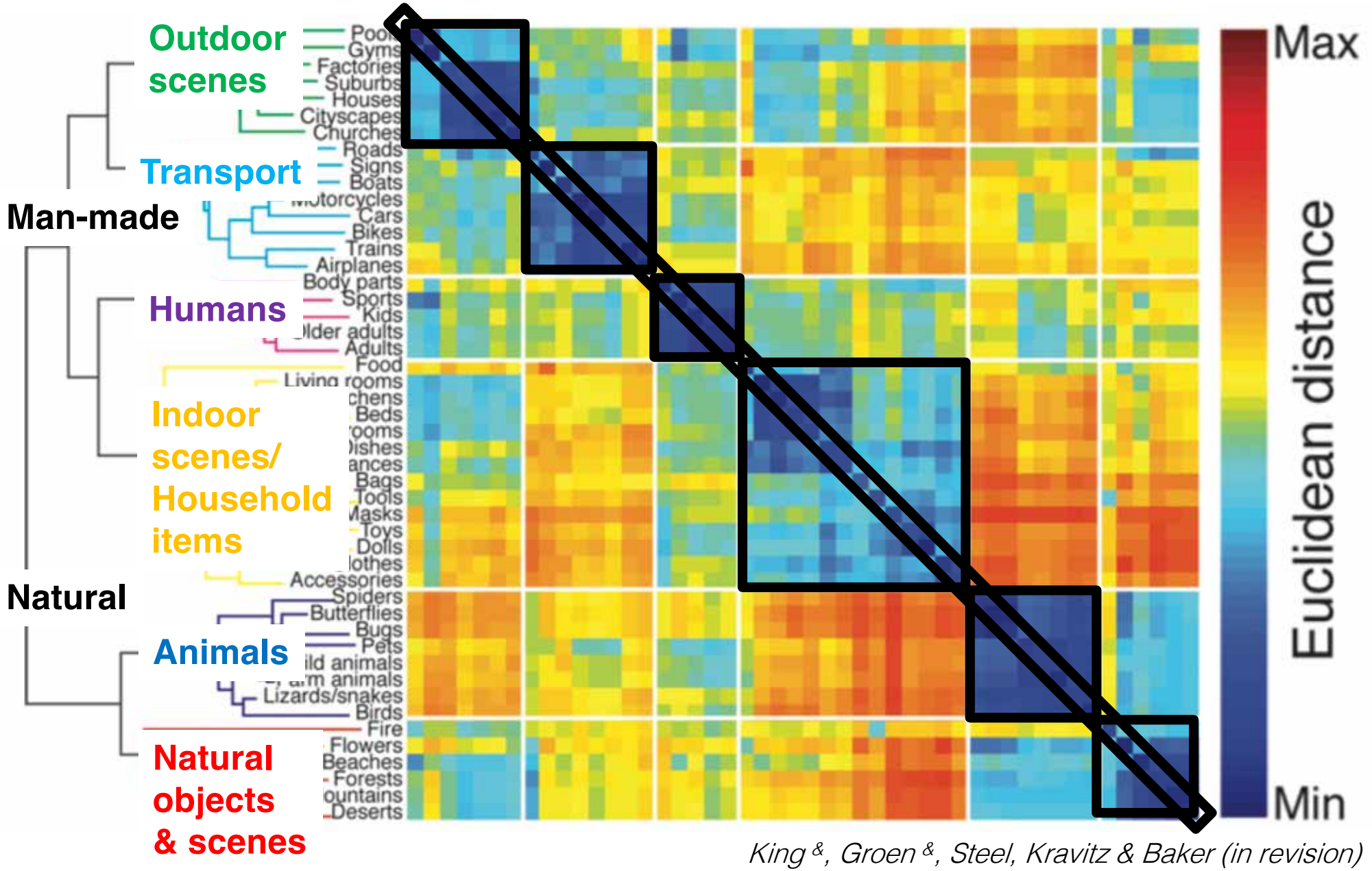
Mur et al., 2013

Free arrangement behavioral task



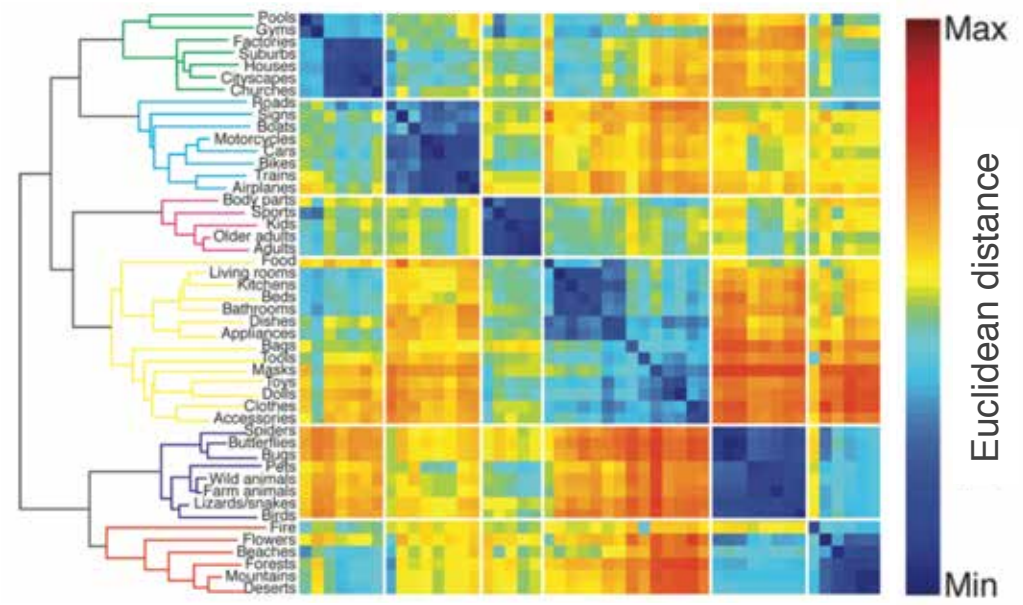
All pairwise comparisons

Behavioral dissimilarity

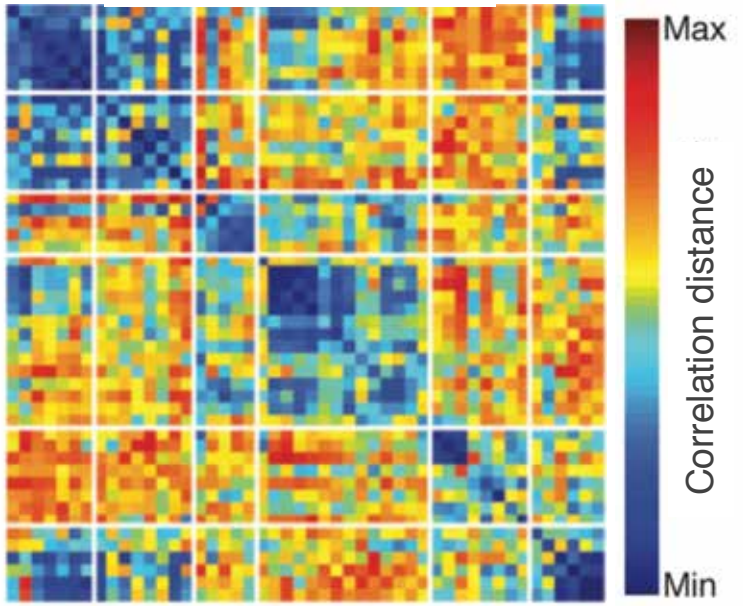


King &, Groen &, Steel, Kravitz & Baker (in revision)

Behavior

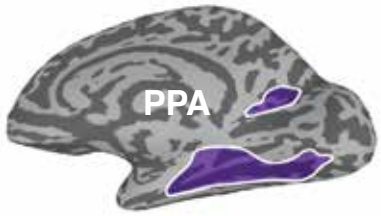


AlexNet layer 8



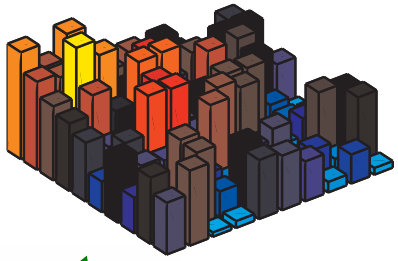
Ventral temporal cortex?

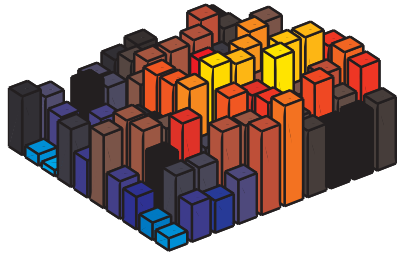
Scene-selective cortex?



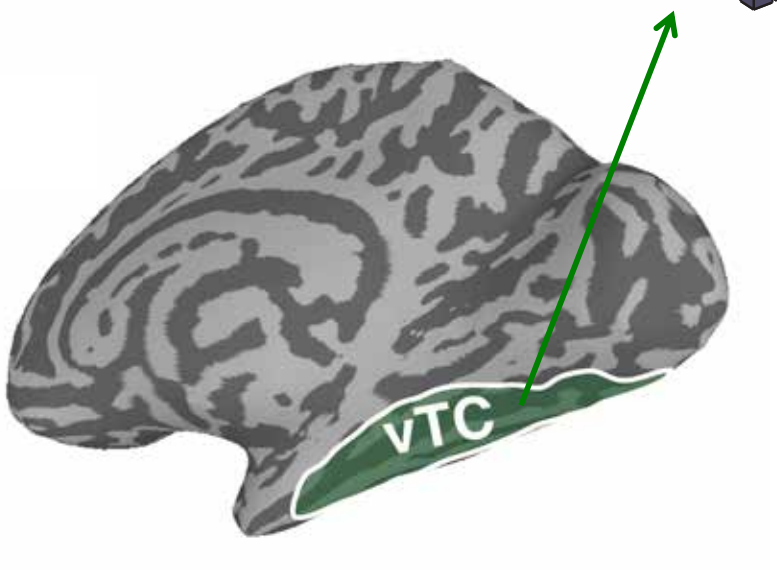
King &, Groen &, Steel, Kravitz & Baker (in revision)

