how far can bayesian theories of vision take us?

Daniel Kersten

Psychology Department, U. Minnesota

Theory of Neural Computation
Mathematical Sciences Research Institute

Berkeley, October 2015

kersten.org

1

It takes just one quick glance to see the fox, a tree trunk, some grass and background twigs.

but the longer we look the more we see…

2

"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."

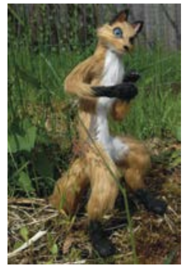
3

two computational problems

Ambiguity: To be sure about any small piece, the visual system has to understand the larger context

4

two computational problems

Versatility: To make an unlimited variety of inferences, to generalize, the visual system needs to represent and access information across multiple scales, feature types and transformations

Inferences about the fox picture involved various:

- levels of abstraction
- spatial scales
- feature types (shape, material)
- relationships between parts, objects, and viewer

A strong “bayesian” assumption is that reliable and versatile visual inferences are based on structured generative, probabilistic knowledge of how virtually any natural image could be produced

…but doesn’t specify what the generative factors are, how they should be used, structured in the brain, or the mechanisms that underly their inferences

working hypothesis

Hierarchical computations within and between visual cortical areas reflects

- the rich, probabilistic, generative structure of image input,

constrained by

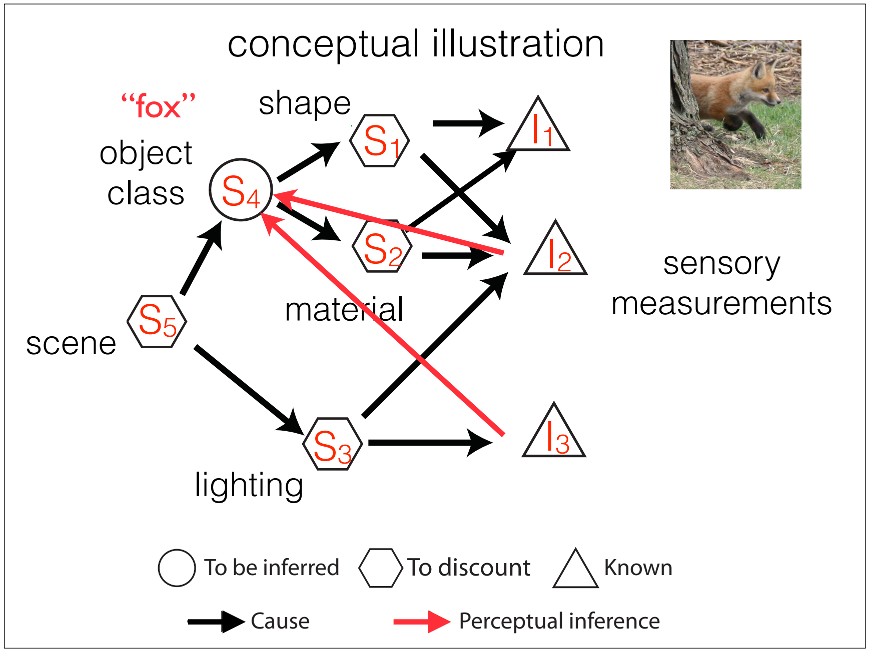
- the generative factors important for behavioral outcomes (hardwired or dynamic)

