Bidirectional processing:

feedfoward & feedback circuits

Inferences about the image involve various inferences:

- types of features & attributes (shapes, material)
- recognition over levels of abstraction (parts, objects, actions, scenes)
  - spatial scales
  - relationships

Descriptions are inferences of object properties and relationships — i.e. causes of image intensities, not of image intensity patterns

A crucial assumption is that these inferences are based on deep, generative knowledge of how virtually any natural image could be produced

computational problems

Need to model uncertainty

vision is concerned with causes of image intensity patterns, but the causes of behavioral relevance are encrypted and confounded

many hypotheses about cause can be consistent with the same local image evidence

local variations in image evidence can be consistent with the same cause

accurate perceptual decisions resolve these ambiguities by combining lots of image evidence with built-in knowledge

computational problems

Need to solve scalability

Solving toy (low-dimensional) problems rarely scales up to deal with the complexity of natural images.

Humans have the capacity to <u>*quickly*</u> deal with an enormous space of possible objects (30 to 300K) as they appear in different contexts in natural images for different tasks.

computational problems

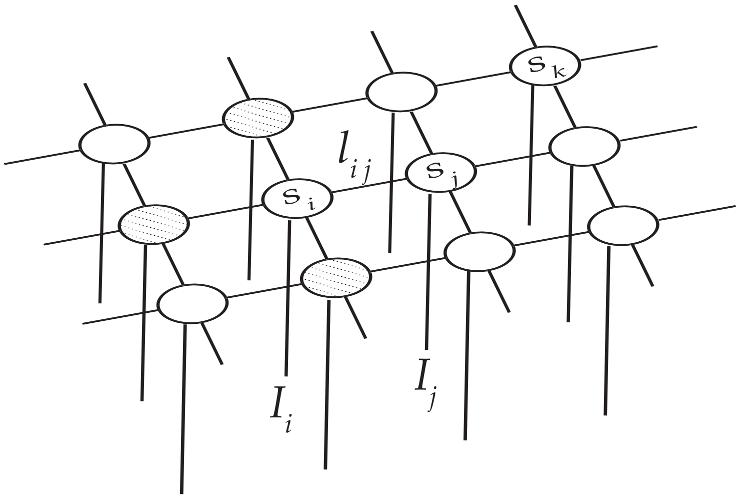
Need to solve task flexibility

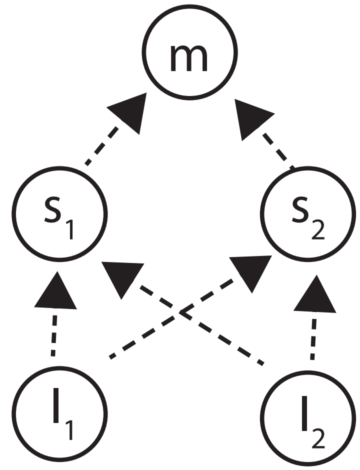
Vision stimulates and support answers to a limitless range of questions. Human vision doesn't just recognize, it interprets scenes.

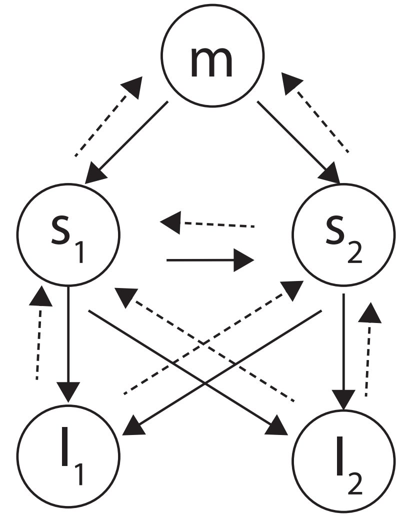
e.g. description of the fox

"One can see that there is an animal, a fox--in fact a baby fox. It is emerging from behind the base of a tree not too far from the viewer, is heading right, high-stepping through short grass, and probably moving rather quickly. Its body fur is fluffy, reddish-brown, relatively light in color, but with some variation. It has darker colored front legs and a dark patch above the mouth. Most of the body hairs flow from front to back...and what a cute smile, like a dolphin."

graphical models







lateral organization,

lateral processing, reciprocal interactions between feature of similar type

prepare to feedforward

indexing?

hierarchical organization:

feedforward processing

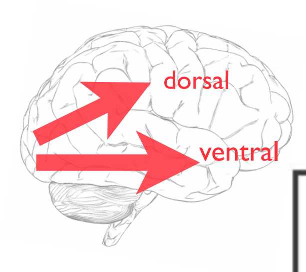
speed

hierarchical organization:

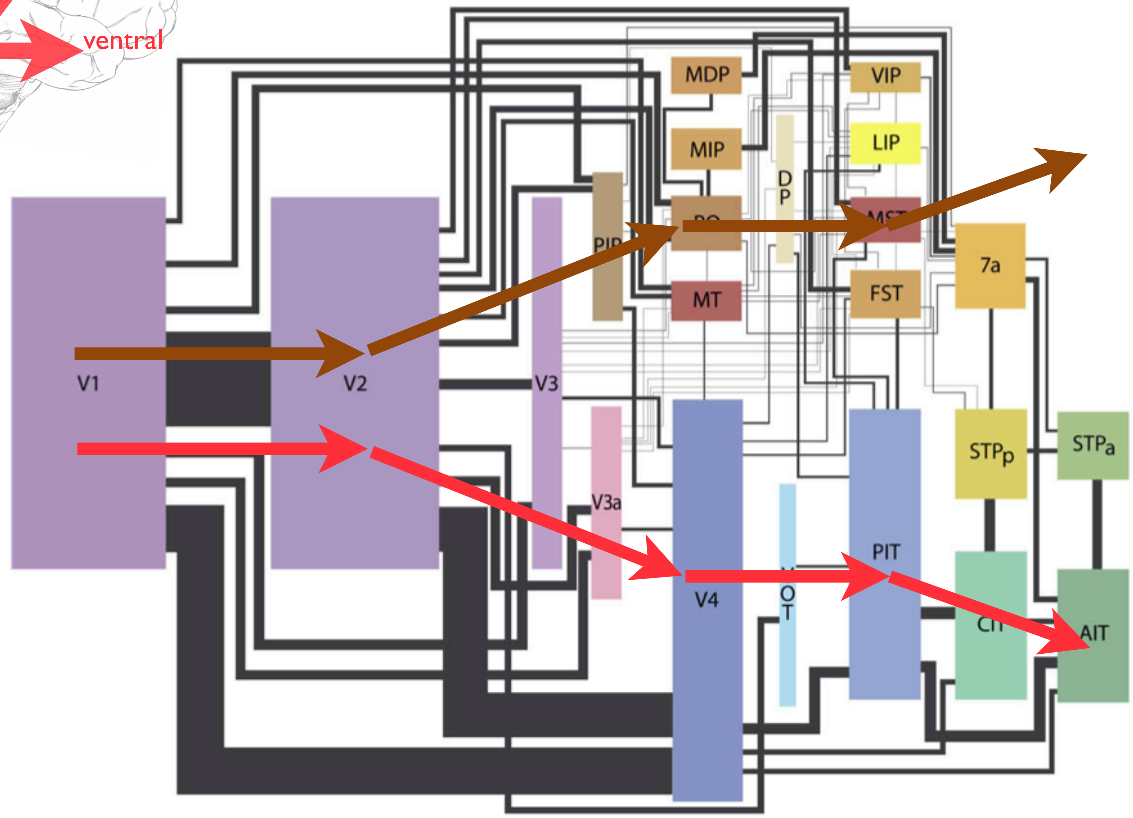
feedforward, feedback & lateral processing

task flexibility & robustness

Feedfoward



from monkey



dorsal

object-viewer properties
for spatial and motor actions

ventral

intrinsic object properties
for identification

feedforward

What determines the different selectivities for pathways and areas?

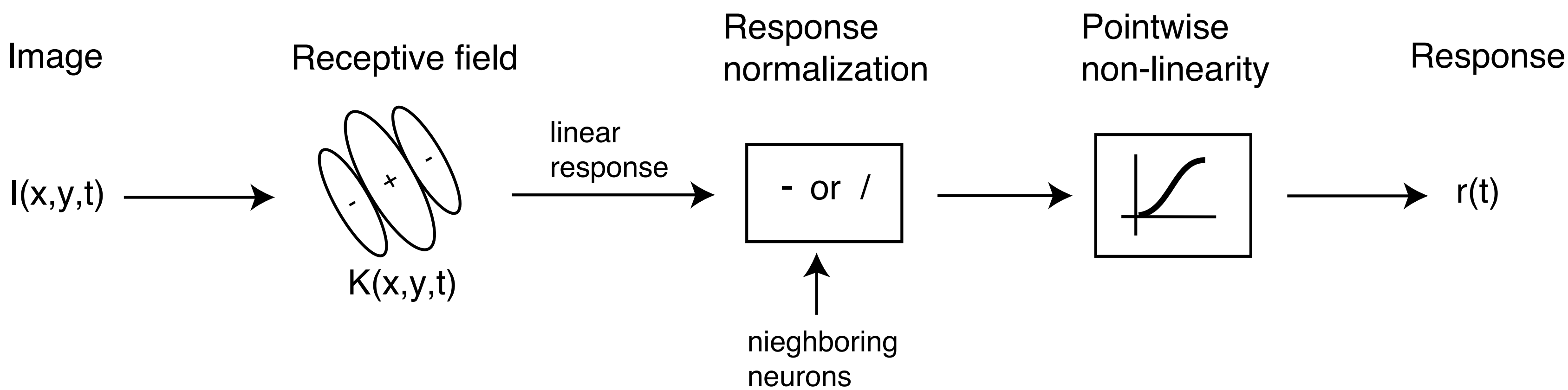
image information required for different basic tasks

...but lots of tasks

Wallisch, P., & Movshon, J. A. (2008). Structure and Function Come Unglued in the Visual Cortex.

“standard” feedforward model

for V1



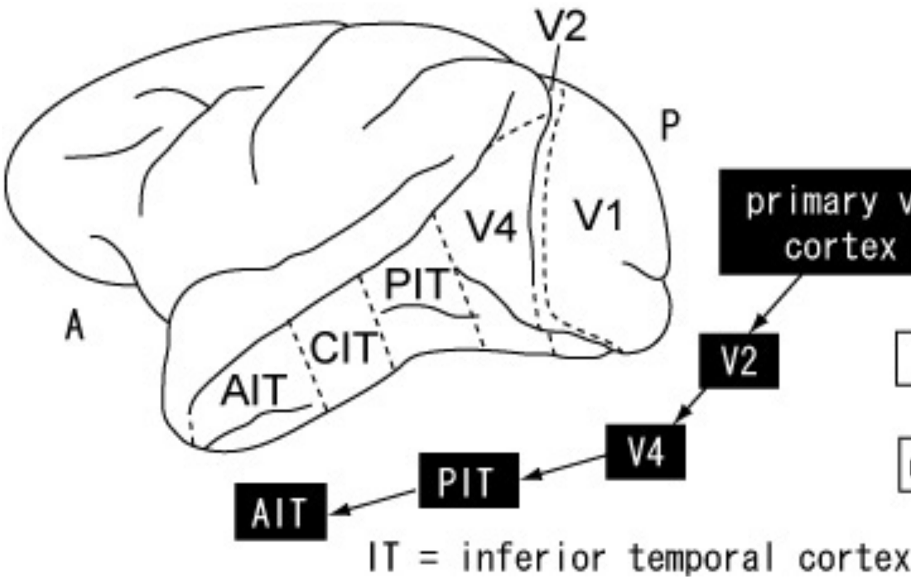
convolution — similar filtering operations repeated over space

Similar filtering operations repeated between subsequent levels

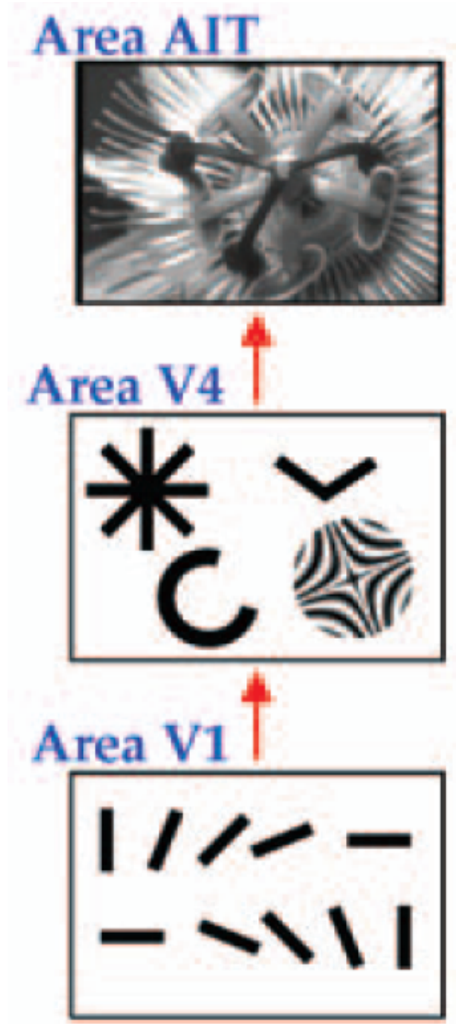
$V_n \rightarrow V_{n+1}$

deep convolutional networks

Hierarchical models of object recognition



bread and butter of ventral stream modeling



Hegde and Felleman, 2007

Hierarchical models for feature extraction for recognition

Local features progressively grouped into more structured representations

- edges => contours => fragments => parts => objects

Selectivity/invariance trade-off

- Increased selectivity for object/pattern type
- Decreased sensitivity to view-dependent variations of translation, scale and illumination

Recall simple & complex cells in V1

Simple cells

- “template matching”, i.e. detect conjunctions, logical “AND”

Complex cells

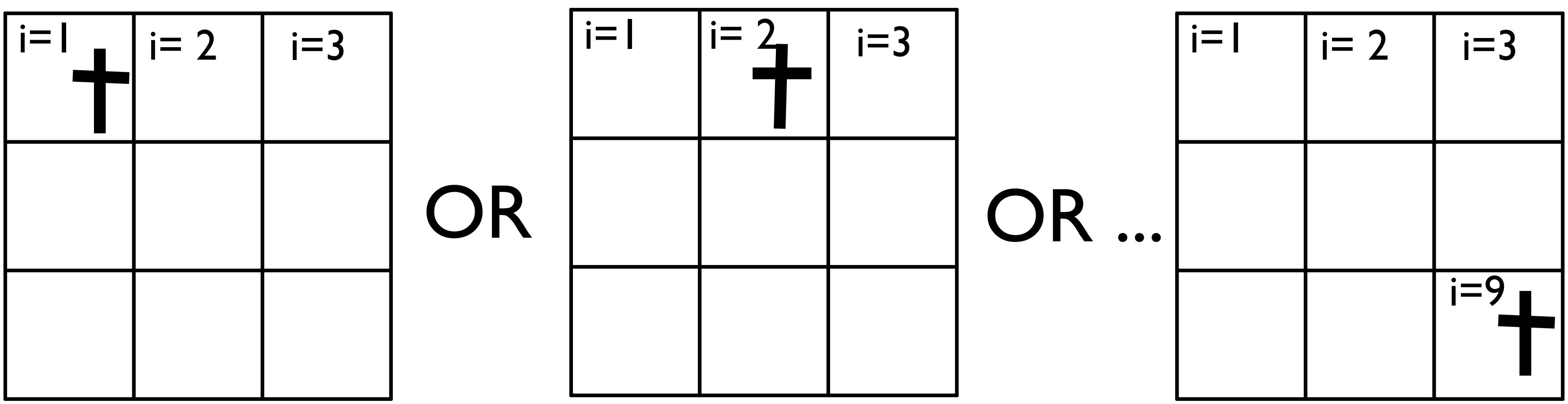
- insensitivity to small changes in position, detect disjunctions, logical “OR”

Recognition as the hierarchical detection of “disjunctions of conjunctions”

Recognize the letter "t"

"t" is represented by the conjunction of a vertical and horizontal bar:

| AND — = t



which can occur at any one of many locations i

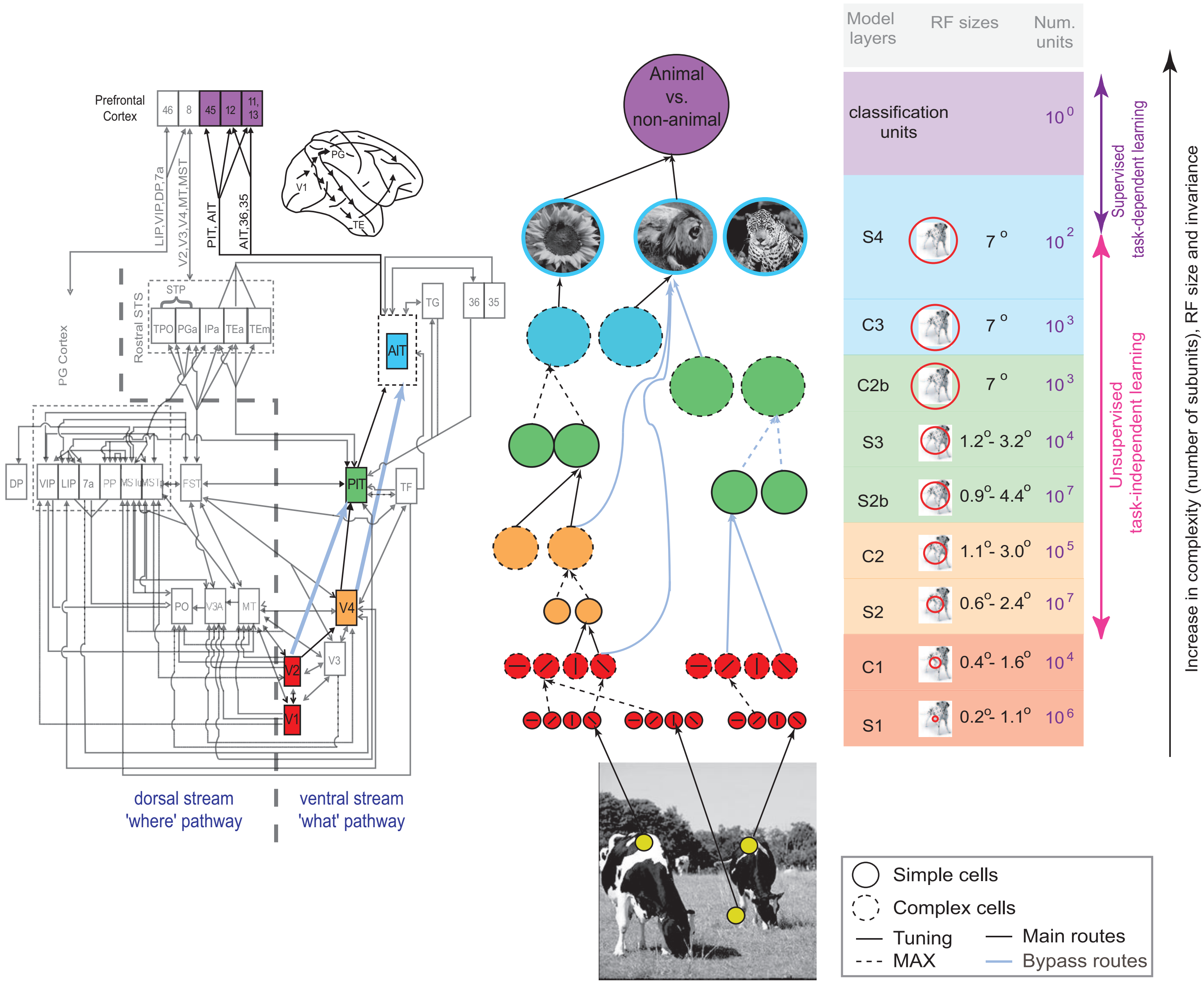
"t": h_1 && v_1 || h_2 && v_2 || h_3 && v_3...

recognition in the ventral pathway

How do neurons compute the ANDs and ORs?

