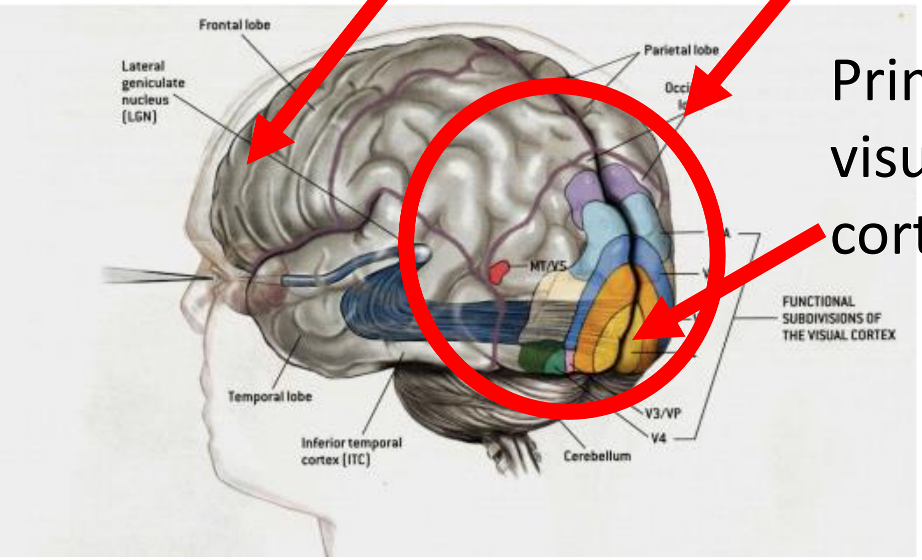
A new path to understanding vision

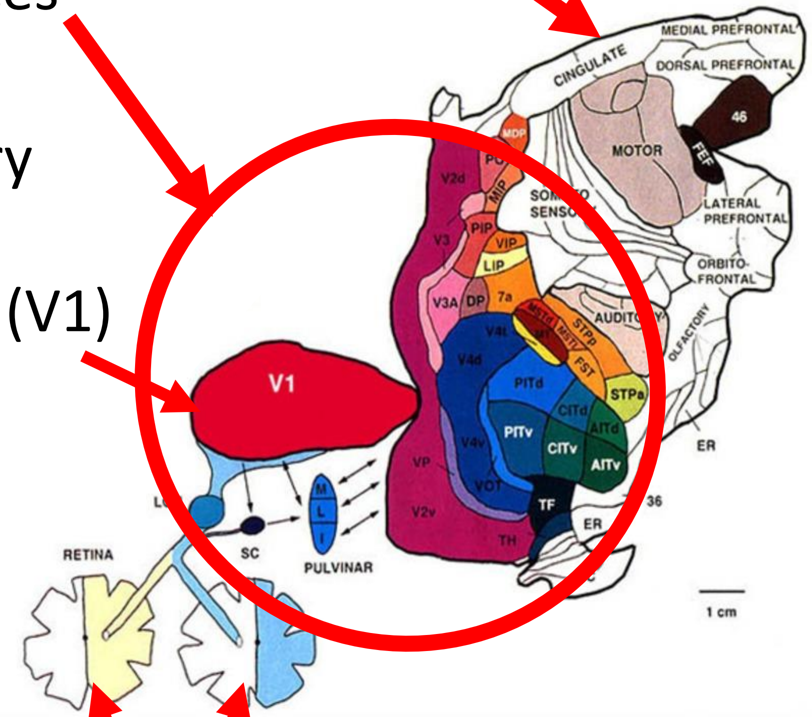
from the perspective of the primary visual cortex

Frontal brain areas

Visual cortices

Primary visual cortex (V1)





Retina

Li Zhaoping

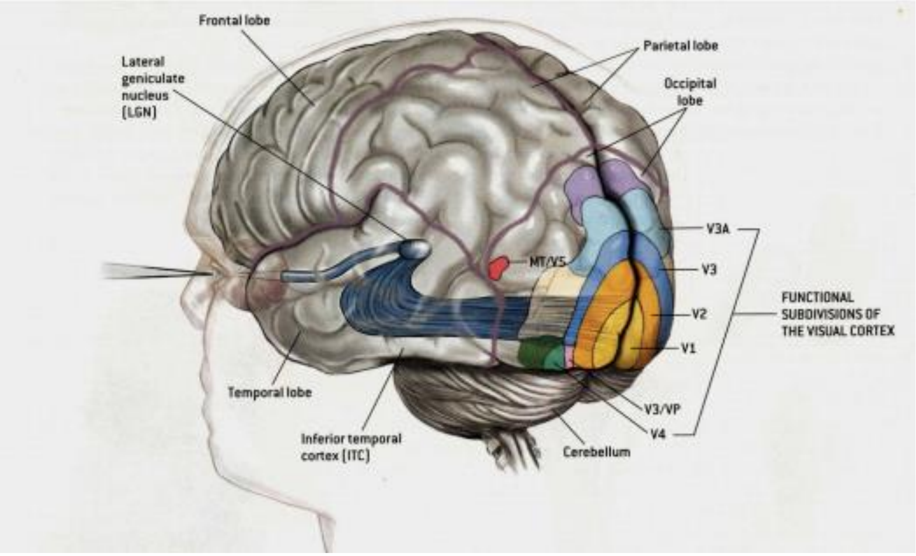
Sep. 14, 2018,
KITP, Santa Barbara, CA, USA

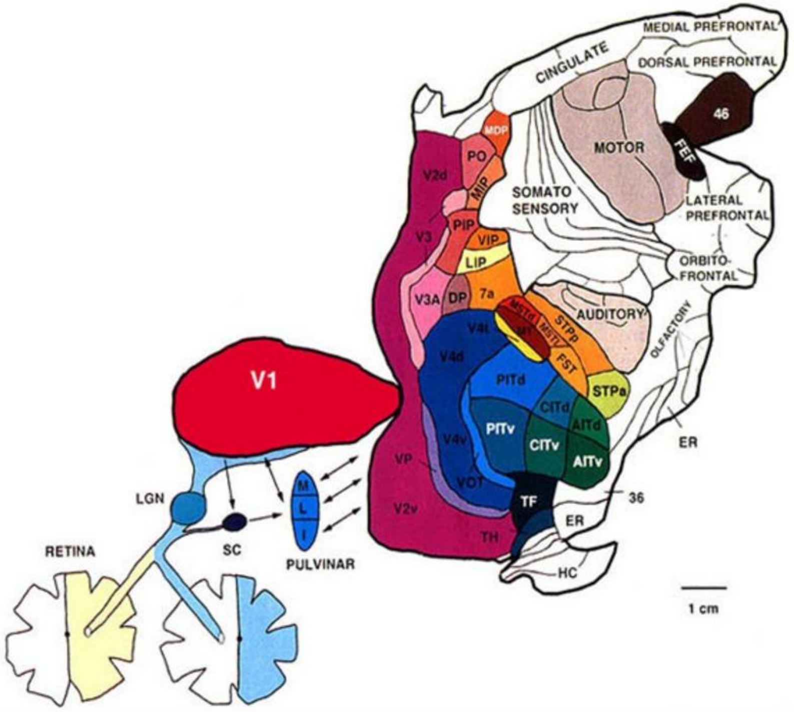
A new path to understanding vision

Traditional paths to understanding vision

(1) Low level vision, mid-level vision, high-level vision

(2) David Marr: primal sketch, 2.5 d sketch, 3-d model.





Li Zhaoping

July 16, 2018, presented at APCV 2018, HangZhou, China

Talk outline

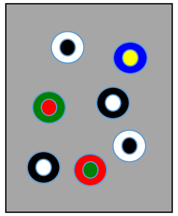
(1) The functional role of the primary visual cortex (V1)

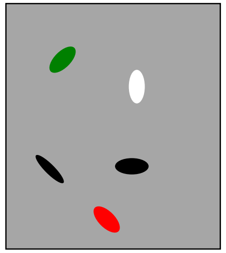
(2) In light of V1’s role →
a new plan to understanding vision

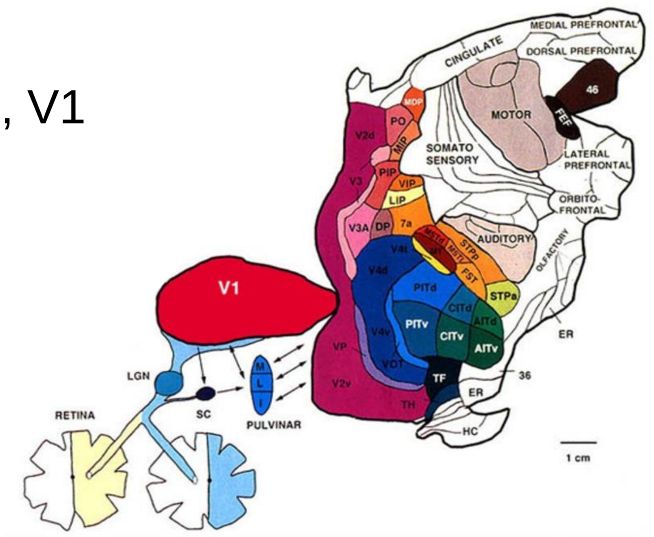
(3) A first example study in this new plan

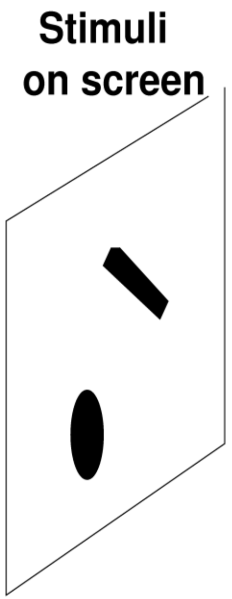
The primary visual cortex (V1)

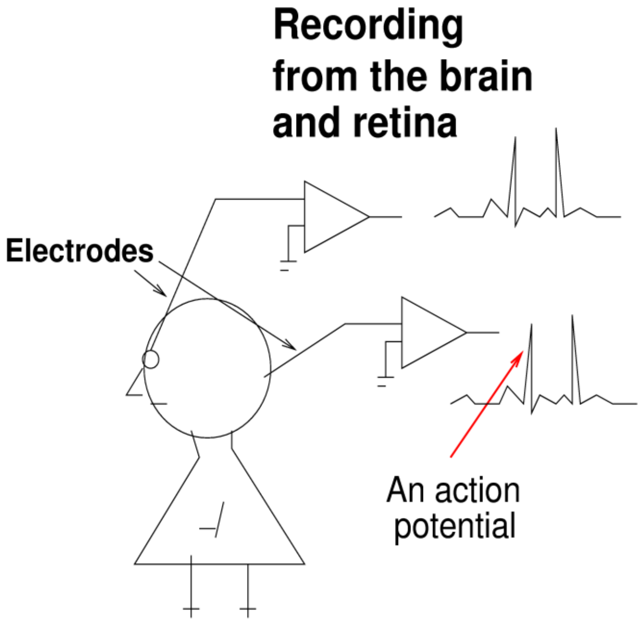
1953, Stephen Kuffler, retina, 1959-- Hubel and Wiesel, V1





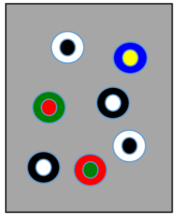


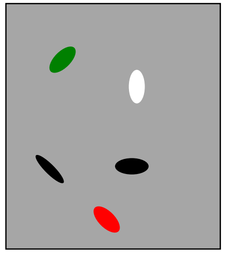


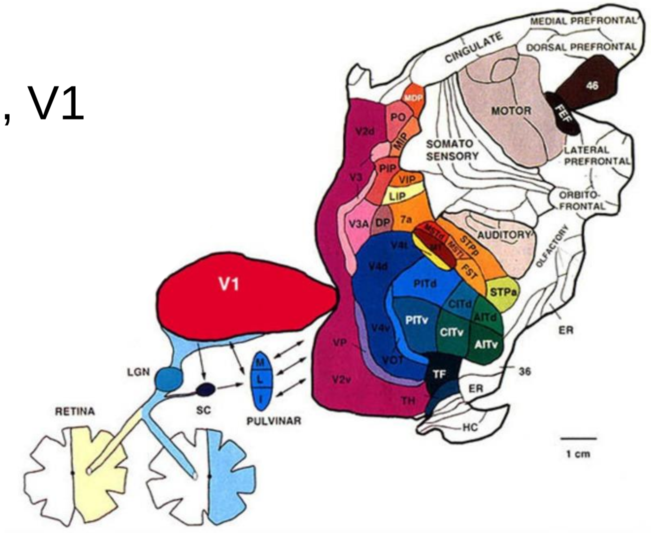


The primary visual cortex (V1)

1953, Stephen Kuffler, retina,   1959-- Hubel and Wiesel, V1







Then …

Experimentally: V1 and beyond

Theoretical/modelling, Reichardt, Marr, etc.