Multi-voxel pattern analysis

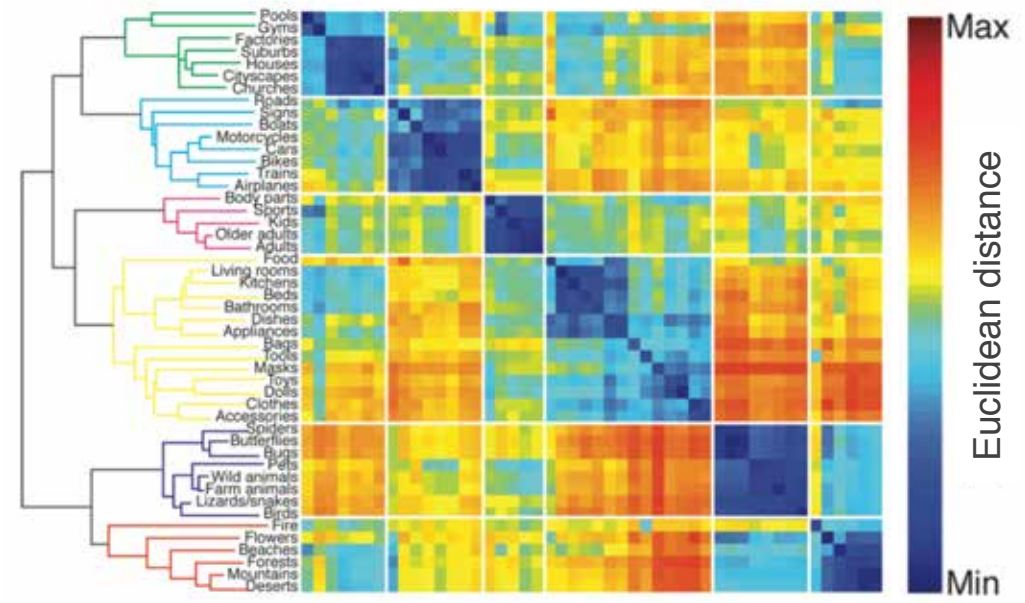




1-correlation

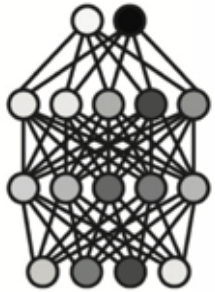


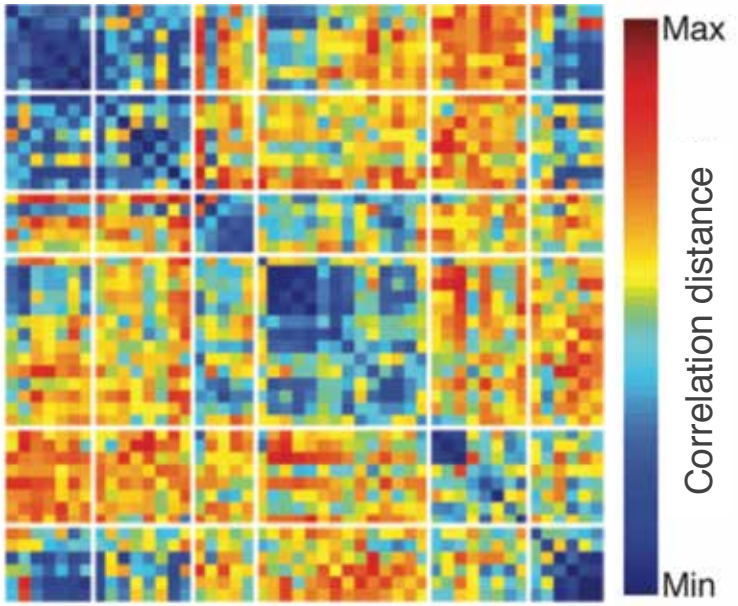
Behavior



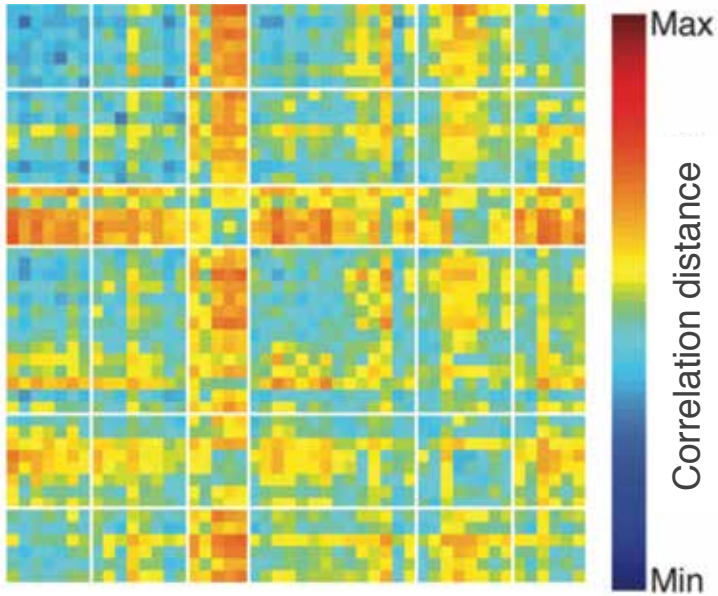
AlexNet layer 8

DNN

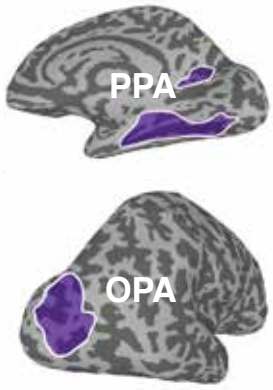


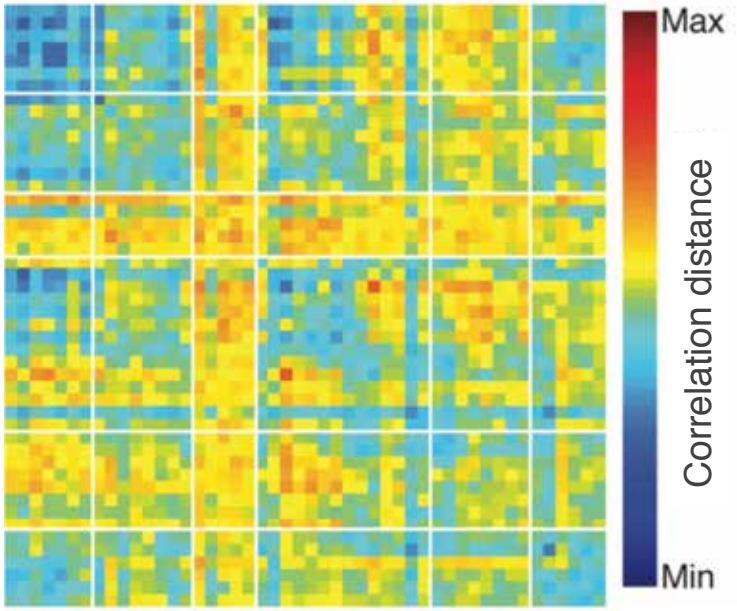


Ventral temporal cortex



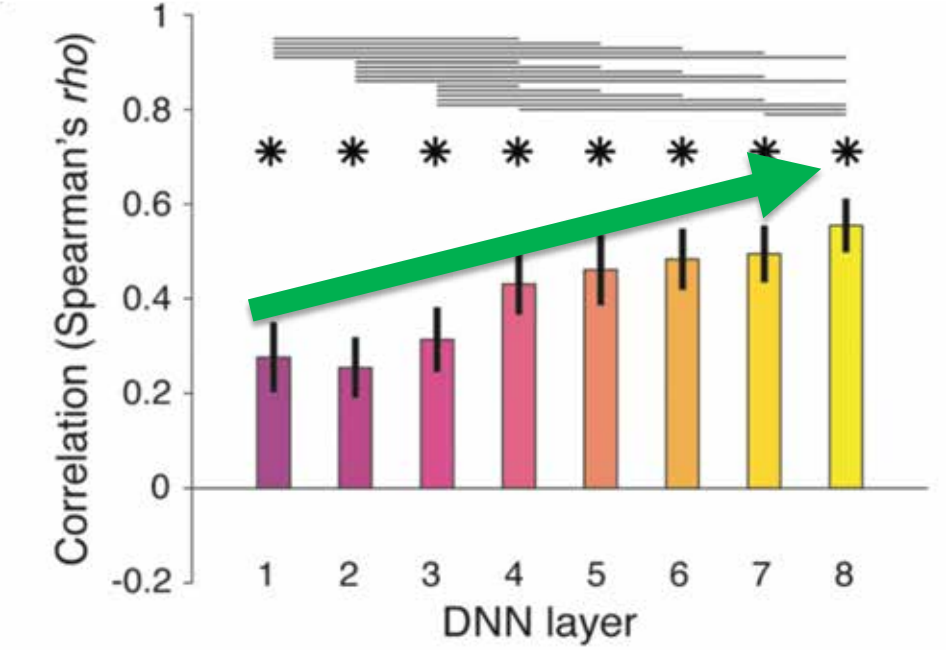
Scene-selective cortex



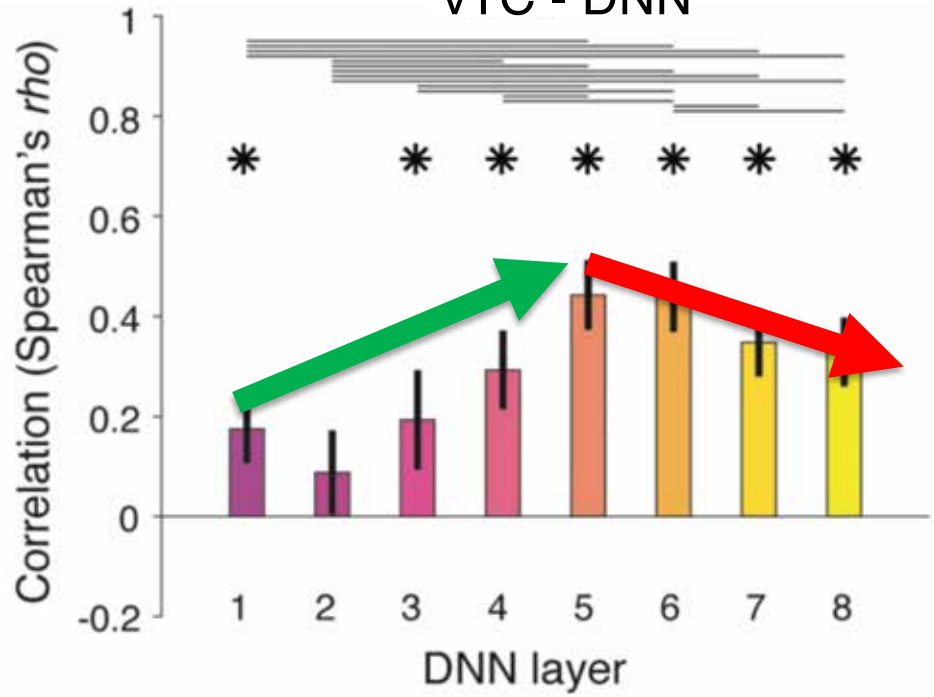


King &, Groen &, Steel, Kravitz & Baker (in revision)

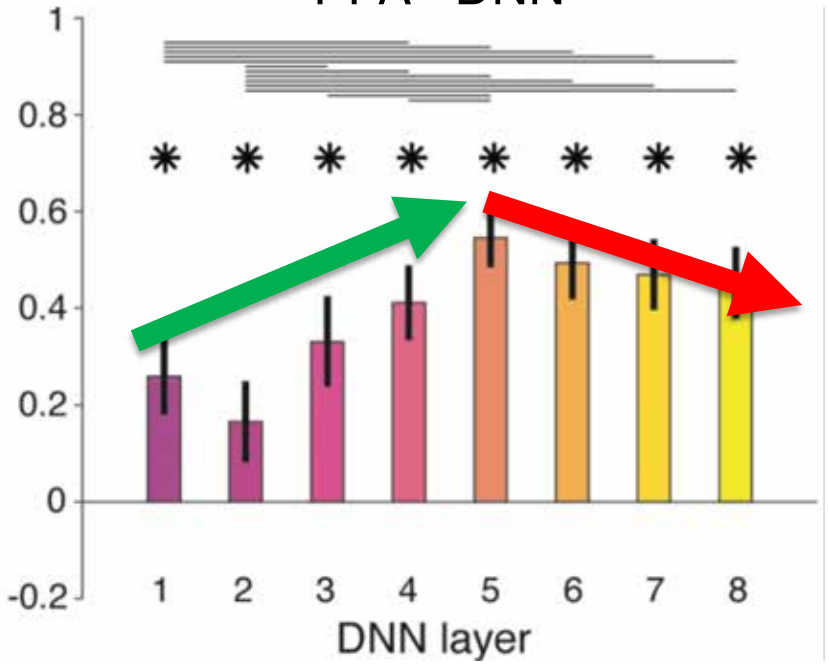
Behavior - DNN



vTC - DNN



PPA - DNN

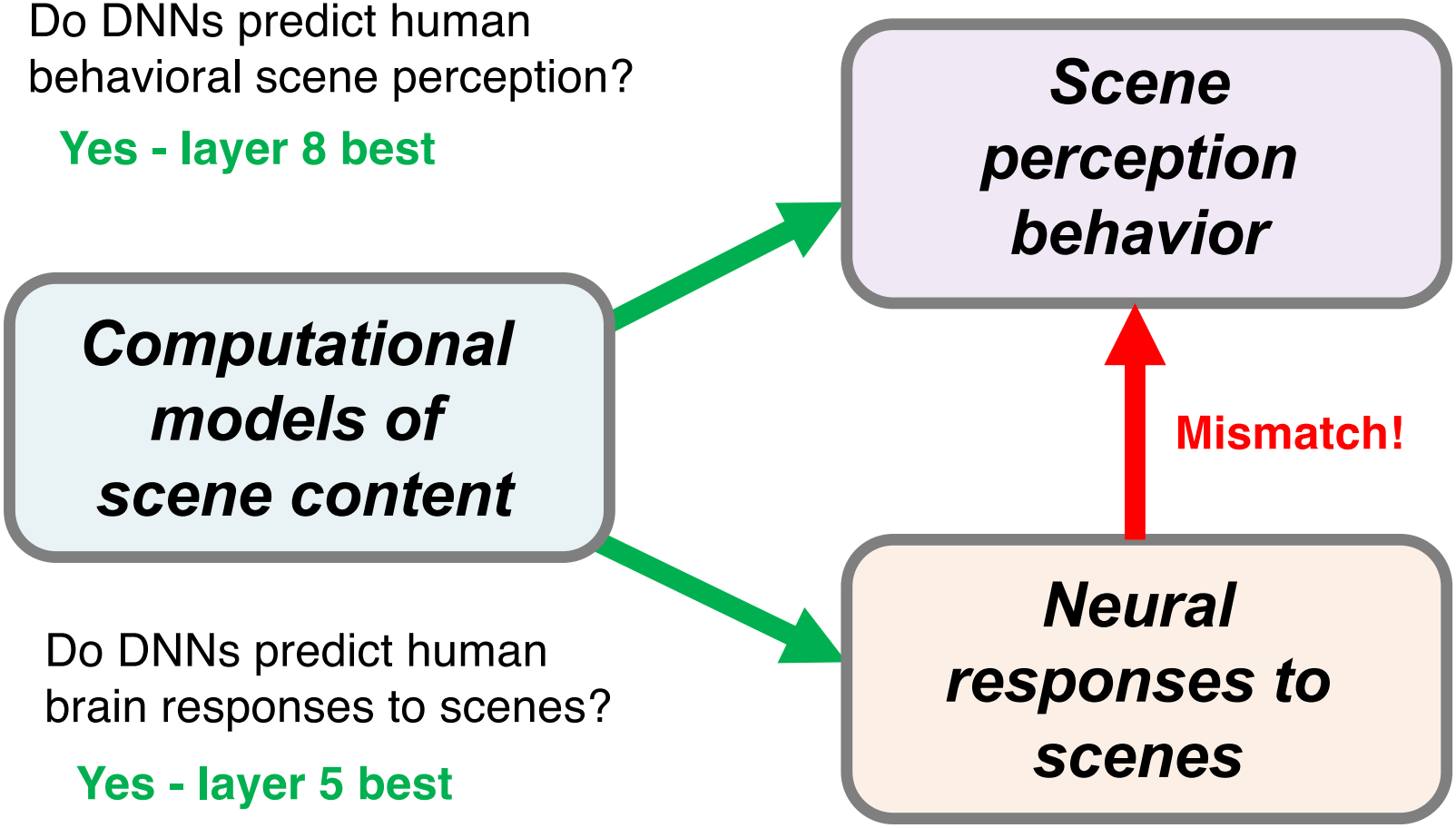


King [&], Groen [&], Steel, Kravitz & Baker (in revision)

Understanding scene perception

Do DNNs predict human behavioral scene perception?

Yes - layer 8 best



Do DNNs predict human brain responses to scenes?

Yes - layer 5 best

Why the mismatch?