the basics

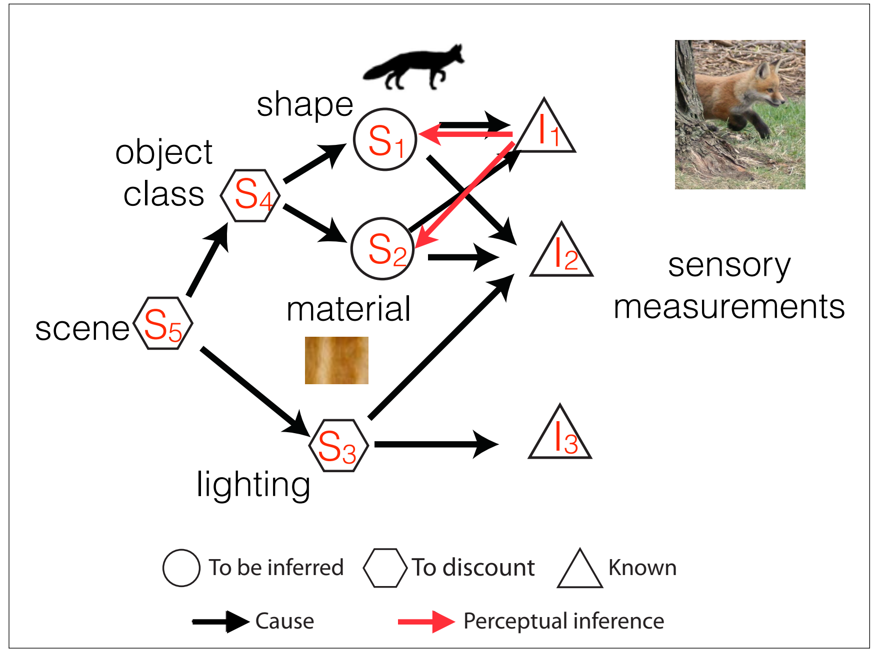
knowledge of the relationships between generative factors, $S = (S_1, S_2,\ldots)$ and image patterns $I = (I_1, I_2\ldots)$ are represented probabilistically:

joint	$p(S_1, S_2,\ldots; I_1, I_2\ldots)$
posterior	$p(S_1, S_2,\ldots \mid I_1, I_2\ldots)$
$\propto$ likelihood x prior	$p(I_1, I_2\ldots \mid S_1, S_2,\ldots) \times p(S_1, S_2,\ldots)$

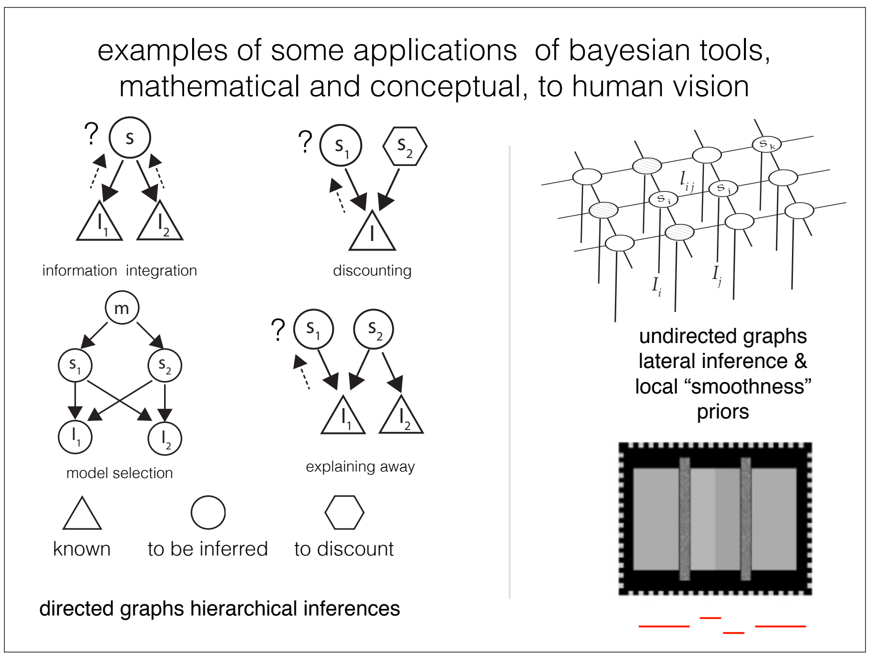
- conditional dependencies structure complex distributions
- the task determines which variables to discount and thus sum over, and the image measurements which variables to fix, and thus condition the posterior
- factoring the posterior into likelihood and prior makes the generative knowledge explicit
- decisions are based on operations over the resulting “simplified” posterior


conceptual illustration
"fox"
object class
shape
material
scene
lighting
S1
S2
S3
S4
S5
I1
I2
I3
sensory measurements
To be inferred
To discount
Known
Cause
Perceptual inference

