The role of top-down processing in segmentation & recognition

Object recognition, given real images

- clutter, occlusion, noise
- role of cortical architecture

Object recognition in real images

Background clutter and occlusion

Object recognition given occlusion, clutter

Linking local information (features) likely to belong to the same object or pattern

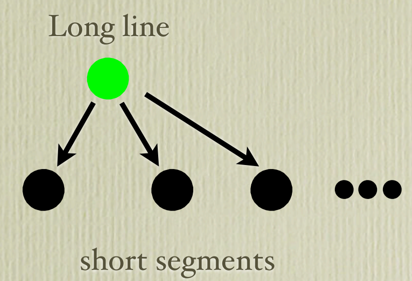
- local ambiguity, noise
- need for generic priors, e.g. smoothness, contour and region-based grouping

Resolving competing explanations

- occlusion, clutter
- need for domain-specific priors

Simple influence graphs

Cue integration

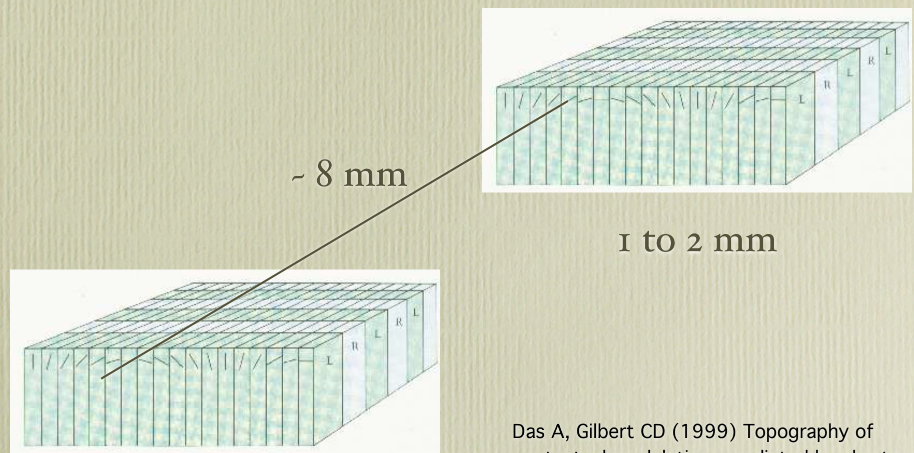


Parent P, Zucker SW (1989) Trace inference, curvature consistency, and curve detection. IEEE Transactions on Pattern Analysis & Machine Intelligence 11:823-839.

Yuille AL, Fang F, Schrater P, Kersten D (2004) Human and Ideal Observers for Detecting Image Curves. In: Advances in Neural Information Processing Systems 16 (Thrun S, Saul L, Schoelkopf B, eds). Cambridge, MA: MIT Press.

Cortical basis?

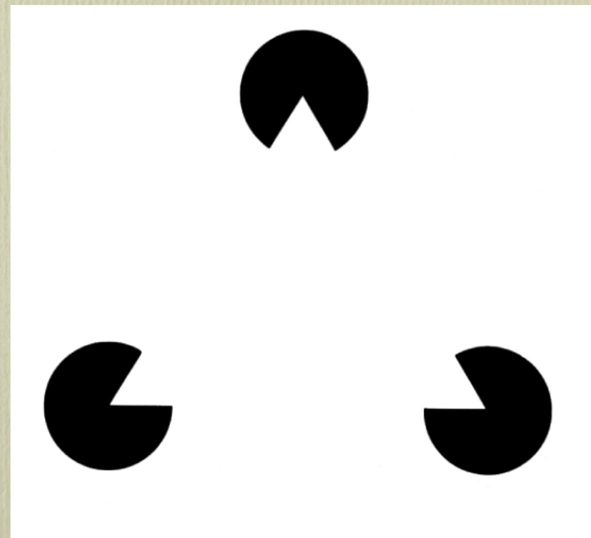
Short segments to long lines? Within-area linkage?



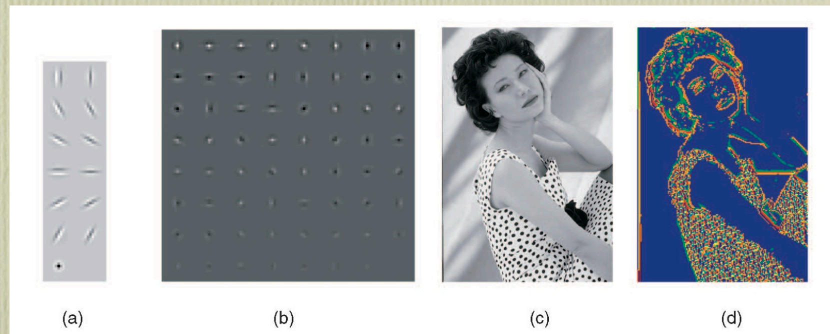
1 to 2 mm

Das A, Gilbert CD (1999) Topography of contextual modulations mediated by short-range interactions in primary visual cortex. Nature 399:655-661.

But what about whole shapes?



Region-based grouping



From: Martin, D. R., Fowlkes, C. C., & Malik, J. (2004). Learning to detect natural image boundaries using local brightness, color, and texture cues. IEEE Trans Pattern Anal Mach Intell, 26(5), 530-549.

Object recognition given occlusion, clutter

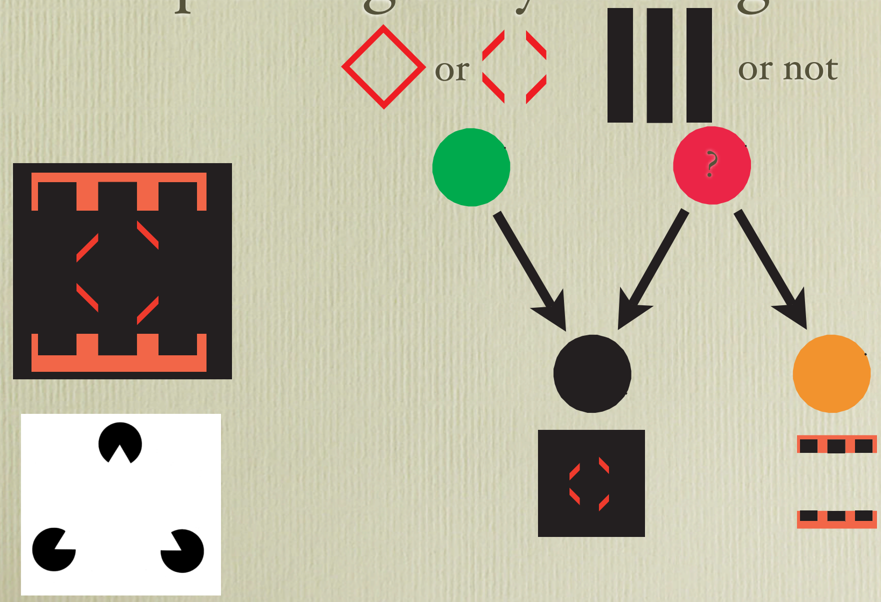
Linking local information (features) likely to belong to the same object or pattern

