A Crash Course on Visual Saliency Modeling: Behavioral Findings and Computational Models

Location and Dates

Conference on Computer Vision and Pattern Recognition CVPR 2013
The Oregon Convention Center in Portland, Oregon, USA

June 24, 2013, 8:30 - 17:15

Speakers

Ali Borji
University of Southern California (USC)
[primary organizor]

borji@usc.edu

Simone Frintrop
University of Bonn, Germany

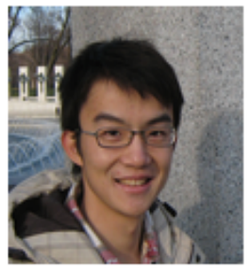
frintrop@iai.uni-bonn.de

Laurent Itti
University of Southern California (USC)

itti@usc.edu

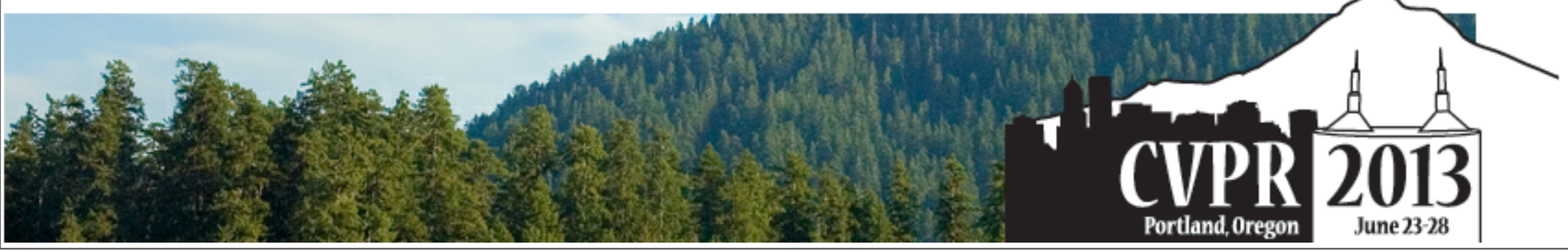
Neil D. Bruce
University of Manitoba, Canada

bruce@cs.umanitoba.ca



Xiaodi Hou
California Institute of Technology (Caltech)

xiaodi.hou@ gmail.com



Schedule

8:30 - 8:45	Introduction to the tutorial	
8:45 - 9:30	Visual attention: Background material	[Ali Borji]
9:30 - 10:15	Attention in daily life	[Ali Borji]
10:15 - 10:30	Break	
10:30 - 11:30	Bayesian and information-theoretic models	[Neil D. Bruce]
11:30 - 12:00	Applications of saliency modeling	[Neil D. Bruce]
12:00 - 13:30	Lunch break	
13:30 - 14:15	Saliency and sparsity	[Xiaodi Hou]
14:15 - 15:00	Towards attentive robots	[Simone Frintrop]
15:00 - 15:30	Attention for 3D object discovery	[Simone Frintrop]
15:30 - 15:45	Break	
15:45 - 16:45	Model comparison and challenges I	[Ali Borji]
16:45 - 17:15	Model comparison and challenges II	[Xiaodi Hou]
17:15 - 18:00	Open forum	

Schedule

8:30 - 8:45	Introduction to the tutorial	
8:45 - 9:30	Visual attention: Background material	[Ali Borji]
9:30 - 10:15	Attention in daily life	[Ali Borji]
10:15 - 10:30	Break	
10:30 - 11:30	Bayesian and information-theoretic models	[Neil D. Bruce]
11:30 - 12:00	Applications of saliency modeling	[Neil D. Bruce]
12:00 - 13:30	Lunch break	
13:30 - 14:15	Saliency and sparsity	[Xiaodi Hou]
14:15 - 15:00	Towards attentive robots	[Simone Frintrop]
15:00 - 15:30	Attention for 3D object discovery	[Simone Frintrop]
15:30 - 15:45	Break	
15:45 - 16:45	Model comparison and challenges I	[Ali Borji]
16:45 - 17:15	Model comparison and challenges II	[Xiaodi Hou]
17:15 - 18:00	Open forum	

Saliency and Life

- Unfortunately, Laurent could not make it because:

Laurent Itti

Jean-Luc Itti

- You work on saliency then you become a father

Plan

- Basic Psychology
- Salience
- Gist
- Some Neurophysiology

Basic Psychology

What is Attention?

Attention is the set of mechanisms that optimize/control the search processes inherent in vision

- **select** spatial region of interest
  - temporal window of interest
  - world/task/object/event model
  - gaze/viewpoint
  - best interpretation/response
- **restrict** task relevant search space pruning
  - location cues
  - fixation points
  - search depth control
- **suppress** spatial/feature surround inhibition
  - inhibition of return

"Everyone knows what attention is."

William James

Attention is the cognitive process of selectively concentrating on one aspect of the environment while ignoring other things. Attention has also been referred to as the allocation of processing resources.

From Wikipedia, the free encyclopedia

Adapted from Bruce, Rothenstein, and Tsotsos; ECCV Tutorial 2008

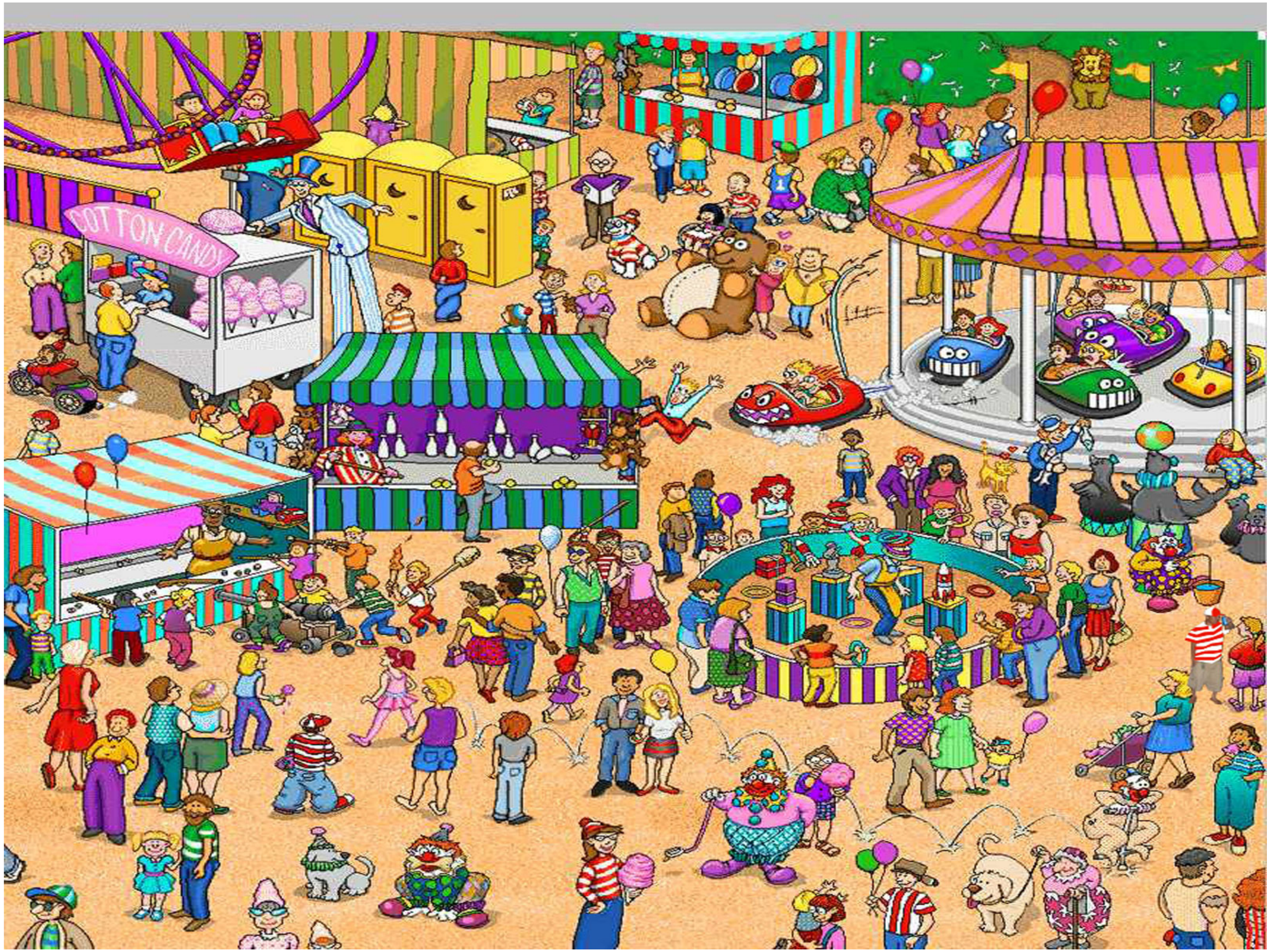
Finding “interesting” information

- In principle, very complex task:
  - Need to attend to all objects in scene?
  - Then recognize each attended object?
  - Finally evaluate set of recognized objects against behavioral goals?

- In practice, survival depends on ability to quickly locate and identify important information.

- Need to develop simple **heuristics** or approximations:
  - **bottom-up** guidance towards salient locations
  - **top-down** guidance towards task-relevant locations
  - applications?

Where is Waldo?

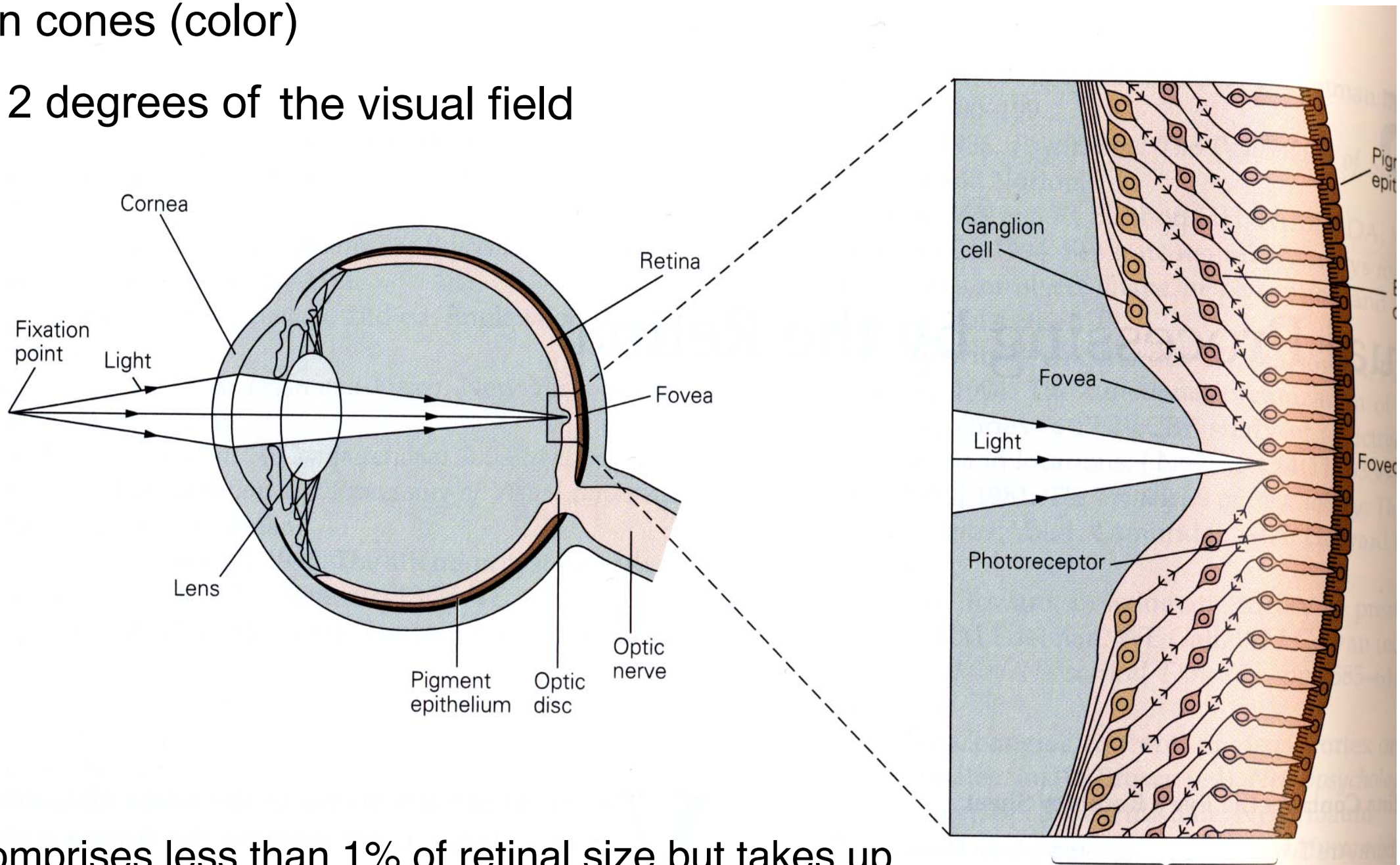


Retinal Structure

120 million rods (intensity)

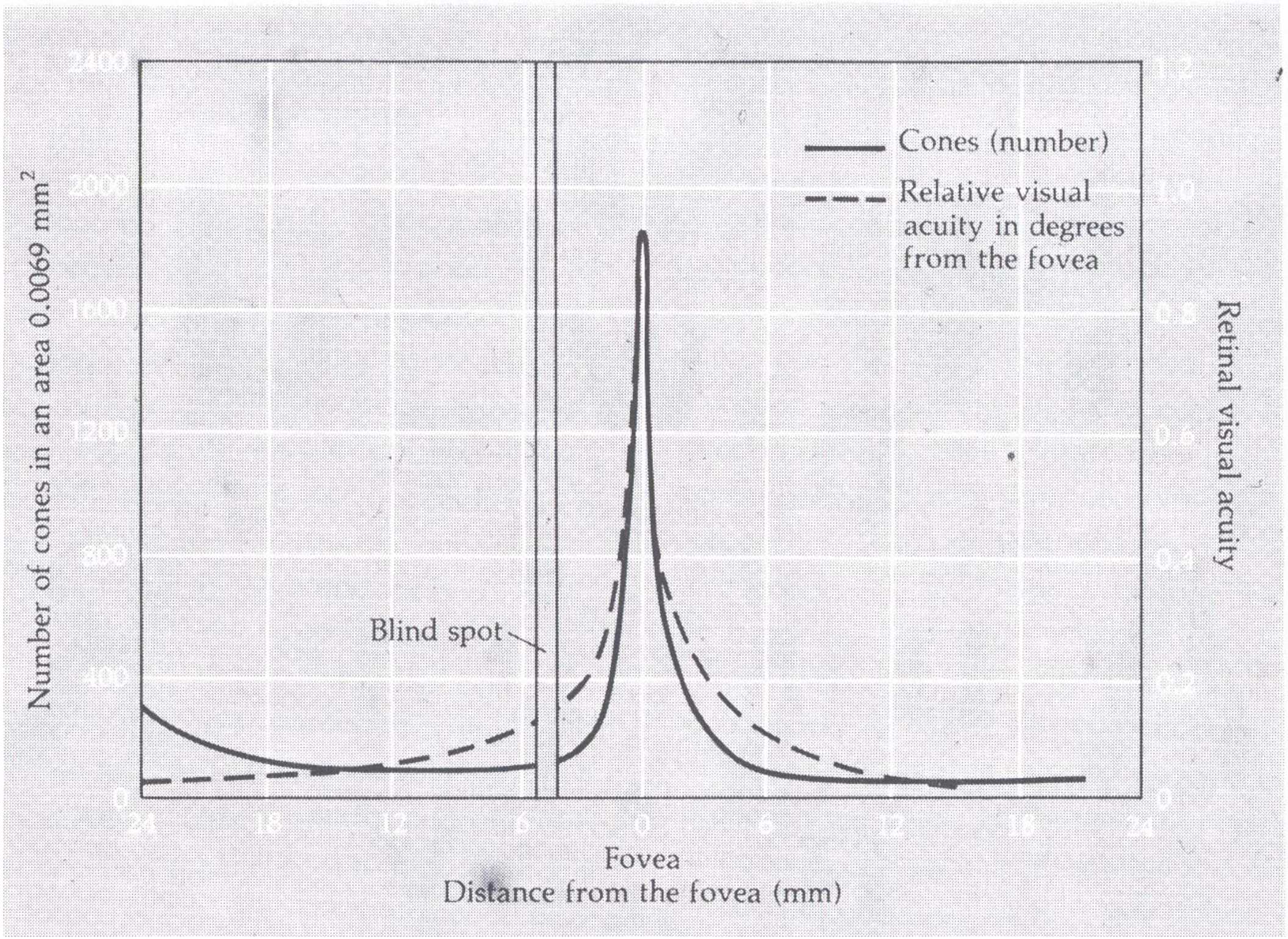
7 million cones (color)

Fovea: 2 degrees of the visual field



Fovea comprises less than 1% of retinal size but takes up over 50% of the visual cortex in the brain.

Visual acuity matches photoreceptor density



Photoreceptor distribution

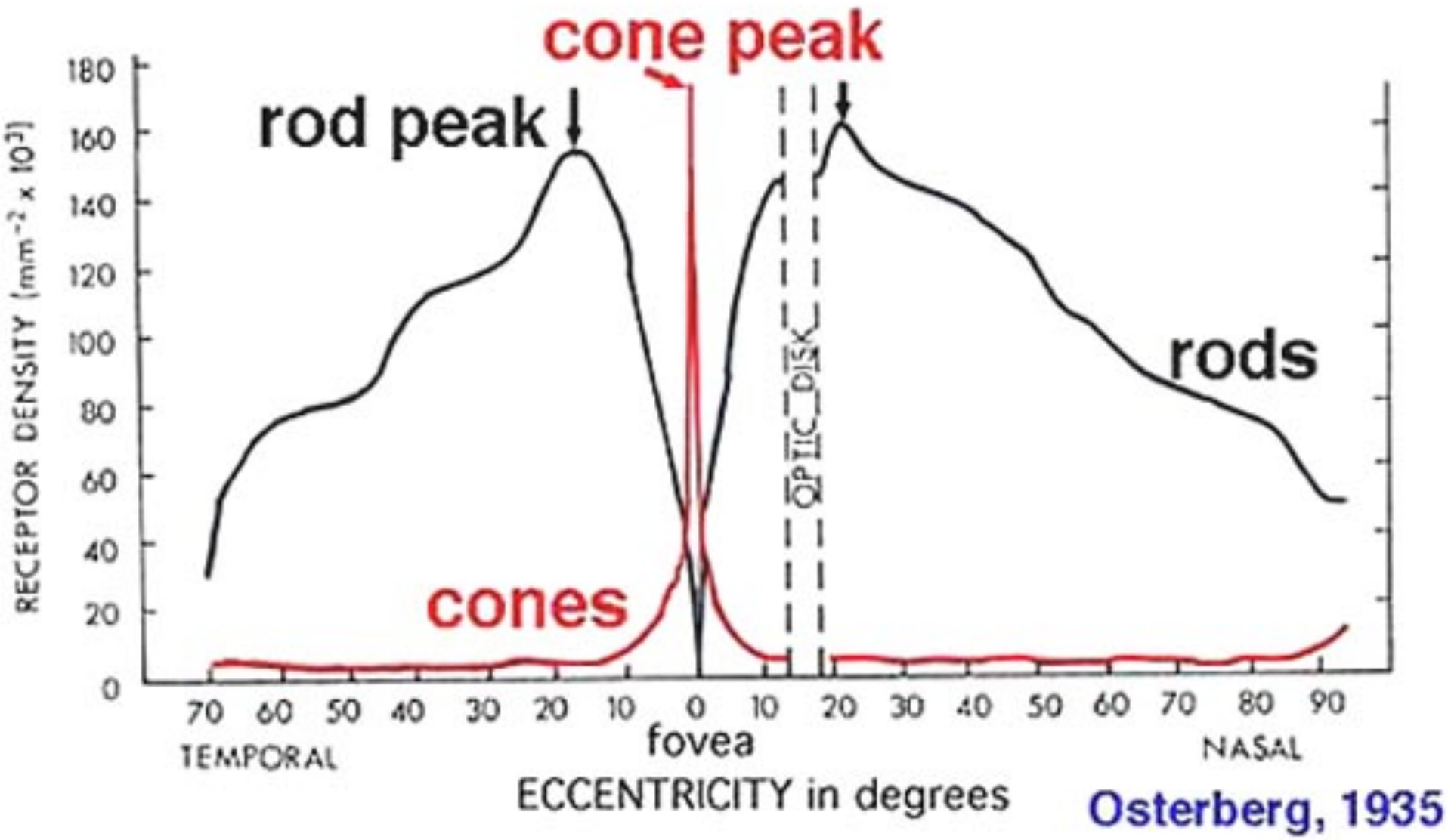


Fig. 20. Graph to show rod and cone densities along the horizontal meridian.

Type of eye movements

Information Gathering	**Stabilizing**
Voluntary (attention)	*Reflexive*
Saccades new location, high velocity (700 deg/sec),	vestibular ocular reflex (vor) body movements
ballistic(?)	
Smooth pursuit object moves, velocity, slow(ish) Mostly 0-35 deg/sec but maybe up to100deg/sec	optokinetic nystagmus (okn) whole field image motion
Vergence change point of fixation in depth slow, disjunctive (eyes rotate in opposite directions) (all others are conjunctive) Note: link between accommodation and vergence	

Fixation: period when eye is relatively stationary between saccades.

Saccades

- Scope: 2 deg (poor spatial res beyond this)
- Duration: 50-500 ms (mean 250 ms)
- Length: 0.5 to 50 degrees (mean 4 to 12)
- Various types (e.g., regular, tracking, micro)

A few definitions

Attention and eye movements:

- overt attention (with eye movements)
- covert attention (without eye movements)

Bottom-up and top-down control:

- bottom-up control
  - based on image features
  - very fast (up to 20 shifts/s)
  - involuntary / automatic
- top-down control [Focus of the second talk]
  - may target inconspicuous locations in visual scene
  - slower (5 shifts/s or fewer; like eye movements)
  - volitional

Control and modulation:

- direct attention towards specific visual locations
- attention modulates early visual processing at attended location

What is attention, then?

