Human Object Perception:

Bottom-up and Top-down

Daniel Kersten
Psychology Department, University of Minnesota
July 2007

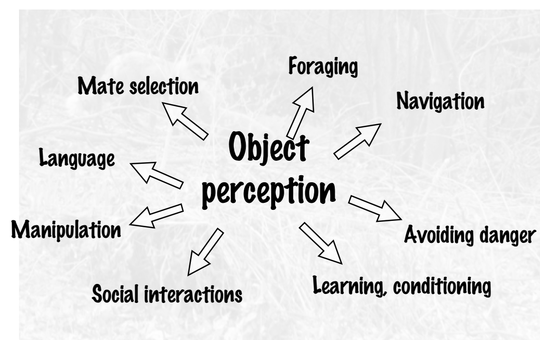
kersten.org

Outline

- Human object perception
- Visual ambiguity & Bayes
- Bottom-up/top-down
- Contextual influences on early cortical processing
  - shape
  - size
  - lightness
  - figure/ground

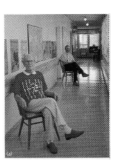
Object perception is critical for a diverse range of functions

Object perception is critical for a diverse range of functions

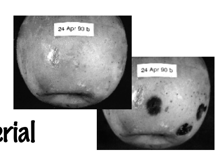


Some jobs of object perception

Size & depth



Figure/ground



Material

www.tfrl.ars.usda.gov

Shape

Outline

- Human object perception
- Visual ambiguity & Bayes
- Bottom-up/top-down
- Contextual influences on early cortical processing

Local ambiguity

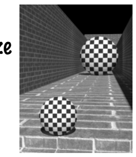
•How to get useful information about objects from ambiguous image intensities?

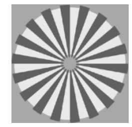
From: Mumford, 2002

Ambiguity with "clean" images too

•Making images "as simple as possible, but no simpler"

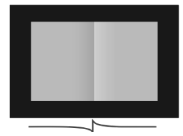
Size

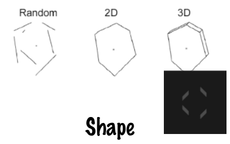




Figure/ground

Material reflectance





Shape

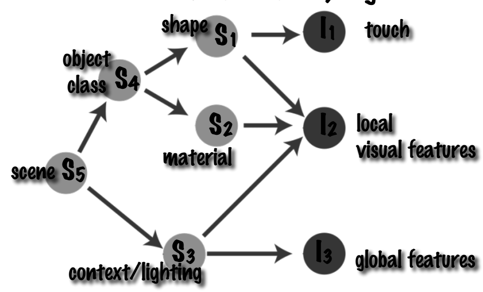
Visual ambiguity resolution from a Bayesian perspective

•Perceptual interpretation should respect how images are generated.

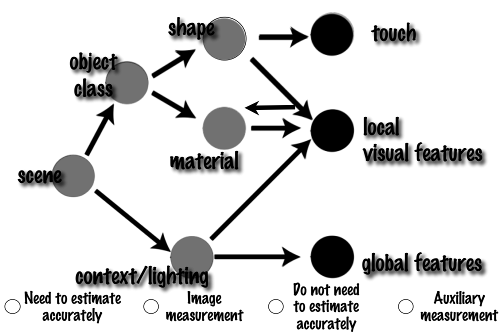
•The percept for scene or object property S should be:

- Consistent with the image I, big likelihood $p(I \mid S)$
- Probable, big prior $p(S)$
- Specifies the joint probability $p(S,I)=p(I \mid S)\,p(S)$
- ...but too complicated! $p(S_1, S_2, S_3,\ldots,I_1, I_2, I_3\ldots)$

Natural sensory input has a rich causal structure, e.g.