- local ambiguity, noise
- need for generic priors, e.g. smoothness

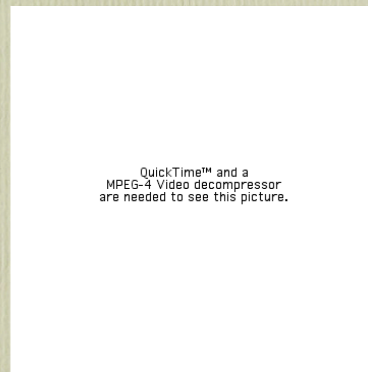
Resolving competing explanations

- occlusion, clutter
- need for domain-specific priors

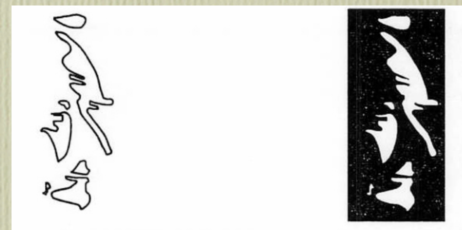
Competing explanations: Explaining away missing data

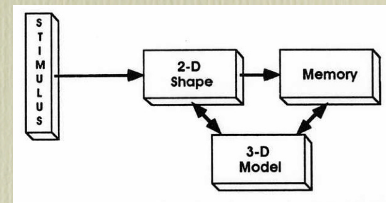


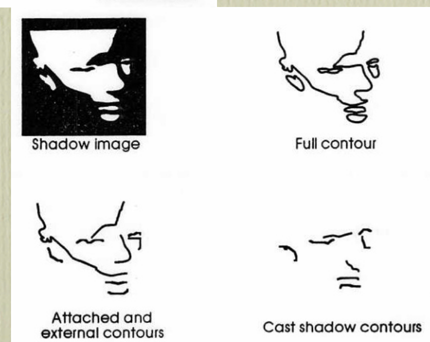
Auxiliary evidence for occlusion



Recognition despite cast shadows

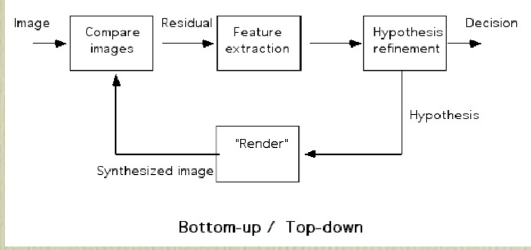




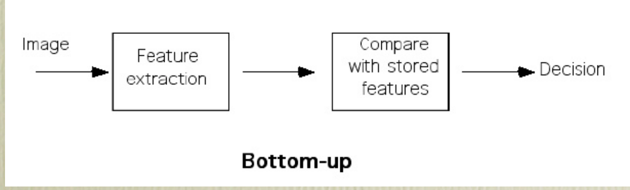


Cavanagh P (1991) What's up in top-down processing? In: Representations of Vision: Trends and tacit assumptions in vision research (Gorea A, ed), pp 295-304. Cambridge, UK: Cambridge University Press.

Suggests...



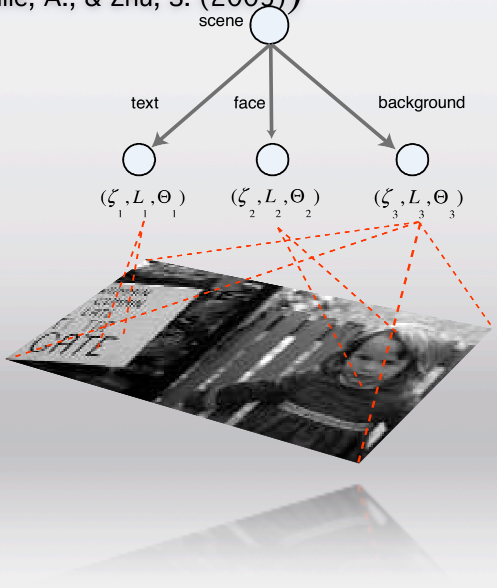
is a more complete picture than this



Doesn’t mean that bottom-up segmentation can’t work, but that achieving high-performance requires a combination of good bottom-up processing with top-down verification.

Neural evidence for top-down role in perceptual organization?

Computer vision
Image parsing: analysis by synthesis
(Tu, Z., Chen, X., Yuille, A., & Zhu, S. (2005))

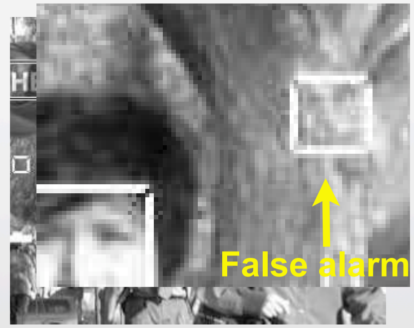


- Find most probable scene description
- Bottom-up "proposals" (cues) to access three types of models (text, faces, background/texture) models
- Verification through top-down synthesis
- If bottom-up proposals are good, synthesis is not needed to find most probable scene
- Flexible graph

Image parsing & "Explaining away"

