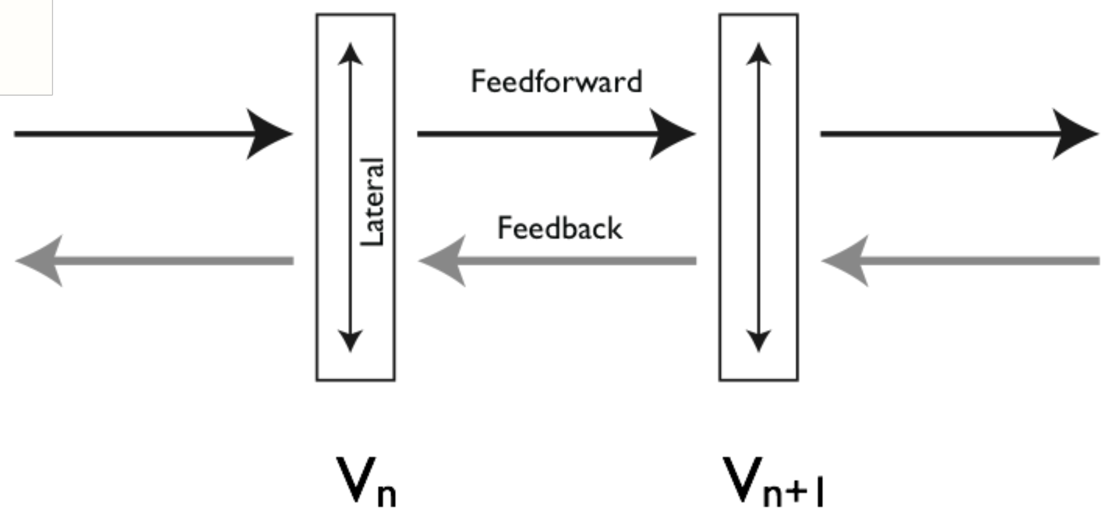
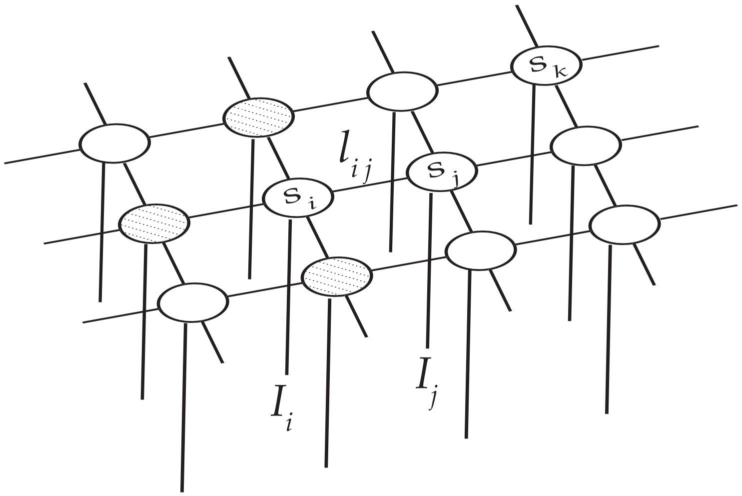
Bidirectional processing:
feedforward & feedback pathways

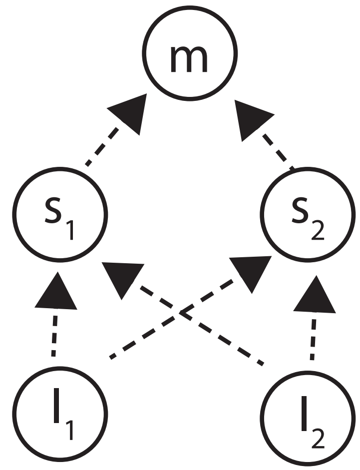
Lecture 27:
Introduction to Neural Networks
Dan Kersten
U of Minnesota

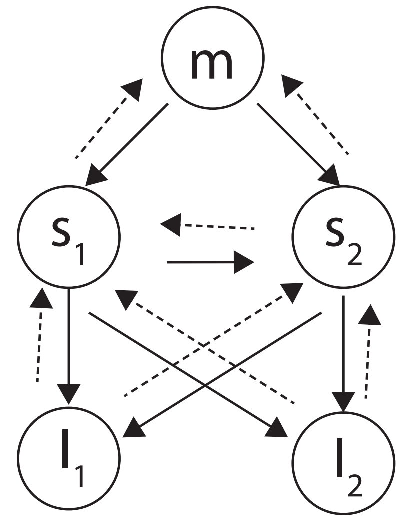


Kersten, D. J., & Yuille, A. L. (2014). Inferential Models of the Visual Cortical Hierarchy. In M. S. Gazzaniga & G. R. Mangun (Eds.), *The New Cognitive Neurosciences, 5th Edition* (pp. 1–22). MIT Press.

graphical models







lateral organization,

lateral processing, reciprocal interactions between feature of similar type

prepare to feedforward

indexing?

hierarchical organization:

feedforward processing

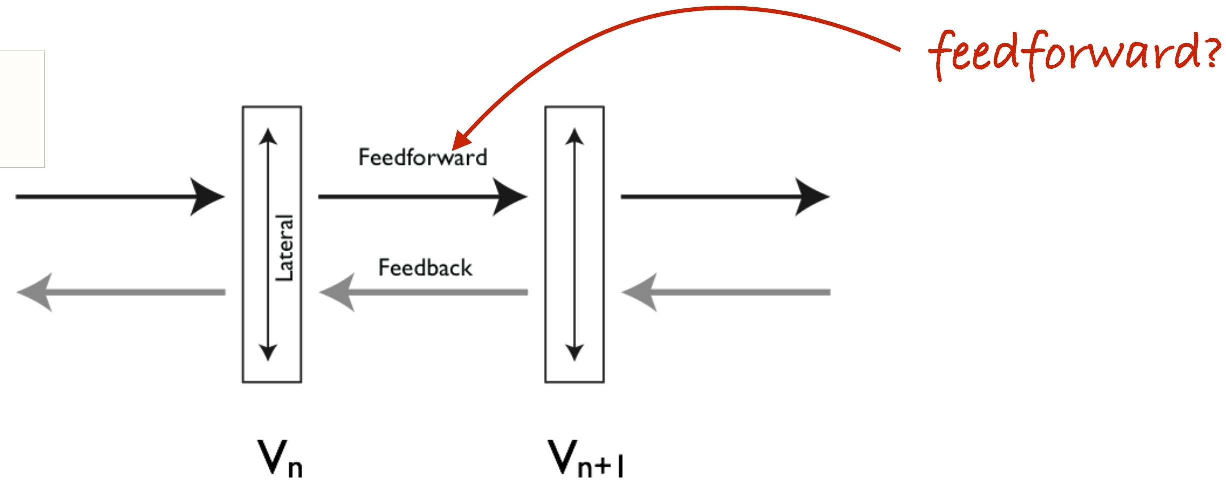
speed

hierarchical organization:

feedforward, feedback & lateral processing

task flexibility & robustness

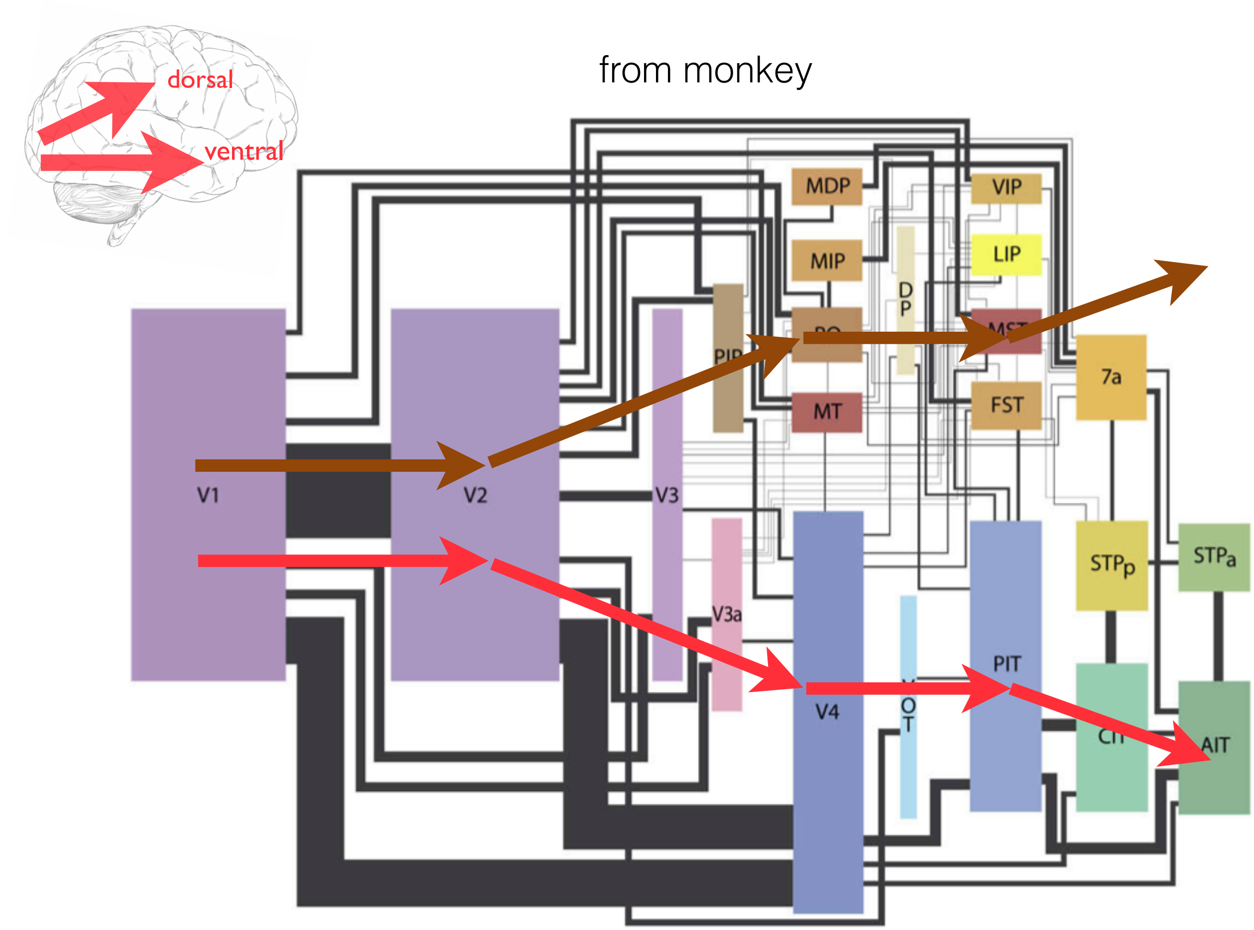
Kersten, D. J., & Yuille, A. L. (2013). Vision: Bayesian Inference and Beyond. In J. S. Werner & L. M. Chalupa (Eds.), *The New Visual Neurosciences* (pp. 1–16). MIT Press.



Lateral organization

- representation and linking of features at a similar level of abstraction
- self-organizing topographical maps
- efficient image coding to explain receptive field properties
- machine learning methods for grouping

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?

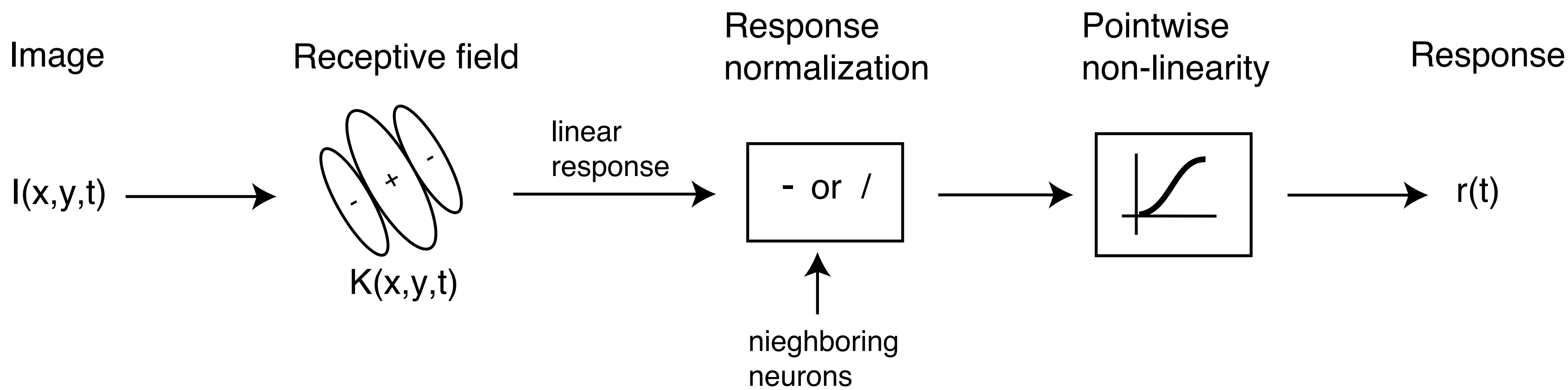
image information required for different basic tasks

...but lots of tasks

Wallisch, P., & Movshon, J. A. (2008). Structure and Function Come Unglued in the Visual Cortex. *Neuron*, *60*(2), 194–197.

“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

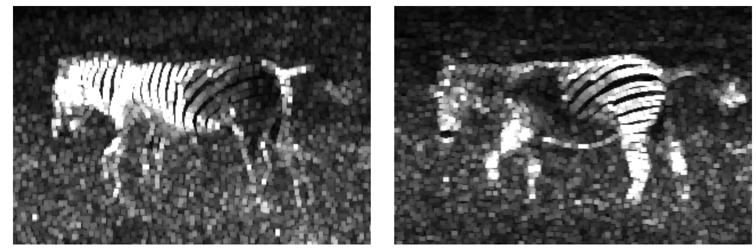
canonical computation?

V1

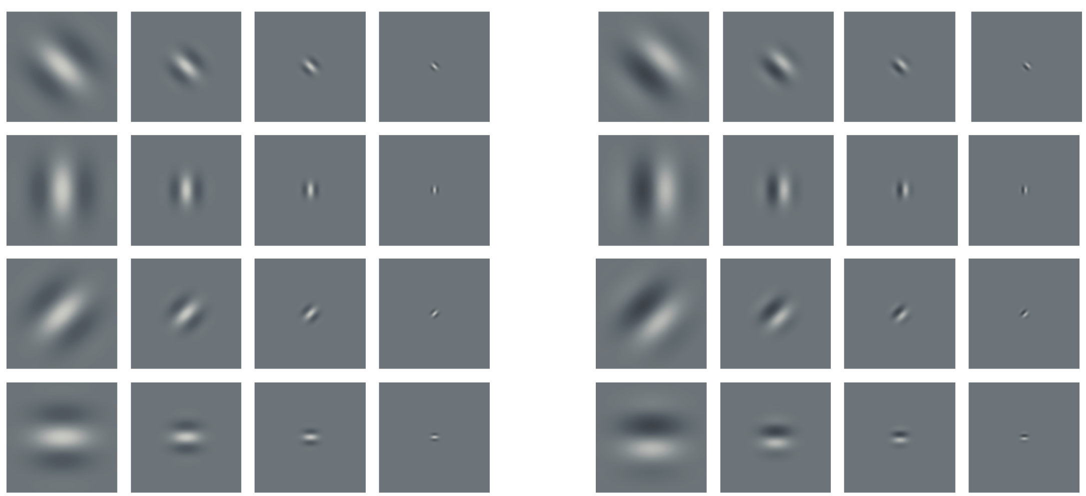
spatial frequency channels — 1968

just 2 in homework #6 

	Input	encoded tensor (size: 1×214×320)
1	ConvolutionLayer	tensor (size: 2×212×318)
2	ElementwiseLayer	tensor (size: 2×212×318)
3	PoolingLayer	tensor (size: 2×209×315)
4	ElementwiseLayer	tensor (size: 2×209×315)
	Output	decoded tensor (size: 2×209×315)



classic models, 16 or more



...but are there more types in V1?

How do tasks constrain feature hierarchies?

