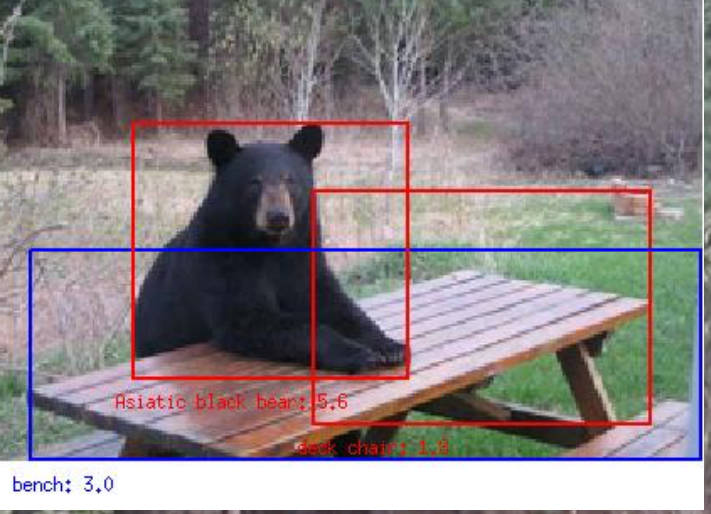
MIT
COMPUTER
VISION

6.819 / 6.869: Advances in Computer Vision

High-level vision

Instructor: Aude Oliva

Lecture TR 9:30AM – 11:00AM
(Room 34-101)



Trevor Darrell

Jitendra Malik

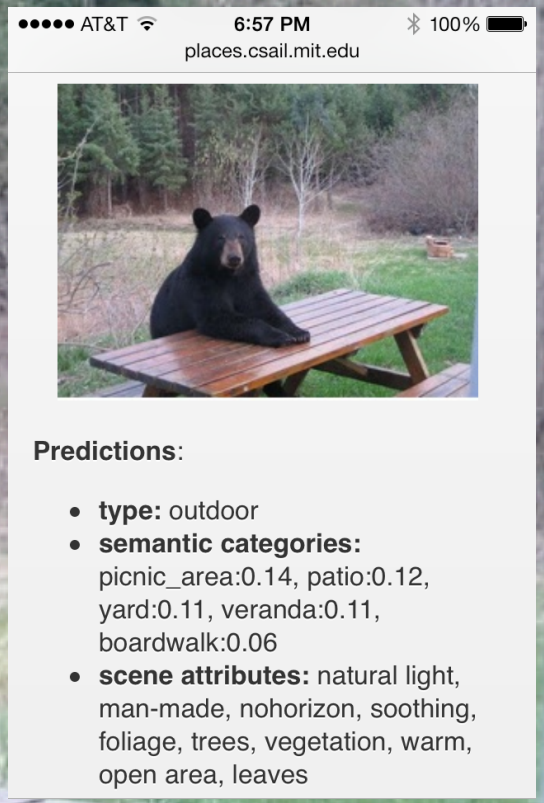
Pietro Perona

Andrew Zisserman

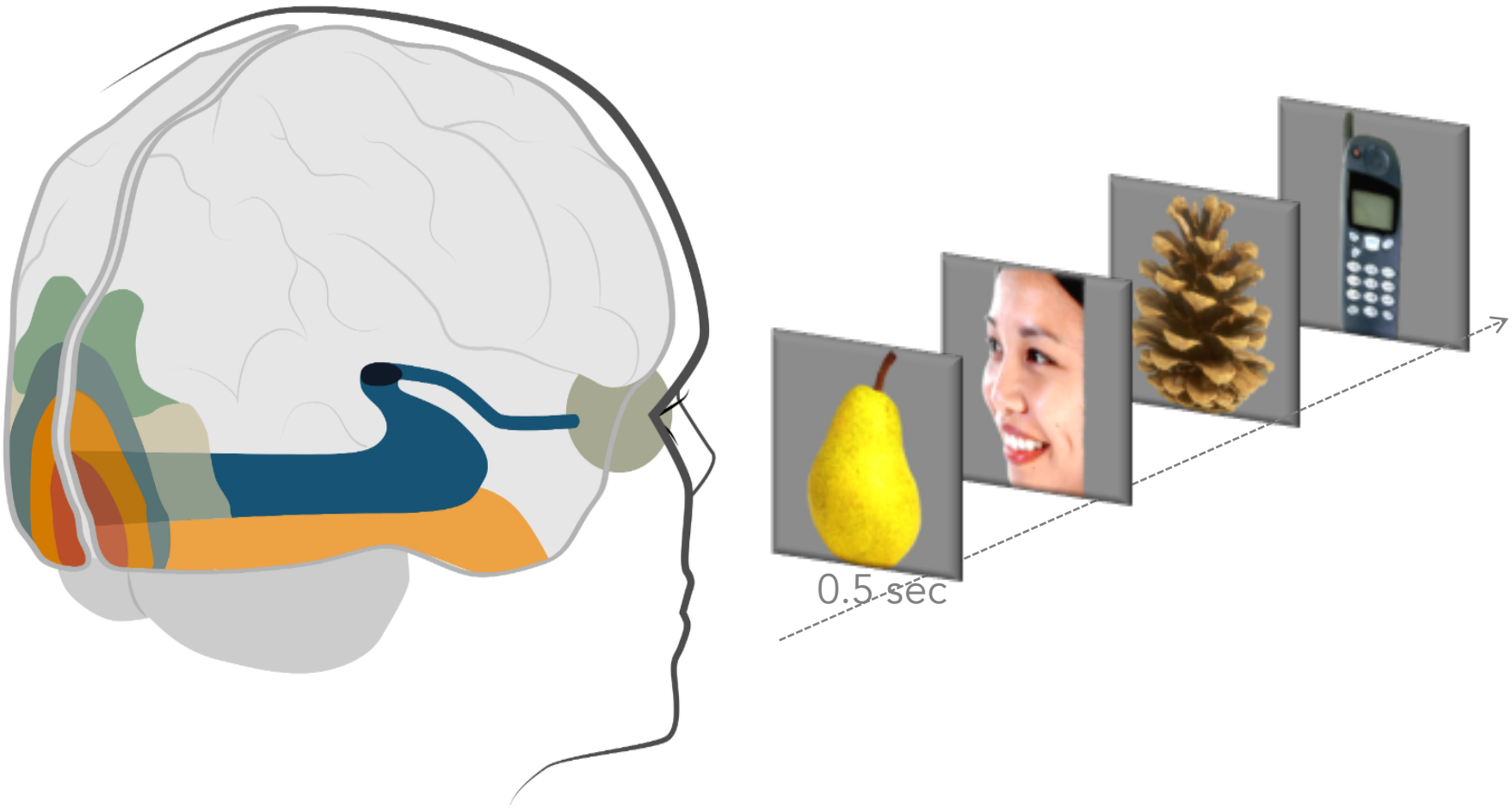
Fei-Fei Li

Aude Oliva

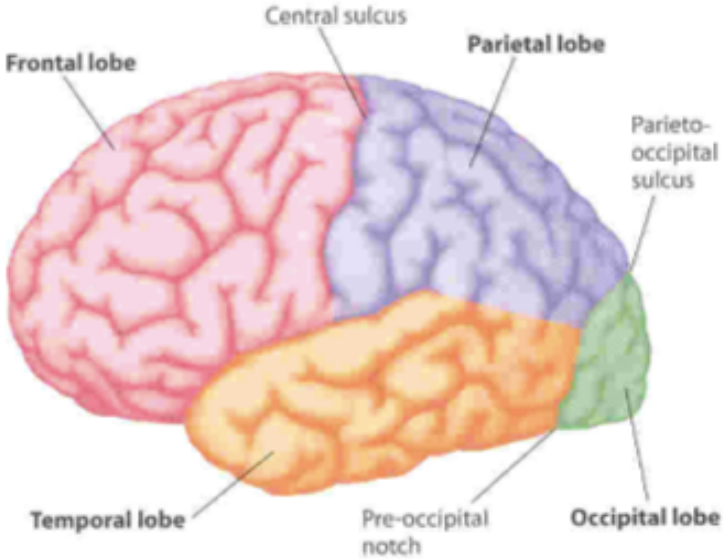
Antonio Torralba

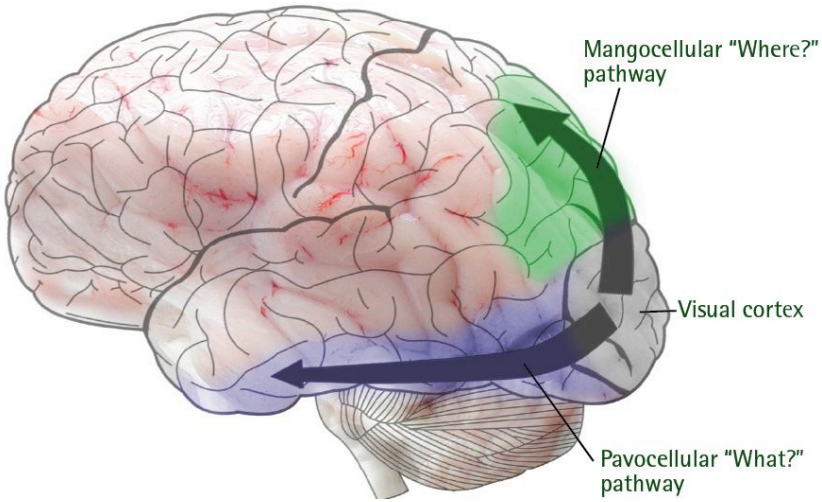


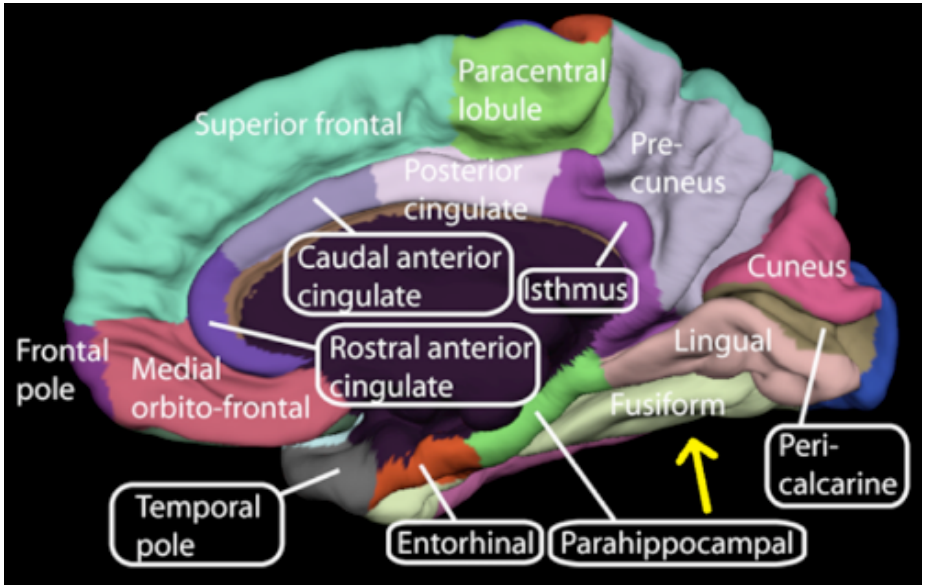
High Computing Visual Engine: Object recognition

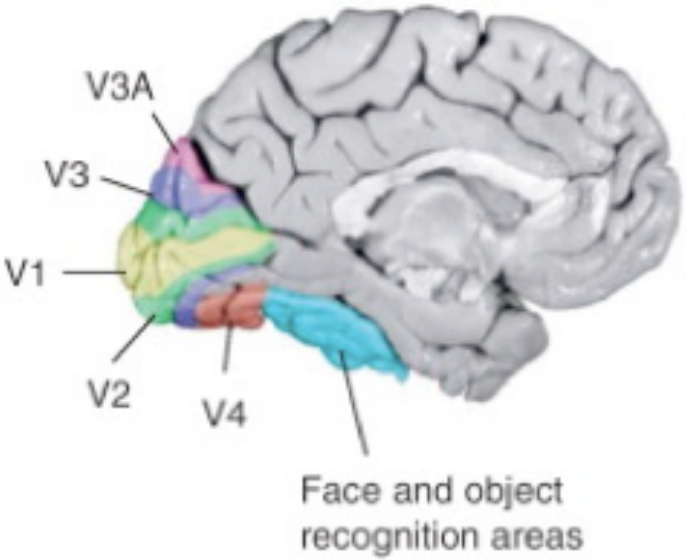


Face & Object recognition: the ventral pathway

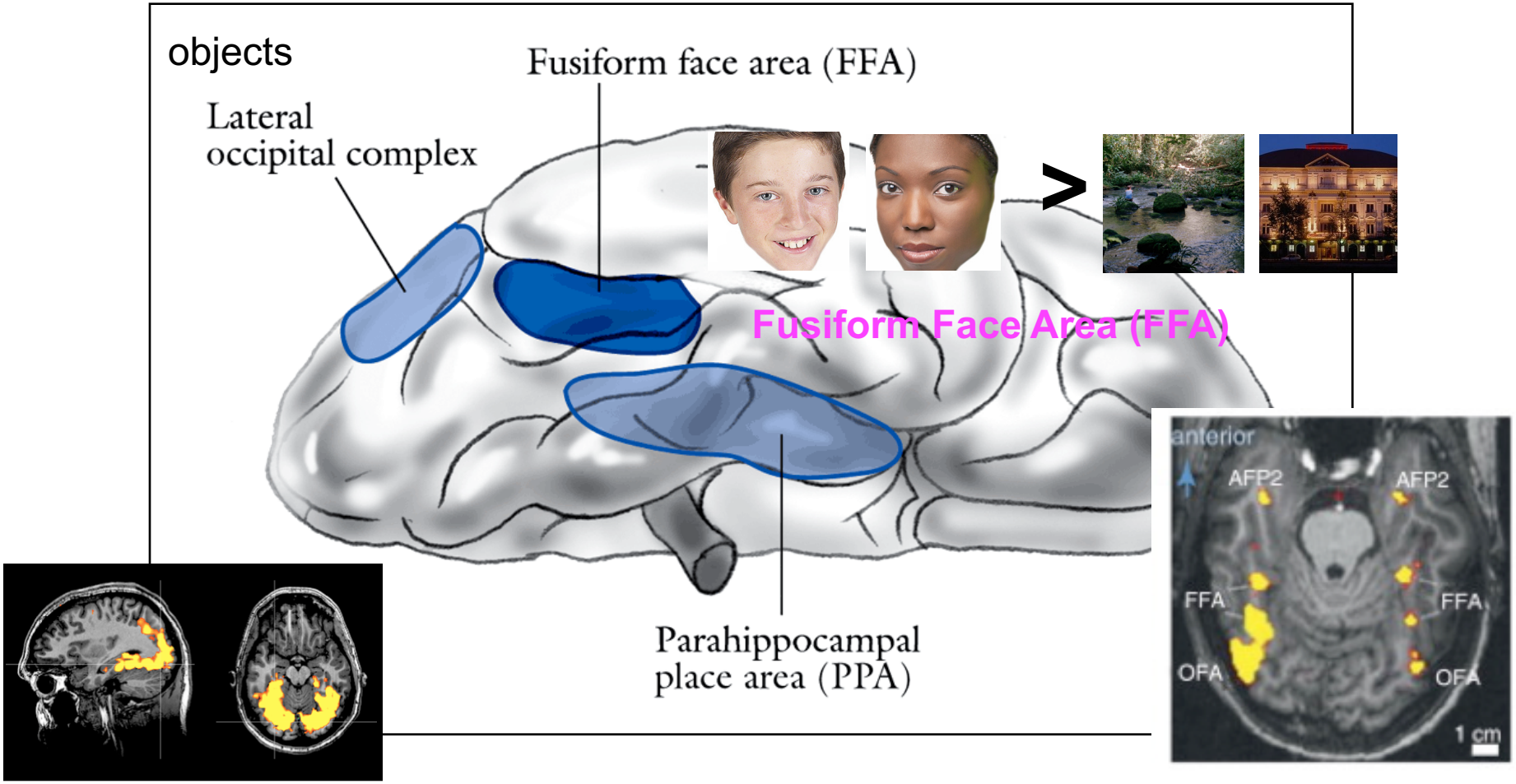








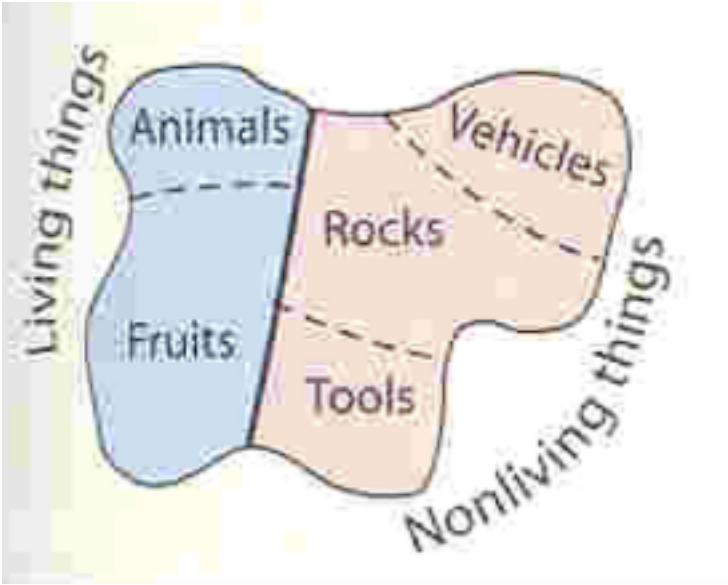
Regions of Interest are common brain sub-regions found in everyone

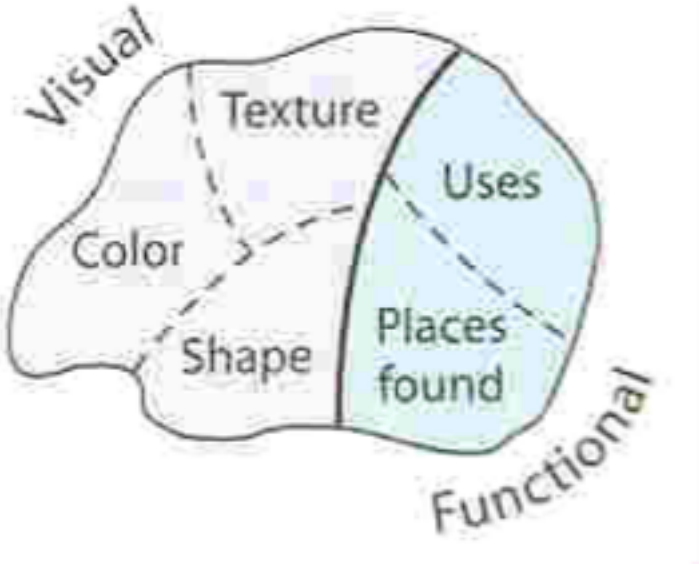


Object recognition

- Categories and Function
- Categorization levels: Basic level and entry level
- Typicality effects
- Canonical Perspective
- Prototype theory
- Object parts segmentation
- Mental Imagery
- Brain regions related to object recognition

Categories or properties ?