2005: How close are we to understand V1?   Olshausen and Field 2005

Do we really know what the early visual system does?

Carandini, Demb, Mante, Tolhurst, Dan, Olshausen, Gallant, Rust, 2005

Standard models of V1 neural receptive field (combining filtering, rectification, squaring, normalization) **captures only 15-35% of the variances in V1 responses**.

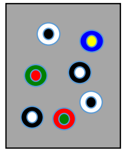
2012, David Hubel, in answer to “What Do You Feel Are the Next Big Questions in the Field?”

”**We have some idea … for the retina, the lateral geniculate body, and the primary visual cortex, but that’s about it**.”

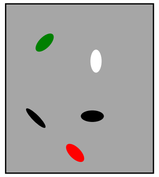
(Hubel & Wiesel 2012, Neuron)

The primary visual cortex (V1)

1953, Stephen Kuffler, retina,

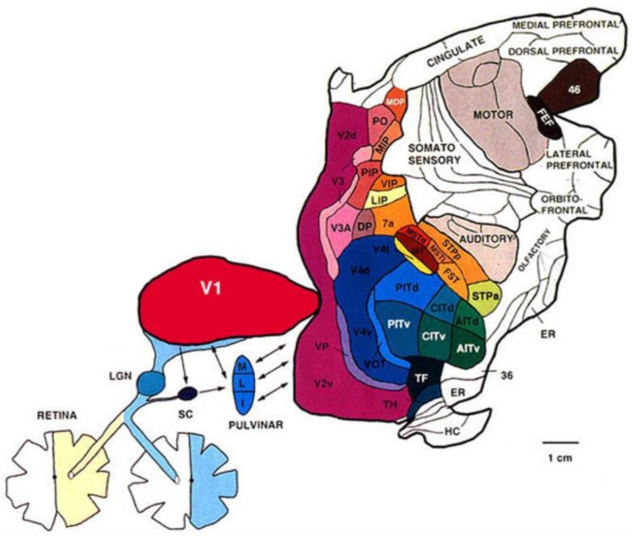


1959-- Hubel and Wiesel, V1



Then …

Experimentally: V1 and beyond
Theoretical/modelling, Reichardt, Marr, etc.



2005: How close are we to understand V1? Olshausen and Field 2005

Do we really know what the early visual system does?
Carandini, Demb, Mante, Tolhurst, Dan, Olshausen, Gallant, Rust, 2005

Standard models of V1 neural receptive field combining linear filtering, rectification and squaring, and response normalization **captures only 15-35% of the variances in V1 responses**.

2012, David Hubel, in answer to "What Do You Feel Are the Next Big Questions in the Field?"

"**We have some idea … for the retina, the lateral geniculate body, and the primary visual cortex, but that's about it**."
(Hubel & Wiesel 2012, Neuron)

Questions:

Is a lack of understanding of V1 hindering our progress beyond V1?

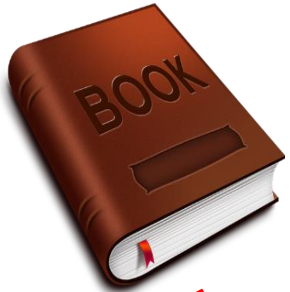
Physiologically

Functionally (in behaviour)?

Information bottlenecks in the visual pathway:

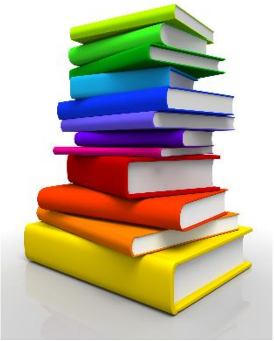
"To be or not to be,
This is the question .."

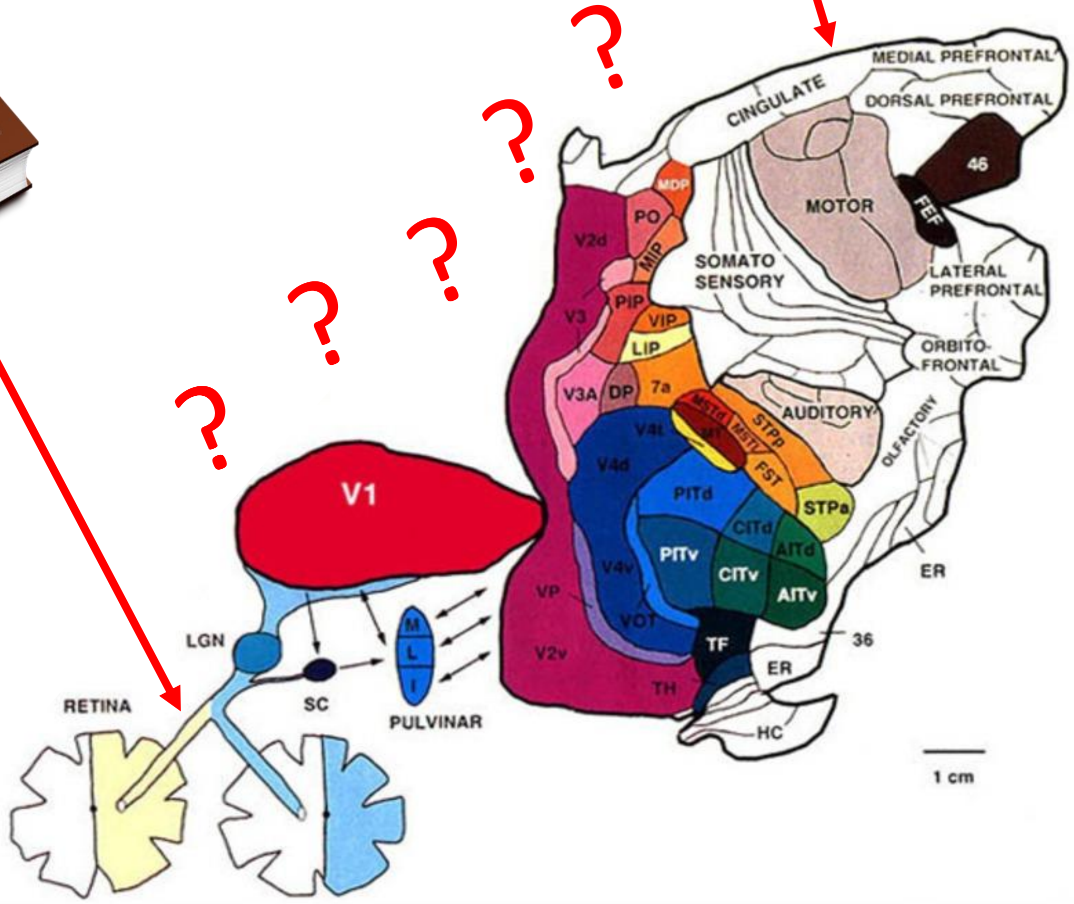
10^7 bits/second
~ 10^6 neurons ,
~10 spikes/neuron
~1 bit/spike



40 bits/second
(Sziklai, 1956)

10^9 bits/second (Kelly 1962)
~ 25 frames/second,
2000x2000 pixels,
1 byte/pixel





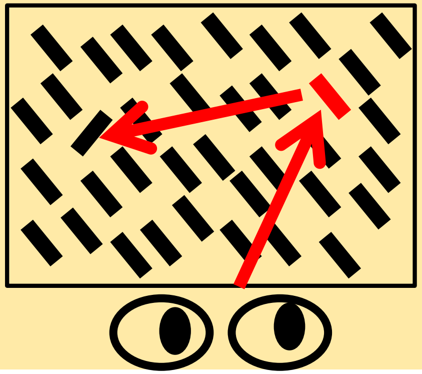
Information bottlenecks in the visual pathway:

We are nearly blind!

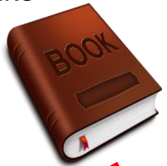
Vision ~ Looking (selecting) + Seeing

top-down vs. bottom-up selection

Task: find a uniquely oriented bar



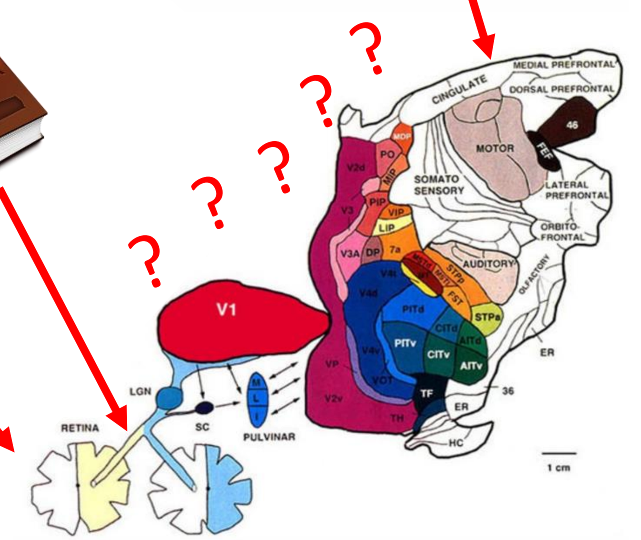
10^7 bits/second
~ 10^6 neurons ,
~10 spikes/neuron
~1 bit/spike



“To be or not to be,
This is the question ..”

40 bits/second
(Sziklai, 1956)

10^9 bits/second
(Kelly 1962)
~ 25 frames/second,
2000x2000 pixels,
1 byte/pixel



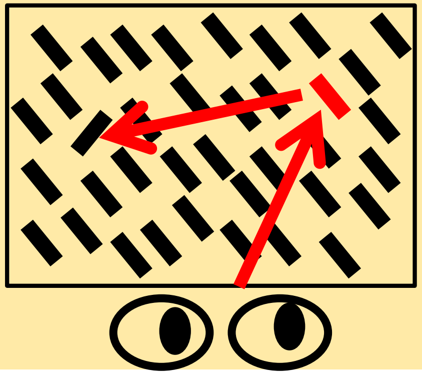
Information bottlenecks in the visual pathway:

Questions:
which brain areas are doing the bottom-up selection?

