NOTETAKER CHECKLIST FORM

(Complete one for each talk.)

Name: Xin Wang  Email/Phone: xinw@salk.edu
Speaker’s Name: Daniel Kersten
Talk Title: How Far can Bayesian Theory of Vision Take Us
Date: 10 / 06 / 15  Time: 10:45 am/ pm (circle one)

List 6-12 key words for the talk: Bayesian; vision; ambiguity; versatility; generative; strong Bayesian assumption; hierarchy; binding

Please summarize the lecture in 5 or fewer sentences: There has been considerable progress in the application of Bayesian concepts and tools to the analysis and modeling of human visual behavior. One key idea is that human vision exploits built-in generative knowledge of image structure. Computational advances for doing bidirectional inference over complex, structured distributions suggest the possibility that related computations may be reflected in the neural architecture of the visual brain. In particular, this talk presents neuroimaging and behavioral evidence that feedback from higher to lower cortical areas reflects the application of generative knowledge.

CHECK LIST

(This is NOT optional, we will not pay for incomplete forms)

☐ Introduce yourself to the speaker prior to the talk. Tell them that you will be the note taker, and that you will need to make copies of their notes and materials, if any.

☐ Obtain ALL presentation materials from speaker. This can be done before the talk is to begin or after the talk; please make arrangements with the speaker as to when you can do this. You may scan and send materials as a .pdf to yourself using the scanner on the 3rd floor.
  - **Computer Presentations**: Obtain a copy of their presentation
  - **Overhead**: Obtain a copy or use the originals and scan them
  - **Blackboard**: Take blackboard notes in black or blue **PEN**. We will **NOT** accept notes in pencil or in colored ink other than black or blue.
  - **Handouts**: Obtain copies of and scan all handouts

☑ For each talk, all materials must be saved in a single .pdf and named according to the naming convention on the “Materials Received” check list. To do this, compile all materials for a specific talk into one stack with this completed sheet on top and insert face up into the tray on the top of the scanner. Proceed to scan and email the file to yourself. Do this for the materials from each talk.

☐ When you have emailed all files to yourself, please save and re-name each file according to the naming convention listed below the talk title on the “Materials Received” check list.
  (YYYY.MM.DD.TIME.SpeakerLastName)

☐ Email the re-named files to notes@msri.org with the workshop name and your name in the subject line.
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

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…

“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.”

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

Hierarchical computations within and between visual cortical areas reflect:
- the rich, probabilistic, generative structure of image input,
  constrained by:
  - the generative factors important for behavioral outcomes (hardwired or dynamic)

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:

$$p(S_1, S_2, \ldots ; I_1, I_2, \ldots)$$

joint

$$p(S_1, S_2, \ldots | I_1, I_2, \ldots)$$

posterior

$$\propto \text{likelihood} \times \text{prior}$$

$$p(I_1, I_2, \ldots | S_1, S_2, \ldots) \times p(S_1, S_2, \ldots)$$

- 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
Examples of some applications of Bayesian tools, mathematical and conceptual, to human vision:

- 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, ...).

Integration of multiple sources of ambiguous information:

- Directed graphs hierarchical inferences.
- Undirected graphs lateral inference & local "smoothness" priors.
- Known to be inferred to discount.
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)

model-dependent human parameter estimation

- human estimation of surface slant from texture—model averaging of isotropic and homogeneous texture models (Knill, 2005)
- vision/auditory localization of sound — model selection (Kording et al., 2007)
- conditioned perception. (Stocker & Simoncelli, 2008)

slant of the scree field?

flexible summaries of hierarchical motion structure

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?

a strong generative hypothesis

computational architectures: probabilistic models on graphs
the “executive metaphor”—Alan Yuille

empirical evidence for inference through generative cortical mechanisms via feedback?

V1 sensitive to local, oriented edges/bars
higher-level visual cortical regions, such as LOC, respond to changes in perceived whole shape

Diamond shape perceived

V1 activity is suppressed when the diamond shape is perceived

Lee & Mumford, 2003
Murray, Kersten, Olshausen, Schrater, & Woods (2002)
psychophysical test of modulation?
use adaptation--psychophysicist's "electrode"

assumption: adapts neurons in early cortical areas, V1
vertical appearance
adapt

assumption: adapts high-level cortical areas
normal appearance
adapt

...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)

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.

...but we haven’t always found localized suppression when local patches “fit” the larger context.

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

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

inferring the size of an object

With background clutter, there was evidence 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.
perceptual estimation of the size of an object

\[ \theta \approx \frac{S}{D} \]

Perceptual effect: \( \sim 17\% \)

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

what was found for an illusory increase in ring size


in terms of inference, what might be going on?

two possible representational assumptions:
physical or angular size?

\[ \theta \approx \frac{S}{D} \]

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 (\( \theta \))?
estimating angular size is also a non-trivial inference

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