Attention is often described as an information processing bottleneck.

Controls access to higher levels of processing, short-term memory and consciousness.

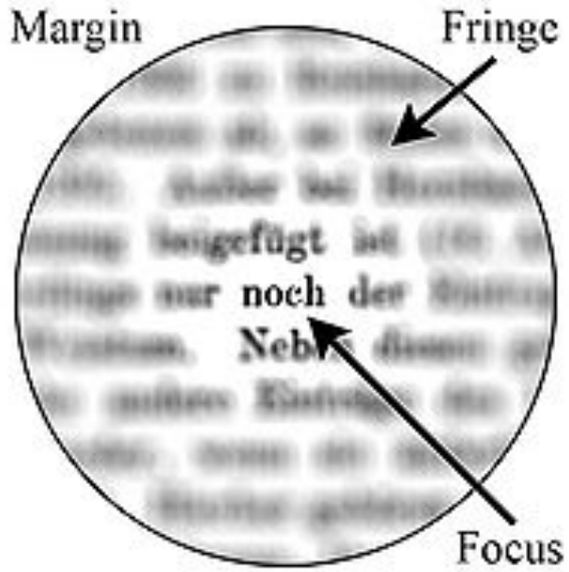
Metaphor of the "spotlight" of attention (Crick, 1984).

Hence, the strategy nature has developed to cope with information overload is to break down the problem of analyzing a visual scene:

- from a massively parallel approach
- to a rapid sequence of circumscribed recognitions.

Pre-attentive processing: low-level visual processing happening in real-time over the entire visual scene

Attentive processing: more detailed analysis of only those scene regions which are attended to



Spotlight Metaphor

The attention bottleneck: selects a fraction of the incoming visual input for detailed processing.

Change Blindness

Rensink, R., O'Regan, K., Clark, J., (1997). To See or Not to See: The Need for Attention to Perceive Changes in Scenes, *Psychological Science*, 8, 368-373.

Precursors: visual memory - Observers were found to be poor at detecting change if old and new displays were separated by an ISI of more than 60-70 ms.
saccades - observers were found to be poor at detecting change, with detection good only for a change in the saccade target

Two conclusions:

- observers never form a complete, detailed representation of their surroundings.
- attention is required to perceive change, and that in the absence of localized motion signals it is guided on the basis of high-level "interest".

http://www.psych.ubc.ca/~rensink/flicker/download/

Rensink et al., Nature, 1997

Saturday, June 22, 2013

Illusions

- We don't see what we think we see
- Change blindness (Door): http://www.youtube.com/watch?v=VkrrVozZR2c&feature=c4-overview-vl&list=PLE9CC1569697BFF96
- Change blindness:
- https://www.youtube.com/watch?v=ubNF9QNEQLA
- Pen & Teller magic tricks: http://www.youtube.com/watch?v=oJhYySXzOq0
- http://www.youtube.com/watch?v=FxJb-Lw8onY
- Tatler magic talk:
- http://www.youtube.com/watch?v=Z3Dpb6Hs9aQ

SimonsLab

HOME PEOPLE RESEARCH VIDEOS RESOURCES PARTICIPATE CONTACT

Welcome to the Simons Lab Website

This is the website for the Visual Cognition Laboratory at the University of Illinois at Champaign-Urbana, headed by Prof. Daniel Simons. Professor Simons is a member of the Department of Psychology and the laboratory is located on the second floor of the Beckman Institute for Advanced Science and Technology.

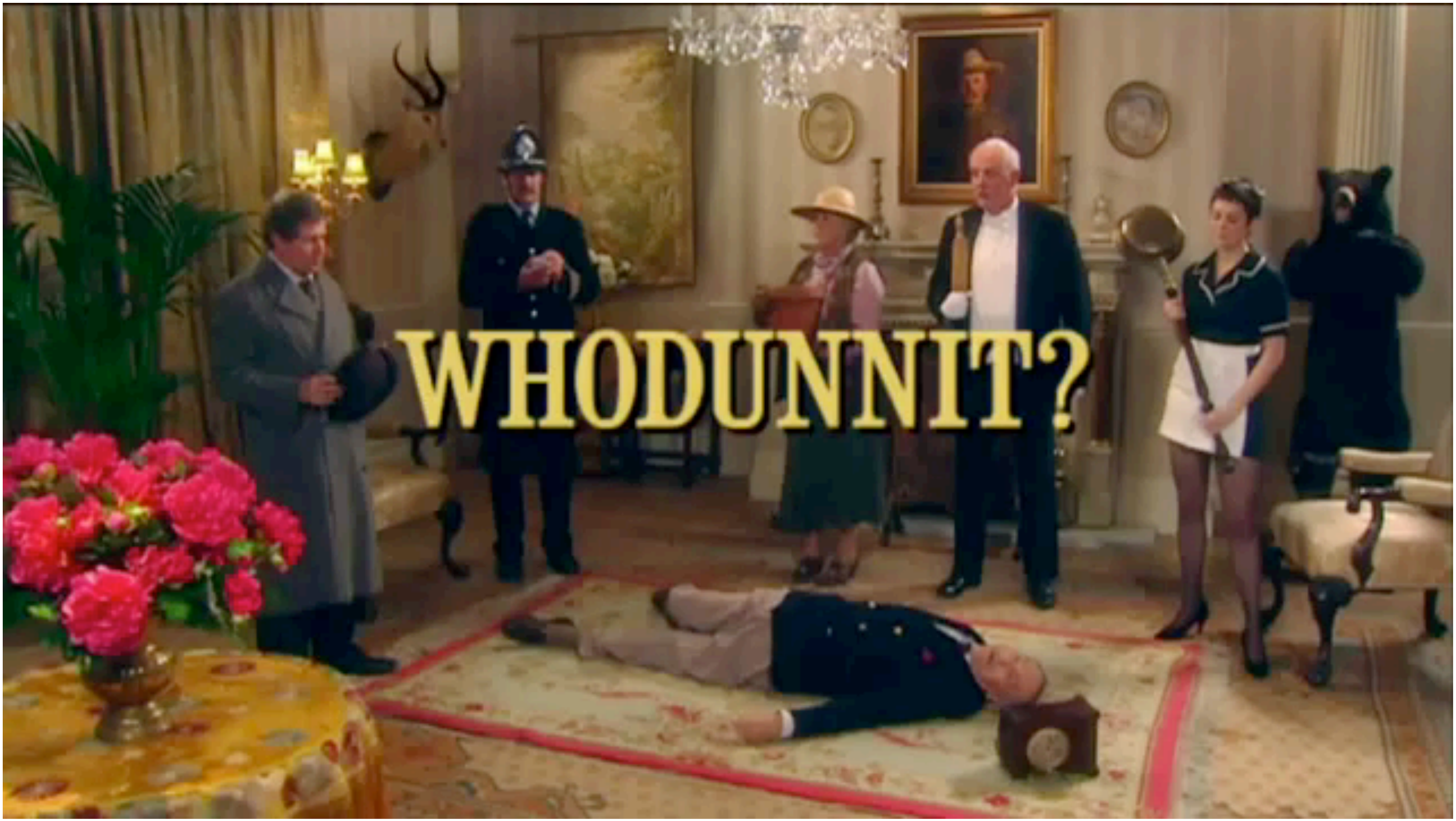
On this site, you can "meet" members of the laboratory, learn about the research we do, view videos, etc.

We decided to host the website on a private domain rather than at the university to make it easier to maintain and to provide a shorter domain name so that people can find it more easily. Older versions of this site hosted at the University of Illinois will redirect to this page.

Latest News:

- New lab website launched on October 4, 2010
- Prof. Simons's personal website launched on September 1, 2010

The Invisible Gorilla


WHODUNNIT?


ROYAL
VARIETY
PERFORMANCE

Inhibition of Return

Posner, M. I., Rafal, R. D., Choate, L. S., and Vaughan, J. (1985). Inhibition of return: Neural basis and function. *Cognitive Neuropsychology*, **2**(3), 211–228

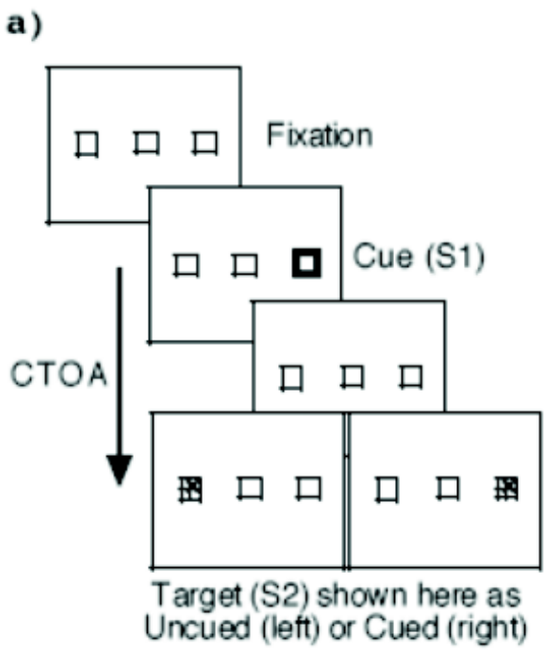
A bias against returning attention to previously attended locations

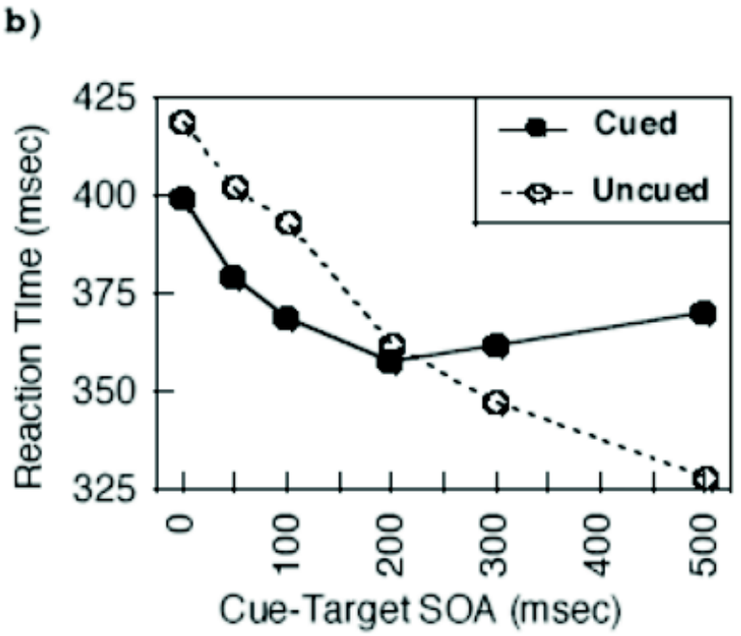
Posner and Cohen (1984) - by discouraging orienting toward previously attended locations in a scene, IOR might serve as a novelty seeking mechanism.
Klein (1988) hypothesized that by biasing orienting away from previously attended locations in the environment IOR could serve to facilitate visual search when the target does not pop out.

IOR can be location or object based
IOR can be task-based - as Yarbus showed, we visit locations several times, presumably until we have found the information we are looking for, and then there is no need to look again

Why go back to anything anyway?
Need to first know the answer to:
what is extracted in a single attentional fixation?
what is remembered and for how long?

Also see
Lupianez, J., Klein, R., Bartolomeo, P. (2006). Inhibition of Return: twenty Years After, *Cognitive Neuropsychology, 23(7), 1003-1014*

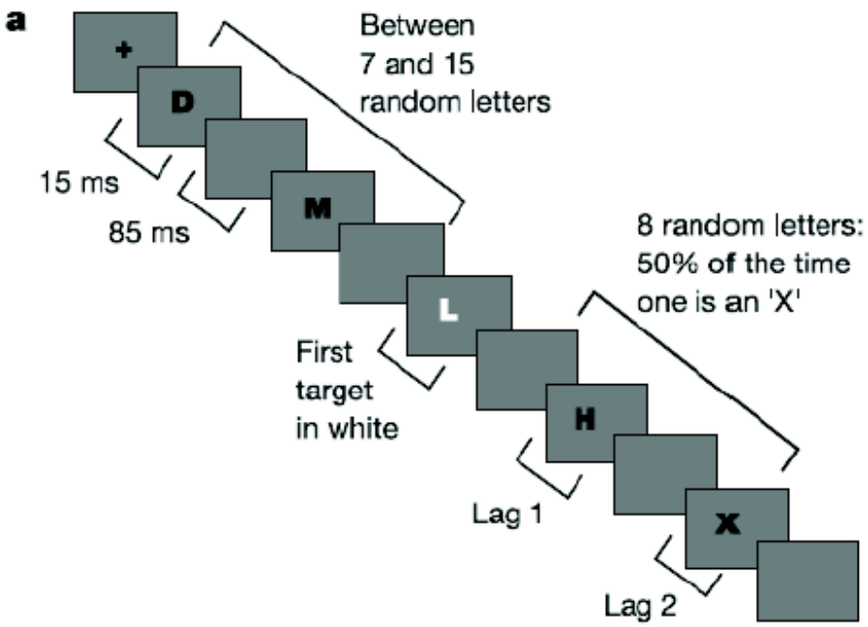


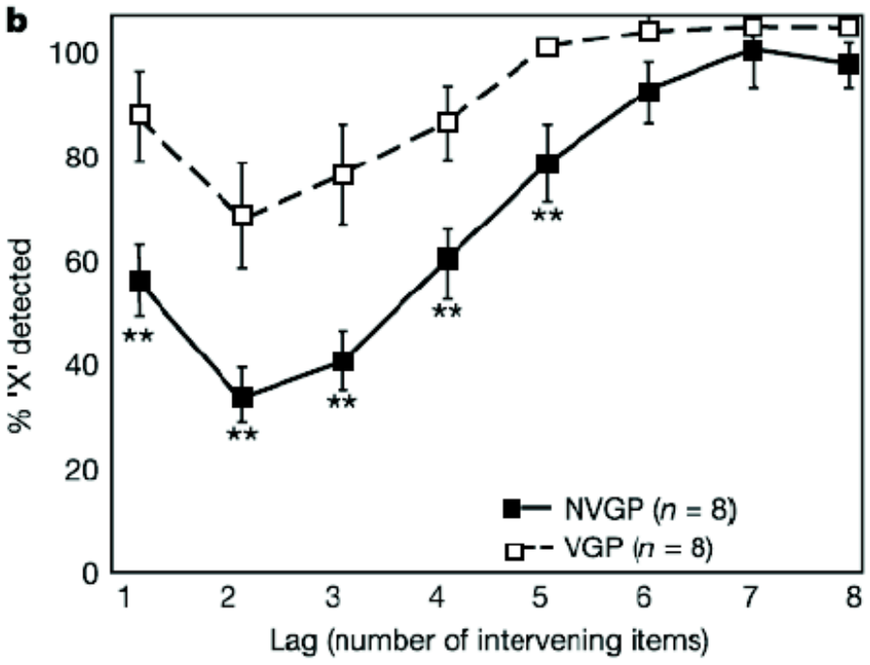


Adapted from Bruce, Rothenstein, and Tsotsos; ECCV Tutorial 2008

Attentional Blink

Raymond, J. E., Shapiro, K. L. Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink?, *J. Experimental Psychology: Human Perception and Performance. 18(3)*, 849-860.





Task: identify a partially specified letter (target) and then detect the presence or absence of a fully specified letter (probe).

- targets are accurately identified

- probes are poorly detected when they are presented during a 270-msec interval beginning 180 msec after the target. Probes presented immediately after the target or later in the RSVP stream are accurately detected.

- not found in conditions if a brief blank interval followed the target

- suggest that the presentation of stimuli after the target but before target-identification processes are complete produces interference at a letter-recognition stage.

from Green, C.S., Bavelier, D., (2003). Action video game modifies visual selective attention, *Nature 42*, 534-537.

!!

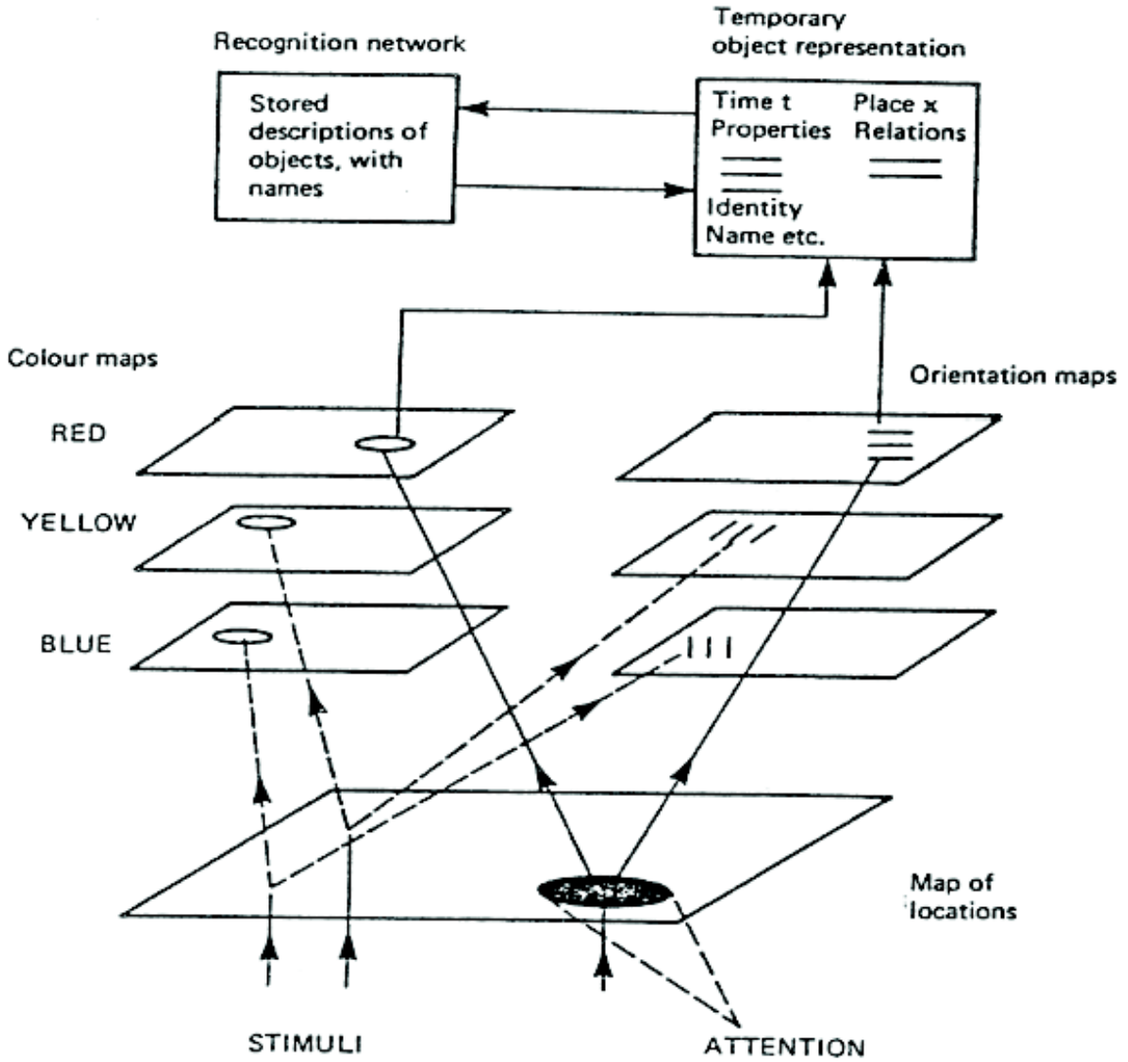
Your target is the letter "R"... which will be followed rapidly by the letter "C".

Feature Integration Theory (FIT)

Treisman, A., Gelade, G. (1980). A feature integration theory of attention, *Cognitive Psychology 12:* 97-136.

Key ideas:

- we can detect and identify separable features in parallel across a display (within the limits set by acuity, discriminability, and lateral interference)
- this early, parallel, process of feature registration mediates texture segregation and figure ground grouping;
- that locating any individual feature requires an additional operation;
- that if attention is diverted or overloaded, illusory conjunctions may occur;
- conjunctions, require focal attention to be directed serially to each relevant location;
- they do not mediate texture segregation, and they cannot be identified without also being spatially localized.



from Treisman & Sato 1990

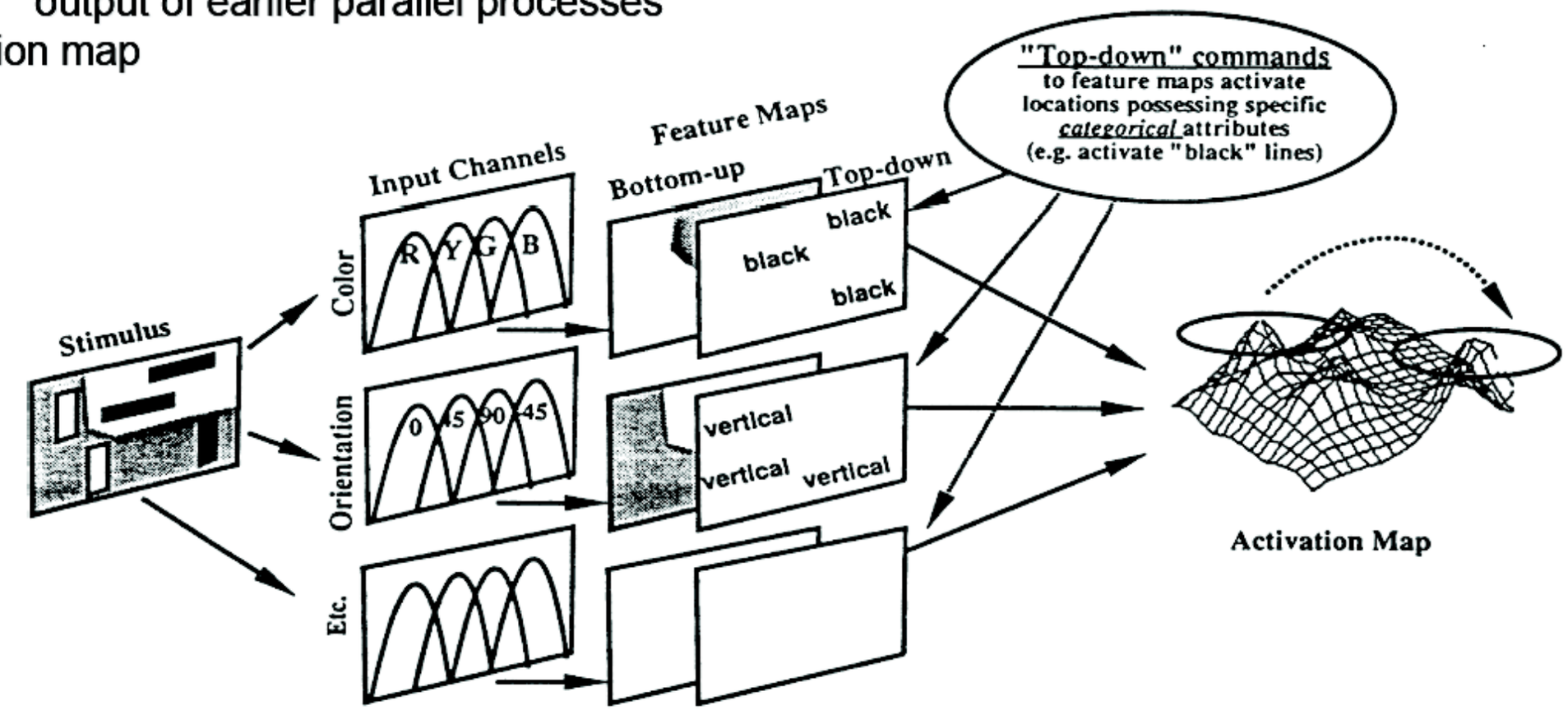
Adapted from Bruce, Rothenstein, and Tsotsos; ECCV Tutorial 2008

Guided Search 1989

Wolfe, J., Cave, K., Franzel, S. (1989). Guided search: An alternative to the feature integration model for visual search, *J. Exp. Psychology: Human Perception and Performance 15*, 419-433.