Organization of knowledge in the brain: Two theories

1) **A category-based approach** suggests that semantic knowledge is organized according to our categories of the world. For instance, one prominent split is between living and non living things
2) **A property-based hypothesis** is that semantic knowledge is organized according to the properties (attributes) of objects. These properties maybe visual or functional

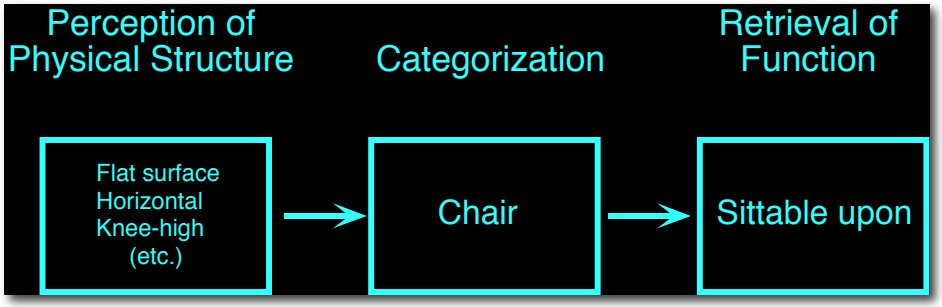
Perception of Object function

Function of an object that an observer can observe directly from its structure

Affordances

throwable

Categorization



sittable-upon

Levels of Object Categorization

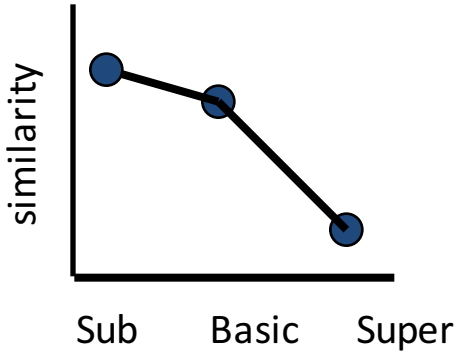
Object Categorization: superordinate level (e.g. animal), basic level (e.g. preferred level, a tiger), subordinate level (e.g. a Persian tiger).

Basic level organization: Three levels of Regularities:

Shape : Highest level categories for which members have *similar shape*

Motor action : Highest level categories for which people interact with its members using *similar motor sequences*

Common attributes: Highest level categories for which they are significant number of attributes in common between pairs of members.



If plot the degree of shape similarity, the degree of similar motor actions, and the number of common attributes: steep decline between the basic and superordinate levels.

Rosch E, M., CB, Gray MD, Johnson DM, Boyes-Braem P,. (1976)
Basic Objects in Natural Categories. Cognitive Psychology 8:382-439.

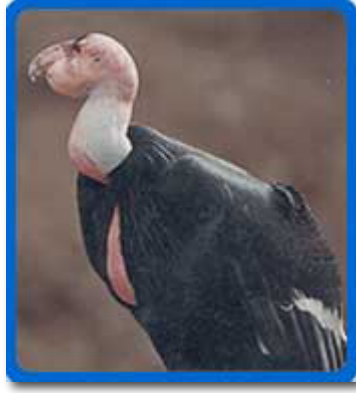
Levels of Categorization

Table 7.3. ***Examples of Nested Category Structures***

Superordinate Level	Basic Level	Subordinate Level	
Musical instrument	Guitar	Folk guitar	Classical guitar
	Piano	Grand piano	Upright piano
Fruit	Peach	Freestone peach	Cling peach
	Grapes	Concord grapes	Green seedless grapes
Tree	Maple	Silver maple	Sugar maple
	Birch	River birch	White birch
	Oak	White oak	Red oak

- Rosch et al (1976) found that when people are shown pictures of objects, they identify objects at a basic level more quickly than they identified objects at higher or lower levels.
- Objects appear to be recognized first at their basic level, and only afterwards they are classified in terms of higher or lowers level categories.
- **Entry-level** (Jolicoeur et al., 1984) corresponds to either basic level or subordinate level depending on the object category and exemplar frequency (a *taxi* instead of car)

Entry-level categories

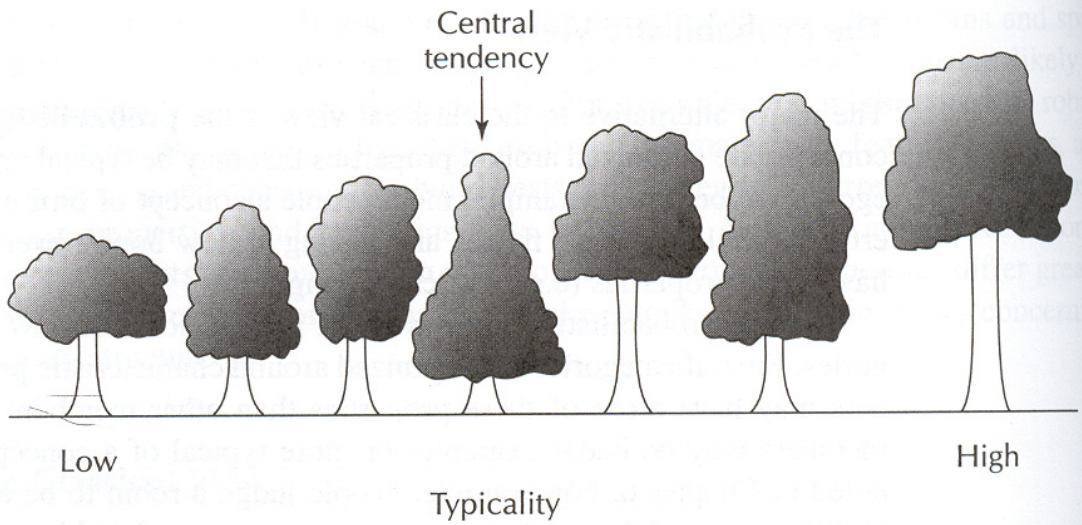


- The level of which objects are first categorized perceptually
- Basic level: “bird” is the basic level category for every bird
- **Entry-level**: “Bird” is the entry level category for <u>typical birds</u>, but subordinate names (chicken, ostrich) are the entry level categories for atypical birds

Typicality: How common an item is

a member of a category

Typical exemplar: a representation of the average (or central tendency)
The <u>representation of a category vary with experience</u>, so does the “typical” exemplar



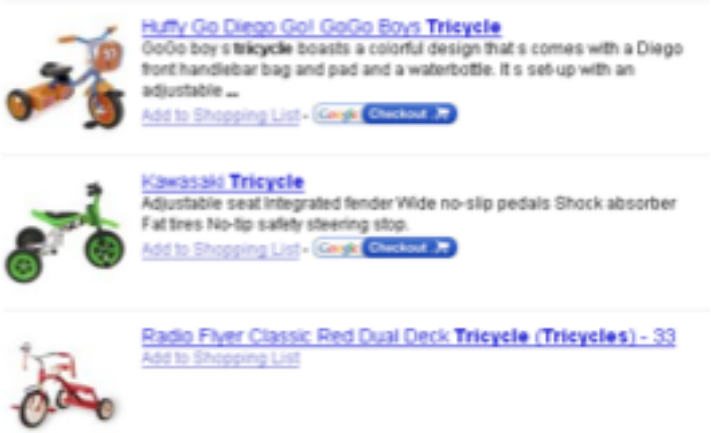
Typicality and central tendency illustrated for different kinds of trees varying in height. If typicality is based on central tendency, then trees of average height should be rated as the most typical. Tree experts, however, rate the tallest trees as most typical.

For tree experts, the internal structure of the category tree is organized around attributes of of weediness and height. Thus, the best exemplars of trees are trees of high heights (and free of weedy characteristics like having weak limbs and being susceptible to disease).

Canonical Perspective

The “best,” most easily identified view of an object.
(Palmer, Rosch & Chase, 1981)

Slide by Palmer



Canonical Viewpoint:

¾ view – or side view

View with all the parts seen

Invariance in object recognition

(1) Objects can be decomposed in parts

(2) Each object has a distinct configuration of parts

(3) Different viewpoint of the same object must show the same configuration of parts.

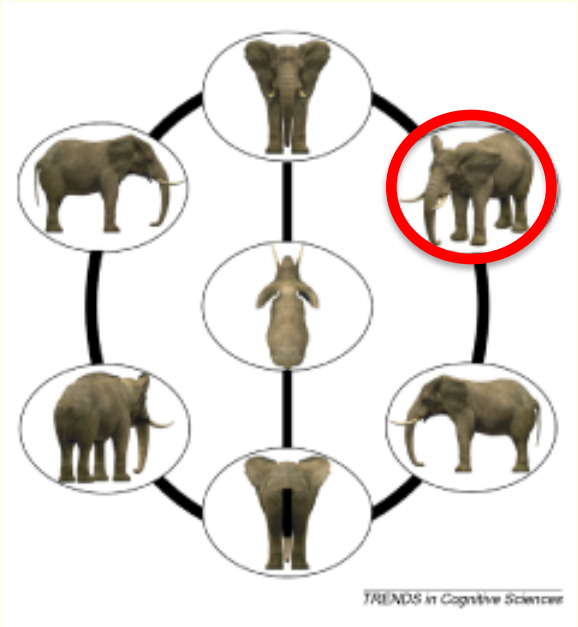


Fig. 1. A viewpoint-dependent object representation of an elephant. Because individual views will only generalize over a limited range of positions, the representation must include multiple views of the same object, each showing different visible features.

Prototype Theory

- Categories are formed on the basis of characteristics features, which describe the typical model of the category
- Whereas a **defining feature** is possessed by **every instance of a category** (e.g. a dog has 4 legs), a characteristic feature need not be (e.g. the ability to fly is typical to birds, but is not a defining feature of a bird).
- According to prototype theory, because an ostrich cannot fly, it seems *less birdlike* than a robin which can fly.

Prototype Theory

- According to the prototype view, an object will be classified as an instance of a category if it is sufficiently similar to the prototype.
- **Evidence for Prototype**:
- **Typicality ratings**: how good are robins as an example of birds
- **Production order of exemplars**: Name all the kinds of bird you can think of
- **Time to verify categorical statements**: True or false: a robin is a bird

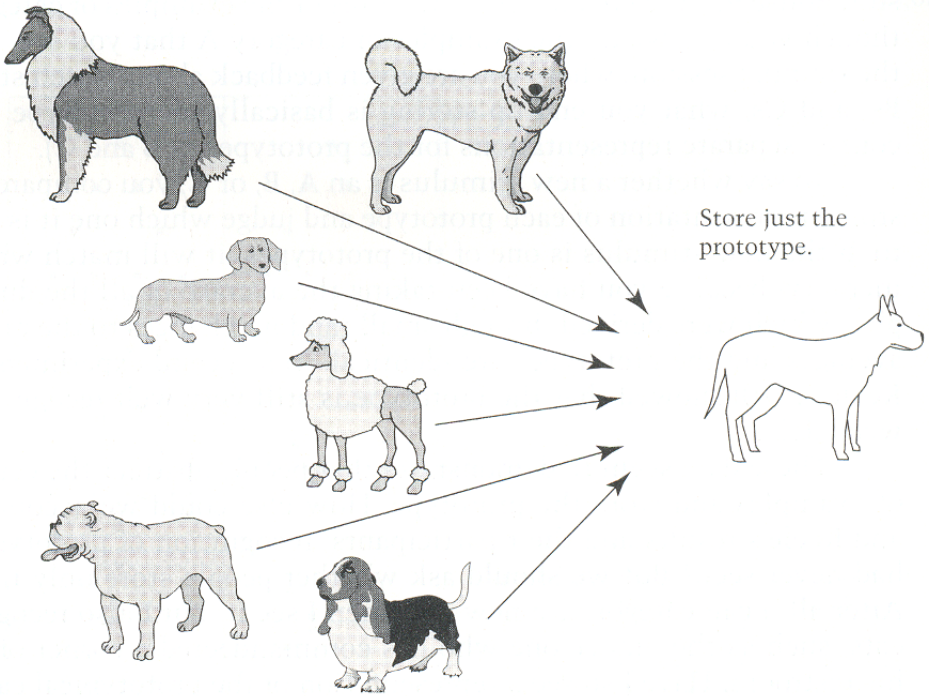


Figure 7.3. *Schematic of the prototype model. Although many exemplars are seen, only the prototype is stored. The prototype is updated continually to incorporate more experience with new exemplars.*

1 Prototype or many Exemplars ?

It may depend on the object class

Prototype Model

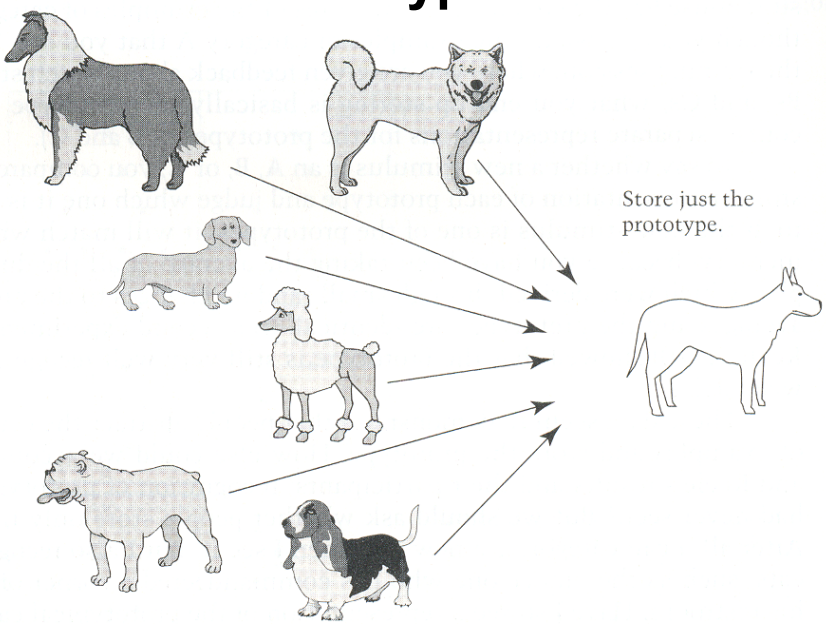


Figure 7.3. *Schematic of the prototype model. Although many exemplars are seen, only the prototype is stored. The prototype is updated continually to incorporate more experience with new exemplars.*

Category judgments are made by **comparing a new exemplar to the prototype.**

Exemplars Model

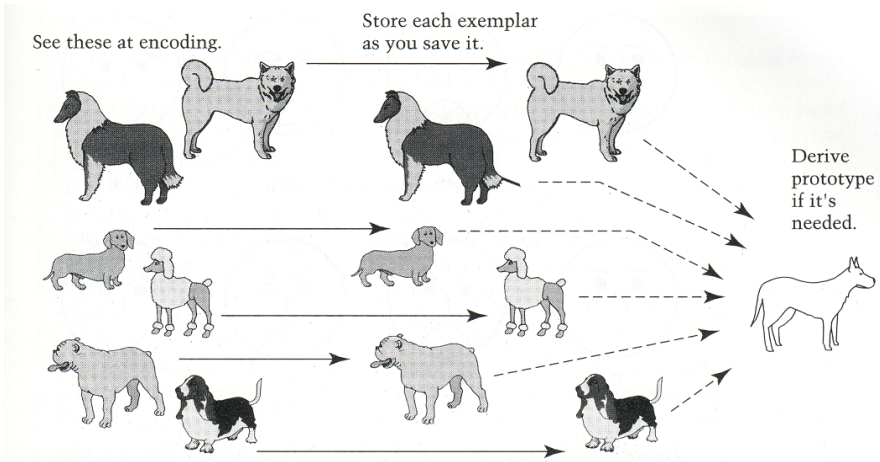


Figure 7.4. *Schematic of the exemplar model. As each exemplar is seen, it is encoded into memory. A prototype is abstracted only when it is needed, for example, when a new exemplar must be categorized.*