What to estimate depends on the task

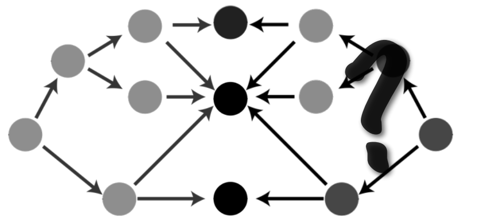


The empirical challenge

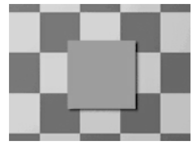
•Generative knowledge

- Test for "built-in" knowledge of causal structure in images
- Find out what human vision "cares about"
- Ideal-observer analysis
  - quantitatively compare human and Bayes-optimal performance

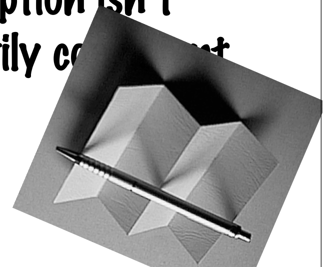
To what extent does human vision have built-in knowledge of image generation?



Perception isn't necessarily co

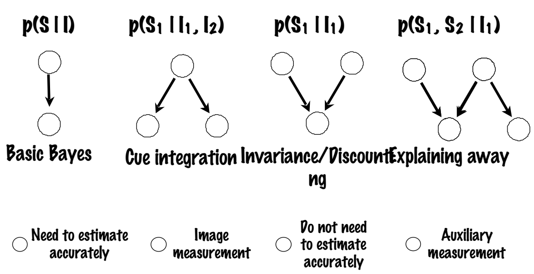


Moving cast shadow consistent with change in depth



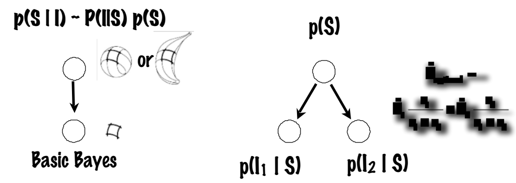
Perception isn't always consistent: Cast shadow inconsistent with perceived shape of folded card

Simple influence graphs for ideal observer analysis



Kersten, D., & Yuille, A. (2003). Bayesian models of object perception. Current Opinion in Neurobiology, 13(2), 1-9.

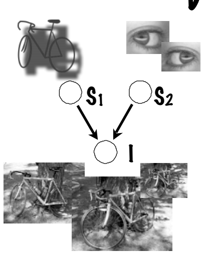
Integration of visual information



A growing number of studies have shown that human integration of visual information is often near optimal

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian Inference. Annual Review of Psychology, 55, 271-304.

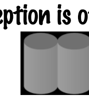
Object invariance & Discounting

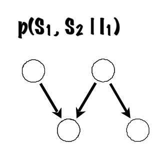


...but human object decisions are not always optimal, reflecting various suboptimal strategies for dealing with confounding variables, like viewpoint or illumination, e.g. Liu et al, 1995

$$p(S_1, I) = \sum_{S_2} p(S_1, S_2, I)$$

Perceptual "explaining away"

- Human perception is often "good at it" 
- Why?
- Evidence for "Bayesian model competition"?

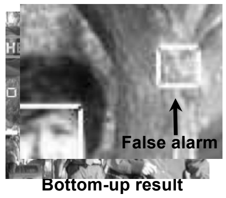


Battaglia, P. W., Schrater, P., & Kersten, D. (2005). Auxiliary object knowledge influences visually-guided interception behavior.

Image parsing & "Explaining away"

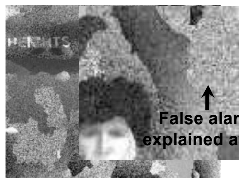
Three models: text, faces, texture

Input



False alarm

Bottom-up result



False alarm explained awa

Synthesized image

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

Outline

- Human object perception
- Visual ambiguity & Bayes
- Bottom-up/top-down
- Contextual influences on early cortical processing
  - size
  - lightness
  - figure/ground
  - shape

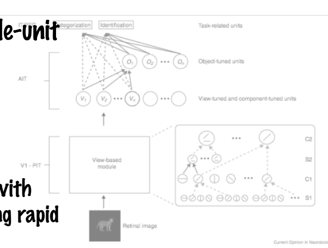
•Bottom-up

"Discriminative" models

- pixels->edges->contours->objects
  - a predominant working hypothesis in visual neuroscience and psychology
- pixels->textures->regions->objects
  - augment with cue-integration (e.g. motion, color, ...)
  - intermediate-level constraints, smoothing, (Gestalt) grouping--"lateral interactions"
- pixels->intermediate-level features->object/scene categories
  - gist
  - fragment-based approaches

Successes

- Consistent with single-unit data



Bottom-up consistent with behavioral data showing rapid responses...