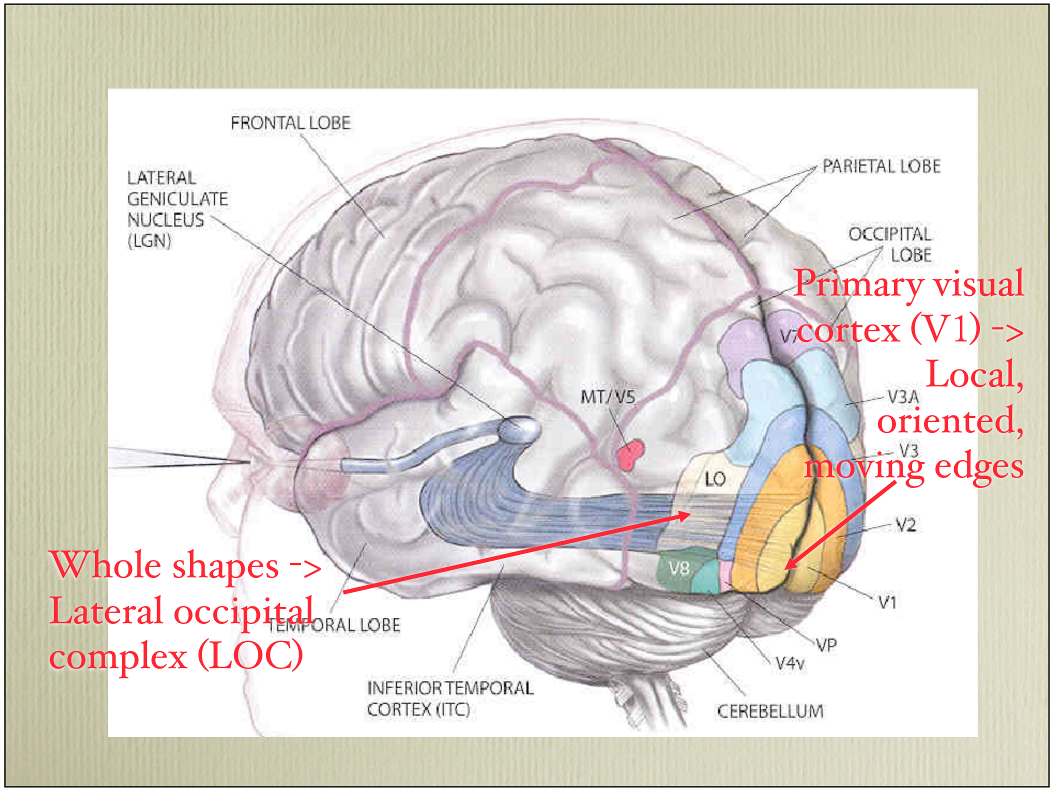
Input

Tu, Z., Chen, X., Yuille, A., & Zhu, S. (2005). Image Parsing: Unifying Segmentation, Detection and Recognition. IJCV, 63(2).

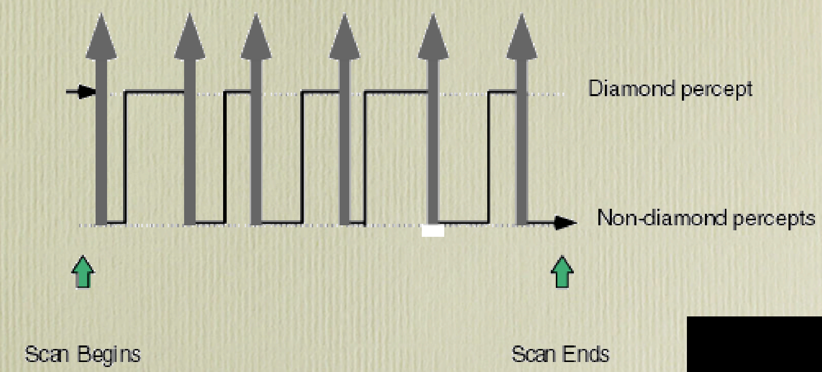


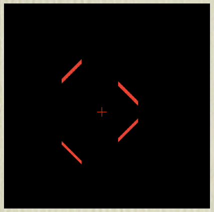
Bottom-up result

Synthesized image

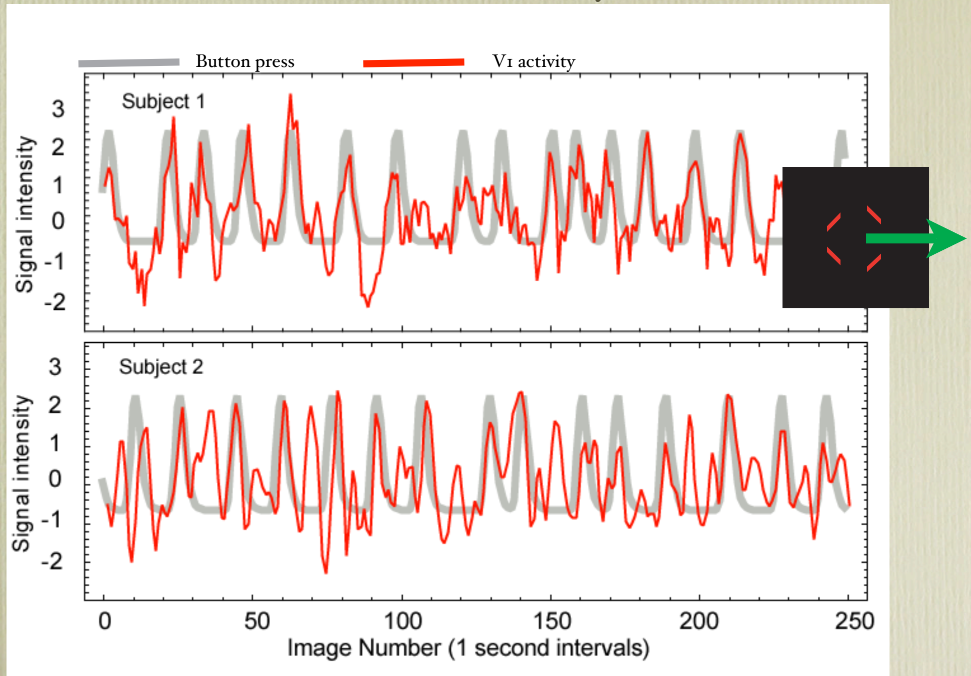

FRONTAL LOBE
LATERAL GENICULATE NUCLEUS (LGN)
PARIETAL LOBE
OCCIPITAL LOBE
Primary visual cortex (V1) -> Local, oriented, moving edges
MT/V5
V3A
V3
LO
V2
V8
V1
VP
V4v
Whole shapes -> Lateral occipital complex (LOC)
TEMPORAL LOBE
INFERIOR TEMPORAL CORTEX (ITC)
CEREBELLUM