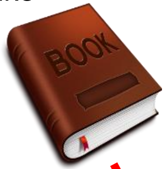
Frontal? Parietal?

top-down vs. bottom-up selection

Task: find a uniquely oriented bar



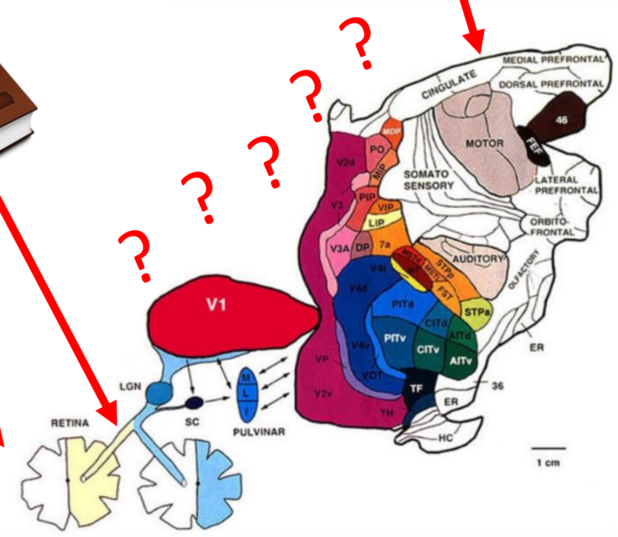
10^7 bits/second
~ 10^6 neurons ,
~10 spikes/neuron
~1 bit/spike



10^9 bits/second
(Kelly 1962)
~ 25 frames/second,
2000x2000 pixels,
1 byte/pixel

"To be or not to be,
This is the question .."

40 bits/second
(Sziklai, 1956)

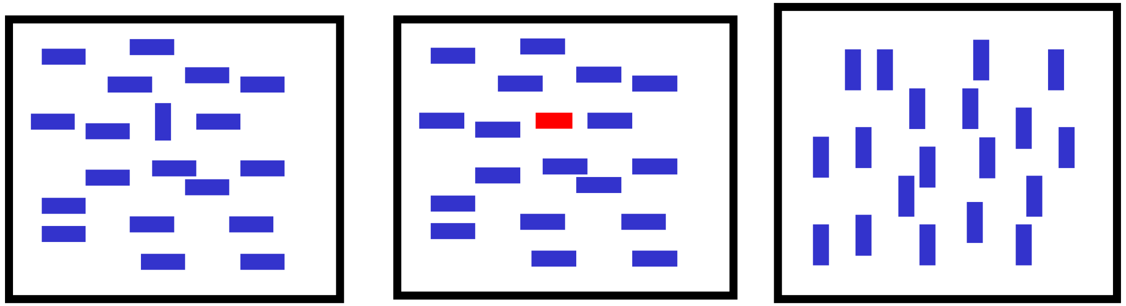


Information bottlenecks in the visual pathway:

Questions:
which brain areas are doing the bottom-up selection?

Frontal? Parietal?

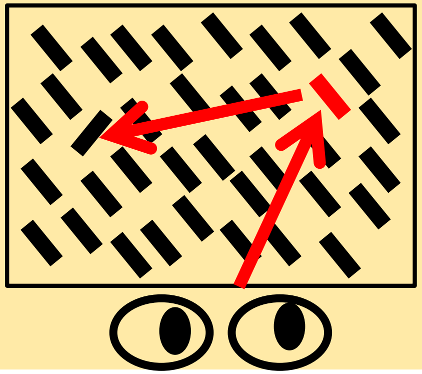
Saliency regardless of visual features



Koch & Ullman 1985, Itti & Koch 2001, etc

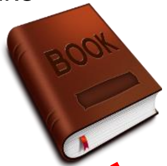
top-down vs. bottom-up selection

Task: find a uniquely oriented bar



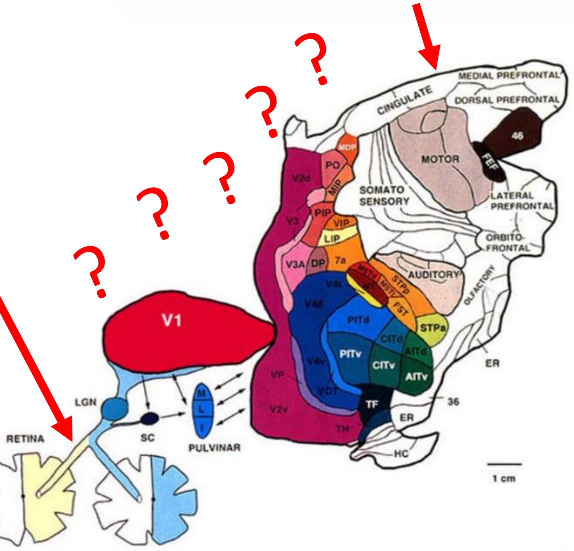
"To be or not to be,
This is the question .."

10^7 bits/second
~ 10^6 neurons ,
~10 spikes/neuron
~1 bit/spike



40 bits/second
(Sziklai, 1956)

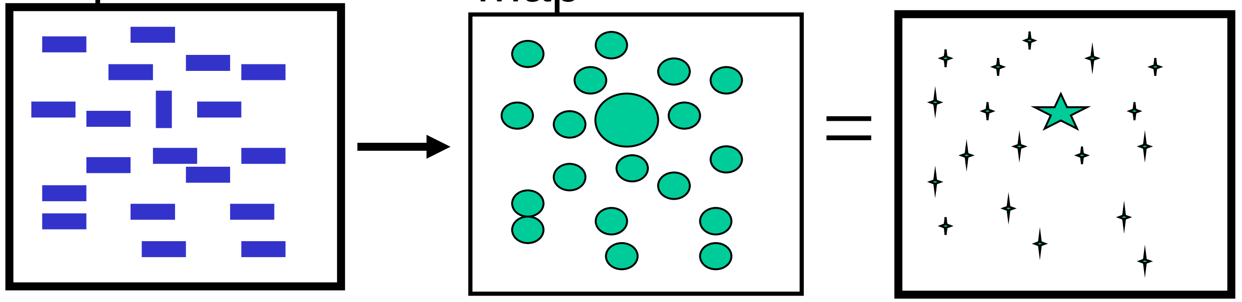
10^9 bits/second
(Kelly 1962)
~ 25 frames/second,
2000x2000 pixels,
1 byte/pixel



The V1 Saliency Hypothesis:

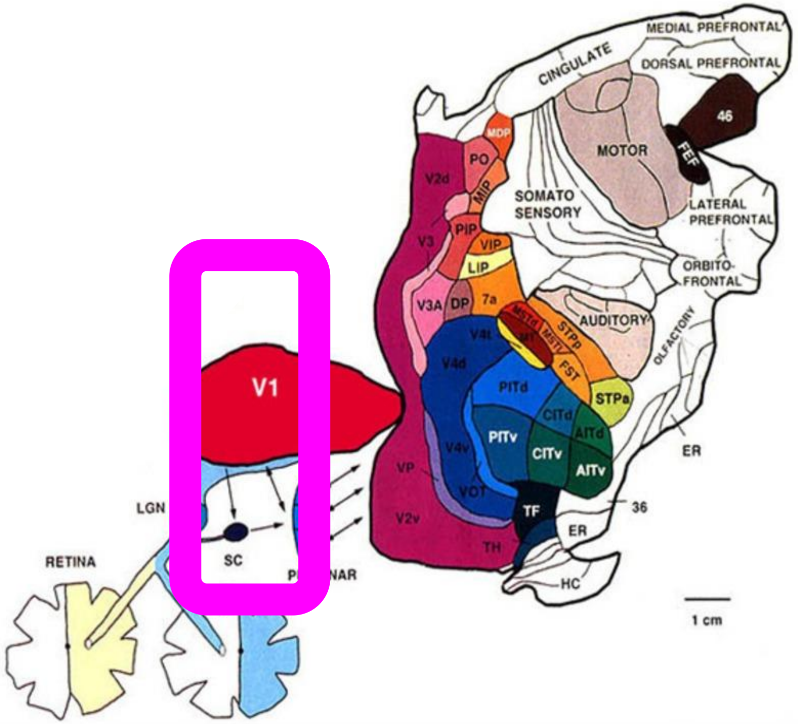
A bottom-up saliency map in the primary visual cortex (Li 1999, 2002)

Retina inputs → Saliency map = V1 firing rates (highest at each location)



→ Winner-take-all

Superior colliculus

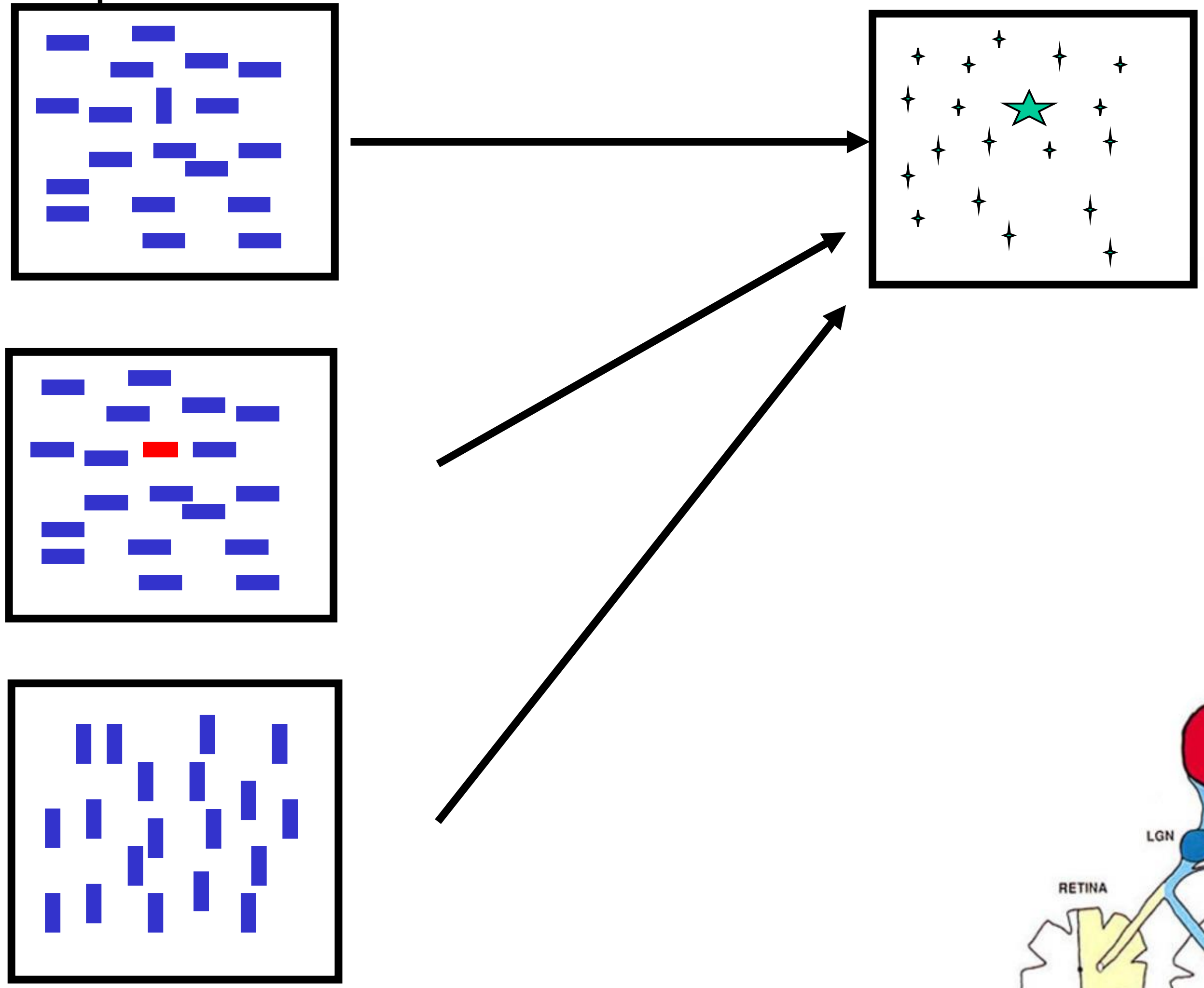


The V1 Saliency Hypothesis:

A bottom-up saliency map in the primary visual cortex (Li 1999, 2002)

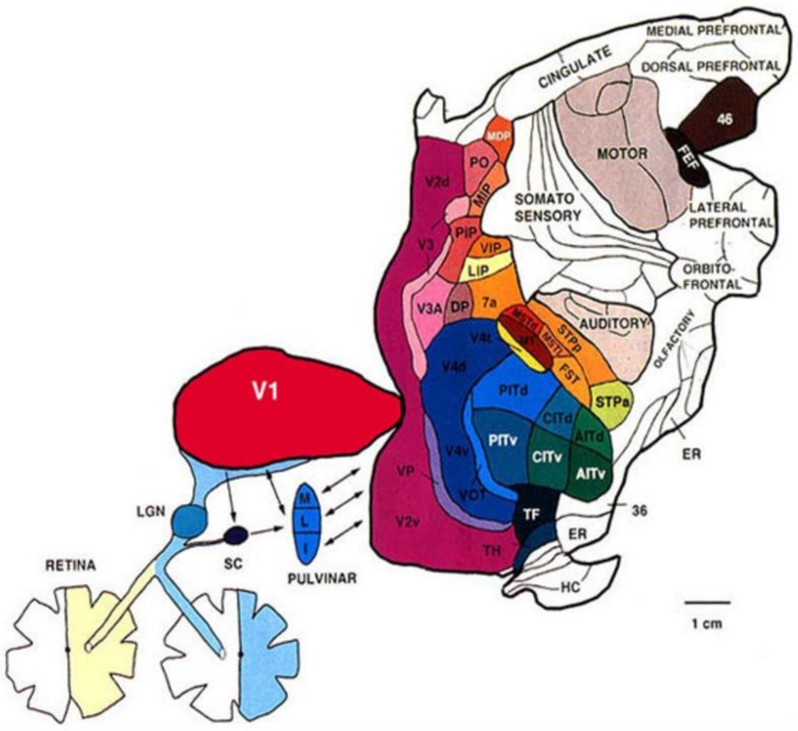
Retina inputs

V1 firing rates (highest at each location)



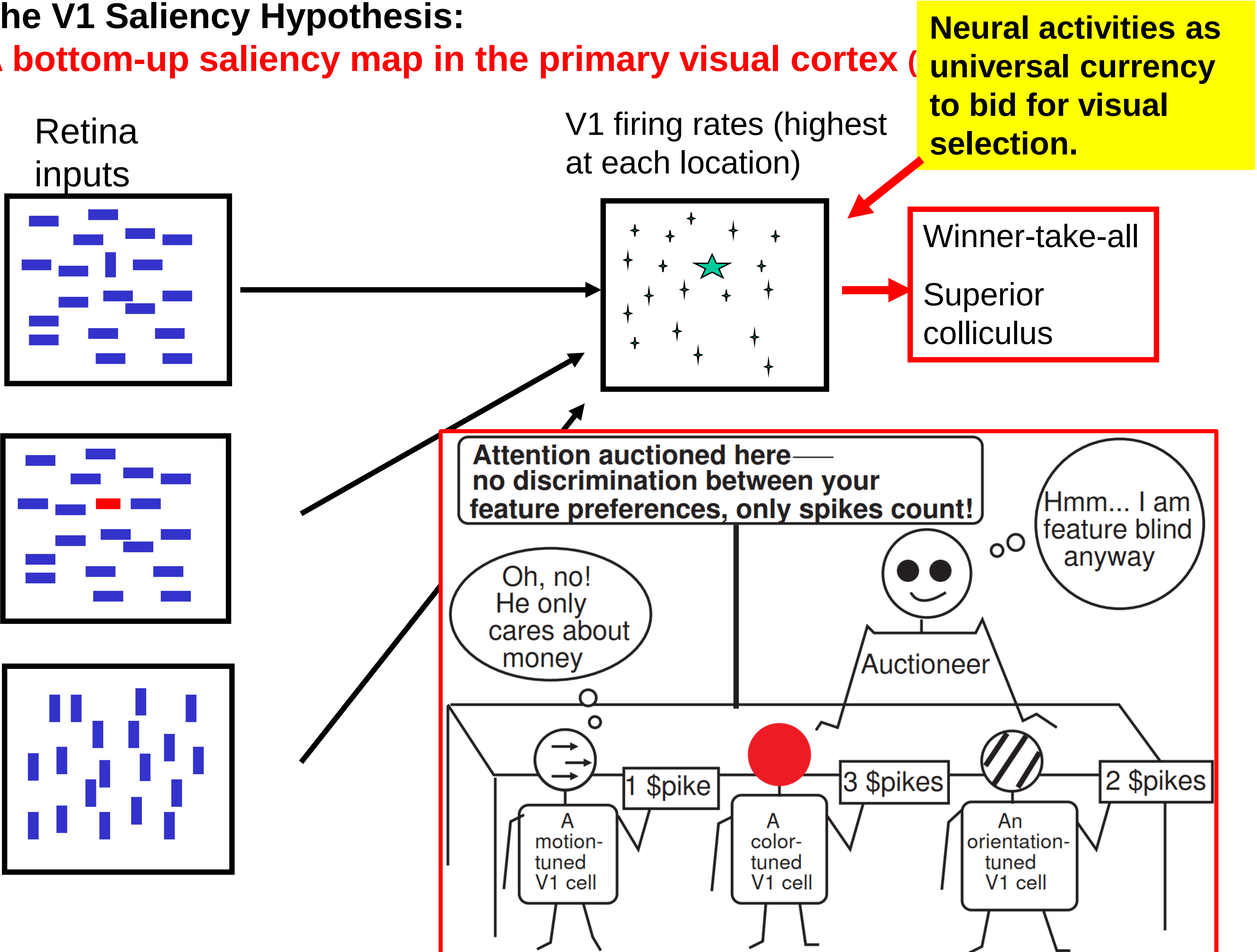
Winner-take-all

Superior colliculus



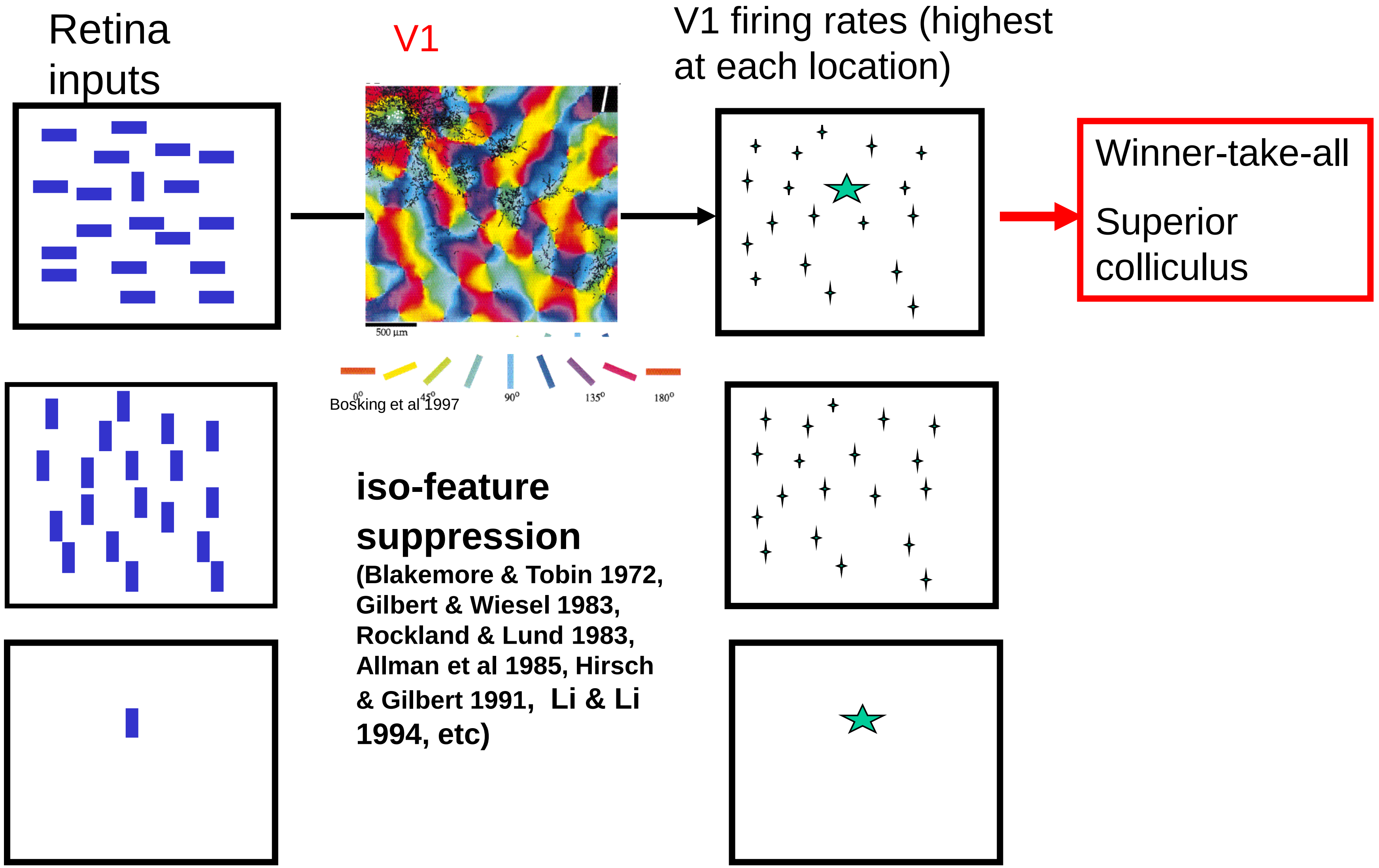
The V1 Saliency Hypothesis:

A bottom-up saliency map in the primary visual cortex (



The V1 Saliency Hypothesis:

A bottom-up saliency map in the primary visual cortex (Li 1999, 2002)