Riesenhuber, M., & Poggio, T. (2002). Neural mechanisms of object recognition. *Curr Opin Neurobiol, 12* (2), 162-168.

Rousselet, G.A., Fabre-Thorpe, M., & Thorpe, S.J. (2002). Parallel processing in high-level categorization of natural images. Nat Neurosci, 5 (7), 629-630.

Problems

- Computation
  - Edge-based/local feature segmentation is hard
  - Some kinds of variation hard to discount
- Enormous versatility of human scene interpretations, and robustness
- Neural data: Doesn't account for cortical backprojections

McDermott, J. (2004). Psychophysics with junctions in real images. *Perception, 33* (9), 1101-1127.

Yuille, A.L., Coughlan, J.M., Wu, Y.N., & Zhu, S.C. (2001). Order Parameter for Detecting Target Curves in Images: How Does High Level Knowledge Help? *Int'l Journal of Computer Vision, 41* (1/2), 9-33.

Top-down Generative inference

- Historically, feedback as mechanism for task-based attention allocation
  - "Spotlight" to control flow of information, complexity
- Feedback could provide global information for resolving ambiguity locally
  - e.g. instantiate a generative model for the input, analysis-by-synthesis

Mackay, Grossberg, Mumford, Hinton, Dayan, Friston, Lee, ...

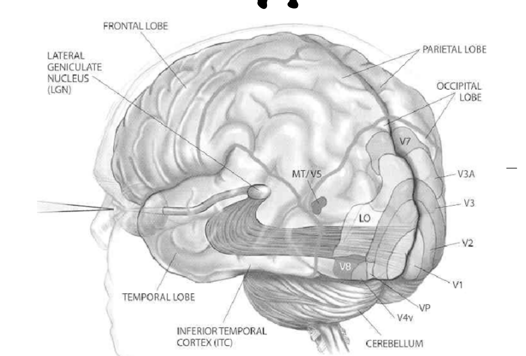
Outline

- Human object perception
- Visual ambiguity & Bayes
- Bottom-up/top-down
- Contextual influences on early cortical processing

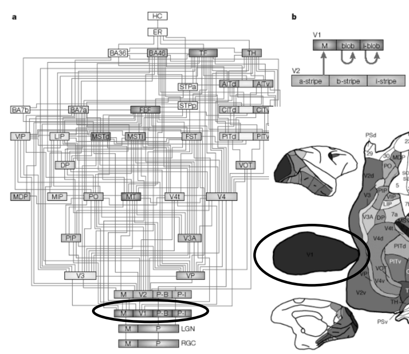
Contextual influences on early cortical processing

- Bird's eye view of visual cortex

V1

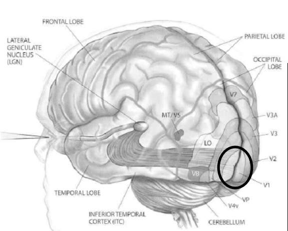


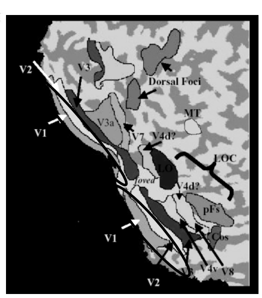
Monkey V1



Felleman & Van Essen, 1991

Human V1





Grill-Spector, K., Kourtzi, Z., & Kanwisher, N. (2001). Vision Res, 41(10-11), 1409-1422.

Primary Visual Cortex: V1

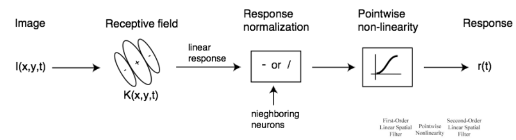
- Lots of sophisticated theory…but most working models begin with simple assumptions:
  - 2D spatially organized local, (almost) linear filters
- Starting point for bottom-up, discriminative models of recognition

V1: "Standard model"

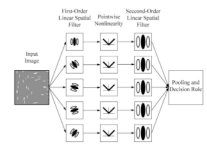
- Banks of localized spatio-temporal filters or receptive fields (i.e. the edge and bar detectors of Hubel & Wiesel; the "gabor filters" of later years)
- Feedforward
  - with perhaps local feedback for normalization/tuning

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. Nat Neurosci, 4(8), 819-825.