A repeating theme:

Local spatial filters (simple and complex cell-like) arranged in a hierarchy can be built up to enable visual recognition



Serre, Oliva & Poggio, 2007

What determines feature hierarchies?

Grouping to form more abstract features, given image regularities that support tasks

— “hand - wire” (Riesenhuber and Poggio, …)

— unsupervised learning based on based on successive discovery of image regularities (Barlow)

- detecting “suspicious coincidences”:
    - Is p(feature A, feature B) >> p(feature A) p(feature B)
    - if so, recode to remove dependence. E.g. contingent adaptation example
- advantage of general features. but perhaps mainly useful at lower levels of the hierarchy

— supervised learning

- — “20 questions” approach (Ephstein et al.)
    - find diagnostic features that distinguish the categories for the most important tasks to determine the top level
    - repeat at a lower level of abstract to find sub-features that distinguish the diagnostic features
    - …and so forth
- deep convolutional networks

What determines feature hierarchies?
An example for one level of abstraction

Need features for rapid, accurate generalization, given a visual task requirement.

Find features of "intermediate complexity", i.e. image "fragments", that are most informative for category distinctions

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. Nature Neuroscience

Object recognition in the context of a task requirement

What do these scenes have in common?

“Up” curbs-- requiring a step up

Distinguish from non “up curbs”

...that do not require a step up and require different actions

Learning based on informative fragments for the task

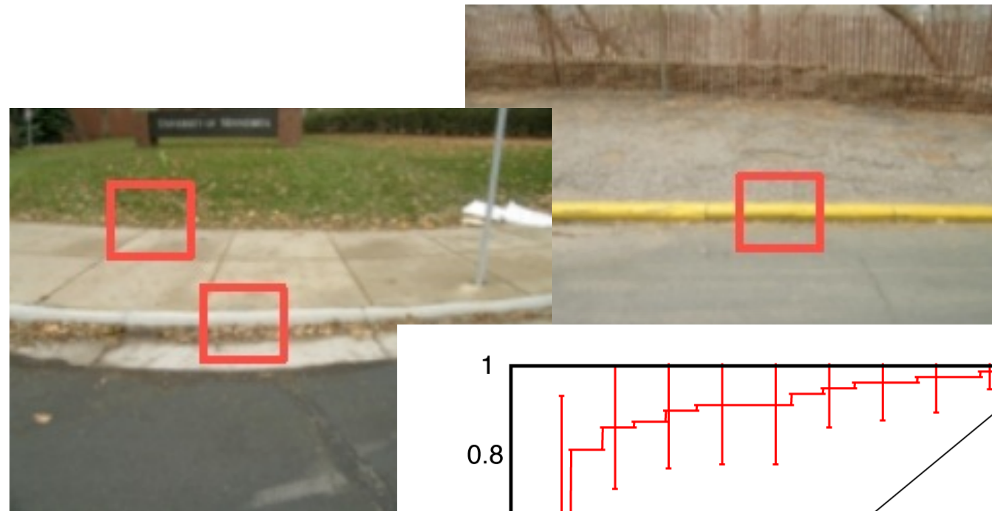
Algorithm finds fragments that maximize mutual information

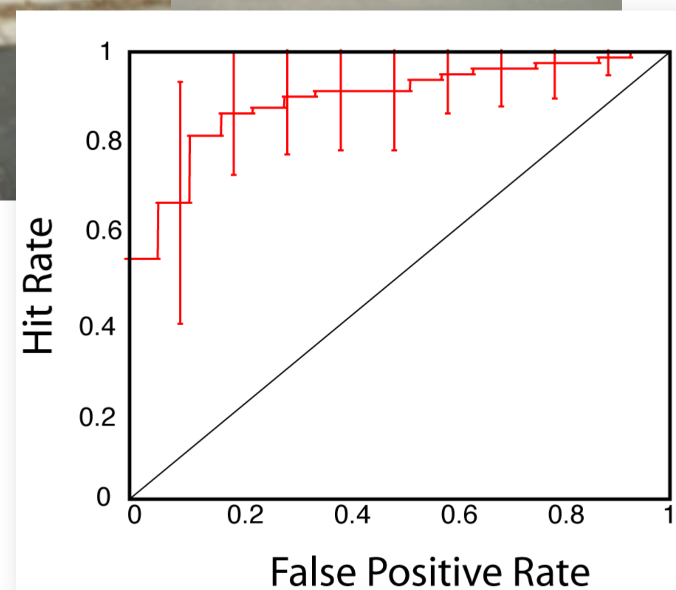
Detect "up curbs" from an approach angle that requires a step.

View-specific

Works well

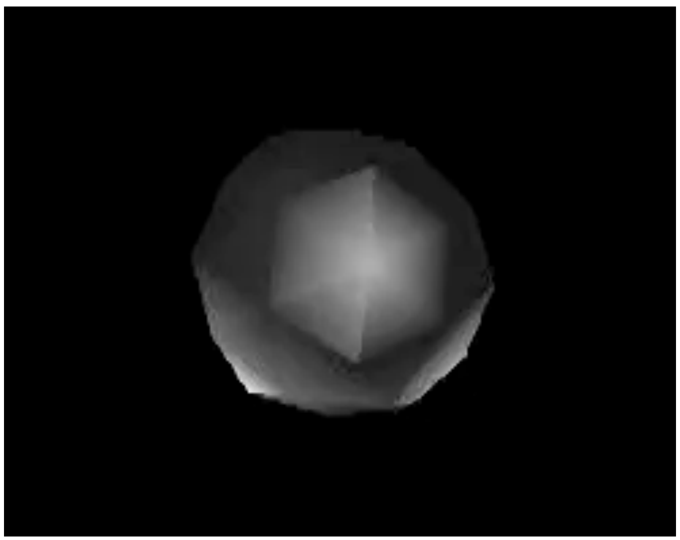
Experimentally tractable



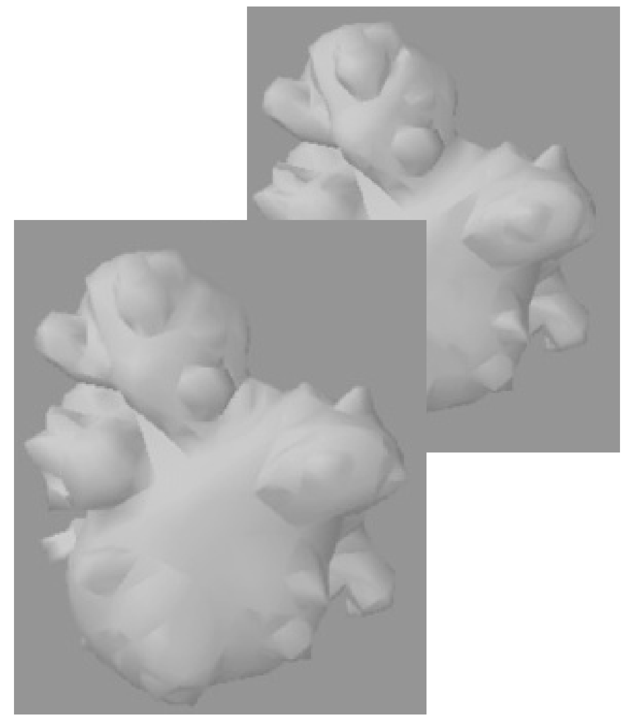


Evgeniy Bart

Do people learn to use fragments of predicted "intermediate complexity"



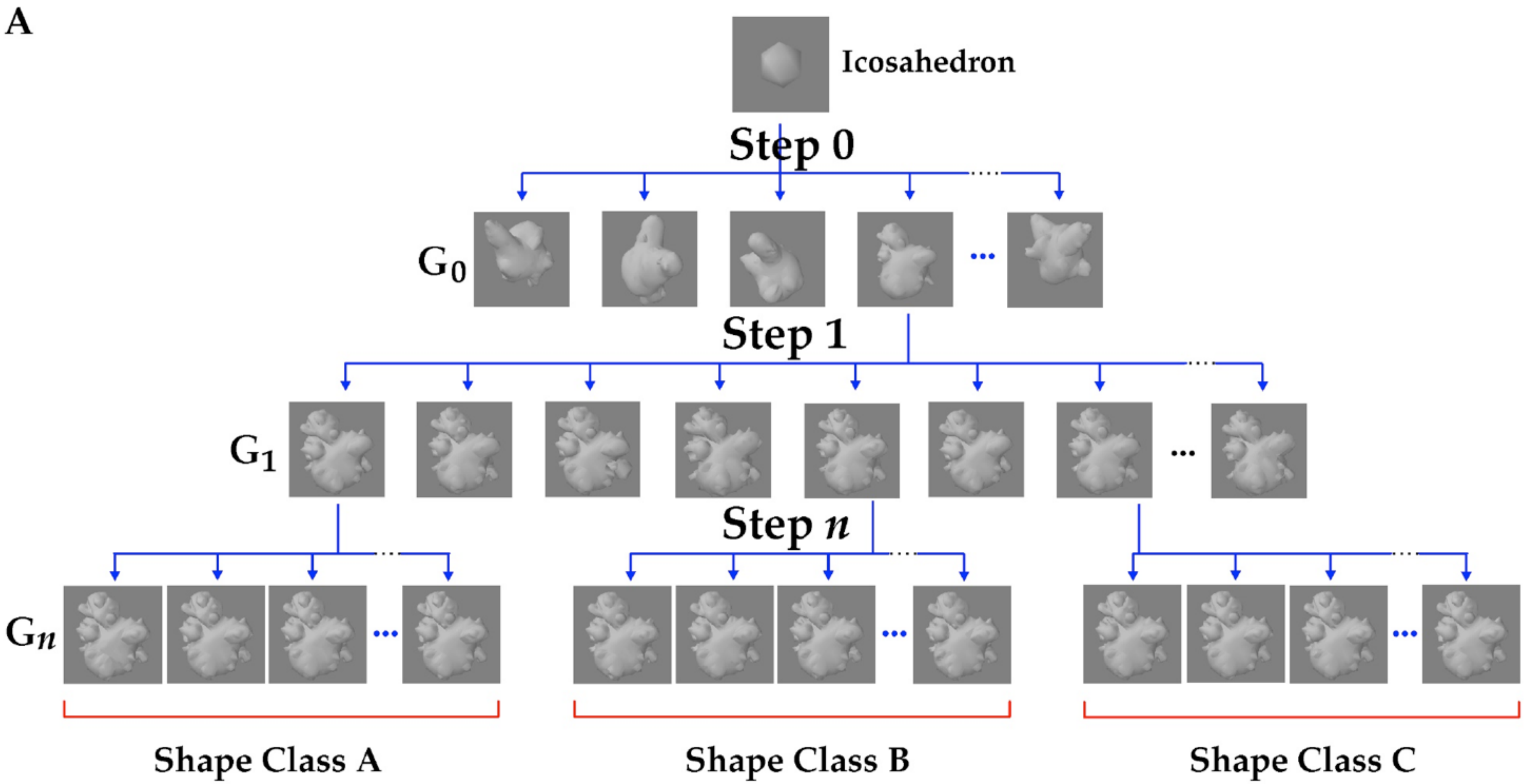
Virtual morphogenesis



Brady, M. J., & Kersten, D. (2003). Bootstrapped learning of novel objects. Journal of Vision, 3(6), 413–422.

Generating naturalistic object classes

Virtual Phylogenesis



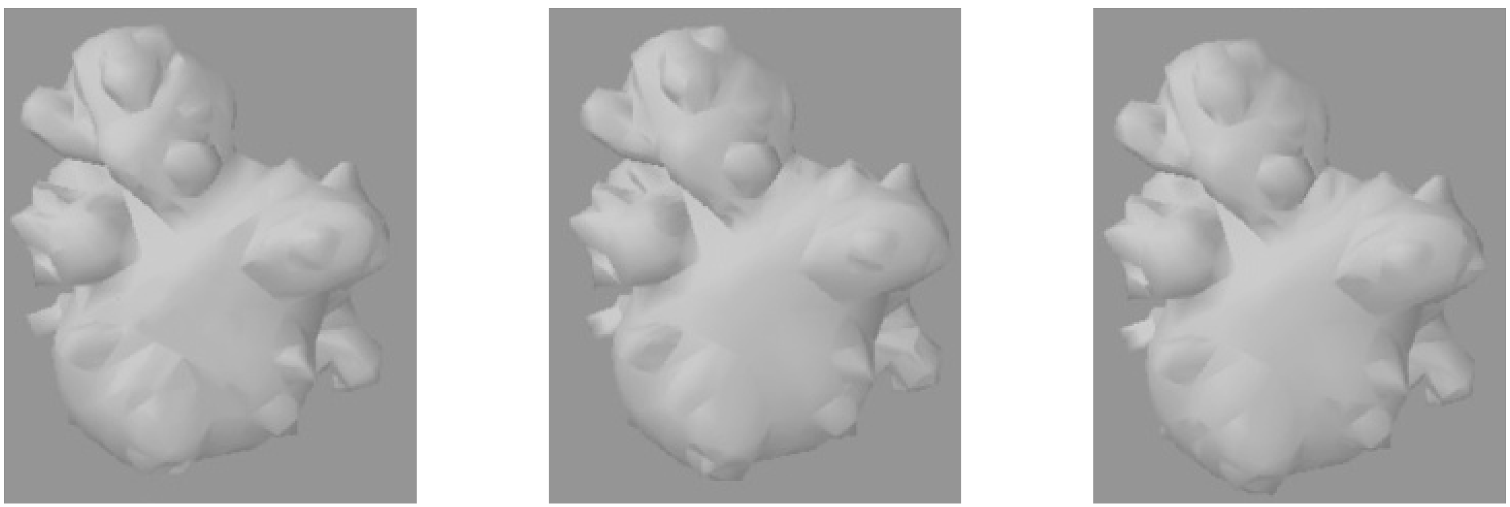
Hegde, J., Bart, E., & Kersten, D. (2008). Fragment-Based Learning of Visual Object Categories. Curr Biol. 18, 597-601

Training

Member of category A or B?

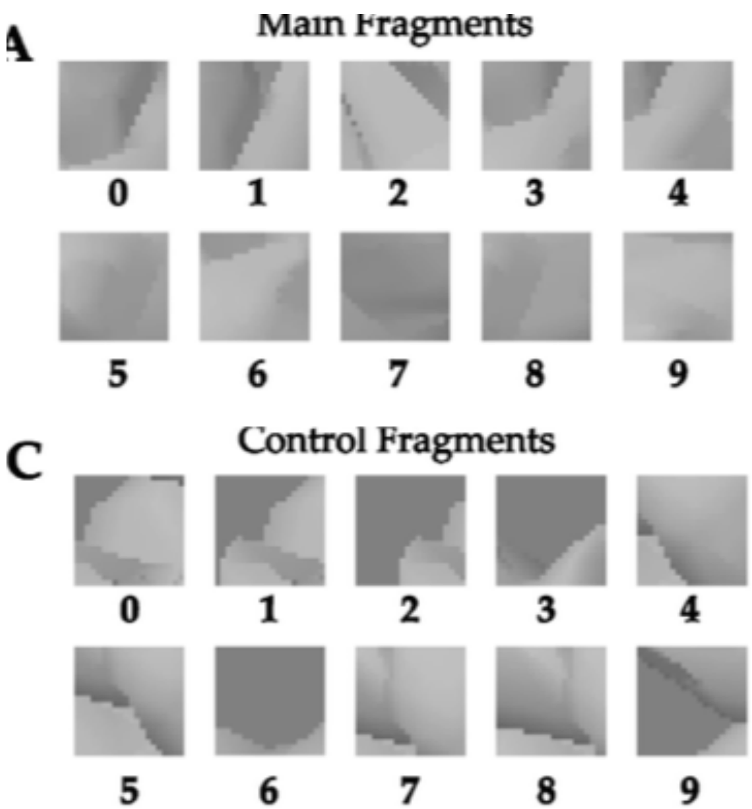
A

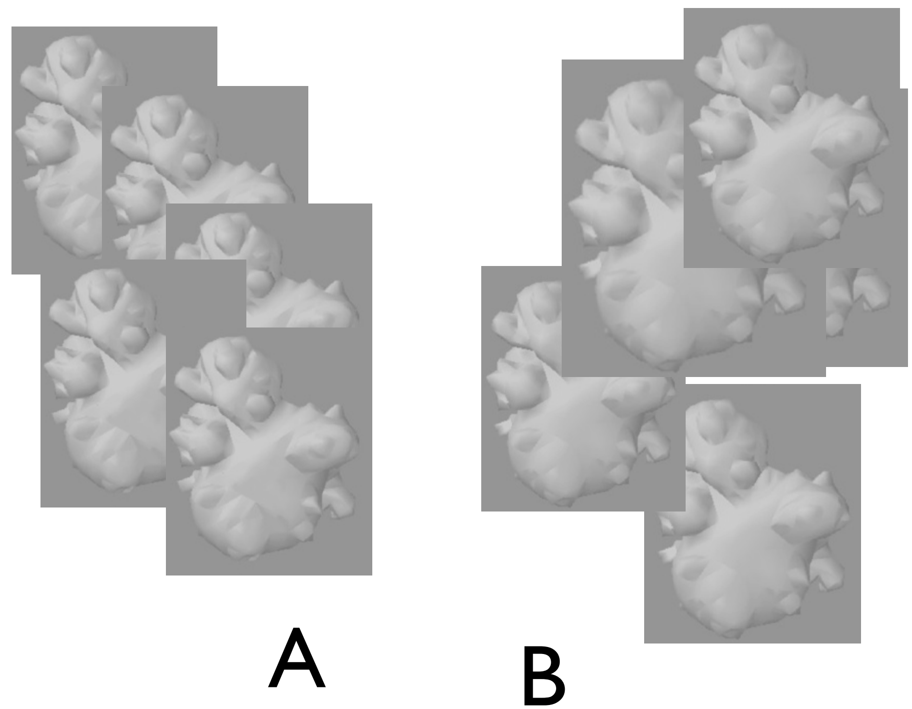
B



Results

Features of intermediate complexity (local image patches) predicted human observers ability to classify new objects from learned categories