Key ideas:

- attentional deployment of limited resources is guided by output of earlier parallel processes
- activation map



Adapted from Bruce, Rothenstein, and Tsotsos; ECCV Tutorial 2008

What's a Feature?
What Attracts Attention?

for a nice summary, see Wolfe, J. (1998). Visual Search, in **Attention** (ed. Pashler, H.), 13–74, University College London, London.

Just about everything someone may have studied can be considered a feature or can capture attention

Wolfe presents the kinds of features that humans can detect 'efficiently':

- Color
- Orientation
- Curvature
- Texture
- Scale
- Vernier Offset
- Size, Spatial Frequency
- Motion
- Shape
- Onset/Offset
- Pictorial Depth Cues
- Stereoscopic Depth

Table1 | **Attributes that might guide the deployment of attention**

Undoubted attributes*	Probable attributes‡	Possible attributes§	Doubtful cases‖	Probable non-attributes¶
-Colour[26,27,37,39,40] -Motion[30,56,57] -Orientation[41,42,58–61] -Size (including length and spatial frequency)[27,62,63]	-Luminance onset (flicker)[64,65] -Luminance polarity[21,66] -Vernier offset[67] -Stereoscopic depth and tilt[68–70] -Pictorial depth cues[71–73] -Shape[27,58,74–80] -Line termination[22,81,82] -Closure[26,77,83–85] -Topological status[77,86,87] -Curvature[27,67,88]	-Lighting direction (shading)[51,89] -Glossiness (luster)[52] -Expansion[90,91] -Number[27,81] -Aspect ratio[27]	-Novelty[28,53,92] -Letter identity (over-learned sets, in general)[93–95] -Alphanumeric category[96–99]	-Intersection[8,58] -Optic flow[29,91] -Colour change[64] -Three-dimensional volumes (such as geons)[100,101] -Faces (familiar, upright, angry and so on)[102–108] -Your name[109] -Semantic category (for example, 'animal','scary')[10]

Attributes are grouped by the likelihood that they are, in fact, sources of guidance of attention. References are representative but not exhaustive. *'Undoubted' meaning that they are supported by many studies with converging methods. ‡Less confidence owing to limited data, dissenting opinions or the possiblity of alternative explanations. §Still less confidence. ‖Unconvincing, but still possible. ¶Suggested guiding features where the balance of evidence argues against inclusion on the list.

Wolfe and Horowitz, NN reviews 2004

For most, subjects can 'select' feature or feature values to attend in advance

Adapted from Bruce, Rothenstein, and Tsotsos; ECCV Tutorial 2008

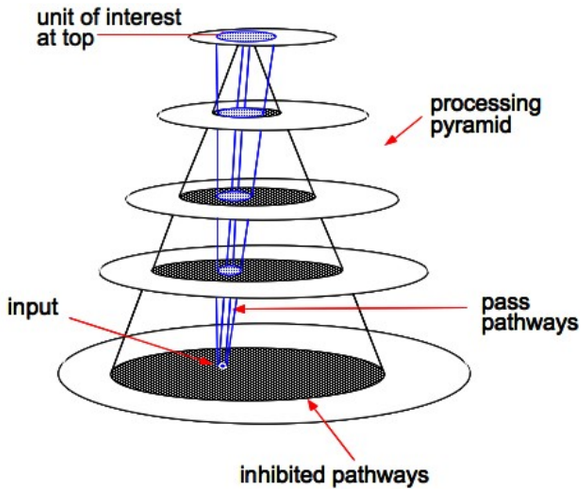
Salience

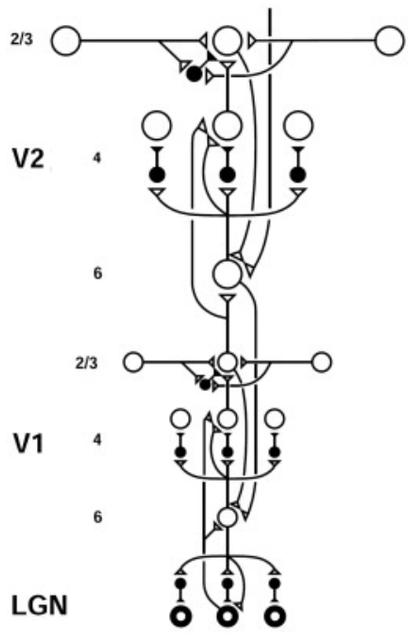
Visual Salience

- Some notion of what is interesting in the world that captures our attention
- is important as it drives a decision we make a couple hundred thousand times a day - where we decide to look.
- The role of Cognitive Science is to create a working model of visual salience
- Several computational models have been proposed over the past 30 years. [Focus of the next talks]

Old Testament

- Biased Competition
- Selective Tuning
- Grossberg
- Deco
- Broadbent 1958
- Deutsch/Norman
- Moray/MacKay Model
- Treisman 1964
- Kahneman 1973
- Milner 1974
- Treisman & Gelade 1980
- Crick 1984
- Wolfe 1989+
- Bundesen 1990+
- Von der Malsburg 1981





Niebur, Koch et al. 1993+

Desimone & Duncan 1995

Deco et al. 2001+

Grossberg 1976+

Koch and Ullman 1985

Fukushima 1986 (Neocognitron

Anderson and Van Essen 1987

Ullman 1995

Cave 1999

Burt 1988

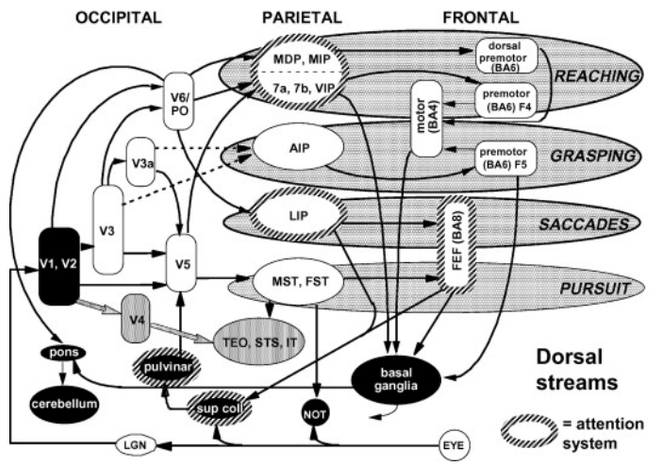
Sandon 1990

Tsotsos 1990+

Ahmad 1991

Mozer 1991

...

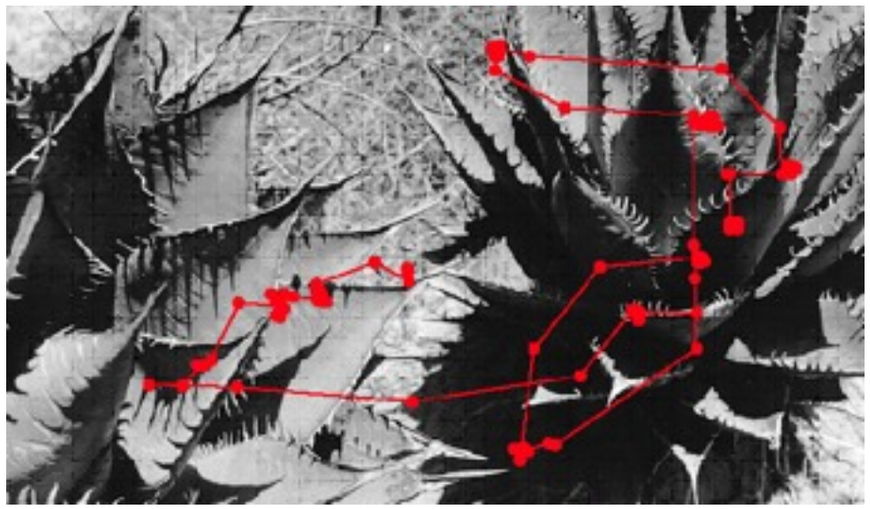


These models aim to model neural/cognitive mechanisms of attention rather than predicting gaze (Referred here as to "New Testament")

See also Bruce, Rothenstein, and Tsotsos; ECCV Tutorial 2008

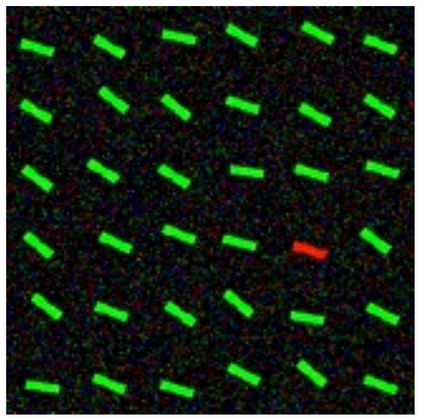
- Local image statistics

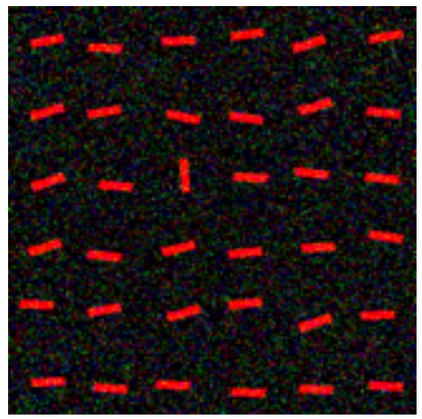
  E.g., Barth et al. ‘98; Reinagel & Zador ‘99;
  Privitera & Stark ‘00; Parkhurst & Niebur ‘03;
  Einhauser et al. `06; Tatler et al. `07



- Spatial outliers – Saliency

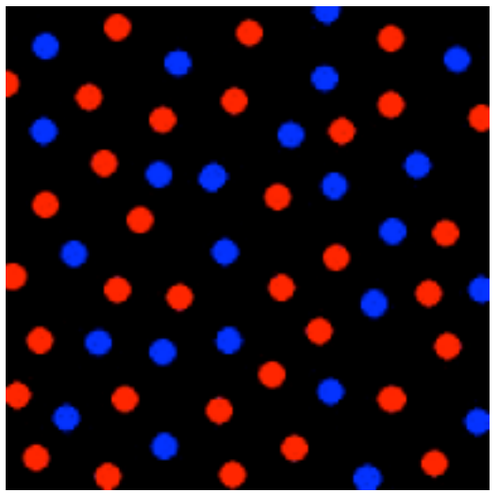
  E.g., Treisman & Gelade ‘80; Koch & Ullman ‘85;
  Tsotsos et al. ‘95; Li, ‘98; Itti, Koch & Niebur `98;
  Burce & Tsotsos `06; Gao & Vasconcelos `07;
  Zhang et al. `07





- Temporal outliers – Novelty

  E.g., Mueller et al. ‘99; Markou & Singh ‘01;
  Theeuwes ‘95; Fecteau & Munoz ‘04

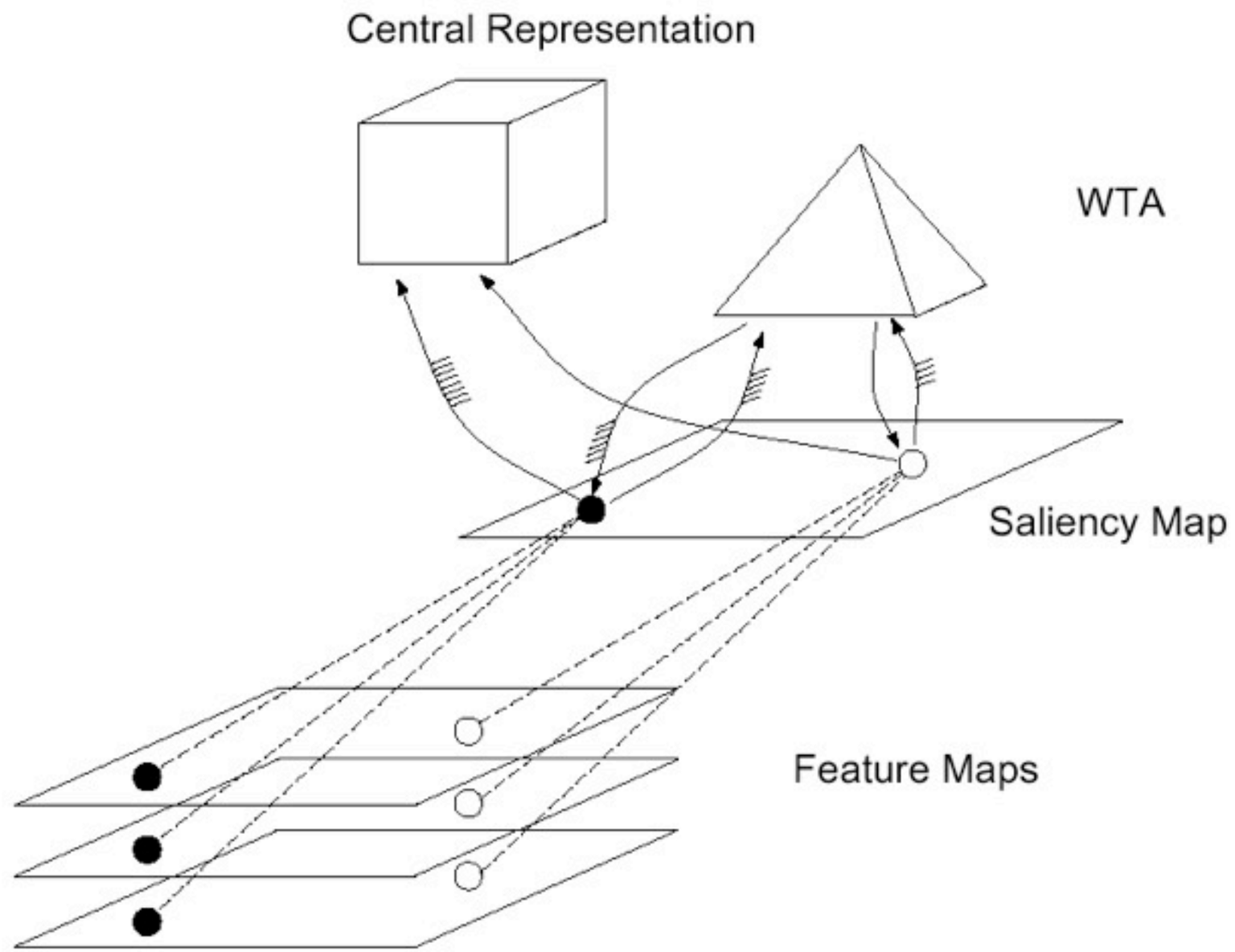


First computational model

Koch & Ullman,
Hum. Neurobiol., 1895

Introduce concept of a single topo-graphic saliency map.

Most salient location selected by a winner-take-all network.



Input image

Multiscale low-level feature extraction

Colours
Red, green, blue, yellow, etc.

Intensity
On, off, etc.

Orientations
0°, 45°, 90°, 135°, etc.

Other
Motion, junctions and terminators, stereo disparity, shape from shading, etc.

Attended location

Inhibition of return

Winner-take-all

Saliency map

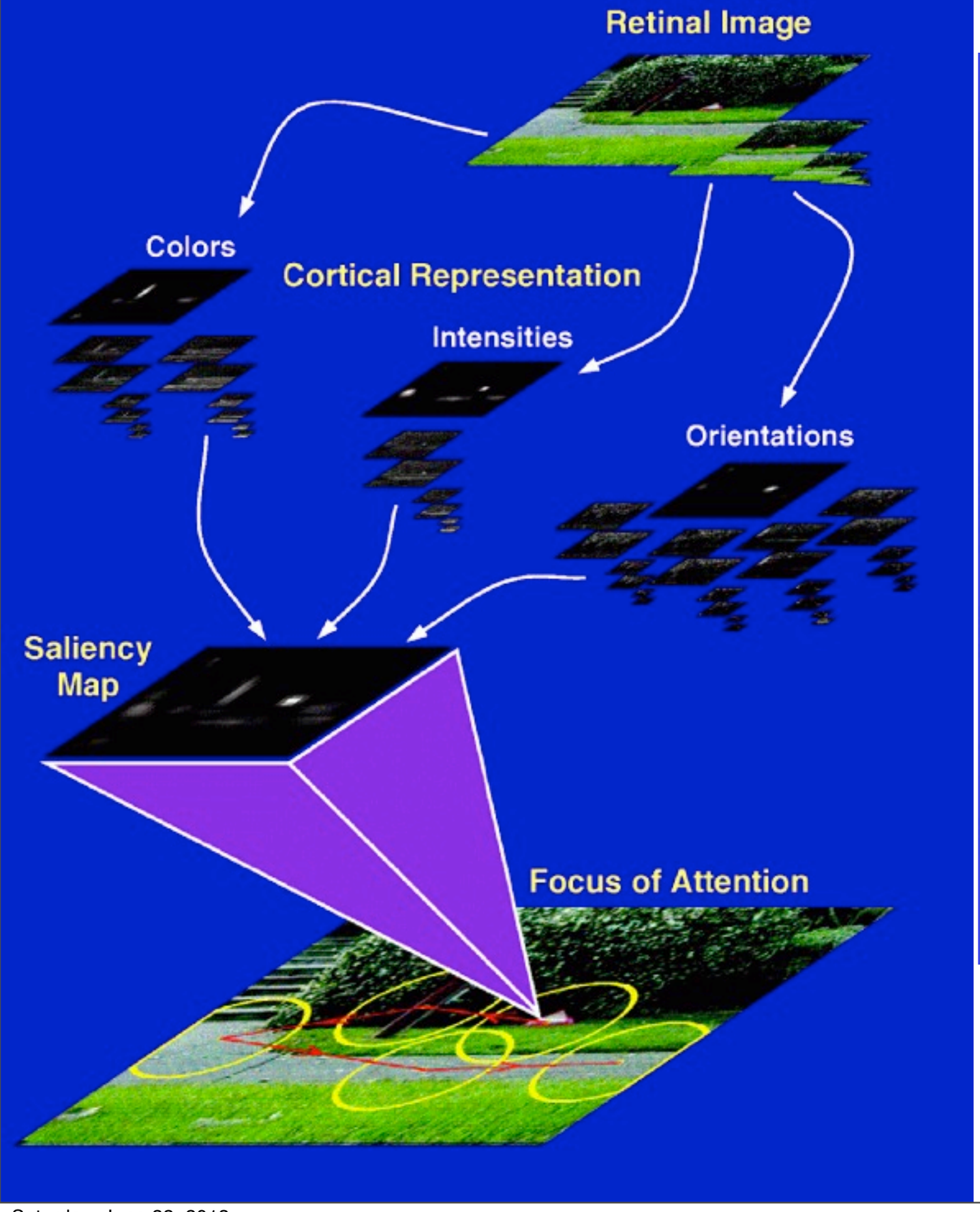
Centre–surround differences and spatial competition