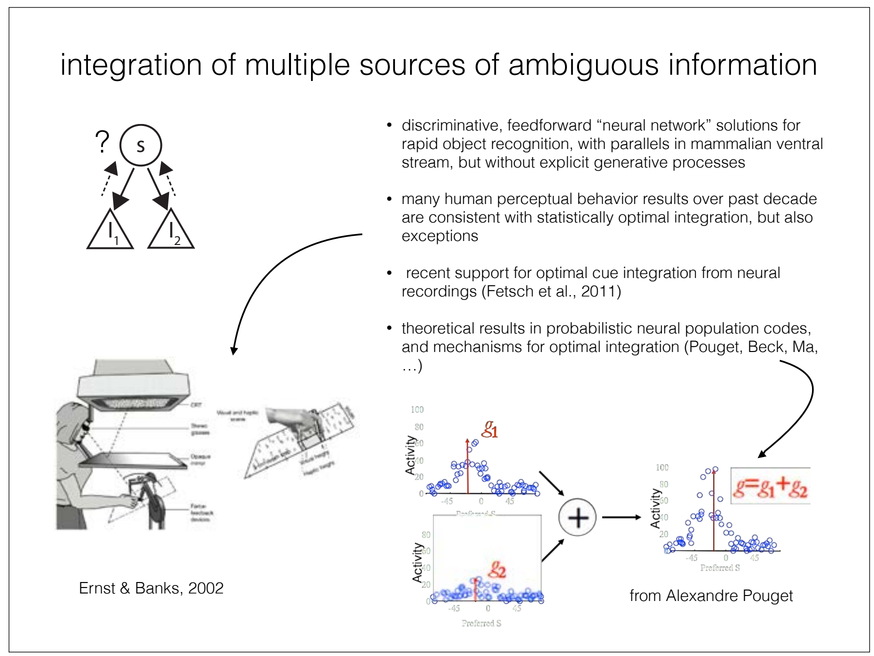
9


shape
object class
material
scene
lighting
S1
S2
S3
S4
S5
I1
I2
I3
sensory measurements
To be inferred
To discount
Known
Cause
Perceptual inference

10

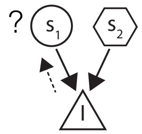

examples of some applications of bayesian tools, mathematical and conceptual, to human vision
information integration
discounting
model selection
explaining away
known
to be inferred
to discount
directed graphs hierarchical inferences
undirected graphs lateral inference & local "smoothness" priors

11


integration of multiple sources of ambiguous information
• discriminative, feedforward "neural network" solutions for rapid object recognition, with parallels in mammalian ventral stream, but without explicit generative processes
• many human perceptual behavior results over past decade are consistent with statistically optimal integration, but also exceptions
• recent support for optimal cue integration from neural recordings (Fetsch et al., 2011)
• theoretical results in probabilistic neural population codes, and mechanisms for optimal integration (Pouget, Beck, Ma, ...)
Ernst & Banks, 2002
Activity
g=g1+g2
from Alexandre Pouget

12

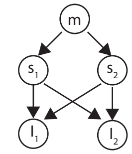
integrating out unwanted information



- core problem of "object constancy", recognition, …
  - implicit in training of feedforward "neural network" solutions for object recognition, e.g. discounting variations in appearance.
- long history in ideal observer analysis of human vision, with applications, e.g. human color constancy
- theoretical results in active marginalization using probabilistic neural population codes (Beck et al., 2011)

13

model-dependent human parameter estimation

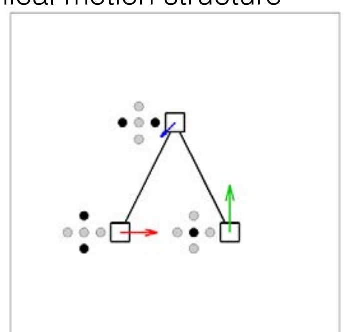


slant of the scree field?

- human estimation of surface slant from texture—model averaging of isotropic and homogeneous texture models (Knill, 2003)
- vision/auditory localization of sound — model selection (Kording et al., 2007)
- conditioned perception. (Stocker & Simoncelli, 2008)
- human velocity estimation depends on the optic flow category. Wu, S., Lu, H., & Yuille, A. (2008)

14

flexible summaries of hierarchical motion structure

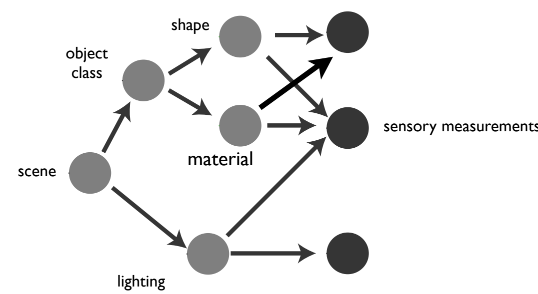


https://sites.google.com/site/hierarchicalmotionperception/home

Gershman, S. J., Tenenbaum, J. B., & Jakel, F. (2015). Discovering hierarchical motion structure. Vision Research, 1–10. http://doi.org/10.1016/j.visres.2015.03.004