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

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

— 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

— 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

Hierarchical models for feature extraction for recognition

Local features progressively grouped into more structured representations

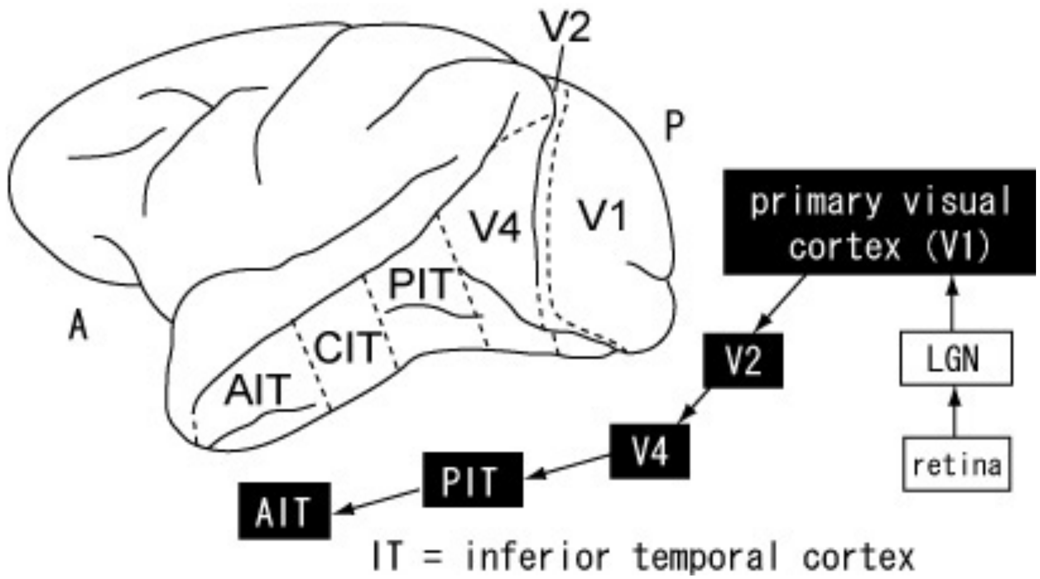
- edges => contours => fragments => parts => objects

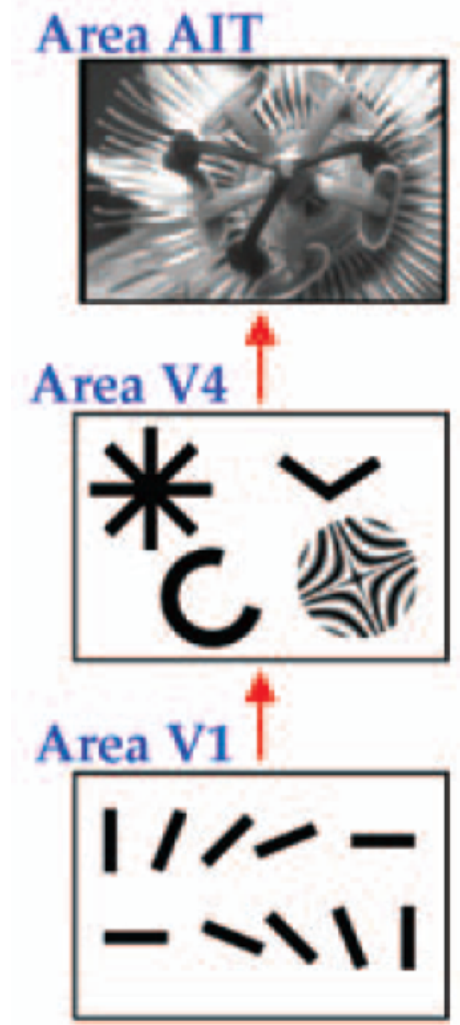
Selectivity/invariance trade-off

- Increased selectivity for object/pattern type

- Decreased sensitivity to view-dependent variations of translation, scale and illumination

Hierarchical models of object recognition





bread and butter of ventral stream modeling

Hegde and Felleman, 2007

Recall simple & complex cells in V1

Simple cells

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

Complex cells

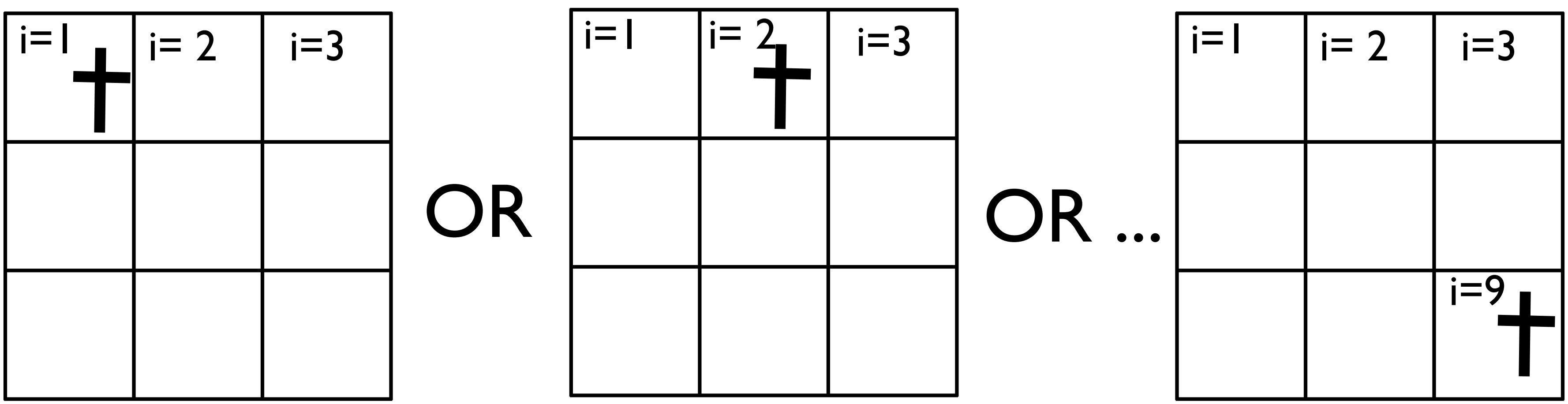
- insensitivity to small changes in position, detect disjunctions, logical “OR”

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

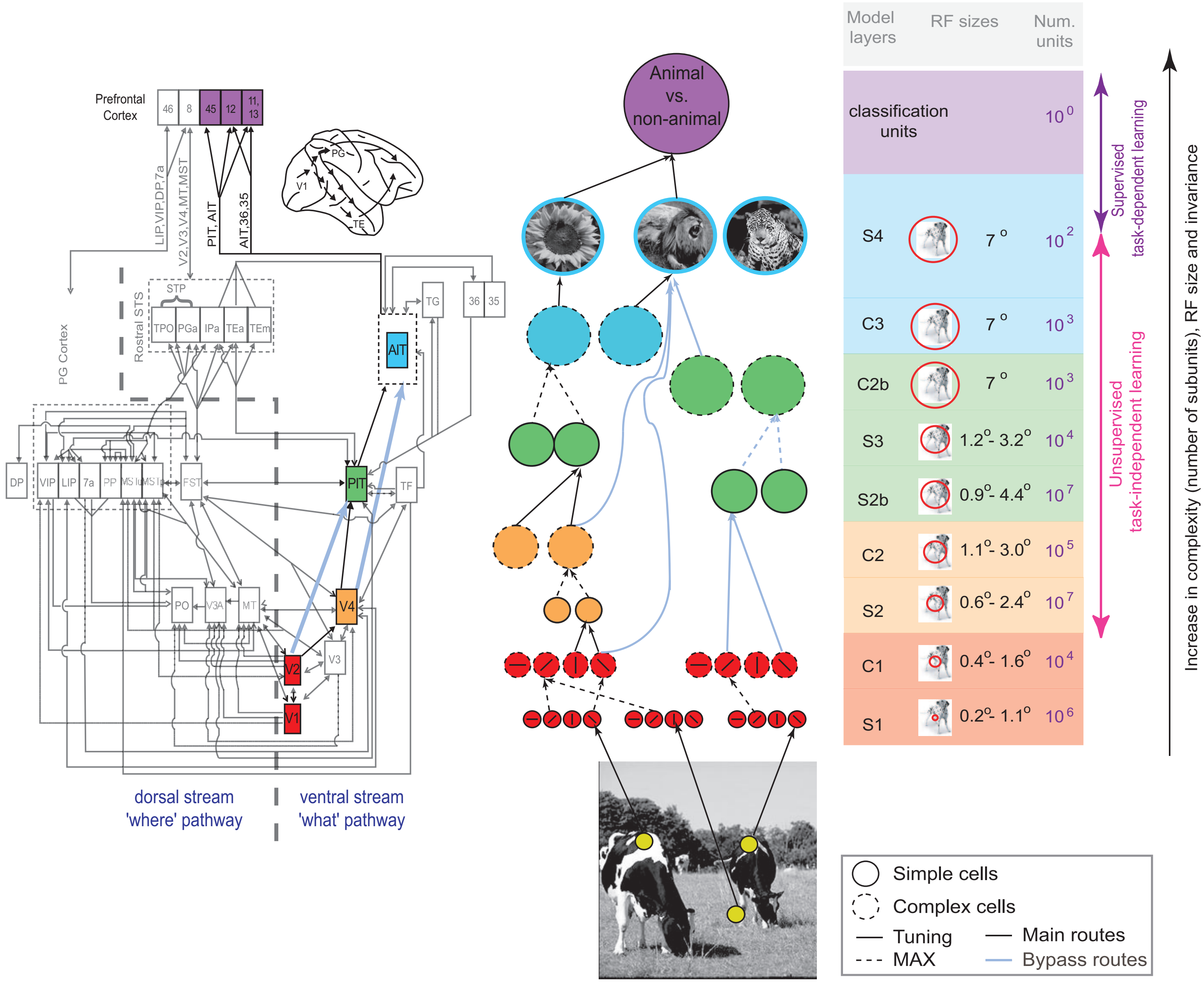
"t": h_1 && v_1 || h_2 && v_2 || h_3 && v_3...

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

How do tasks constrain feature hierarchies?

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

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

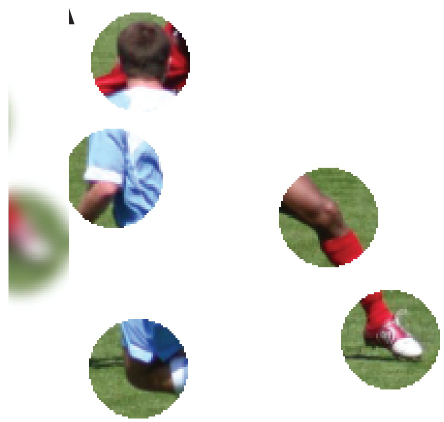
— supervised learning

- deep convolutional networks
- — “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

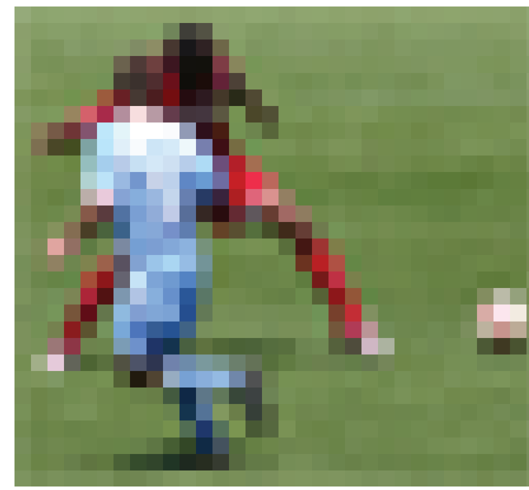
— unsupervised learning based on based on successive discovery of image regularities (Barlow)

- 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

Higher-level features?

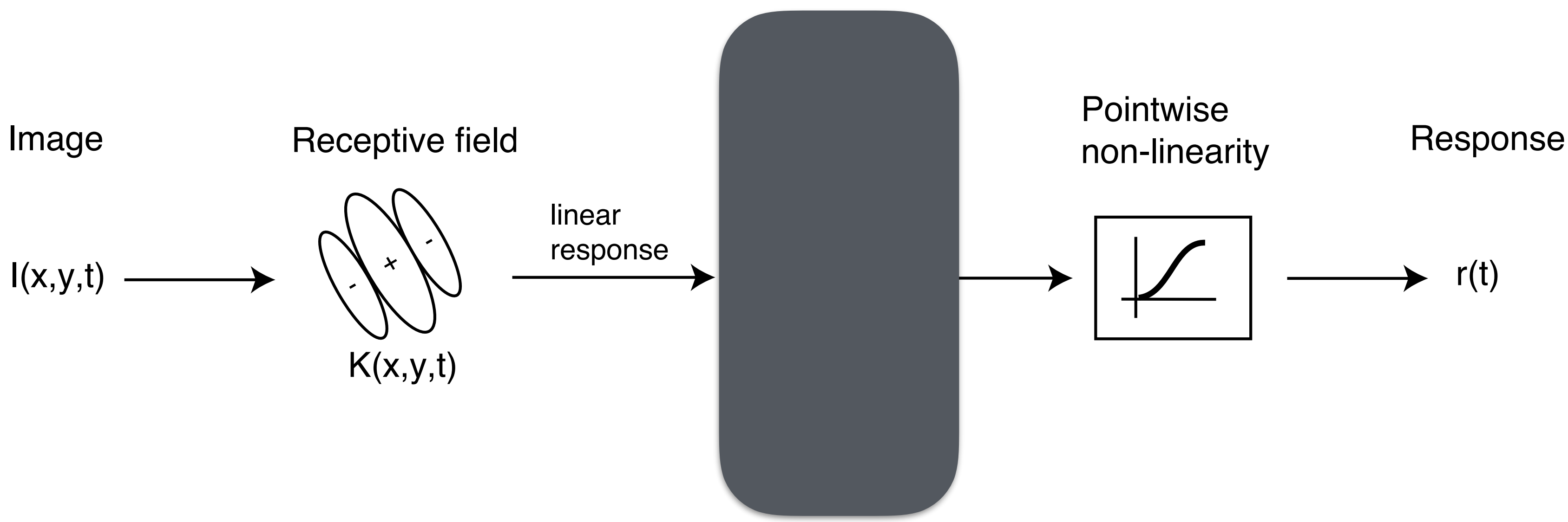


B



Higher-level features?
(machine) learn them?

	Input	encoded tensor (size: 3 × 51 × 51)
1	ConvolutionLayer	tensor (size: 40 × 47 × 47)
2	ElementwiseLayer	tensor (size: 40 × 47 × 47)
3	PoolingLayer	tensor (size: 40 × 24 × 24)
4	ConvolutionLayer	tensor (size: 20 × 20 × 20)
5	ElementwiseLayer	tensor (size: 20 × 20 × 20)
6	PoolingLayer	tensor (size: 20 × 10 × 10)
7	FlattenLayer	vector (size: 2000)
8	DotPlusLayer	vector (size: 500)
9	ElementwiseLayer	vector (size: 500)
10	DotPlusLayer	vector (size: 3)
11	SoftmaxLayer	vector (size: 3)
	Output	decoded vector (size: 3)



...one problem, not clear how the features learned generalize across tasks

How do tasks constrain feature hierarchies?

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

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

— supervised learning

- deep convolutional networks
- — “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

— 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

How do tasks constrain 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

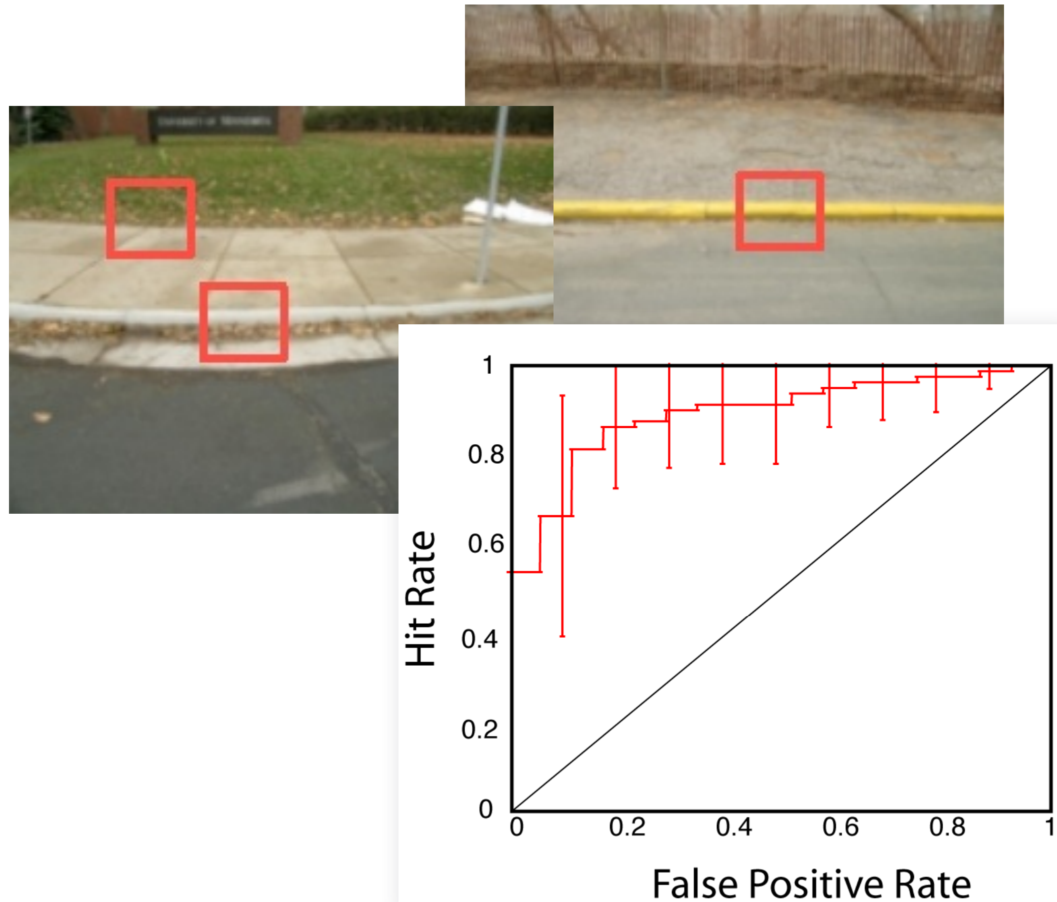
Algorithm finds fragments that maximize mutual information