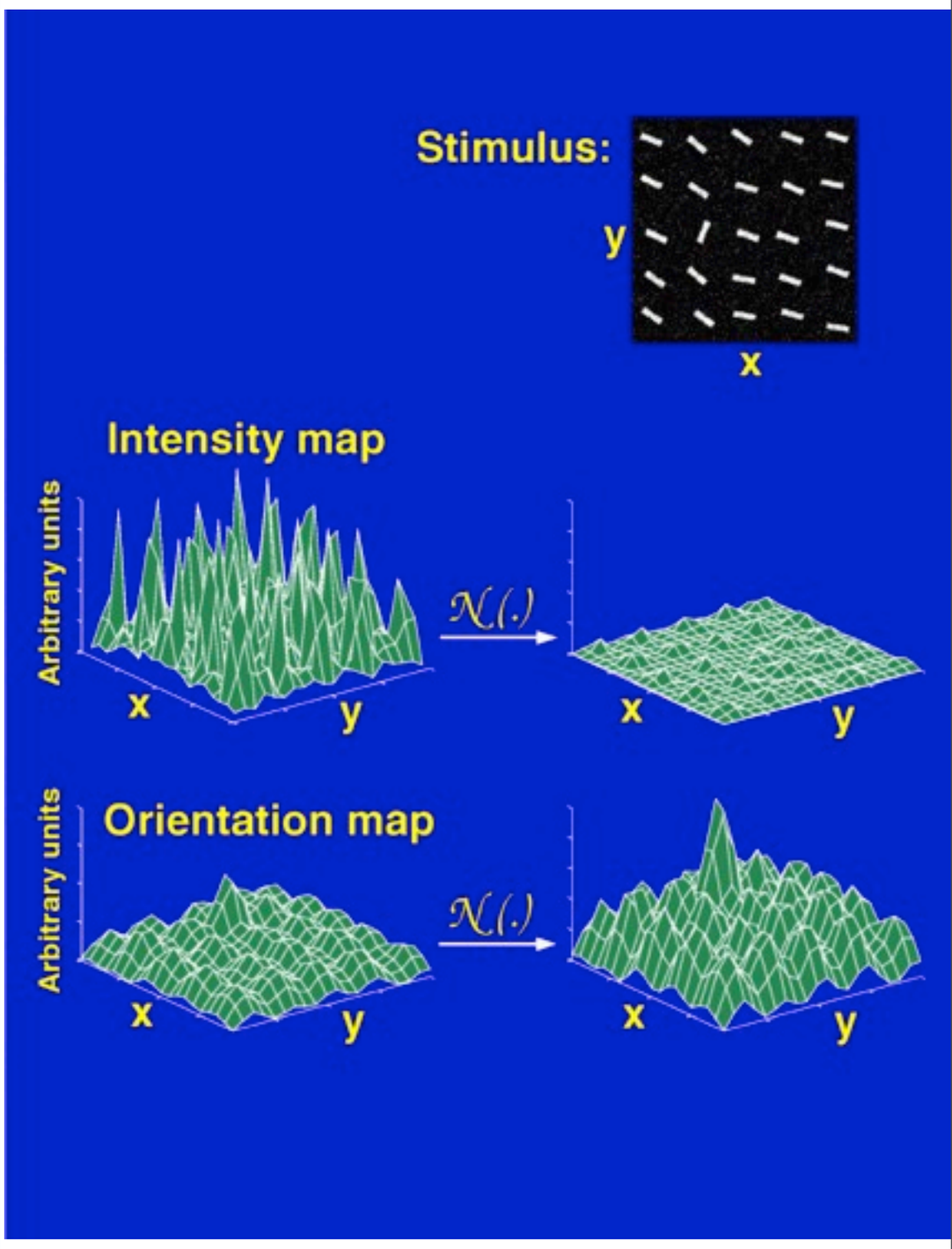
Feature maps

Feature combinations

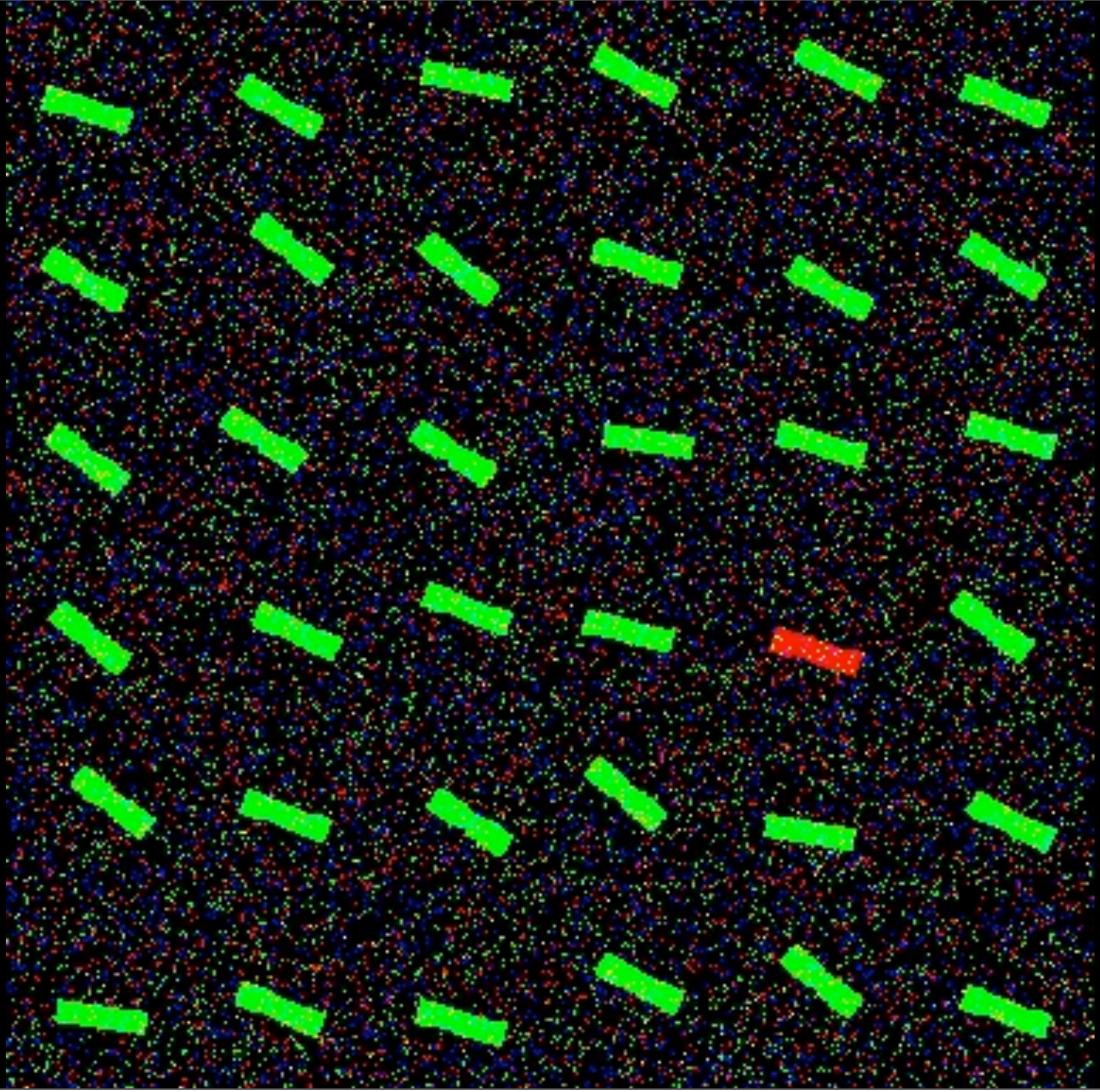
Top-down attentional bias and training

Itti & Koch, Nat. Rev. Neurosci., 2001

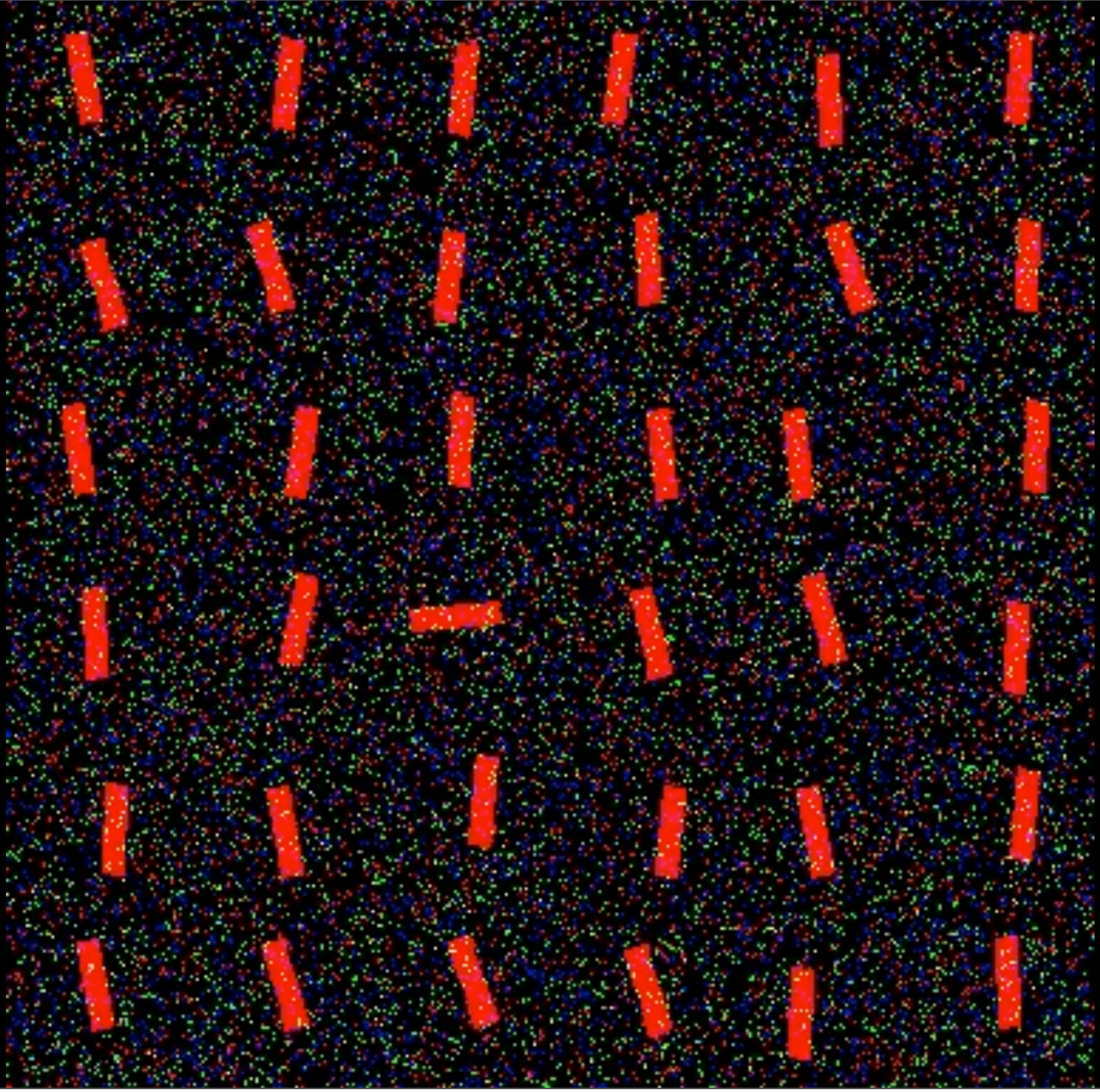


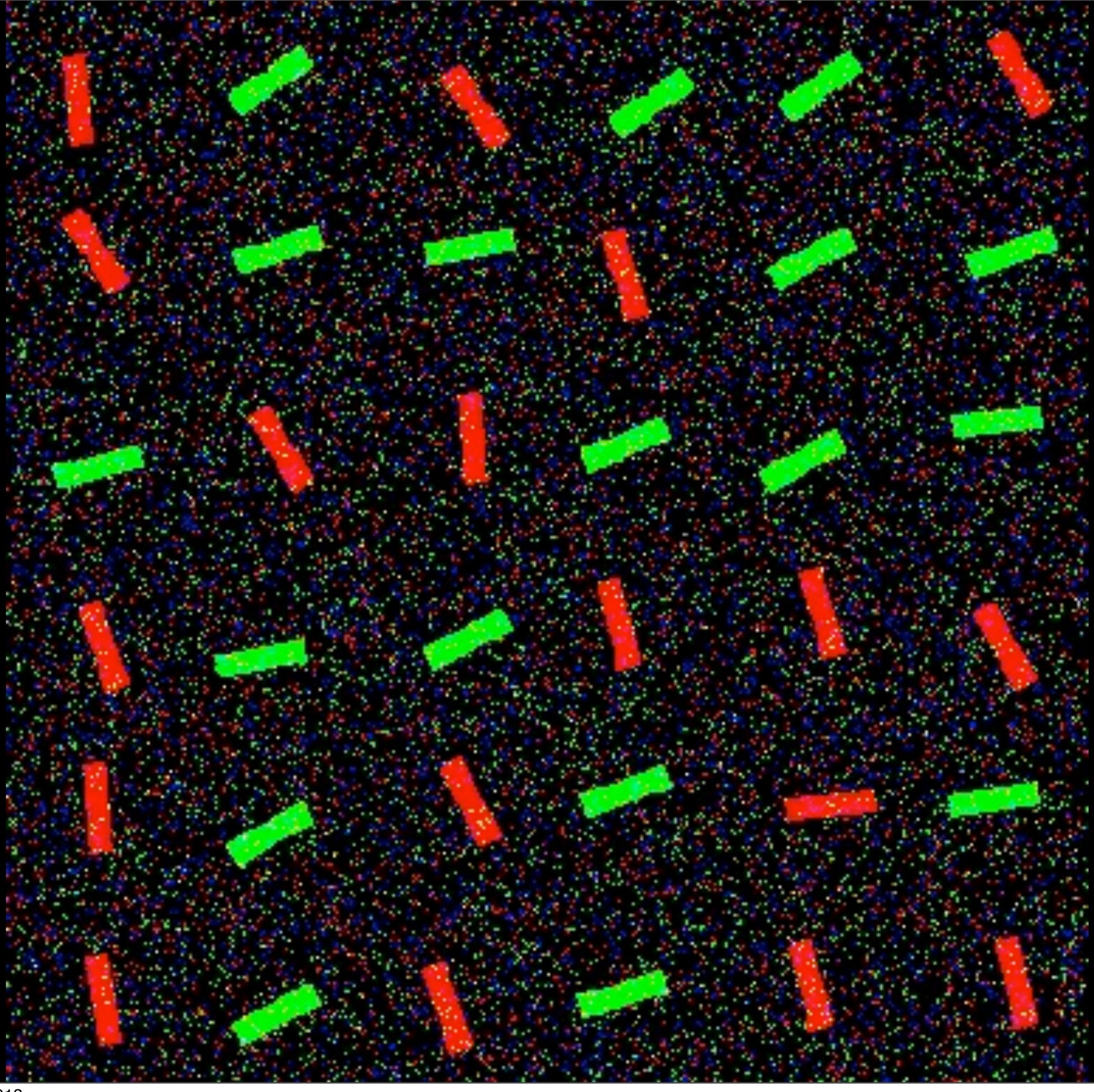


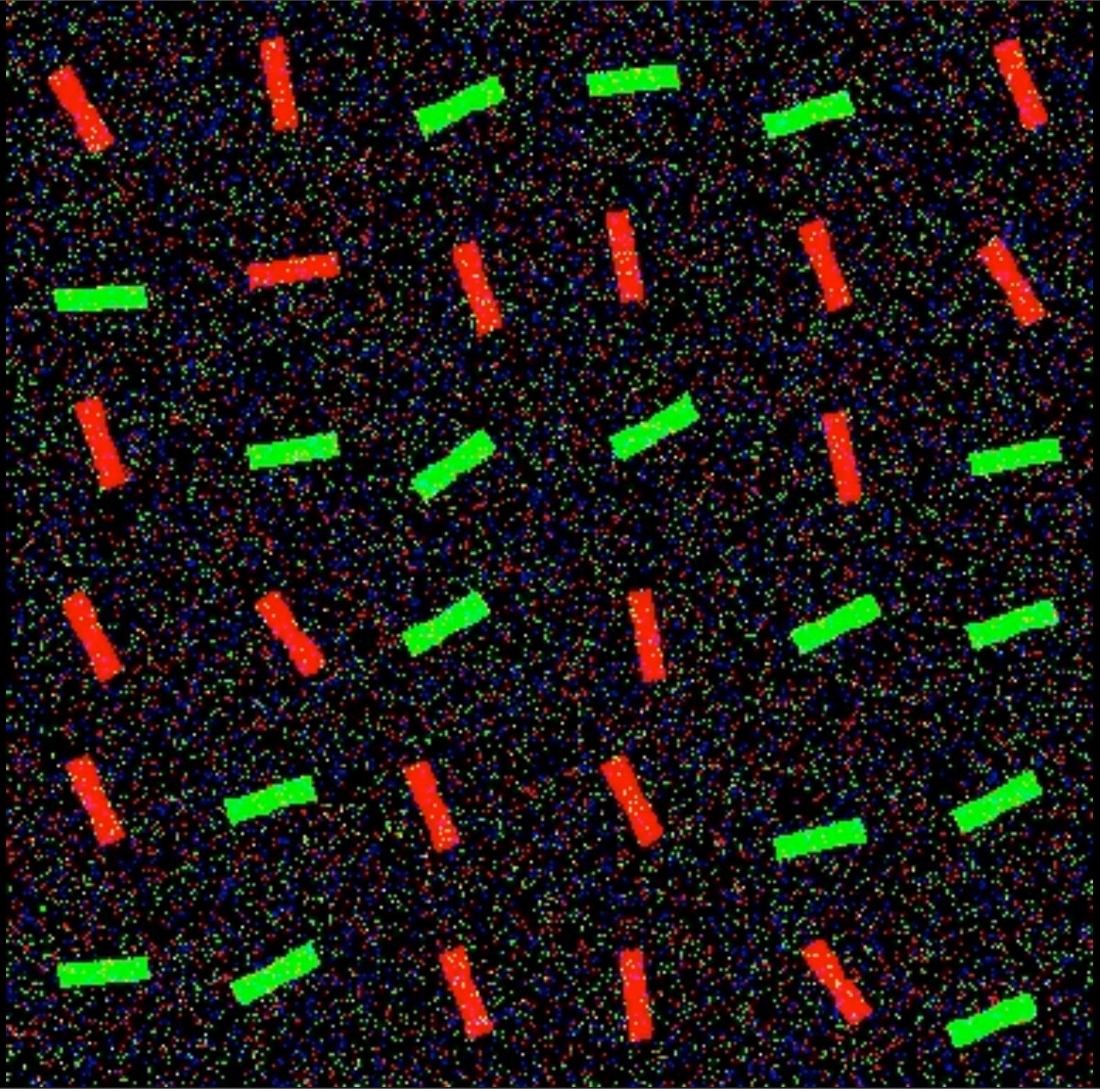
Itti, Koch & Niebur,
IEEE PAMI 1998



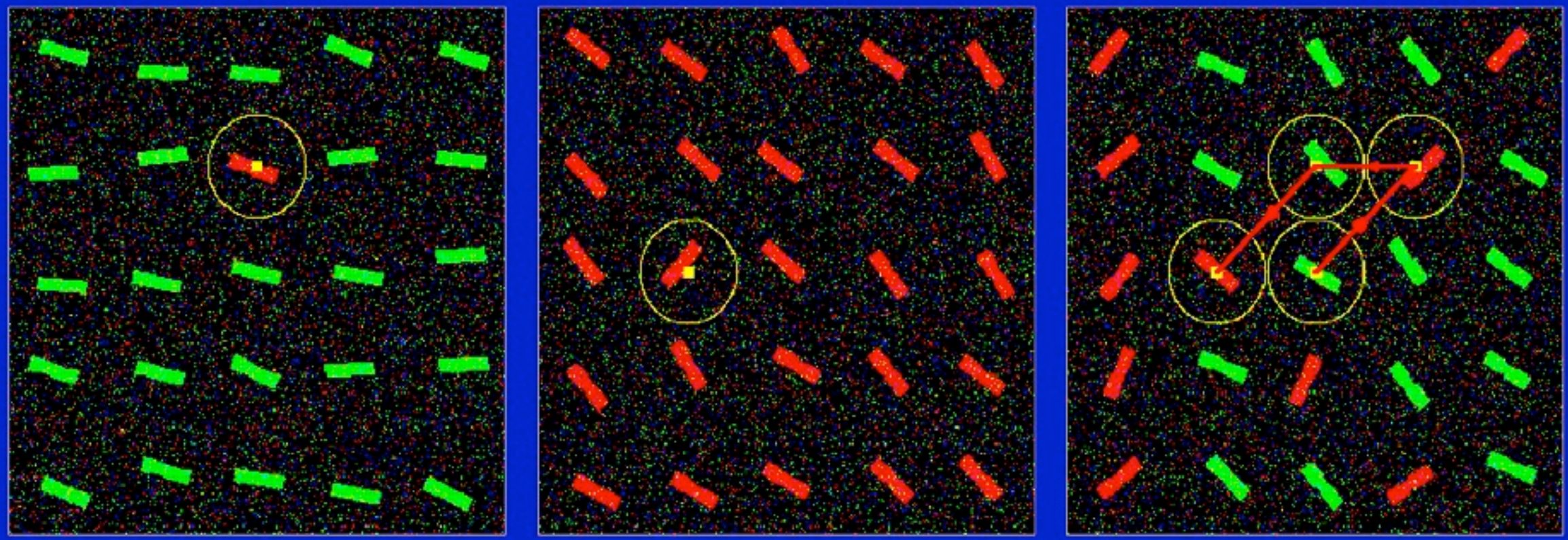
Saturday, June 22, 2013

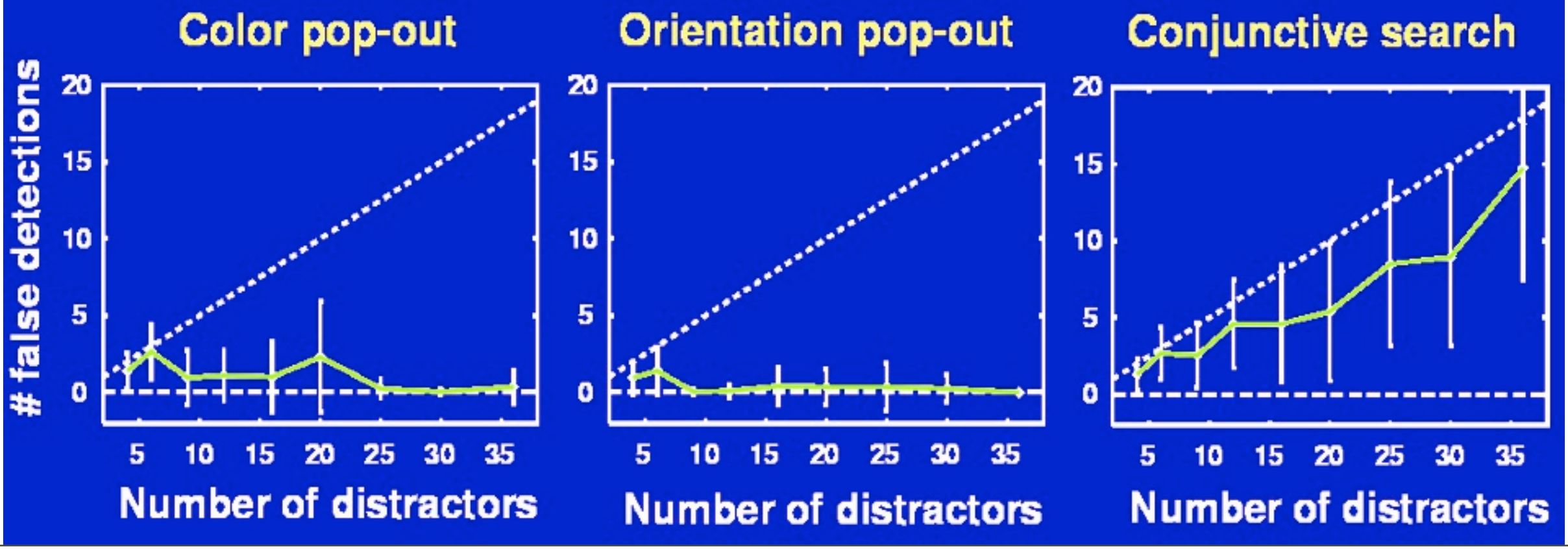


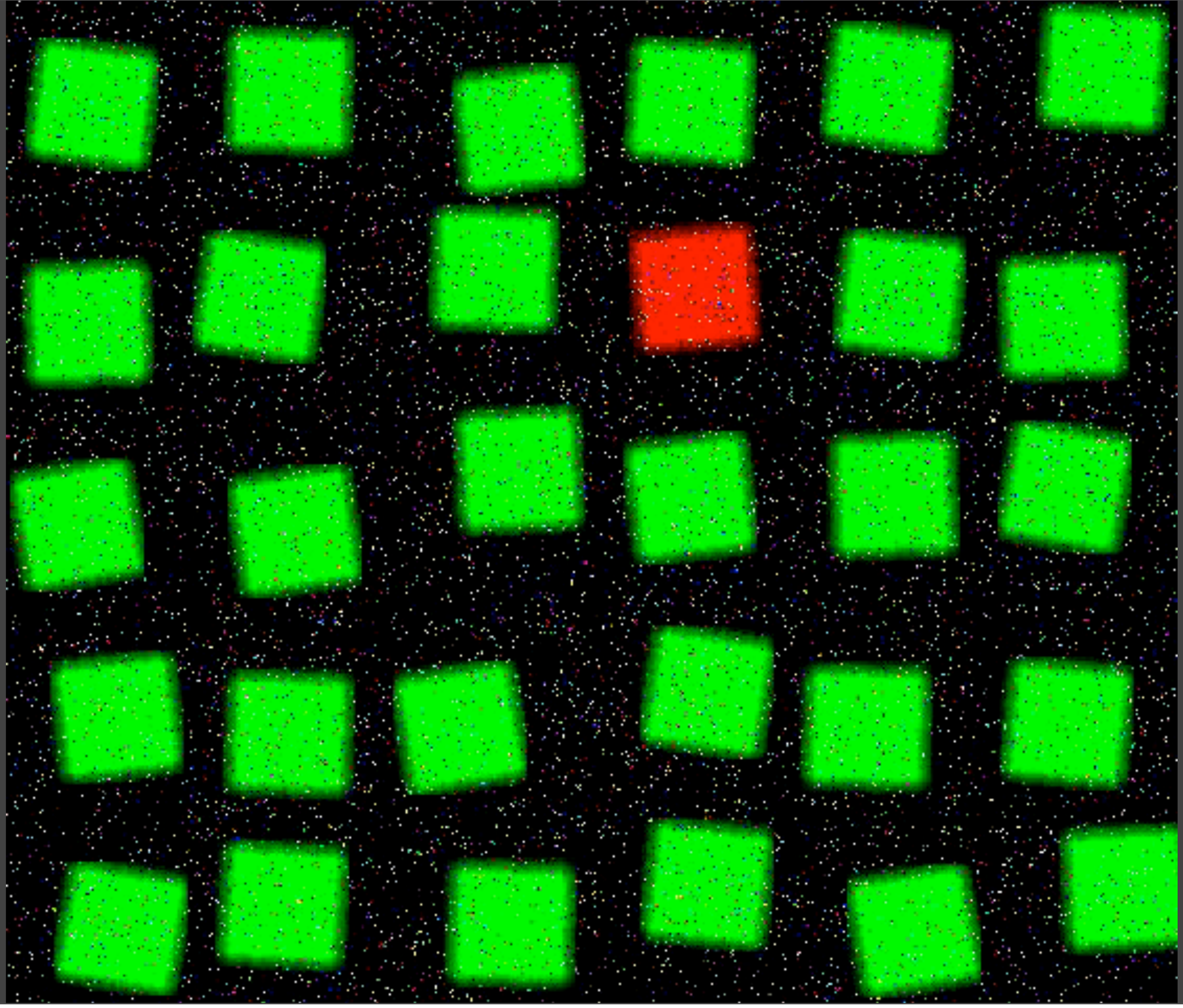




Simulated Psychophysics



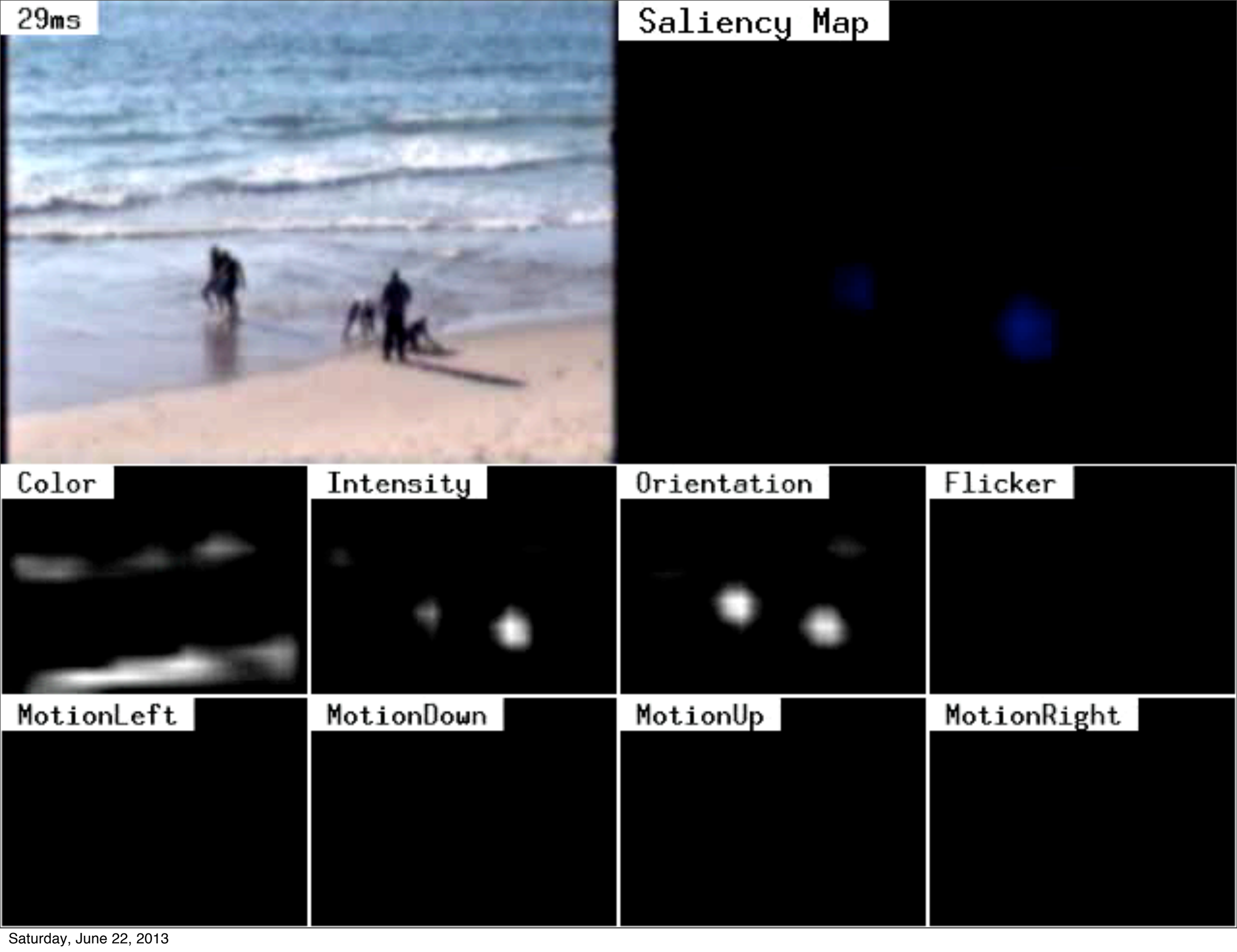




Itti, Koch & Niebur, 1998

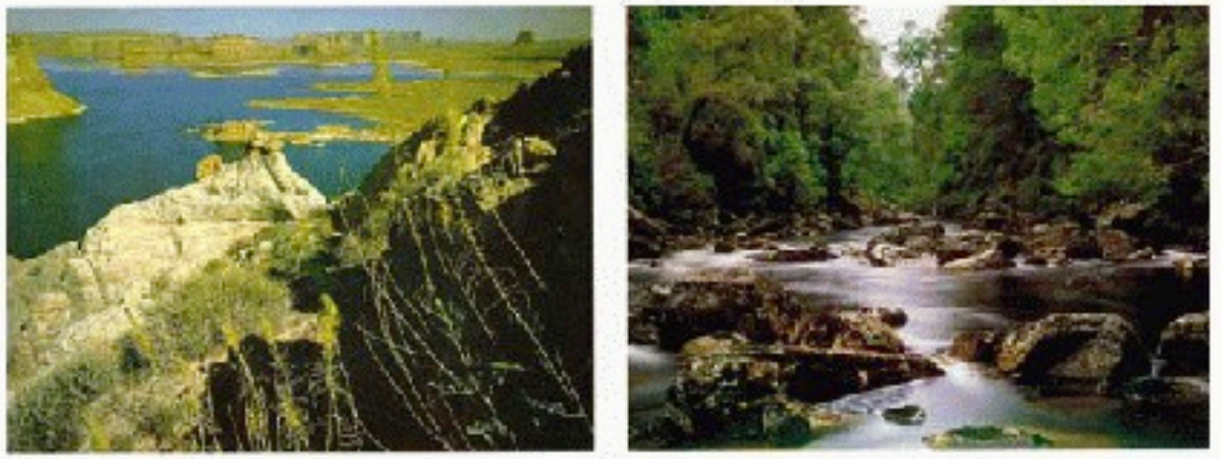
Also see:

Treisman & Gelade, 1980
Wolfe, Cave & Franzel, 1989
Hamker, 1999
Heinke & Humphreys, 1997
Zhaoping, 1999
Krummenacher et al., 2001
Torralba, Oliva et al., 2006
Tsotsos et al., 1995
Tatler et al., 2005
Underwood et al., 2007
Zhang et al., 2008
Kanan, Tong, Zhang & Cottrell, 2009


29ms
Saliency Map
Color
Intensity
Orientation
Flicker
MotionLeft
MotionDown
MotionUp
MotionRight

Home Interiors

Natural Landscapes



Buildings and City Scenes

Fractals

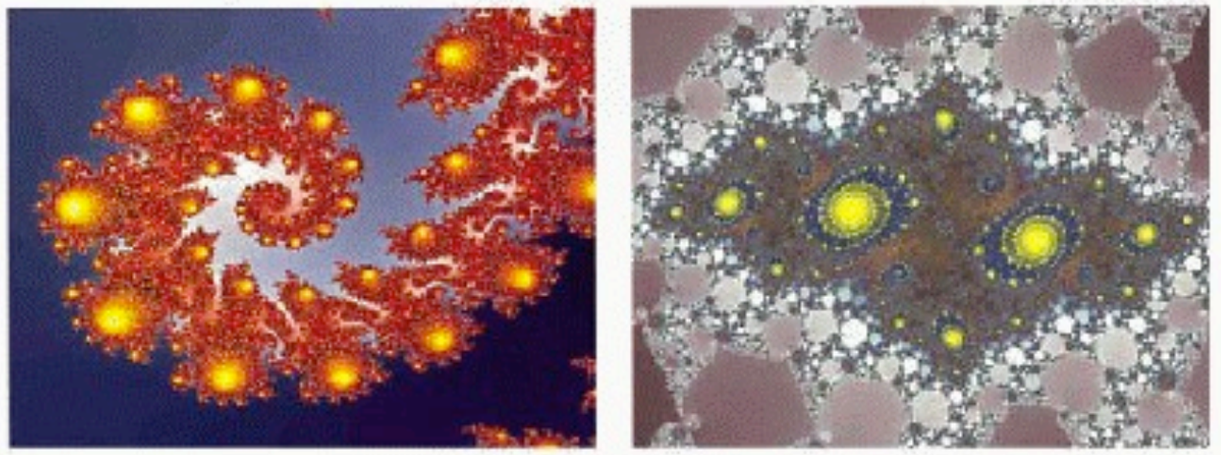


Fig. 2. Examples of the four classes of images used in the experiment.

Parkhurst et al. 2002