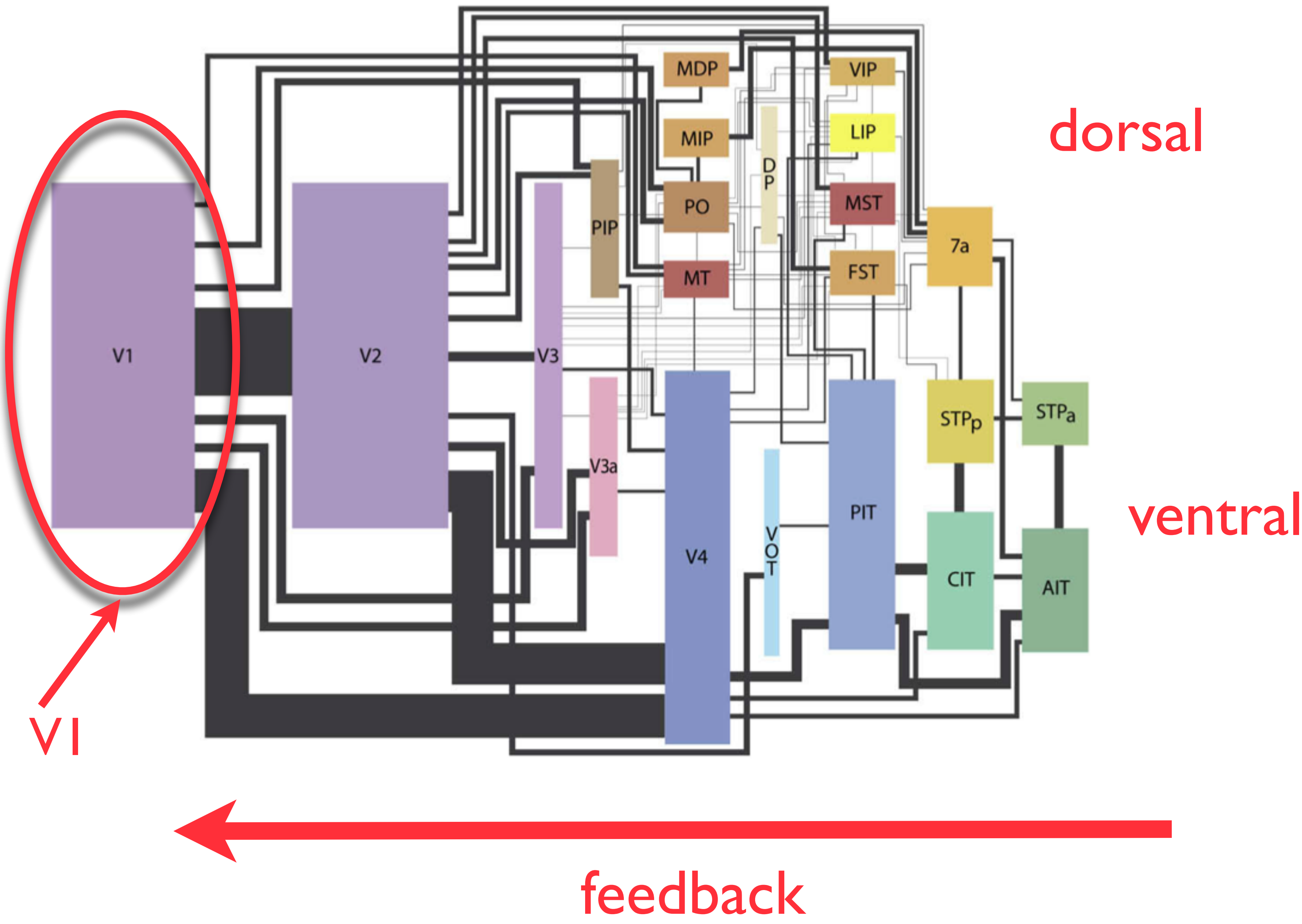
Hegde, J., Bart, E., & Kersten, D. (2008). Fragment-Based Learning of Visual Object Categories. Curr Biol. 18, 597-601

Puzzle?

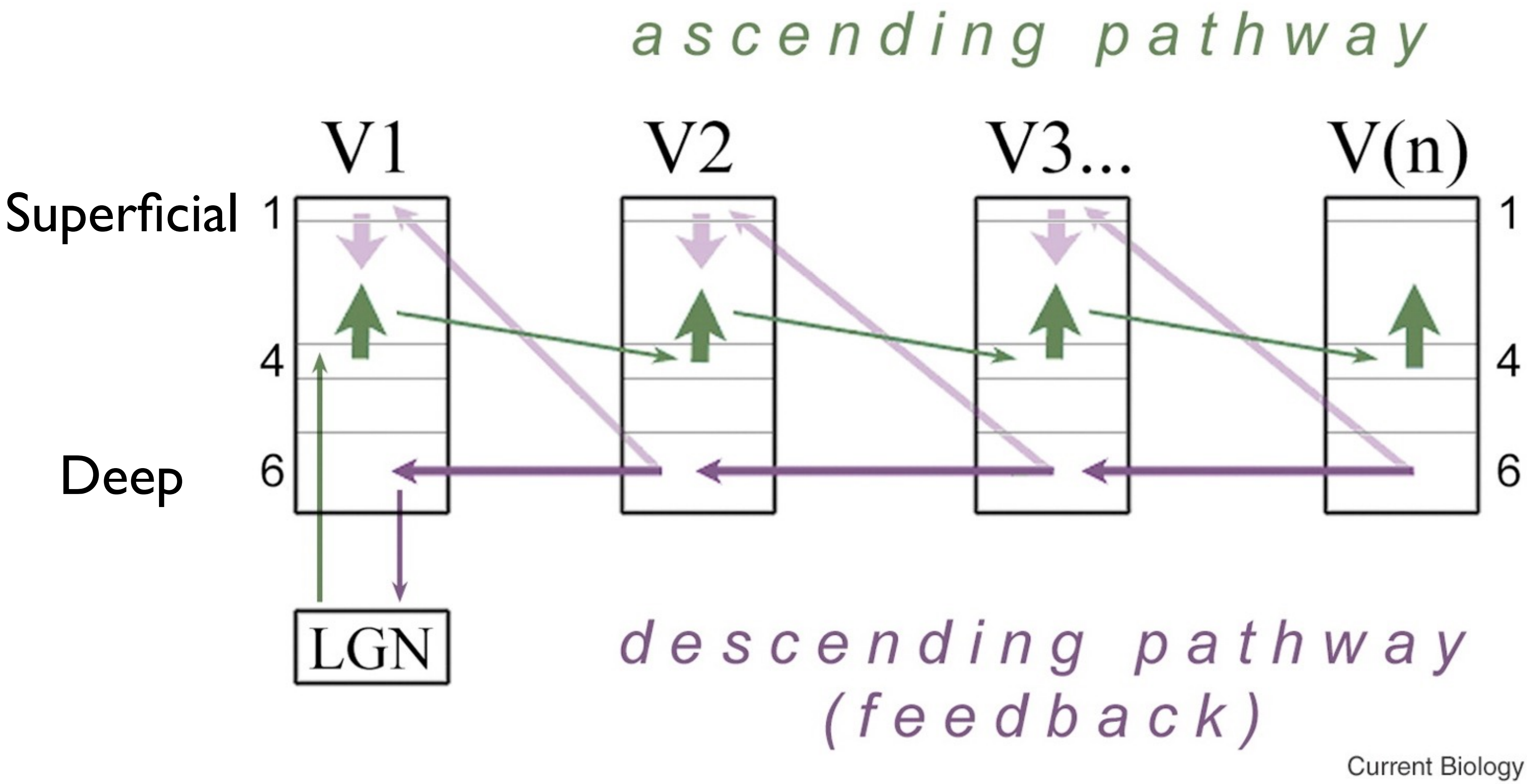
How to learn when objects aren’t experienced
in isolation?

This is the bootstrapping problem
(Brady & Kersten)

Feedback



MDP
VIP
MIP
LIP
D P
PIP
PO
MST
7a
MT
FST
V1
V2
V3
STPp
STPa
V3a
V4
V O T
PIT
CIT
AIT
dorsal
ventral
V1
feedback



Shipp, S. (2007). Structure and function of the cerebral cortex. CURBIO, 17(12), R443–9. doi:10.1016/j.cub.2007.03.044

Two computational strategies

Discriminative mechanisms

$p(\text{object} \mid \text{image})$
feedforward

- Computational/behavioral speed and accuracy requires effective diagnostic features to deal with the enormous variation within a pattern/object category

VanRullen, R., & Thorpe, S. J. (2001). The time course of visual processing: from early perception to decision-making. *Journal of Cognitive Neuroscience*, *13*(4), 454–461.

Generative mechanisms

$p(\text{image} \mid \text{object}) \times p(\text{object})$*
feedback

- Provide flexibility, generalization

* recall bayes: $p(\text{object} \mid \text{image}) \propto p(\text{image} \mid \text{object}) \times p(\text{object})$

Feedback functions

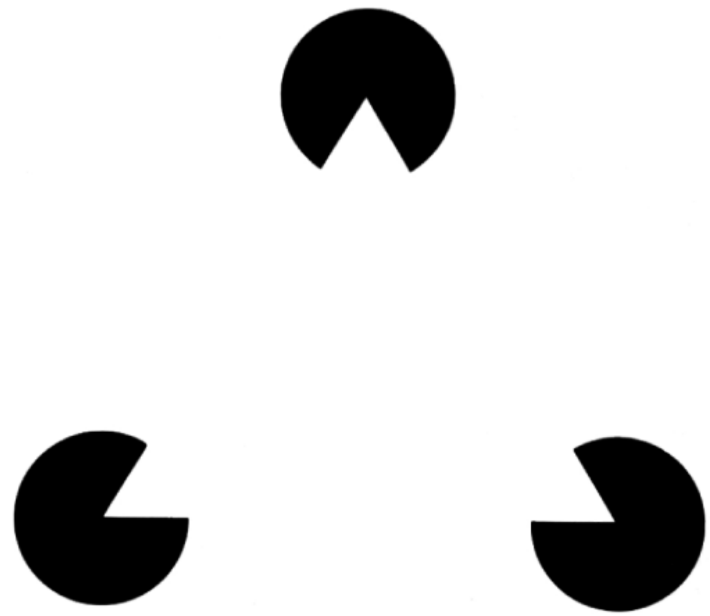
Disambiguation

- suppress explained input
- enhance explained input

The executive metaphor

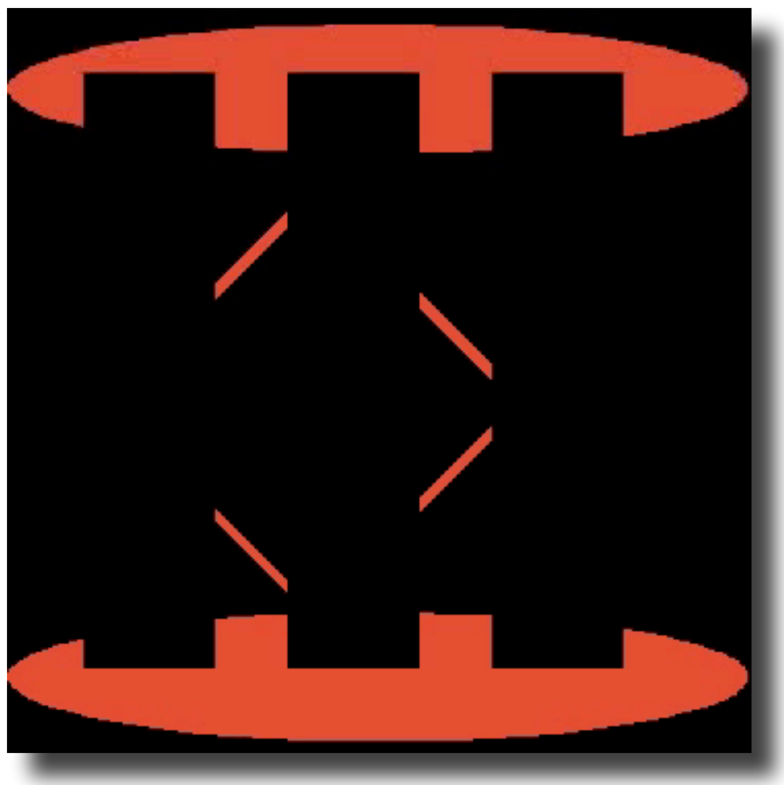
- expertise at various levels of abstraction

motivation: missing data

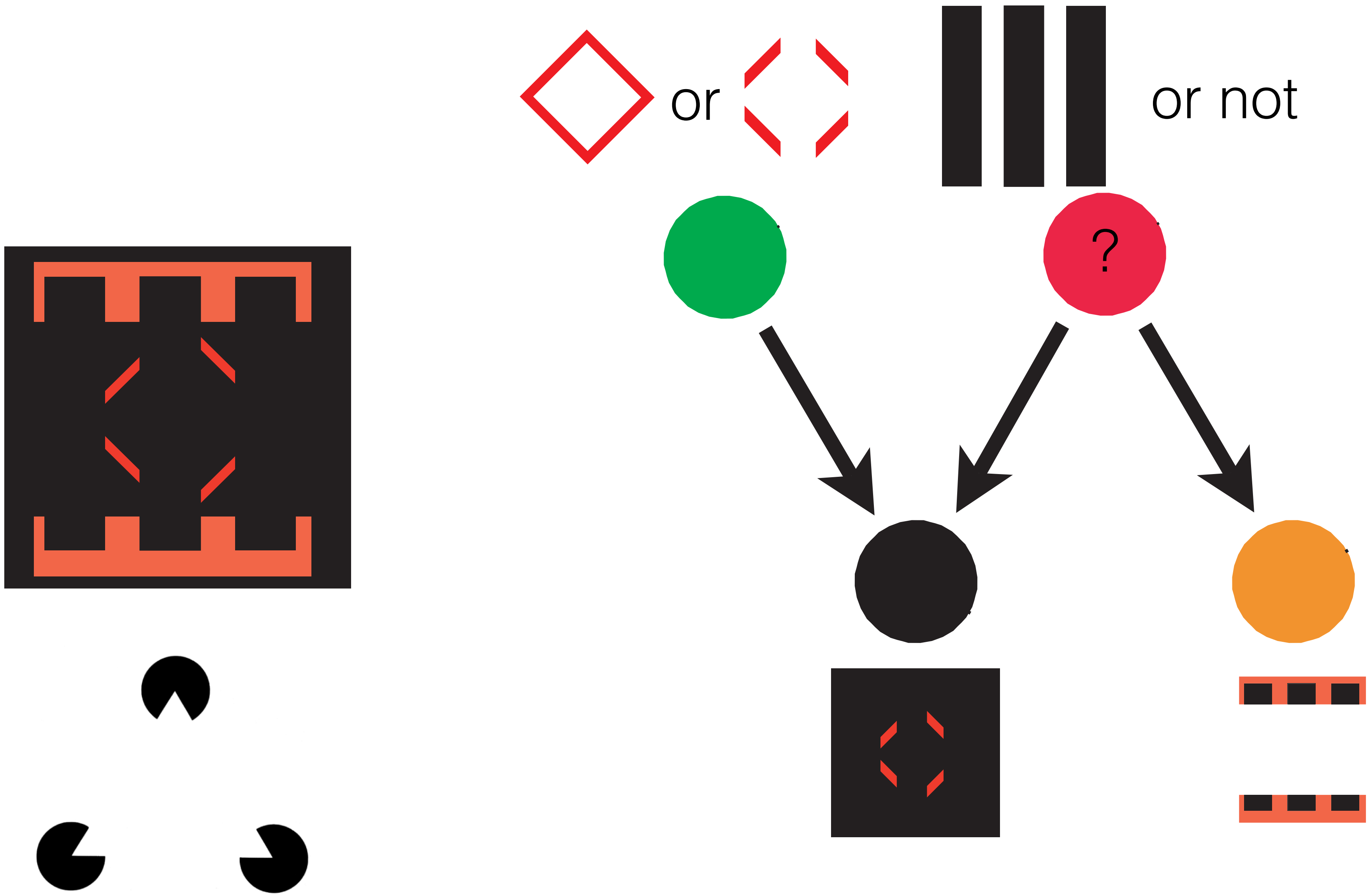


Top-down, generative models?