15

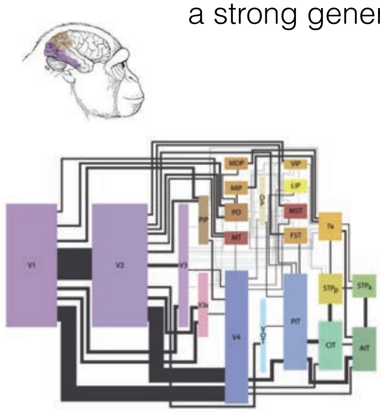
so far these are applications of bayesian concepts/tools to model perceptual behavior



Can the black arrows just be used to represent the confounding variables for the problem to be solved? Or are human visual inferences based on feedback mechanisms that operate on internal generative models of the world?

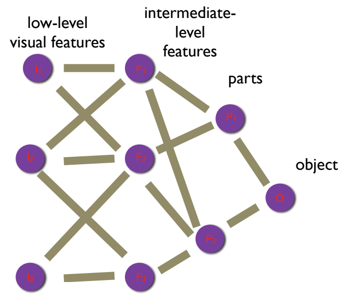
16

a strong generative hypothesis



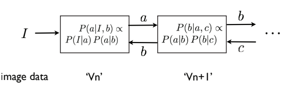
feedforward

feedback



computational architectures:
probabilistic models on graphs

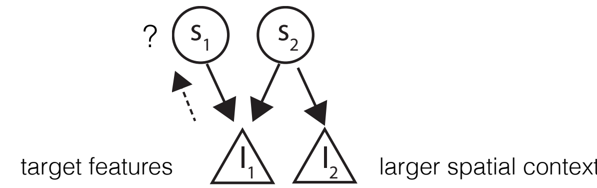
the "executive metaphor"—Alan Yuille



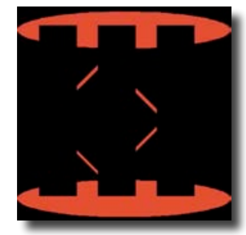
Wallisch, P., & Movshon, J. A. (2008). Structure and Function Come Unglued in the Visual Cortex. *Neuron*, *60*(2), 194–197.

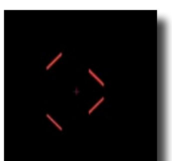
Lee & Mumford, 2003

empirical evidence for inference through generative cortical mechanisms via feedback?



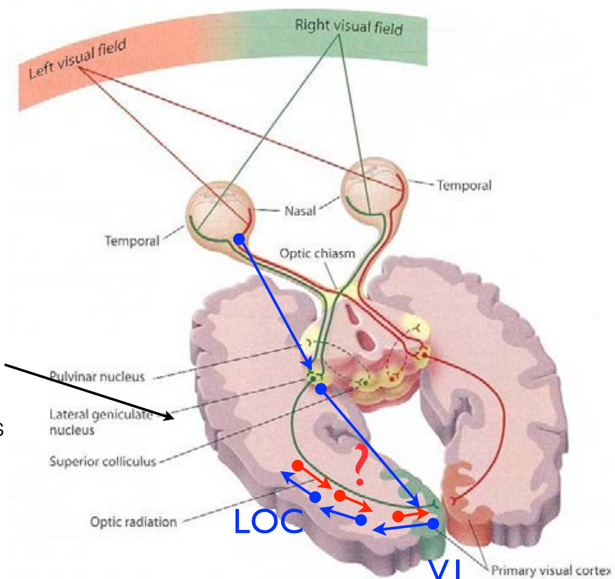
"perceptual explaining away"
…lots of perceptual examples



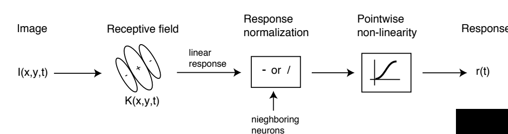


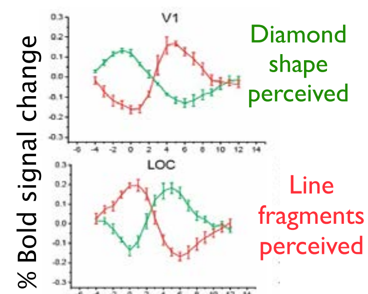
V1 sensitive to local, oriented edges/bars

higher-level visual cortical regions, such as LOC, respond to changes in perceived whole shape