Local filters: Spatial receptive fields

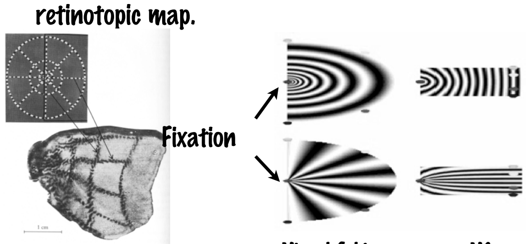


Applications: "back pocket models" for texture segmentation (Chubb & Landy; Landy & Graham)



V1: Spatial organization

•A topographic mapping that transforms retinal coordinates to V1 cortical coordinates--a retinotopic map.



Tootell et al., 1982 Visual field V1

V1

- "What is the other 85% of V1 doing"?
  - Olshausen, B. A., & Field, D. J. (2005). Neural Comput, 17(8), 1665-1699.
- Feedforward processing by local, oriented filters would imply little or no effect of global structure, but we know...
  - within area connections
  - between area connections
  - figure-ground modulation of neural responses

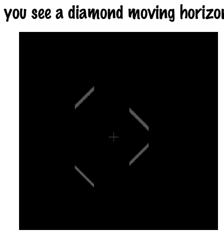
For reviews, see:
Bullier, J. (2001). Integrated model of visual processing. Brain Res Brain Res Rev, 36(2-3), 96-107.
Friston, K. (2005). A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci, 360(1456), 815-836.

Contextual influences on early cortical processing

•V1 & spatial representation and size
•Early cortical response to lightness
•Figure/ground
•Shape

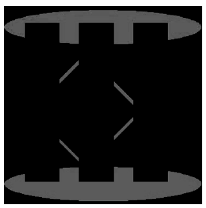
Perceptual organization

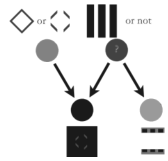
Do you see a diamond moving horizontally?



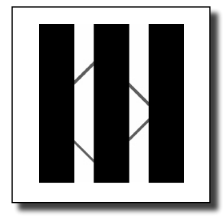
Lorenceau, J., & Shiffrar, M. (1992). The influence of terminators on motion integration across space. Vision Res, 32(2), 263-273.

Auxiliary evidence for occlusion





Yet stronger evidence for occlusion



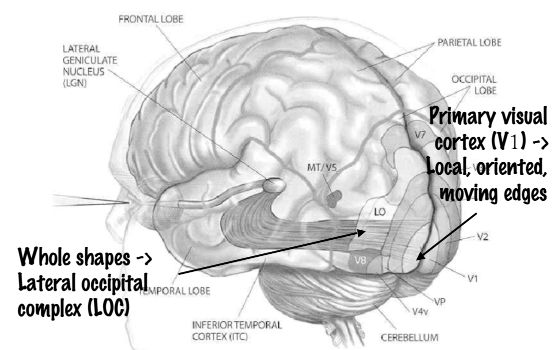
Example of: Perceptual "explaining away"

V1 activity & perceptual organization

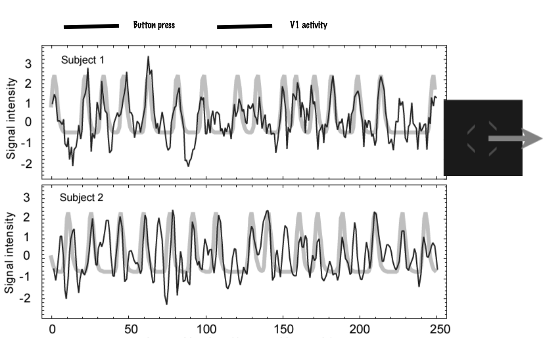
•Use BOLD functional MRI to localize cortical activity that is correlated with the competing perceptual hypotheses of

- Coherent diamond vs. less coherent line fragments

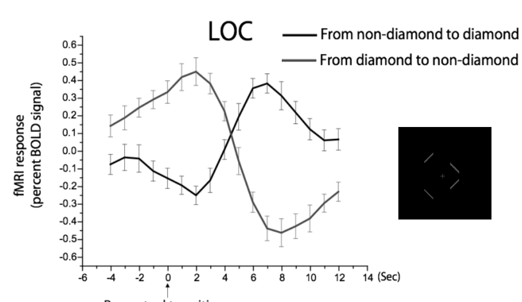
Scott Murray, Dan Kersten, Bruno Olshausen, Paul Schrater and David Woods 2002. Proc Natl Acad Sci U S A, 99, 15164-15169.