Category judgments are made by **comparing a new exemplar to all the old exemplars of a category or to the exemplar that is the most appropriate**

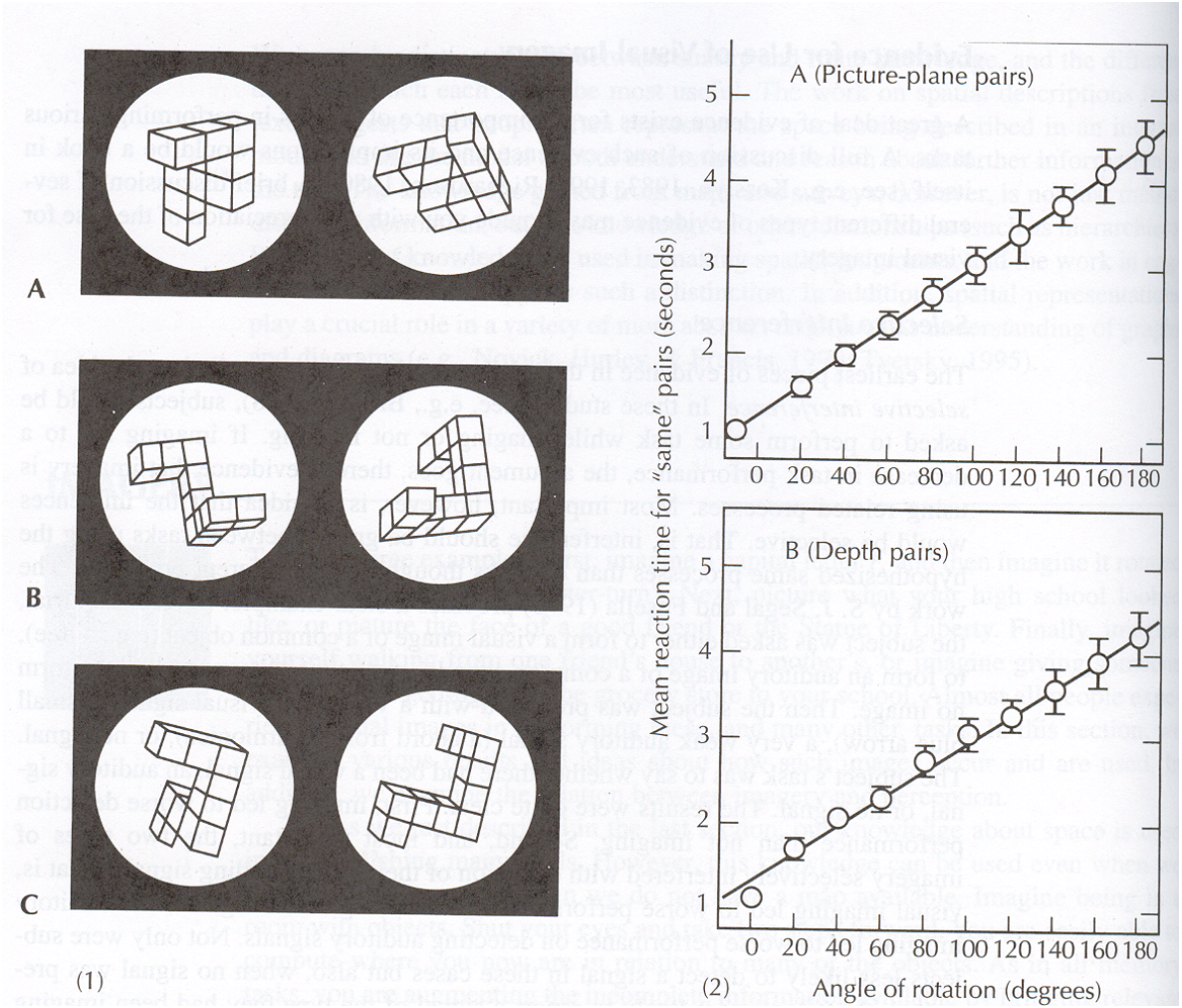
Mental representations

- Functional-equivalence hypothesis: Although we do not construct images that are exactly identical to percept, we construct images that are functionally equivalent to percepts

Principles of Mental Imagery

- Mental transformations of images and mental movements across images corresponds to transformations and movements across physical objects and space
- Spatial relations among elements of a visual image are analogous to these relations in the physical space (for instance, easier to see the details of a larger image in your mind than a smaller one)
- The same regions of the brain are involved in manipulating mental imagery and visual percepts (like imagining manipulating an object)

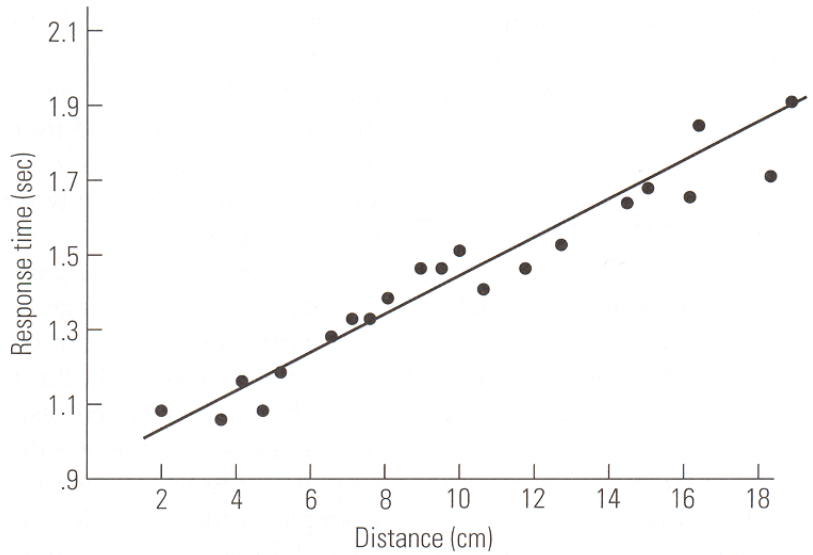
Shepard & Metzler's (1971) Mental rotation experiment



These findings are functionally equivalent to what we might expect if the participants had been rotating physical objects in space

Image Scanning

Map of an imaginary island with various landmarks to determine whether mental scanning across the image of a map was functionally equivalent to perceptual scanning of a perceived map



Participants had to "scan" from one point on their mental image to another point; they pressed a button to indicate when their "mind's eye" had arrived at its destination. Response times in this task were closely related to the "distance" participants had to scan across on the image, implying that mental images are similar to actual pictures in how they represent positions and distance. [After Kosslyn, 1983.]

Image Scanning

Participants seem to have encoded the map in the form of an image and actually to have scanned that image

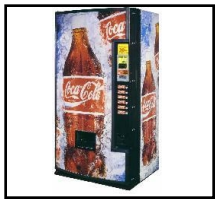
Object Size

What's the template size?

Hypothesis 1:

1 template size

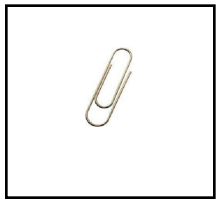
For all objects

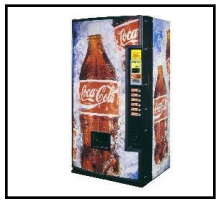


Hypothesis 2:

- consistent visual size

- *different* for each object

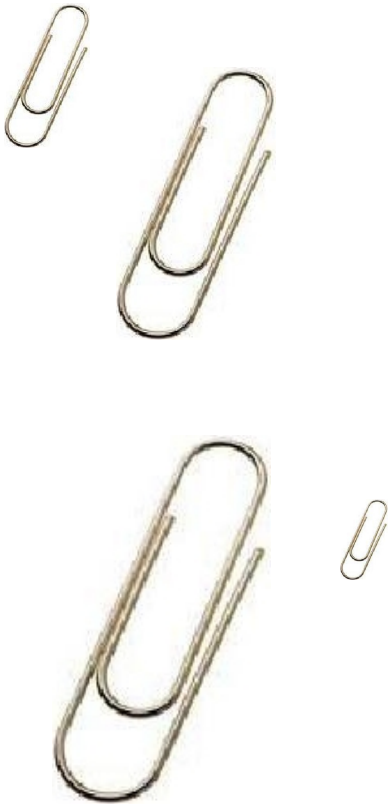


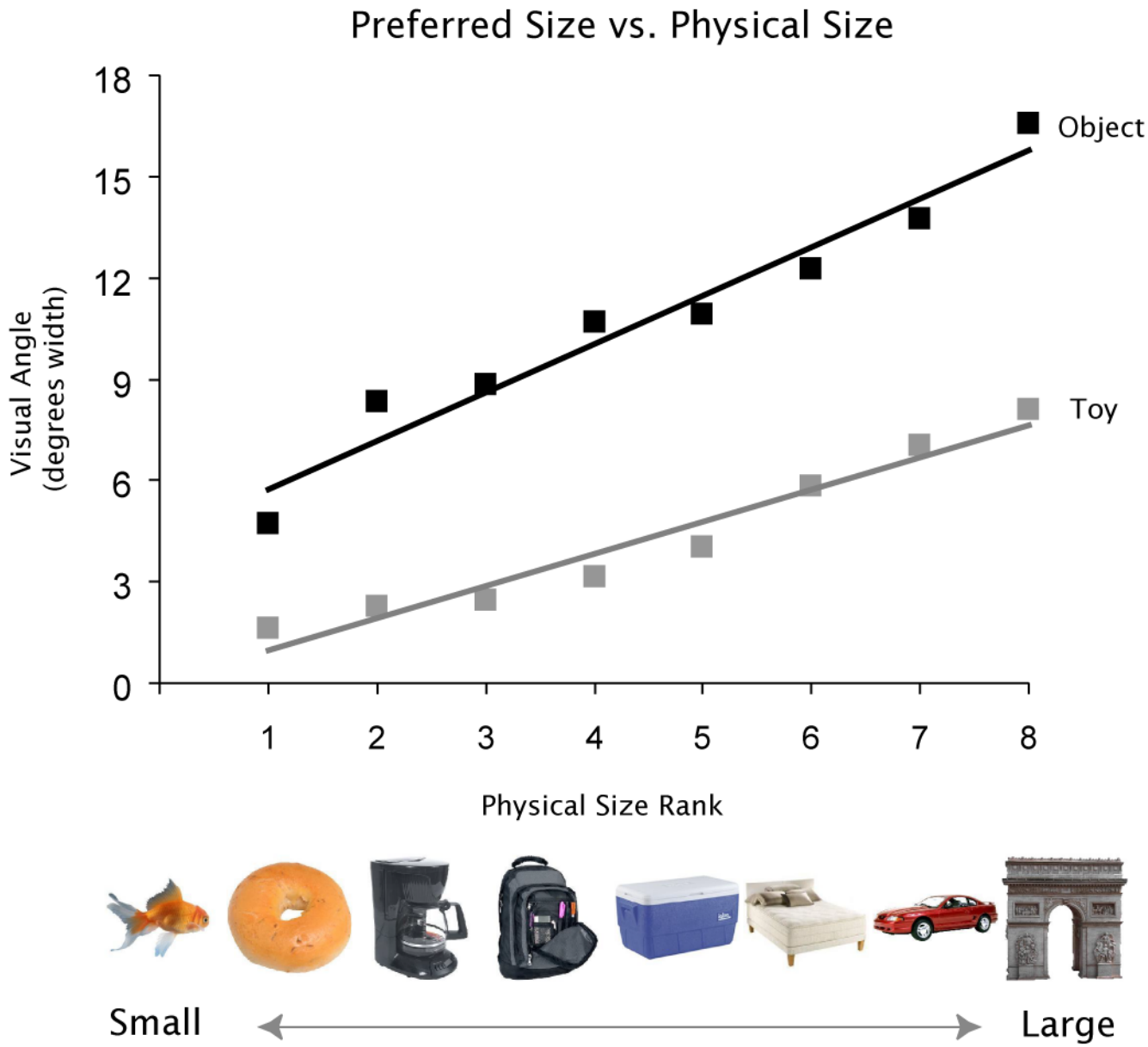


Hypothesis 3:

-no consistent

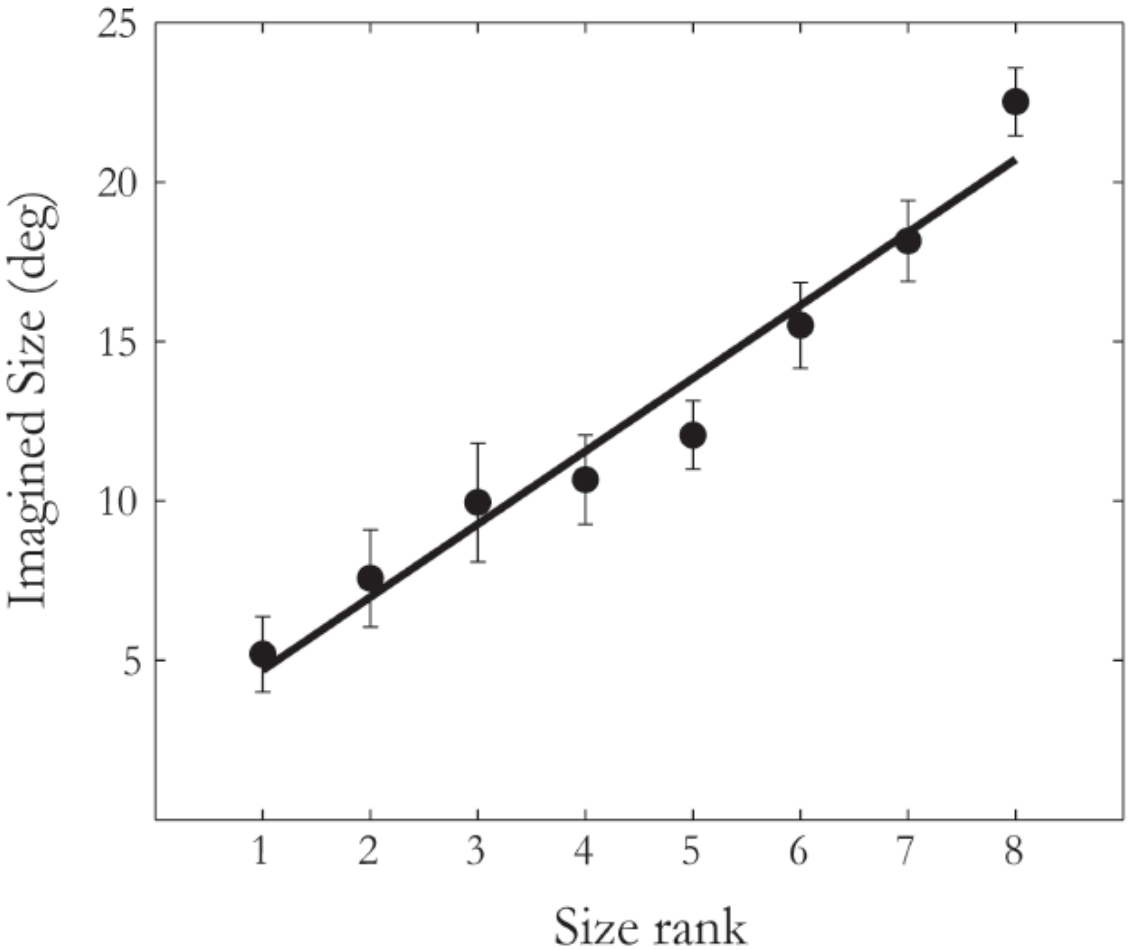
visual size at all

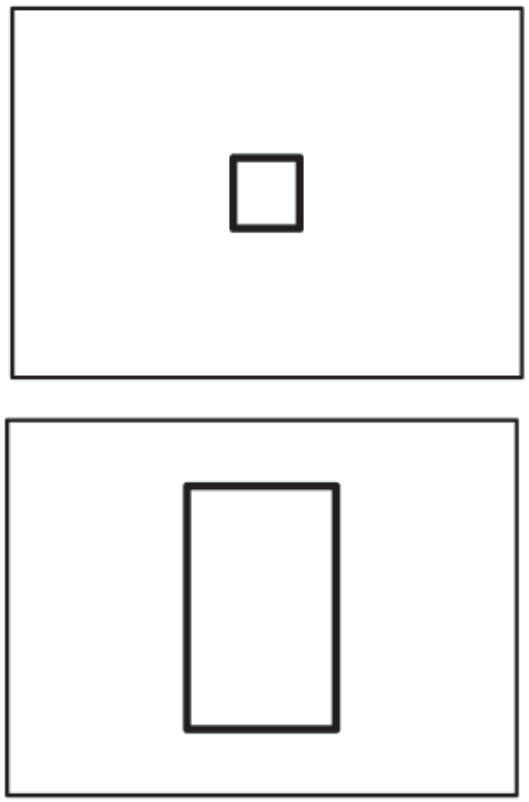




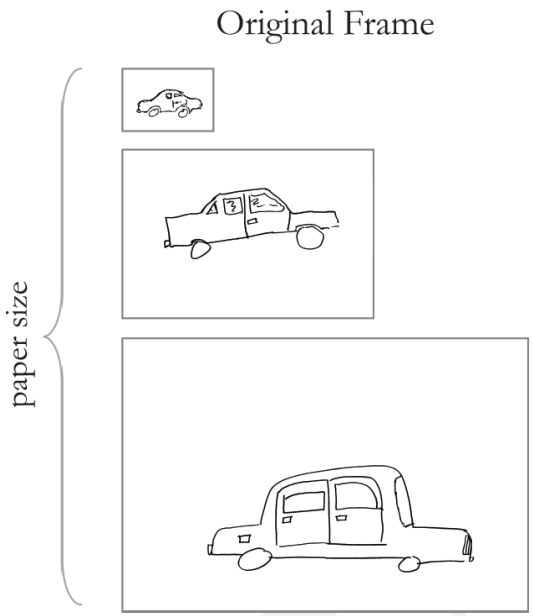
Konkle & Oliva (2011). Canonical visual size for real-world objects.