Button presses

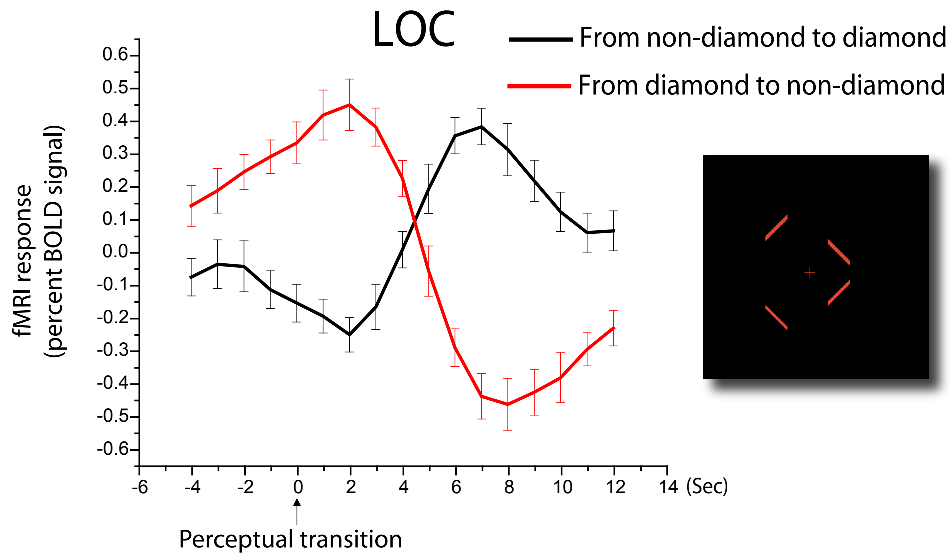




Perceptual organization correlates with reduced V1 activity

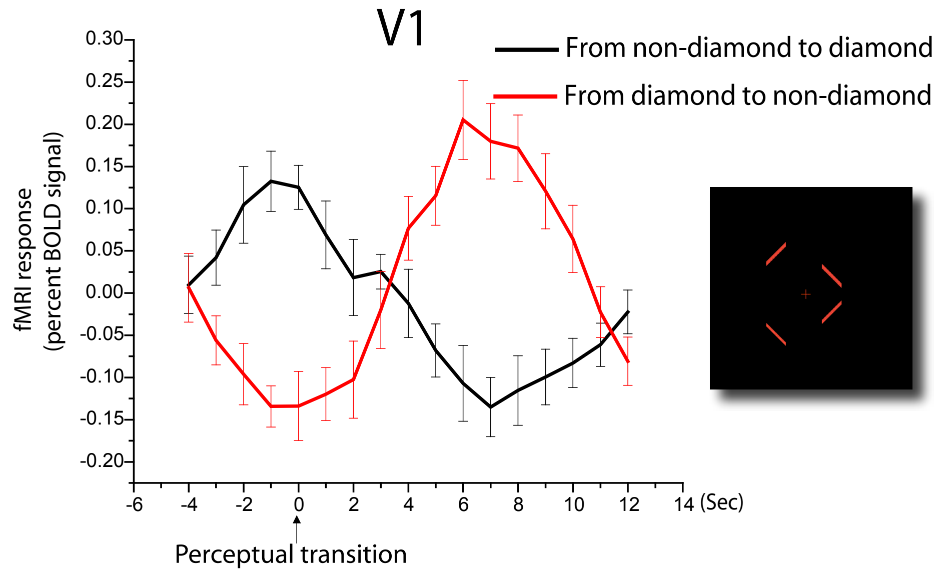


Perceptual organization is correlated with *increased* LOC activity

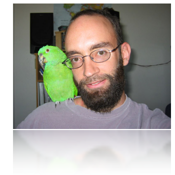


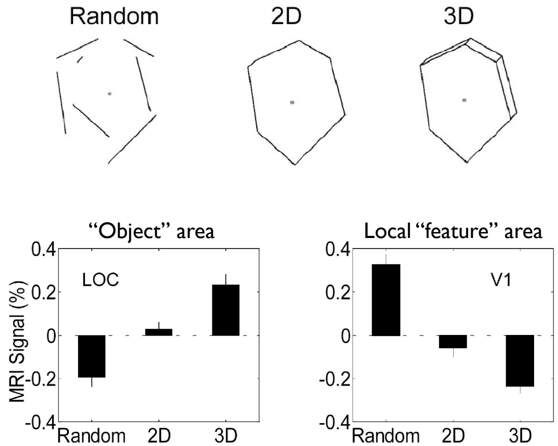
From: Fang, Murray, He & Kersten, 2004, International Congress of Psychology, Beijing

Perceptual organization is correlated with *decreased* V1 activity



Shape perception can reduce V1 activity





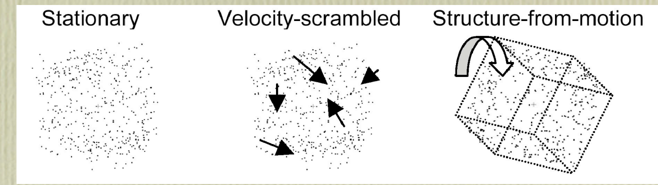
Explanation?

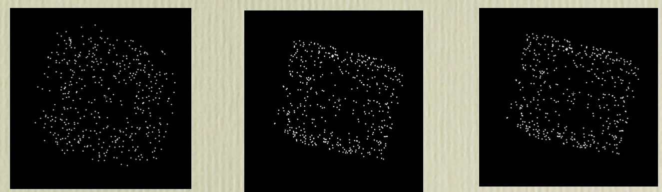
Many...