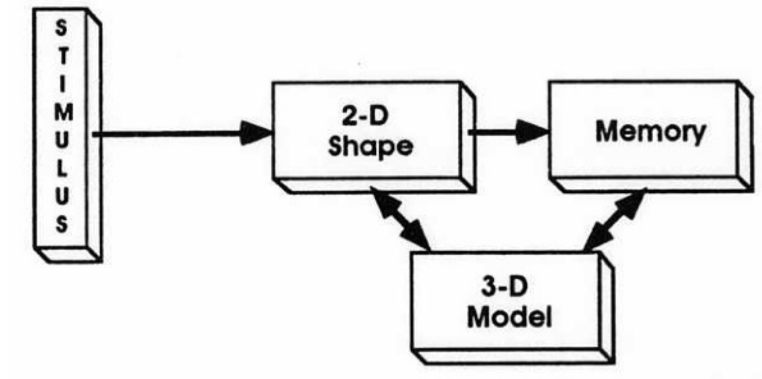
missing data & occlusion

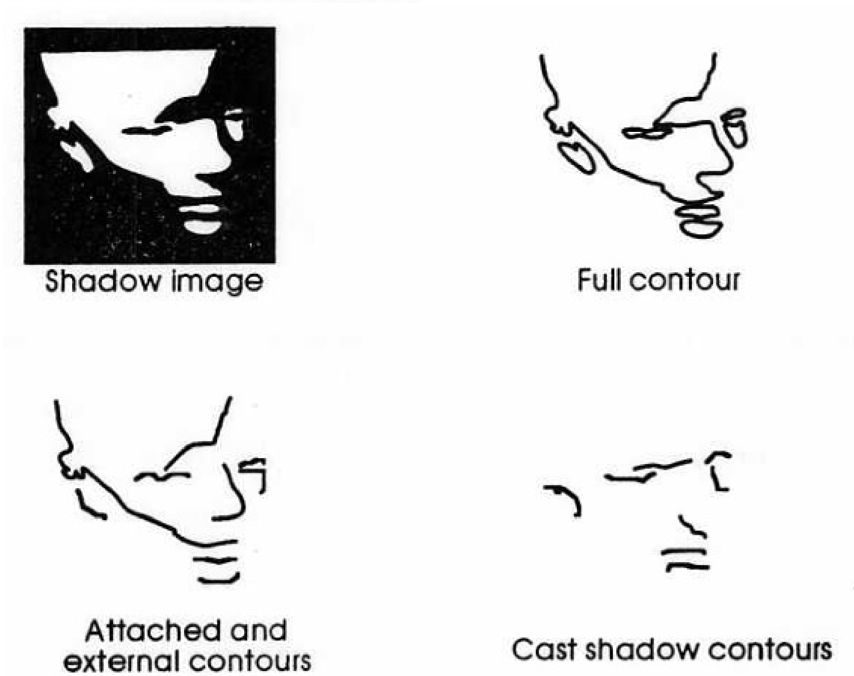


Perceptual "explaining away"



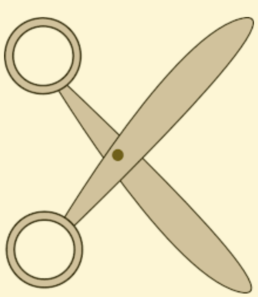
Extraneous data: recognition despite cast shadows





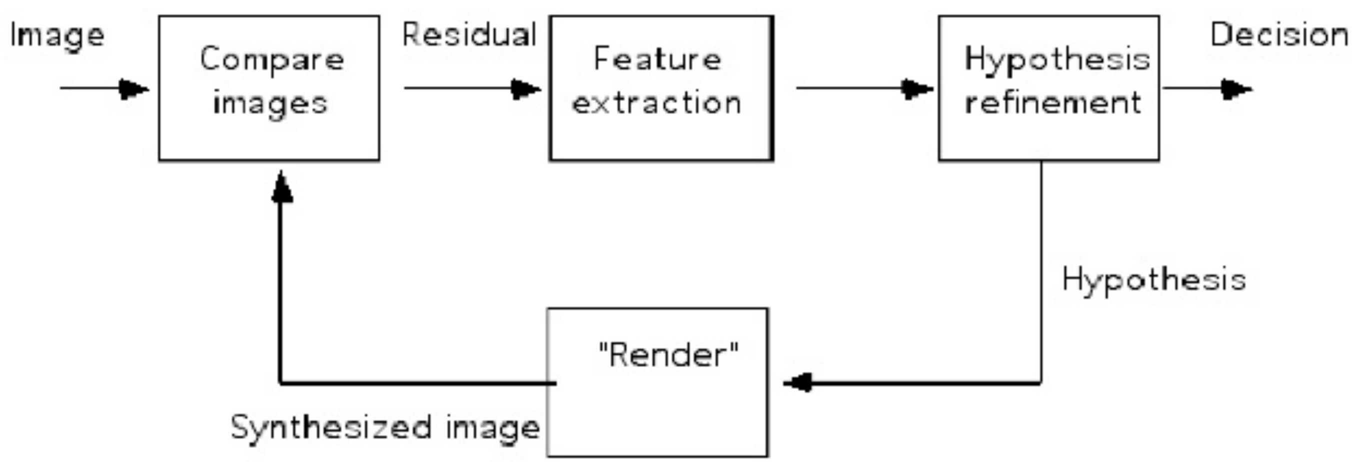
Cavanagh P (1991) What's up in top-down processing? In: Representations of Vision: Trends and tacit assumptions in vision research (Gorea A, ed), pp 295-304. Cambridge, UK: Cambridge University Press.

Object variations that haven't been seen before



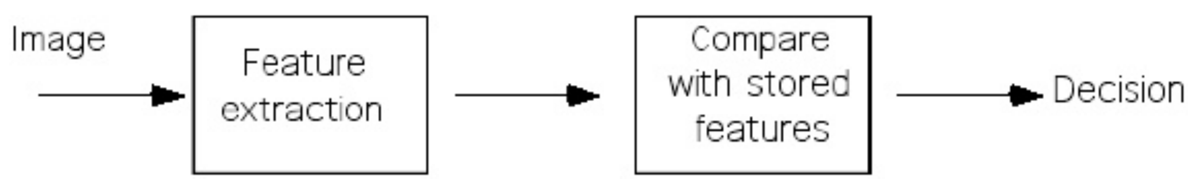
can recognize as scissors AND estimate an articulation

Suggests...



Bottom-up / Top-down

is a more complete picture than this



Bottom-up

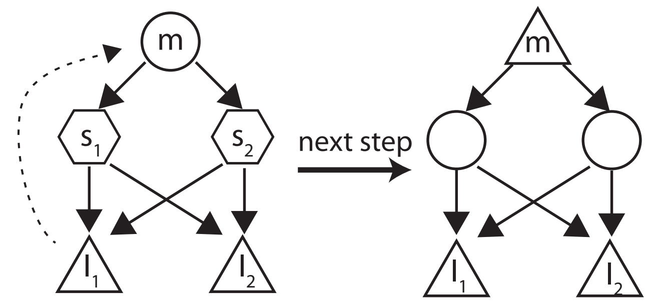
Doesn't mean that feedback is necessary for recognition (Thorpe et al.)

But top-down feedback may be important for

- achieving high-performance given uncertainty, noise, clutter
- task flexibility
- learning new object models

feedback functions?

Coarse-to-fine

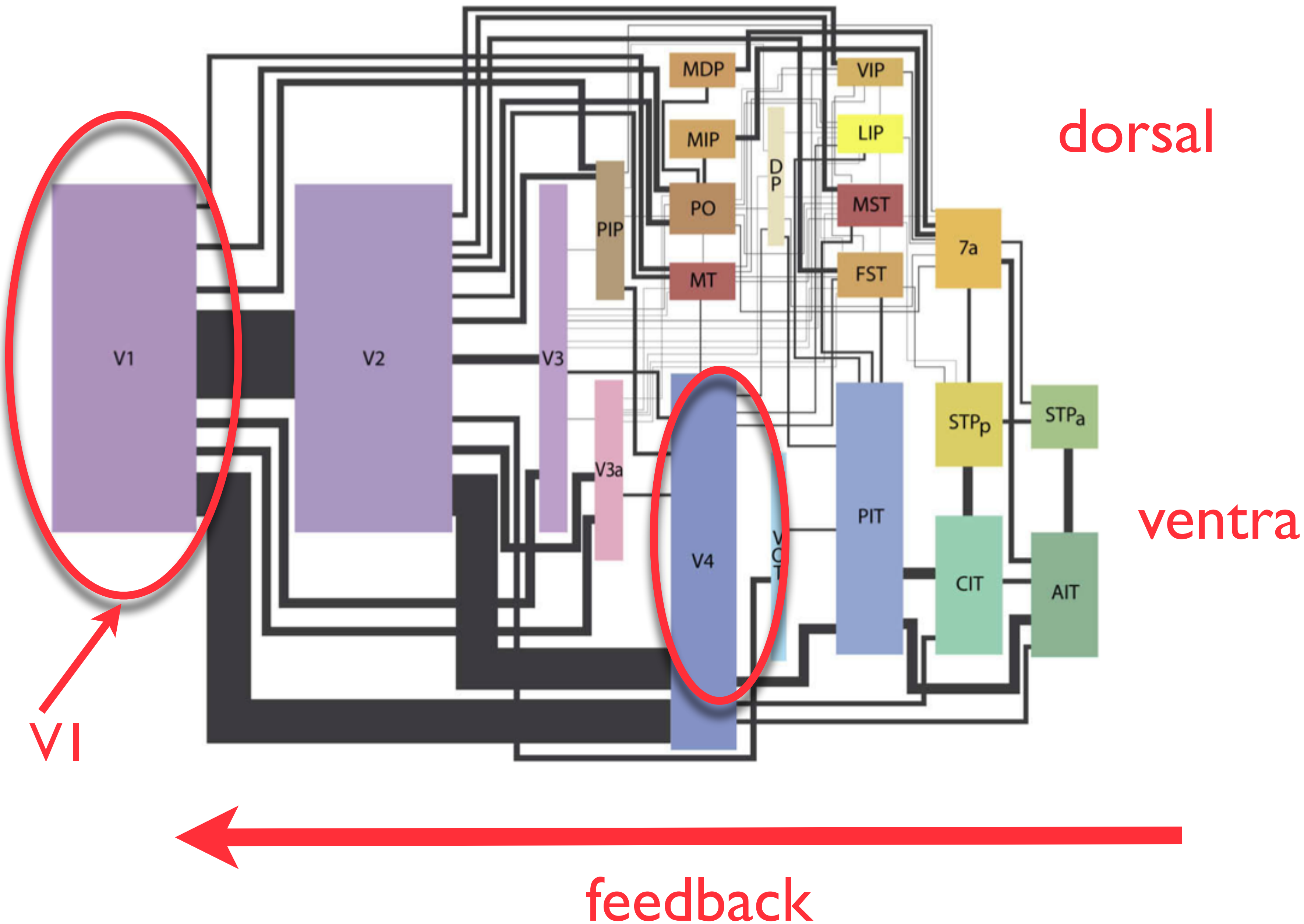


- E.g. is it a fox? If so, where is its nose?

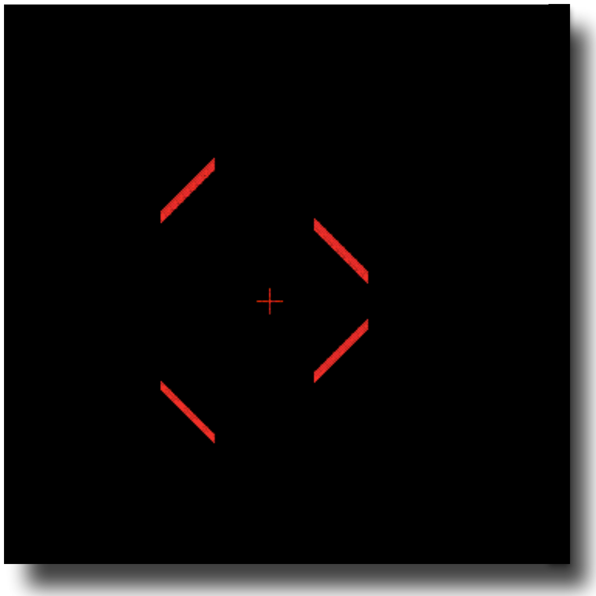
Ambiguity reduction through top-down prediction

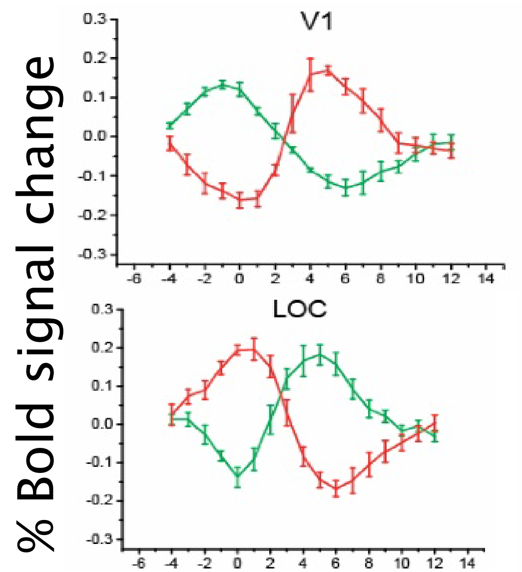
Hierarchically organized representations & expertise

Feedback


MDP
VIP
MIP
LIP
DP
PO
MST
PIP
7a
MT
FST
V1
V2
V3
STPp
STPa
V3a
V4
VOT
PIT
CIT
AIT
dorsal
ventral
V1
feedback

perceptual organization reduces activity in V1





Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002).; Fang, F., Kersten, D., & Murray, S. O. (2008).

...but non-retinotopic voxels are also suppressed (Wit et al., 2012)

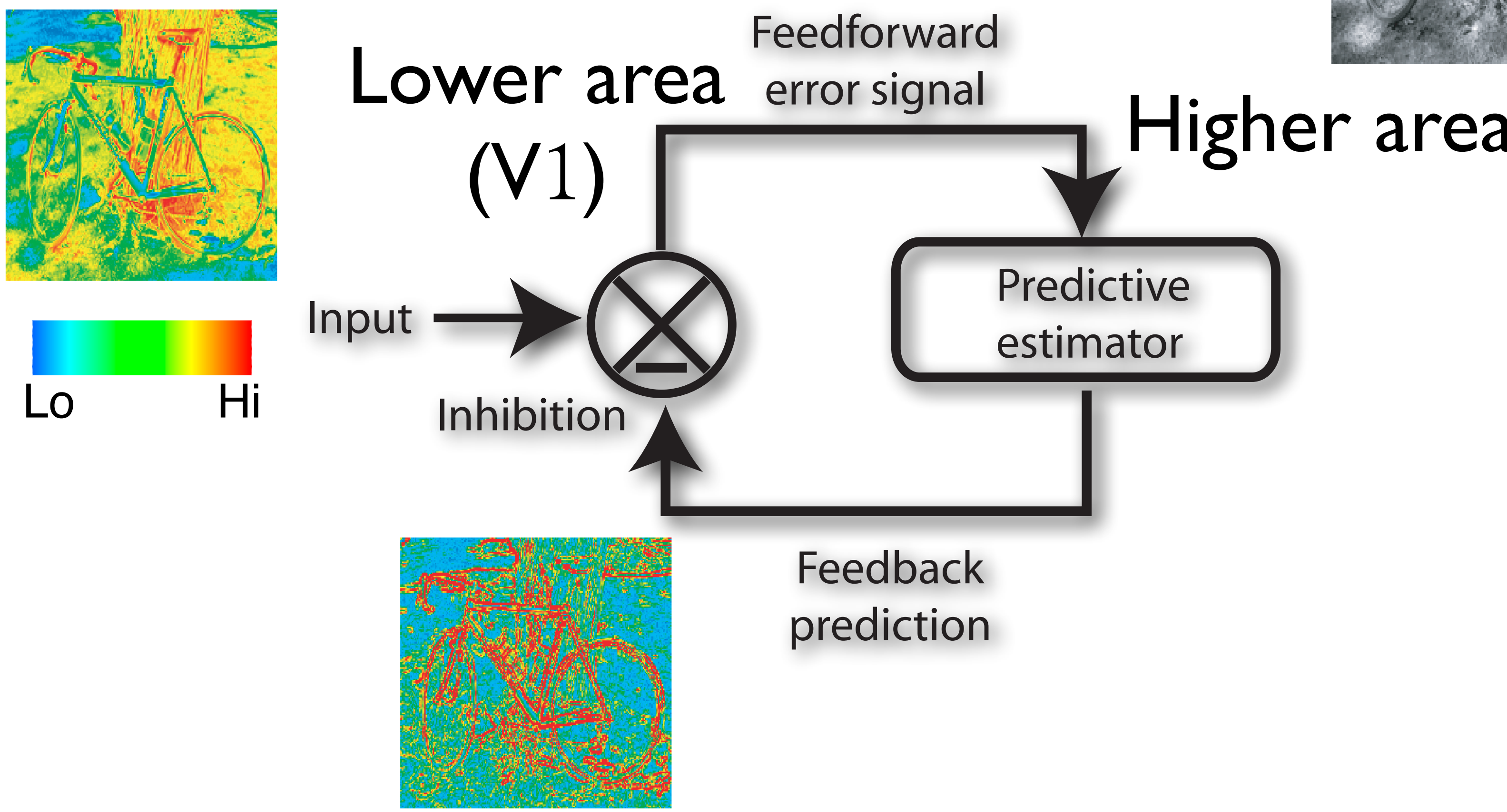
Behavioral evidence for top-down reduction of early activity? Use perceptual adaptation--the psychophysicist's electrode

Disambiguation?