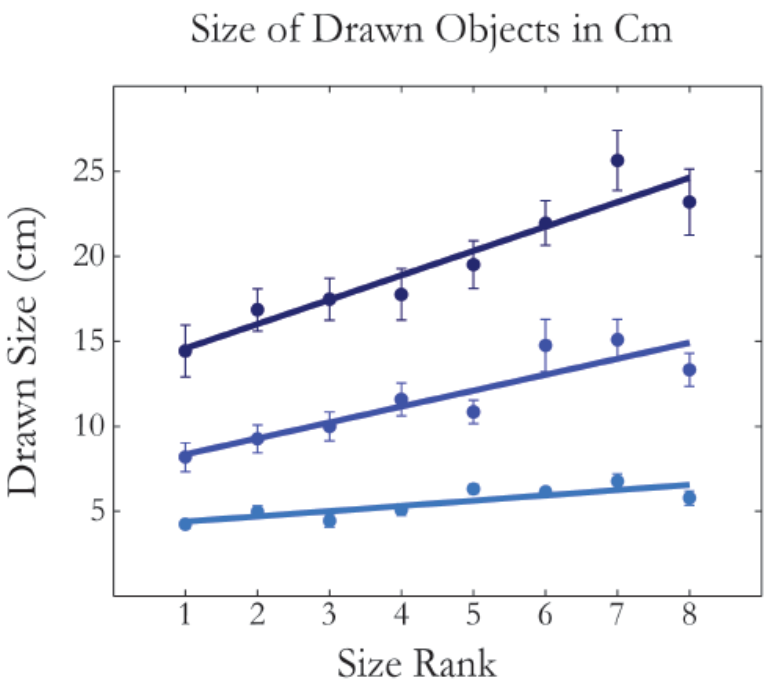
Mental imagery: No picture on screen

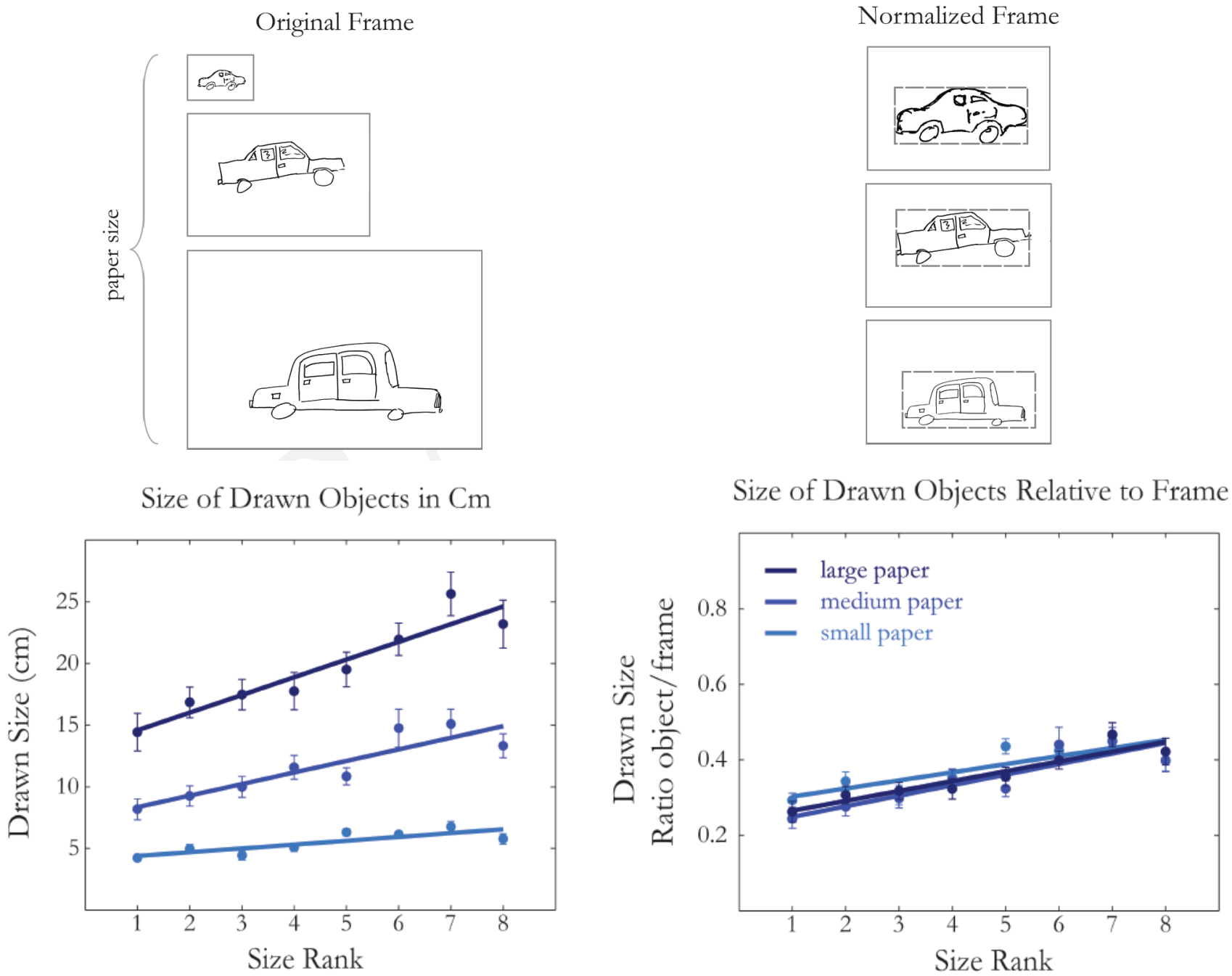




Konkle & Oliva (2011). Canonical visual size for real-world objects.


Original Frame
paper size

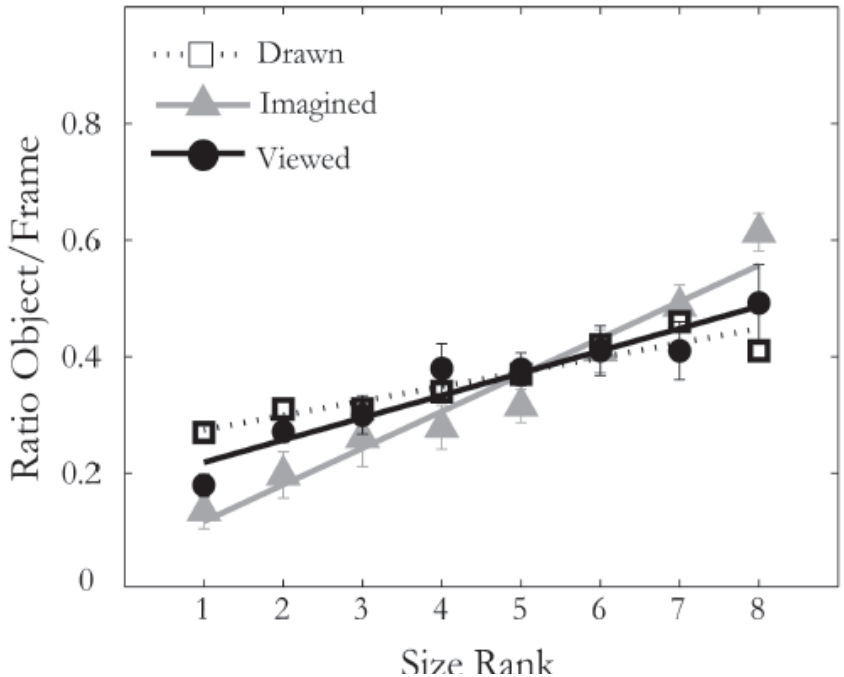

Size of Drawn Objects in Cm
Drawn Size (cm)
25
20
15
10
5
1
2
3
4
5
6
7
8
Size Rank

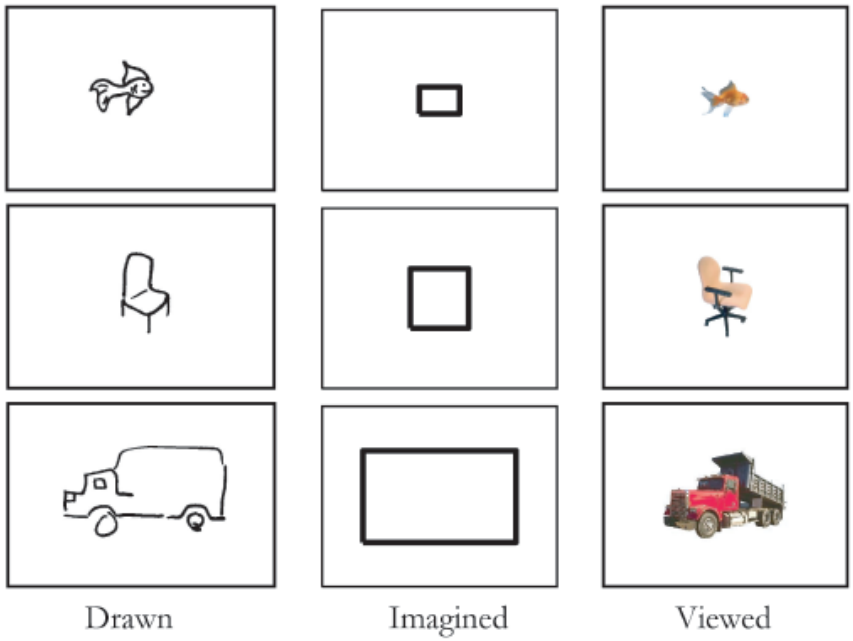


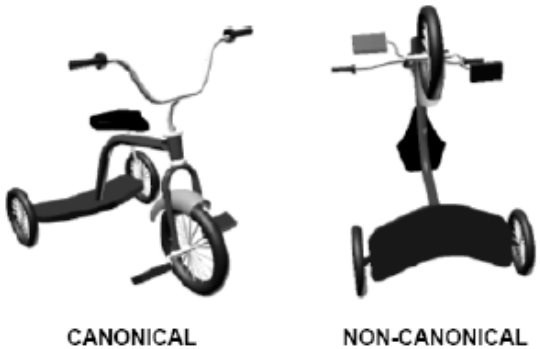
Konkle & Oliva (2011). Canonical visual size for real-world objects.

Canonical size:

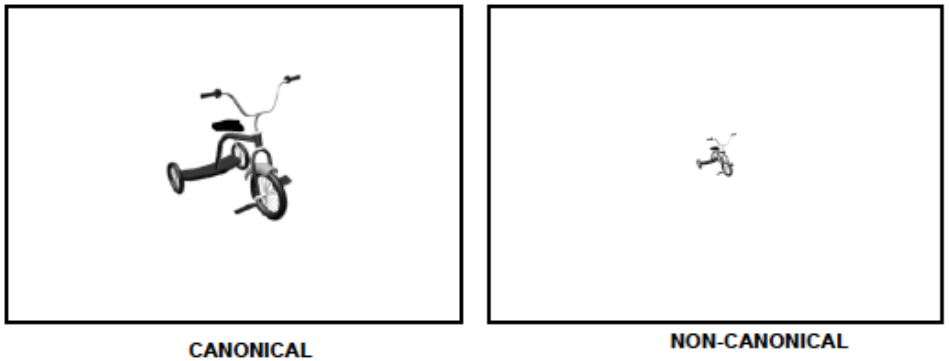
Object representations are not size invariant







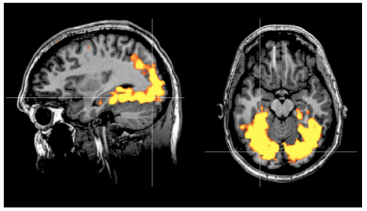
Palmer, Rosch and Chase, 1981



Konkle & Oliva, 2011

Neural Representation of Objects

Are different patches of cortex more likely to process objects of different real-world sizes?

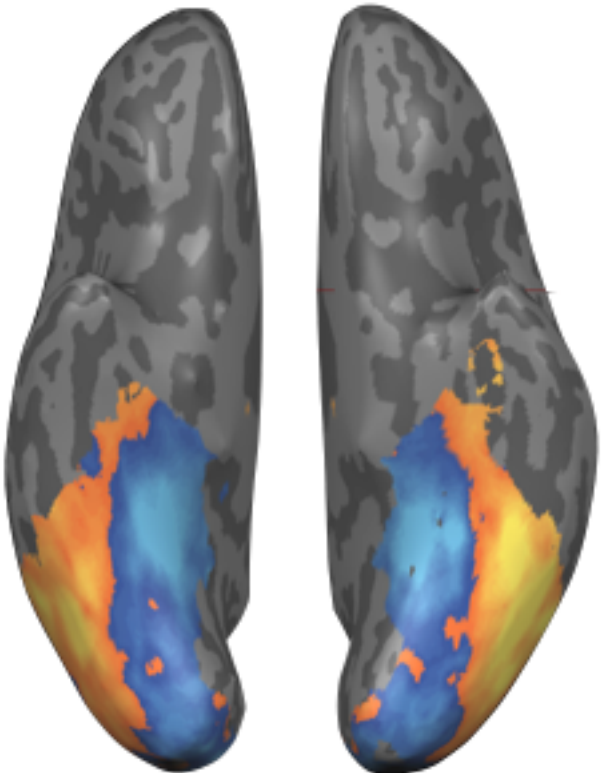


Different cortical regions for small and big objects

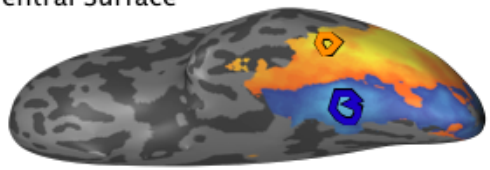
Small Objects

Big Objects

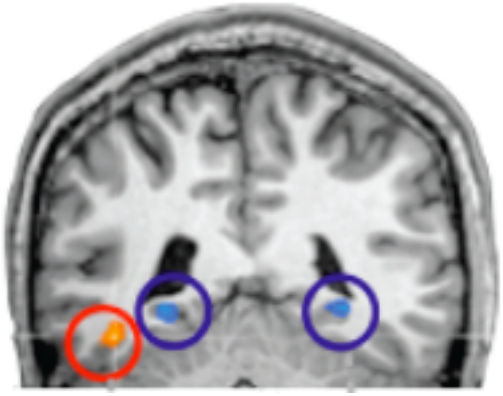
View of the bottom surface of an "inflated" brain



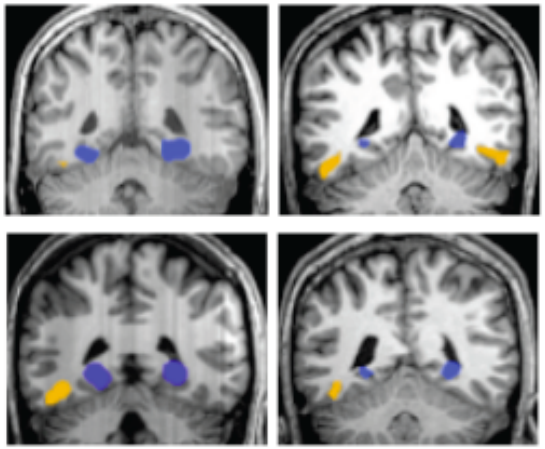
Ventral Surface



Whole brain analysis (n=12)