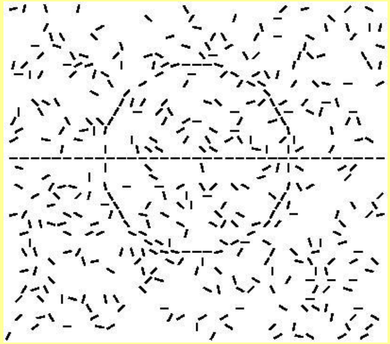


V1's saliency computation on other visual stimuli

Visual input

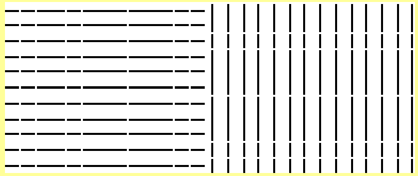
V1 model output

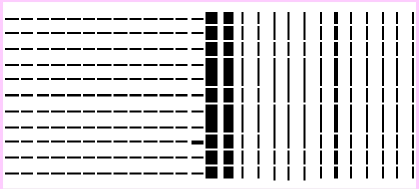
Smooth contours in noisy background



Texture segmentation

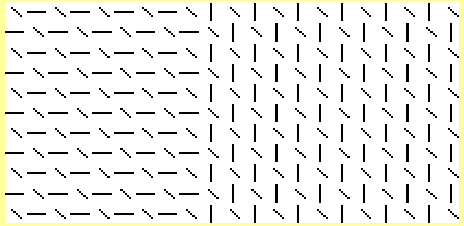
--- simple textures

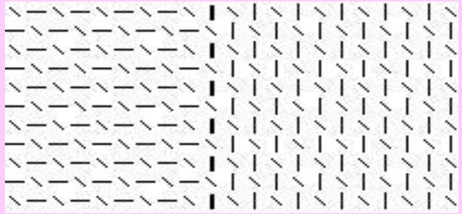




Texture segmentation

--- complex textures



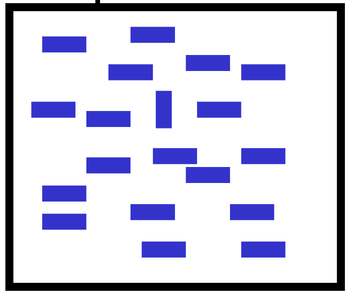


The smooth contours and the texture borders are the most salient

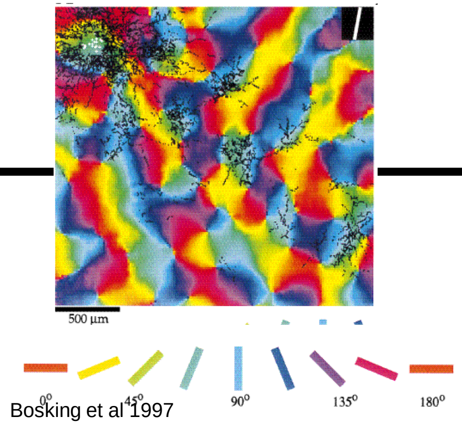
The V1 Saliency Hypothesis:

A bottom-up saliency map in the primary visual cortex (Li 1999, 2002)

Retina inputs

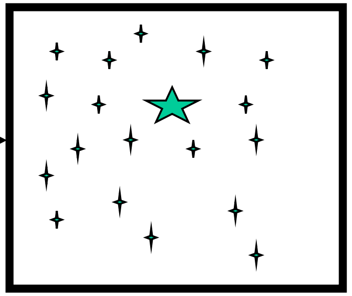


V1



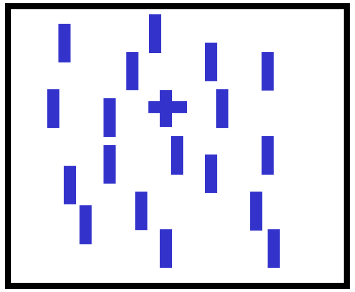
Bosking et al 1997

V1 firing rates (highest at each location)



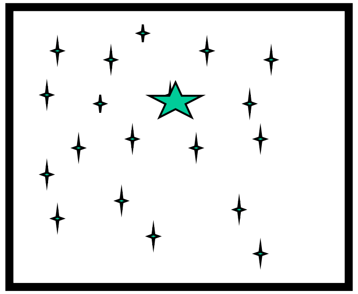
Winner-take-all

Superior colliculus



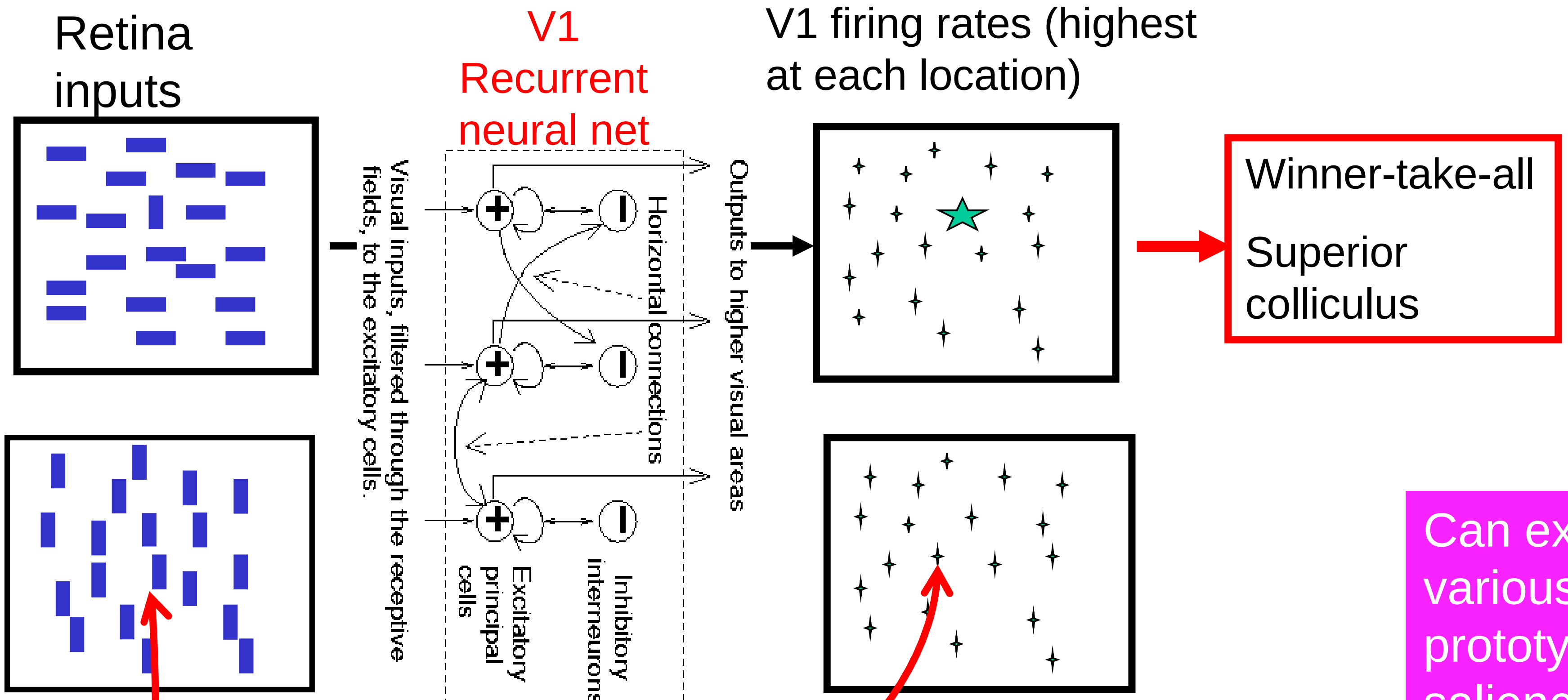
iso-feature suppression

(Blakemore & Tobin 1972, Gilbert & Wiesel 1983, Rockland & Lund 1983, Allman et al 1985, Hirsch & Gilbert 1991, Li & Li 1994, etc)



The V1 Saliency Hypothesis:

A bottom-up saliency map in the primary visual cortex (Li 1999, 2002)



Can explain various prototypical saliency behavior

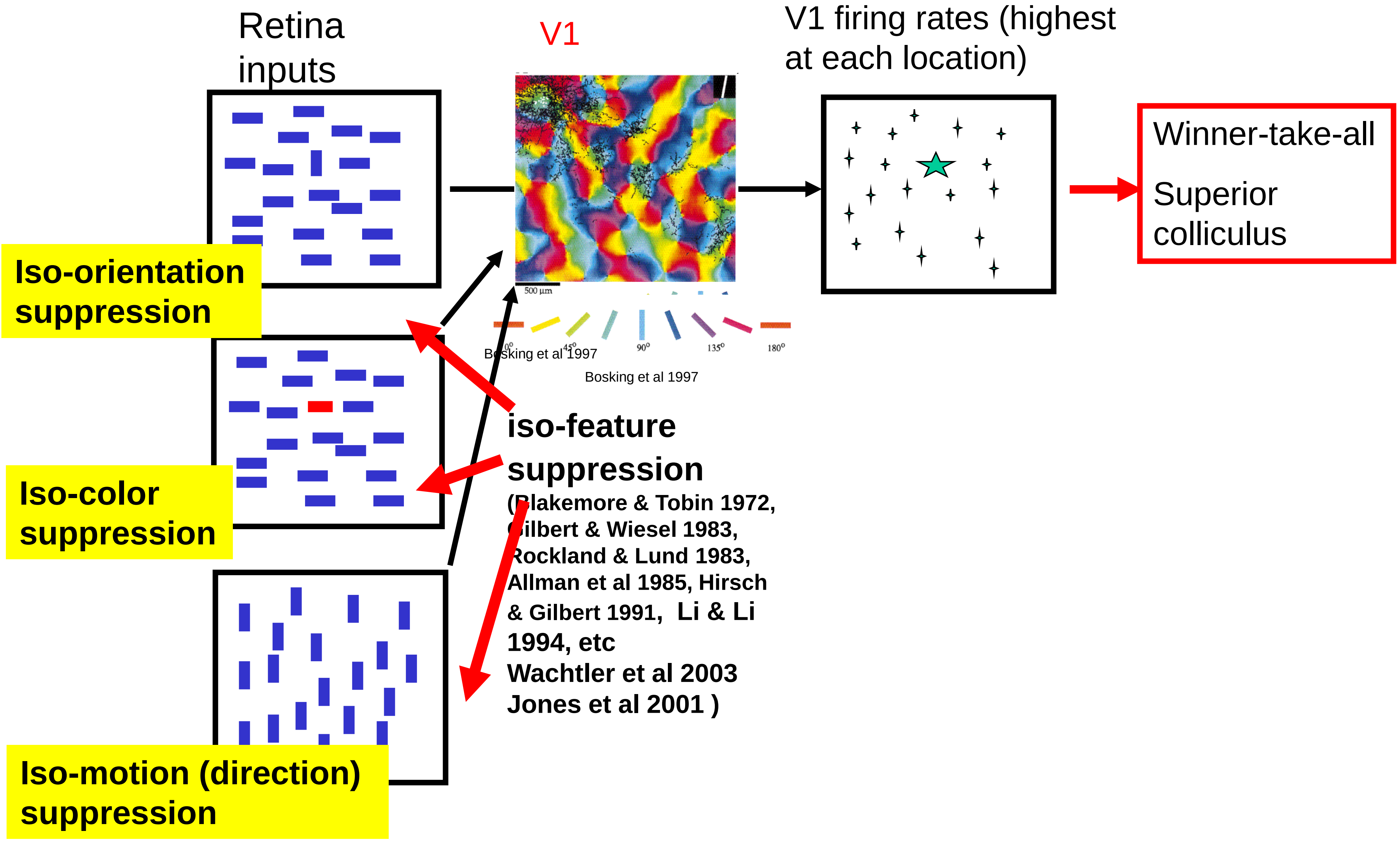
$$\dot{x}_{i\theta} = -\alpha_x x_{i\theta} - g_y(y_{i,\theta}) - \sum_{\Delta\theta\neq 0}\psi(\Delta\theta)\, g_y(y_{i,\theta+\Delta\theta}) + J_o g_x(x_{i\theta}) + \sum_{j\neq i,\theta'} \mathrm{J}_{i\theta,j\theta'}\, g_x(x_{j\theta'}) + I_{i\theta} + I_o + I_{\text{noise}},$$

$$\dot{y}_{i\theta} = -\alpha_y y_{i\theta} + g_x(x_{i\theta}) + \sum_{j\neq i,\theta'} \mathrm{W}_{i\theta,j\theta'}\, g_x(x_{j\theta'}) + I_c + I_{\text{noise}}.$$