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

View-specific

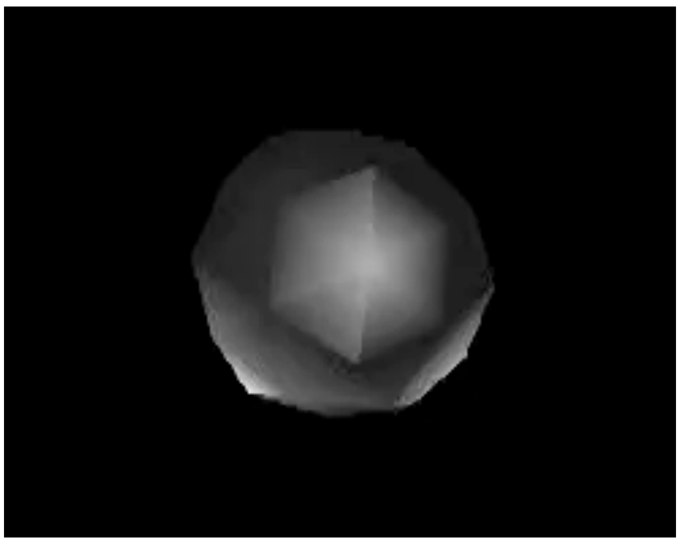
Works well

Experimentally tractable

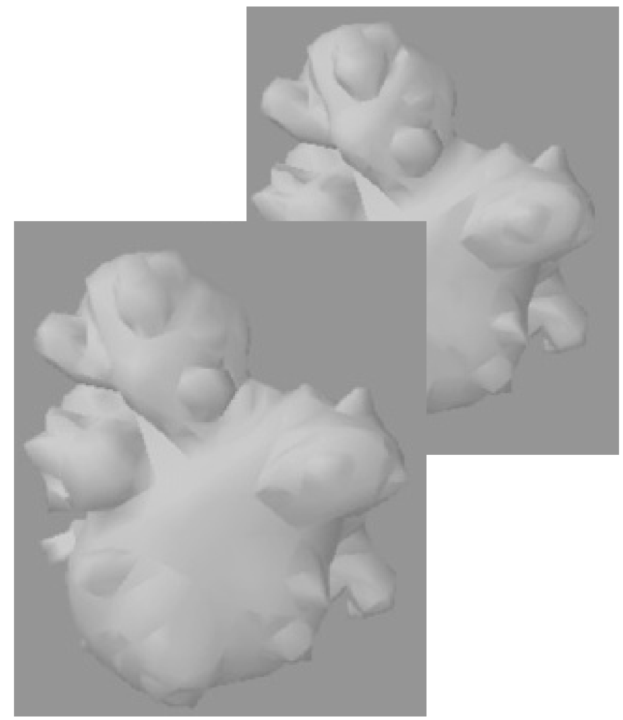


Evgeniy Bart

Do people learn to use fragments of predicted “intermediate complexity”



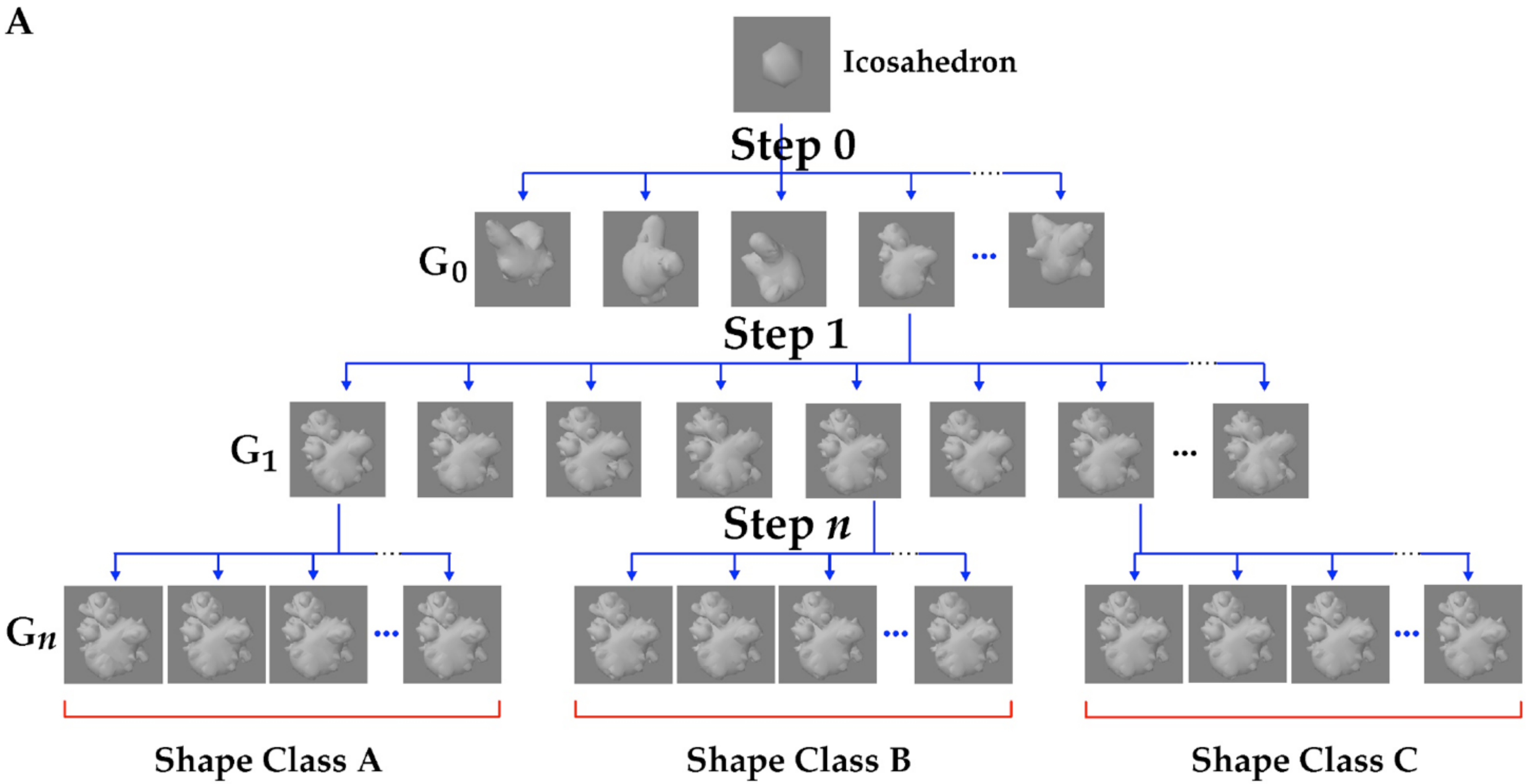
Virtual morphogenesis



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

Generating naturalistic object classes

Virtual Phylogenesis



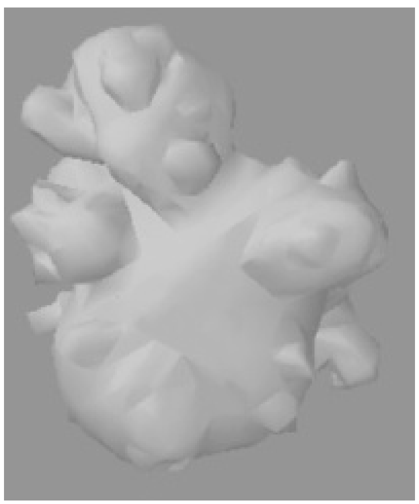
Hegde, J., Bart, E., & Kersten, D. (2008). Fragment-Based Learning of Visual Object Categories. Curr Biol. 18, 597-601

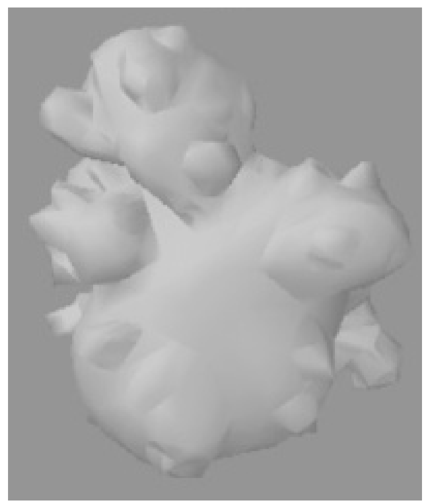
Training

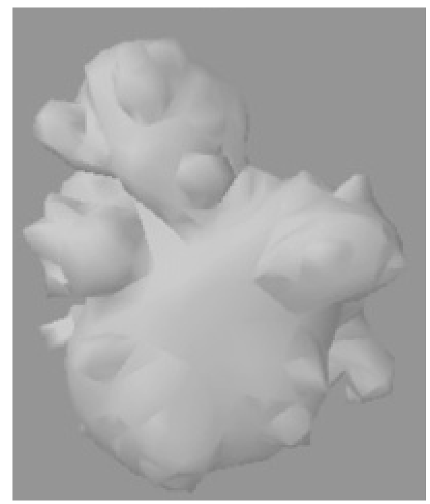
Member of category A or B?

A

B

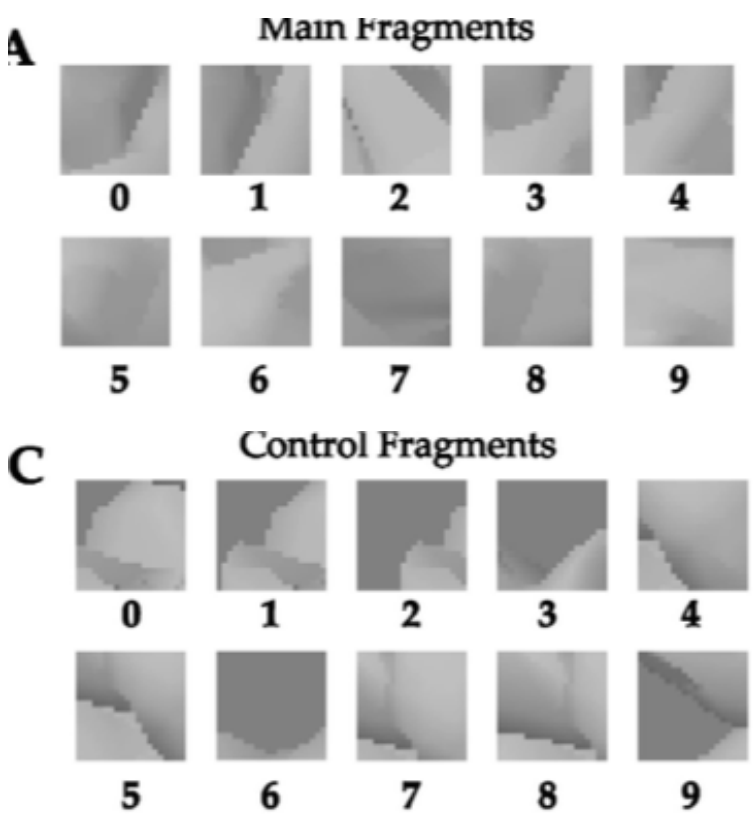


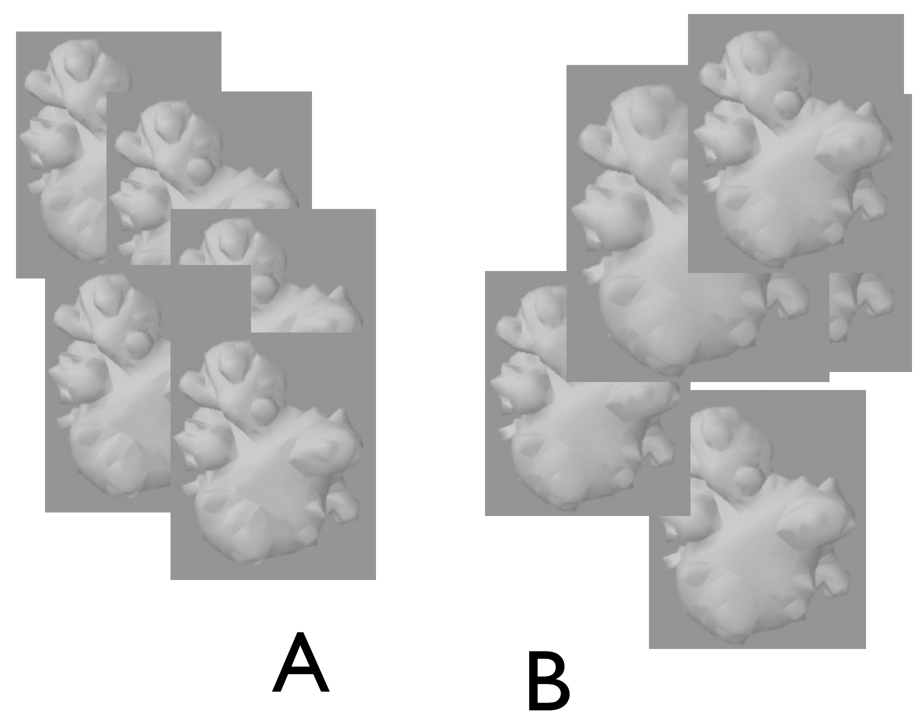




Results

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

How do tasks constrain feature hierarchies?

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

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

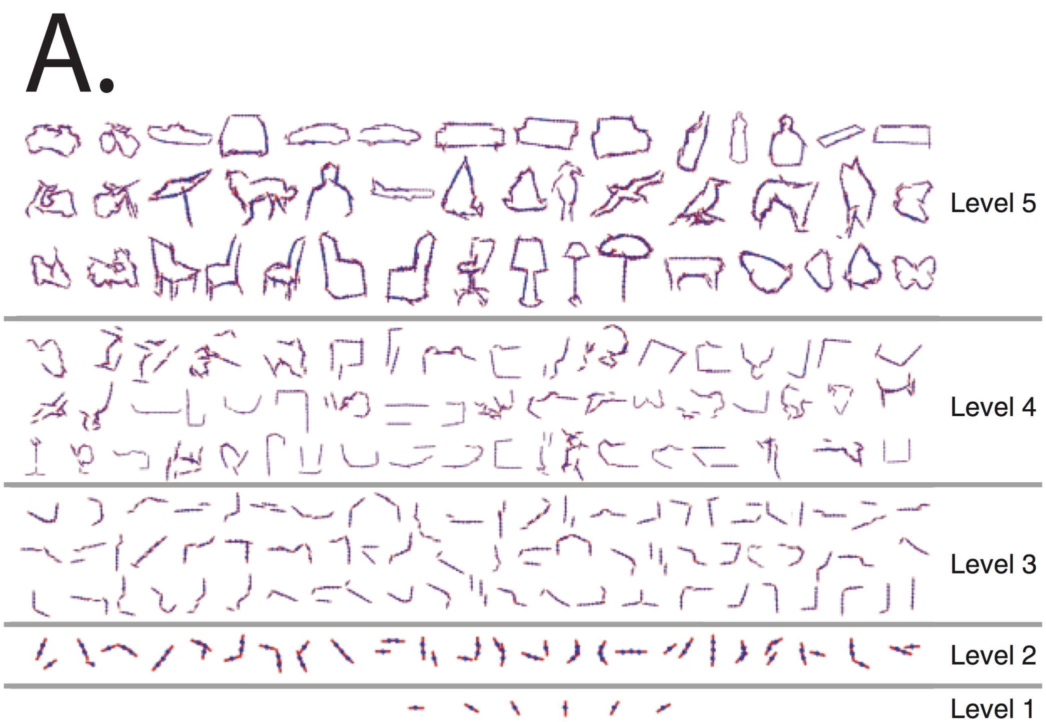
— supervised learning

- deep convolutional networks
- — “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

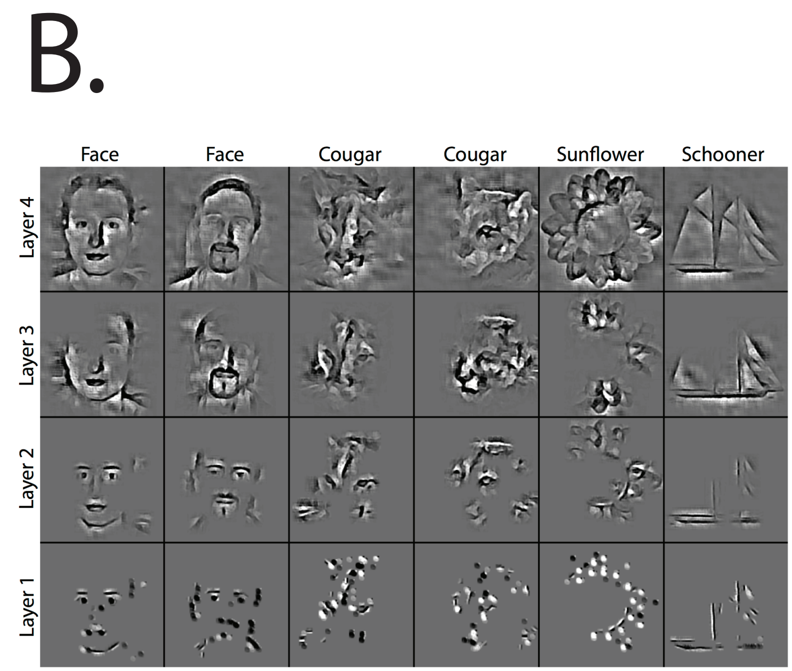
— unsupervised learning based on based on successive discovery of image regularities (Barlow)

- 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

unsupervised learning



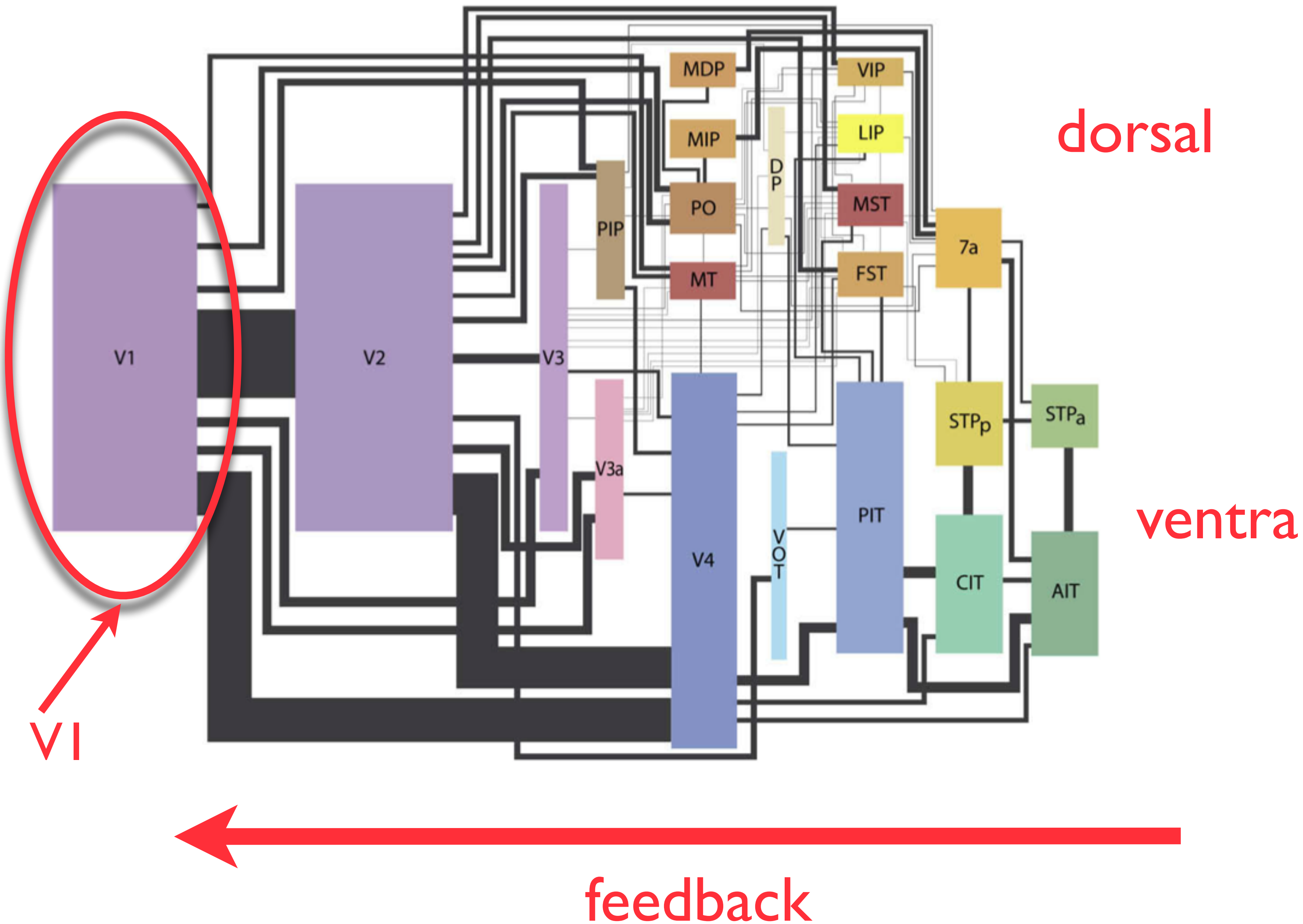
Zhu, L., Chen, Y., Torralba, A., Freeman, W., & Yuille, A. (2010). Part and appearance sharing: Recursive compositional models for multi-view multi-object detection (pp. 1919–1926). Presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.