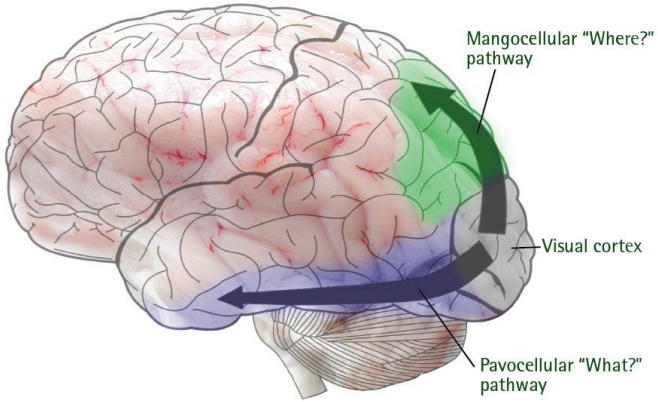


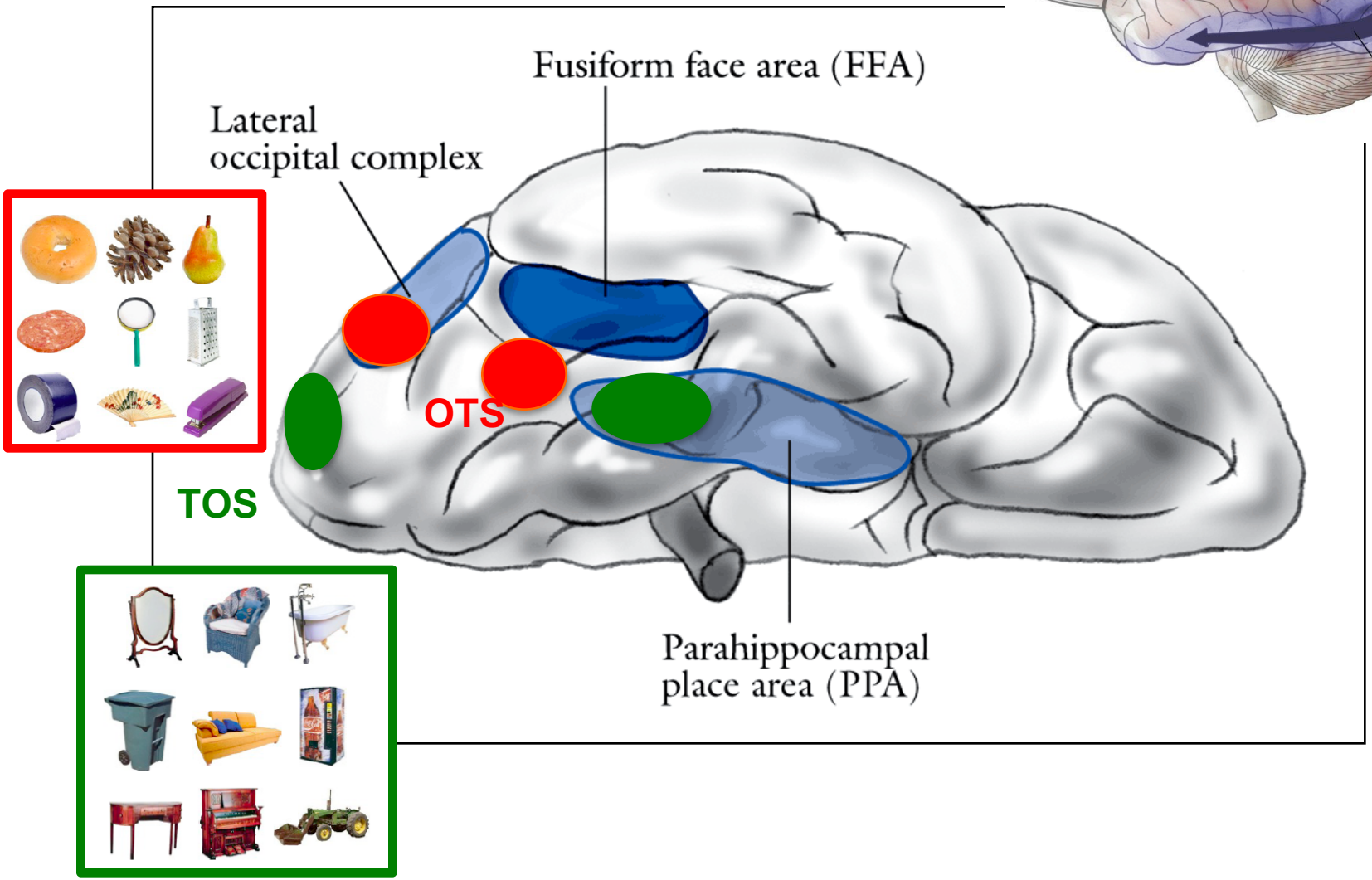
Individual brains



Konkle & Oliva (2012). *Neuron*

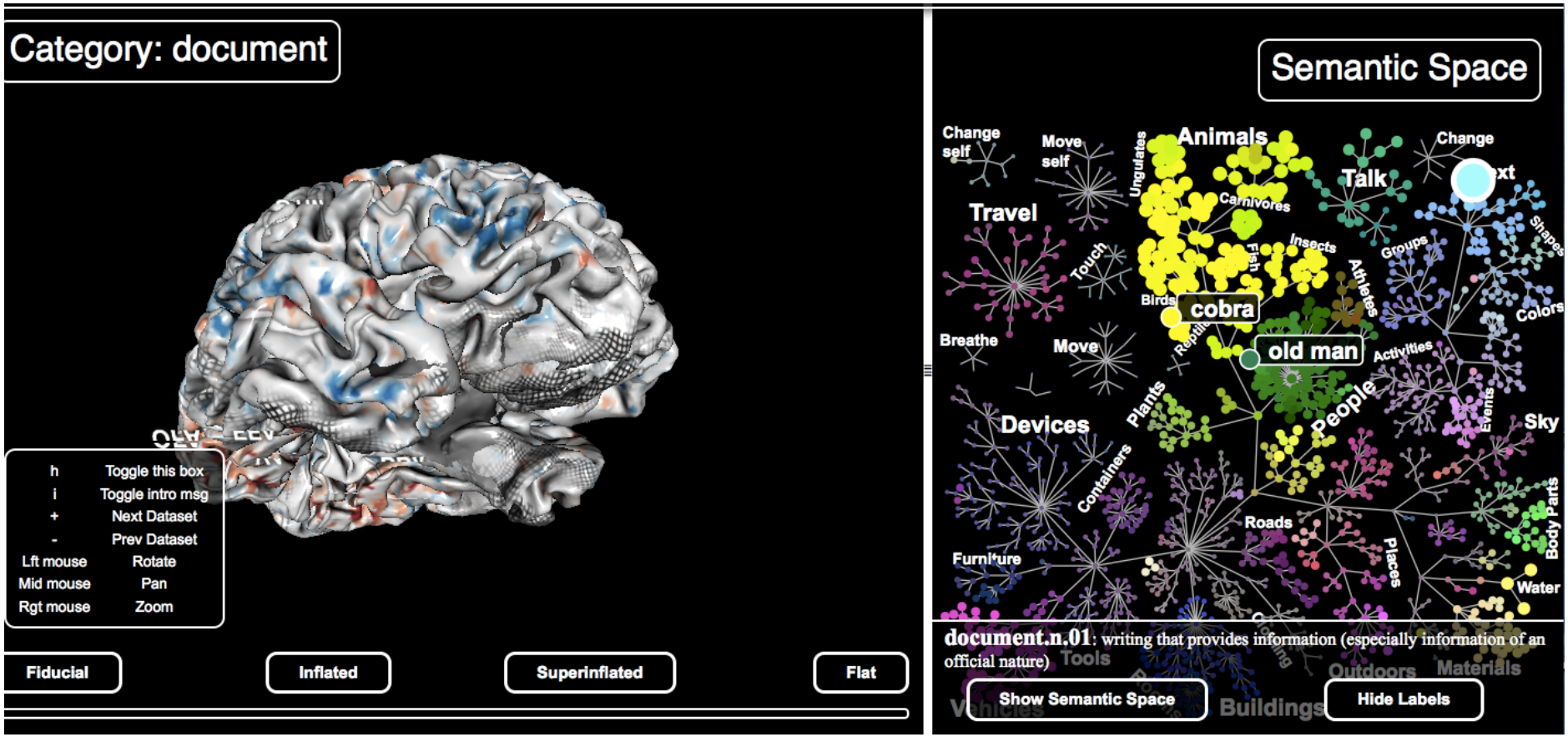
Object Recognition: Multiple sub-regions





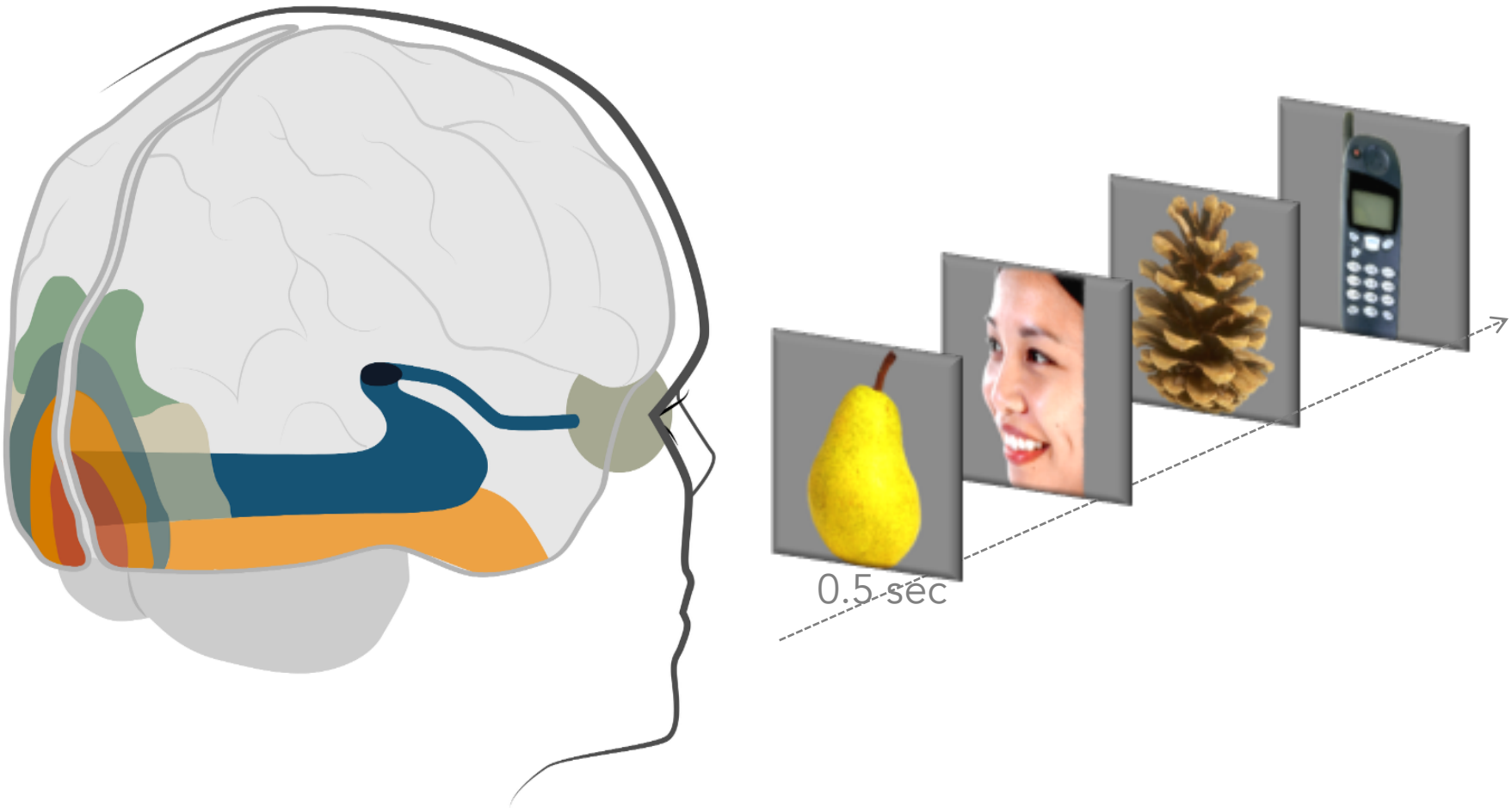
Konkle & Oliva (2012) Neuron; Trioani et al, (2012) Cortex

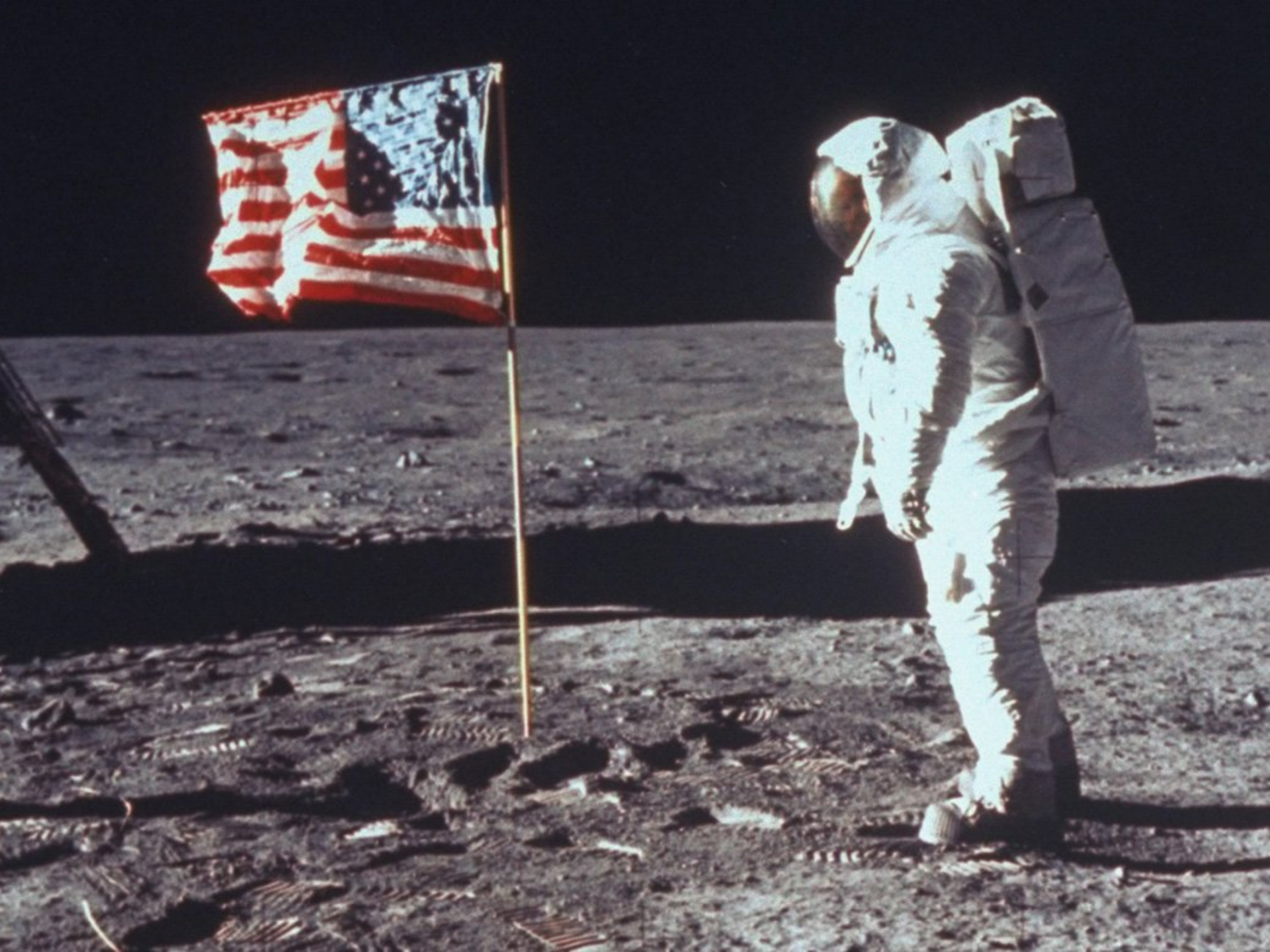
Distributed View of Object Representations

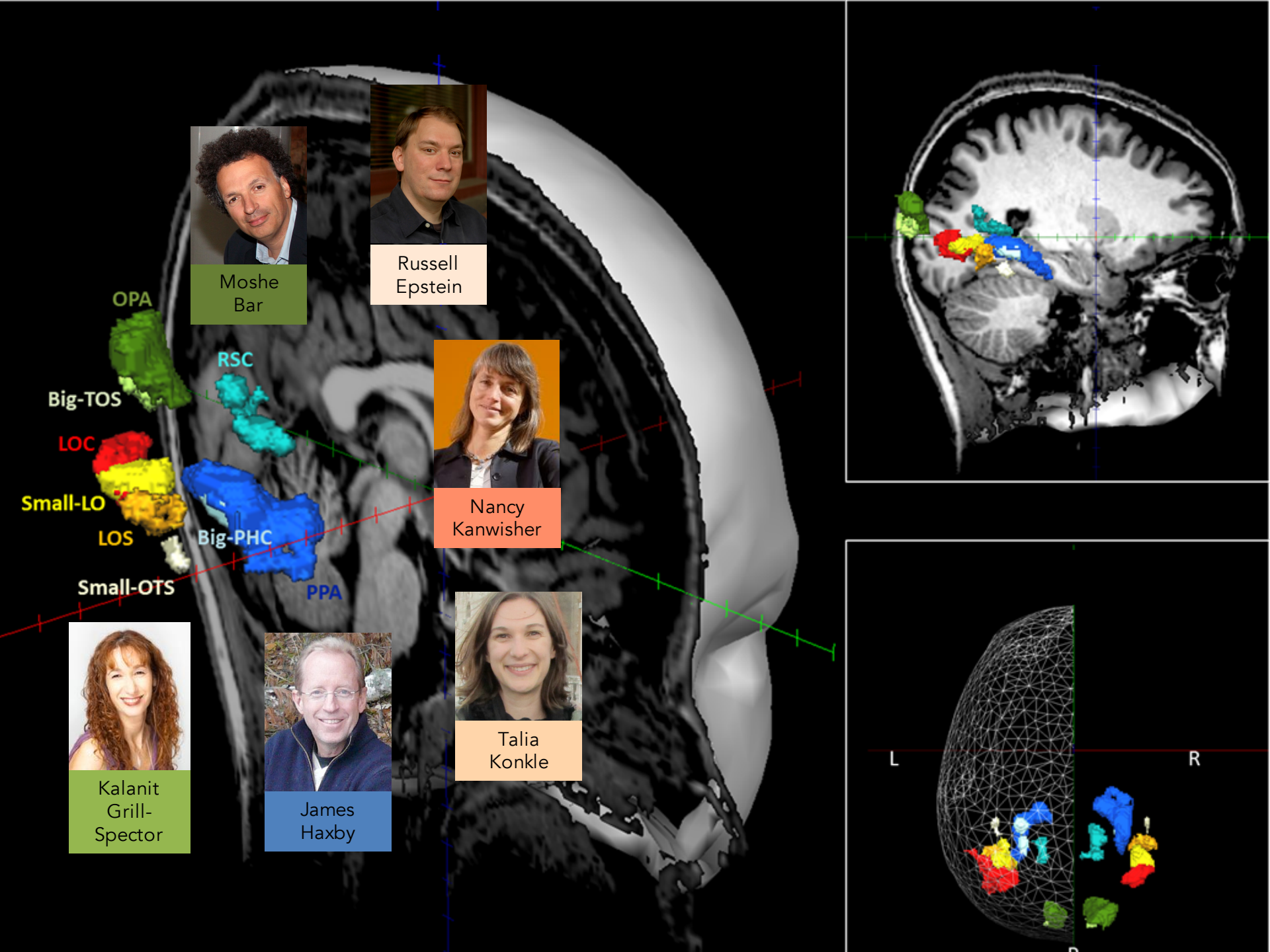


Huth, Nishimoto, Vu and Gallant (2012), "A continuous semantic space describes the representation of thousands of object and action categories across the human brain", Neuron 2012. (web: http://gallantlab.org/semanticmovies/)