A neural circuit model

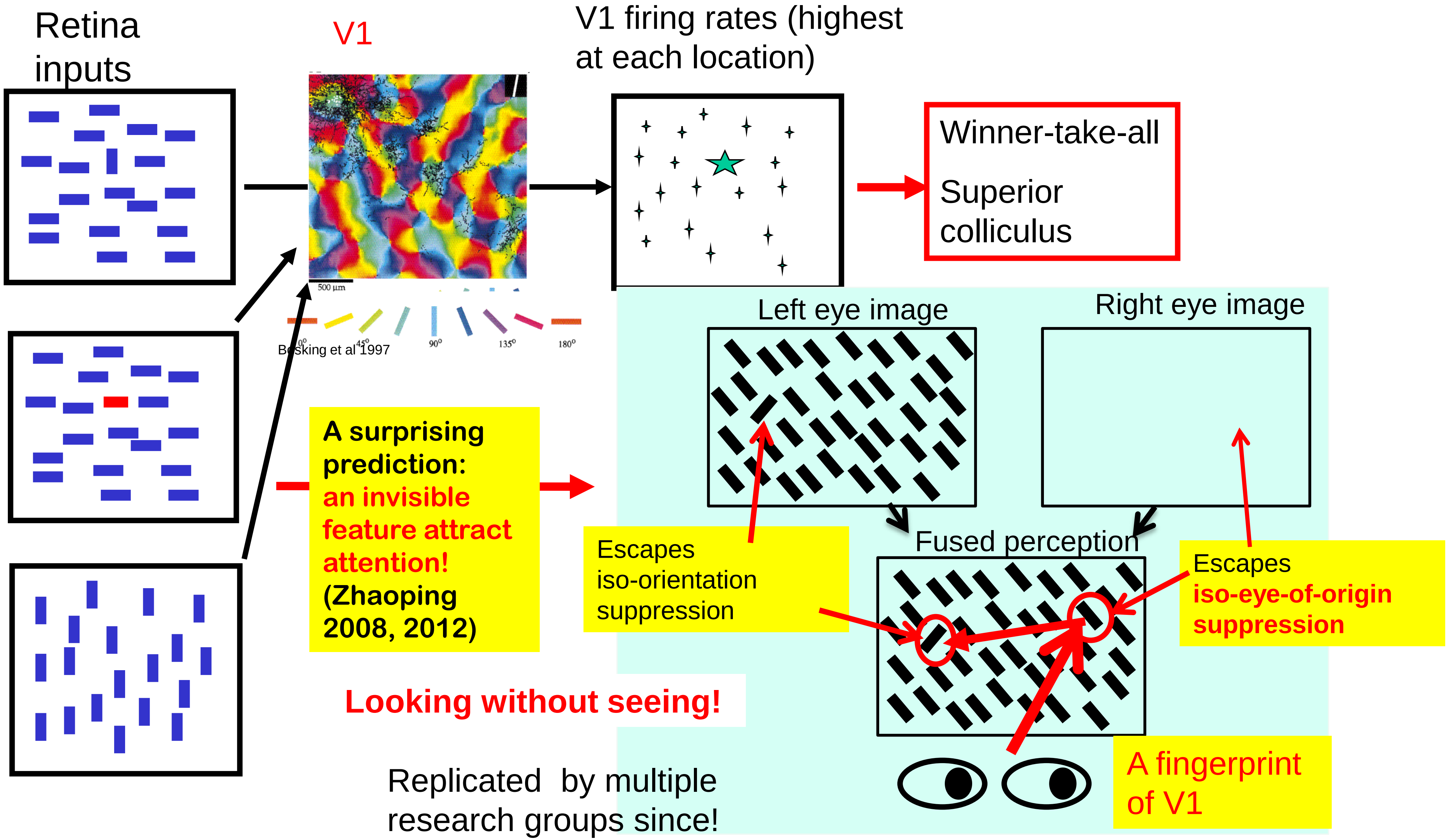
Local connections → global outcome

The V1 Saliency Hypothesis:
A bottom-up saliency map in the primary visual cortex (Li 1999, 2002)



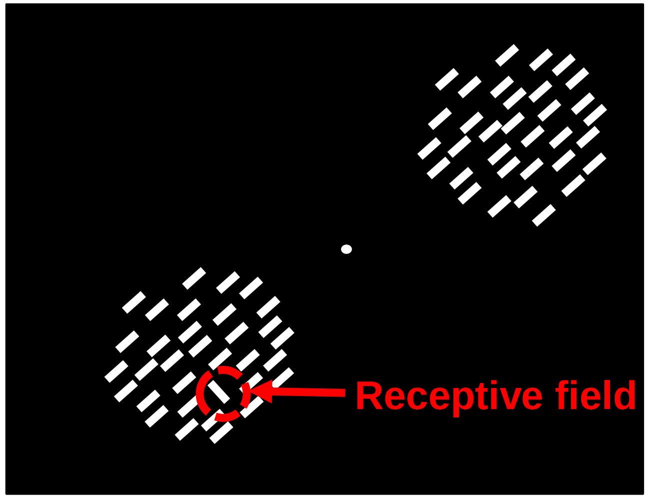
The V1 Saliency Hypothesis:

A bottom-up saliency map in the primary visual cortex (Li 1999, 2002)

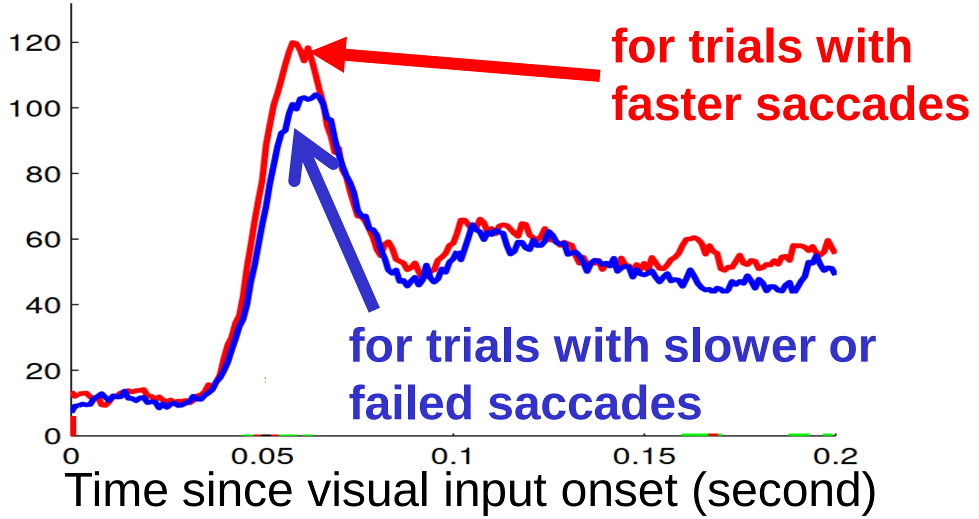


Testing the V1 theory on behaving monkeys --- Yan, Zhaoping, & Li, in press, PNAS 2018.

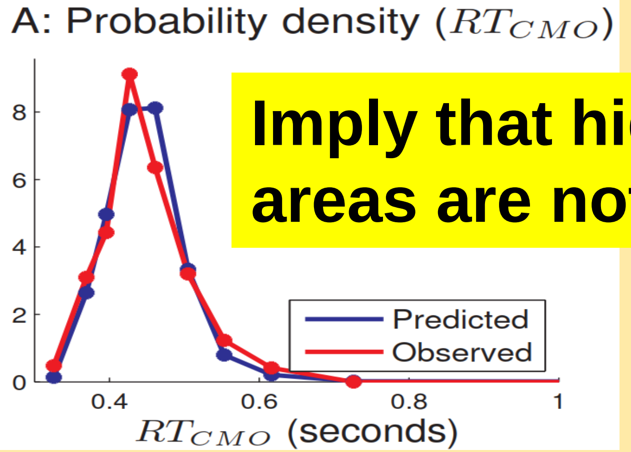
Saccade to an uniquely oriented bar ASAP



V1 neural responses to input stimulus (spikes/sec)



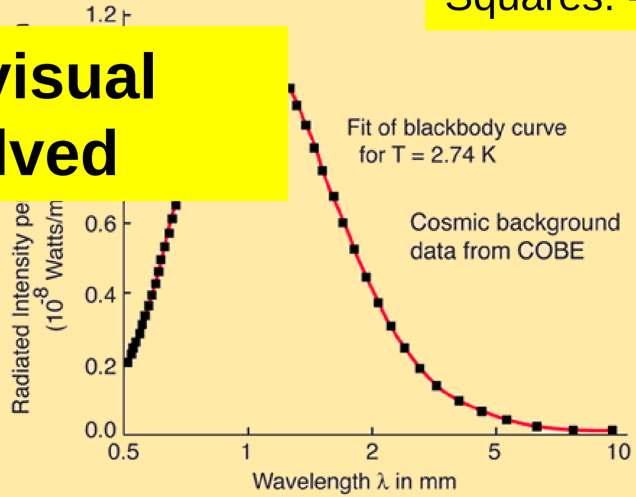
Quantitative, zero-parameter, predictions from theory