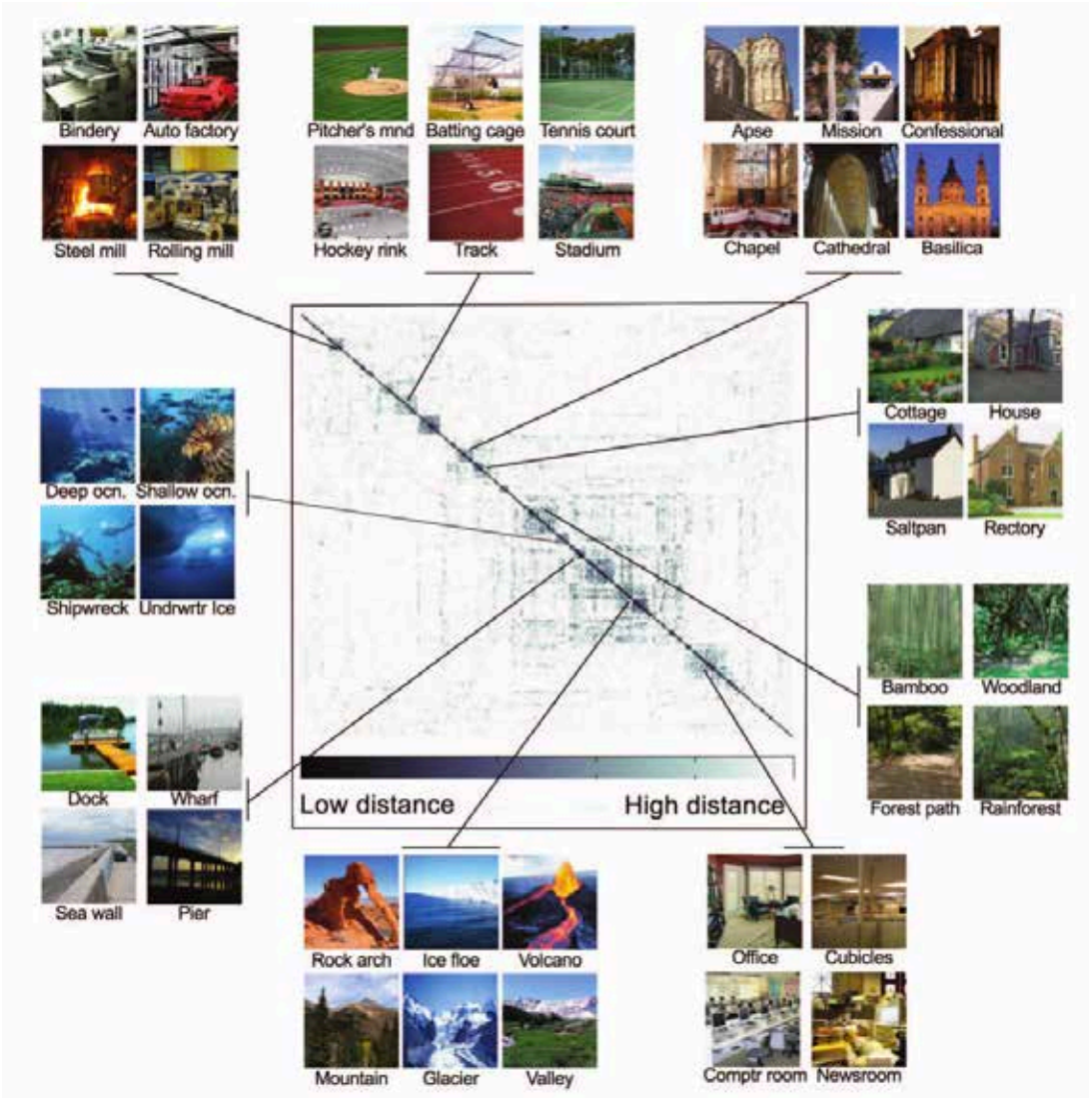
- Nature of stimuli (mix of objects-in-context and scenes)?
- Multi-arrangement task?
- Alternative models for behavior?

Large-scale scene perception

- Mechanical Turk
- SUN database
- > 300 scene categories

SAME or DIFFERENT CATEGORY?

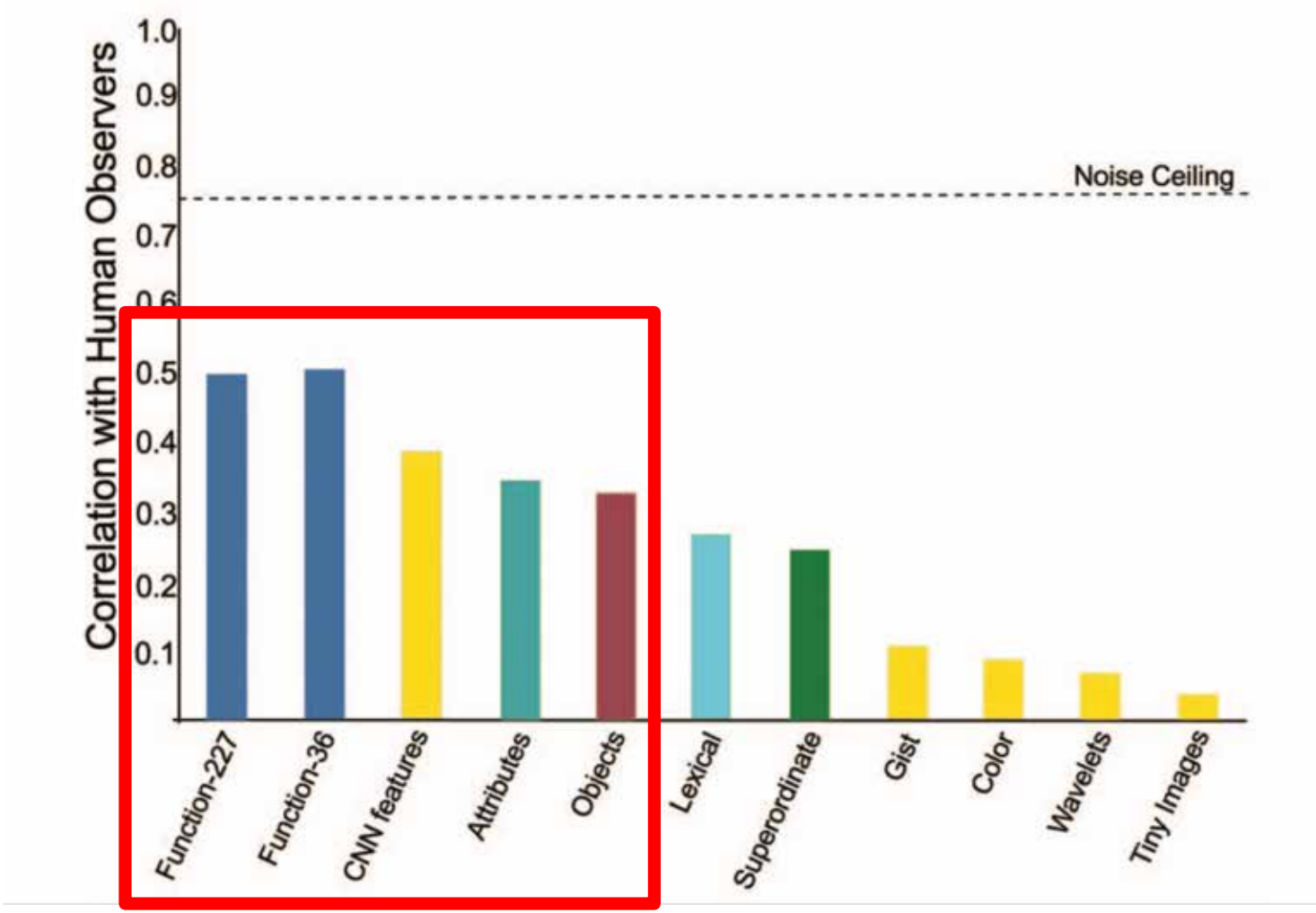
> 2000 participants
> 5 million trials



Greene et al., 2016 (JEP:General)

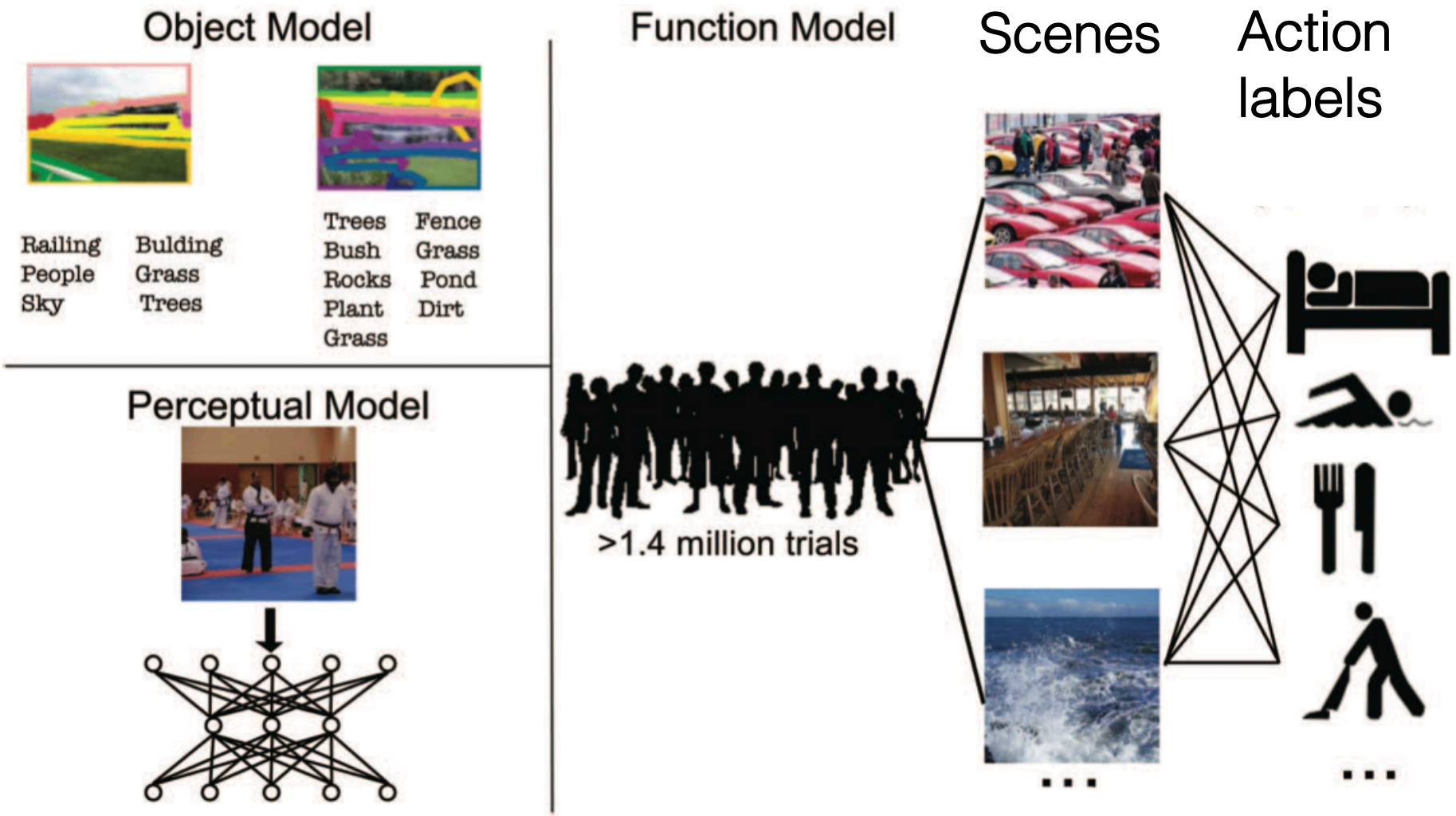
Comparing behavior to multiple models

Correlation with scene categorization



Greene et al., 2016 (JEP:General)

Three best models of scene perception



Convolutional Neural Network
(AlexNet; 7th layer)

Greene et al., 2016 (JEP:General)

Study 2

- **Goal:** determine how well the top three models predict fMRI responses and behavioral multi-arrangement task

!! Inherent correlations make it difficult to disambiguate models, even within large sets of naturalistic images

(Lescroart, Stansbury & Gallant, 2015; Malcolm, Groen & Baker, 2016)

→ Minimize model correlations through stimulus selection: Iterative sampling from SUN database until we obtain maximally different predictions

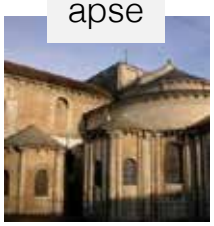
→ Use variance partitioning to identify unique contribution of each model