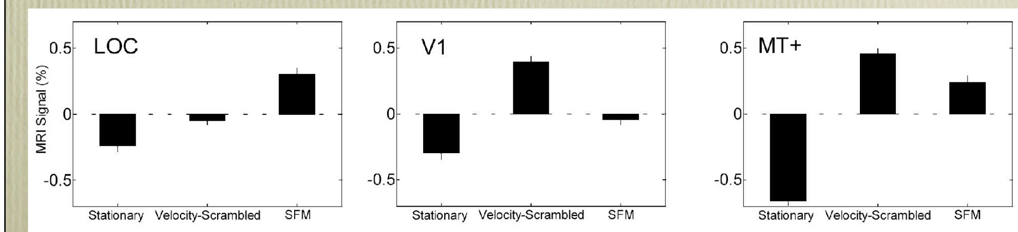
- "Explaining away" through predictive coding
- Sparse coding

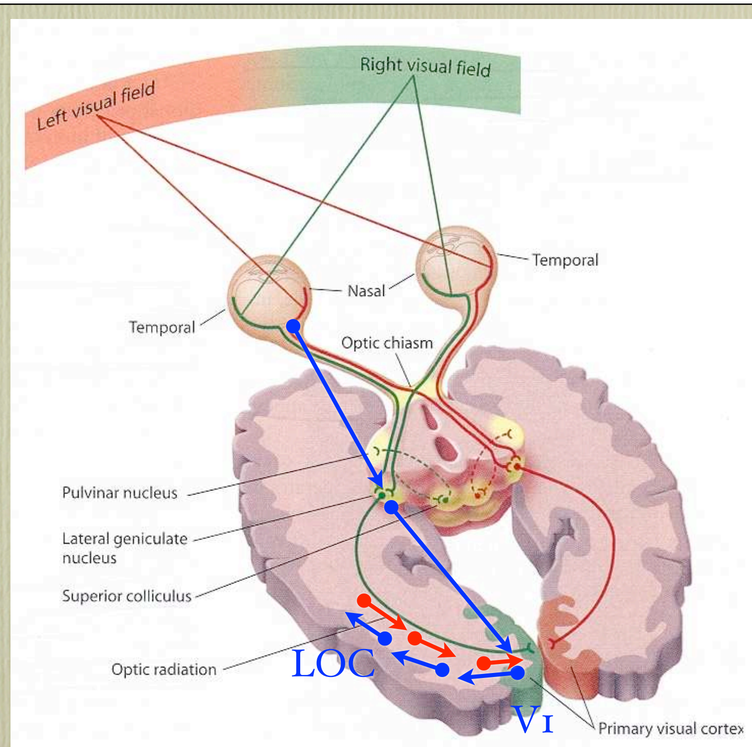
Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. Proc Natl Acad Sci U S A, 99, 15164-15169.

Structure from motion








Left visual field
Right visual field
Temporal
Nasal
Temporal
Optic chiasm
Pulvinar nucleus
Lateral geniculate nucleus
Superior colliculus
Optic radiation
LOC
V1
Primary visual cortex

Cortical Mechanism? ...some speculation

1. Feedforward: local features to objects

2. Feedback models

   a. Feedforward + attention:

      competitive selection of features

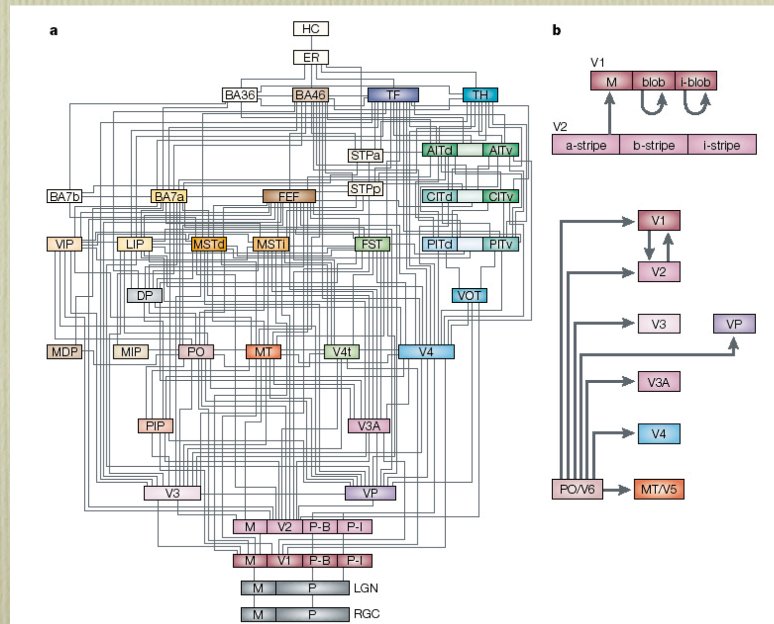
   b. Predictive coding

   c. Sparsification

Internal generative models

MacKay DM (1956) The epistemological problem for automata. In: Automata Studies (Shannon CE, McCarthy J, eds), pp 235-250. Princeton: Princeton University Press.

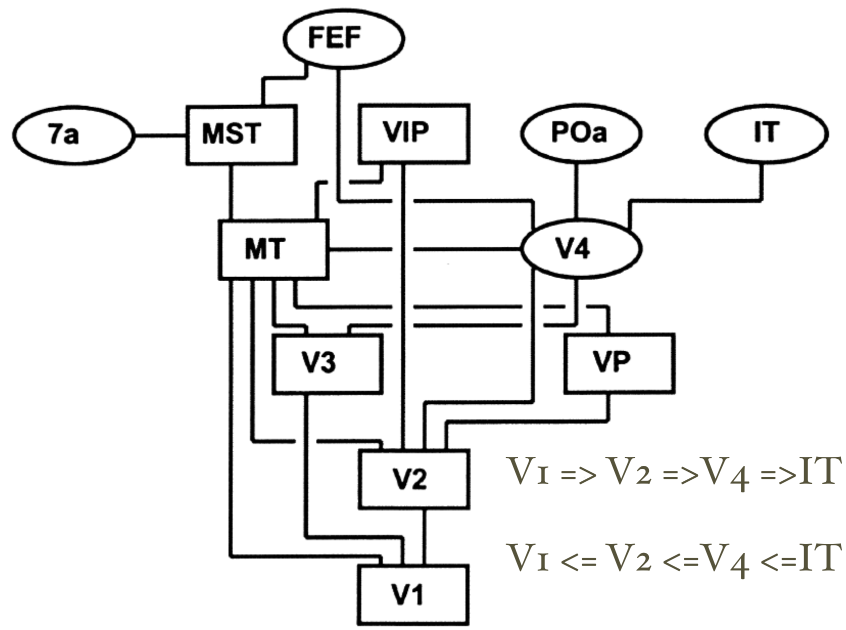
Cortical organization



Cortical organization

- Organization of visual cortices is a hierarchy
- Depends on distinct feedforward/feedback pathways
- Different laminar specificity
- More backward connections
- Backward connections more diffuse

Object recognition?