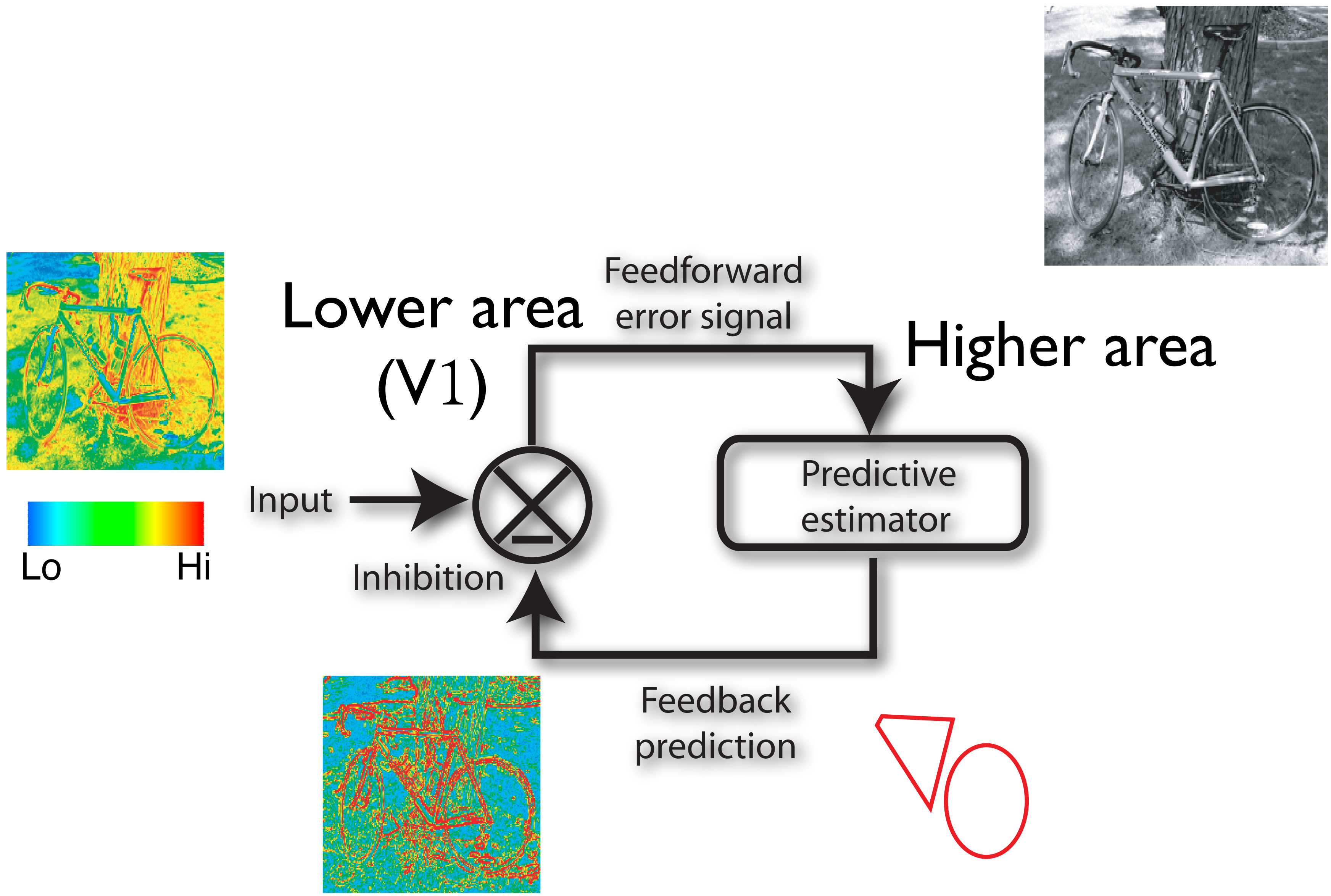
Predictive coding: suppress lower-level features that are consistent with a confident high-level interpretation. Reduce metabolic costs, signal new unexplained incoming information.

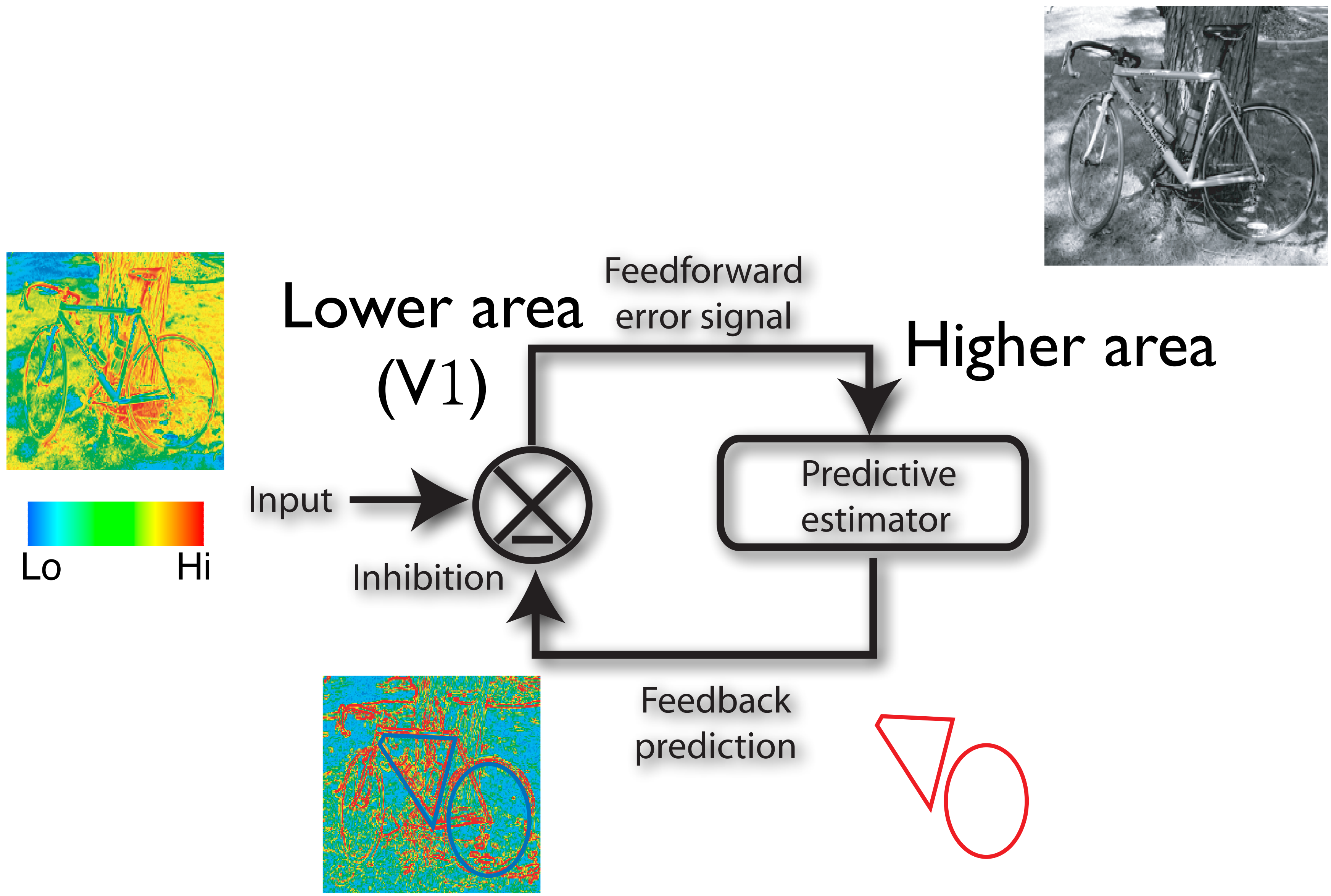
Analysis-by-synthesis. Bind lower-level information that might be required for executive tasks, e.g. fine-grain. : enhance lower-level consistent features and/or suppress inconsistent ones. Useful for representation and interpretation of novel patterns? Dealing with clutter?

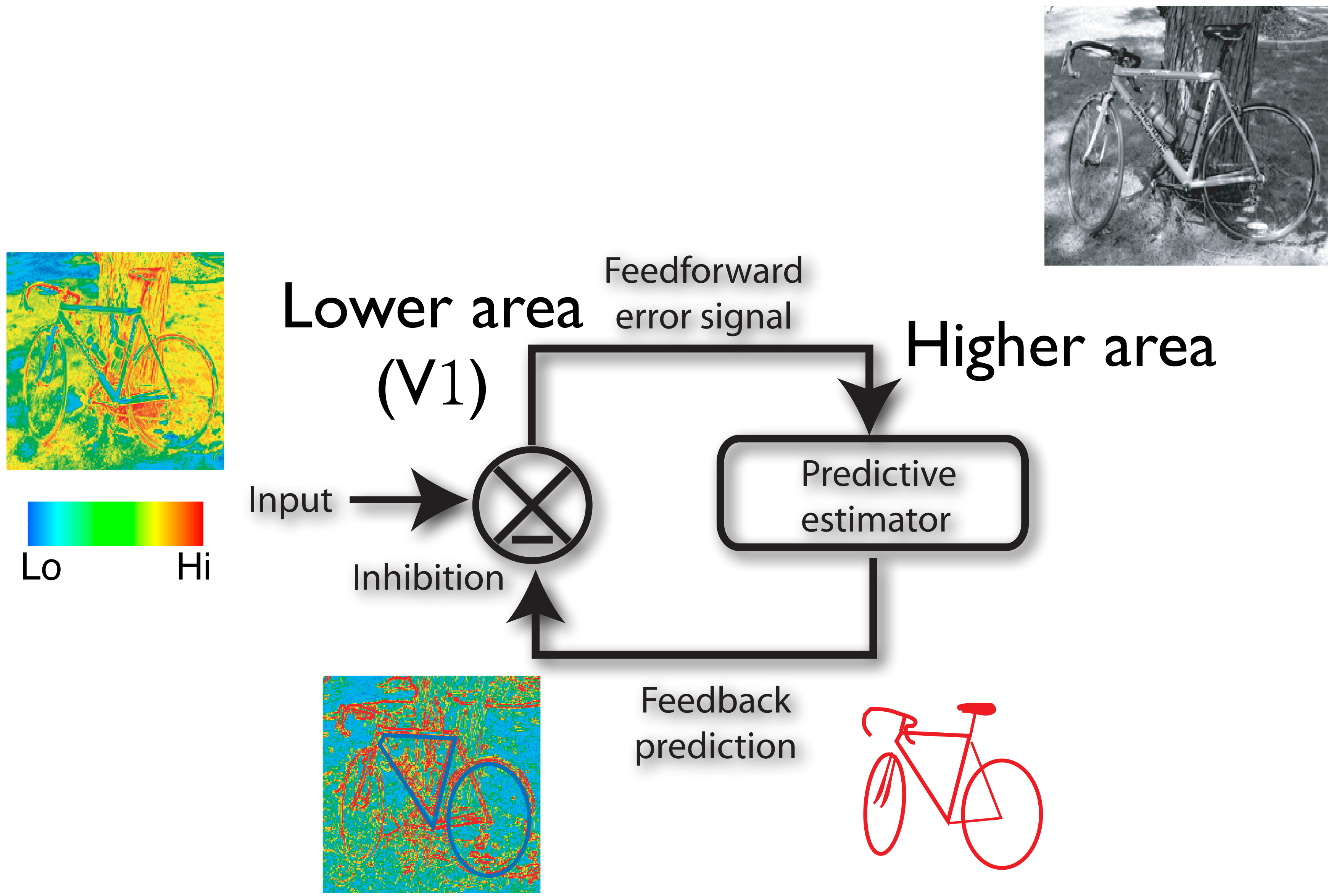
“predictive coding” through suppression of consistent features at lower levels



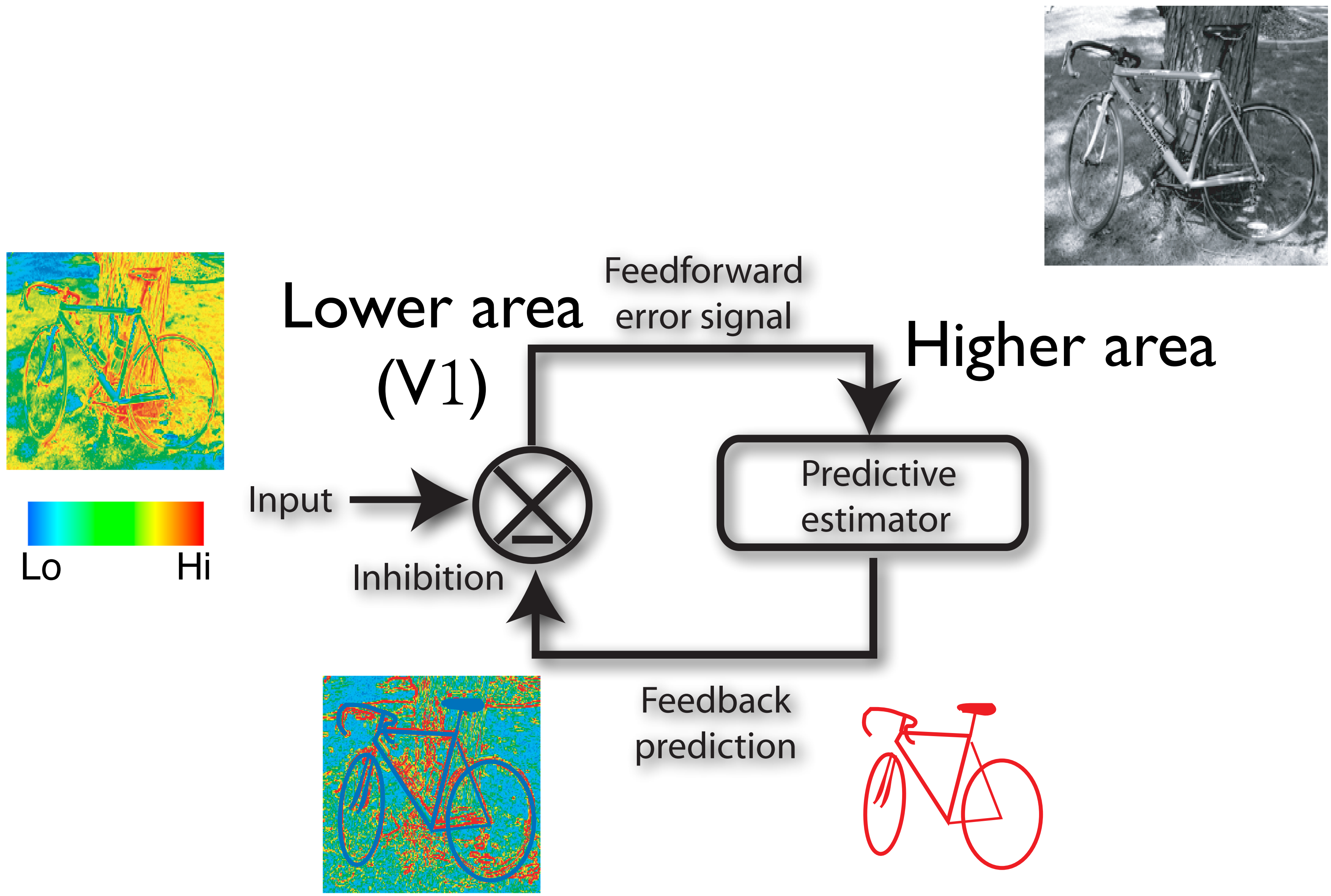
e.g. Rao, R. P., & Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. Neural Comput, 9(4), 721-763.


Lower area
(V1)
Feedforward
error signal
Higher area
Input
Inhibition
Predictive
estimator
Feedback
prediction
Lo
Hi

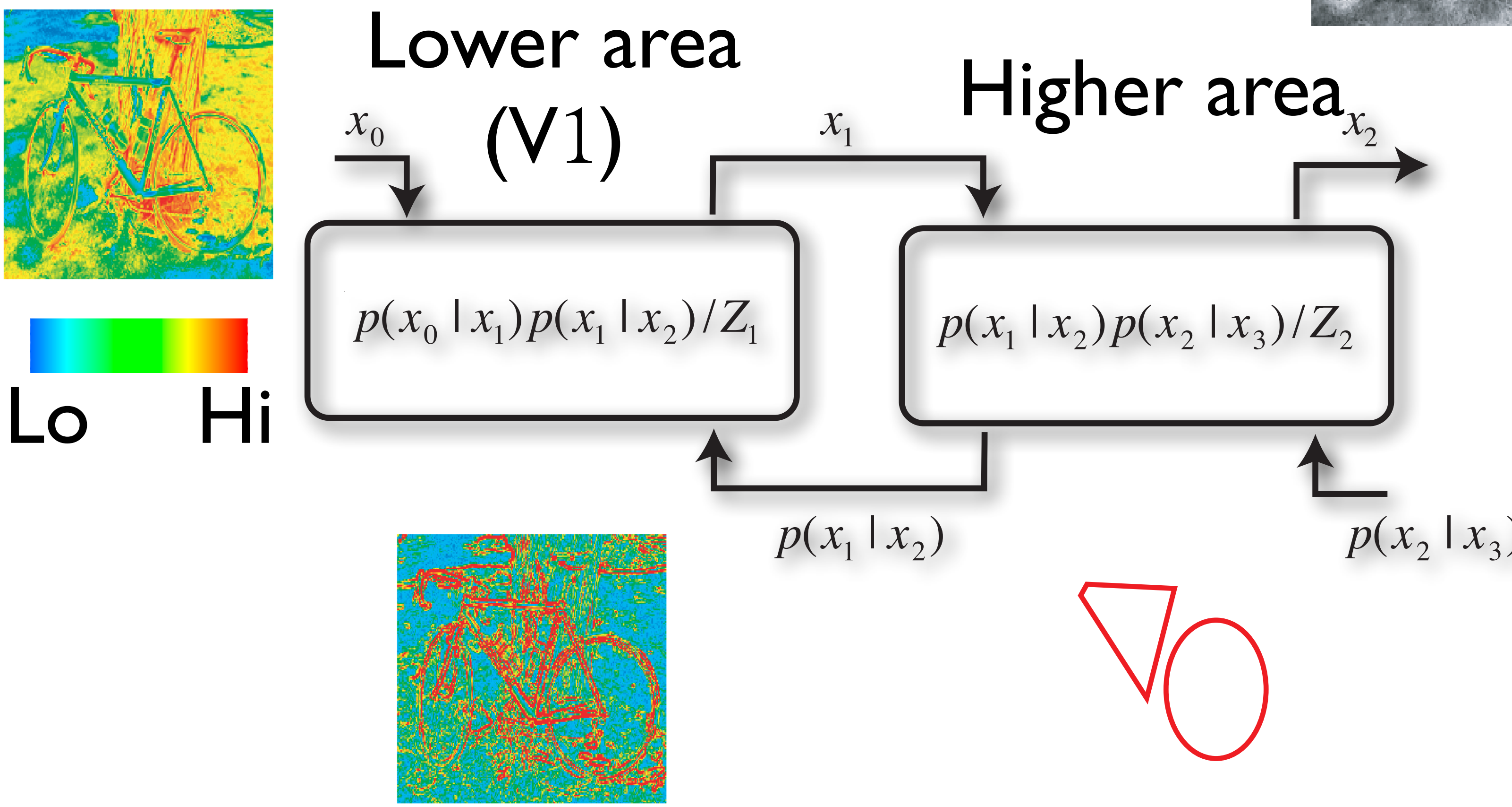

Lower area
(V1)
Feedforward
error signal
Higher area
Input
Predictive
estimator
Inhibition
Lo
Hi
Feedback
prediction