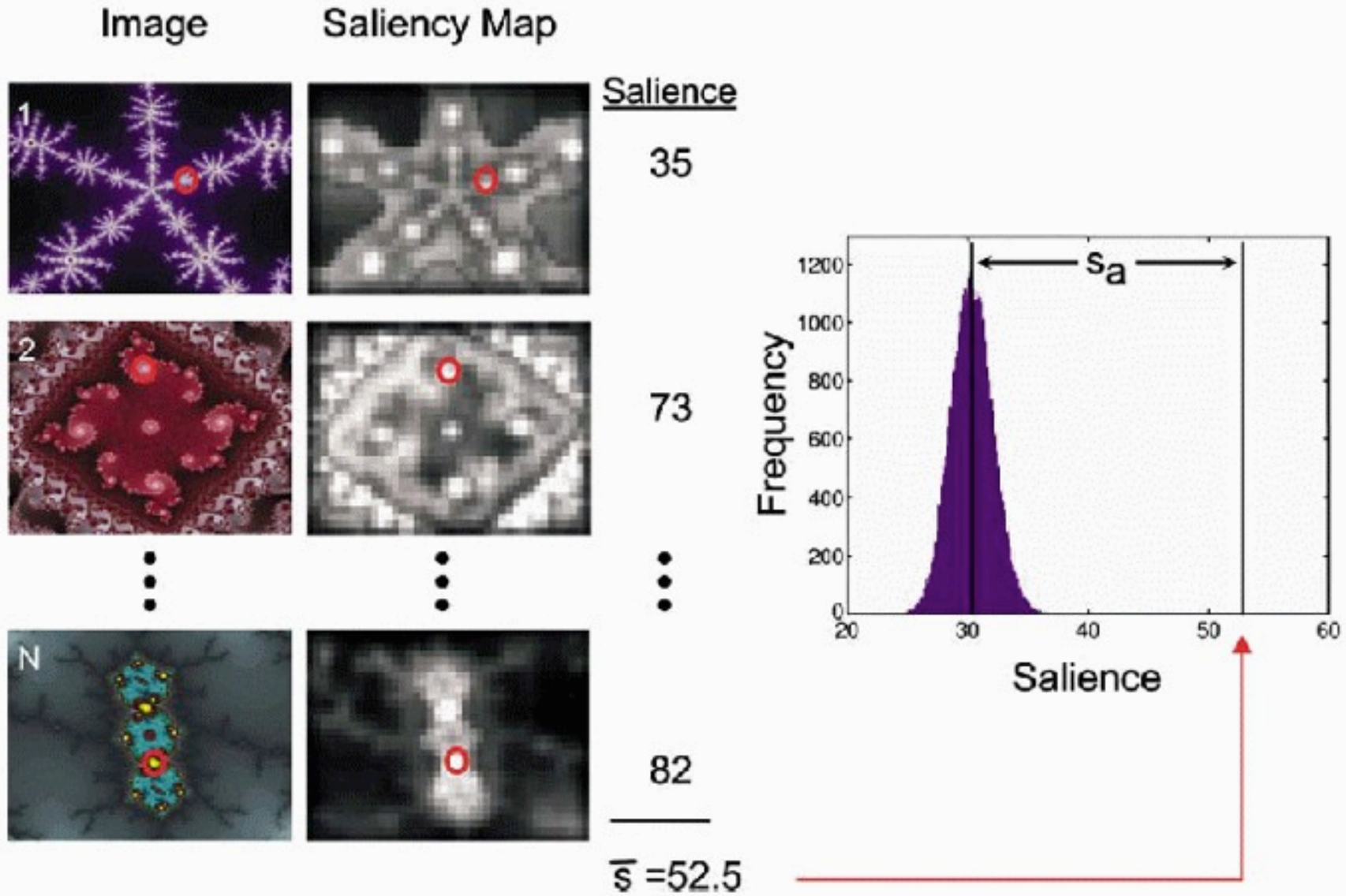


Fig. 3. The method for quantifying the correlation between stimulus salience and fixation locations is illustrated for one image database. The location of the first fixation after stimulus onset is extracted from the eye movement record and indicated by a red circle on each image (left). A saliency map is generated for each image in the database and the saliency at the first fixation location is extracted (center). The mean of the extracted salience values ($\bar{s}$) is calculated across images and compared to the distribution of $\bar{s}$ expected by chance (right). The distance between the $\bar{s}$ obtained as a fixation location and the mean $\bar{s}$ expected by chance alone is referred to as the chance-adjusted salience s_a.

Parkhurst et al. 2002

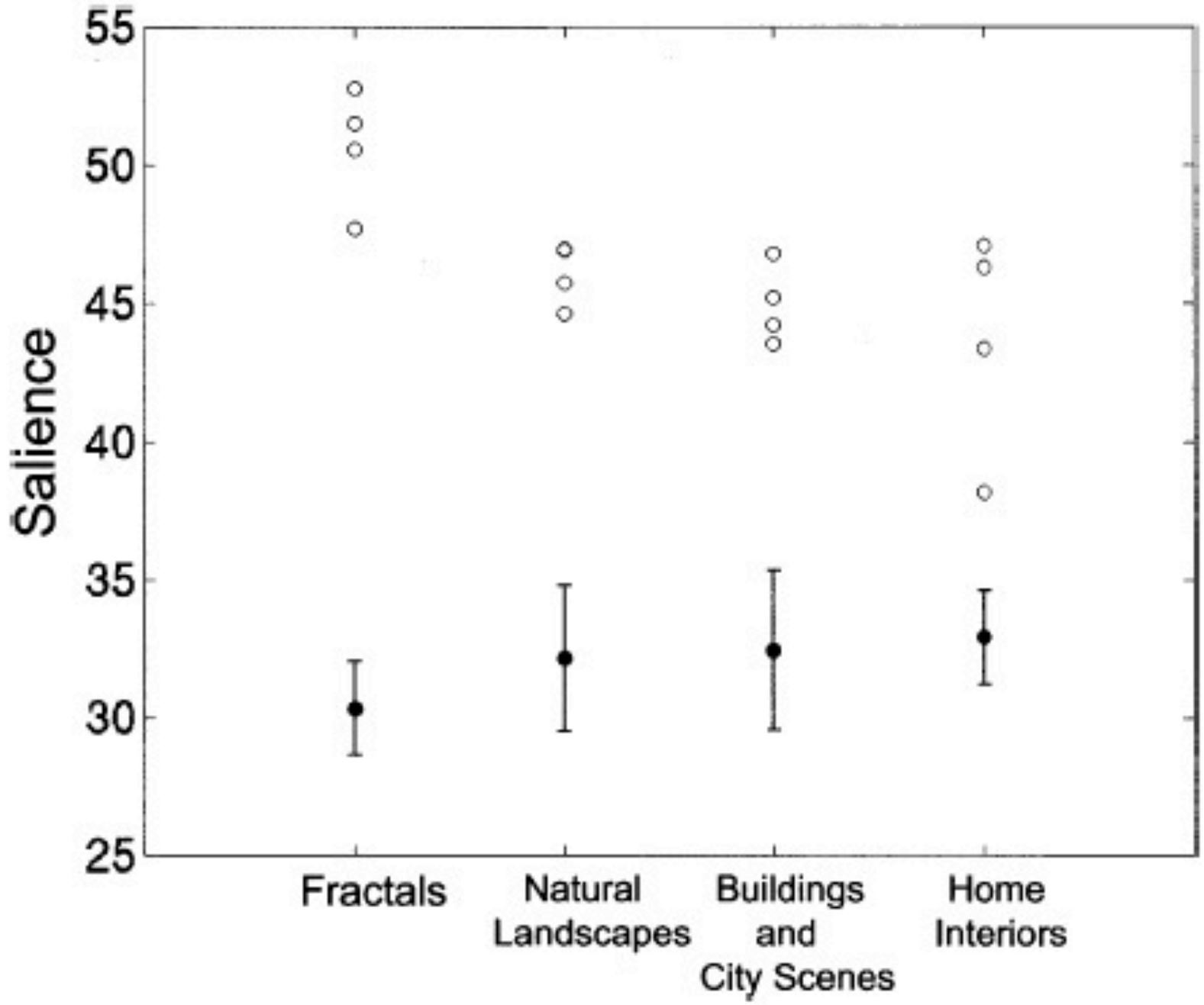


Fig. 4. The mean salience at the first fixation location is shown as an open circle for each participant within each database. The mean salience expected by chance for each database is shown as a closed circle with errorbars indicating plus/minus one standard error of the mean. Each observation significantly differs from chance. Stimulus dependence for the fractal images was the highest.