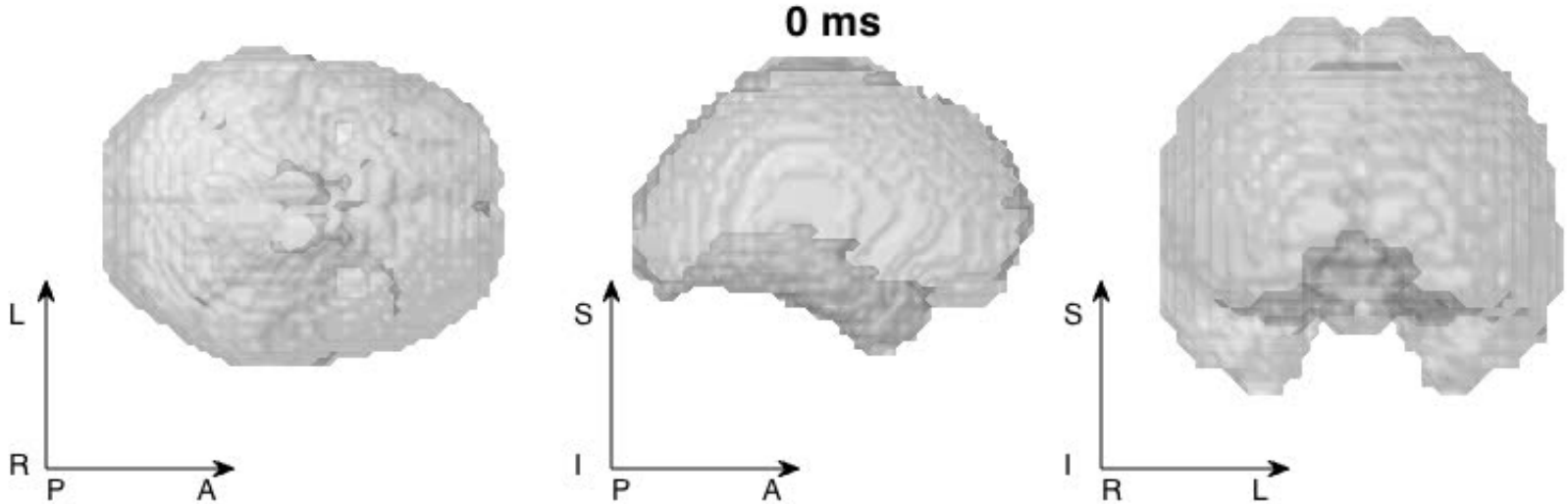
High Computing Visual Engine: Object recognition






Moshe Bar
Russell Epstein
Nancy Kanwisher
Talia Konkle
James Haxby
Kalanit Grill-Spector
OPA
RSC
Big-TOS
LOC
Small-LO
LOS
Big-PHC
Small-OTS
PPA
L
R

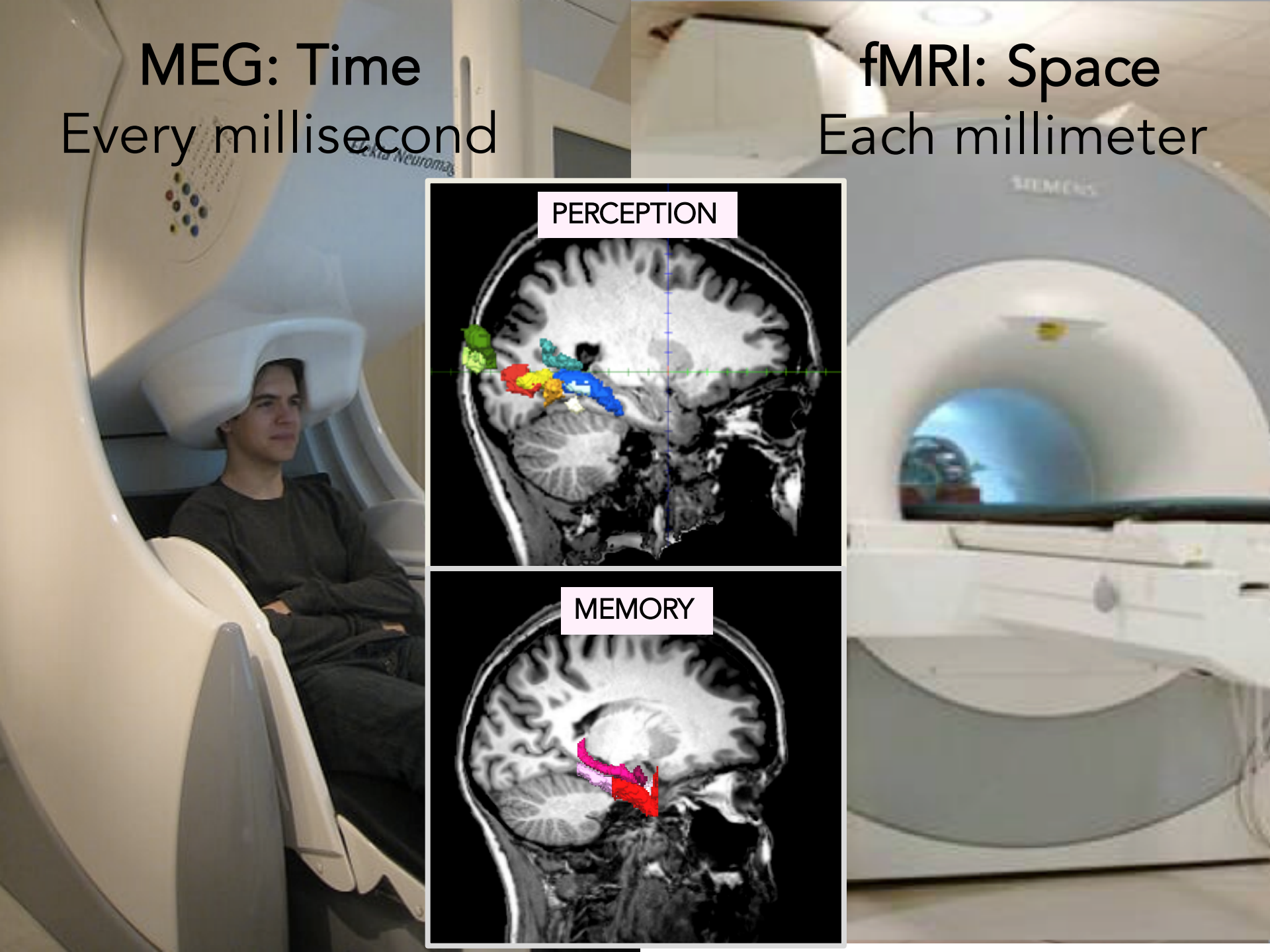
Dynamics of Seeing



Cichy, Pantazis, Oliva

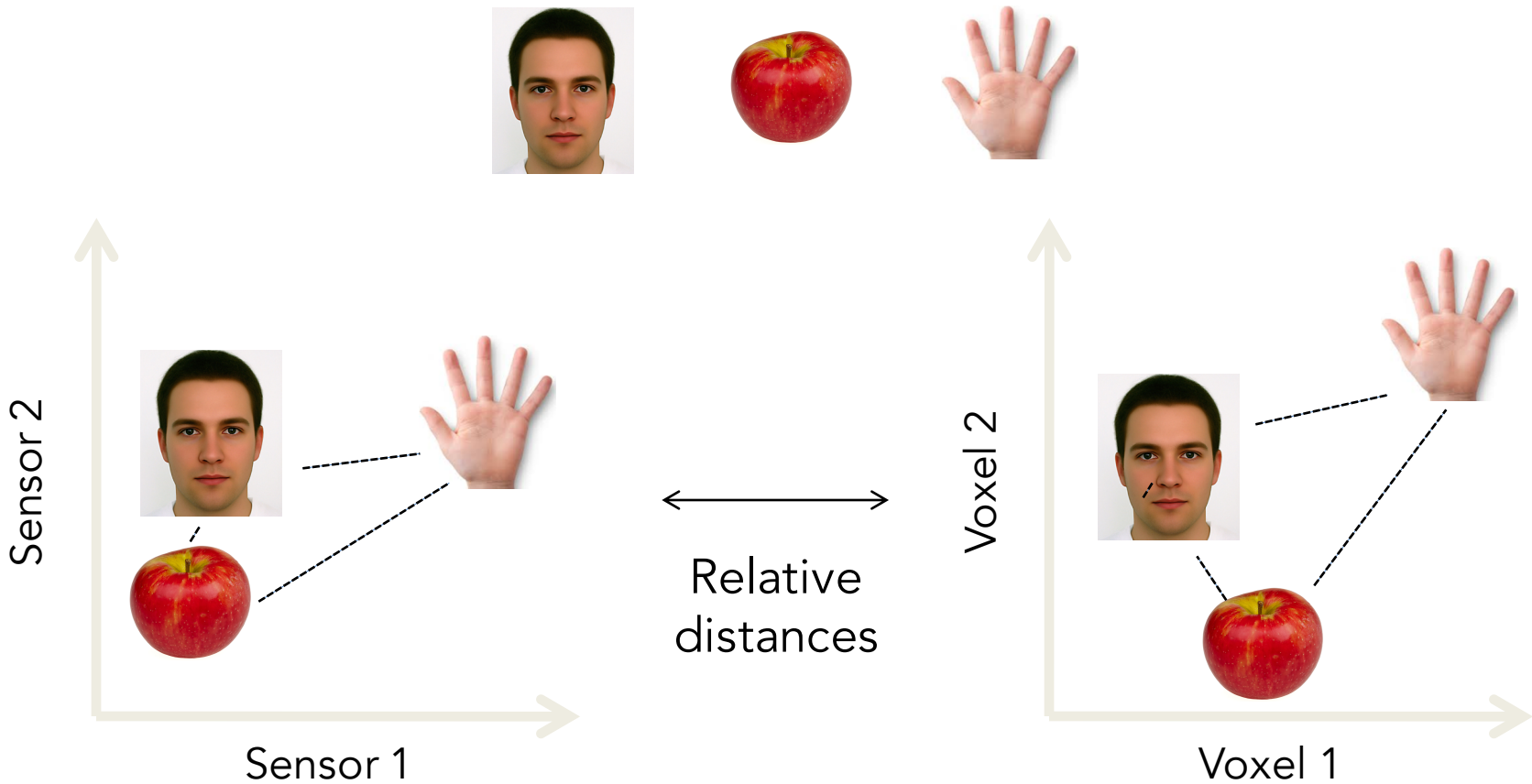
MEG: Time
Every millisecond

fMRI: Space
Each millimeter



Representational Geometry

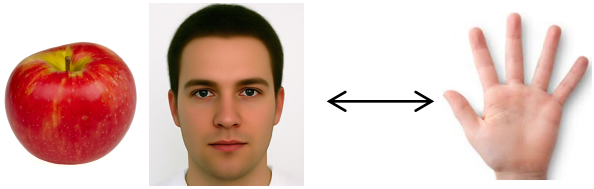
Nikolaus Kriegeskorte (2008)



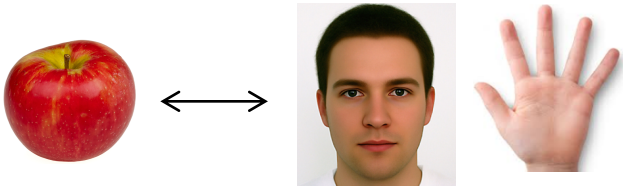
Shepard et al., 1980; Kruskal and Wish., 1978; Edelman et al. 1998; Kriegeskorte et al., 2008; Mur et al., 2009; Liu et al., 2013

Representational Geometry

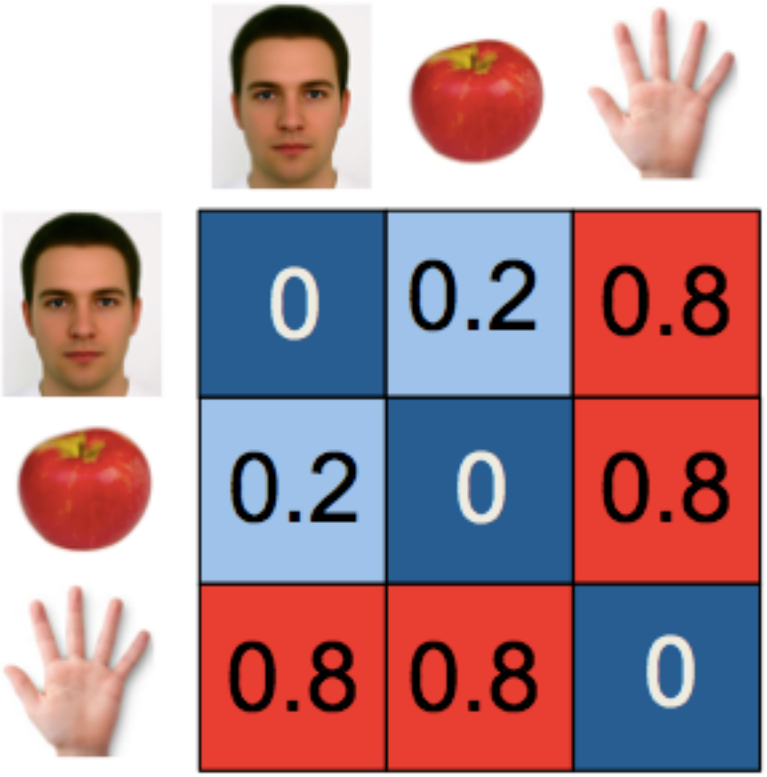
Nikolaus Kriegeskorte (2008)