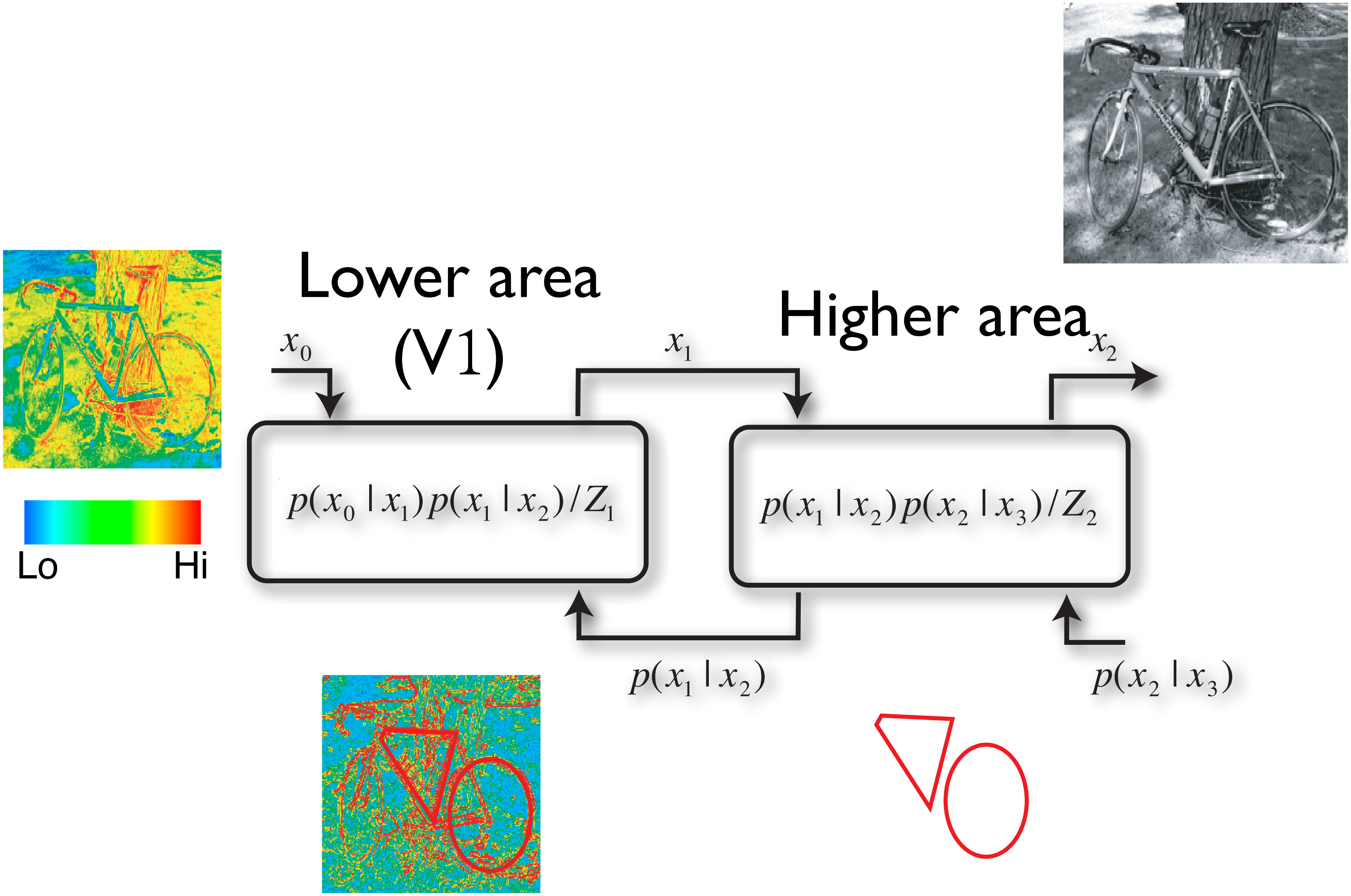
Lower area
(V1)
Feedforward
error signal
Higher area
Input
Inhibition
Predictive
estimator
Feedback
prediction
Lo
Hi


Lower area
(V1)
Feedforward
error signal
Higher area
Input
Predictive
estimator
Inhibition
Lo
Hi
Feedback
prediction

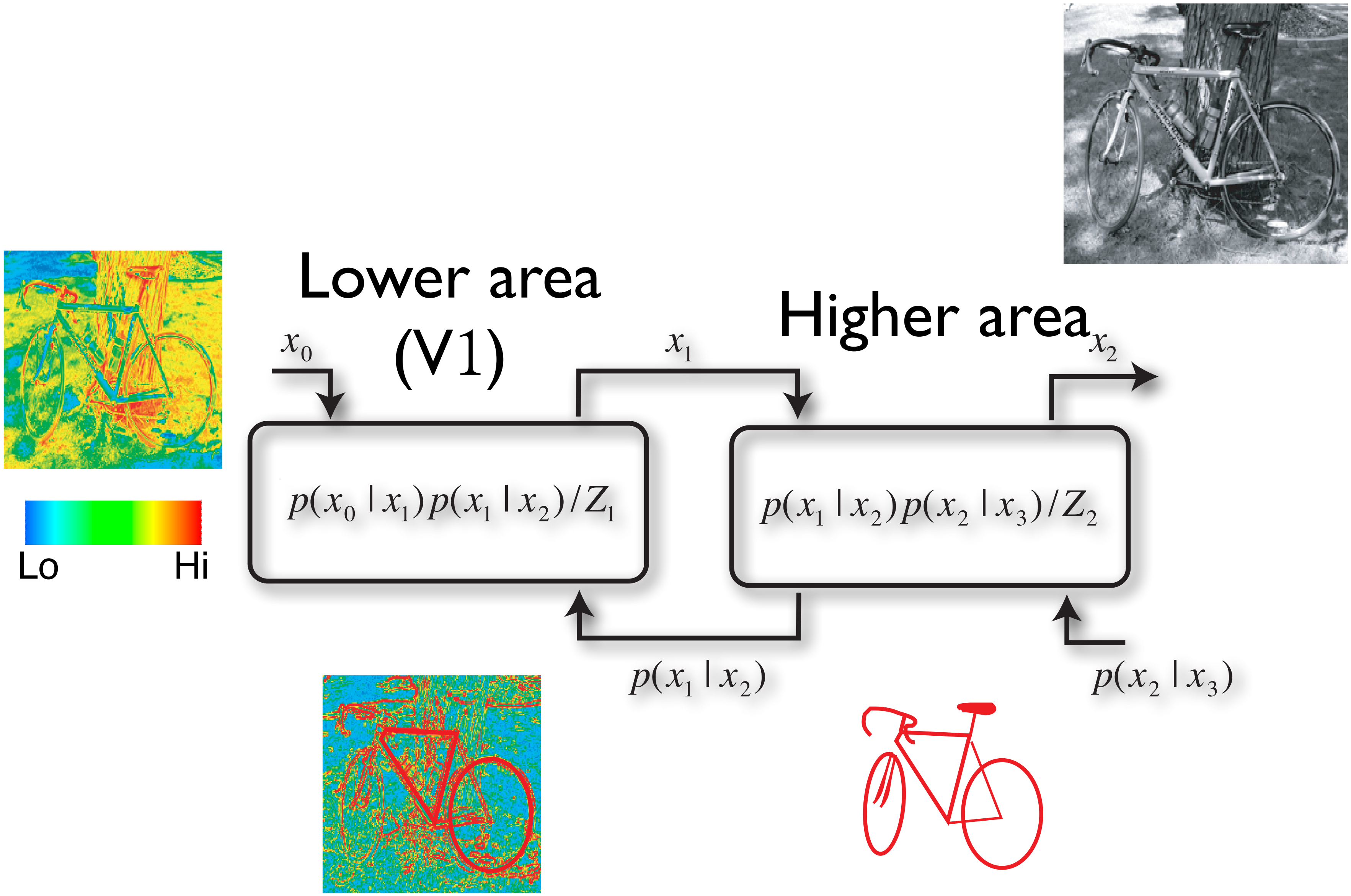
binding through enhancement of consistent features at lower levels



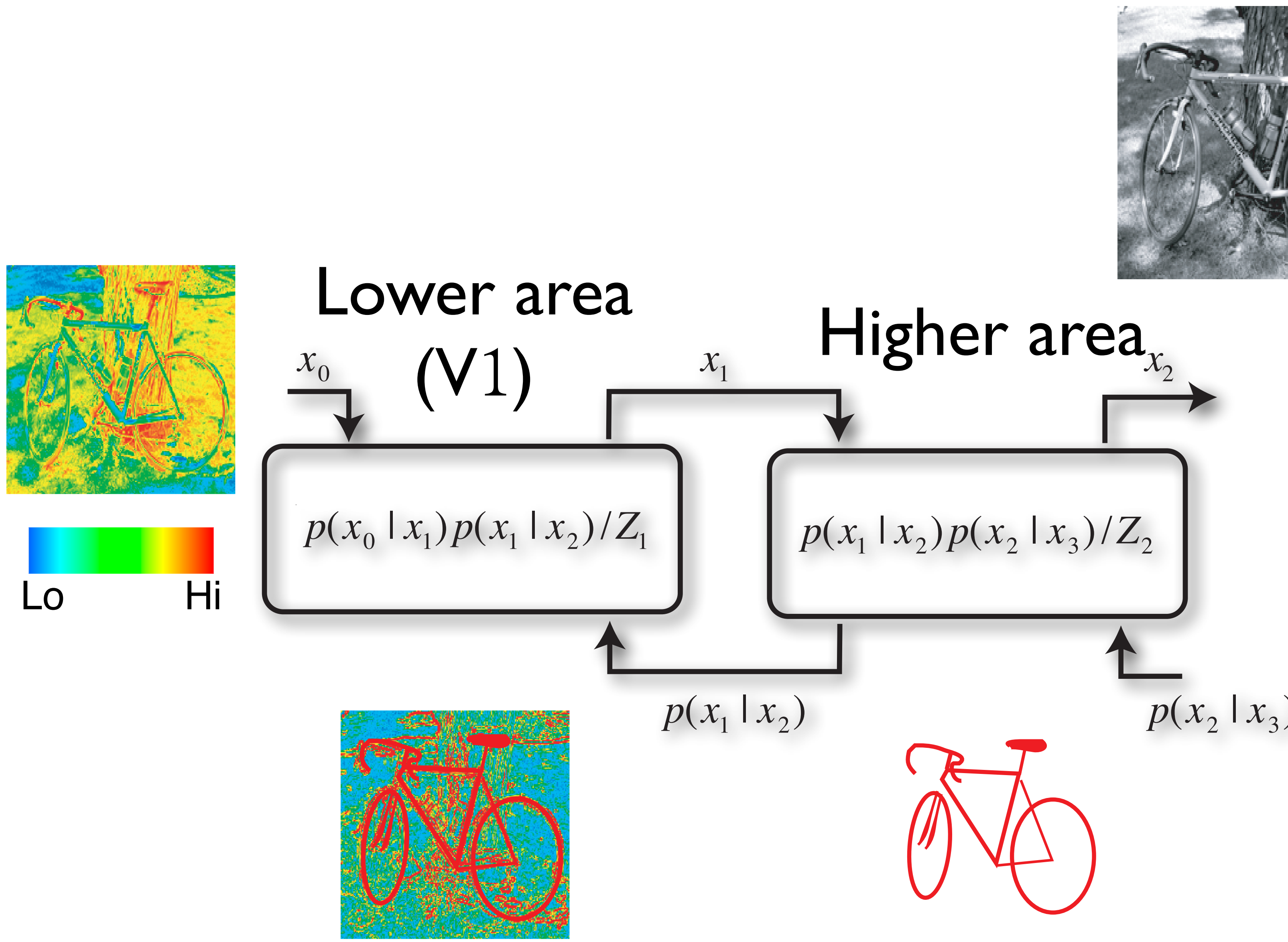
Lee & Mumford, 2003, JOSA



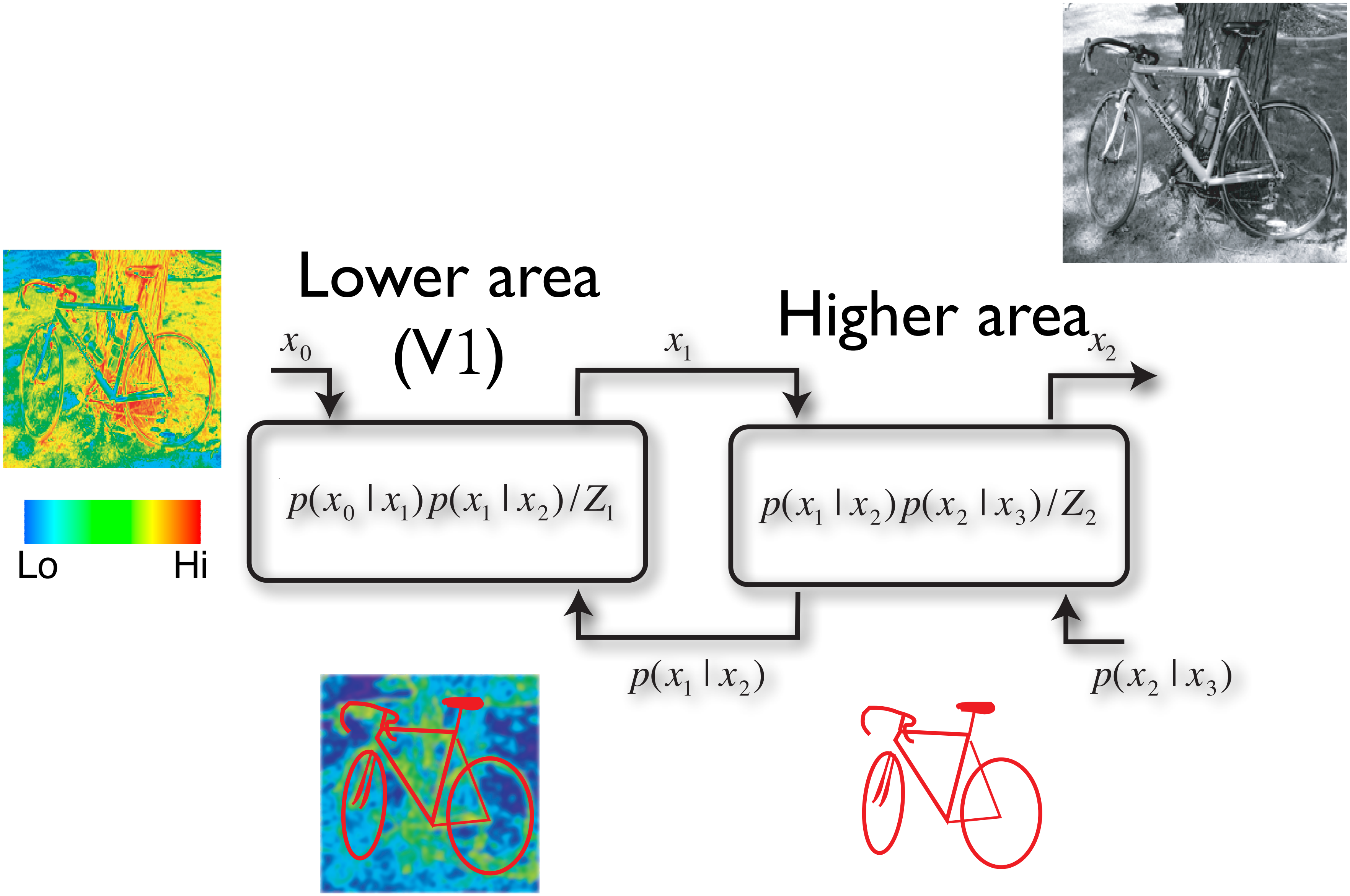
Lower area
(V1)
Higher area
x_0
x_1
x_2
$p(x_0 \mid x_1)p(x_1 \mid x_2)/Z_1$
$p(x_1 \mid x_2)p(x_2 \mid x_3)/Z_2$
$p(x_1 \mid x_2)$
$p(x_2 \mid x_3)$
Lo
Hi

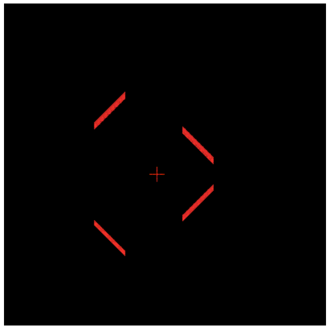


Lower area
(V1)
Higher area
x_0
x_1
x_2
$p(x_0 \mid x_1)p(x_1 \mid x_2)/Z_1$
$p(x_1 \mid x_2)p(x_2 \mid x_3)/Z_2$
$p(x_1 \mid x_2)$
$p(x_2 \mid x_3)$
Lo
Hi



Lower area
(V1)
Higher area
x_0
x_1
x_2
$p(x_0 \mid x_1)p(x_1 \mid x_2)/Z_1$
$p(x_1 \mid x_2)p(x_2 \mid x_3)/Z_2$
$p(x_1 \mid x_2)$
$p(x_2 \mid x_3)$
Lo
Hi


Lower area
(V1)
Higher area
x_0
x_1
x_2
$p(x_0 \mid x_1) p(x_1 \mid x_2) / Z_1$
$p(x_1 \mid x_2) p(x_2 \mid x_3) / Z_2$
$p(x_1 \mid x_2)$
$p(x_2 \mid x_3)$
Lo
Hi

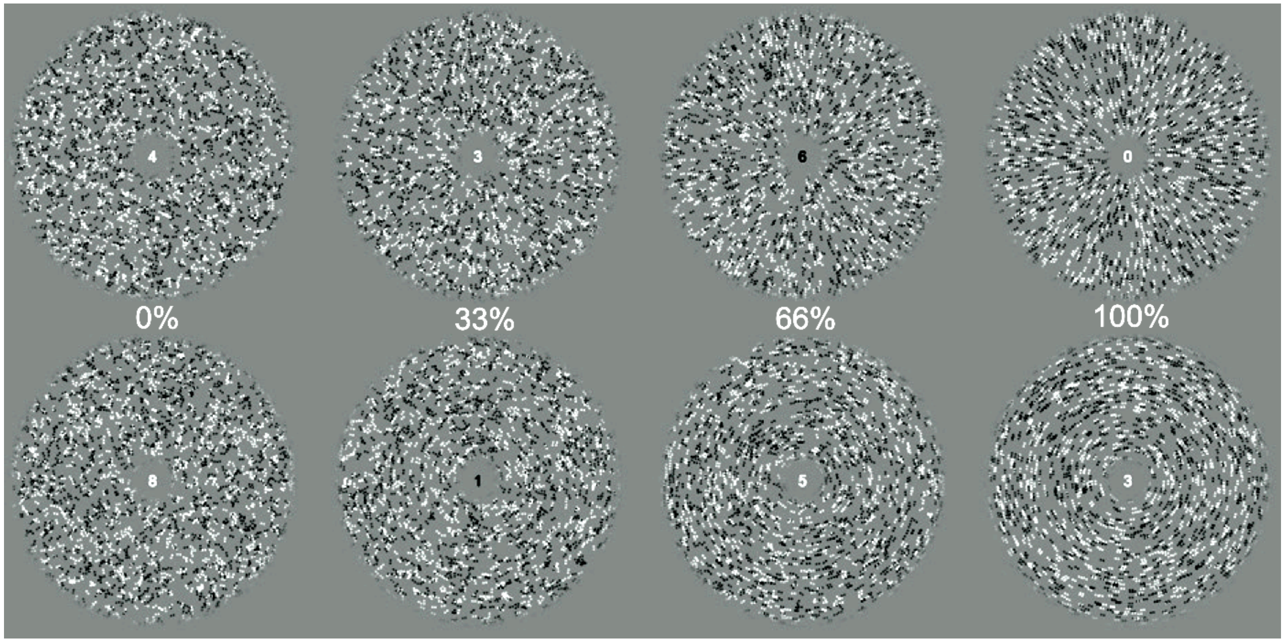


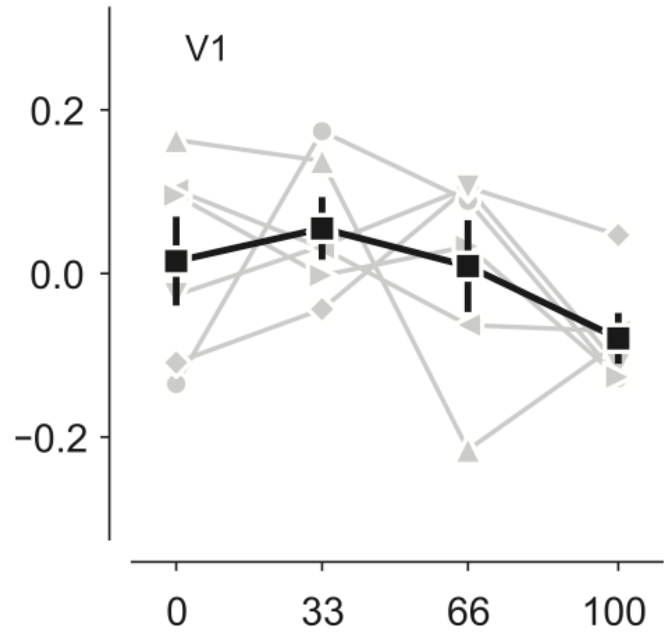
global suppression?