Parkhurst et al. 2002

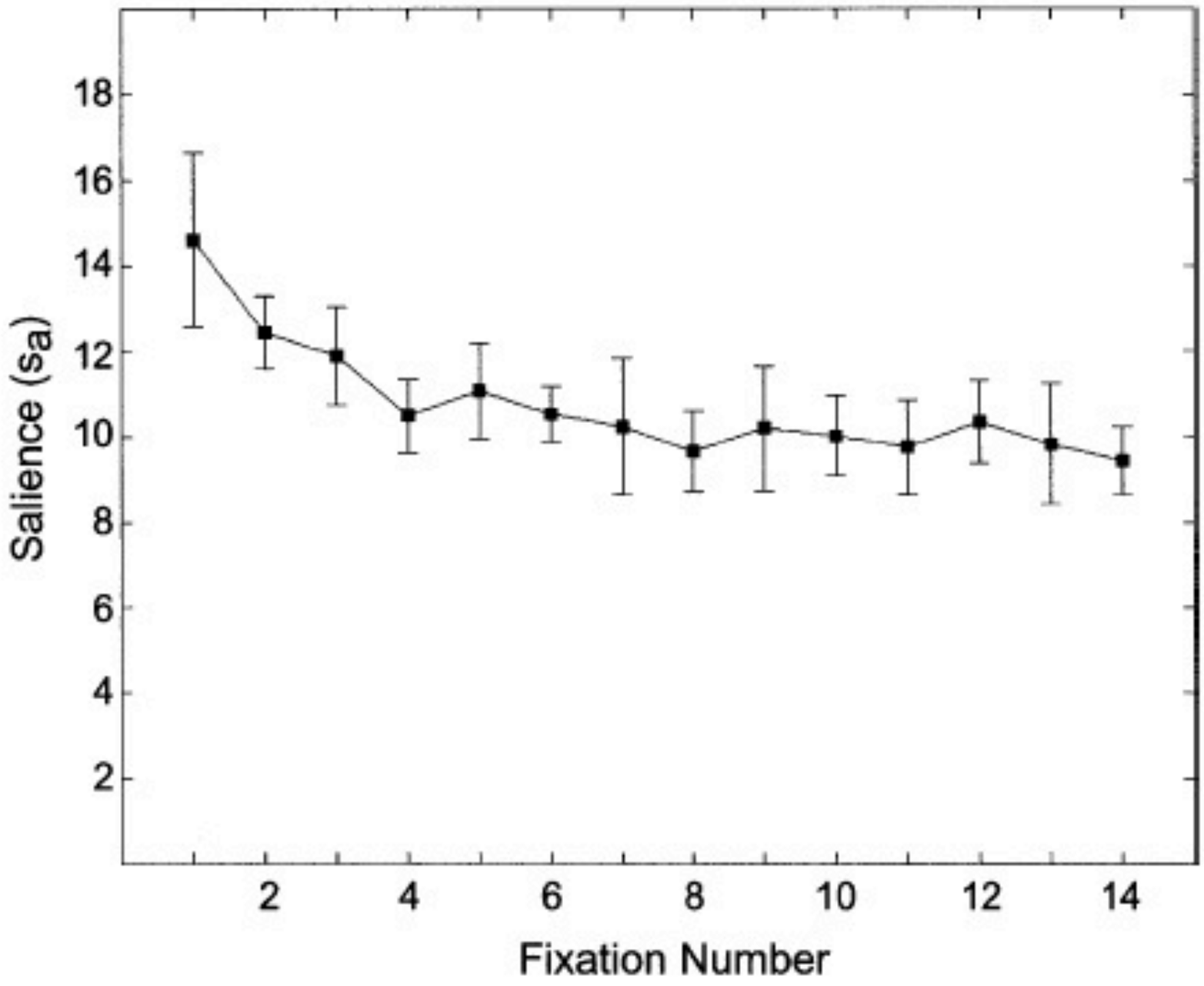


Fig. 5. The mean chance-adjusted salience for all databases is shown averaged across participants as a square where the errorbars represent plus or minus one standard error of the mean. Stimulus dependence is greatest for early fixations, but remains highly above chance levels throughout the trial.

Parkhurst et al. 2002

Gist

Gist of a Scene

- Biederman, 1981: from very brief exposure to a scene (120ms or less), we can already extract a lot of information about its global structure, its category (indoors, outdoors, etc) and some of its components.

- "riding the first spike:" 120ms is the time it takes the first spike to travel from the retina to IT!

- Thorpe, van Rullen: very fast classification (down to 27ms exposure, no mask), e.g., for tasks such as "was there an animal in the scene?"

Scene Context

- Scene-constrained targets detected faster, with fewer eye movements

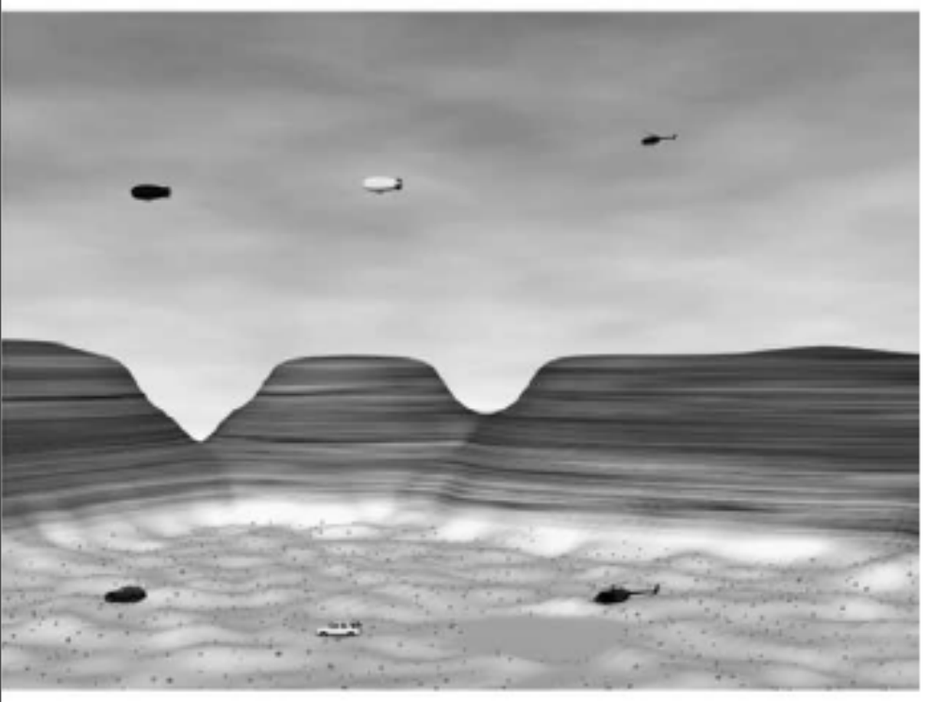
- Strategy

1st: check target-consistent regions

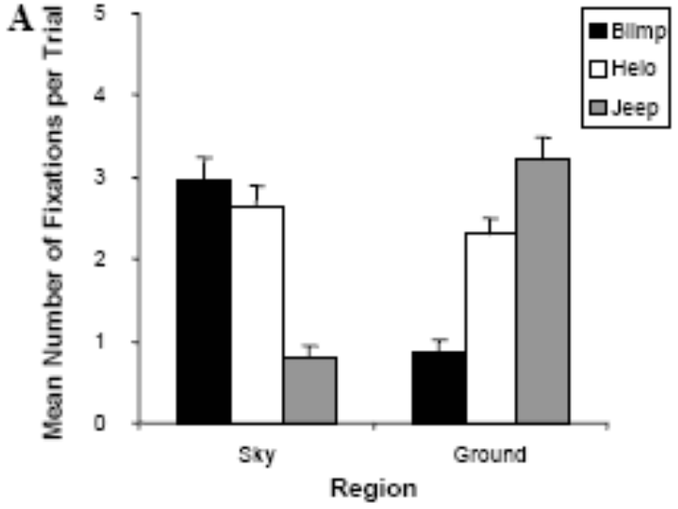
2nd: check target-inconsistent regions

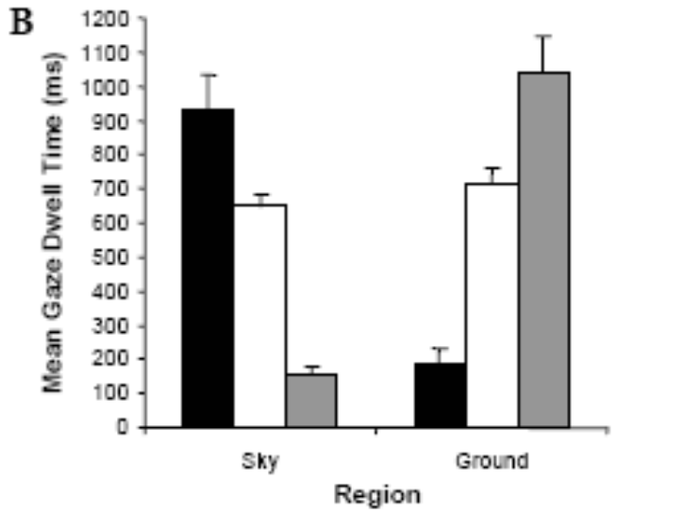
Scene Context

Neider and Zelinsky 2005

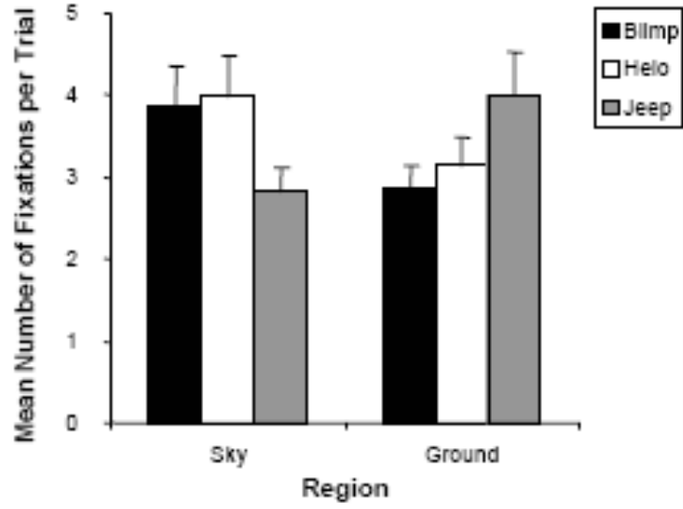


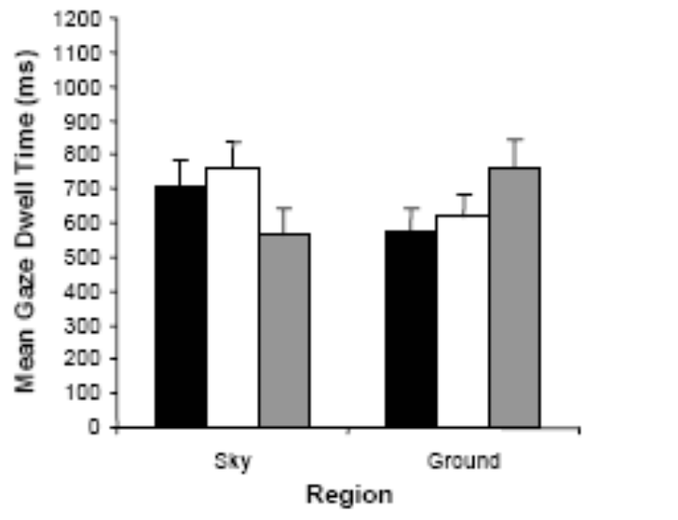
Target Presence





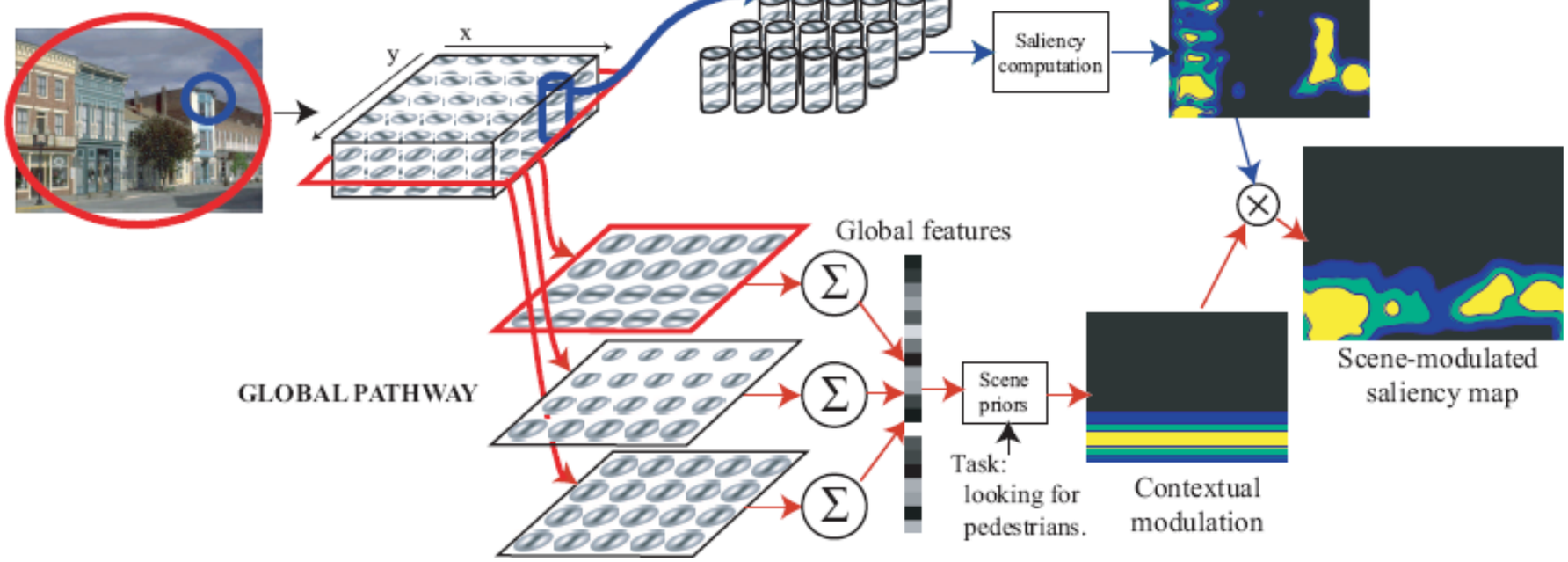
Target Absence





Scene Context

- "Gist" can provide image height prior



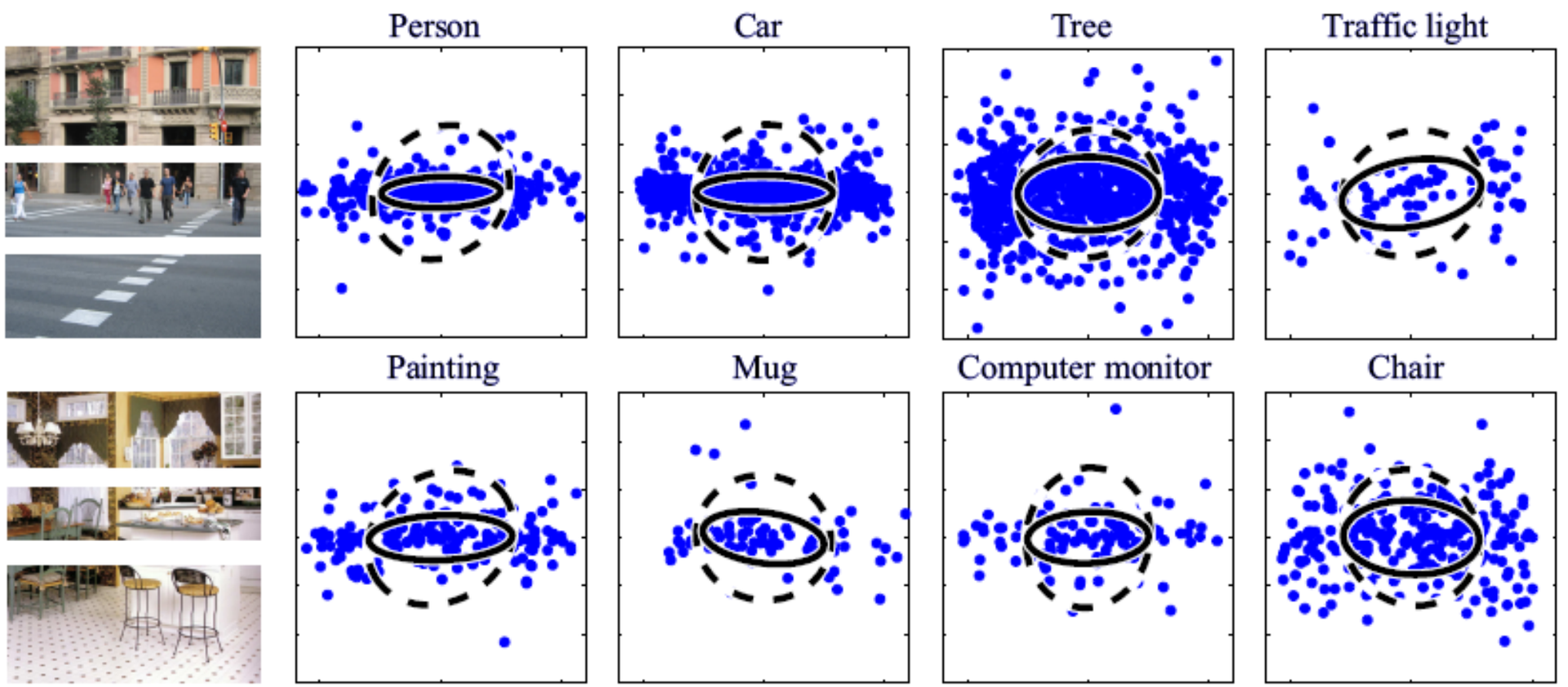
Aude Oliva

Antonio Torralba

Saliency = inverse probability ^(0.05) * gaussian

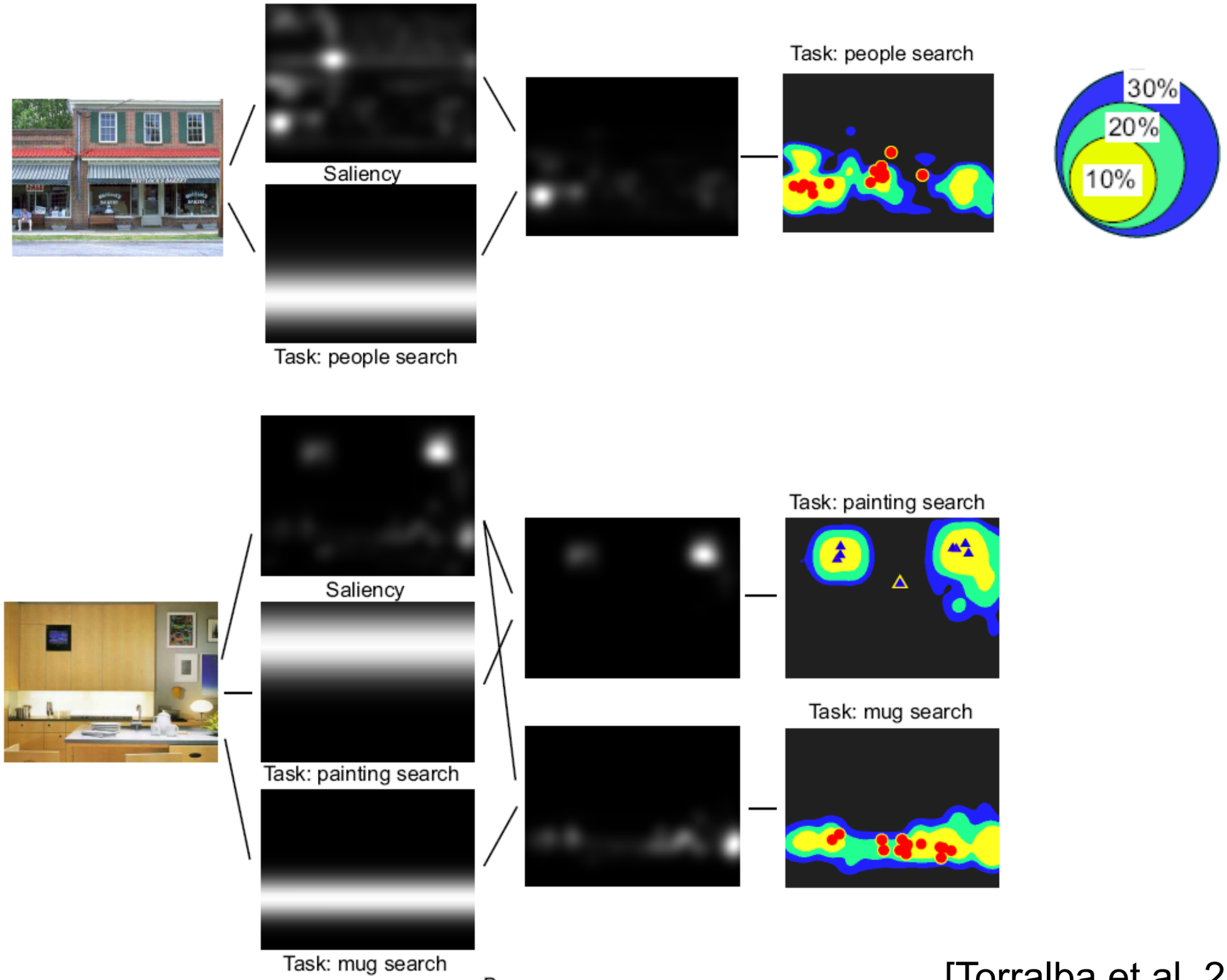
[Torralba et al. 2006]