Stimuli by function

access_road

airplane_cabin

apse



badminton_court

bamboo_forest

bar

batting_cage

bindery

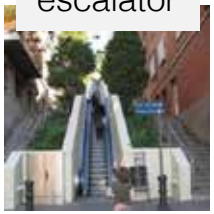
bus_depot

butte

control_tower

dolmen

escalator



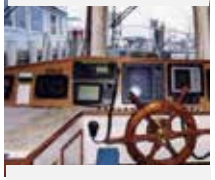
hedgerow

lido_deck

naval_base

pier

pilothouse



playroom

pump_room

putting_green

skyscraper

stadium

stilt_house

tea_garden

underwater_pool

volcano

volleyball_court

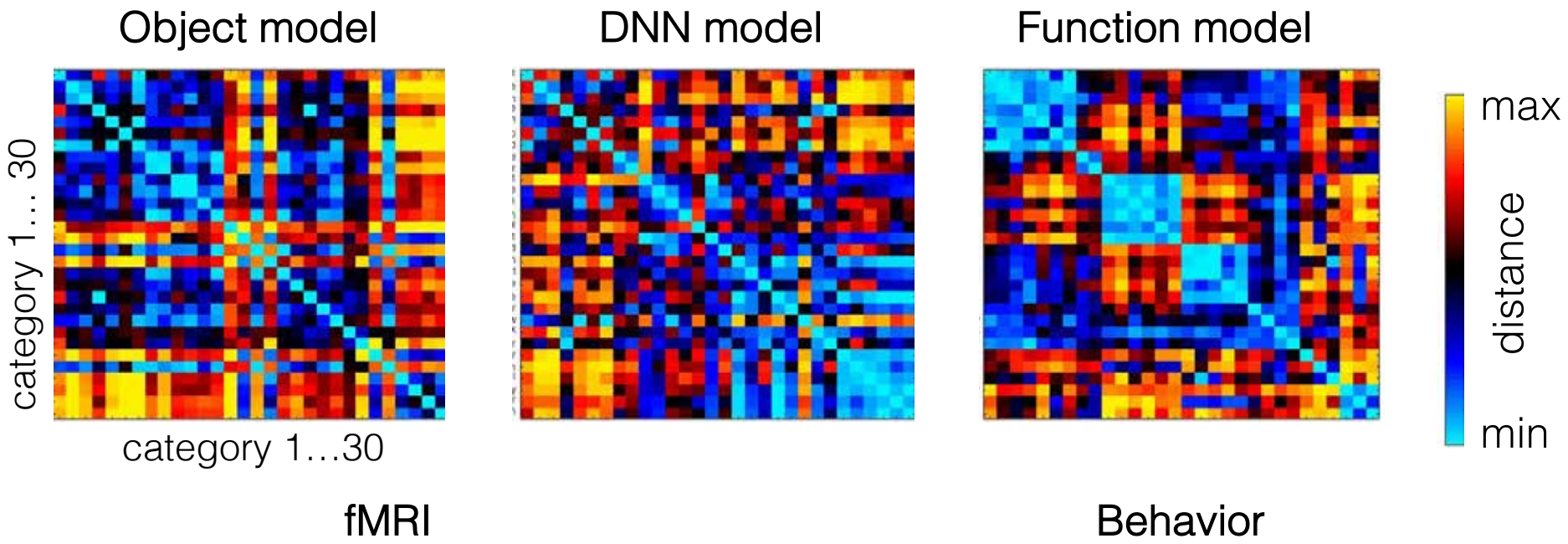
woodland

youth_hostel

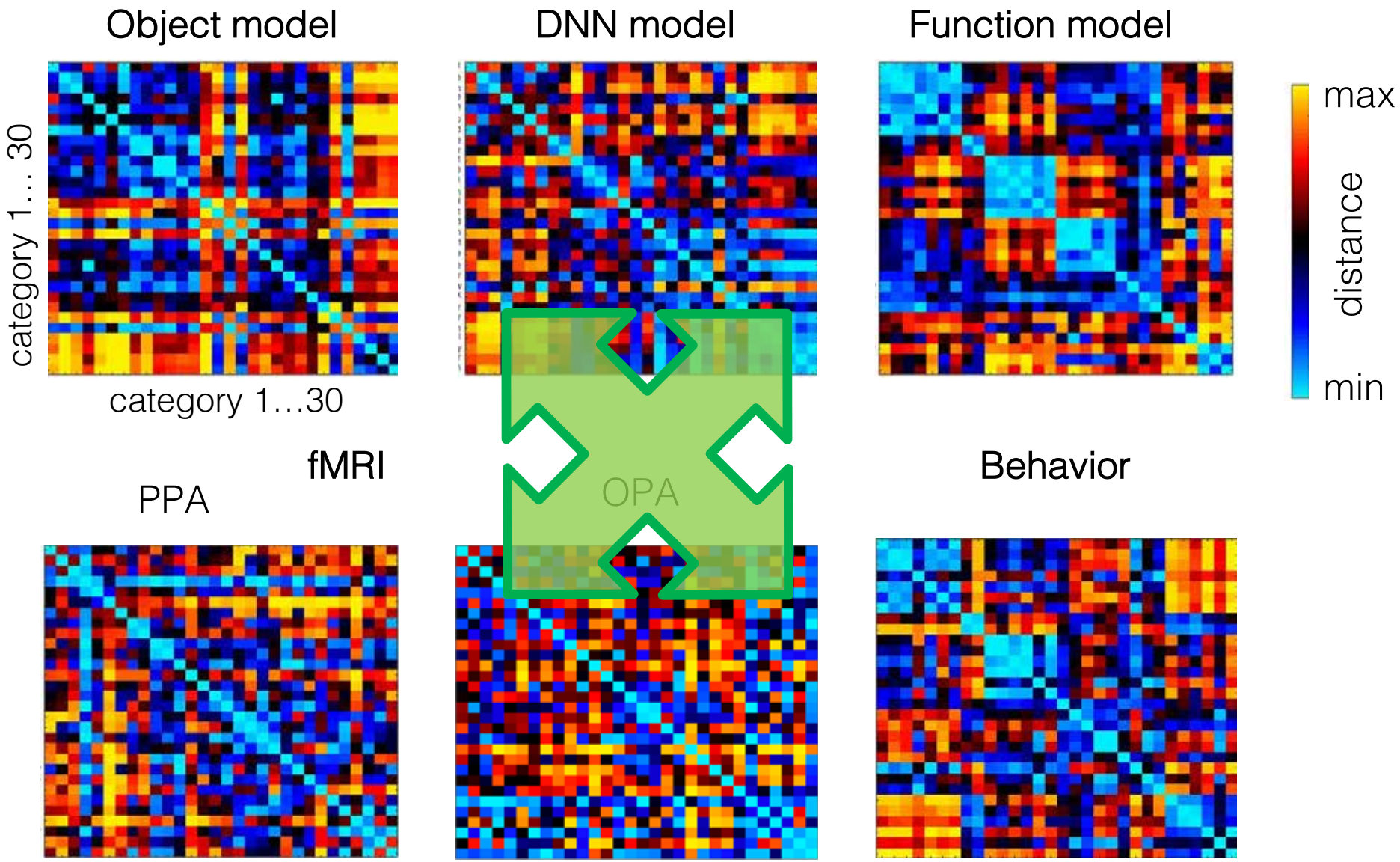
Stimuli by DNN

Stimuli by Objects

Model predictions

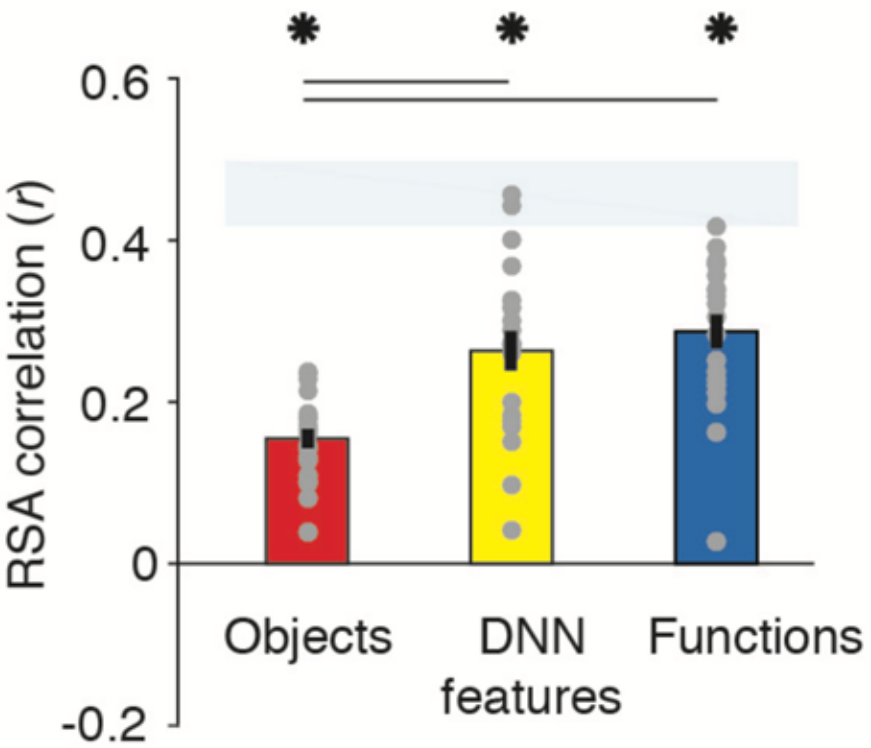


Model predictions

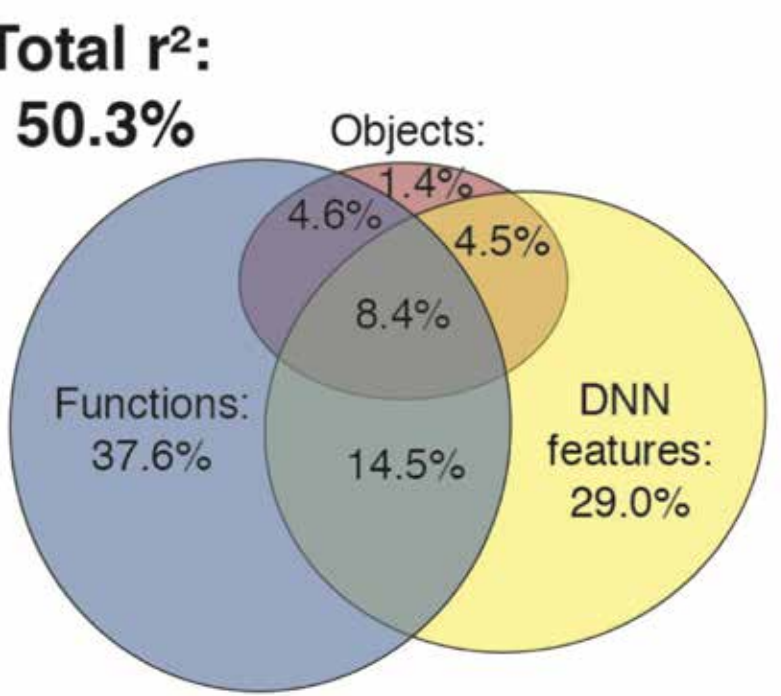


Model correlations with behavior

Correlations



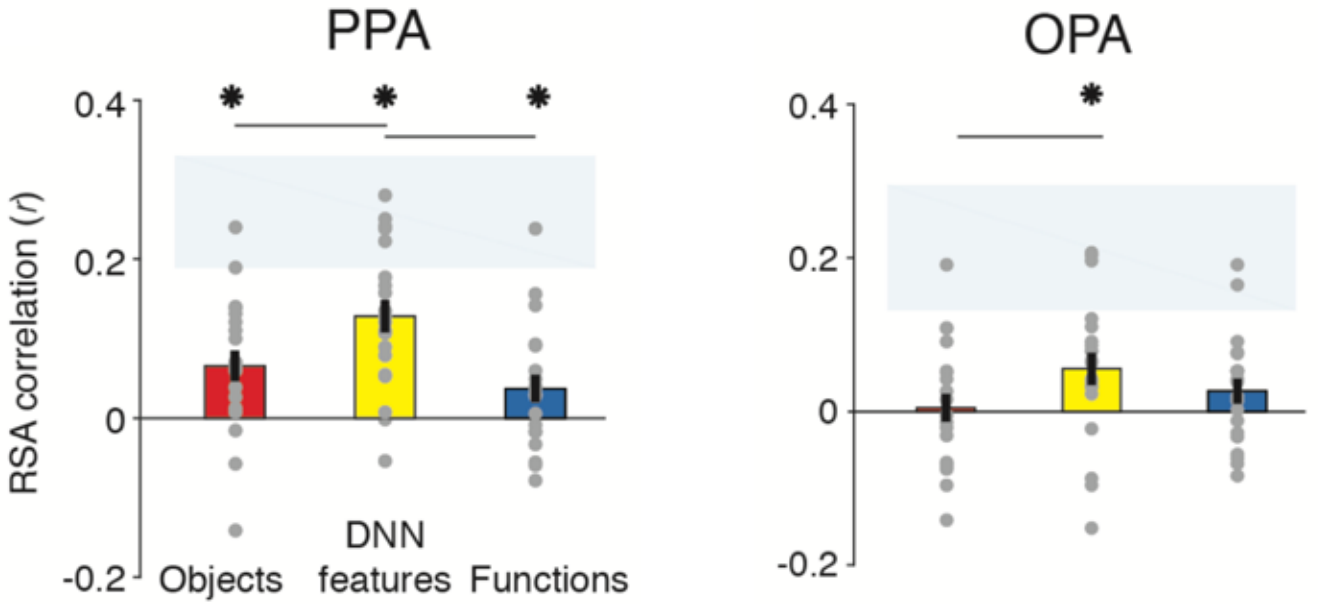
Variance partitioning



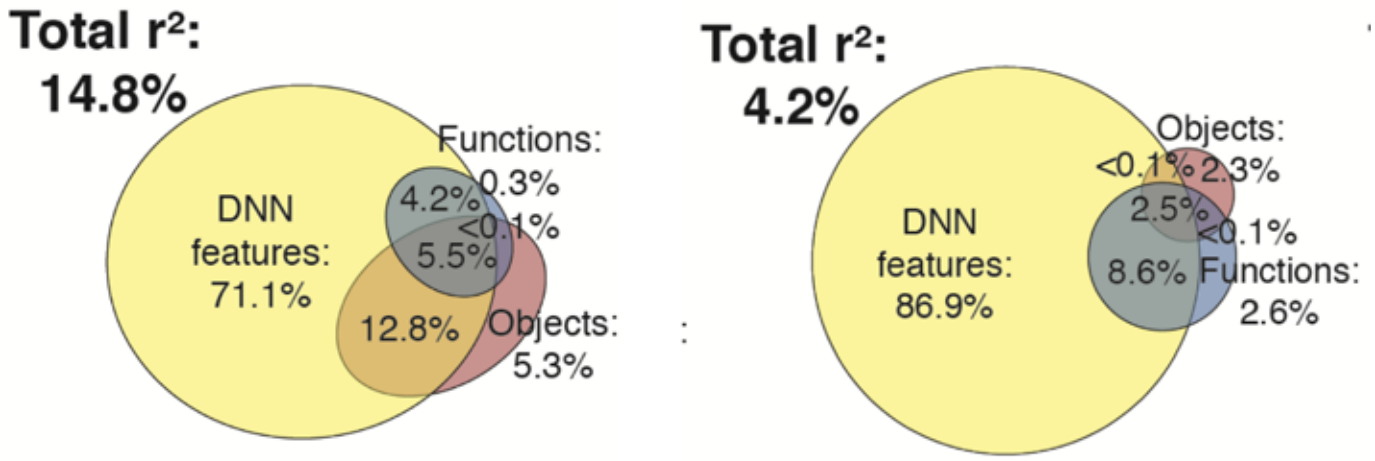
Groen, Greene, Baldassano, Fei-Fei, Beck & Baker, 2018, eLife