Initially, we interpreted the diamond results as local suppression of consistent features. …but both our subsequent measurements and those of others showed suppression of voxels not in retinotopic correspondence with the line segments during perceptual coherence (e.g. Wit et al., 2012). The suppression was not necessarily localized to the features being "explained" by high-level hypotheses.

Further, we now know that "the degree of perceptual organization" doesn't always measurably modulate V1 MR activity even at a coarse global level.

For example…





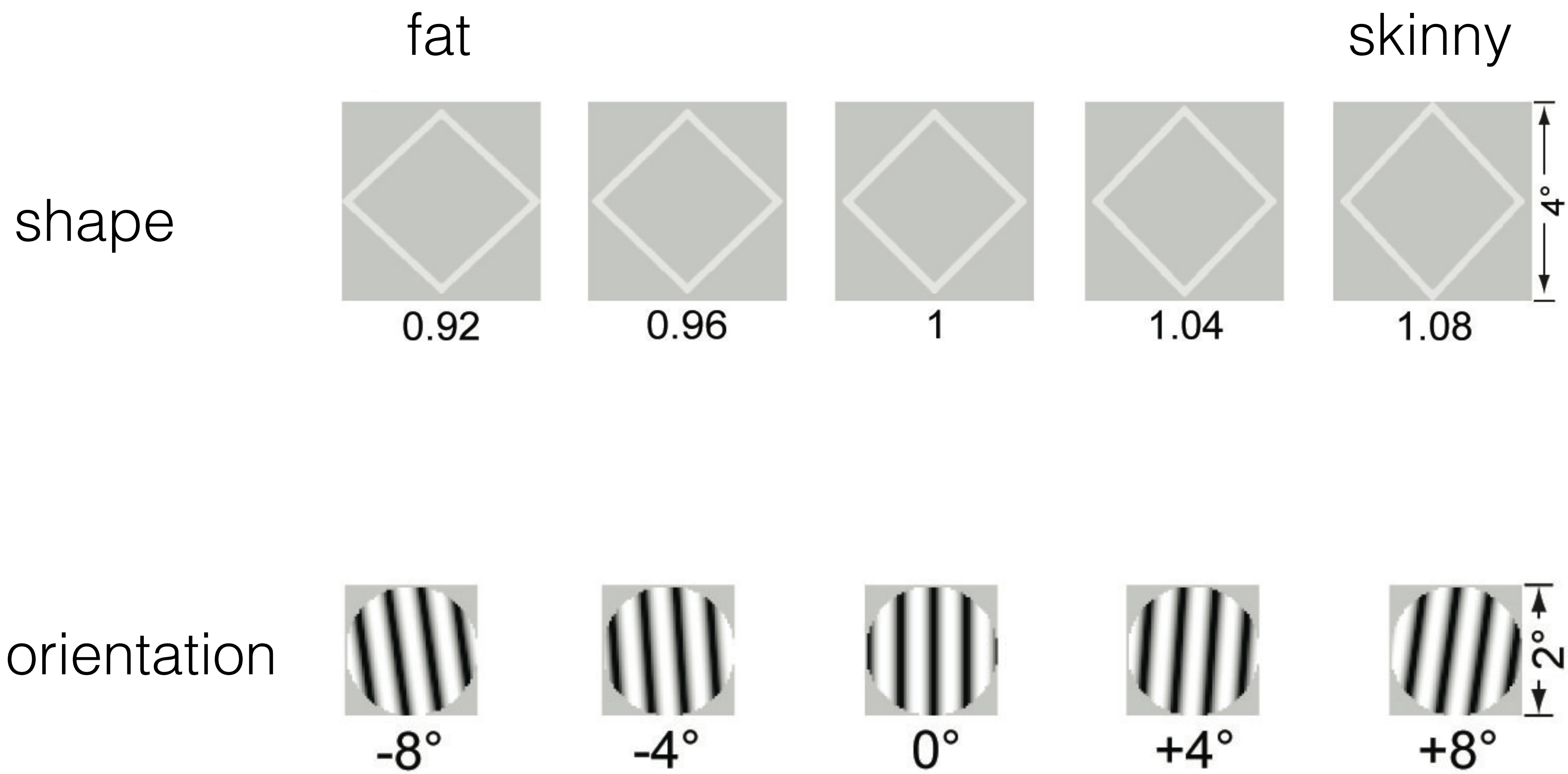
Mannion, Kersten, D. J., & Olman, C. A. (2013). Consequences of polar form coherence for fMRI responses in human visual cortex. NeuroImage.

Why no significant modulation of V1?

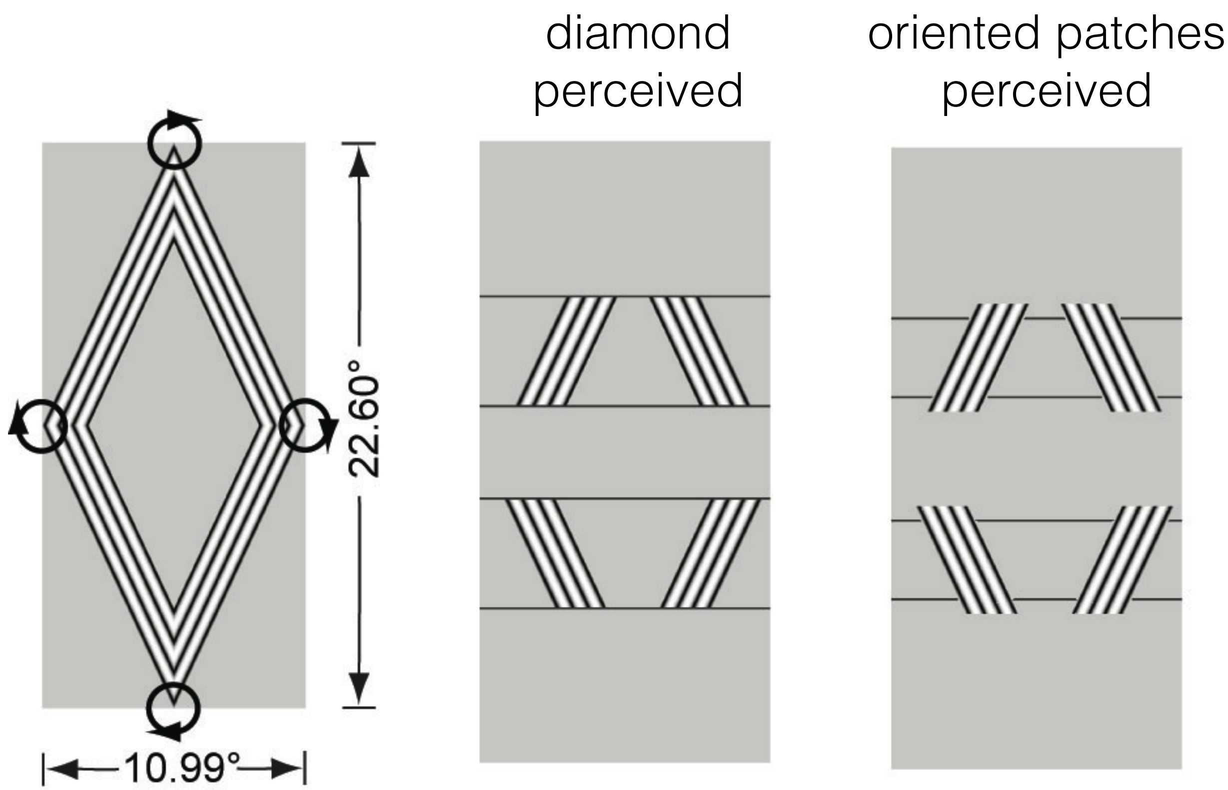
Distracting attentional task

Area V3, dorsal and ventral mid-level areas, and the human MT complex do show increased response with increasing coherence. (No effect on LOC.)

Shape vs. orientation adaptation

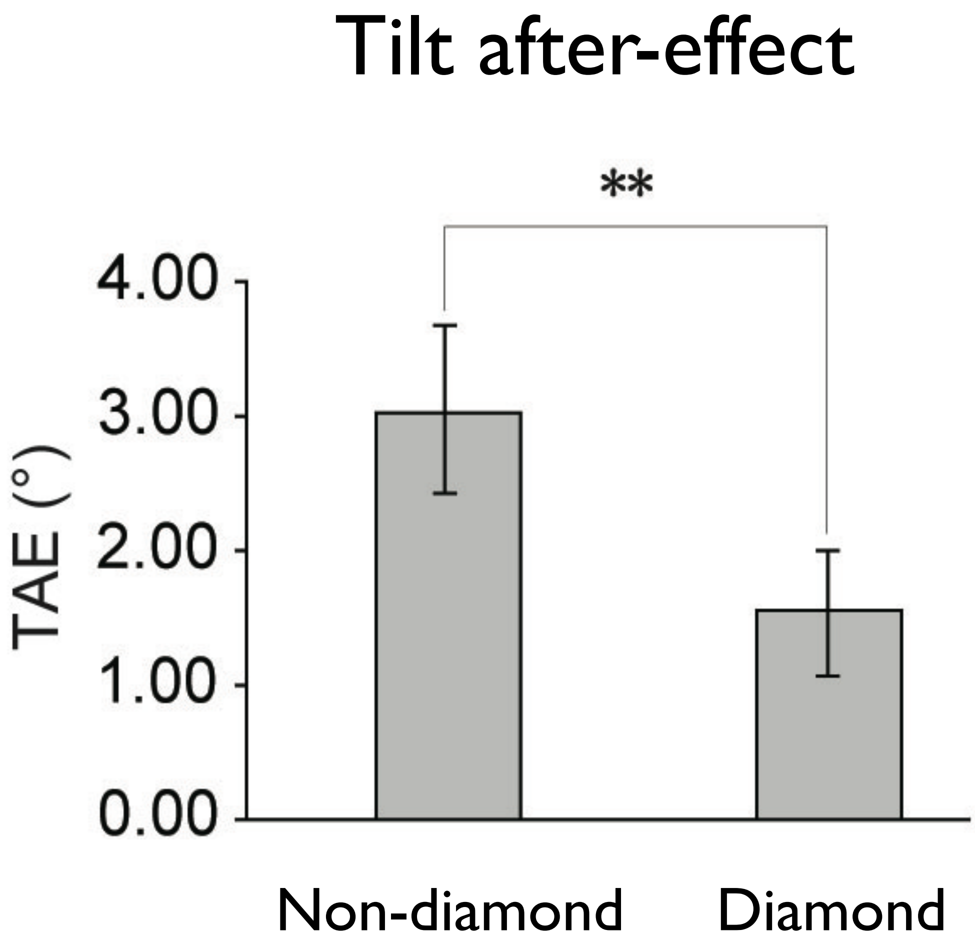


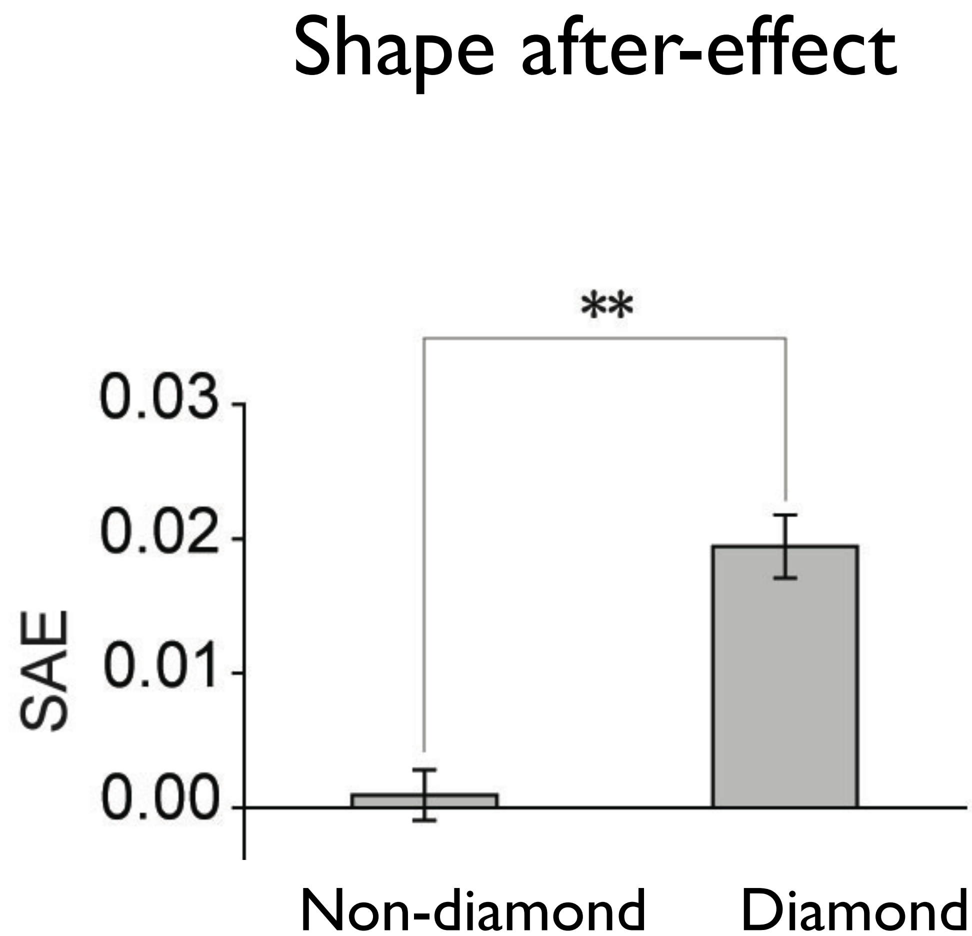
We found opposite modulation of high- and low-level visual aftereffects as a consequence of perceptual grouping



Perceptual grouping ("diamond percept") reduces the strength of adaptation to local tilt, while amplifying the effect of adaptation to a whole shape, consistent with localized lower-level, feature-specific modulation.

He, D., Kersten, D., & Fang, F. (2012). Opposite modulation of high- and low-level visual aftereffects by perceptual grouping. Current Biology, 22(11), 1040–1045.