V1 => V2 =>V4 =>IT

V1 <= V2 <=V4 <=IT

Forward connections

- Sparse axonal bifurcations
- Topographically organized
- Originate in supragranular layers (I,II,III)
    - III => adjacent columns
    - II => other cortical areas
- Terminate in layer IV

Friston K (2003) Learning and inference in the brain. Neural Netw 16:1325-1352.

Feedback connections

- Lots of axonal bifurcation
- Diffuse topography
- Originate in infragranular (V, VI) layers
- Mainly terminate in supragranular layers (I,II,III)

Friston K (2003) Learning and inference in the brain. Neural Netw 16:1325-1352.

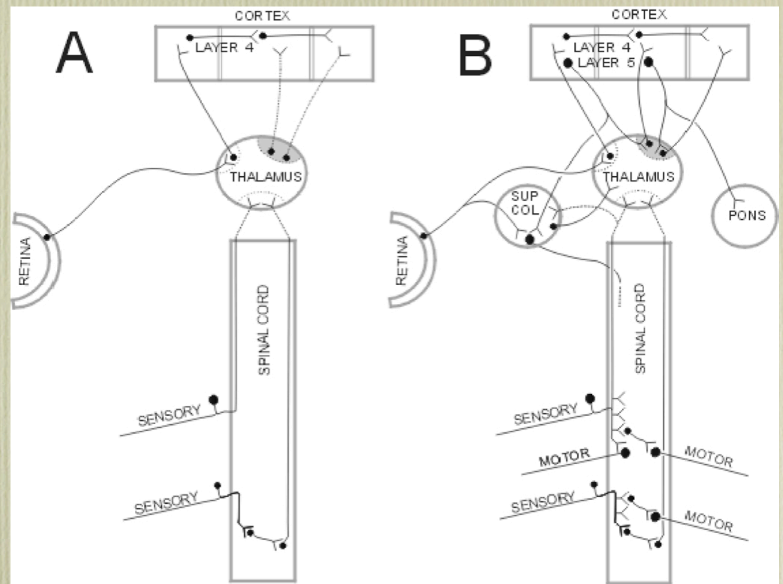


Figure courtesy of Ray Guillery

Internal generative models

Analysis-by-synthesis

Predictive coding

- High-level object models project back predictions of the incoming data

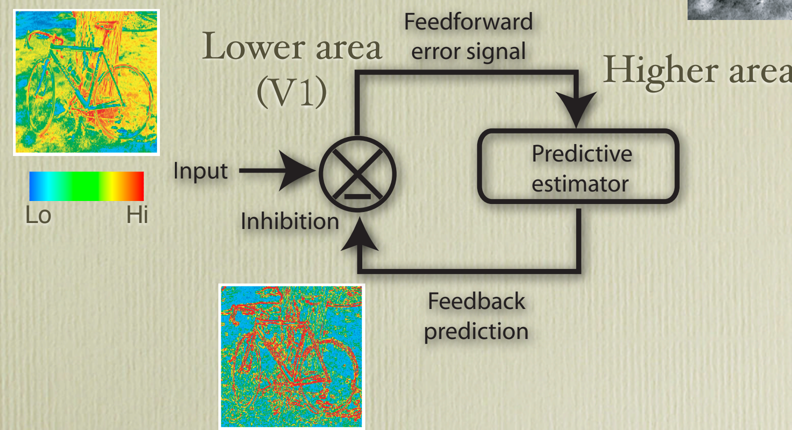
  Poor fit, high residual => high activity

Sparsification

- A good high-level fit tells earlier areas to "stop gossiping"

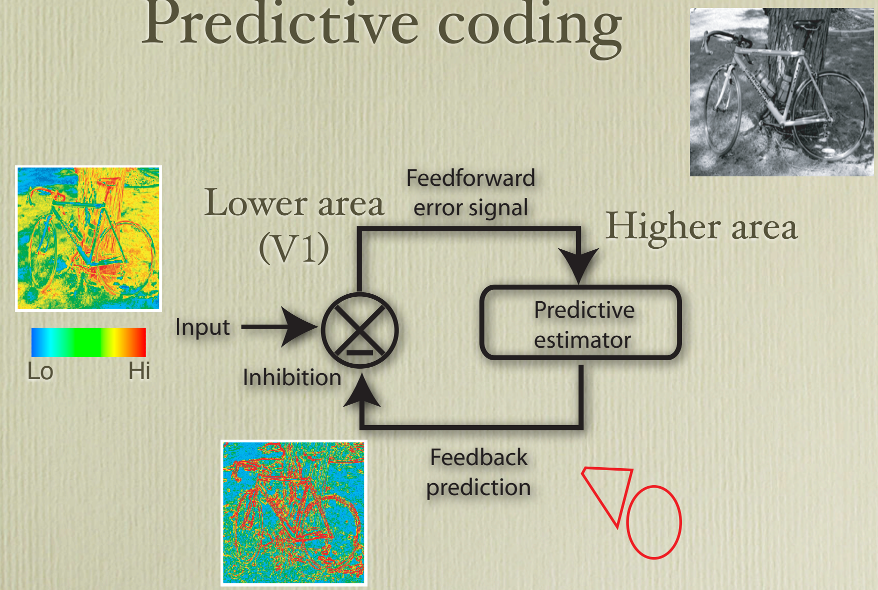
  Amplify the activity for early features that belong to object, suppress the rest