Model correlations with brain

Correlations



Variance partitioning

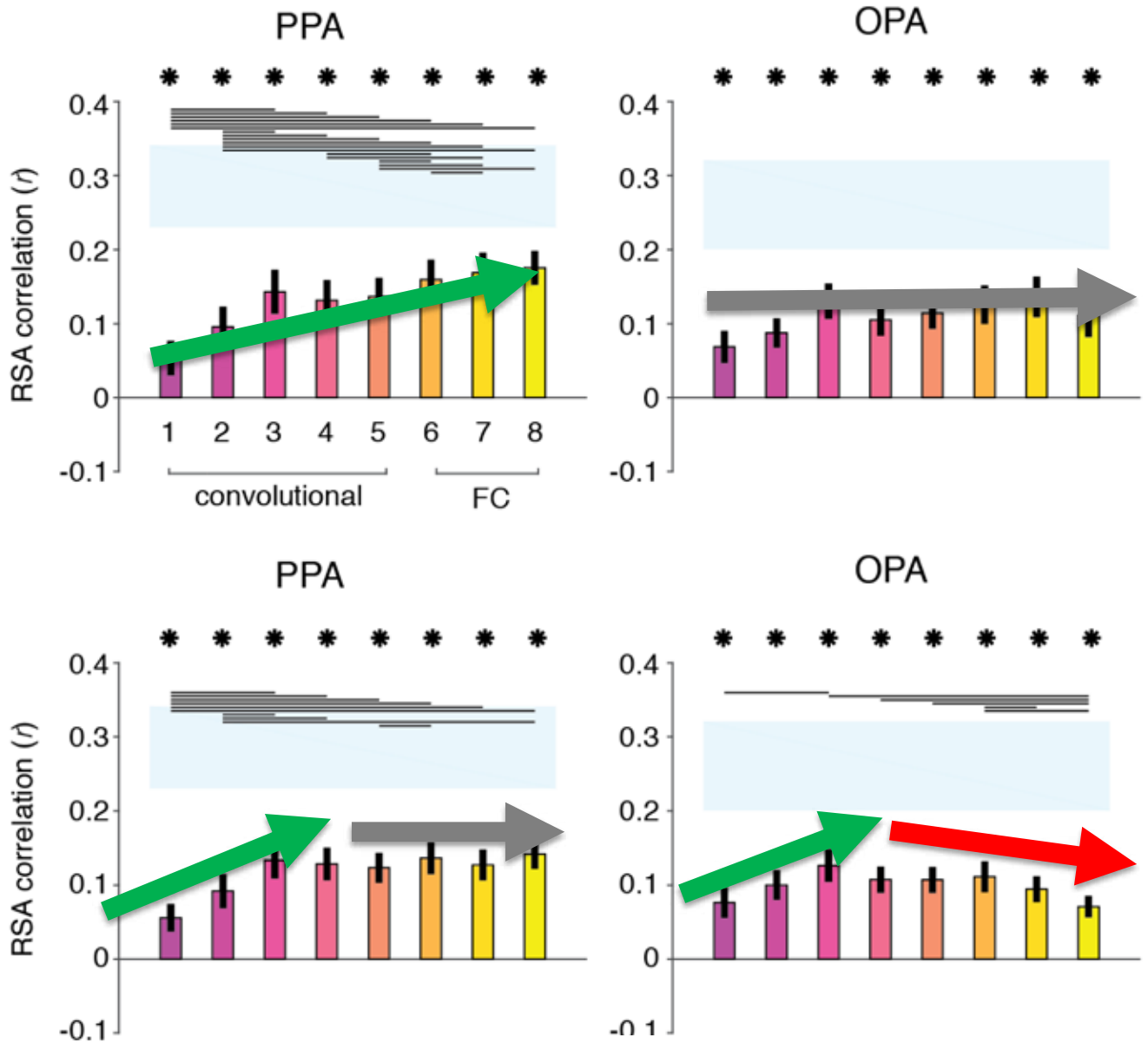


Groen, Greene, Baldassano, Fei-Fei, Beck & Baker, 2018, eLife

DNN correlations in scene cortex

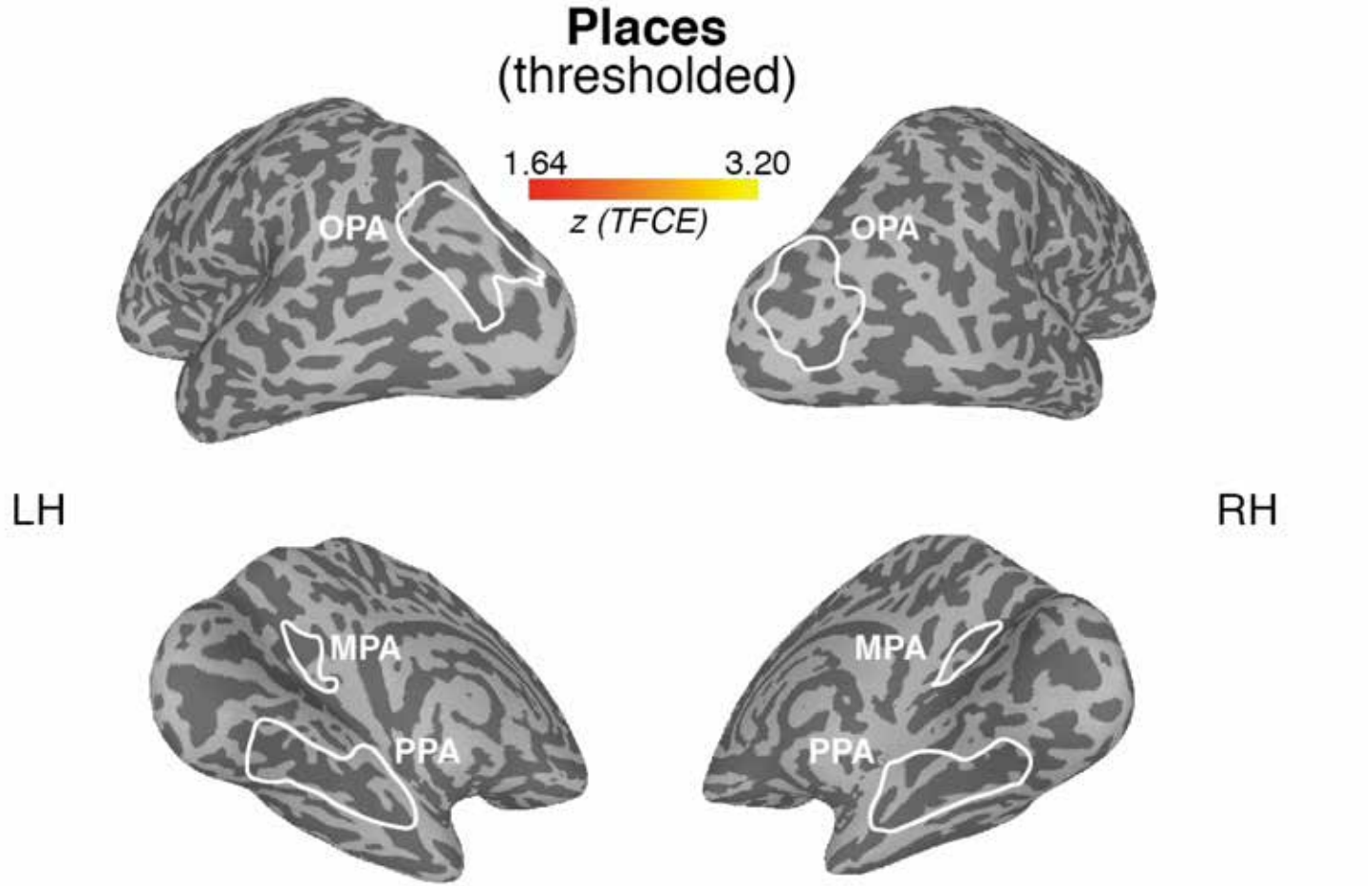
Object-trained
(ReferenceNet)

Scene-trained
(Places 205)



Groen, Greene, Baldassano, Fei-Fei, Beck & Baker, 2018, eLife

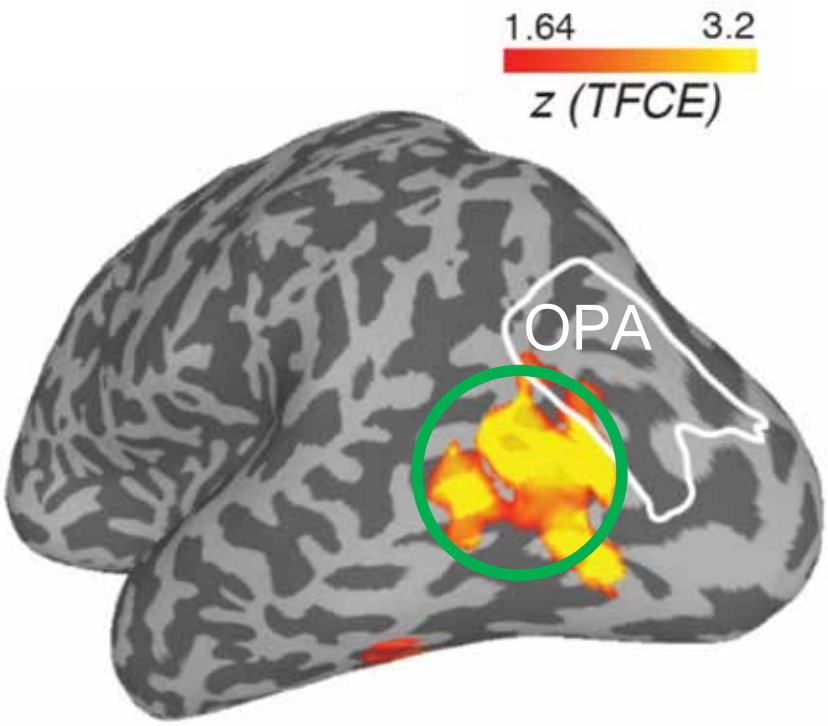
DNN correlations in scene cortex



Groen, Greene, Baldassano, Fei-Fei, Beck & Baker, 2018, eLife

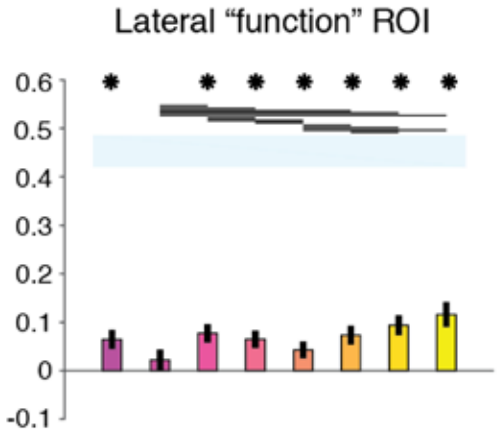
Where are the functions?

Unique correlations with function model outside scene cortex

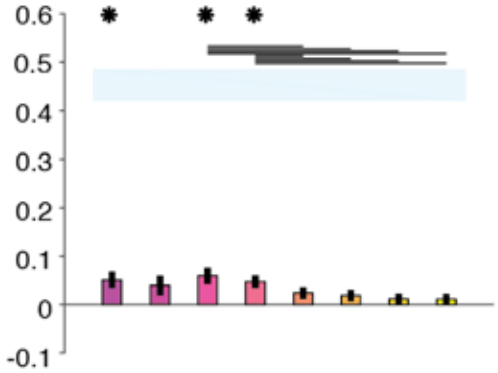


LOTC: Action perception
(Lignau & Downing, 2015)

Object-trained
(ReferenceNet)



Scene-trained
(Places 205)

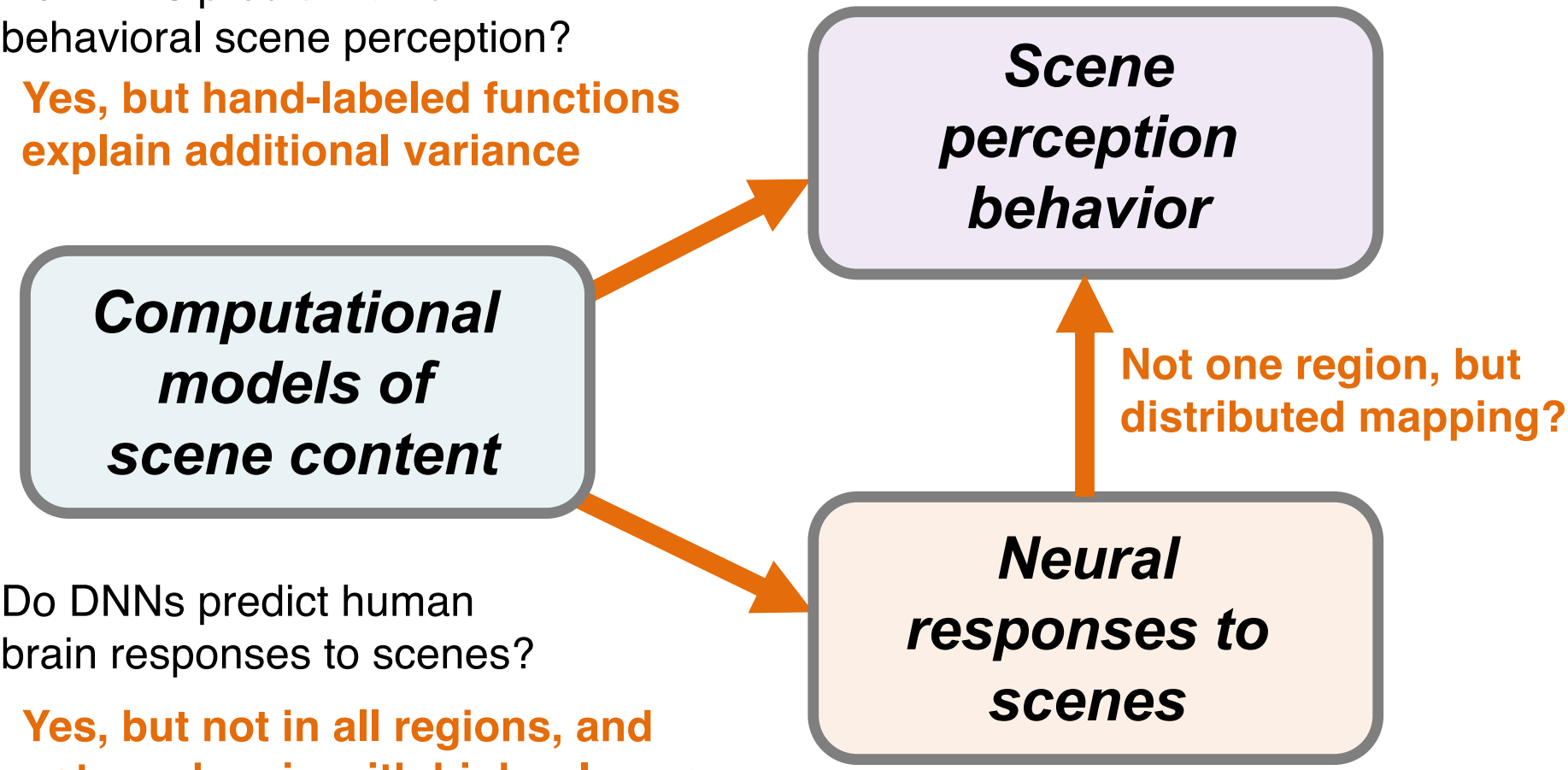


Groen, Greene, Baldassano, Fei-Fei, Beck & Baker, 2018

Understanding scene perception

Do DNNs predict human behavioral scene perception?