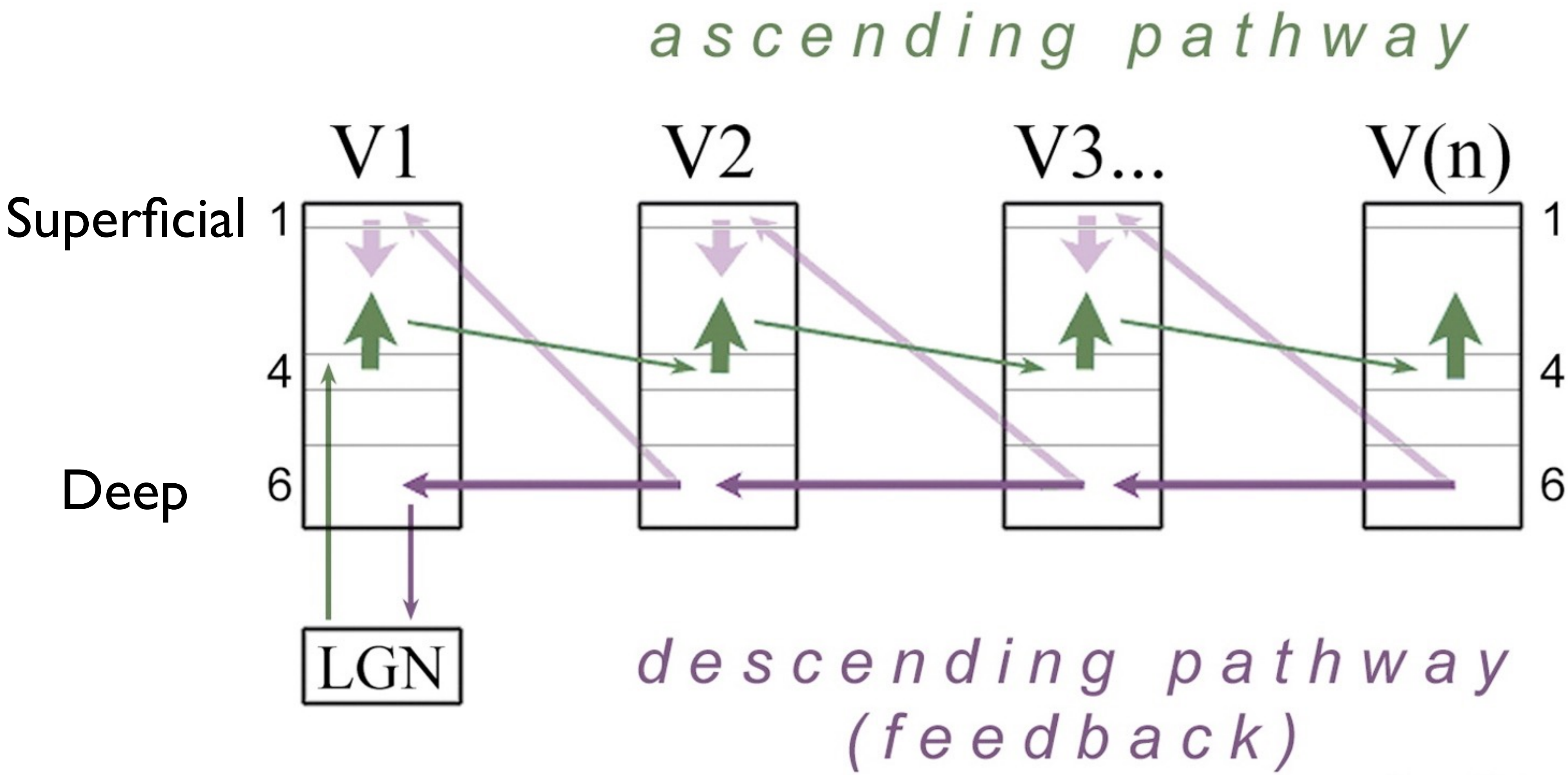
Zeiler, M. D., Taylor, G. W., & Fergus, R. (2011). Adaptive deconvolutional networks for mid and high level feature learning. *Computer Vision (ICCV), 2011 IEEE International Conference on*, 2018–2025.

task general

Feedback



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

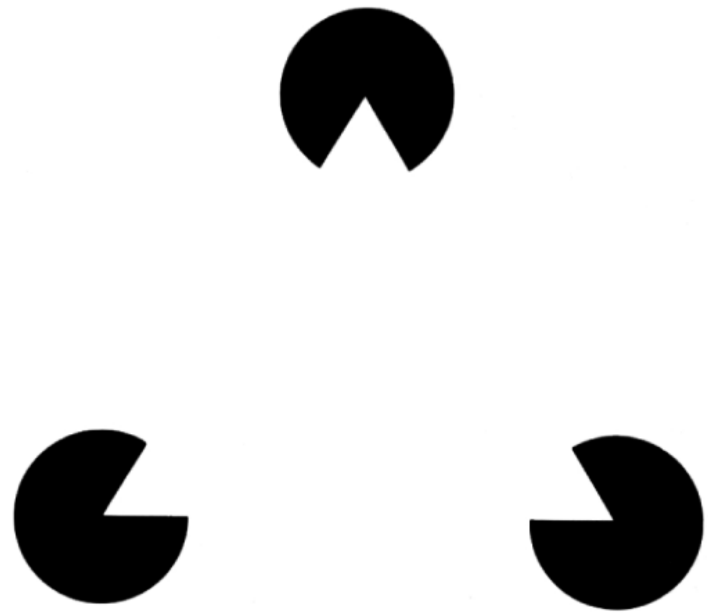
Disambiguation

- suppress explained input
- enhance explained input

The executive metaphor

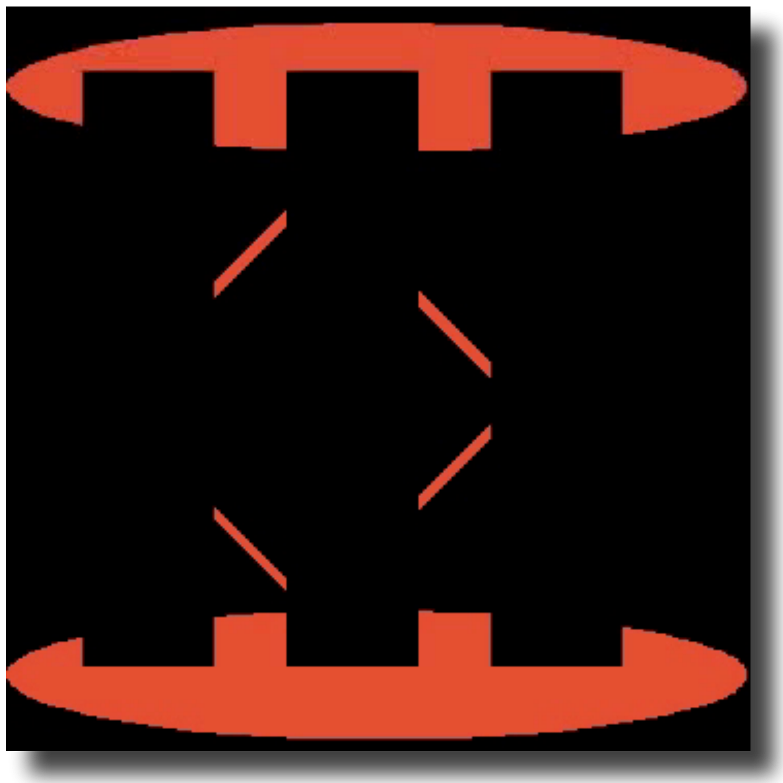
- expertise at various levels of abstraction

motivation: missing data

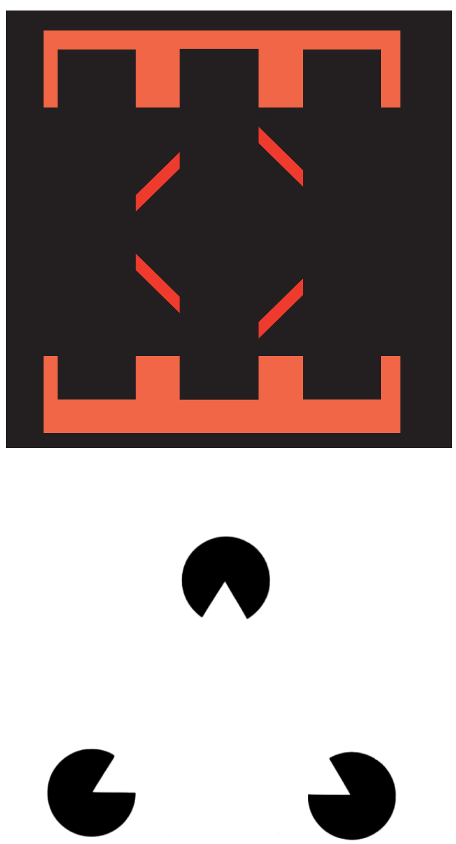


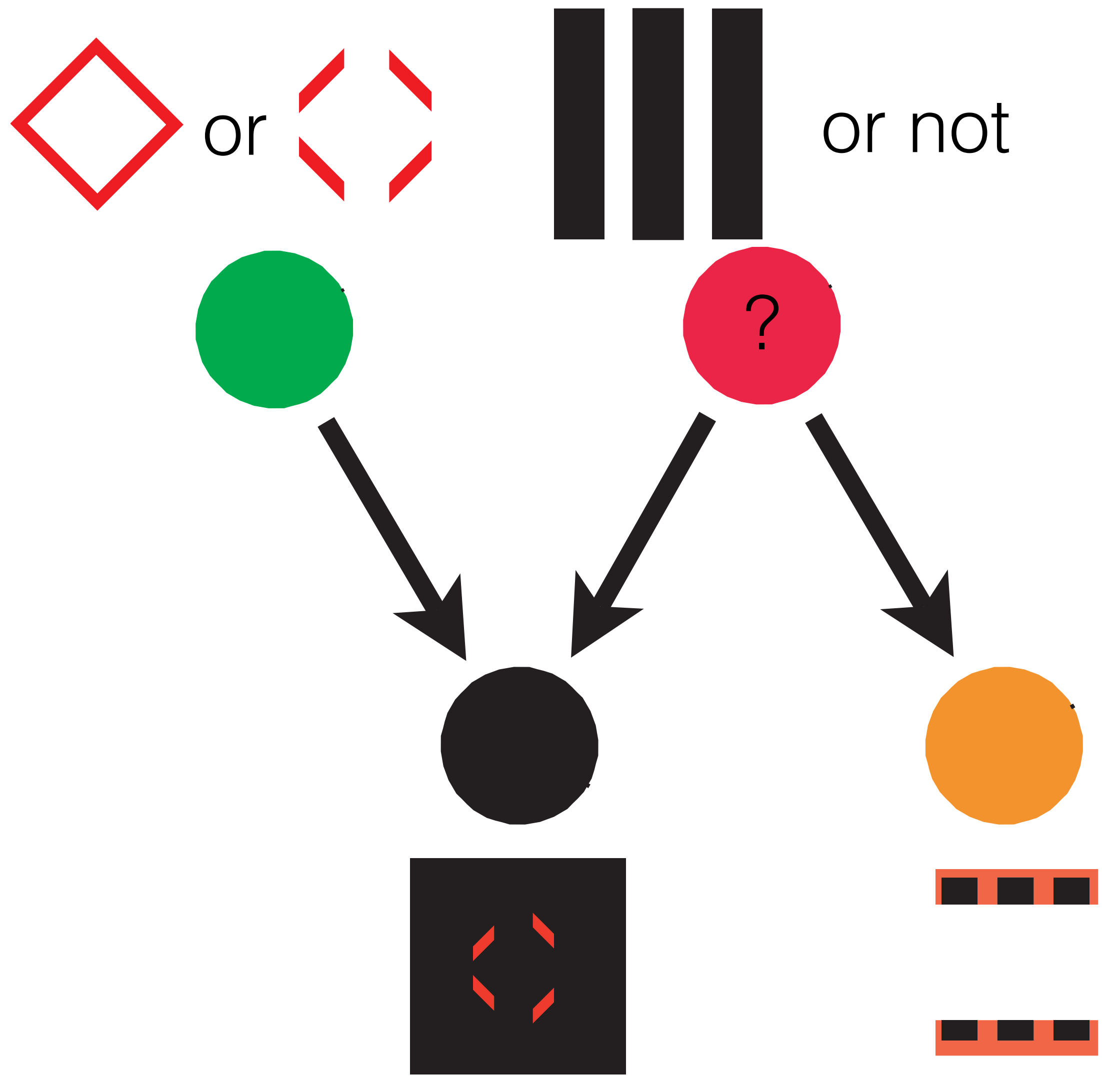
Top-down, generative models?

missing data & occlusion

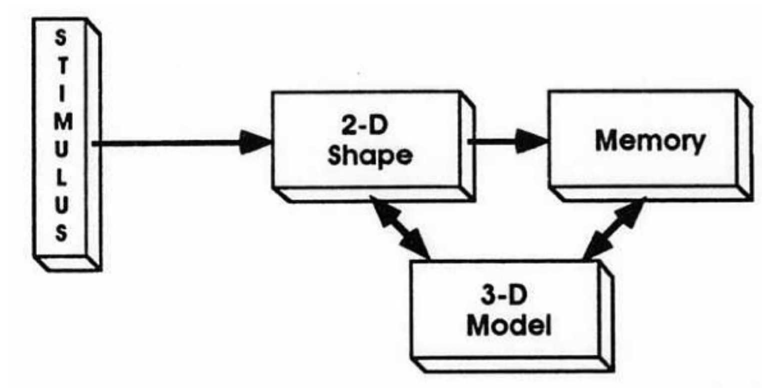


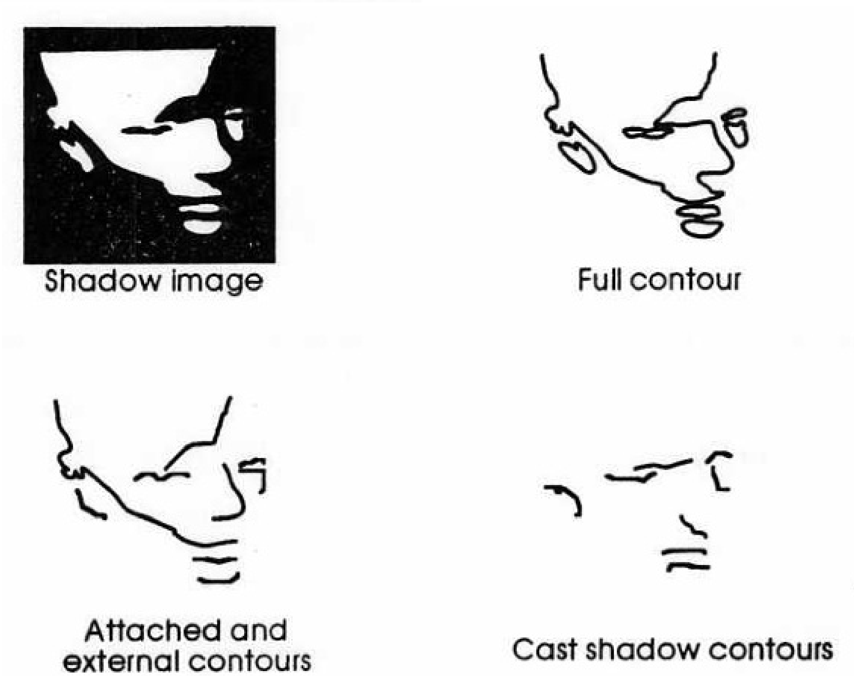
Perceptual "explaining away"