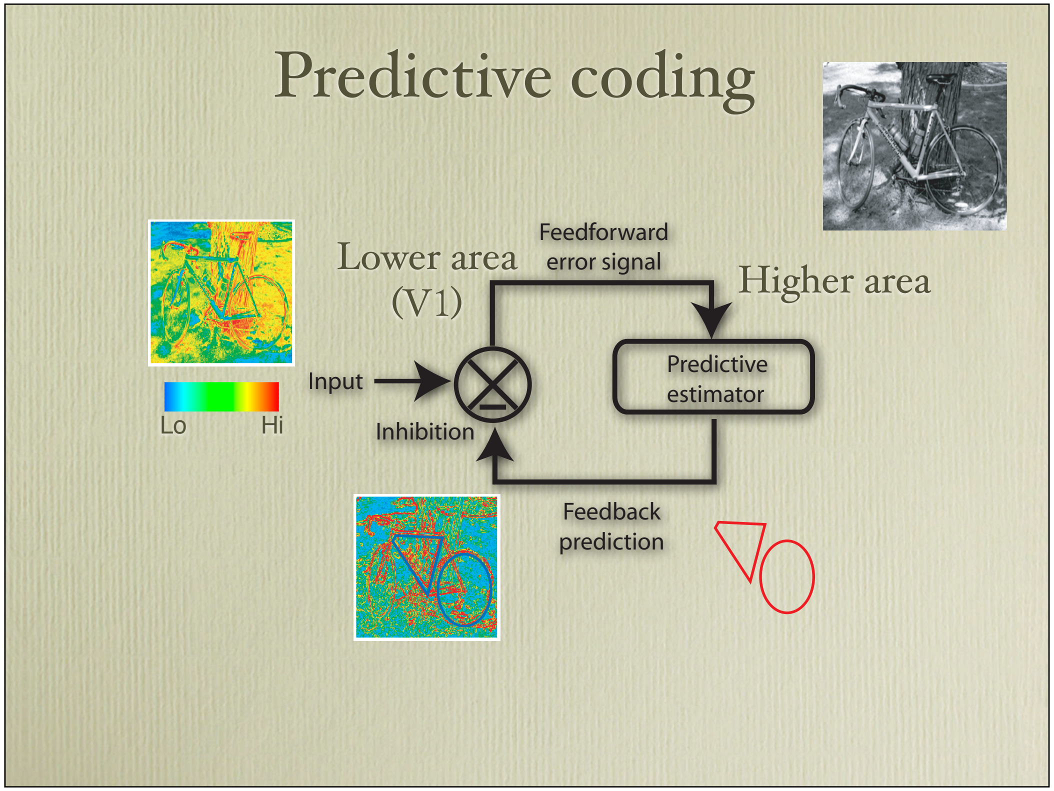
Predictive (top-down) processes in the brain?



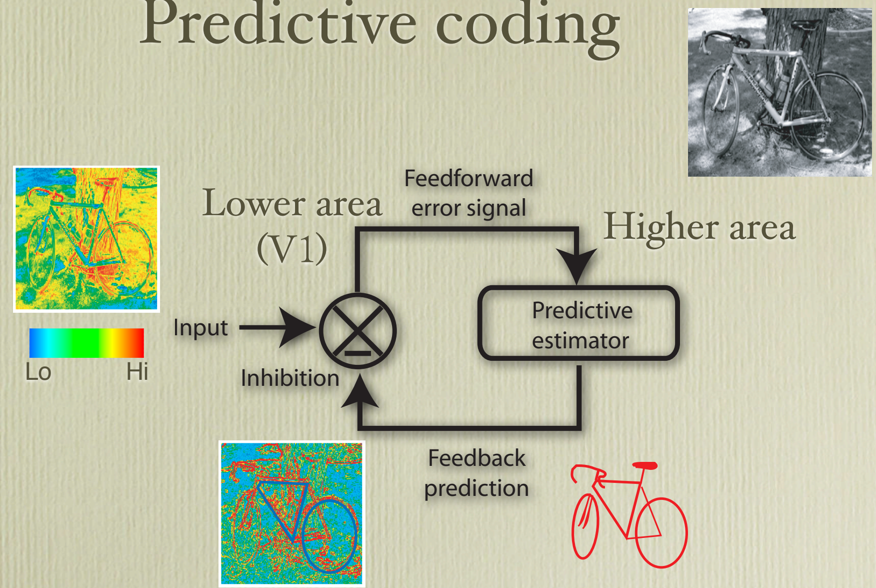
e.g. Rao, R. P., & Ballard, D. H. (1997). Dynamic model of visual recognition predicts neural response properties in the visual cortex. Neural Comput, 9(4), 721-763.

Predictive coding

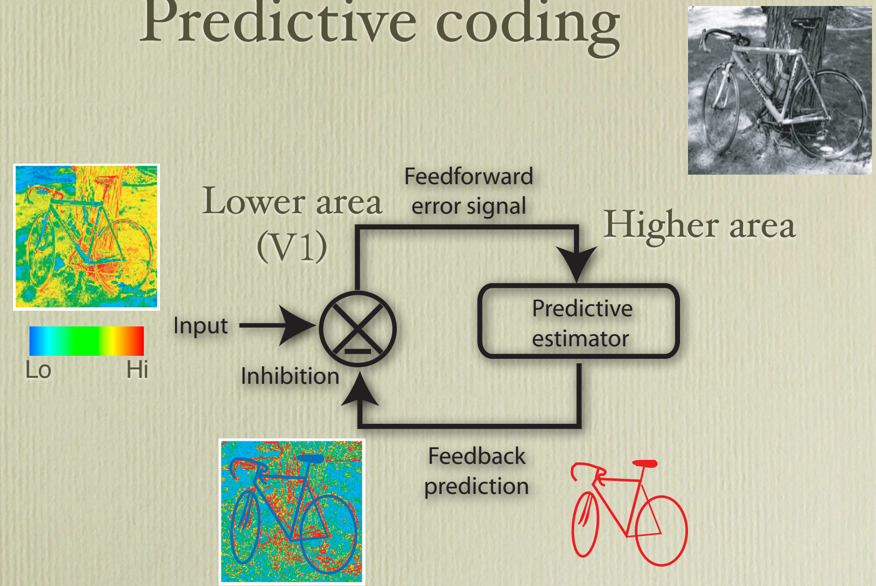



Predictive coding
Lower area
(V1)
Feedforward
error signal
Higher area
Input
Predictive
estimator
Lo
Hi
Inhibition
Feedback
prediction