Yes, but hand-labeled functions explain additional variance



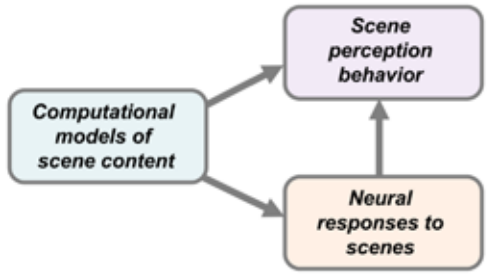
Do DNNs predict human brain responses to scenes?

Yes, but not in all regions, and not much gain with higher layers

Outline

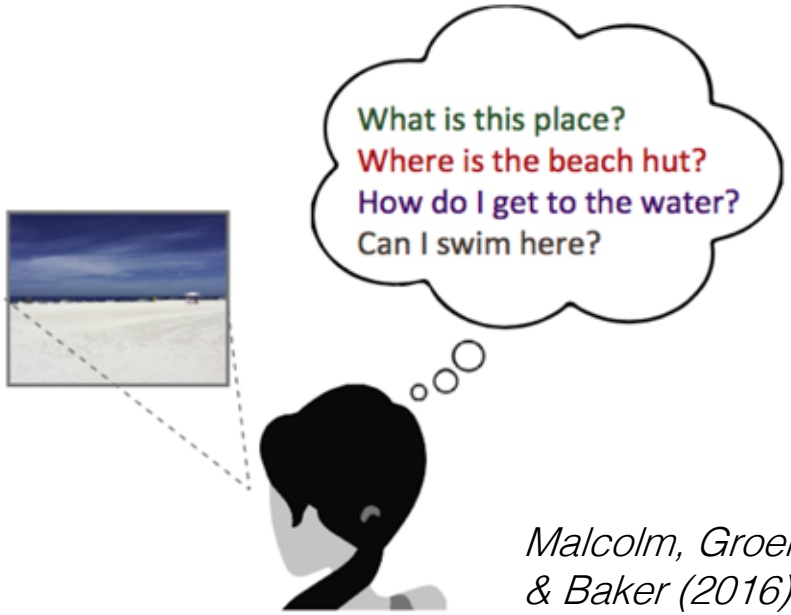
- Scene vs. object perception
- fMRI study 1: objects-in-context
- fMRI study 2: comparing multiple models
- How to move forward?

How can we close the loop?



1. DNNs trained on other tasks than object or scene recognition

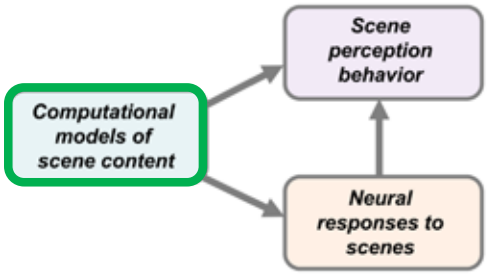
Recognition
Visual search
Navigation
Action

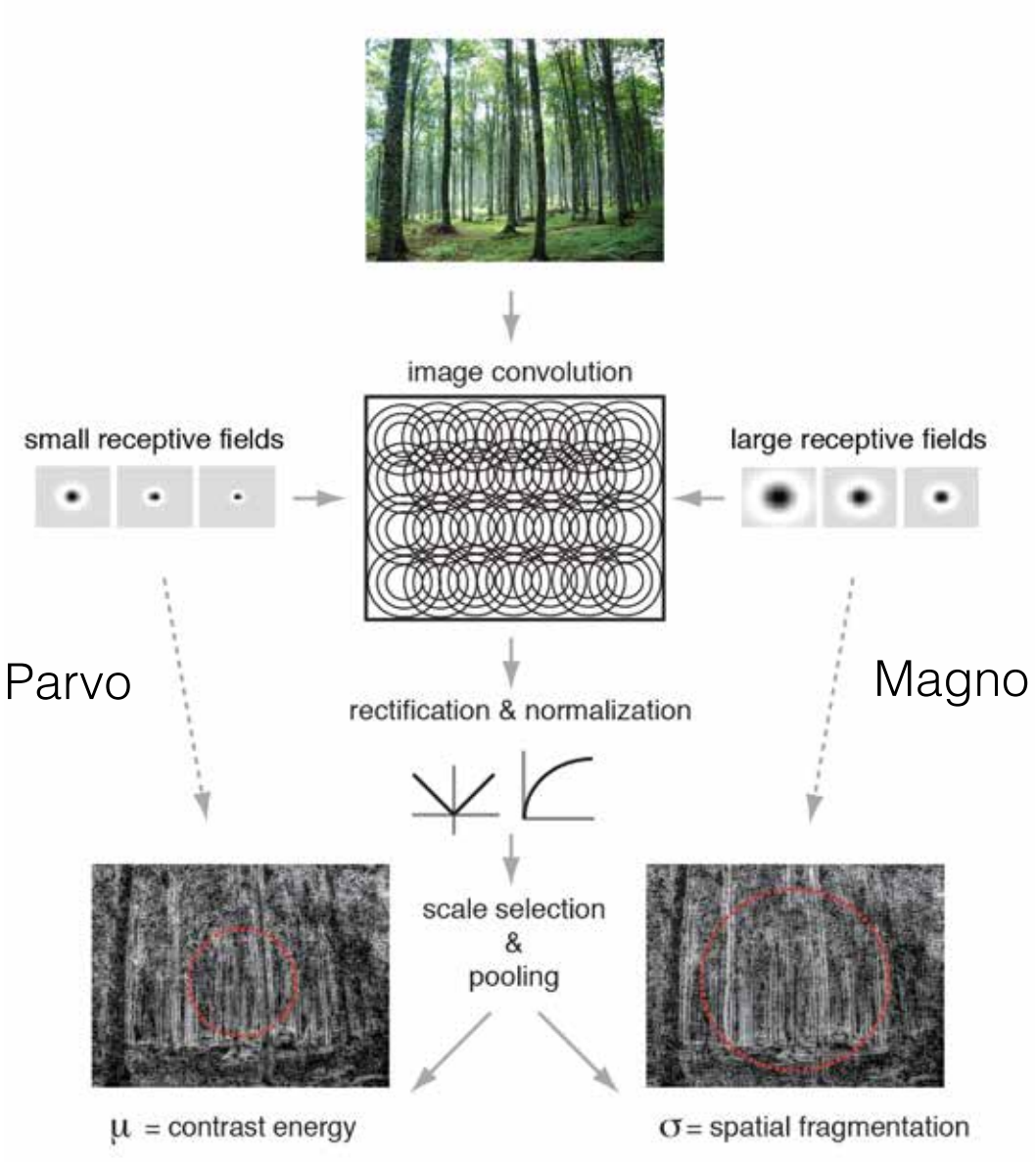


Malcolm, Groen & Baker (2016) TICS

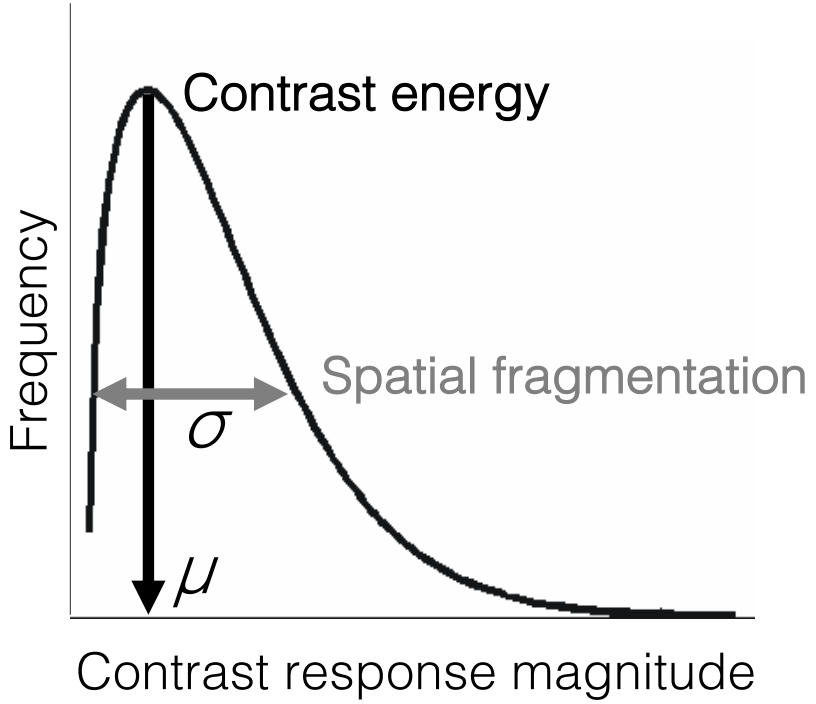
2. Other types of DNNs (e.g. RNNs), conceptual models
3. Use simpler and more explicit computational models

Neurophysiologically plausible model of scene statistics



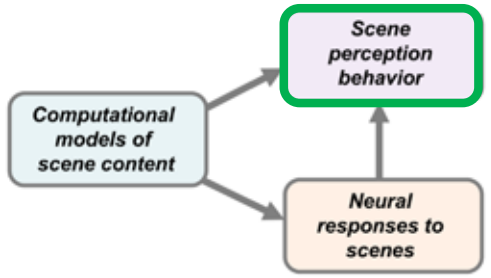


Population response to local contrast

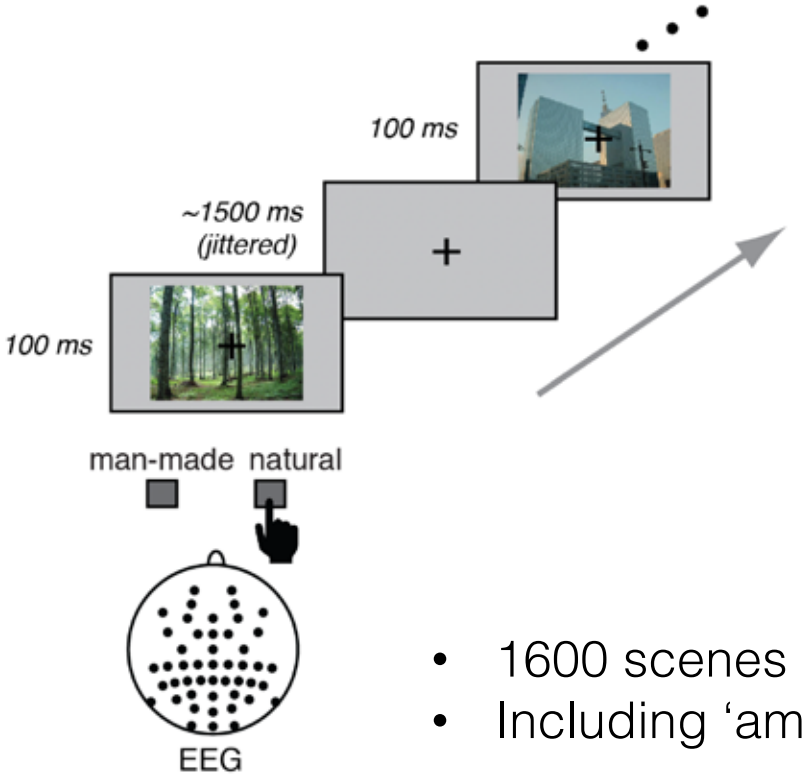


Scholte et al., (2009);
Groen et al., (2012a/b, 2013, 2017)

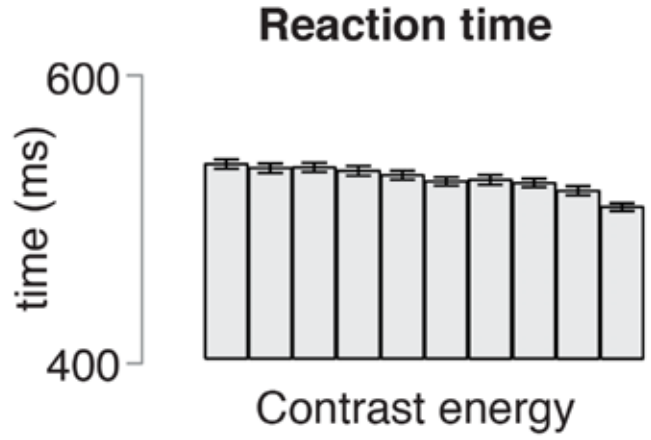
Model predicts behavioral categorization