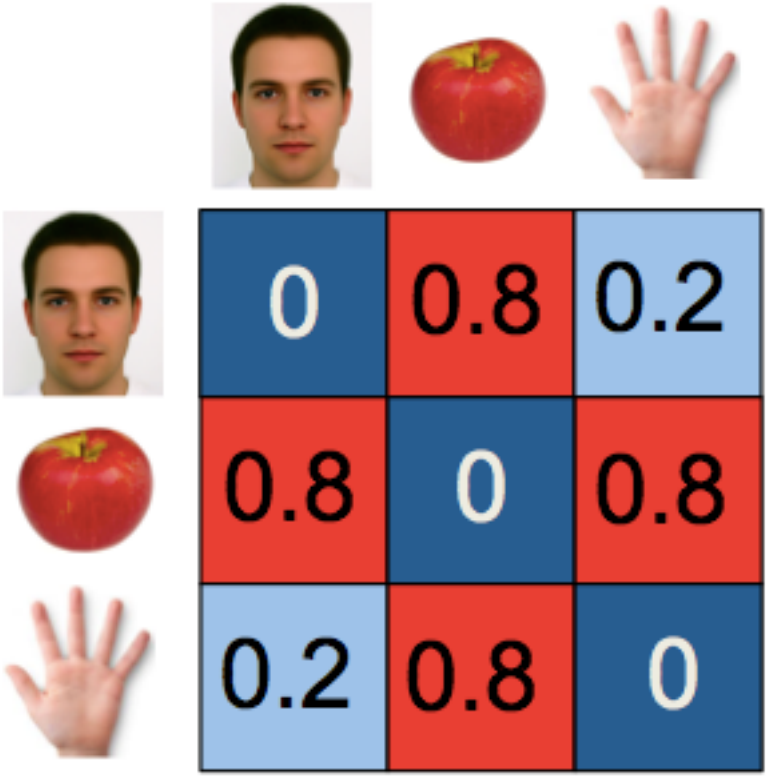


Round shape



Body part

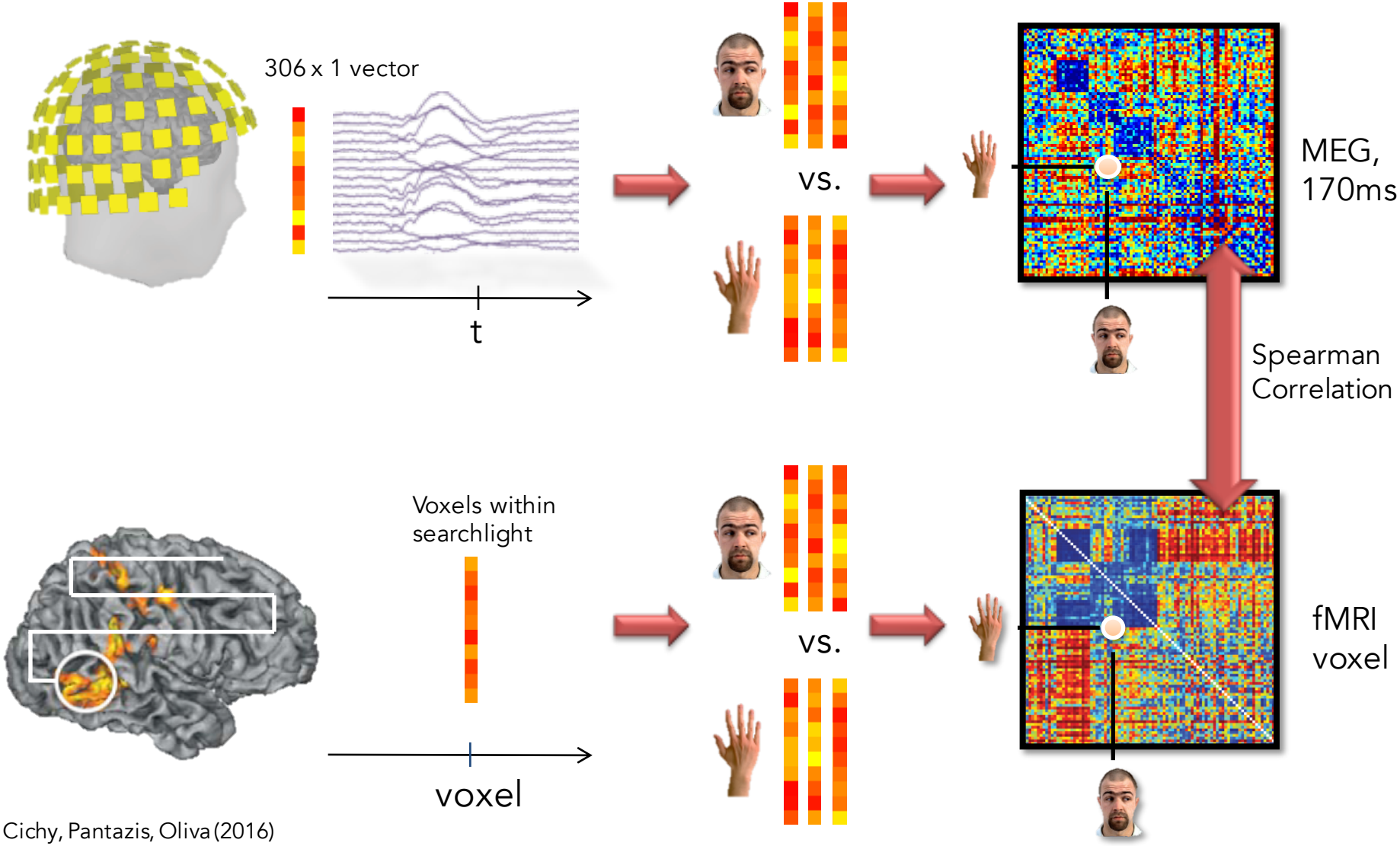




"RDMs as a hub to relate different representations across sensors and models"

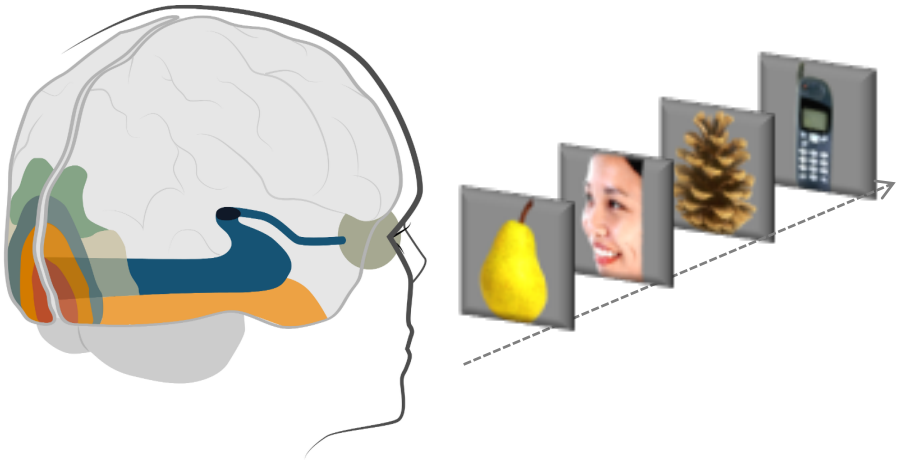
Time-specific fMRI searchlight analysis

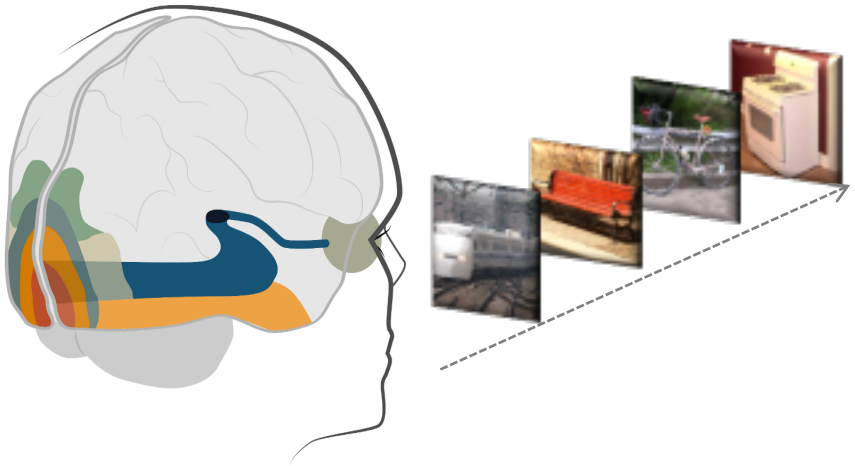
A spatially unbiased view of the relations in similarity structure between MEG and fMRI

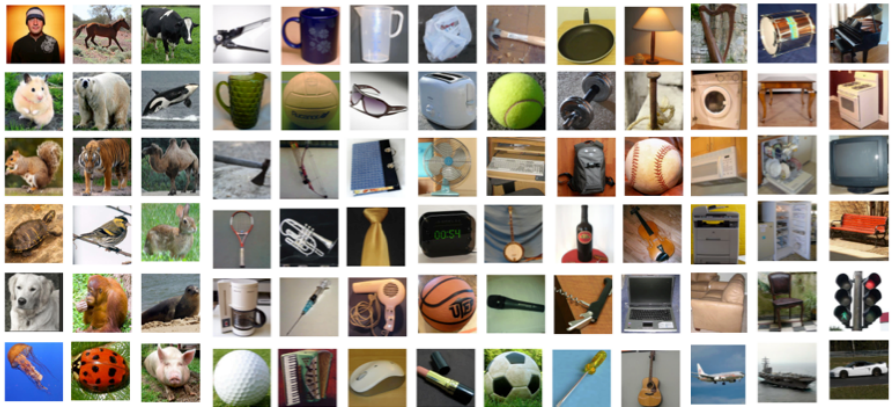


Cichy, Pantazis, Oliva (2016)

Object recognition



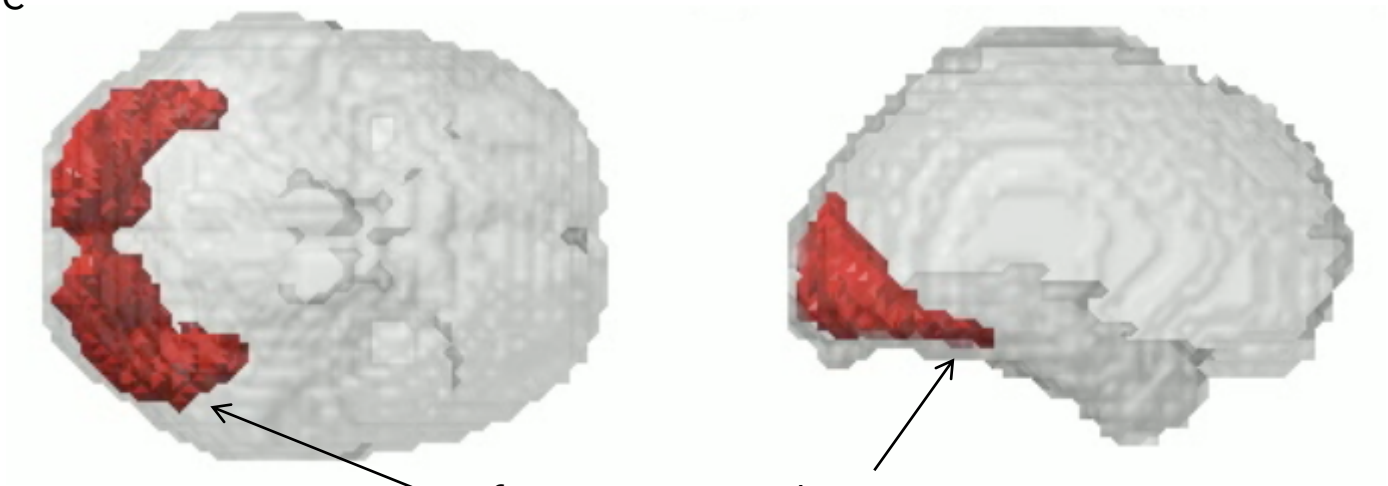




Spatiotemporal maps of correlations between MEG and fMRI

100 msec

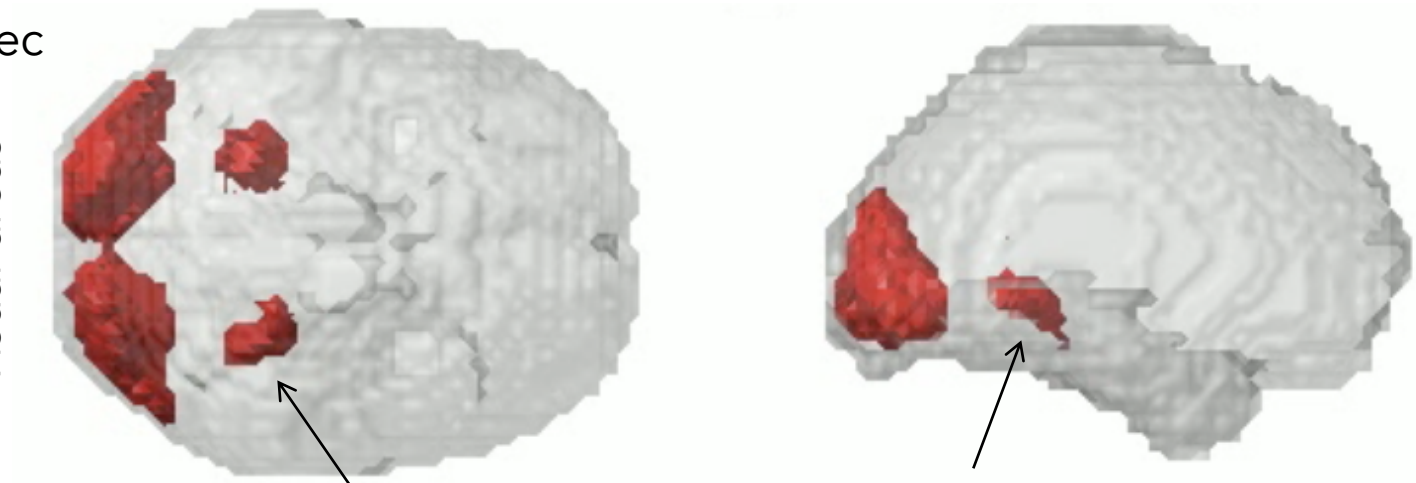
Visual areas



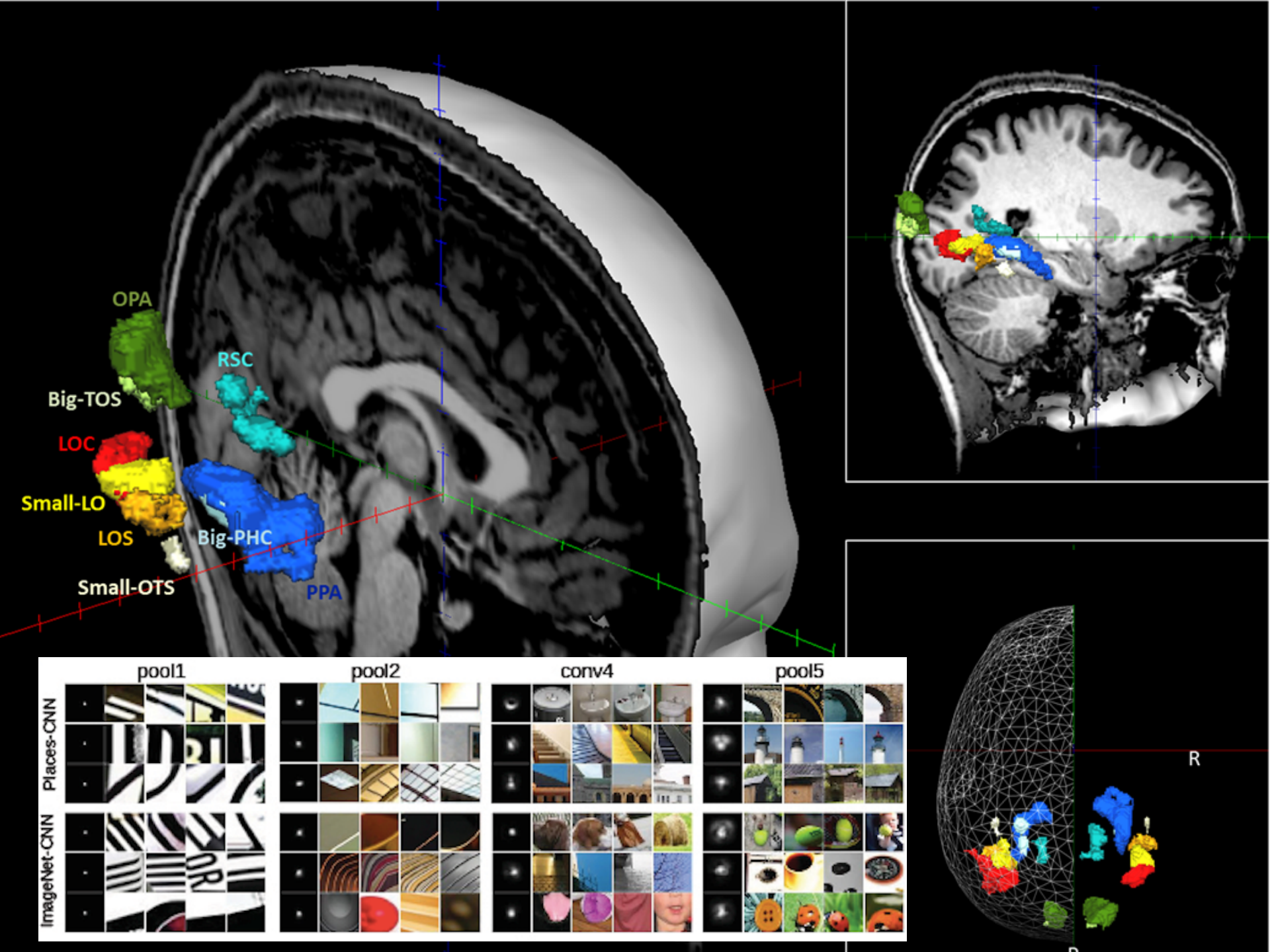
Inferior-temporal cortex

100 msec

Visual areas

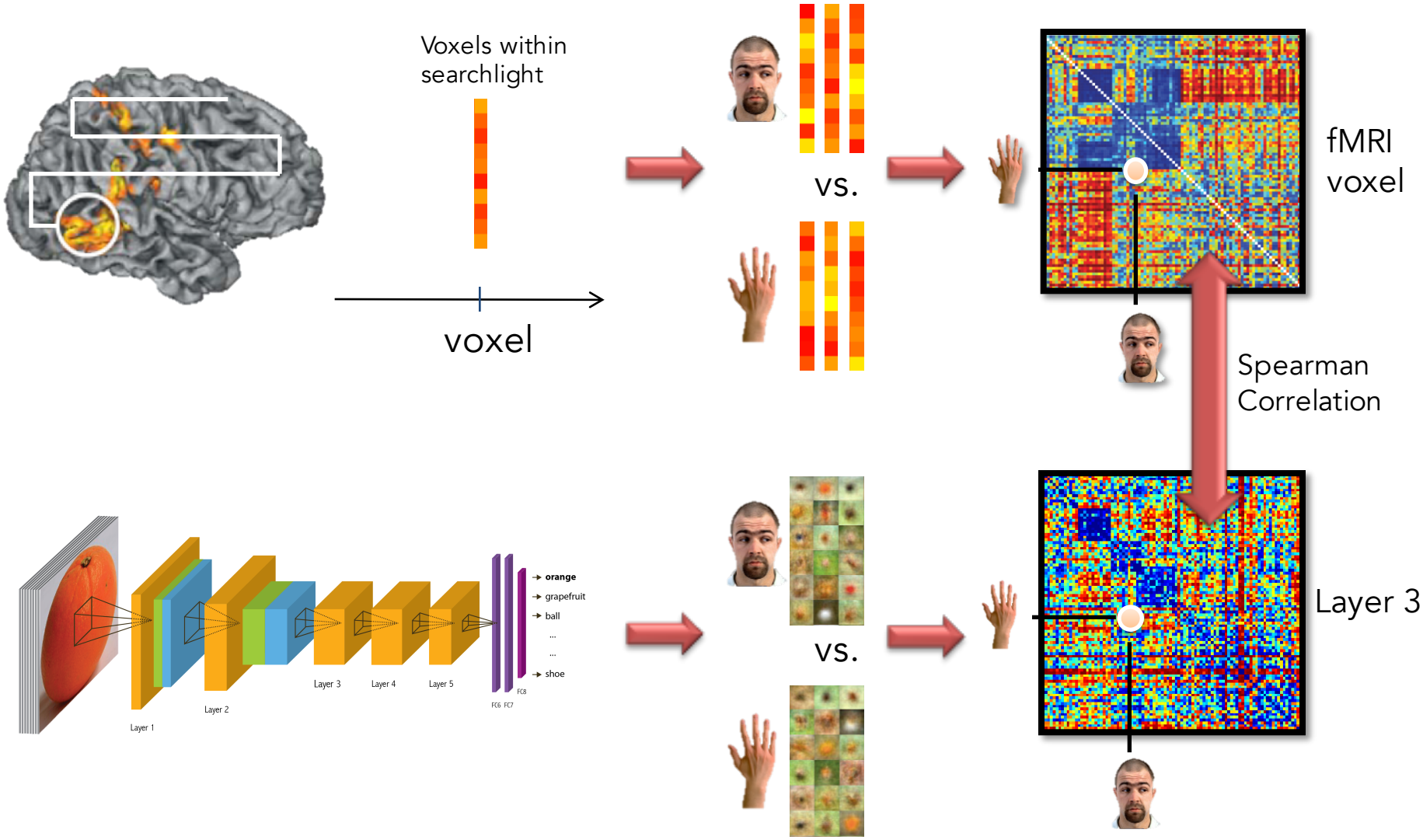


Parahippocampal cortex


OPA
RSC
Big-TOS
LOC
Small-LO
LOS
Big-PHC
Small-OTS
PPA
pool1
pool2
conv4
pool5
Places-CNN
ImageNet-CNN
R