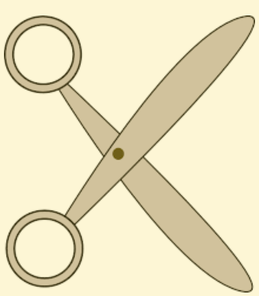
Extraneous data: recognition despite cast shadows





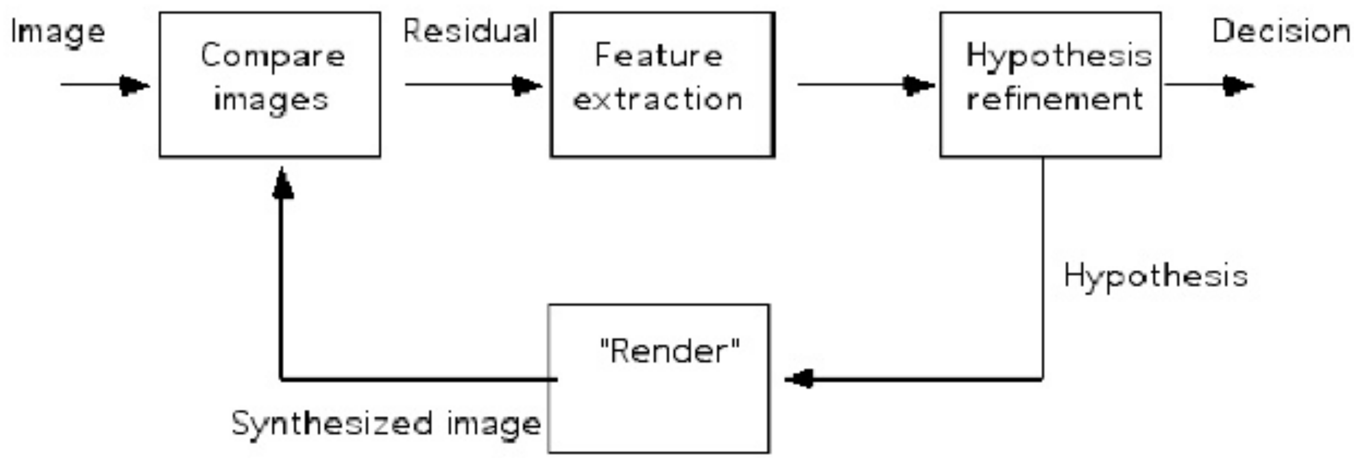
Cavanagh P (1991) What's up in top-down processing? In: Representations of Vision: Trends and tacit assumptions in vision research (Gorea A, ed), pp 295-304. Cambridge, UK: Cambridge University Press.

Object variations that haven't been seen before



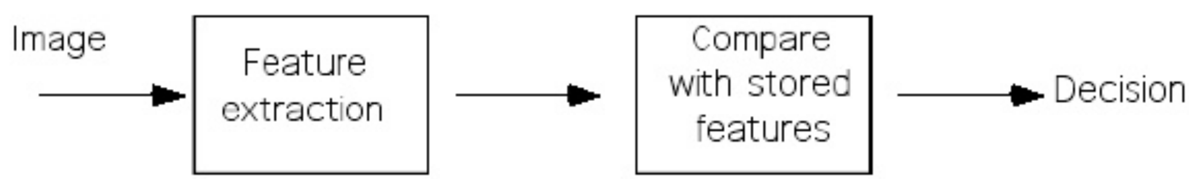
can recognize as scissors AND estimate an articulation

Suggests...



Bottom-up / Top-down

is a more complete picture than this



Bottom-up

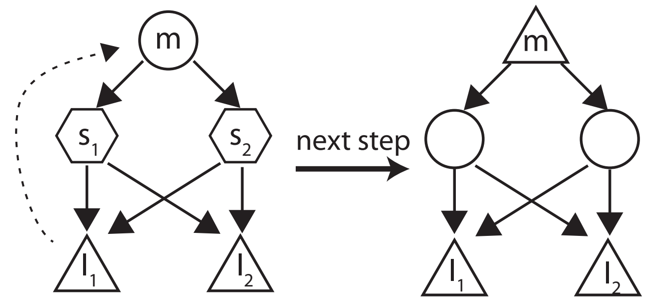
Doesn’t mean that feedback is necessary for recognition (Thorpe et al.)

But top-down feedback may be important for

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

feedback functions?

Coarse-to-fine

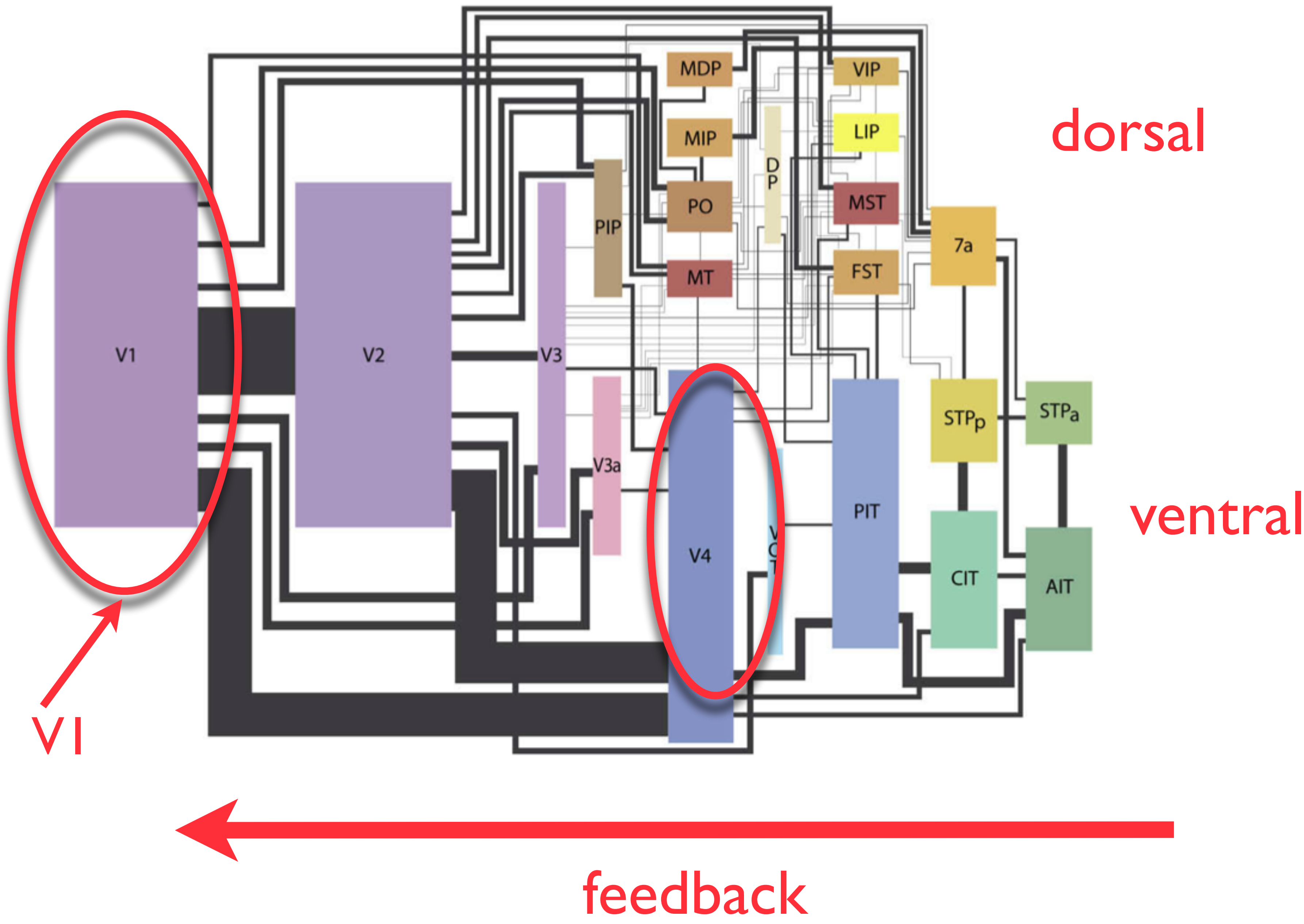


- E.g. is it a fox? If so, where is its nose?

Ambiguity reduction through top-down prediction

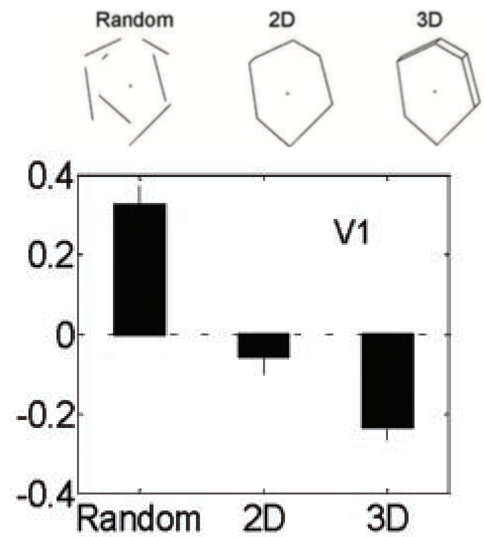
Hierarchically organized representations & expertise

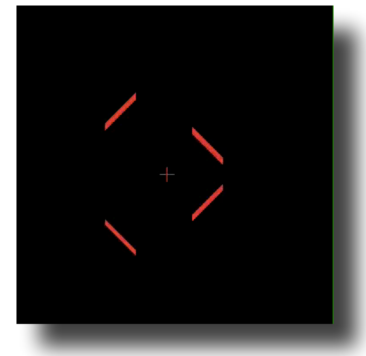
Feedback


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

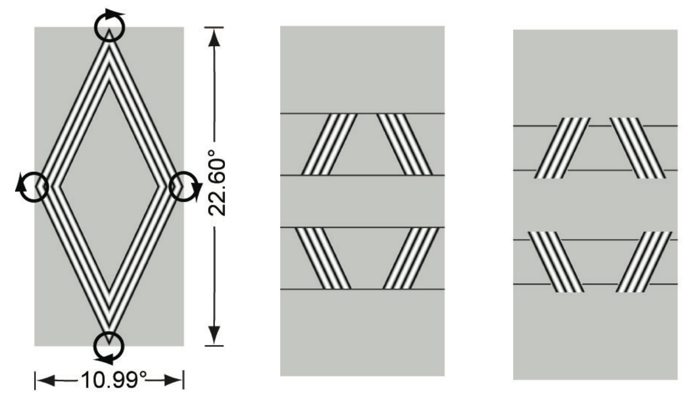
evidence for local, feature-specific feedback ?

- fMRI has shown localized relative suppression in V1 to edges when edges appeared to be perceptually “well explained” by whole shape (Murray et al., 2002).





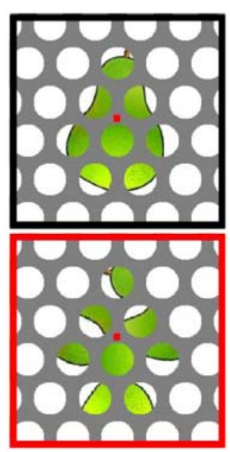
- human perceptual adaptation experiments show suppression to oriented lines—a local “feature”— when a whole shape is perceived. (He, Kersten, & Fang;2012)



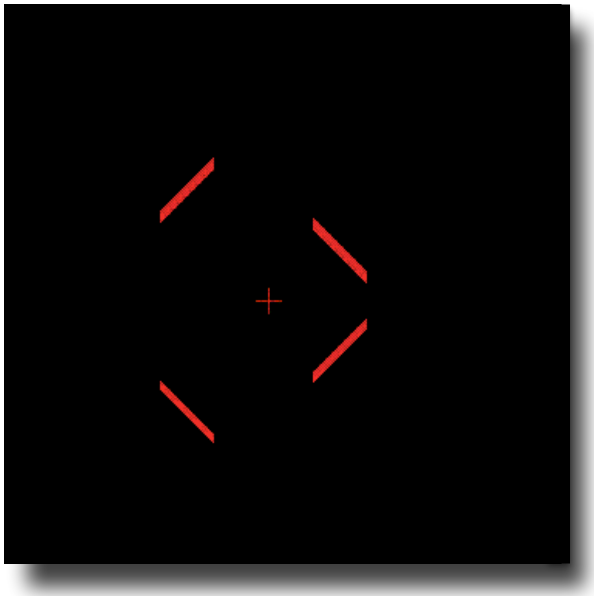
diamond perceived

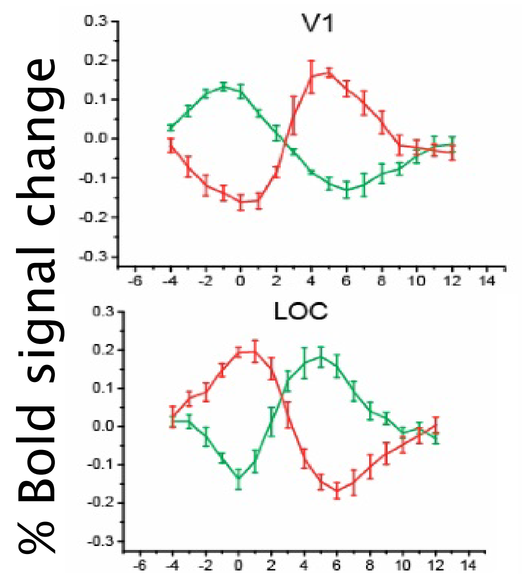
oriented patches perceived

- ultra-high resolution fMRI shows increased V1 activity to scrambled vs. whole shapes (Olman, Harel, Feinberg, He, Ugurbil, & Yacoub; (2012)



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)

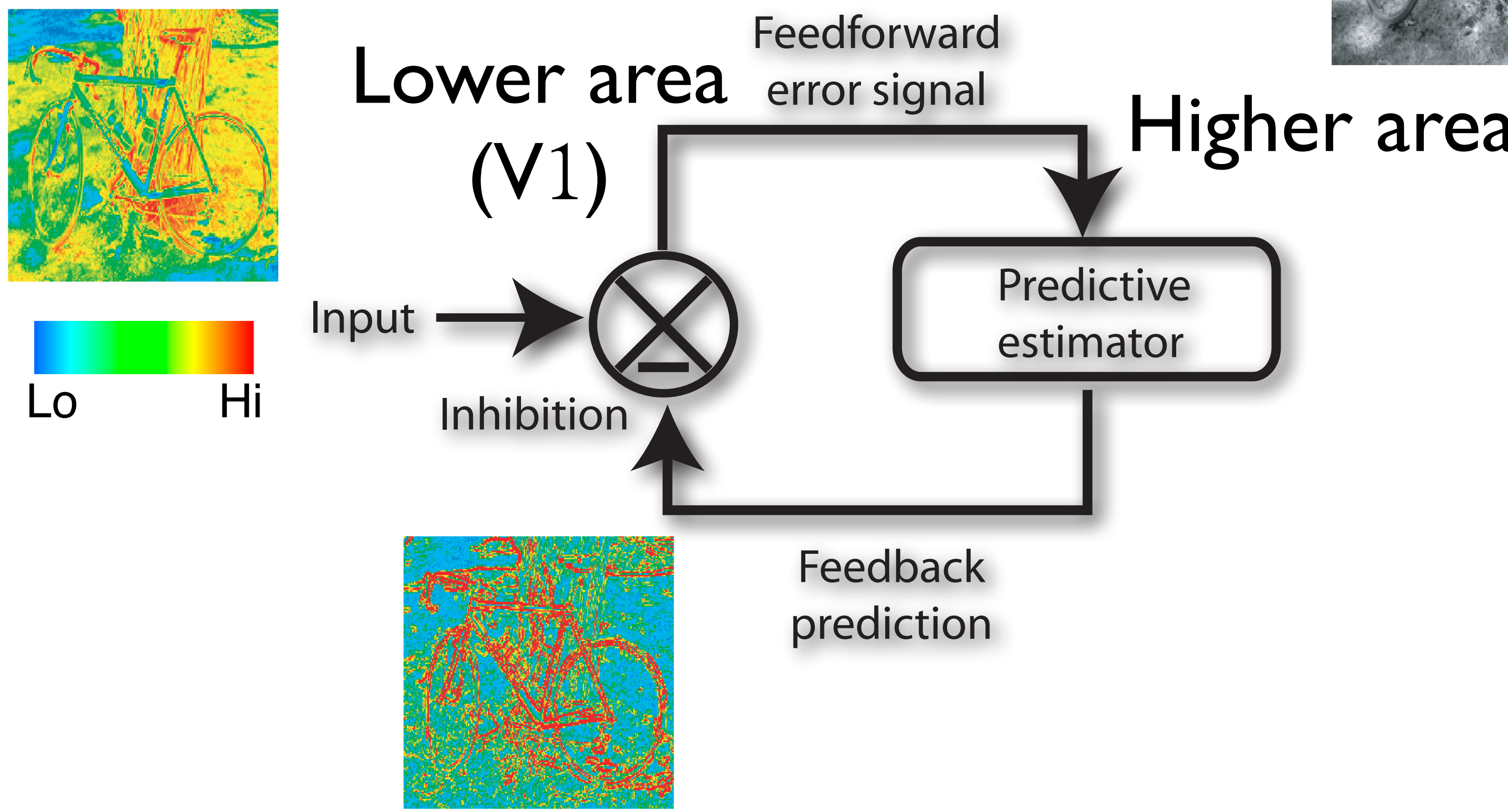
Behavioral evidence for top-down reduction of early activity? Use perceptual adaptation--the psychophysicist's electrode

Disambiguation?