Scene Context

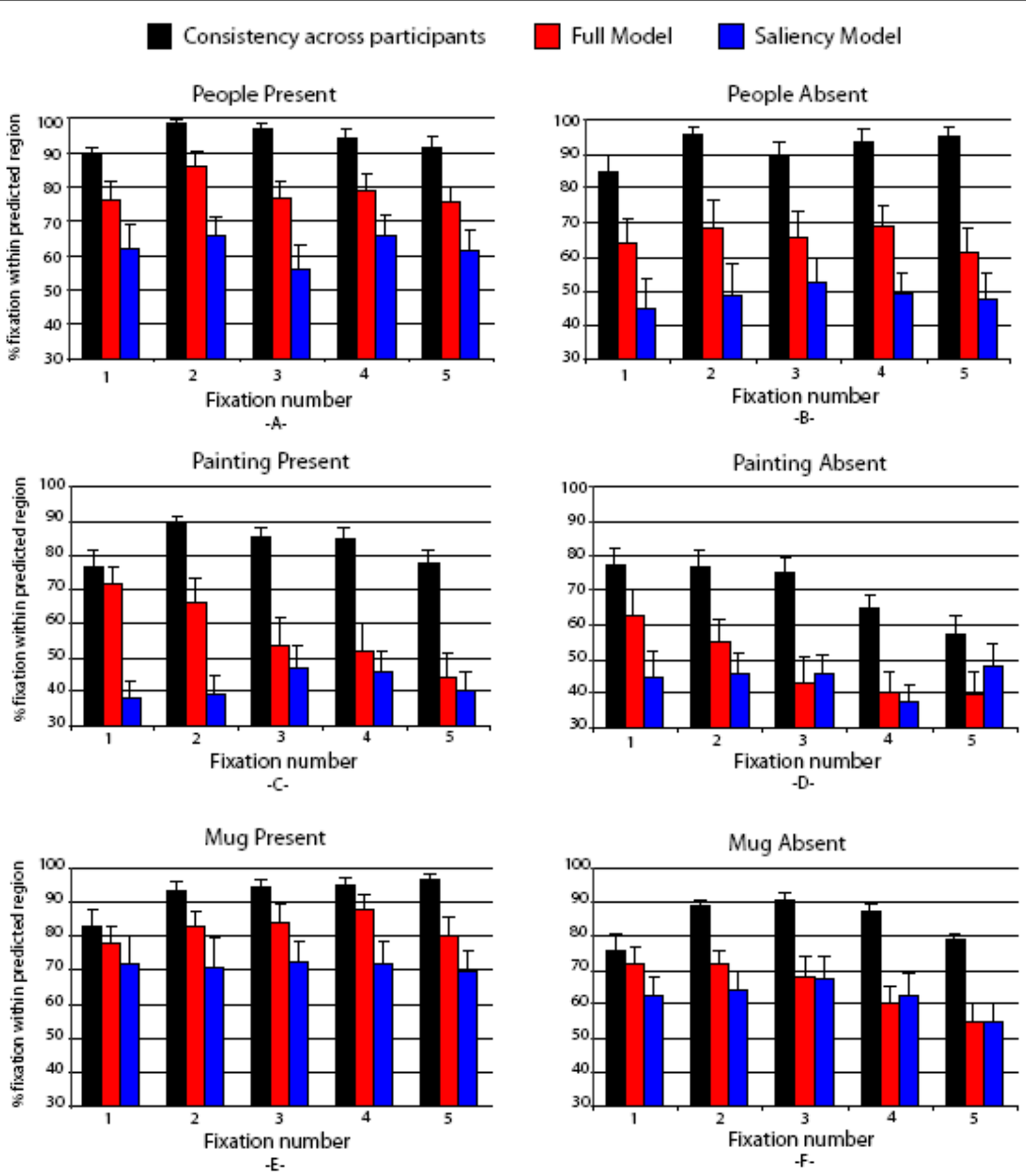


[Torralba et al. 2006]

Scene Context



[Torralba et al. 2006]



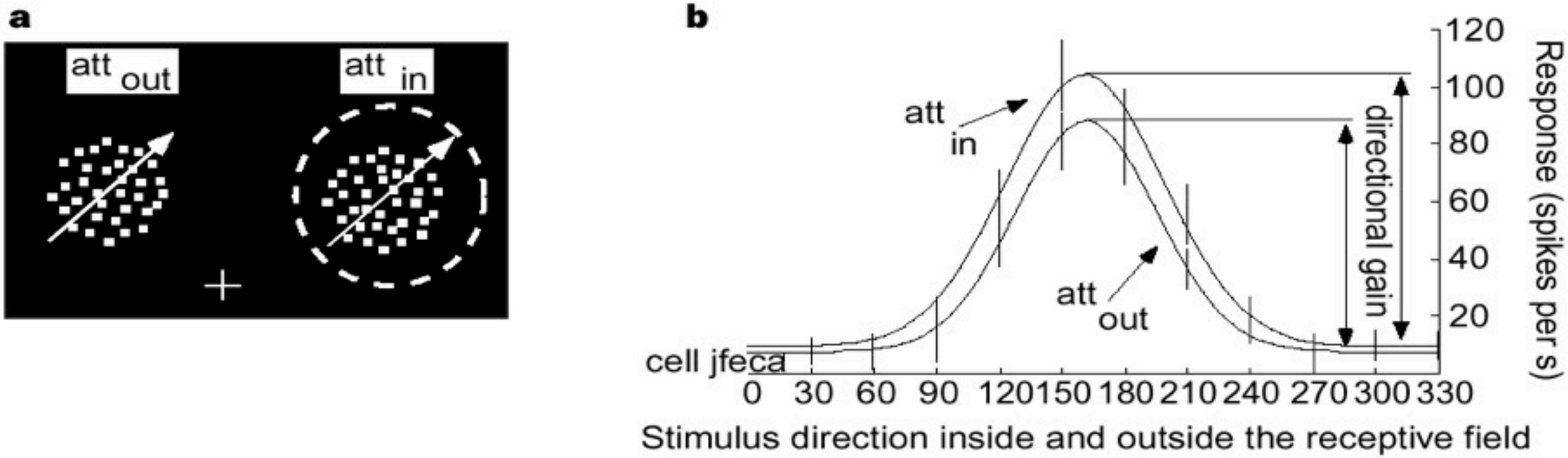
Scene Context

[Torralba et al. 2006]

Some Neurophysiology

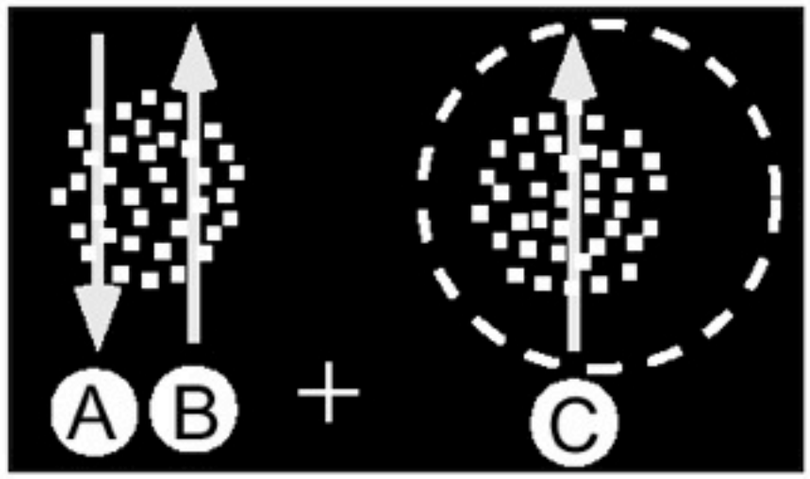
Attention enhances representation

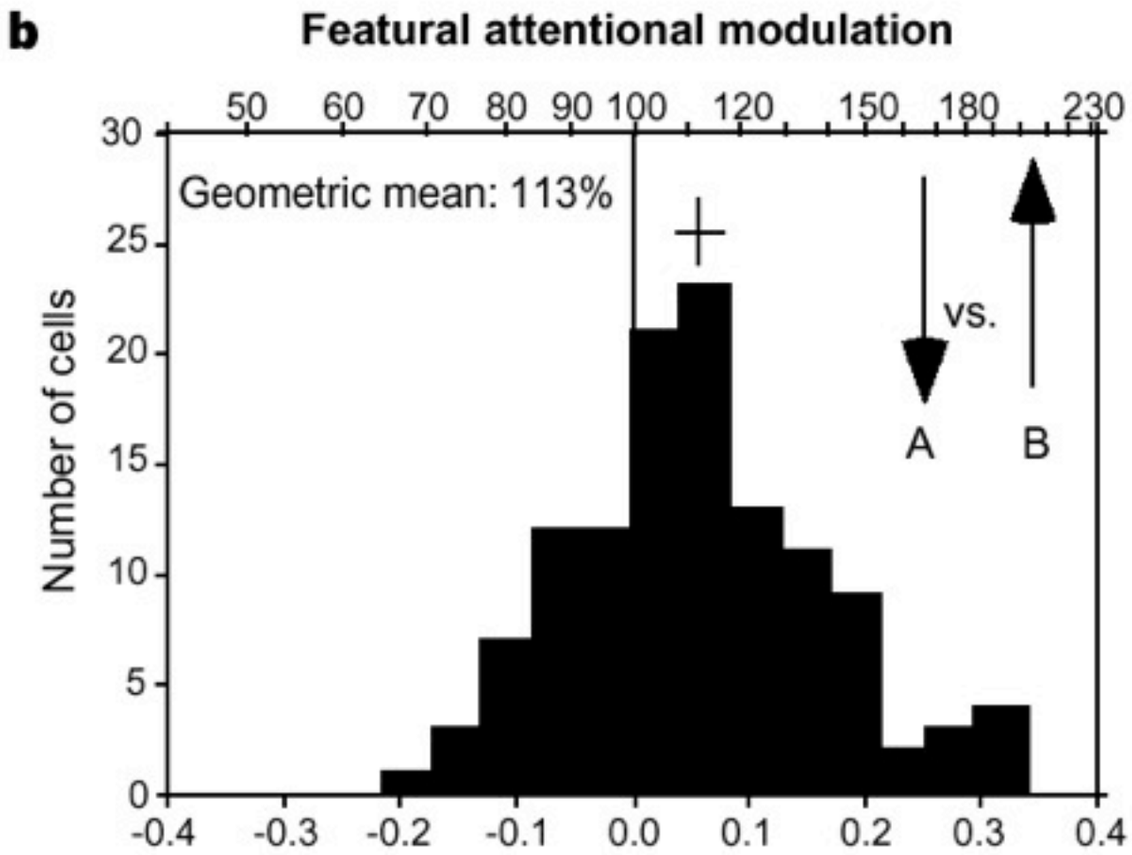
- Top-down attentional modulation: Early visual representation of stimuli enhanced if one voluntarily attends to
  - Stimulus location
  - Stimulus features (e.g., color, drift speed)

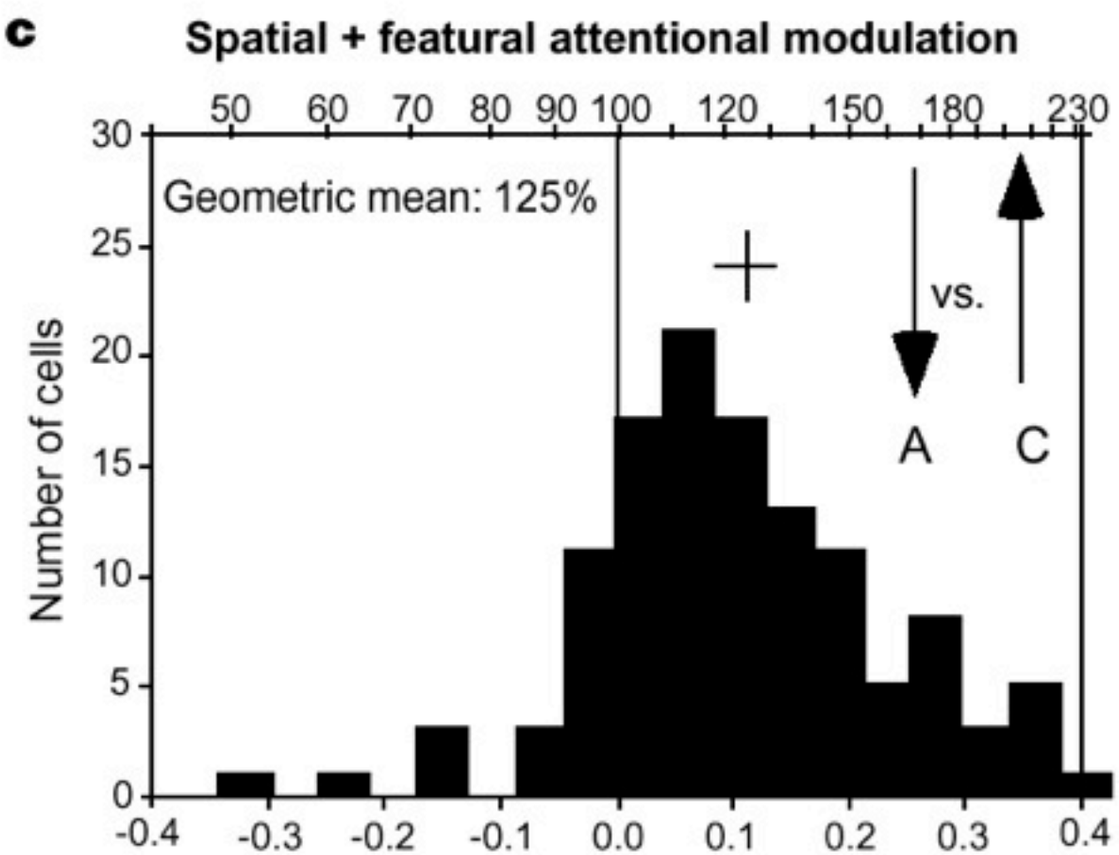


Spatial effect: neural activity higher when attention overlaps with neuron's RF area MT

Treue & Martinez-Trujillo, Nature 1999

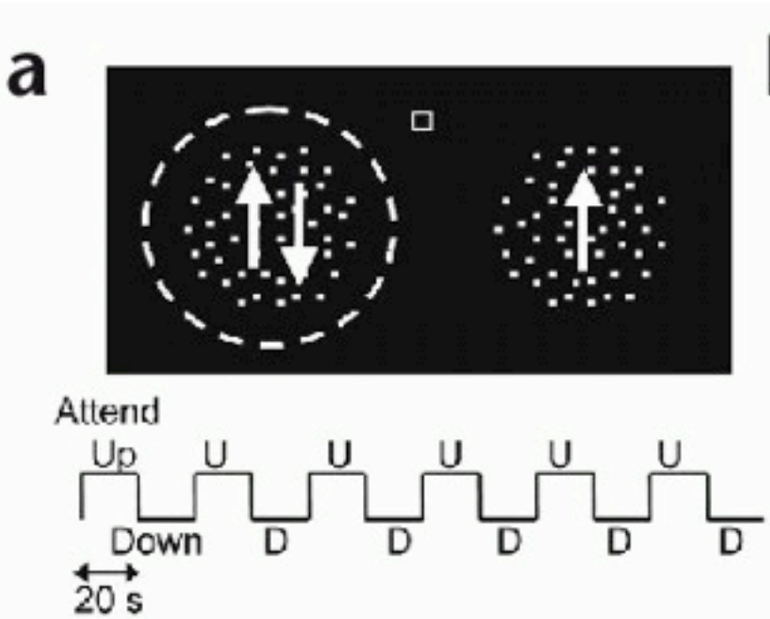


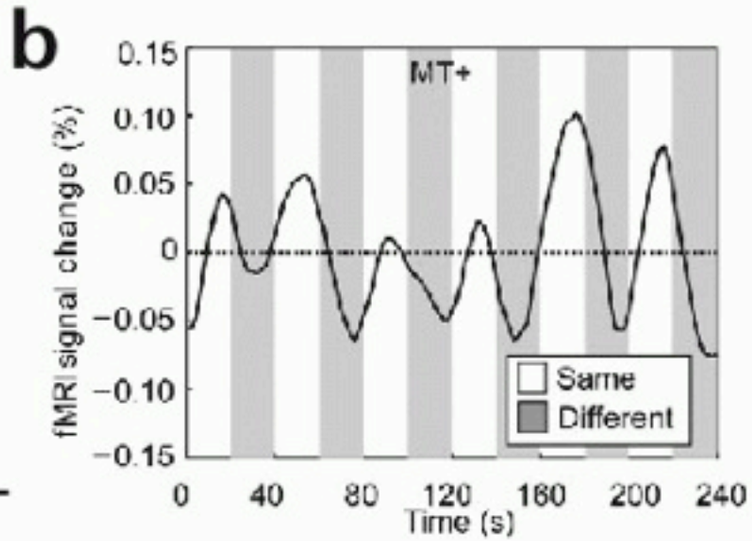


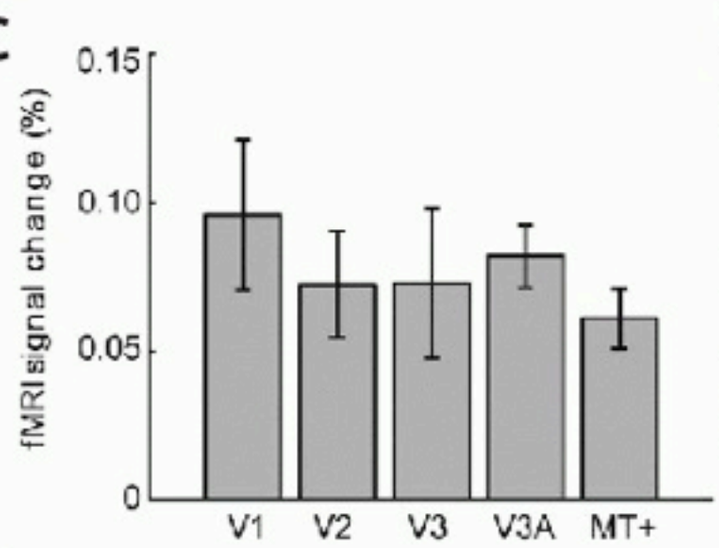


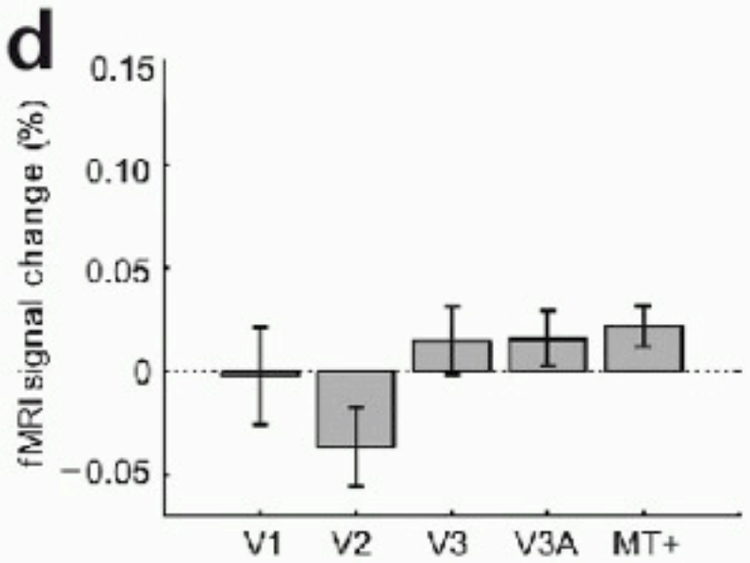
Feature effect: neural activity (on right side) higher when attending (on left side) to preferred direction of neuron on right side