Murray, Kersten, Olshausen, Schrater, & Woods (2002)



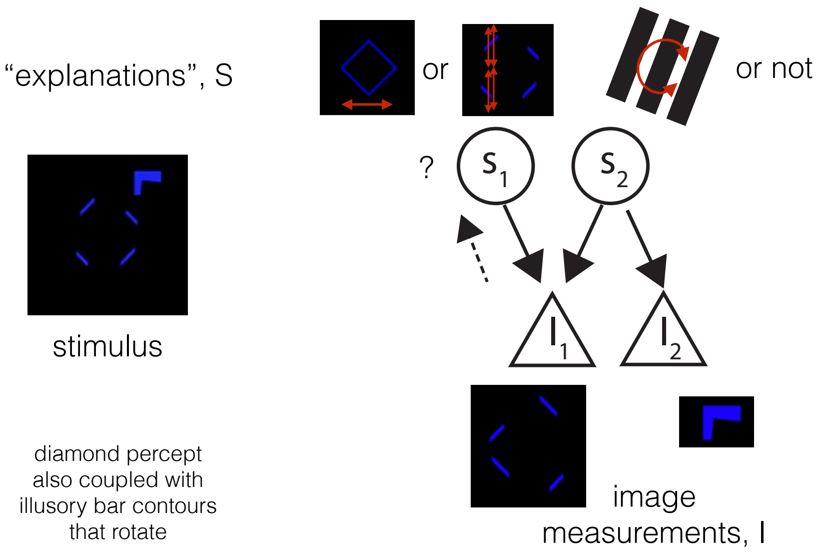


Fang, Boyaci, Kersten, Murray (2008)

Murray, Kersten, Olshausen, Schrater, & Woods (2002)

V1 activity is suppressed when the diamond shape is perceived

Mumford; Rao & Ballard



diamond percept
also coupled with
illusory bar contours
that rotate

21

…but is modulation spatially localized to voxels in V1 that correspond retinotopically to the target features?

.. some fMRI results suggest not (cf. Wit et al., 2012)

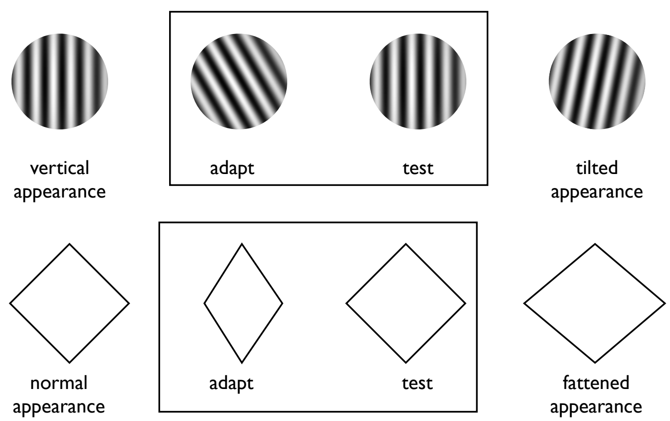
22

psychophysical test of modulation?

use adaptation--psychophysicist’s “electrode”

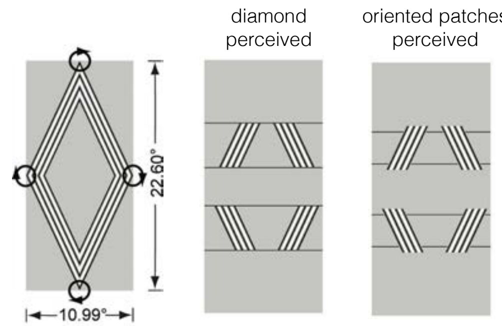
assumption: adapts neurons in early cortical areas, V1

assumption: adapts high-level cortical areas



23

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.