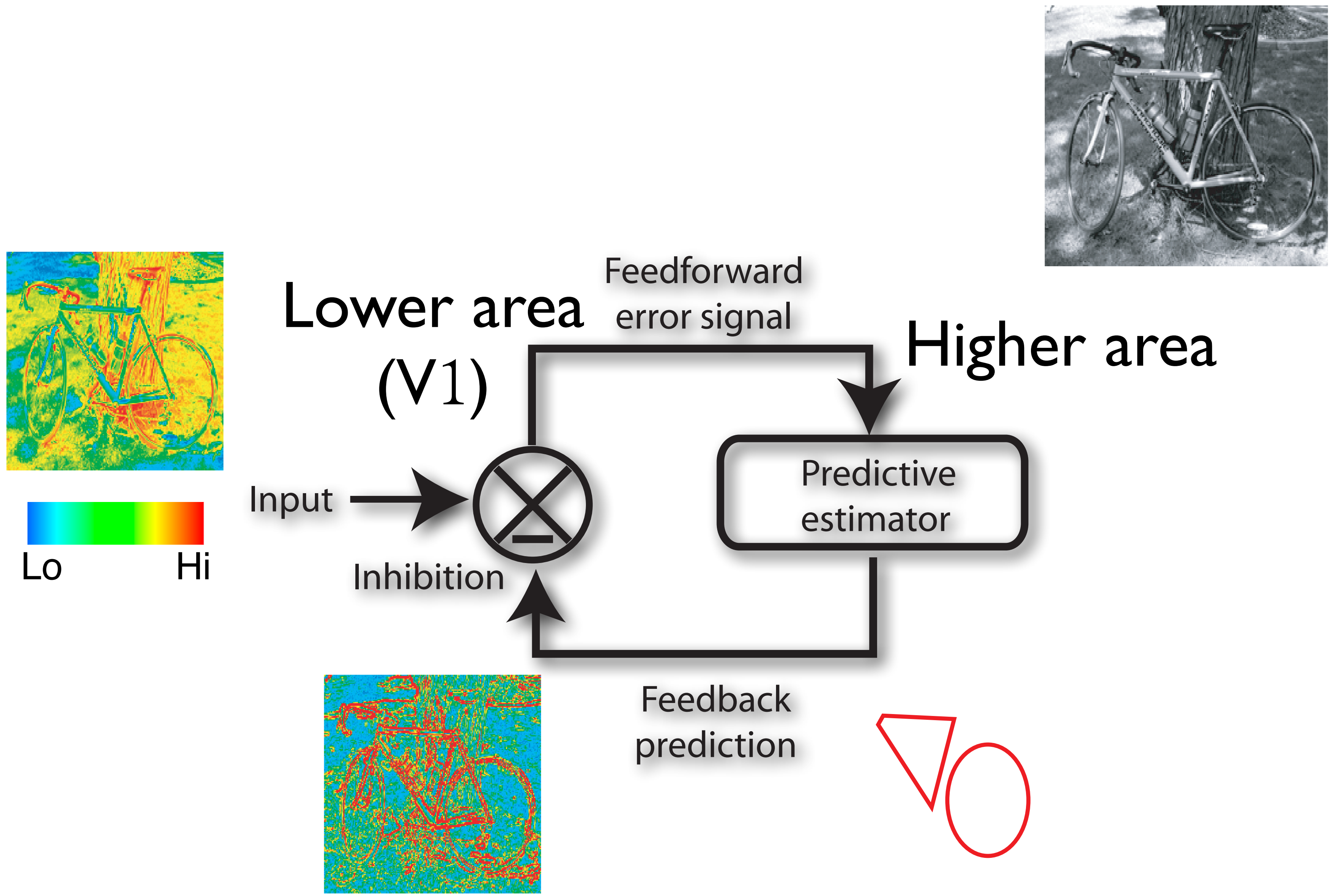
Predictive coding: suppress lower-level features that are consistent with a confident high-level interpretation. Reduce metabolic costs, signal new unexplained incoming information.

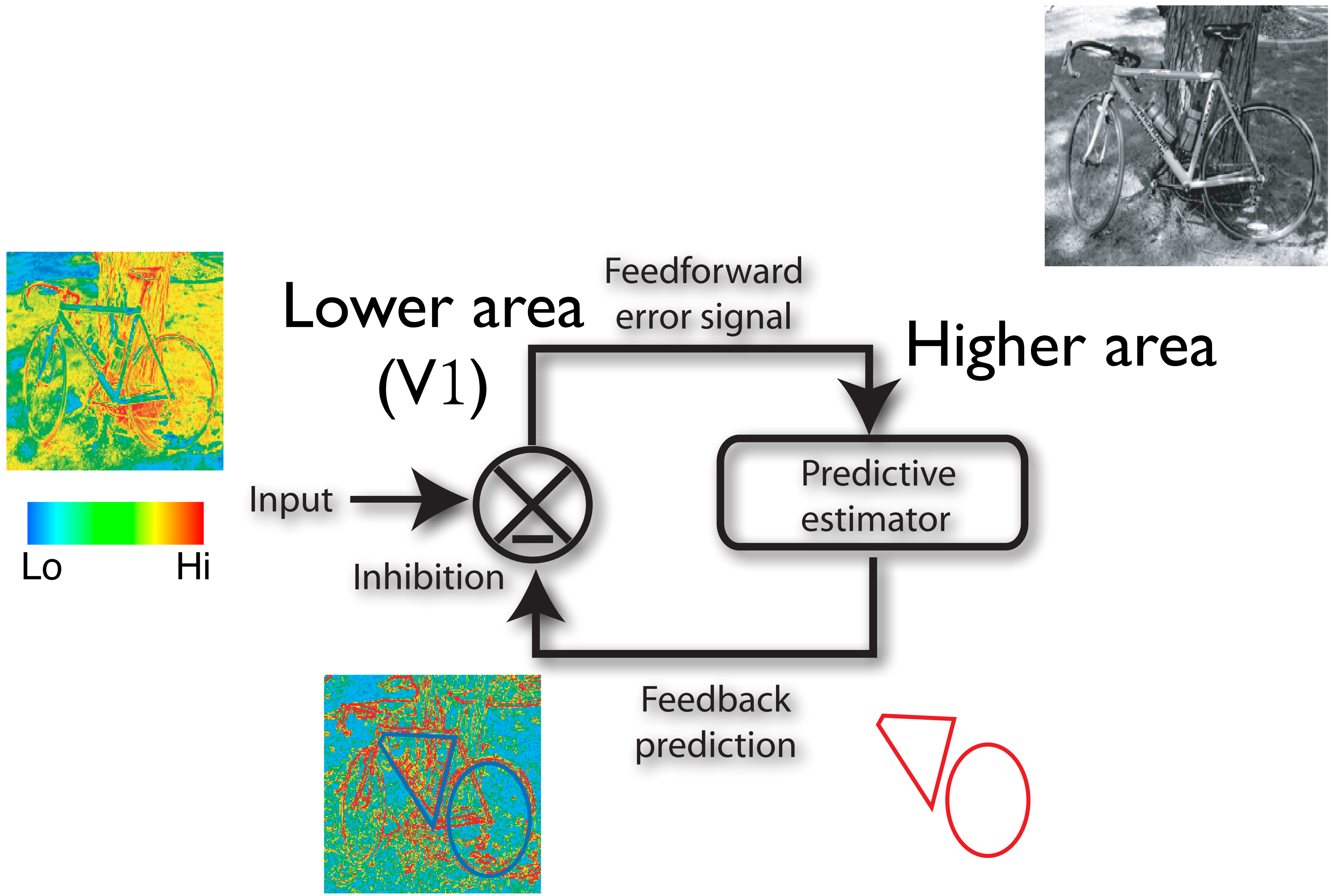
Analysis-by-synthesis. Bind lower-level information that might be required for executive tasks, e.g. fine-grain. : enhance lower-level consistent features and/or suppress inconsistent ones. Useful for representation and interpretation of novel patterns? Dealing with clutter?

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

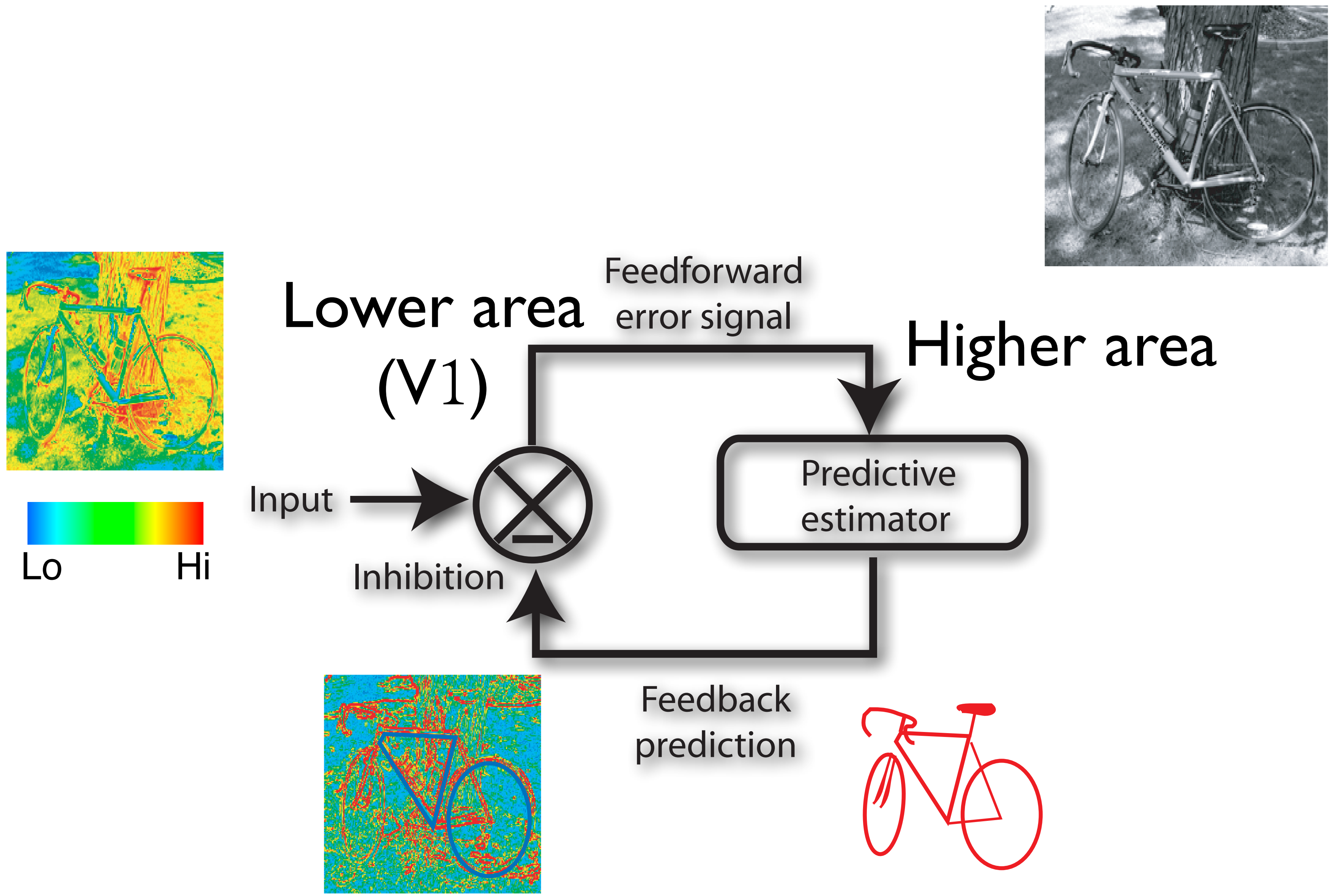


e.g. Rao, R. P., & Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. Neural Comput, 9(4), 721-763.

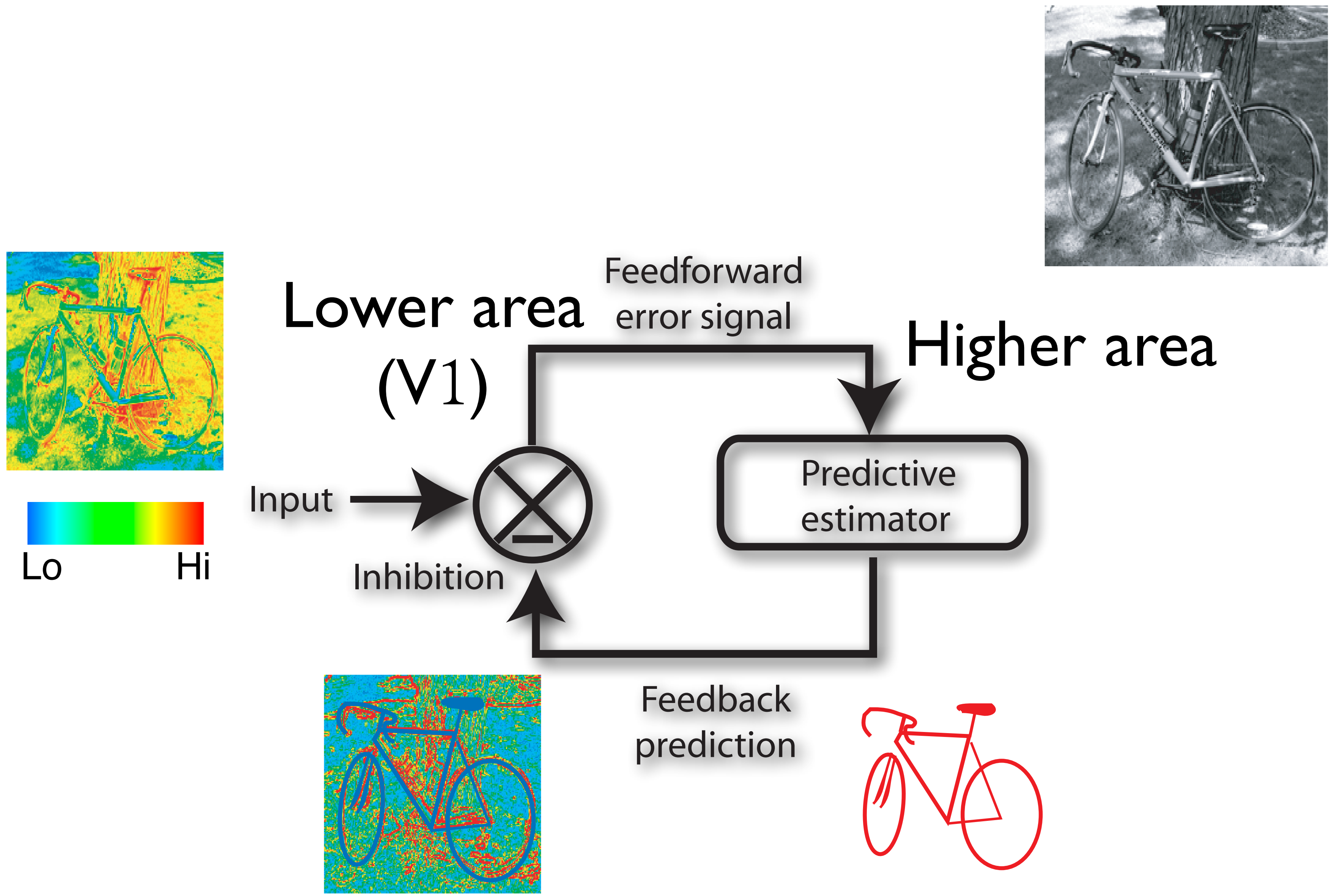

Lower area
(V1)
Feedforward
error signal
Higher area
Predictive
estimator
Input
Inhibition
Feedback
prediction
Lo
Hi



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

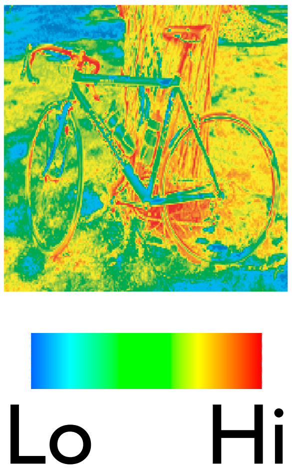


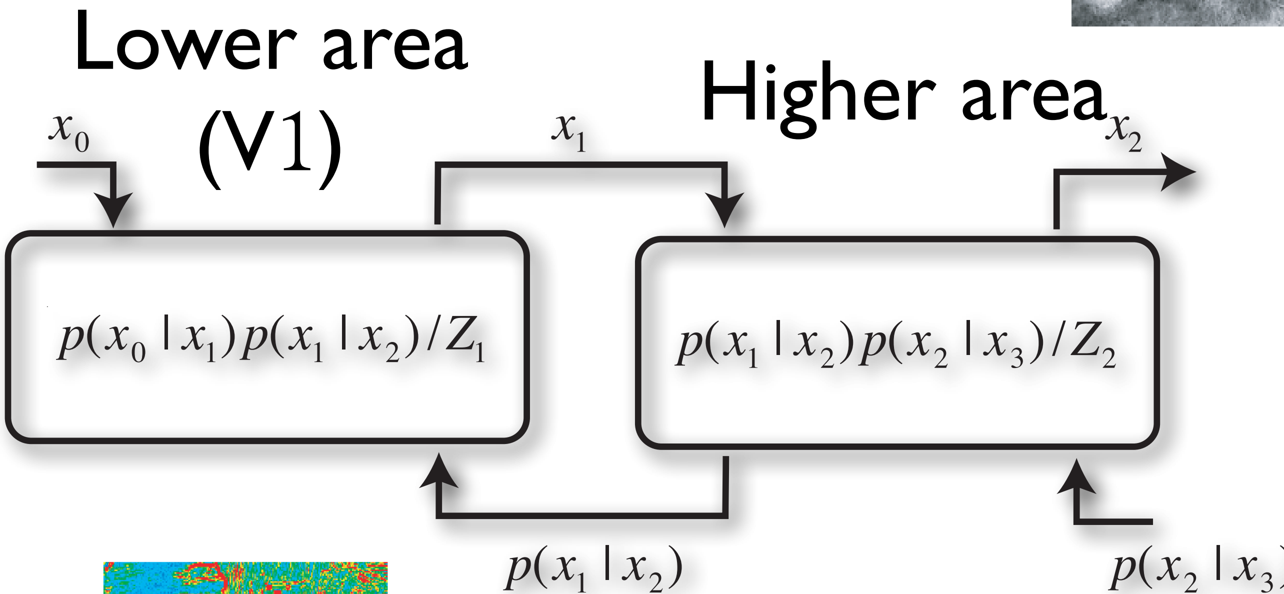
Lower area
(V1)
Feedforward
error signal
Higher area
Input
Predictive
estimator
Inhibition
Lo
Hi
Feedback
prediction