Algorithmic-specific fMRI searchlight analysis

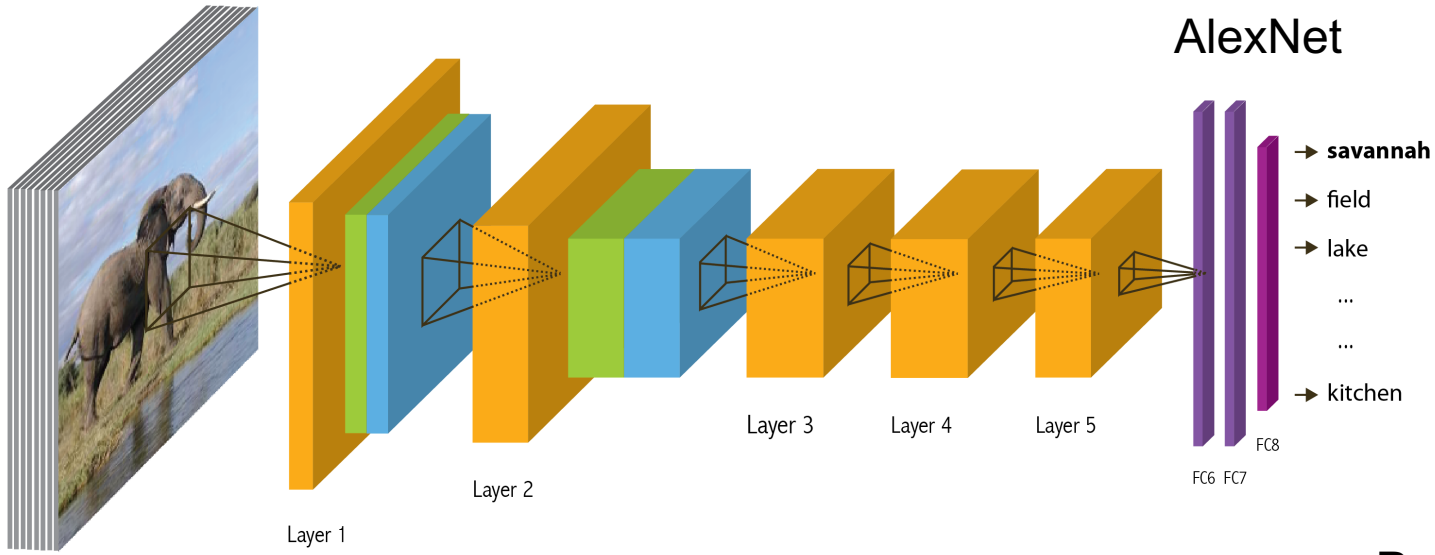
A spatially unbiased view of the relations in similarity structure between deep architectures and fMRI

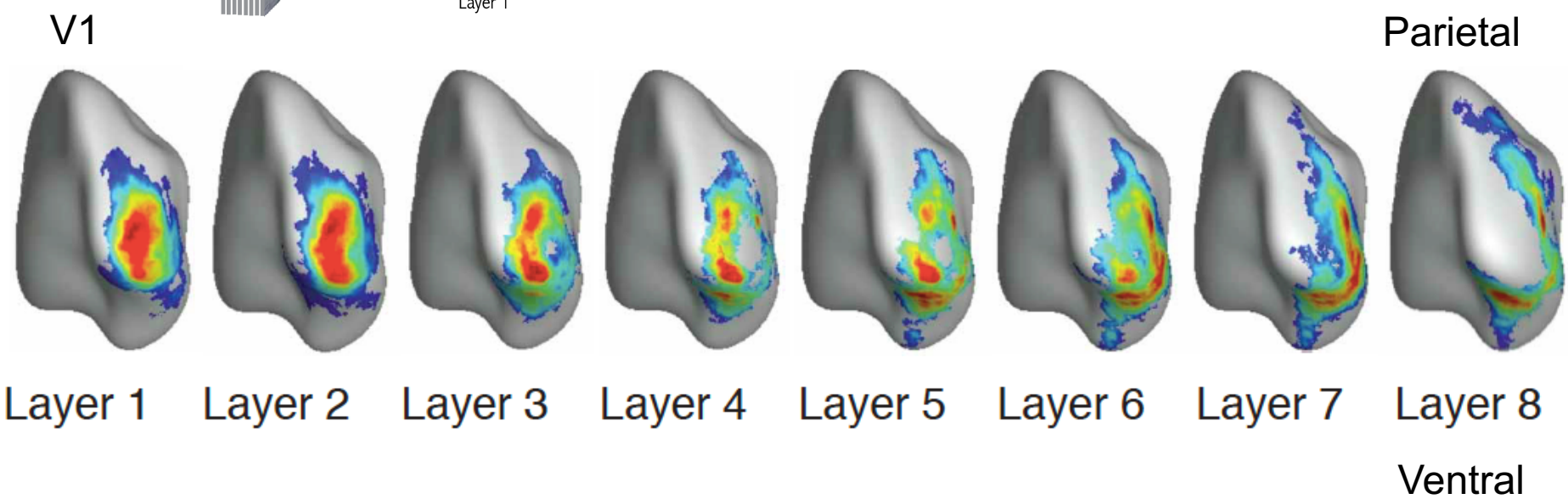


Cichy, Khosla, Pantazis, Torralba, Oliva (2016)

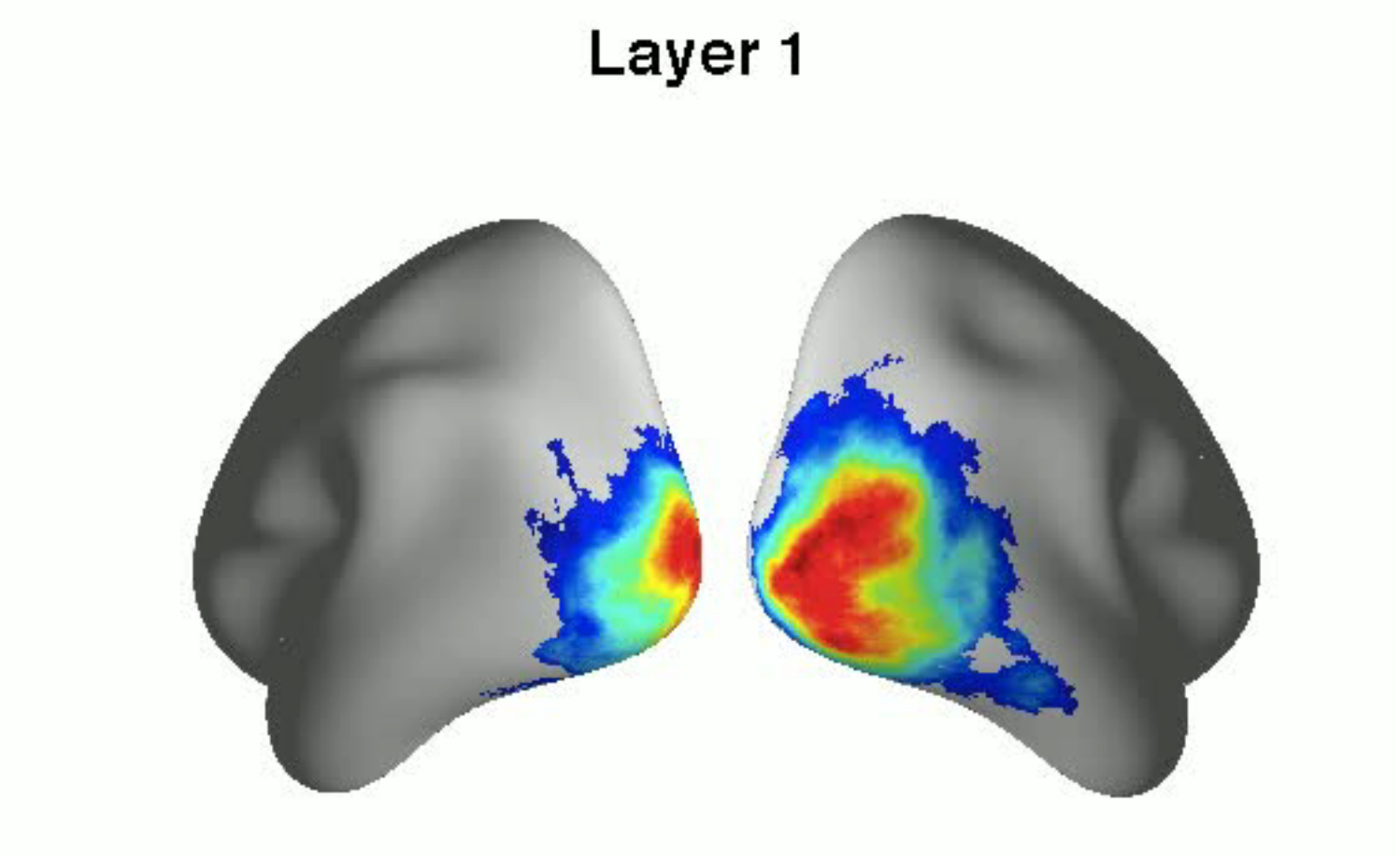
See also Kaligh-Razavi & Krigeskorte (2014)

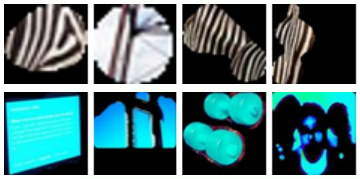
Spatiotemporal map of correlations between human brain and model layers



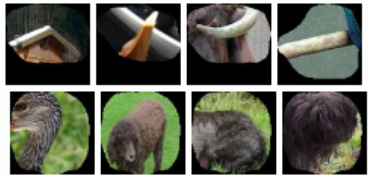


Spatiotemporal maps of correlations between human brain and CNN layers





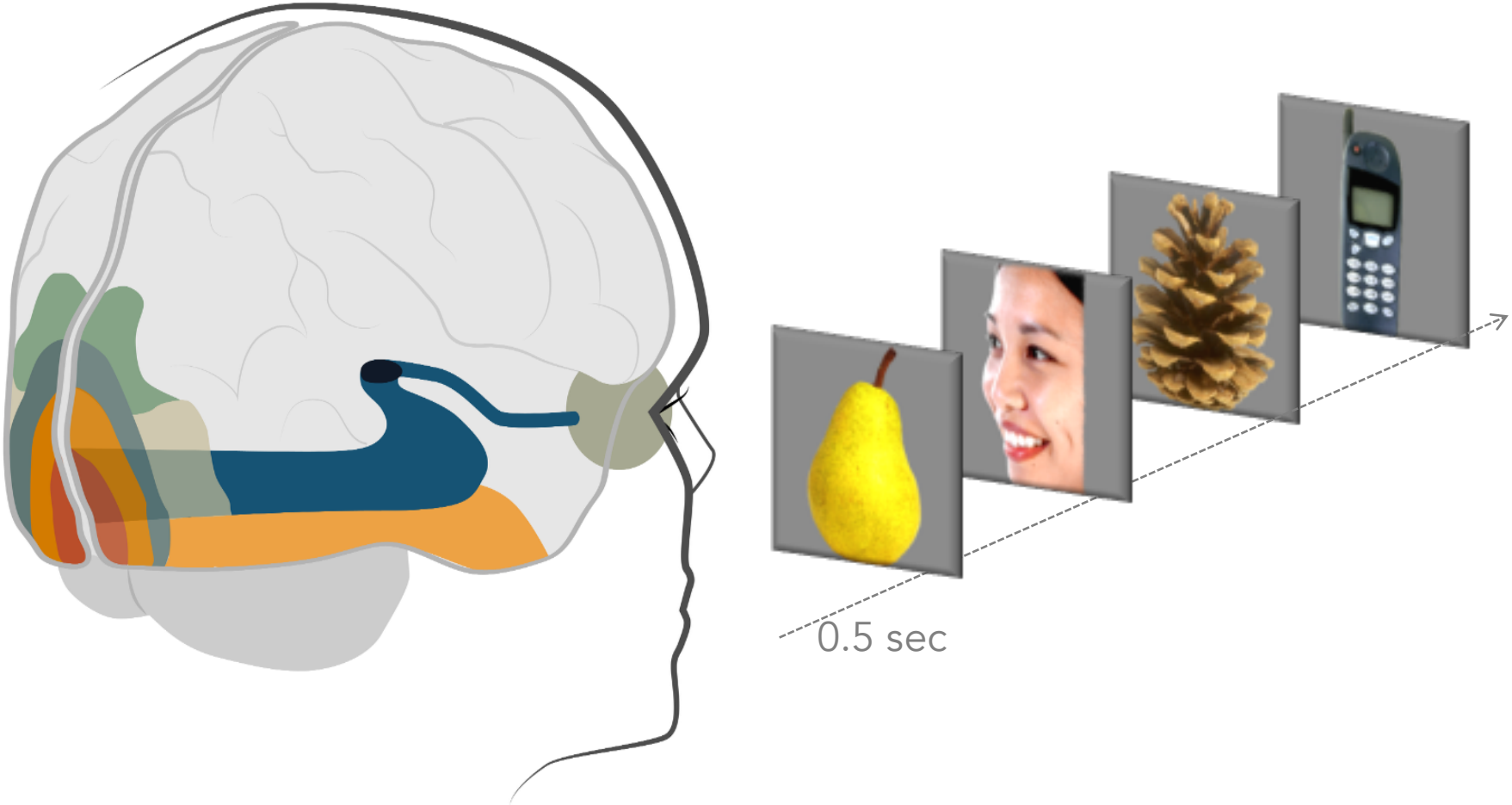
Layers 1-2



Layers 2-4

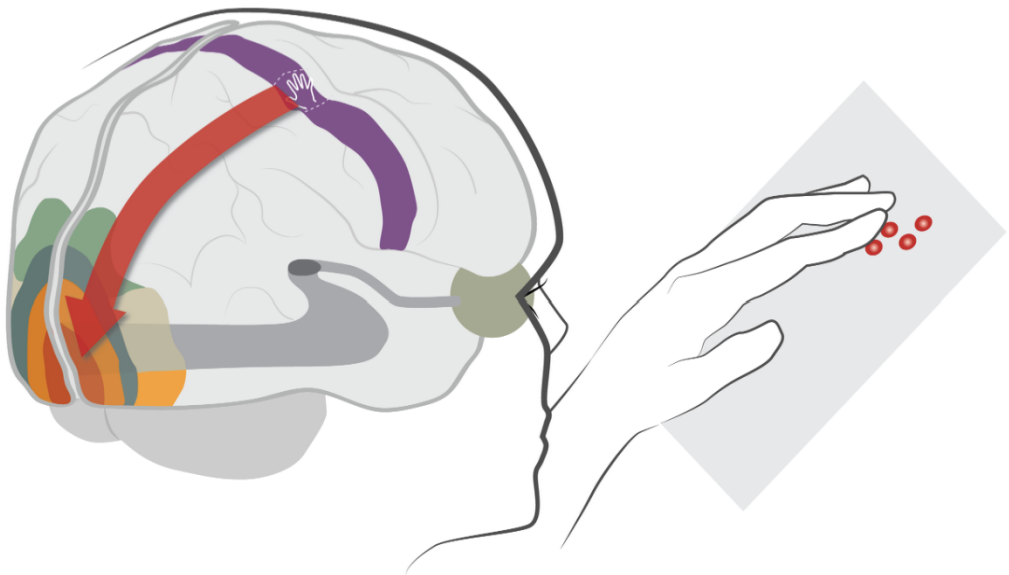
Layers 5-8

High-Powered Machine: Principles



I. Plasticity

Nothing is lost, everything is transformed



Feeling touch with the “visual” brain

Teng, Cichy, Pantazis, Oliva

II. Growth

Immediate

Short-term

Long-term