Perceptual organization correlates with reduced V1 activity

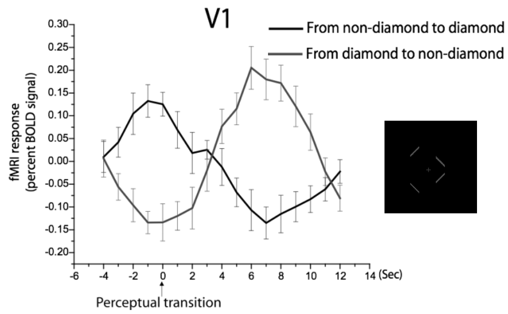


Perceptual organization is correlated with increased LOC activity

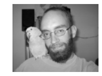


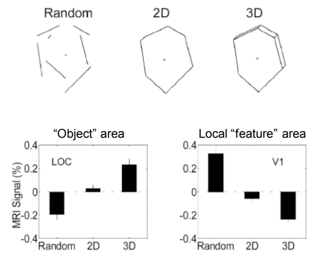
From: Fang, Murray, He & Kersten, 2004, International Congress of Psychology, Beijing

Perceptual organization is correlated with decreased V1 activity



Shape perception can reduce V1 activity





Consistent with top-down theories, but...

1) But more than 1 theory
2) We don't fully understand the perceptual conditions that produce this
3) Representational assumptions

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.

Internal generative models
Analysis-by-synthesis

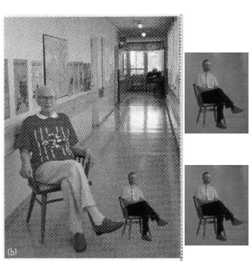
- Predictive coding
  - High-level object models project back predictions of the incoming data
    - Poor fit, high residual => high activity
    - Good fit => low activity ("shut up")
- Sparsification
  - A good high-level fit tells earlier areas to "stop gossiping"
    - Amplify the activity for early features that belong to object, suppress the rest

---

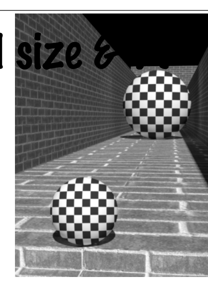
Contextual influences on early cortical processing

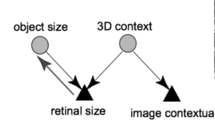
- V1 & spatial representation and size
- Early cortical response to lightness
- Figure/ground
- Shape

Size & depth



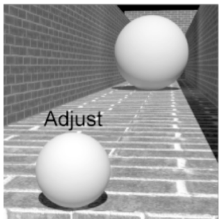
Perceived size &

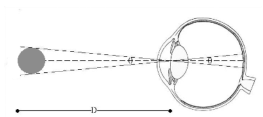




Murray, S. O., Boyaci, H., & Kersten, D. (2006). The representation of perceived angular size in human primary visual cortex. Nat Neurosci, 9(3), 429-434.

Perceived angular size

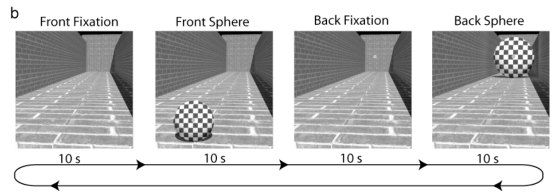




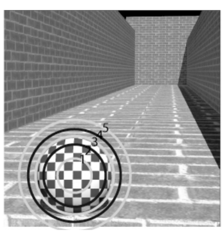
- Perceptual effect: ~17%

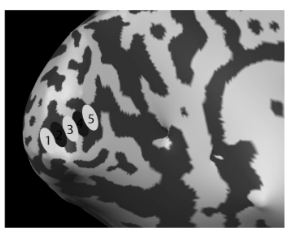
http://www.tc.umn.edu/~boyac003/Vision/SizeAppletLarge.html