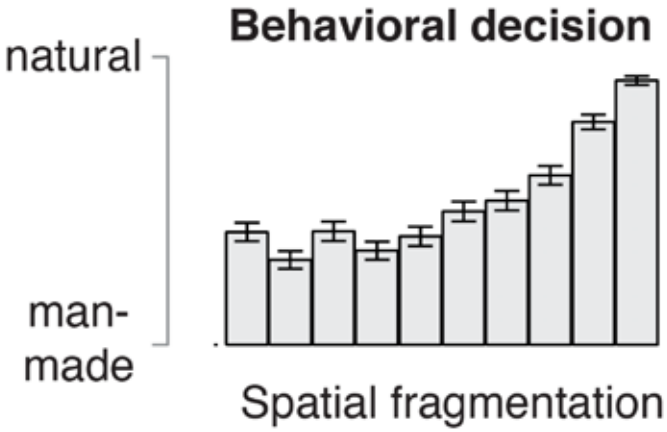


Man-made/natural categorization task



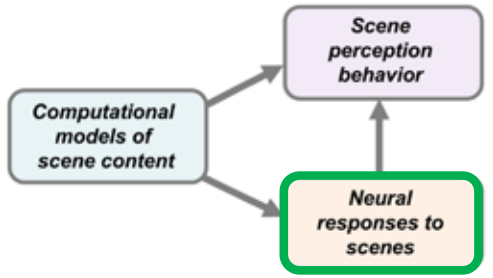
- 1600 scenes
- Including 'ambiguous'

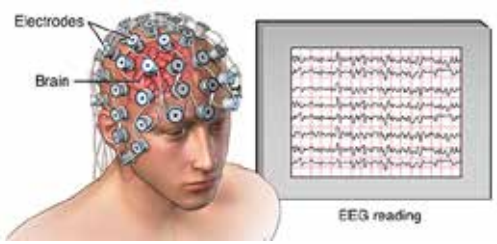


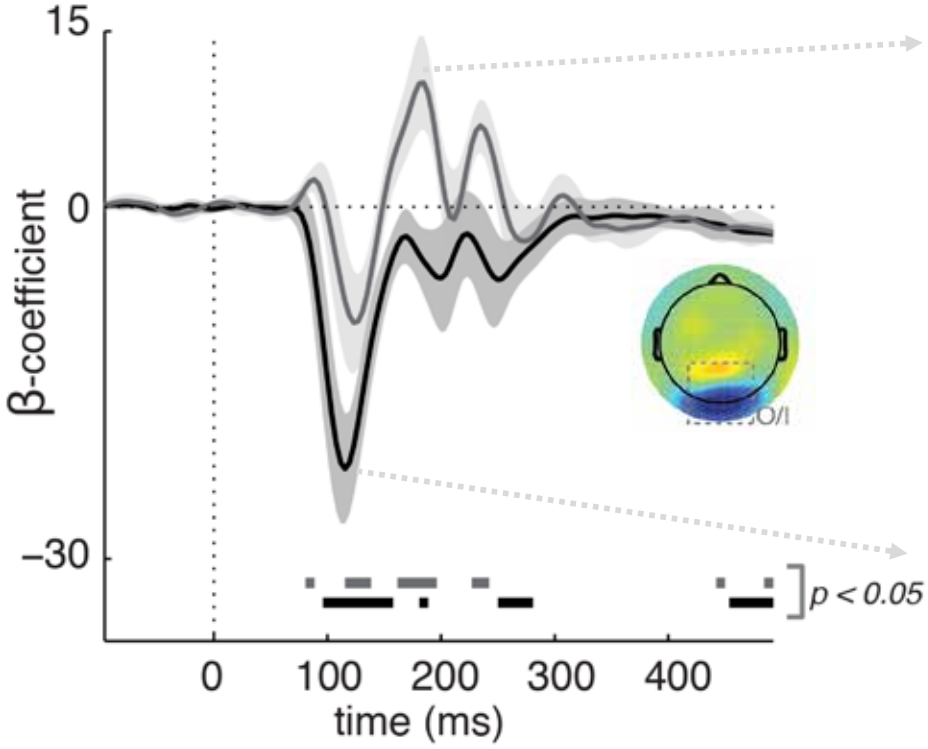


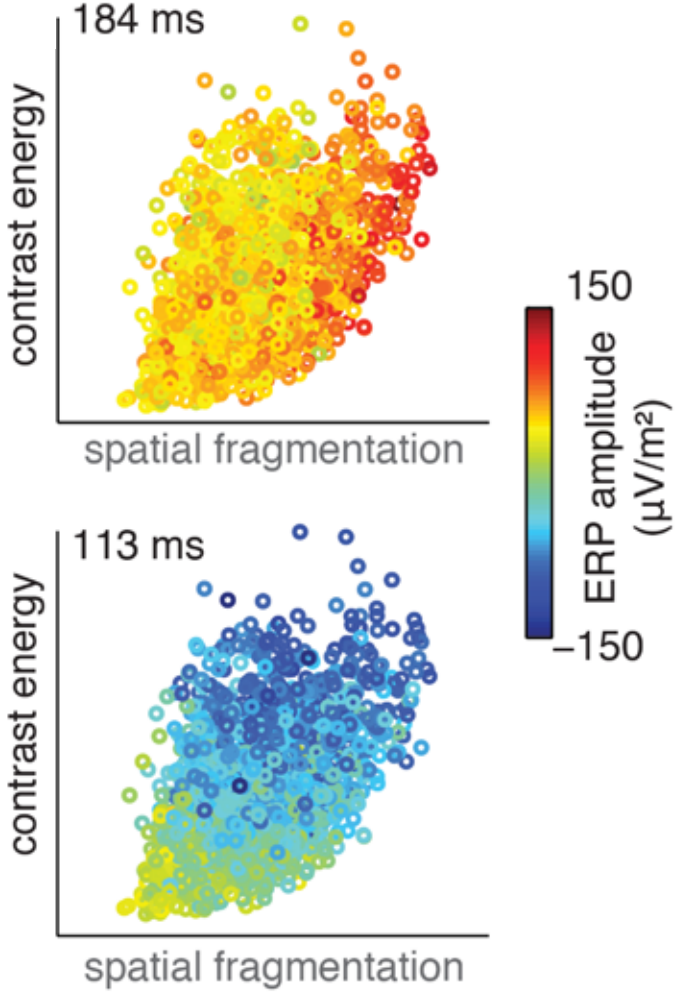
Groen et al., (2013) J Neurosci

Model predicts EEG responses…



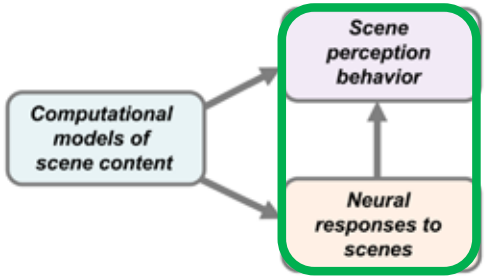




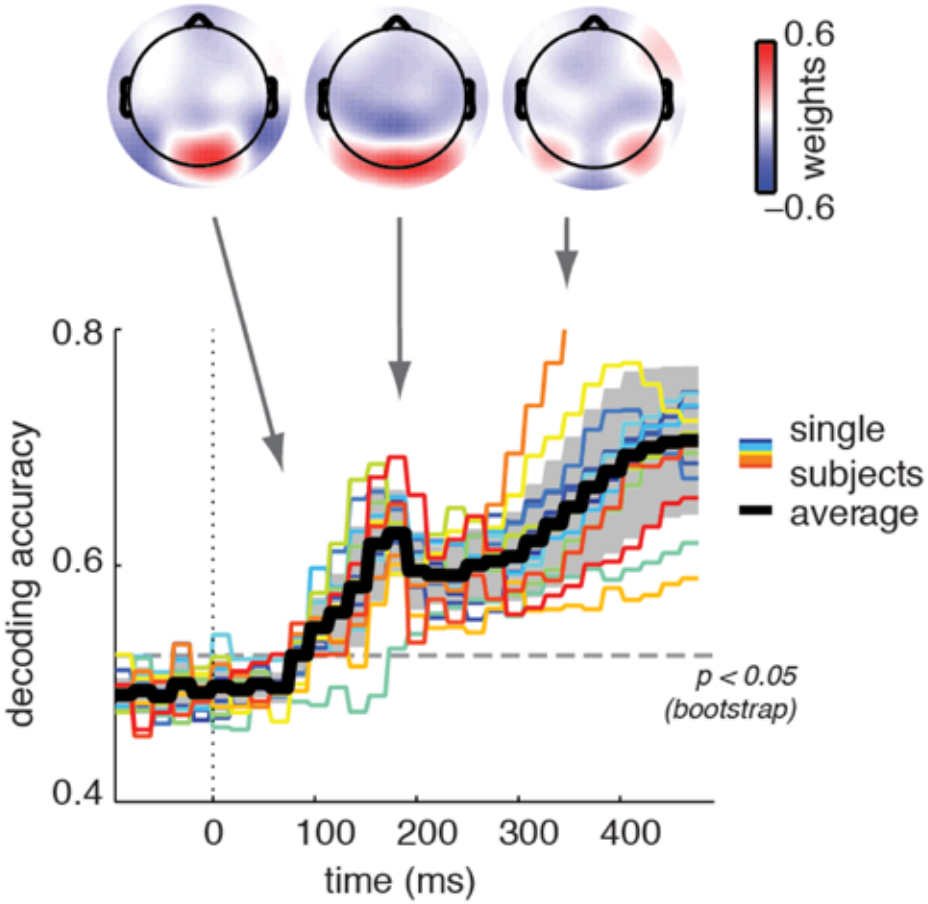


Groen et al., (2013) J Neurosci

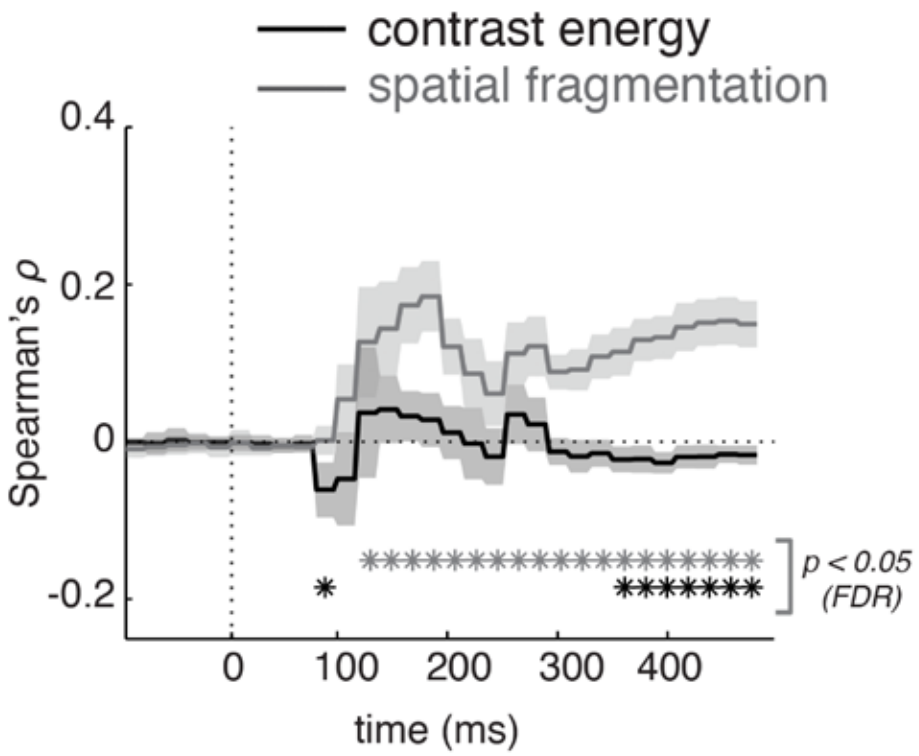
.. and decoding of behavior from EEG



Decoding accuracy

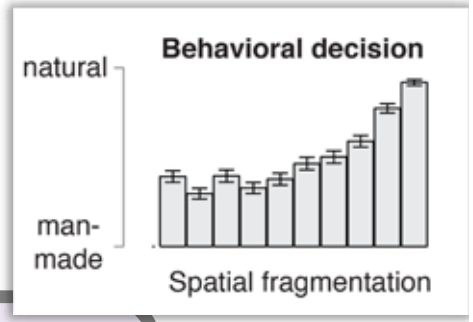


Correlation with discriminating component



Groen et al., (2013) J Neurosci

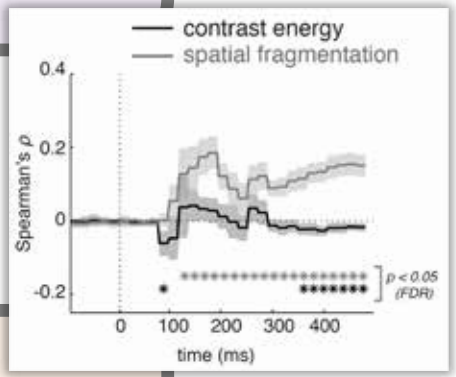
"Closed loop"