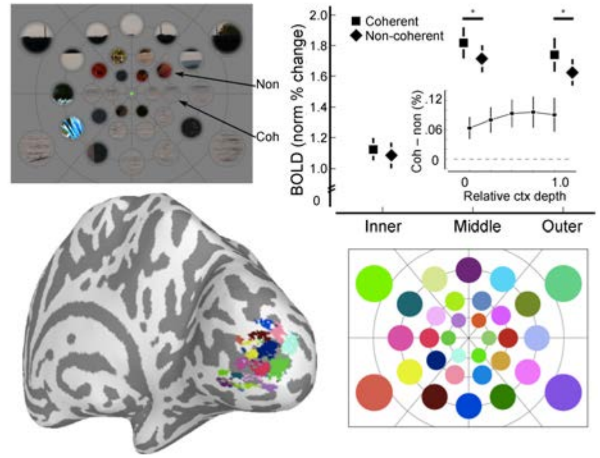
He, D., Kersten, D., & Fang, F. (2012). Opposite modulation of high- and low-level visual aftereffects by perceptual grouping. Current Biology, 22(11), 1040–1045.

24

…but we haven't always found localized suppression when local patches "fit" the larger context

25

some patches are consistent with scene (Coh) and some not (Non)



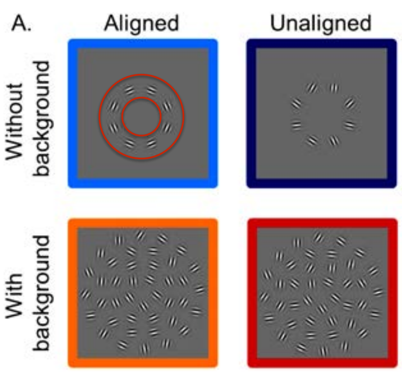
~1mm fMRI in V1

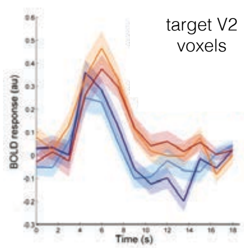
Mannion, Kersten & Olman

26

perhaps context-dependent suppression of V1 voxel activity depends the complexity of the parsing/ segmentation problem?

~2mm fMRI in V1



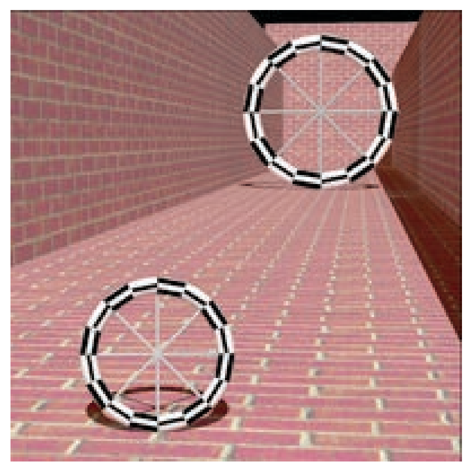


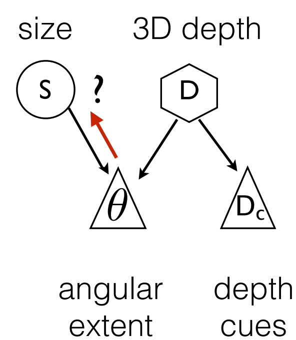
With background clutter, there was evidence of of increased V1-V2 correlations when perceiving aligned versus when perceiving unaligned contours.

Responses in early visual areas to contour integration are context dependent. Cheng Qiu, Philip Burton, Daniel Kersten, Cheryl A. Olman

27

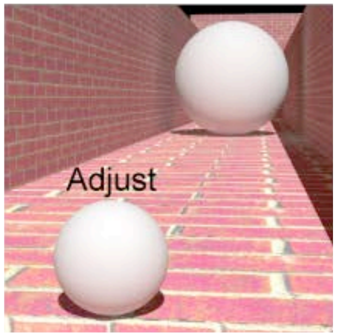
inferring the size of an object

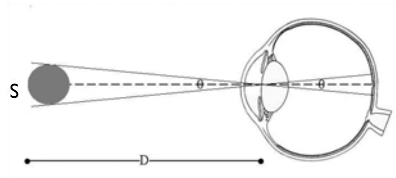




28

perceptual estimation of the size of an object





$$\theta \approx S/D$$

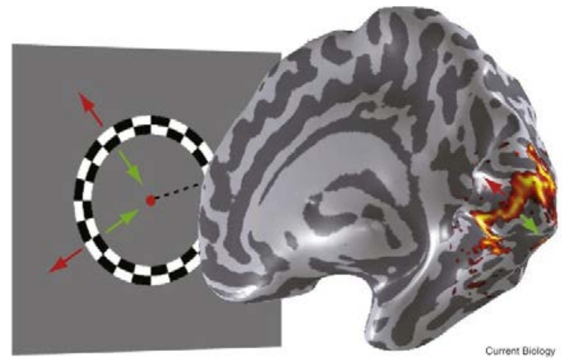
Perceptual effect: ~17%

http://vision.psych.umn.edu/users/boyaci/Vision/SizeAppletLarge.html

29

does 3D context modulate the size of the "neural image" in human V1?