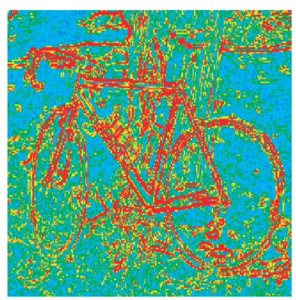


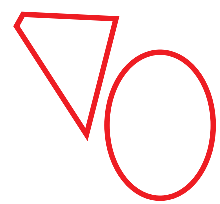
Lower area
(V1)
Feedforward
error signal
Higher area
Input
Predictive
estimator
Inhibition
Lo
Hi
Feedback
prediction

binding through enhancement of consistent features at lower levels

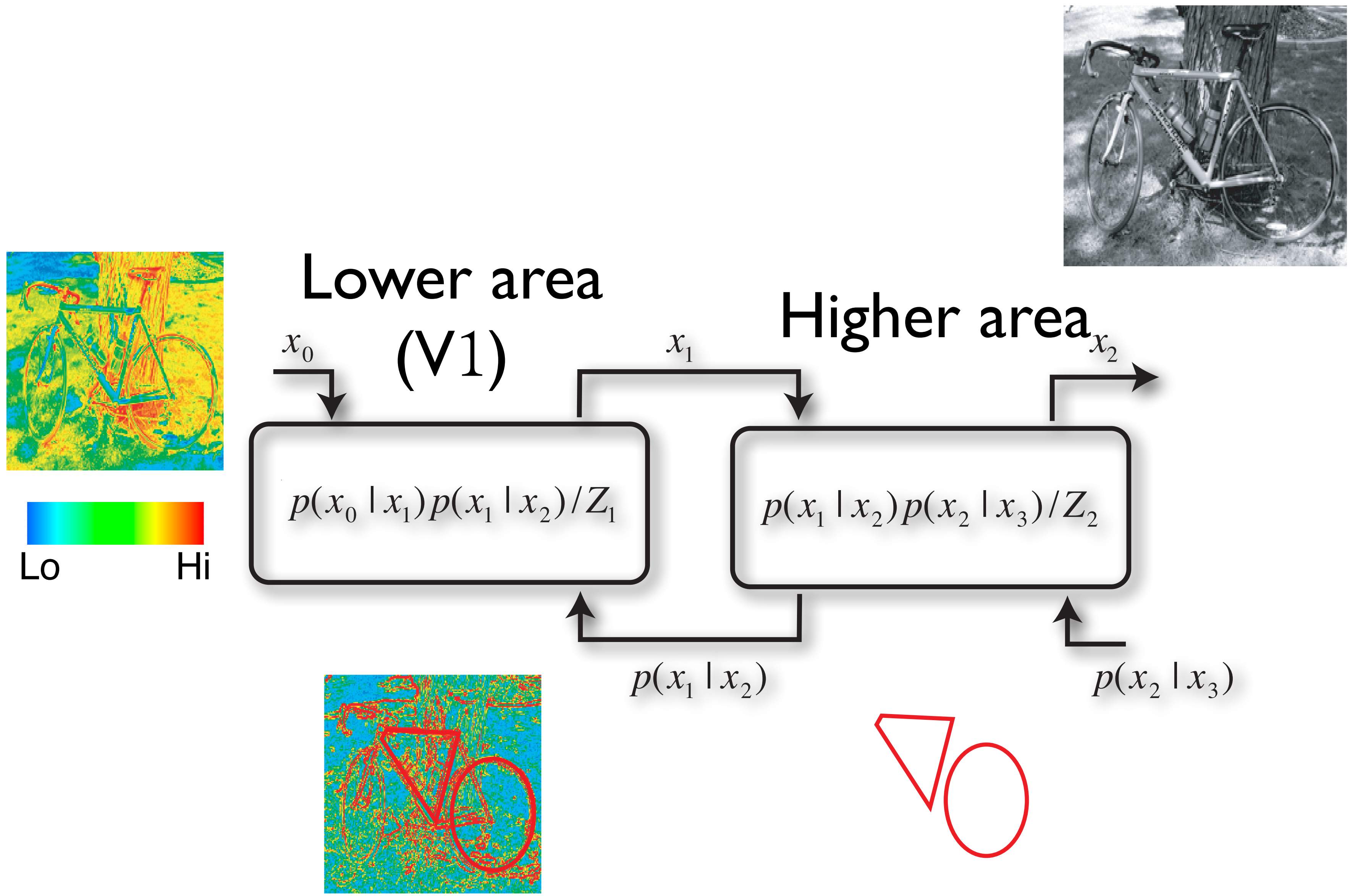




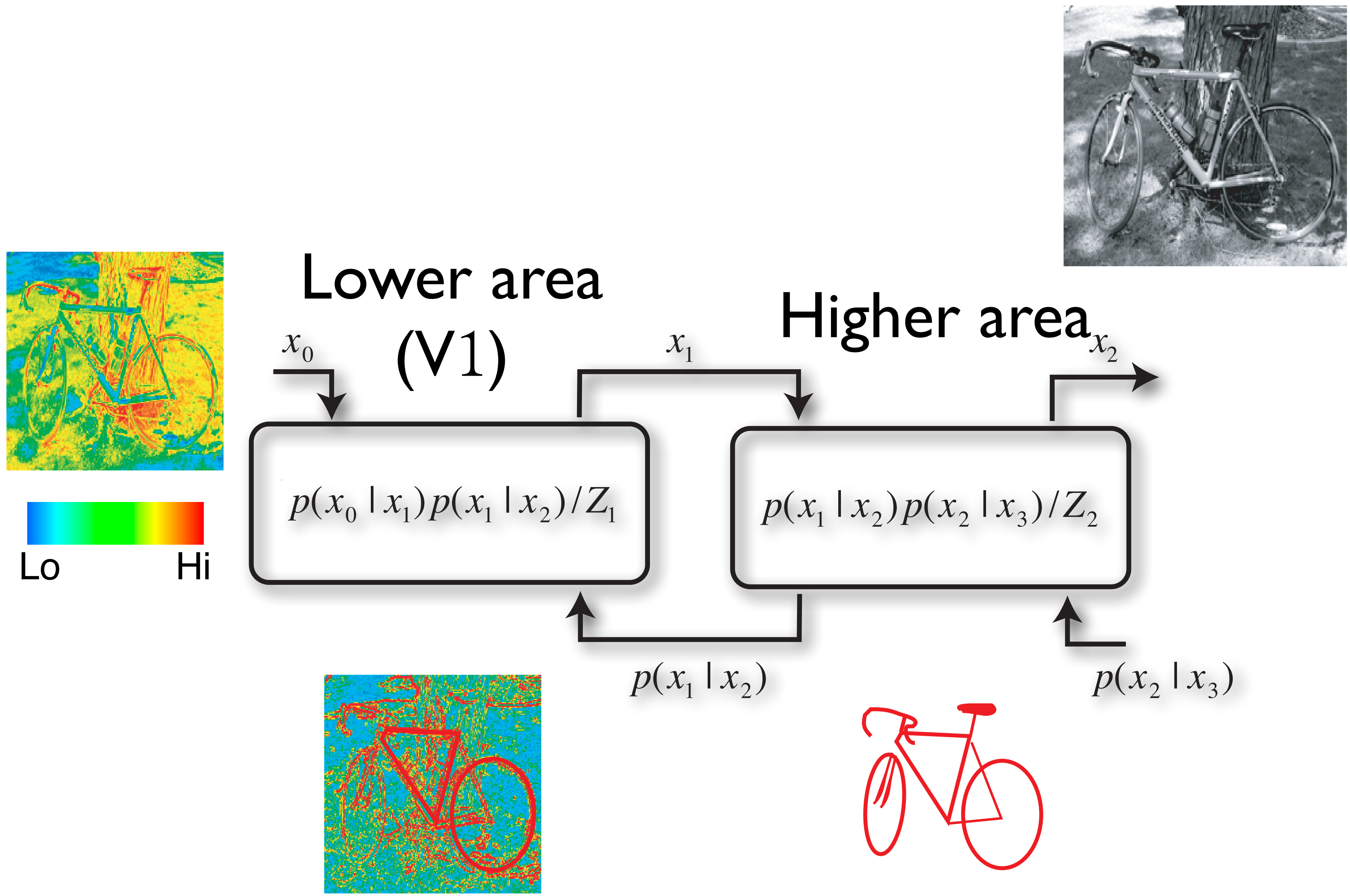


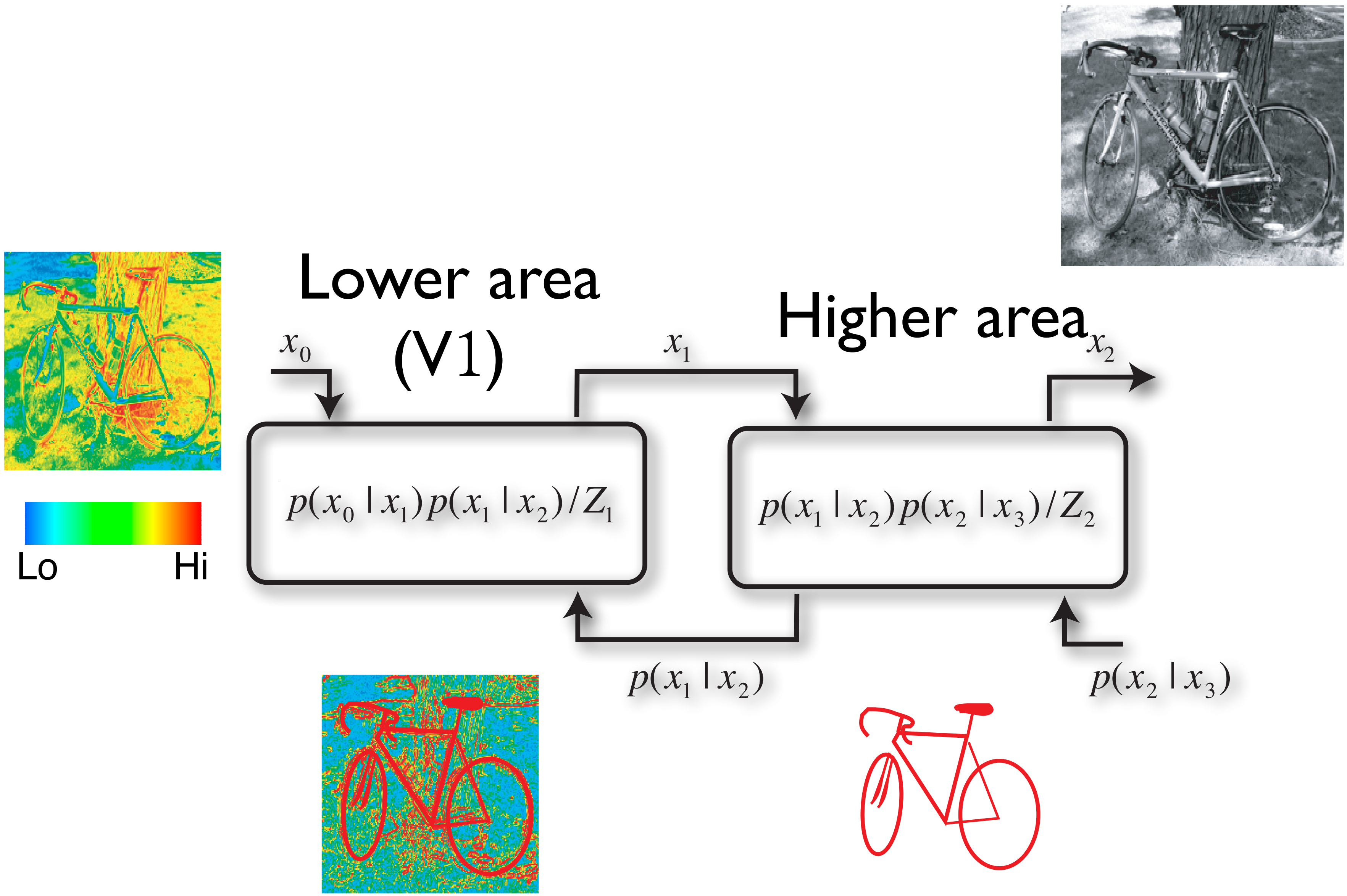


Lee & Mumford, 2003, JOSA

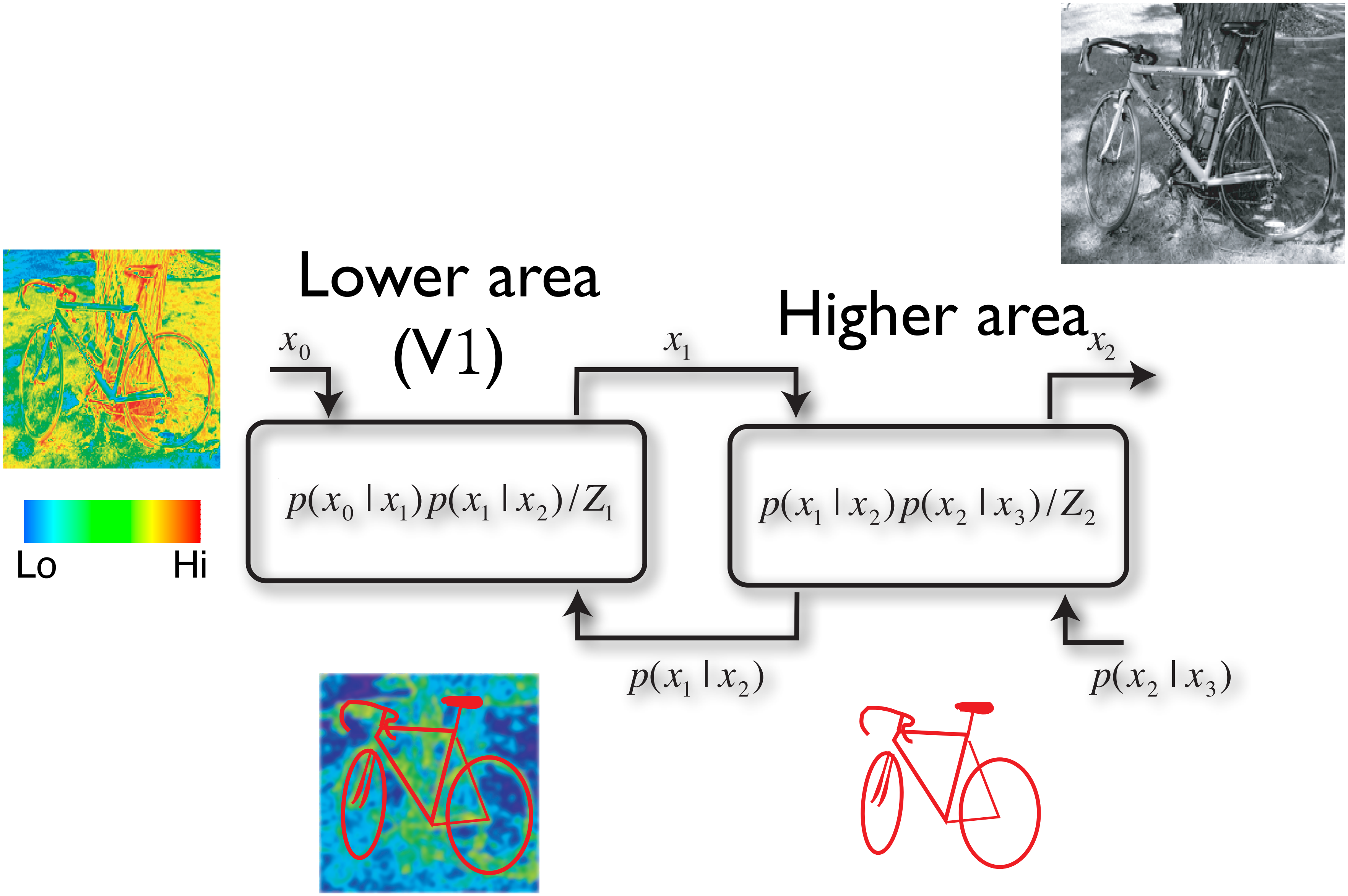


Lower area
(V1)
Higher area
x_0
x_1
x_2
$p(x_0 | x_1) p(x_1 | x_2) / Z_1$
$p(x_1 | x_2) p(x_2 | x_3) / Z_2$
$p(x_1 | x_2)$
$p(x_2 | x_3)$
Lo
Hi