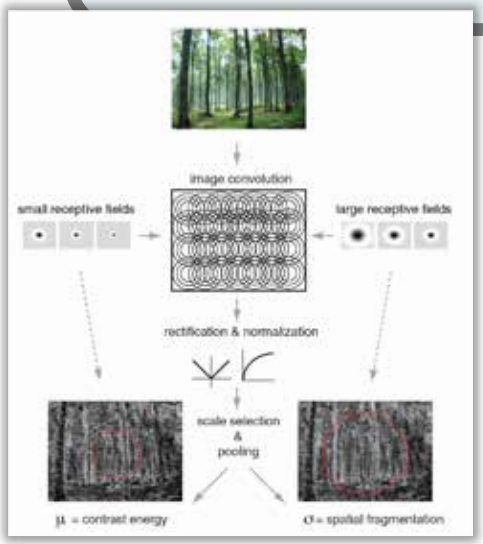


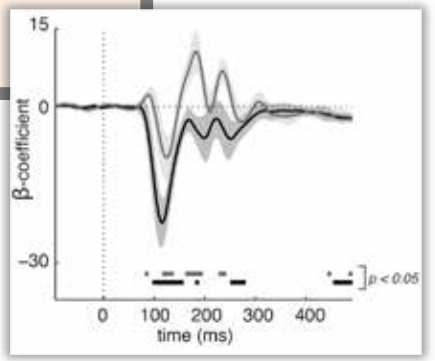
Scene perception behavior

Computational models of scene content

Neural responses to scenes

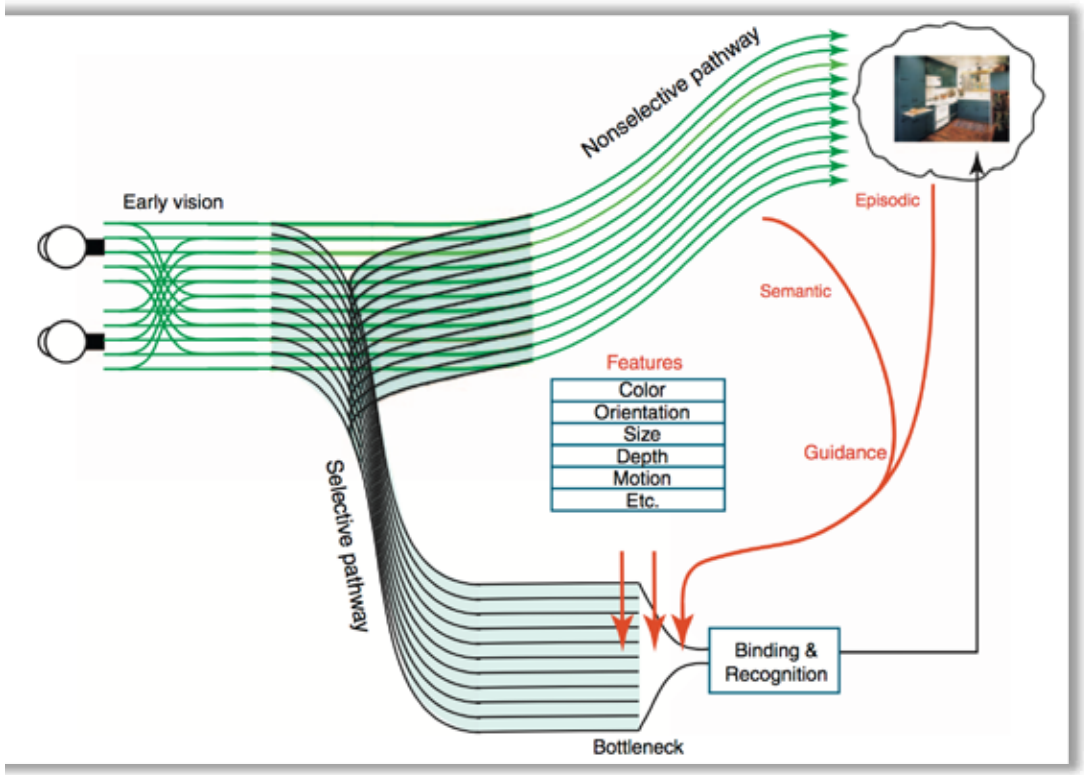






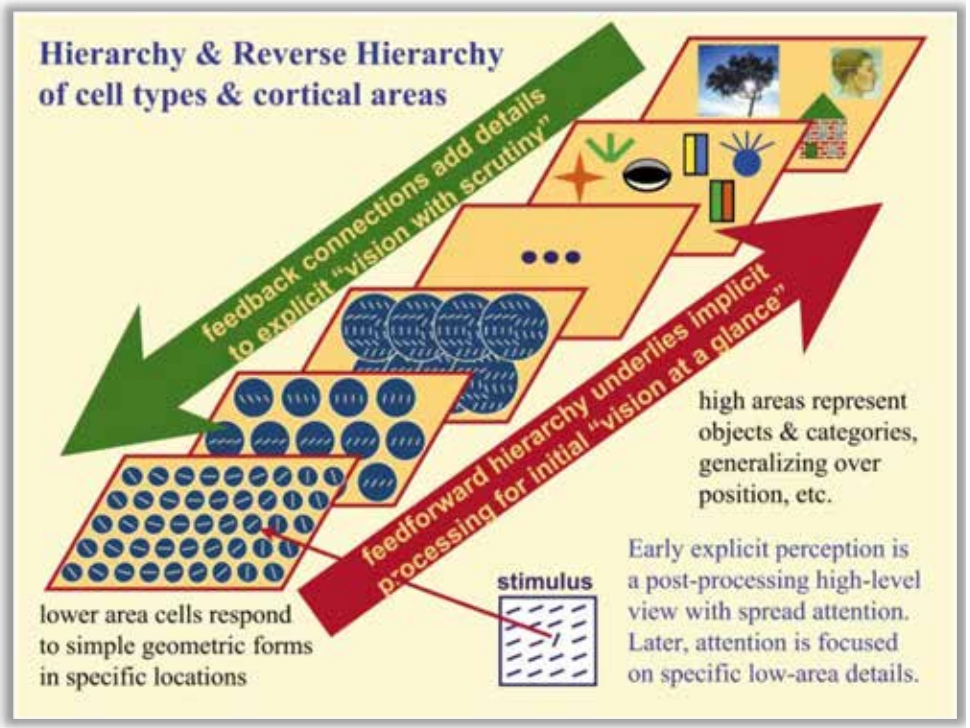
Theories of scene processing

Global vs. local



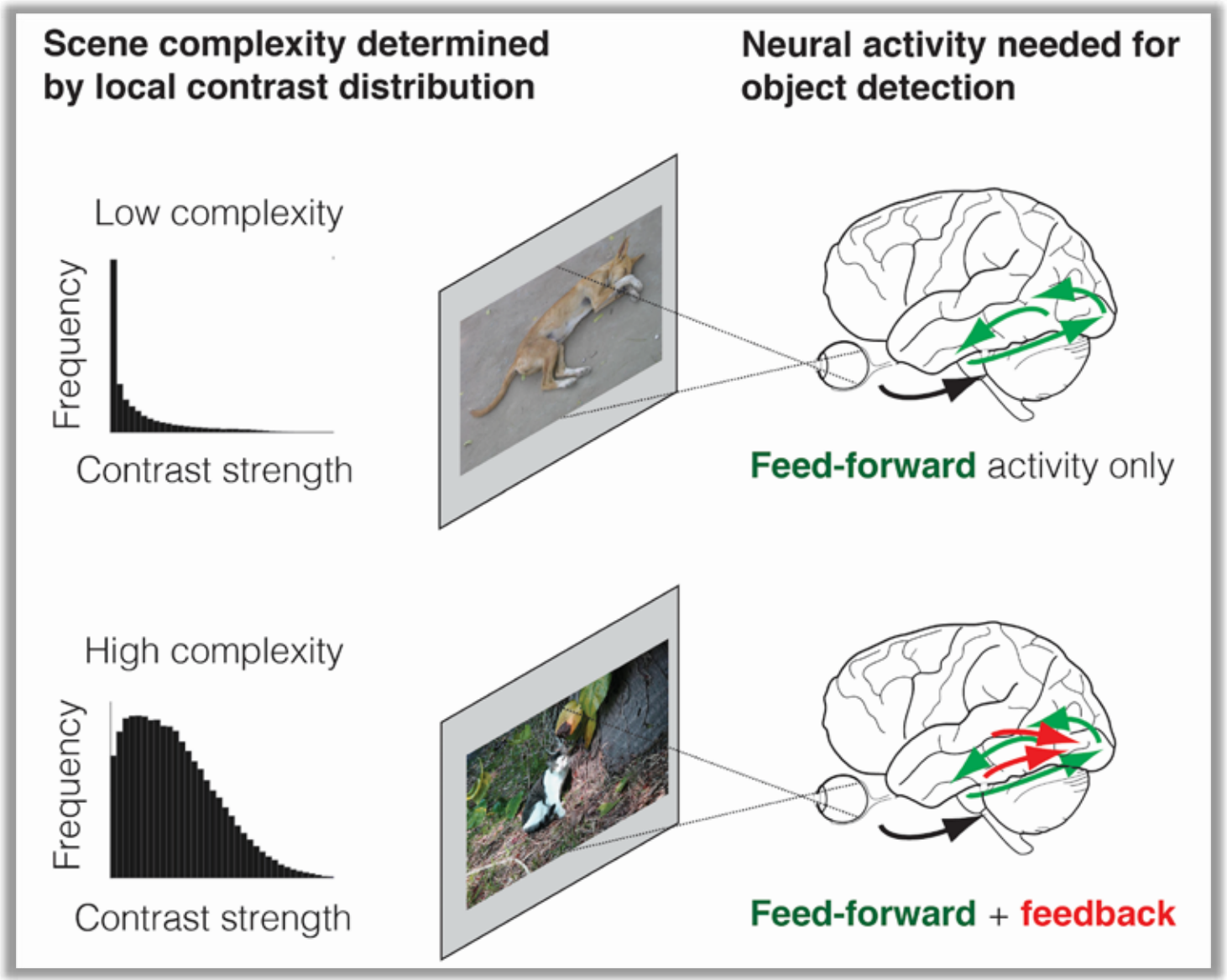
Wolfe et al., 2011

Coarse-to-fine

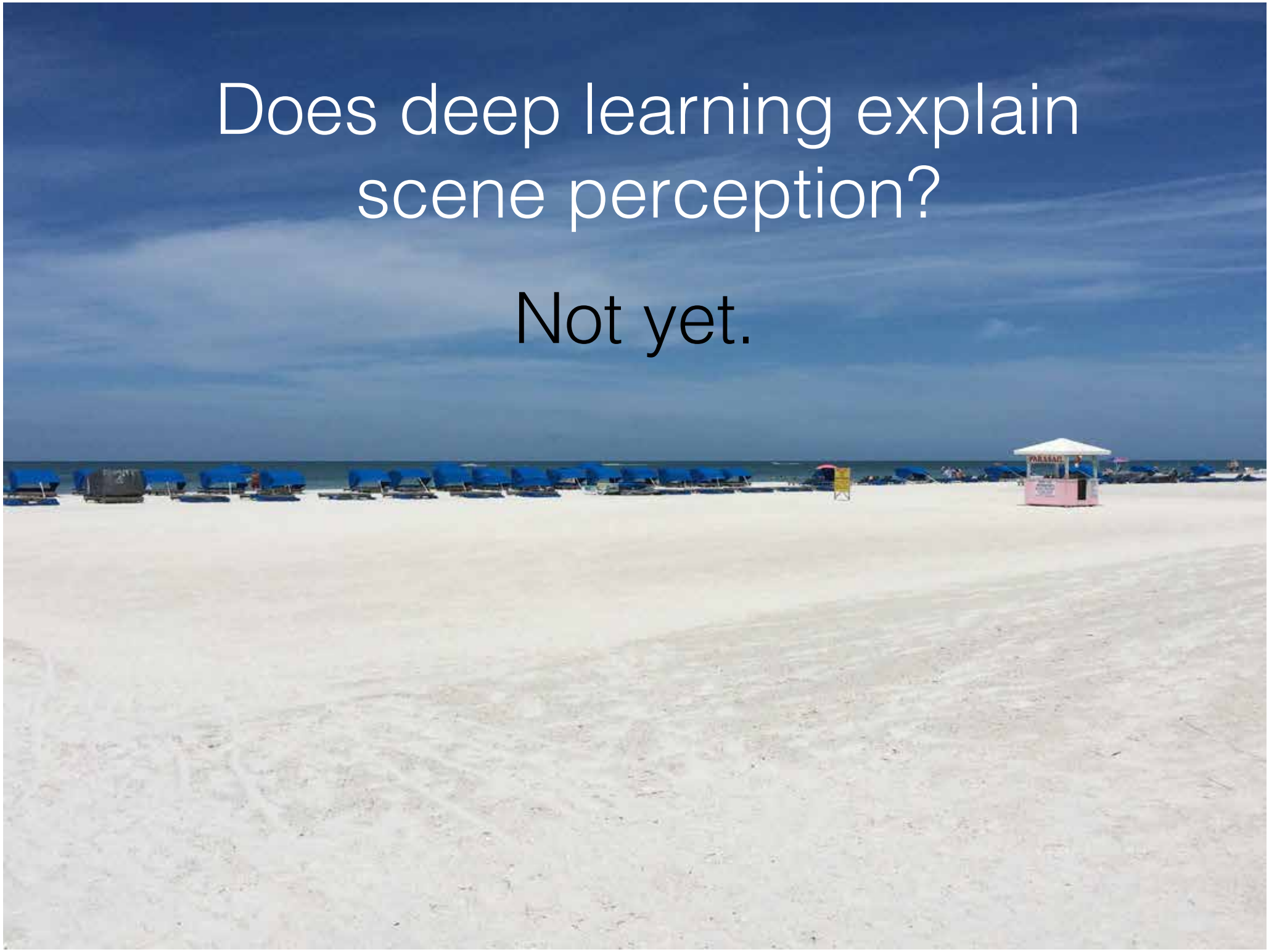


Hochstein & Ahissar 2002

Dynamic feedback for object recognition



Groen et al., 2018, PLoS Comp Biol


Does deep learning explain scene perception?
Not yet.

Why not?

- Scene perception entails more than object recognition
- High-level information not captured by object-trained DNNs (e.g. functions), is necessary to fully account for scene perception behavior
- Low-level information captured by image statistics, or early DNN layers, may be important for global scene representation
- DNNs capacity to predict (fMRI) responses in vTC does not automatically extend to predicting human scene perception

Thank you!

NYU
- Jon Winawer
- Noah Benson
- Dora Hermes
- Stephanie Montenegro
- Jing-Yang Zhou

NIH
- Chris Baker
- Brett Bankson
- Martin Hebart
- Dwight Kravitz
- Marcie King
- Edward Silson

UvA
- Victor Lamme
- Steven Scholte
- Sennay Ghebreab
- Sara Jahfari
- Noor Seijdel

- Chris Baldassano (Columbia U)
- Morgan Barense (U Toronto)
- Diane Beck (U Illionois)
- Tessa Dekker (UCL)
- Michelle Greene (Bates College)
- Assaf Harel (Wright University)
- Fei-Fei Li (Stanford University)
- Natalia Petridou & BAIR team (UMC Utrecht)
- Kandan Ramakrishnan (MIT)

NWO Netherlands Organisation for Scientific Research

NIH National Institute of Mental Health

www.irisgroen.com