Perceptual grouping ("Diamond") reduces the strength of adaptation to local tilt, while amplifying the effect of adaptation to a whole shape

Consistent with predictive coding interpretation

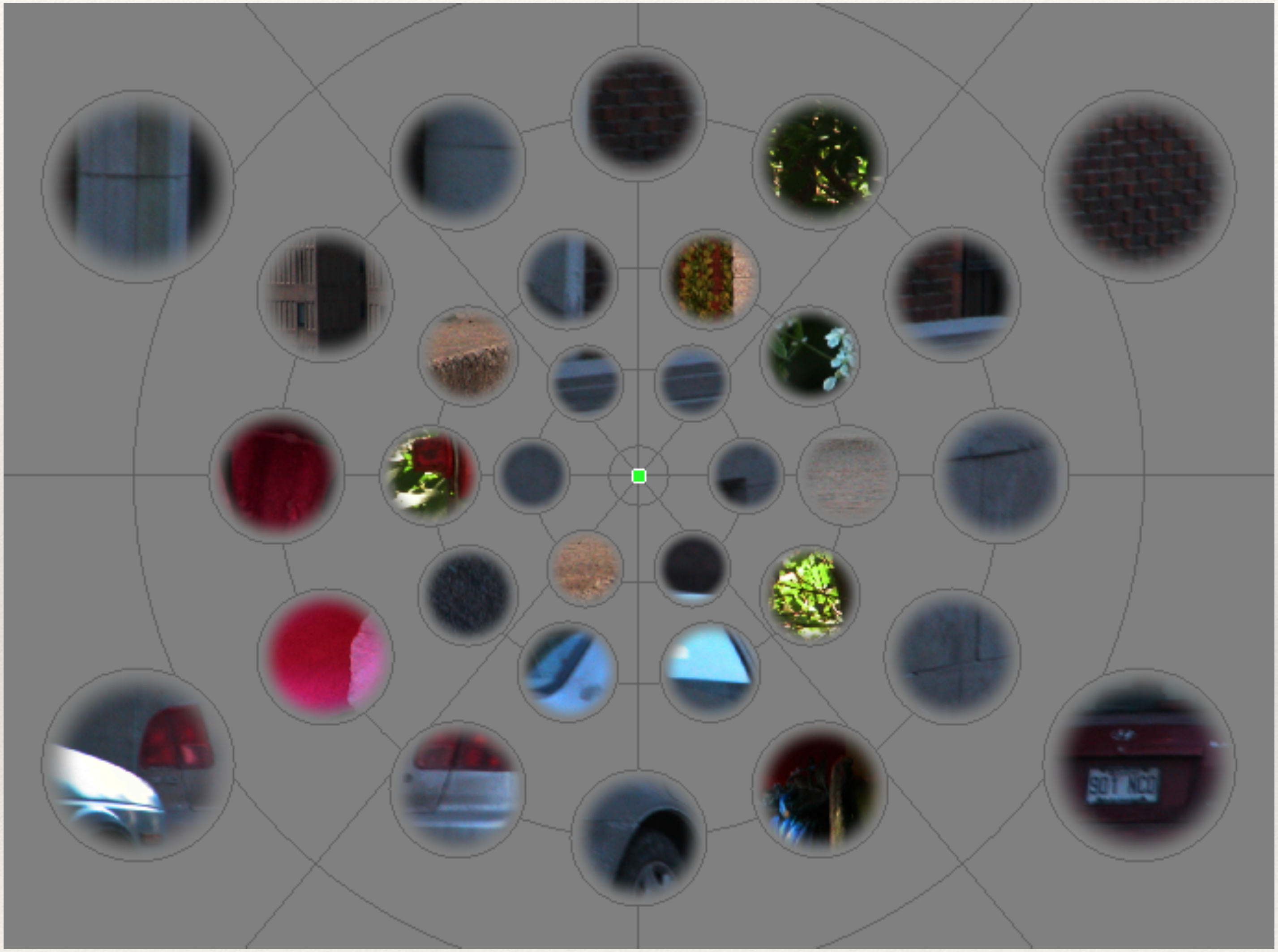
…but more studies are needed

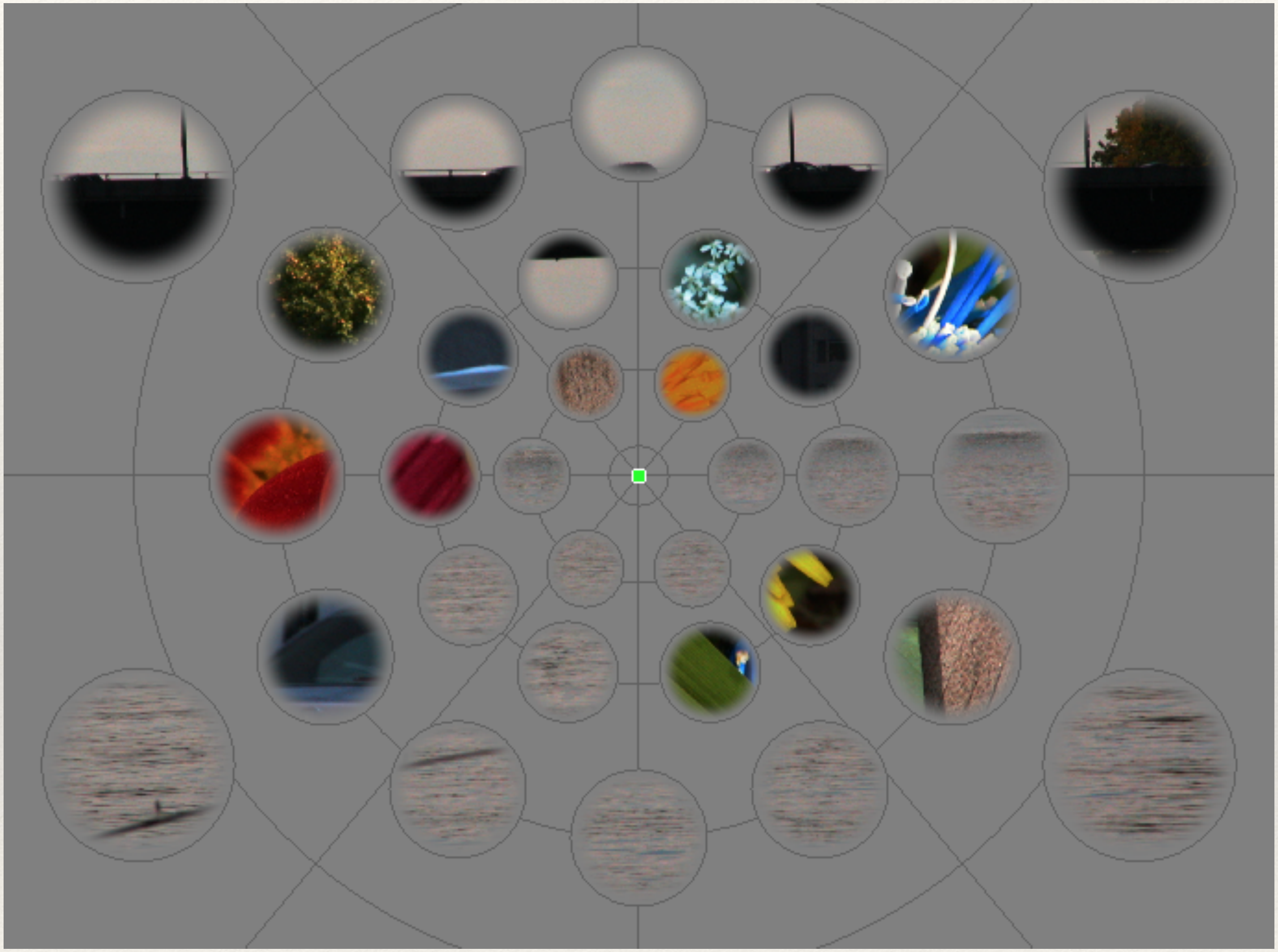
Local enhancement?

Perceptual organization of scenes

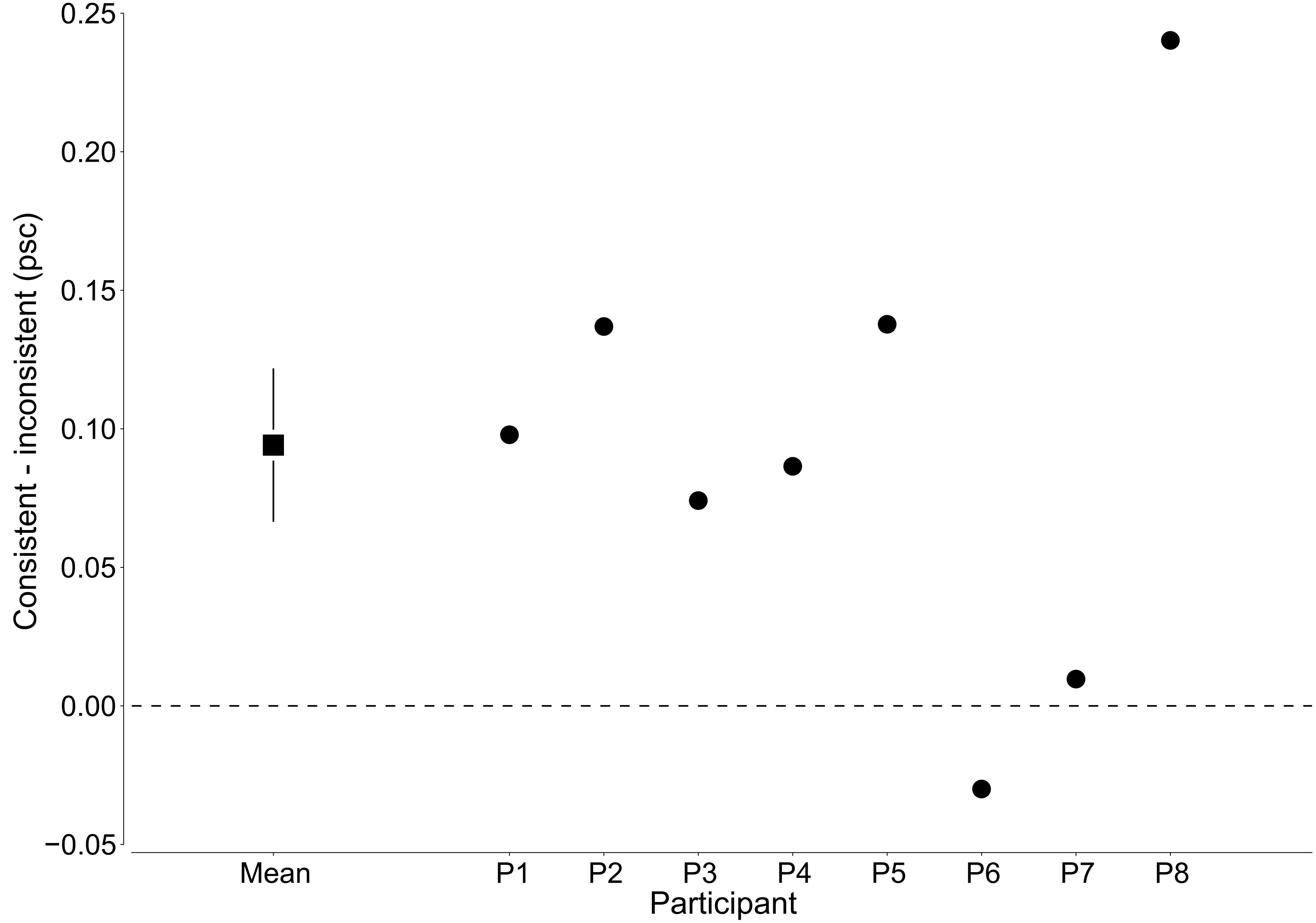
Damien Mannion, Daniel Kersten & Cheryl Olman

Damien Mannion, Daniel Kersten & Cheryl Olman





early visual area



Damien Mannion, Daniel Kersten & Cheryl Olman

Feedback

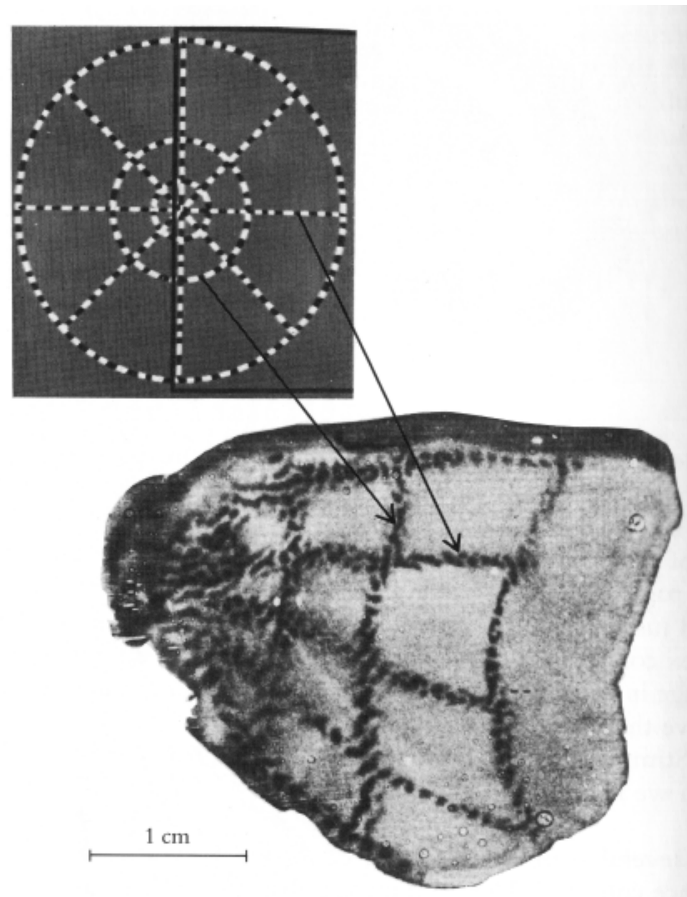
The executive metaphor

- Attention, hierarchically organized expertise.
    - E.g. V1: Feature-specific tasks, Huk & Heeger, 2000; Working memory (Harrison & Tong, 2009); Perceptual learning (Hochstein & Ahissar,2002); Foveal V1 as a high-resolution spatial buffer (Lee et al. 1998,; Williams et al., 2008); Task-dependent changes in early receptive fields (McManus et al., 2011);
- Use of built-in generative knowledge?
    - The “perceived size and V1” puzzle

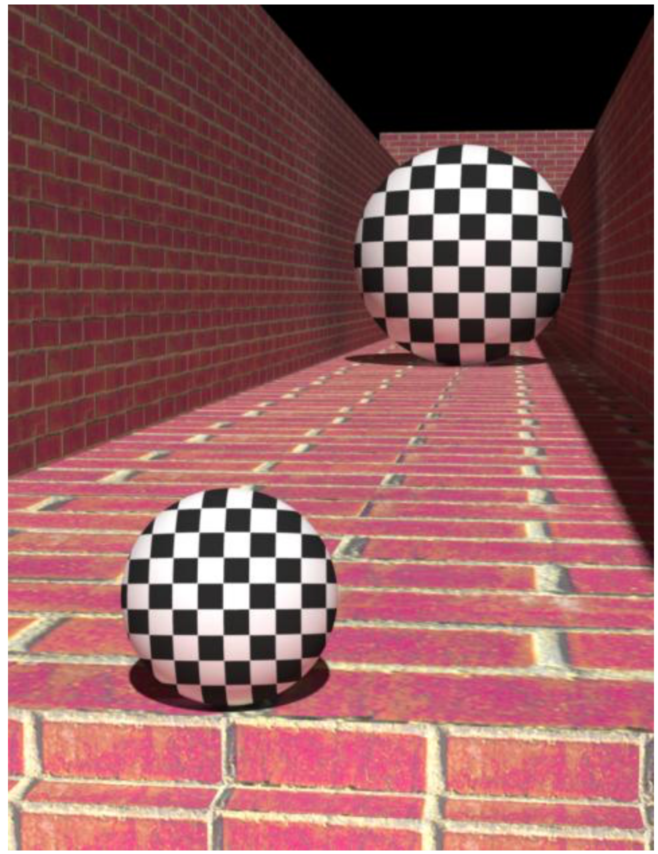
recall global organization of V1

global: hypercolumns arranged retinotopically

neurons receiving information from nearby points in the world are near on cortical surface

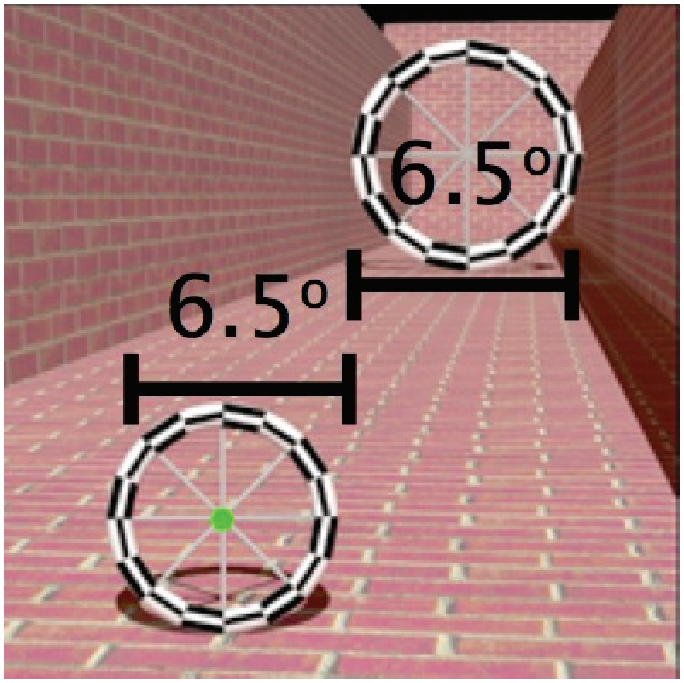


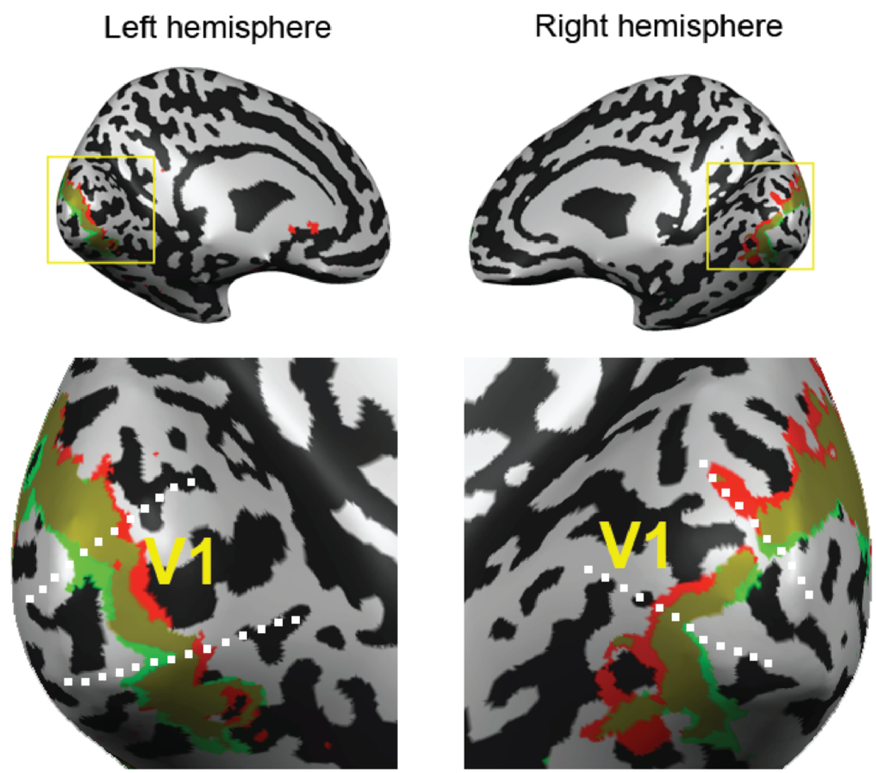
Feedback: Executive metaphor?



Fang, Boyaci, Kersten, & Murray, S. O. (2008). Attention-dependent representation of a size illusion in human V1. Current Biology

Feedback: Executive metaphor?





Fang, Boyaci, Kersten, & Murray, S. O. (2008). Attention-dependent representation of a size illusion in human V1. Current Biology