fMRI BOLD response

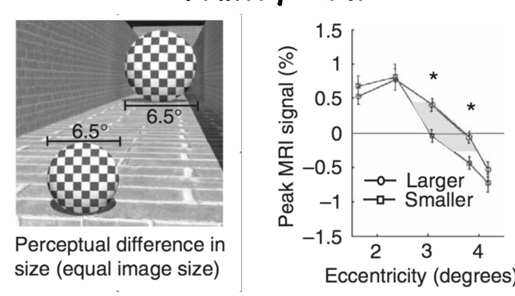


Localizing spatial extent



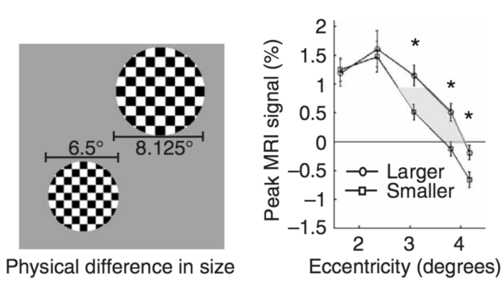


"3D" peak response curves: "Hallway" data



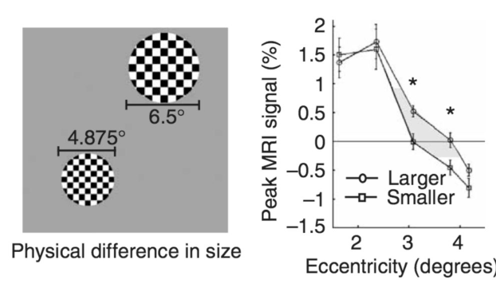
- fMRI effect: ~20%

"2D" peak response curves



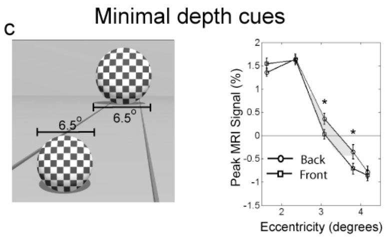
- fMRI effect: ~20%

"2D" peak response curves



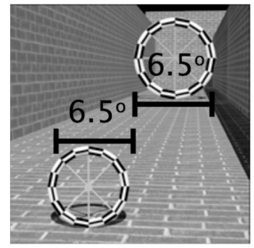
- fMRI effect: ~15%

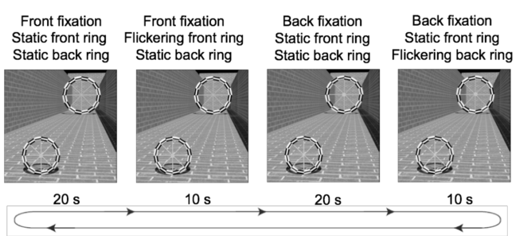
Reduced perceptual effect
(Ponzo)



Perceived angular size and spatial extent of fMRI activation both reduced

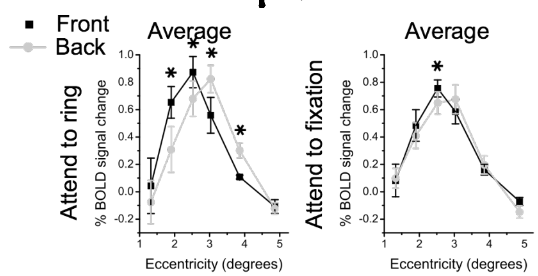
Line response & attention





Localizer rings

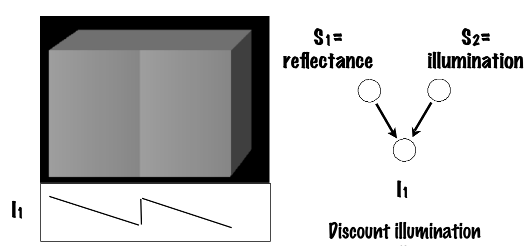
Shift in cortical line response



Contextual influences on early cortical processing

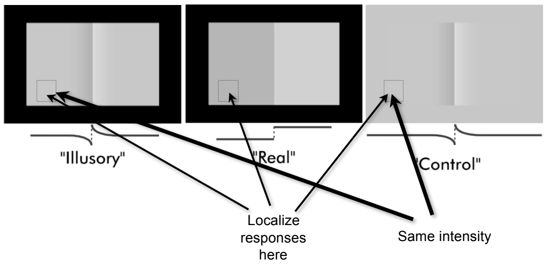
- V1 & spatial representation and size
- Early cortical response to lightness
- Figure/ground
- Shape