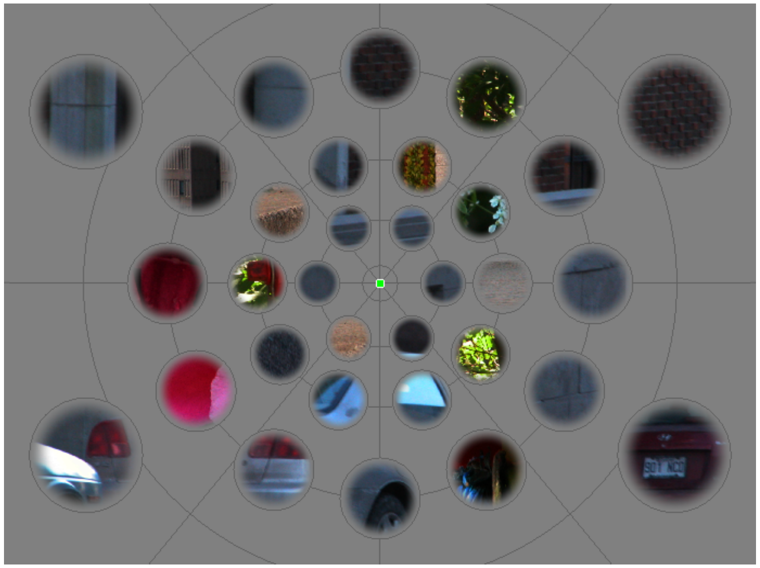

Lower area
(V1)
Higher area
Lo
Hi
x_0
x_1
x_2
$p(x_0 \mid x_1)p(x_1 \mid x_2)/Z_1$
$p(x_1 \mid x_2)p(x_2 \mid x_3)/Z_2$
$p(x_1 \mid x_2)$
$p(x_2 \mid x_3)$

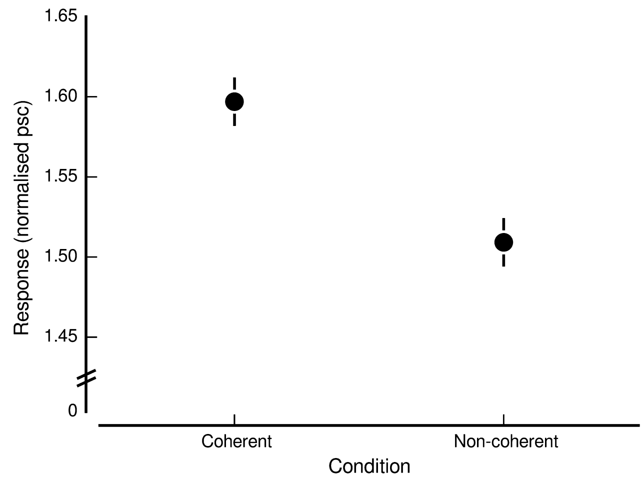


Lower area
(V1)
Higher area
x_0
x_1
x_2
$p(x_0 \mid x_1)p(x_1 \mid x_2)/Z_1$
$p(x_1 \mid x_2)p(x_2 \mid x_3)/Z_2$
$p(x_1 \mid x_2)$
$p(x_2 \mid x_3)$
Lo
Hi


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

Damien Mannion, Daniel Kersten & Cheryl Olman

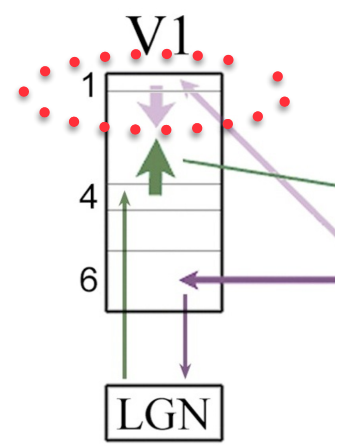




Mannion, D. J., Kersten, D. J., & Olman, C. A. (2015). Scene coherence can affect the local response to natural images in human V1.

Larger fMRI responses to peripheral patches belonging to the perceived “coherent” image

Preference for coherent patches found in more superficial layers of V1



Muckli, L., De Martino, F., Vizioli, L., Petro, L. S., Smith, F. W., Ugurbil, K., Goebel, R. and Yacoub E. (2015). Contextual Feedback to Superficial Layers of V1.

Feedback

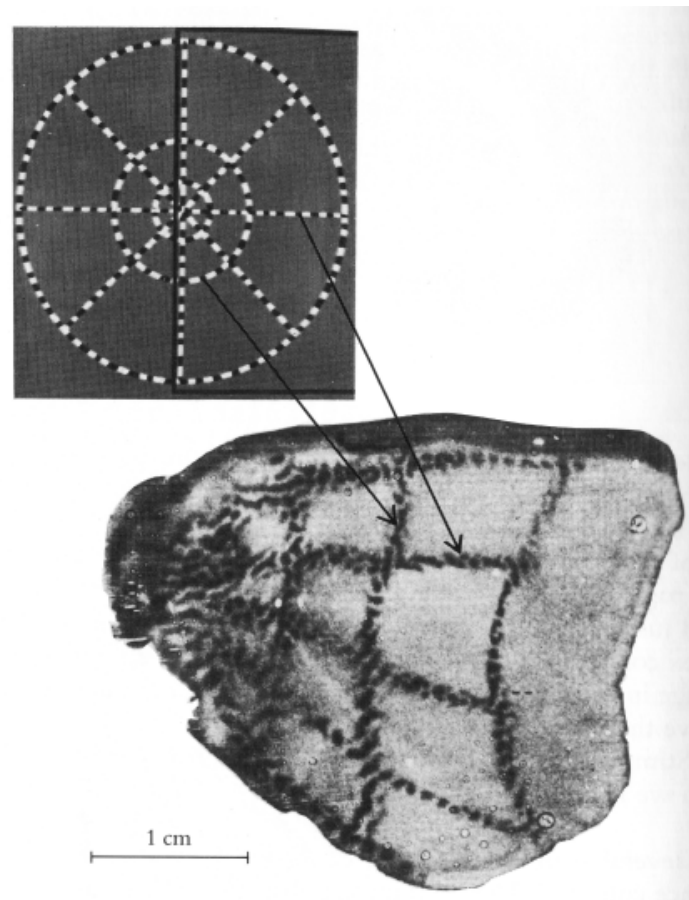
The executive metaphor

- Hierarchically organized expertise?
  - E.g. V1: Feature-specific tasks, Huk & Heeger, 2000; Working memory (Harrison & Tong, 2009); Perceptual learning (Hochstein & Ahissar,2002); Task-dependent changes in early receptive fields (McManus et al., 2011);
  - Foveal V1 as a high-resolution spatial buffer (Lee et al. 1998,; Williams et al., 2008);
  - Fan, X., Wang, L., Shao, H., Kersten, D., & He, S. (2016). Temporally flexible feedback signal to foveal cortex for peripheral object recognition. PNAS.
- Use of built-in generative knowledge?
  - The “perceived size and V1” puzzle

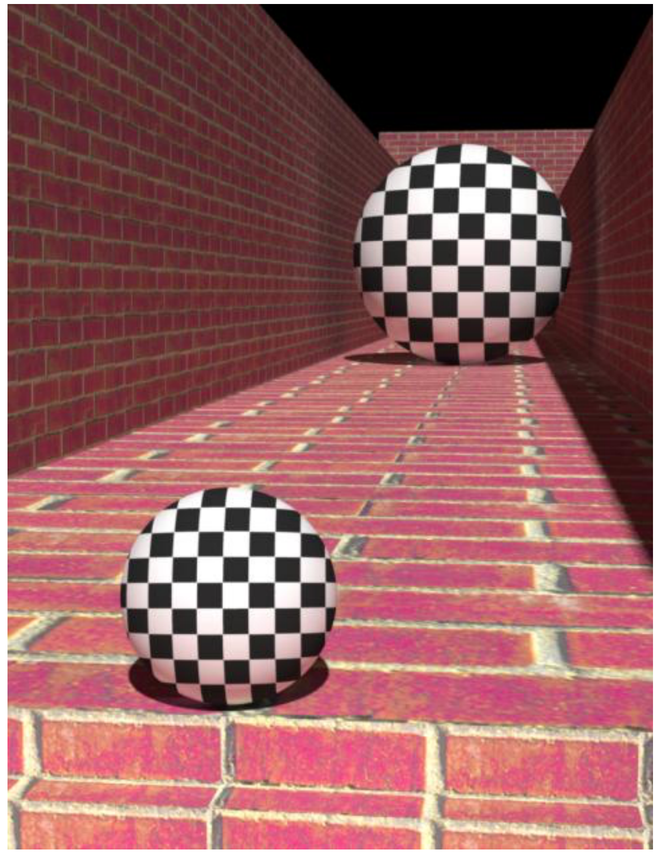
recall global organization of V1

global: hypercolumns arranged retinotopically

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



Feedback: Executive metaphor?

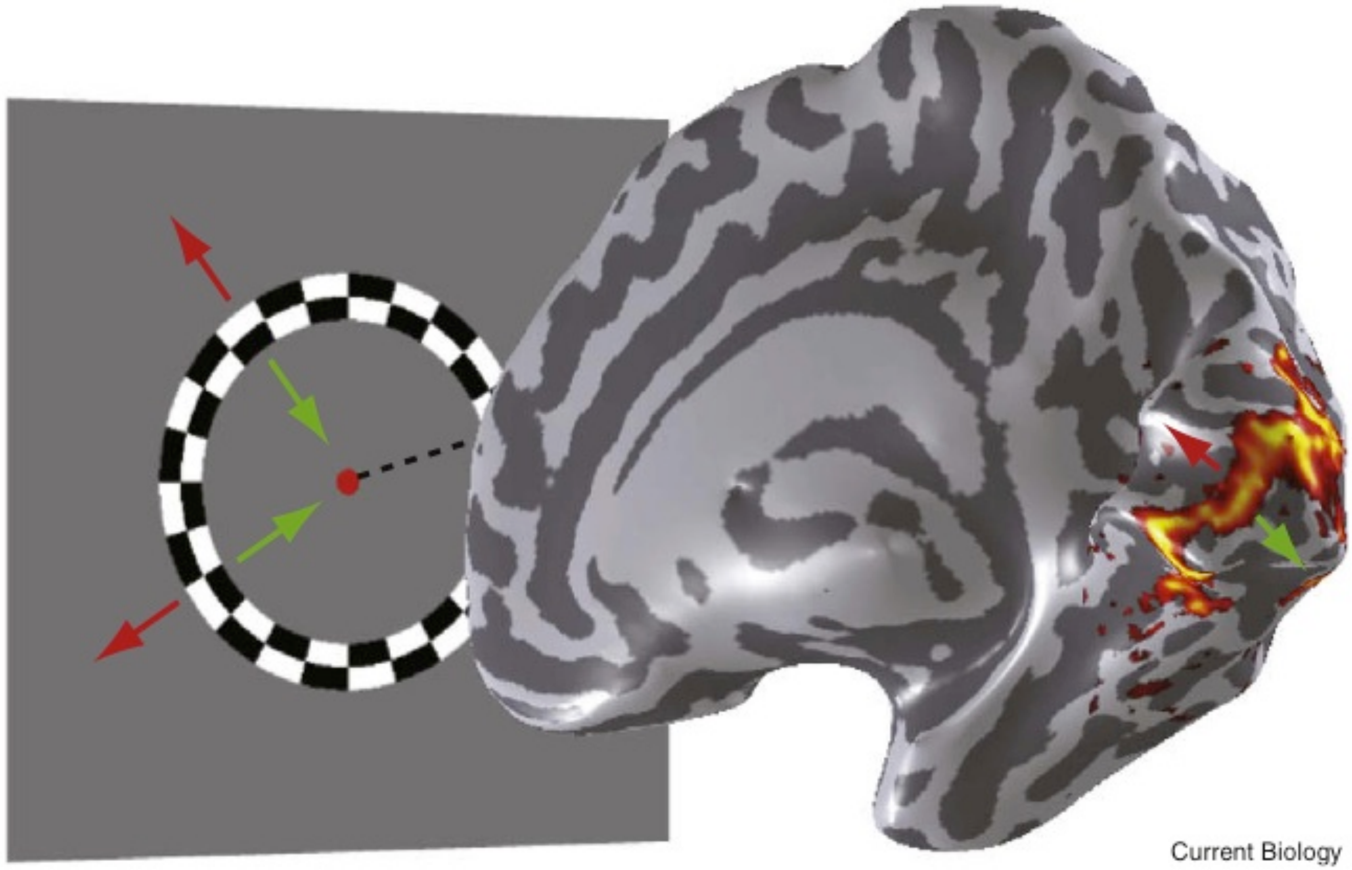


Fang, Boyaci, Kersten, & Murray, S. O. (2008). Attention-dependent representation of a size illusion in human V1. Current Biology

Same angular size, different physical size

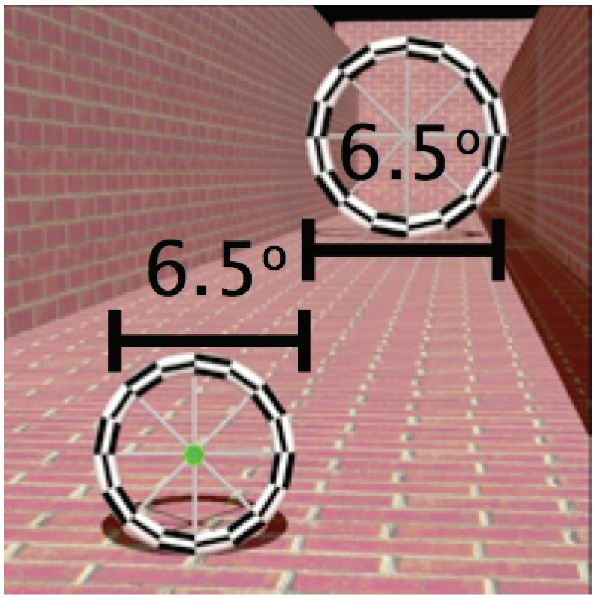
Encyclopedia of Perception, Goldstein Ed., 2009

V1 has a retinotopic map, so for an actual increase in ring size, θ ,in the image, we expect:



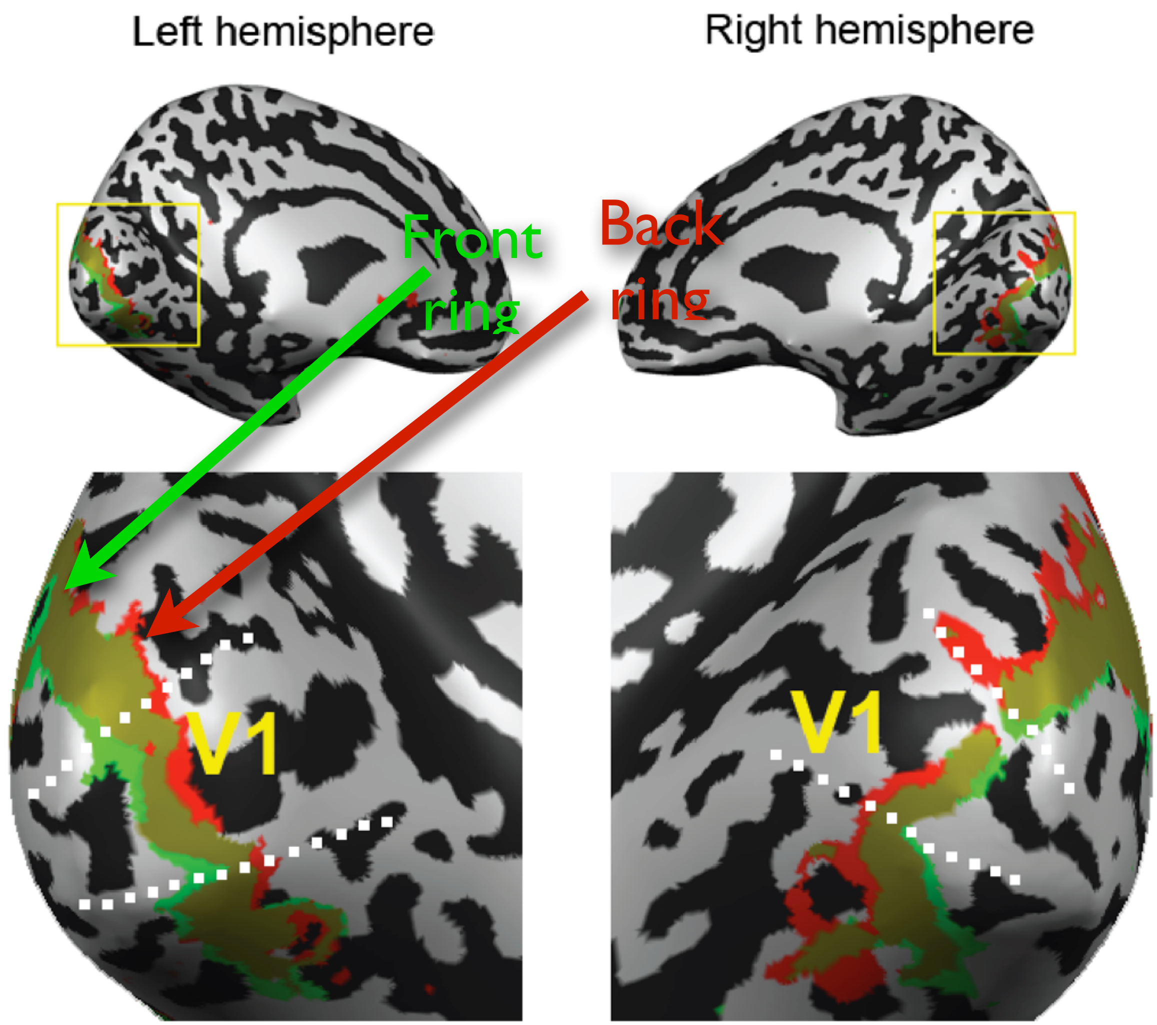
Huk, A. C. (2008) Visual Neuroscience: Retinotopy meets Percept-otopy, Current Biology, 18, 21, R1005-1007.

what was found for an illusory increase in ring size



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

Ni, A. M., Murray, S. O., & Horwitz, G. D. (2014). Object-Centered Shifts of Receptive Field Positions in Monkey Primary Visual Cortex. *Current Biology*, 1–6



attend-to-ring condition