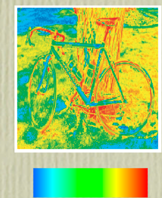
Predictive coding



Predictive coding



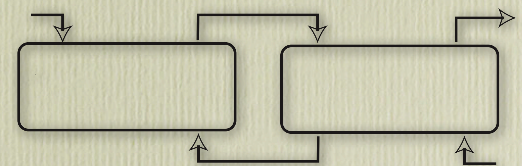
Sparsification
“Stop gossiping”

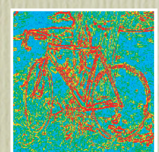


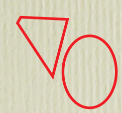
Lo Hi

Lower area
(V1)

Higher areas
(LOC)



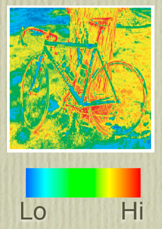


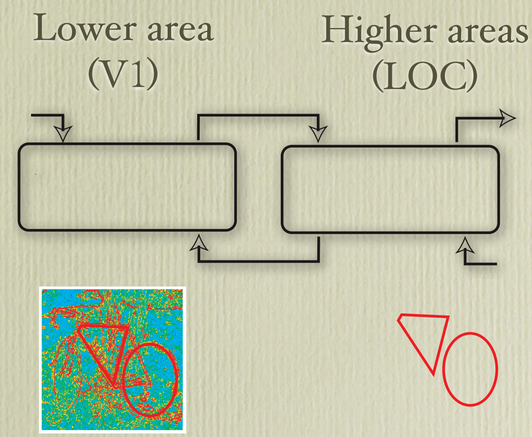


Grossberg S (1994) 3-D vision and figure-ground separation by visual cortex. Percept Psychophys 55:48-121.

Sparsification

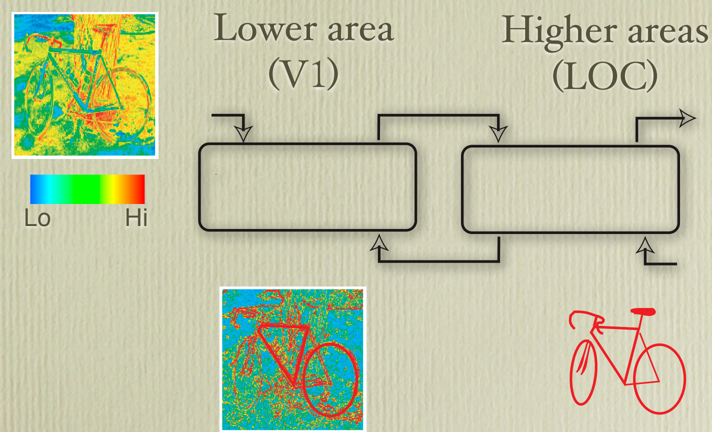
"Stop gossiping"





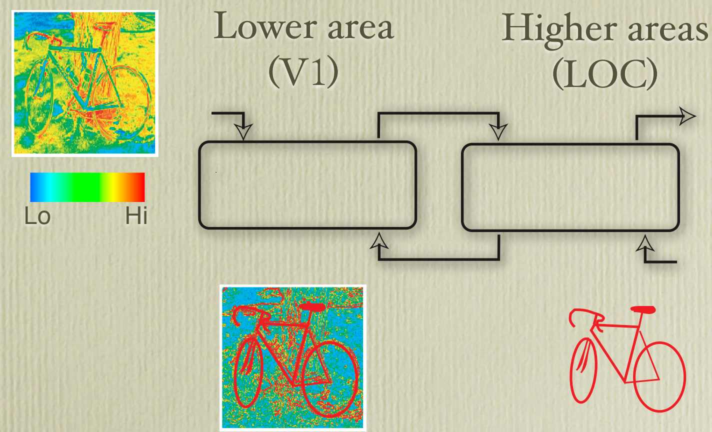
Sparsification

"Stop gossiping"



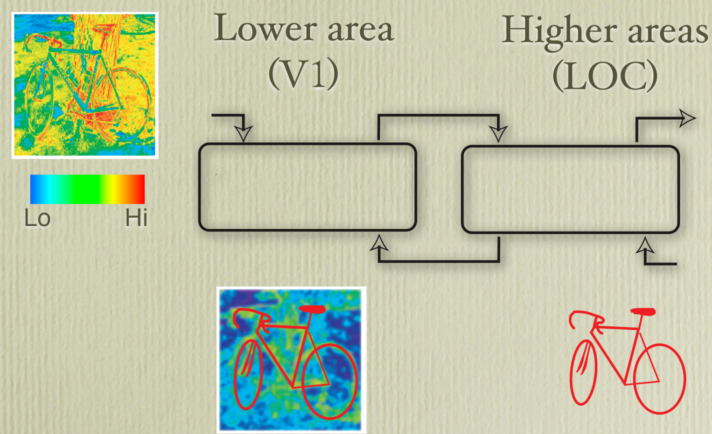
Sparsification

“Stop gossiping”

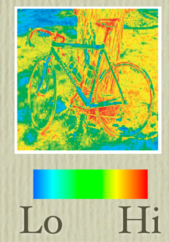


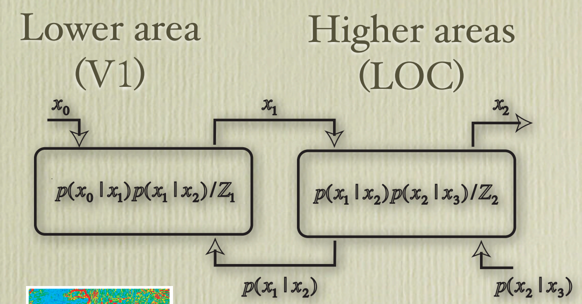
Sparsification

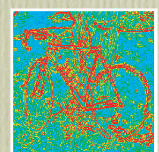
“Stop gossiping”



Bayesian Interpretation Sparsification







Lee & Mumford, 2003, JOSA

Particle filtering ideas: Isard M, Blake A (1998) Condensation -- conditional density propagation for visual tracking. International Journal of Computer Vision 29:5--28.

Summary

Common patterns of neocortex structure

- Has inspired lots of models of cortical information processing

Key target problem?

- Object perception given occlusion, clutter

fMRI and object grouping given occlusion

- consistent with feedback, but...