Lightness constancy as reflectance estimation

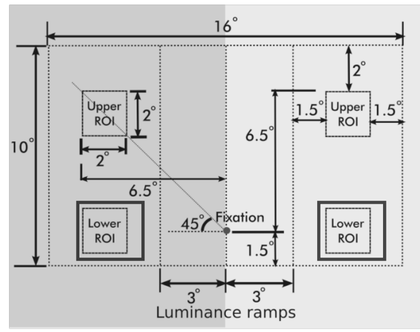


Discount illumination gradients

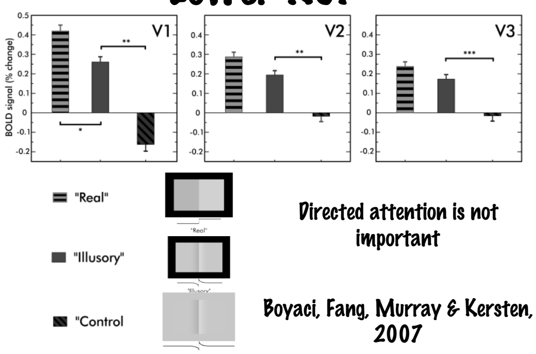
Compare 3 conditions



Some details



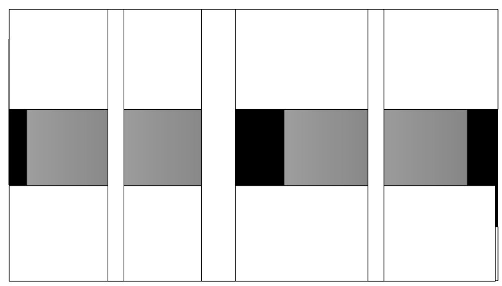
Lower ROI



Directed attention is not important

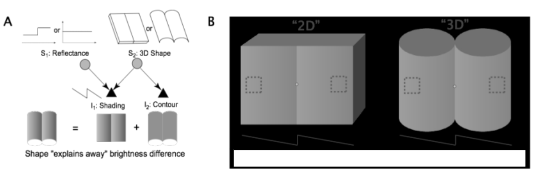
Boyaci, Fang, Murray & Kersten, 2007

Lightness & "explaining away"



Knill & Kersten (1991)

How will shape modulate early cortical responses?



I don't know yet...

Computational challenge: Quantitative testable theory of shape/material in explaining away

Summary of contextual influences on early cortical areas

- Grouping features into shapes can reduce activity in V1
- Depth modulates 2D topographic processing in V1
- V1 responds to lightness change in the absence of local intensity change
- fMRI evidence for distinct neural populations for edge assignment in V2

thanks...

Theory: Paul Schrater, Pascal Mamassian, Peter Battaglia & Alan Yuille

fMRI: Huseyin Boyaci, Fang Fang, Sheng He, Jay Hegdé, Scott Murray, Cheryl Olman,

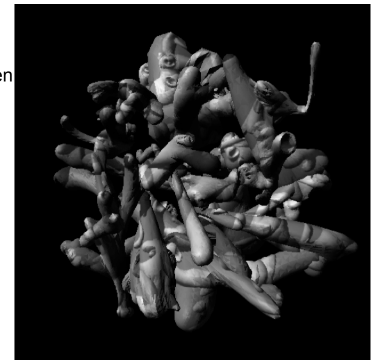
CMRR: Kamil Ugurbil, Liming Shen, Xiaoping Hu, Essa Yacoub

...and the past mentors & members

...and of course, NIH, NSF & ONR

NIH R01 EY015261, NIH P41 RR08079, NSF SBR-9631682, N00014-05-1-0124

Camouflage
challengedemos.kersten.org



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