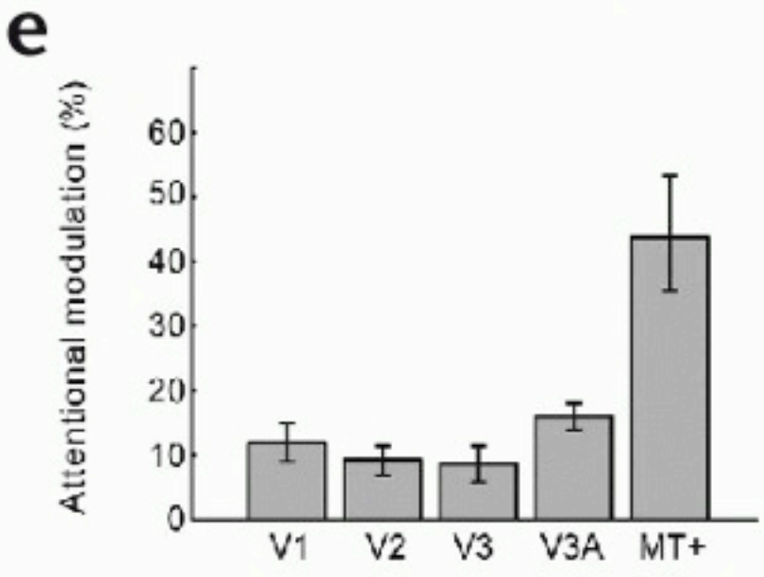
Treue & Martinez-Trujillo, Nature 1999





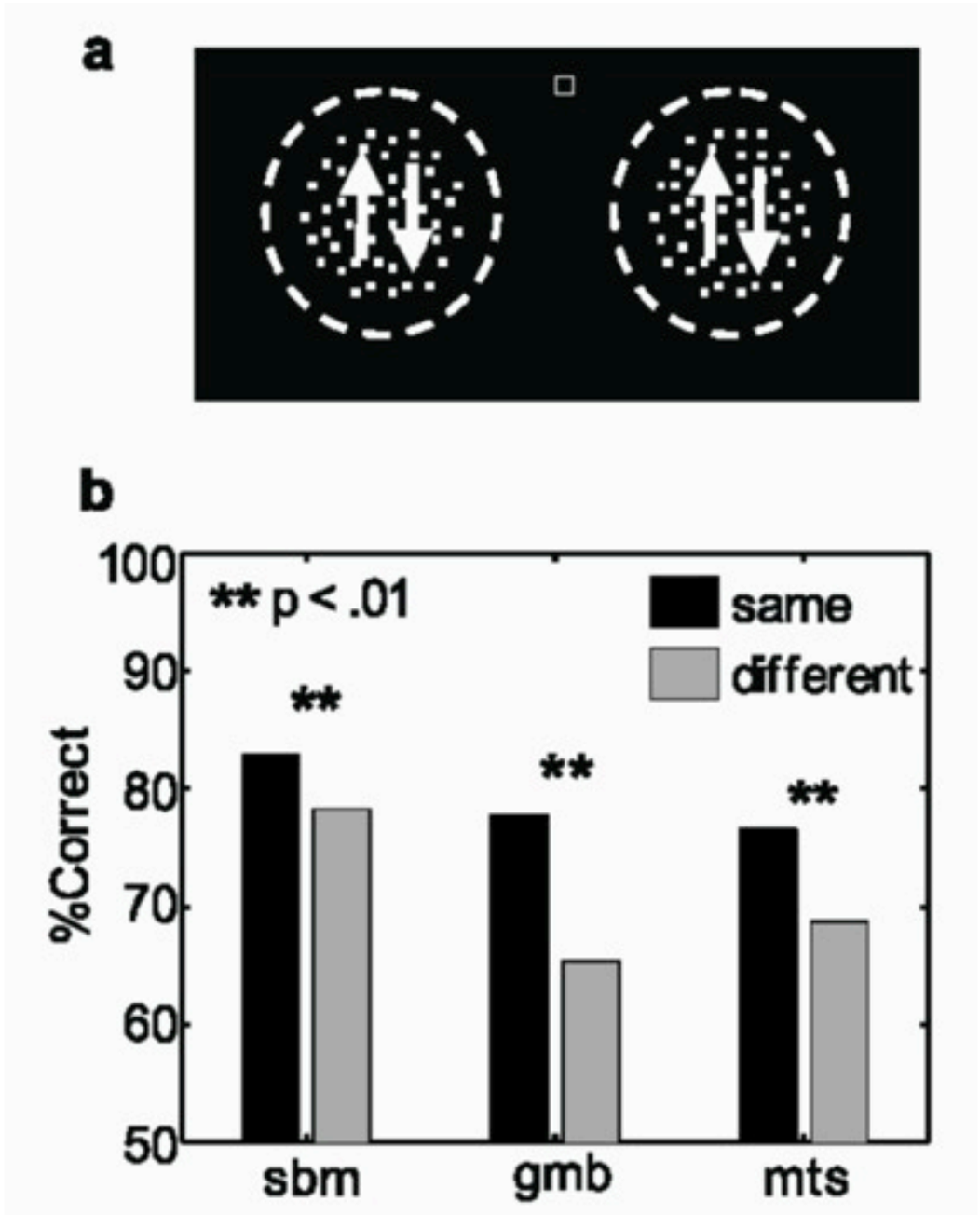


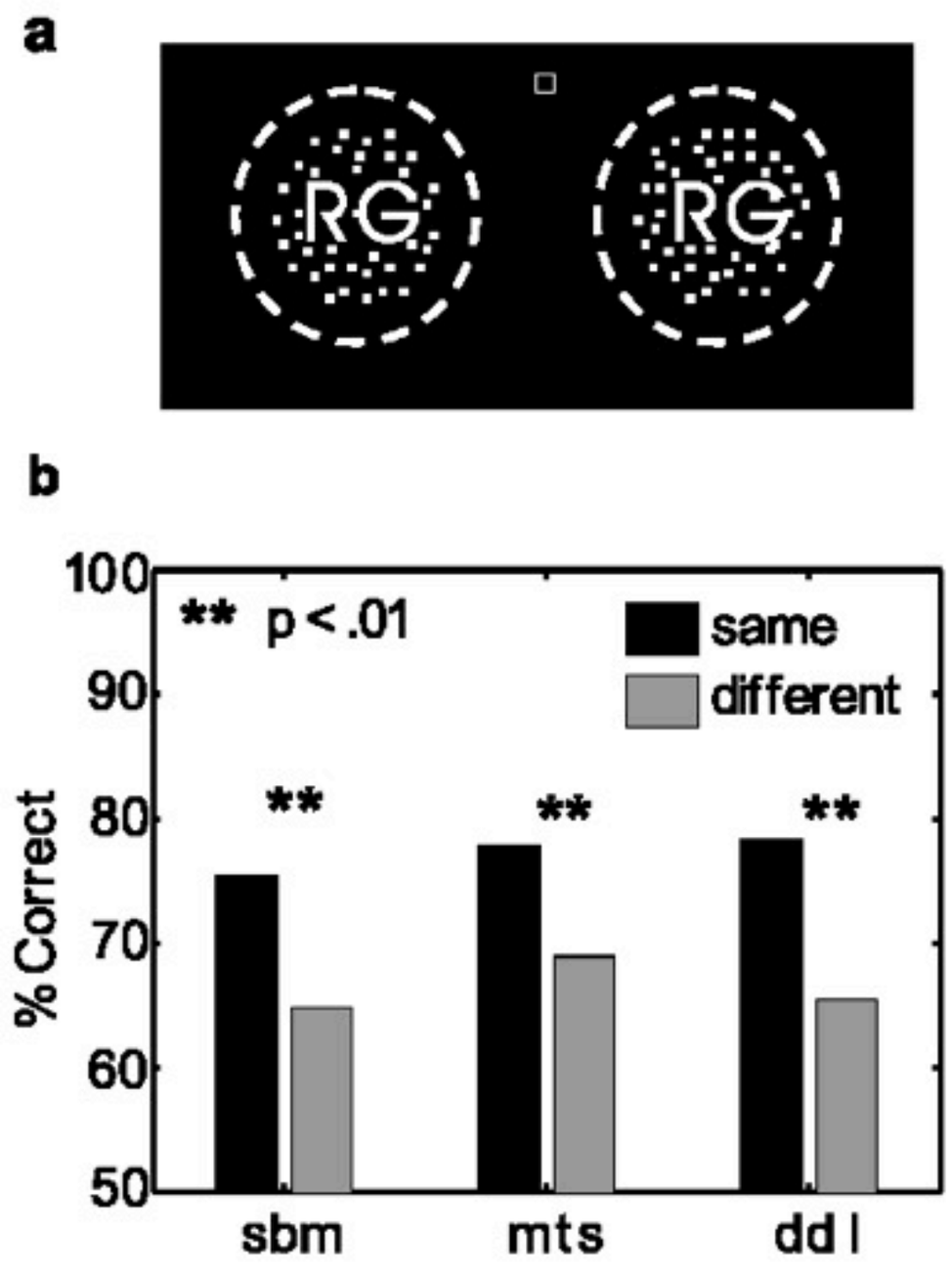




Featural effect: higher activation in MT+ (right side) when attending (left side) to same motion direction as right-side stimulus

Saenz, Buracas & Boynton, Nature Neuroscience 2002





Featural effect in dual-task: Better accuracy (% correct) when simultaneously discriminating drift speed (or luminance) in same direction (or color) on both sides of the display

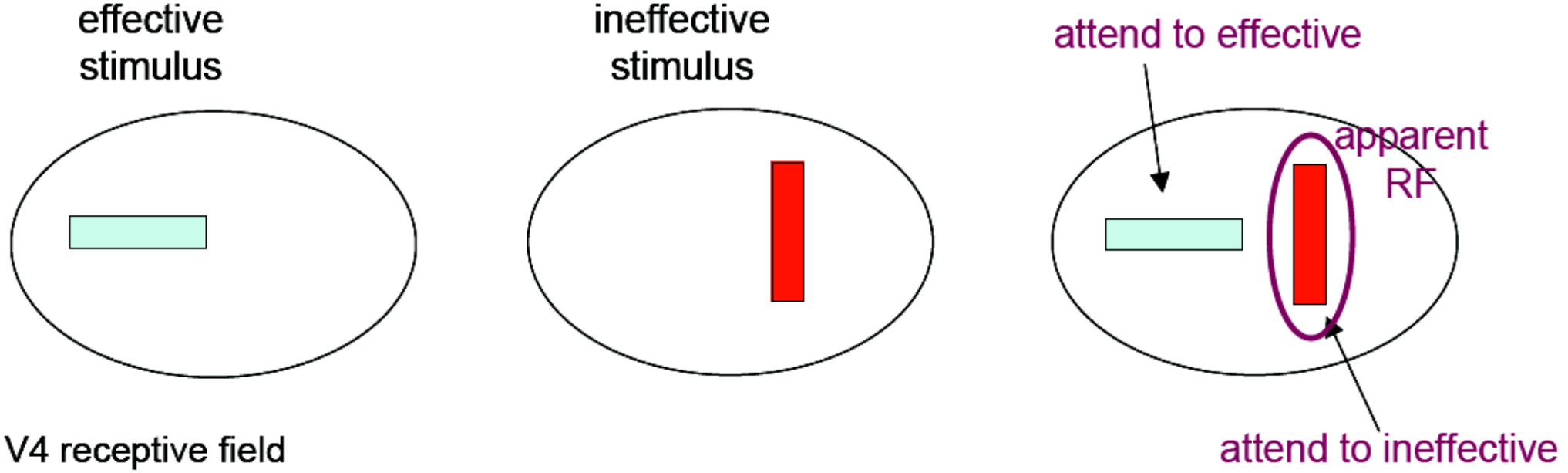
Saenz, Buracas & Boynton, Vision Research 2003

Shrinkwrap Model

Moran,J., Desimone,R. (1985). Selective Attention Gates Visual Processing in the Extrastriate Cortex, *Science 229*, 782-784.

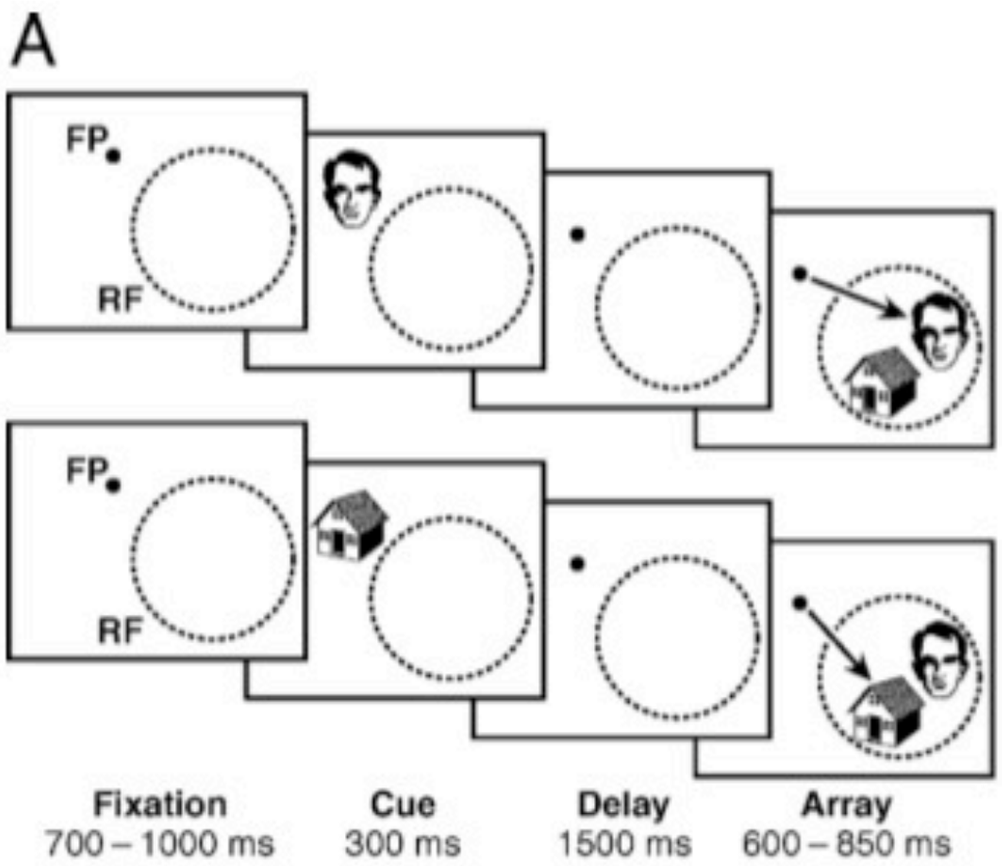
recording from V1, V4 and IT neurons of macaque
stimuli are effective and ineffective and placed inside and outside receptive field of recorded neuron

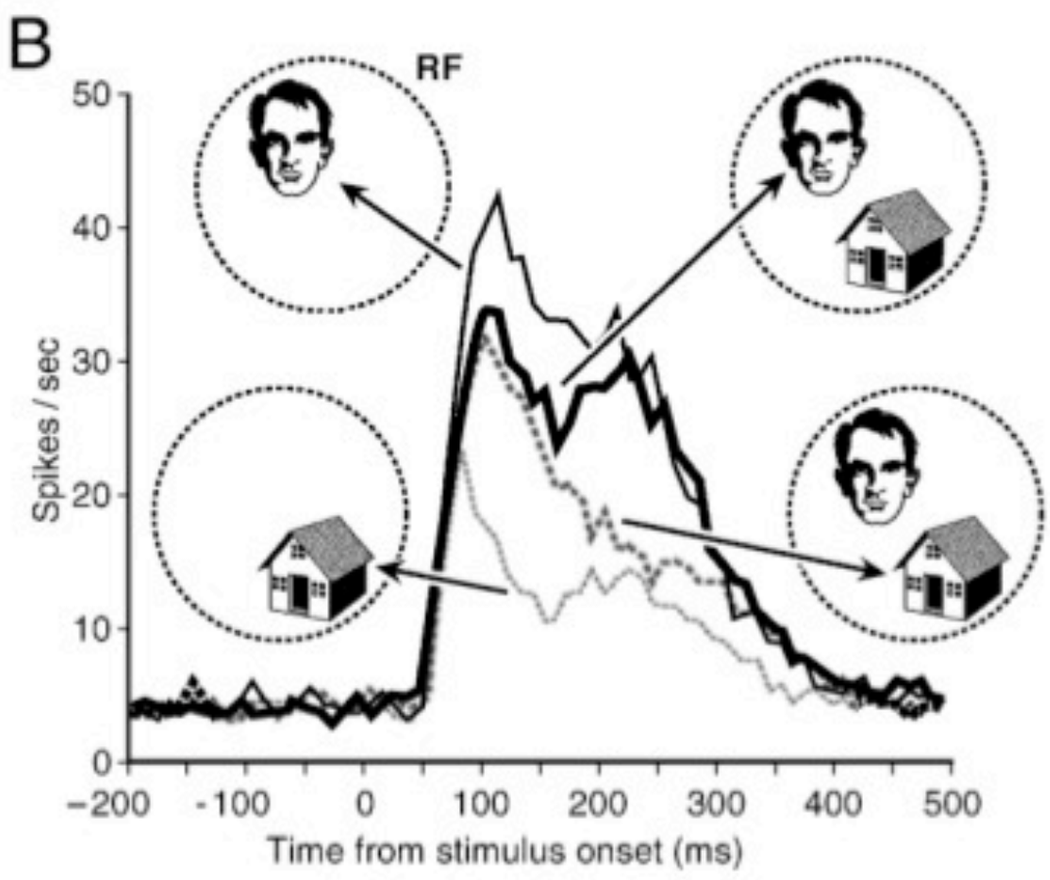
found largest effect for V4, smaller for IT and almost no effect for V1 neurons



for review see
Kastner S, Ungerleider LG. (2000). Mechanisms of visual attention in the human cortex. *Annu. Rev. Neurosci.* 23:315–41

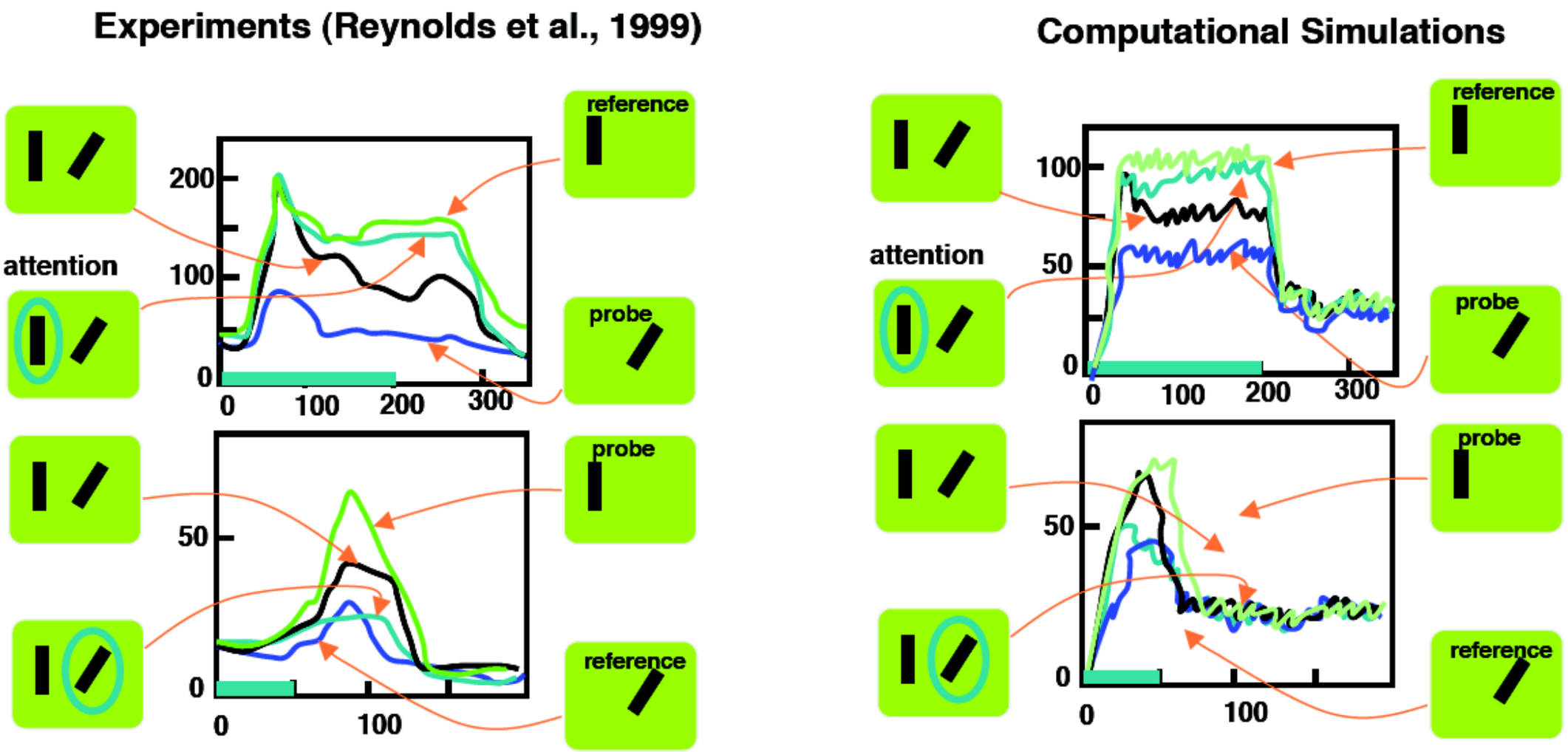
Adapted from Bruce, Rothenstein, and Tsotsos; ECCV Tutorial 2008





Chelazzi et al., 2001

Deco, Rolls, et al. 2001+



Adapted from Bruce, Rothenstein, and Tsotsos; ECCV Tutorial 2008

Saliency Map Locus

The neural correlate of the saliency map (if it exists at all) remains an open question:

Superior Colliculus

A.A. Kustov, D.L. Robinson, Shared neural control of attentional shifts and eye movements, Nature 384 (1996) 74–77.

R.M. McPeek, E.L. Keller, Saccade target selection in the superior colliculus during a visual search task, J. Neurophysiol. 88 (2002) 2019–2034.

G.D. Horwitz, W.T. Newsome, Separate signals for target selection and movement specification in the superior colliculus, Science 284 (1999) 1158–1161.

LGN

C. Koch, A theoretical analysis of the electrical properties of an X-cell in the cat LGN: does the spine-triad circuit subserve selective visual attention? AI Memo 787, MIT, February, 1984.

S.M. Sherman, C. Koch, The control of retinogeniculate transmission in the mammalian lateral geniculate nucleus, Exp. Brain Res. 63 (1986) 1–20.

V1

Z. Li, A saliency map in primary visual cortex, Trends Cog. Sci. 6 (1) (2002) 9–16.

V1 and V2

D.K. Lee, L. Itti, C. Koch, J. Braun, Attention activates winner-take-all competition among visual filters, Nat. Neurosci. 2 (4) (1999) 375–381.

Pulvinar

S.E. Petersen, D.L. Robinson, J.D. Morris, Contributions of the pulvinar to visual spatial attention, Neuropsychologia 25(1987) 97–105.

M.I. Posner, S.E. Petersen, The attention system of the human brain, Annu. Rev Neurosci. 13 (1990) 25–42.

Frontal Eye Fields

D.L. Robinson, S.E. Petersen, The pulvinar and visual salience, Trends Neurosci.15 (4) (1992) 127– 132.

Parietal Cortex

K.G. Thompson, N.P. Bichot, J.D. Schall, Dissociation of visual discrimination from saccade programming in macaque frontal eye field, J. Neurophysiol. 77 (1997) 1046–1050

Adapted from Bruce, Rothenstein, and Tsotsos; ECCV Tutorial 2008

Some References

- Borji and Itti PAMI 2013
- Baluch and Itti 2012
- Navalpakkam and Itti VR, 2005
- Neurobiology of Attention, Editors Itti, Rees & Tsotsos, Elsevier Press, 2005
- Book edited by Christof Koch
- Maria Carasco
- Miguel Eckstein
- Ken Nakayama
- Alex Toet
- Shultz
- Ben Tatler
- Kastner S, Ungerleider LG. (2000). Mechanisms of visual attention in the human cortex. Annu. Rev. Neurosci. 23:315–41
- Henderson
- Hayhoe
- Models of Overt Attention (Geisler and Cormac)
- Itti and Koch NN

See recent spacial editions of JOV and VR for review papers on attention and eye movements

Some topics

- Eye movements
- Covert Attention
- Auditory Attention
- Overt Attention
- Visual Search
- Salience
- Optical Metaphors
- Neural Modulation
- Control of Attention
- Attention and Recognition, Binding

Related Fields

- Active vision
- Active learning
- Ego centric vision
- First person vision
- Feature learning
- Points/Regions of Interest Detection
- Feature Learning
- Category Learning
- Optimal Search
- Optimal foraging

Active Vision

"Active sensing is the problem of intelligent control strategies applied to the data acquisition process which will depend on the current state of data interpretation including recognition." Ruzena Bajcsy 1985

- to move to fixation point/plane or to track motion
- to see a portion of the visual field otherwise hidden due to occlusion
  - manipulation
  - viewpoint change
- to see a larger portion of the surrounding visual world
  - exploration
- to compensate for spatial non-uniformity of a processing mechanism
  - foveation
- to increase spatial resolution or to focus
  - sensor zoom or observer motion
  - adjust camera depth of field, stereo vergence
- to disambiguate or to eliminate degenerate views
  - induced motion (kinetic depth)
  - lighting changes (photometric stereo)
  - viewpoint change
- to achieve a "pathognomonic" view
  - viewpoint change
- to complete a task
  - multiple fixations

Focus of the next talk

Adapted from Bruce, Rothenstein, and Tsotsos; ECCV Tutorial 2008

Review: Finding "interesting" information

- In principle, very complex task:
  - Need to attend to all objects in scene?
  - Then recognize each attended object?
  - Finally evaluate set of recognized objects against behavioral goals?

- In practice, survival depends on ability to quickly locate and identify important information.

- Need to develop simple **heuristics** or approximations:
  - **So far: bottom-up** guidance towards salient locations
  - **Next: top-down** guidance towards task-relevant locations
  - **Next:** applications?