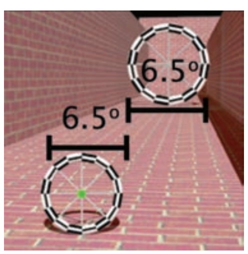
V1 has a retinotopic map, so for an actual increase in ring size in the image, we expect:

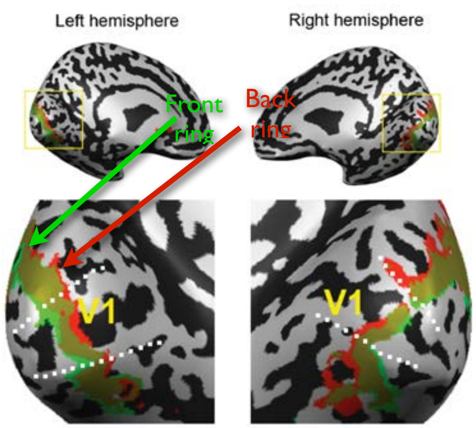


Huk, A. C. (2008) Visual Neuroscience: Retinotopy meets Percept-otopy, Current Biology, 18, 21, R1005-1007.

30

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

attend-to-ring condition

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

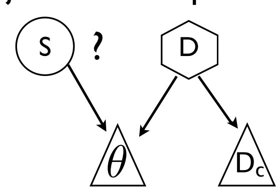
31

in terms of inference, what might be going on?

two possible representational assumptions: physical or angular size?

$$\theta \approx S/D$$

object size **depth**

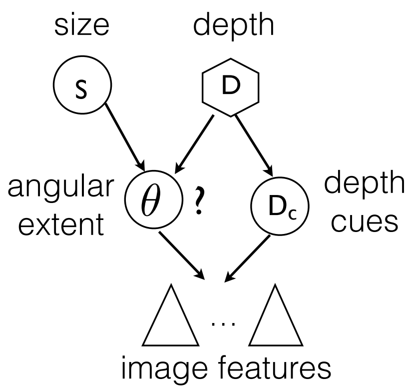


angular extent **depth cues**

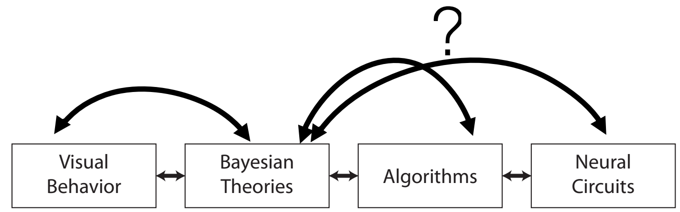
Does the shift of spatial extent in V1 represent the neural representation of an estimate of physical size (S) or a bias in the estimate of angular size (θ)?

32

estimating angular size is also a non-trivial inference



how far can bayesian theories of vision take us?



Bayesian theories of human visual behavior are here to stay

Progress closely tied to progress in computer vision,

But augmented by physiological constraints, attentional mechanisms, …

The jury is out on whether and how strong, hierarchical generative models of scene structure (the "executive metaphor") may be instantiated in neural brain architecture

…but the robustness and versatility of human vision suggest such processes

Bayes provides conceptual tools for managing uncertainty given specific task requirements at an abstract level...but we need more.

In particular, a better understanding of human-oriented generative models, compositional structure, and the algorithms/control structures for accessing information for a enormously diverse range of tasks

To explain how the longer we look, the more we see

Thanks to my collaborators

Huseyin Boyaci, Bilkent U
Fang Fang, Peking U
Damien Mannion, U of New South Wales
Scott Murray, U Washington
Cheryl Olman, U Minnesota
Cheng Qiu, U Minnesota
Alan Yuille, UCLA

Supported by the WCU (World Class University) program of the Ministry of Education, Science and Technology through the National Research Foundation of Korea (R31-10008), NIH R01 EY015261 and ONR N000141210883.