Koene & Zhaoping 2007
Zhaoping & Zhe 2015,

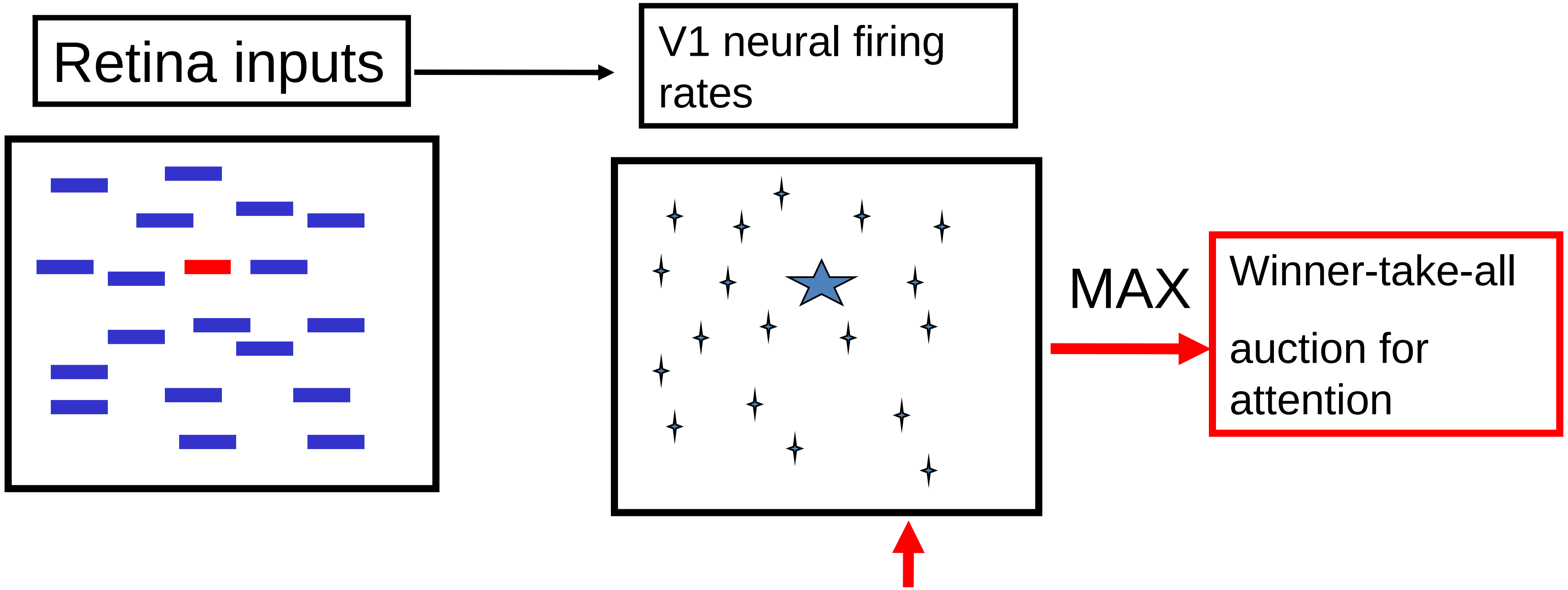
Imply that higher visual areas are not involved

Solid curve ---Planck's law
Squares: --- data points



Now the zero parameter quantitative prediction (Zhaoping & Zhe 2012,2015)

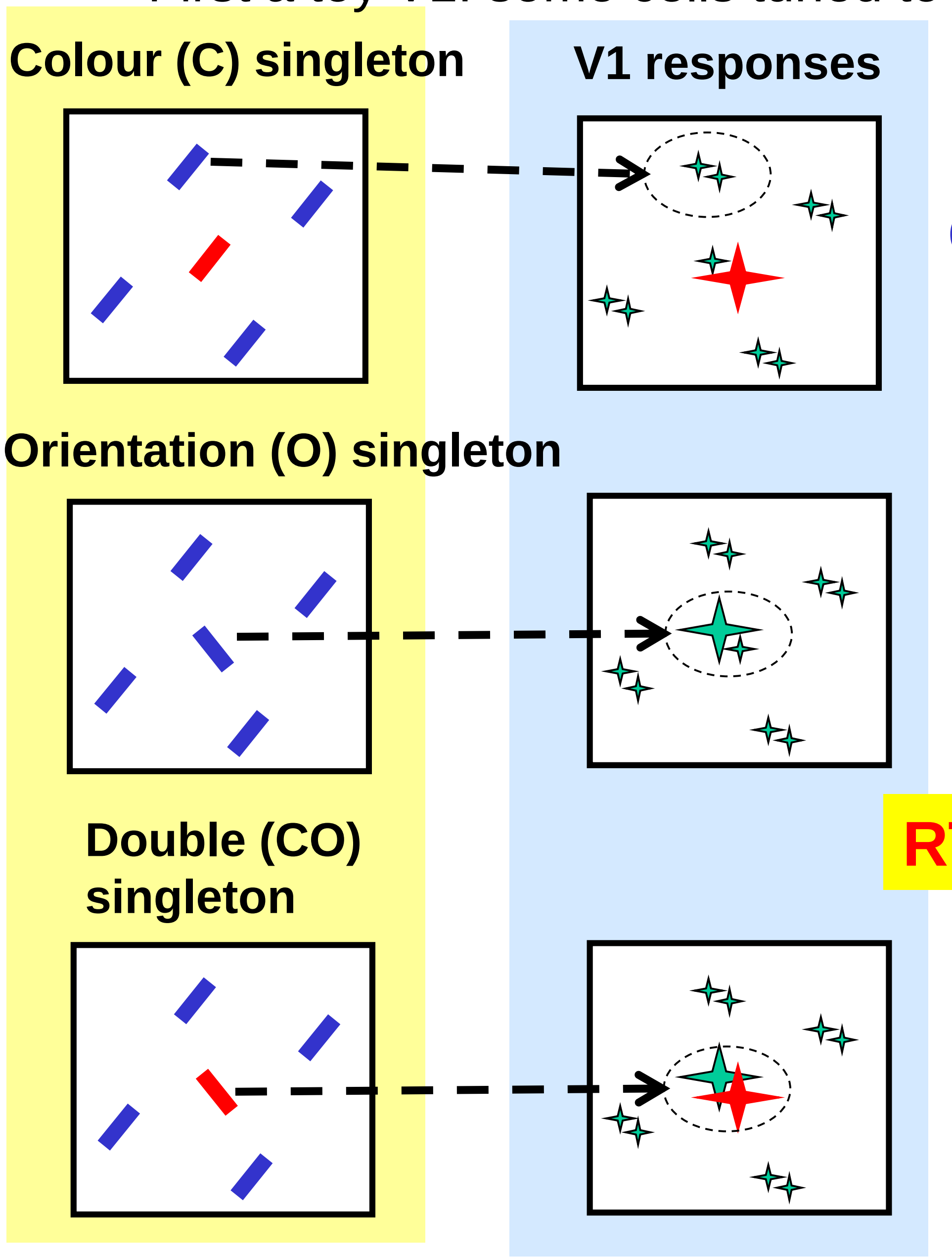
First, recall from the theory:
maximum firing, (not summation of firing rates),
at a location determines its saliency



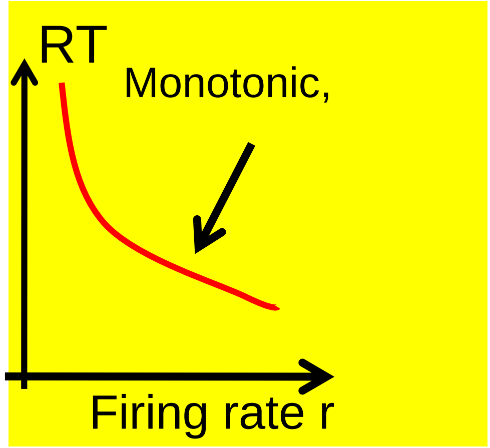
Neural activities as universal currency to bid for visual selection.
The receptive field of the most active V1 cells is selected

Predicting RT:

First a toy V1: some cells tuned to orientation, others tuned to color



RT_C = 500 ms ,
Color (C) cell
Max response: r_C



RT_O = 600 ms,
Orientation (O) cell
Max response: r_O

RT_{CO}= 500 ms = min (RT_C, RT_O)

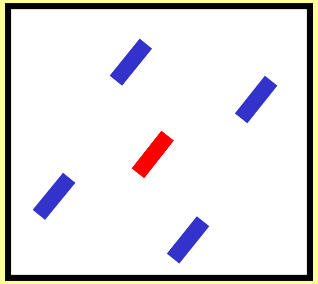
RT_{CO} = ?

C cell r_C
O cell r_O

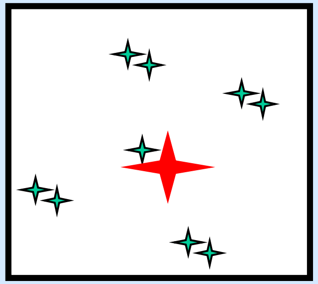
Predicting RT:

First a toy V1: some cells tuned to orientation, others tuned to color

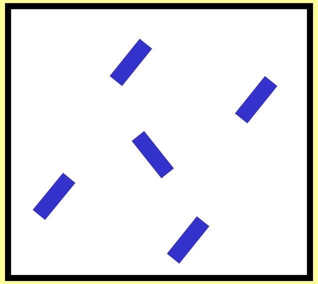
Colour (C) singleton

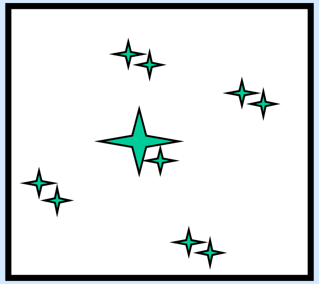


V1 responses

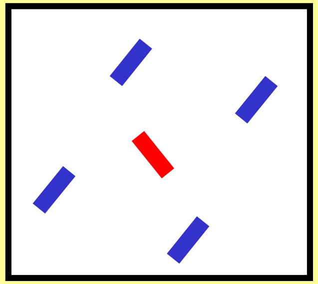


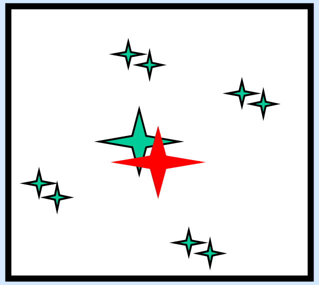
Orientation (O) singleton

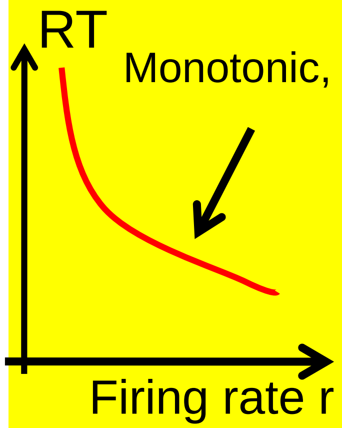




Double (CO) singleton



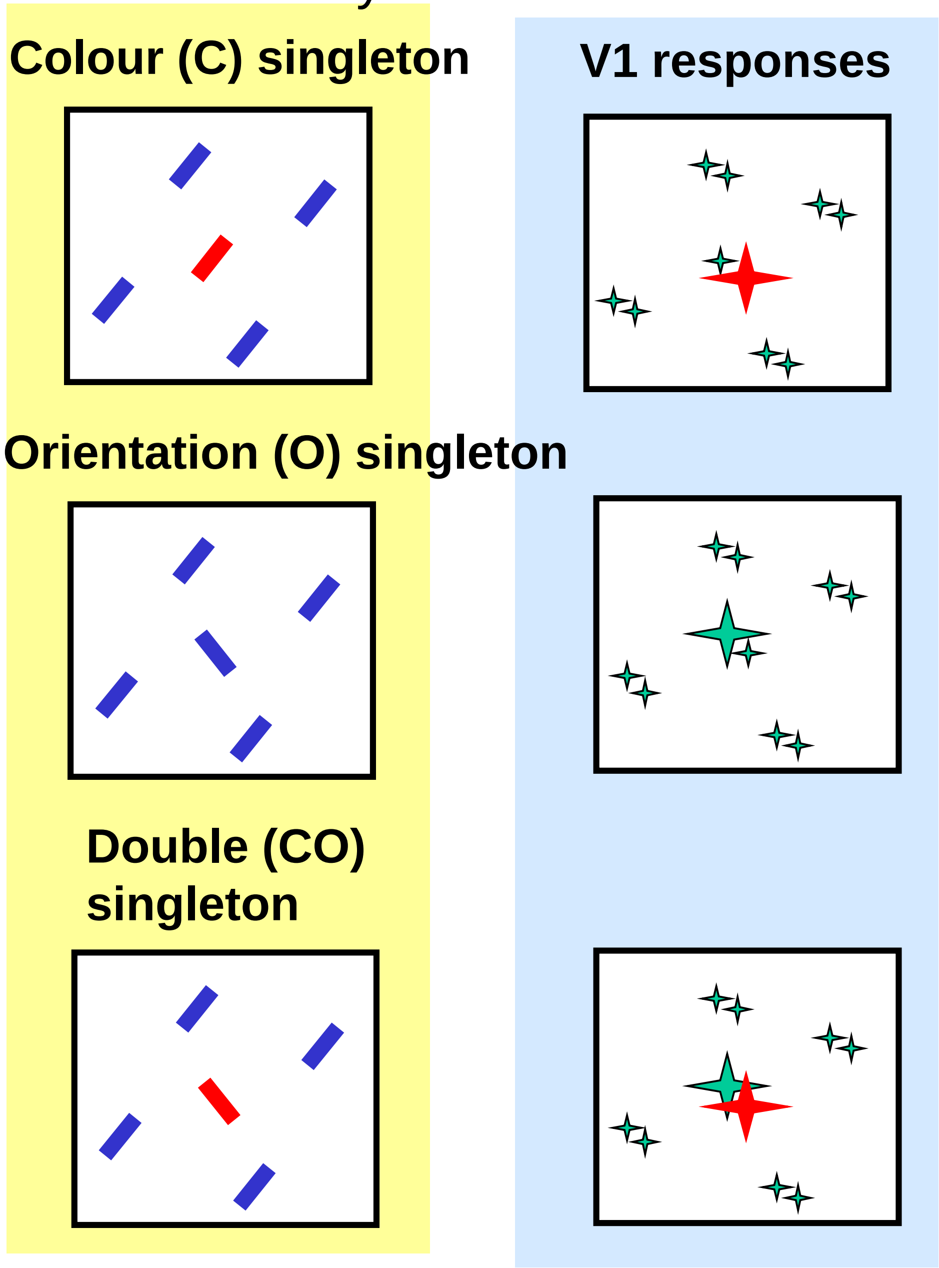




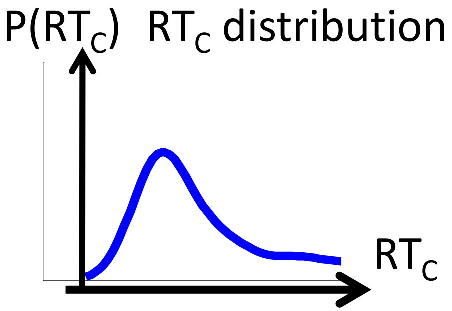
$$RT_{CO} = \min(RT_C, RT_O)$$

Predicting RT:

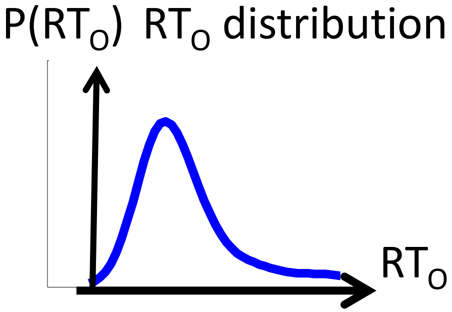
First a toy V1: some cells tuned to orientation, others tuned to color



Stochastic
C cell responses
10, 9, 11, …



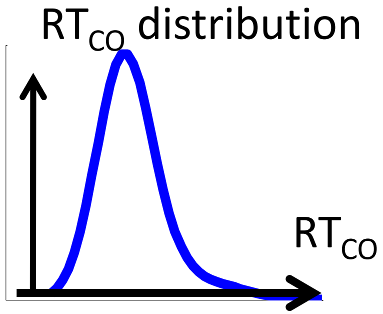
O cell responses
11, 9, 8, 10, …



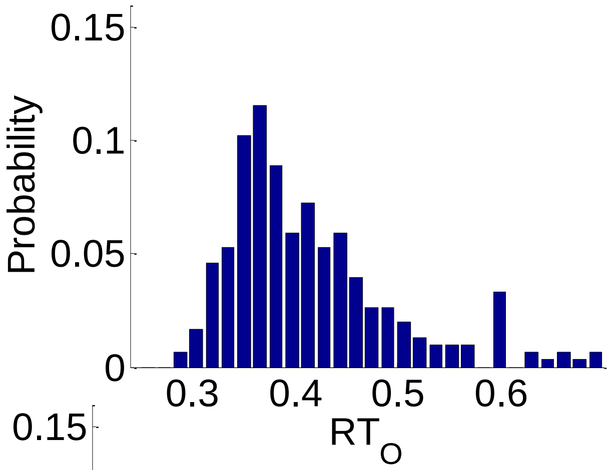
$RT_{CO} = \min(RT_C, RT_O)$

$P(RT_{CO})$

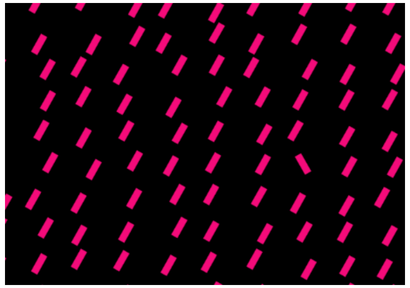
$= P[\min(RT_C, RT_O)]$

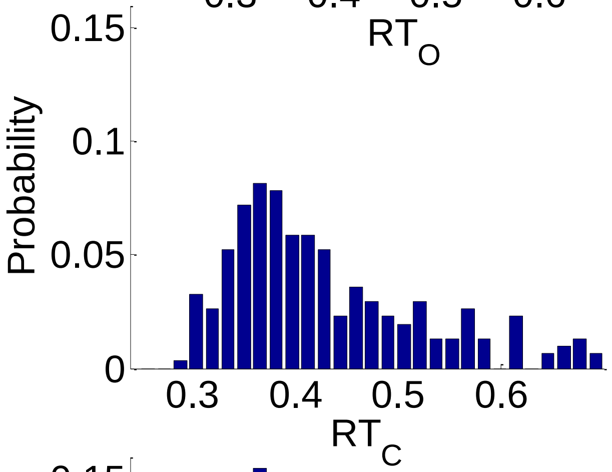


Behavioral data from Koene and Zhaoping (2007)

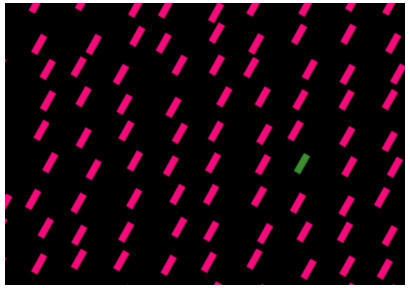


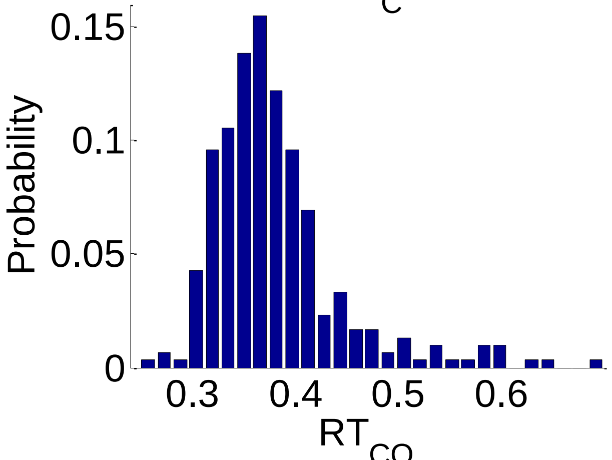
O target



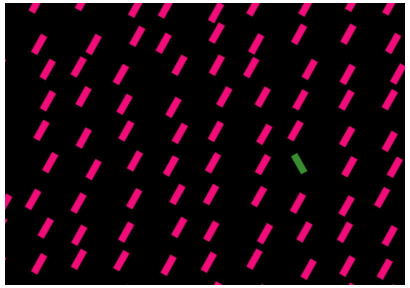


C target



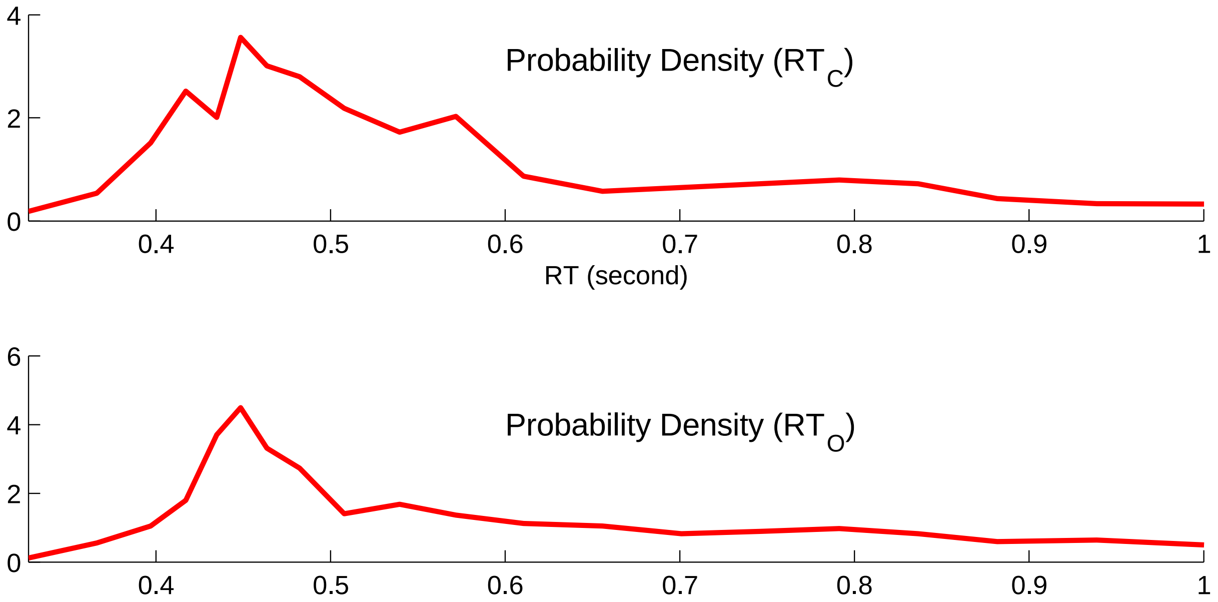


CO target



About 300 trials per condition

Predict RT_{CO} from $RT_{CO} = \min(RT_C, RT_O)$

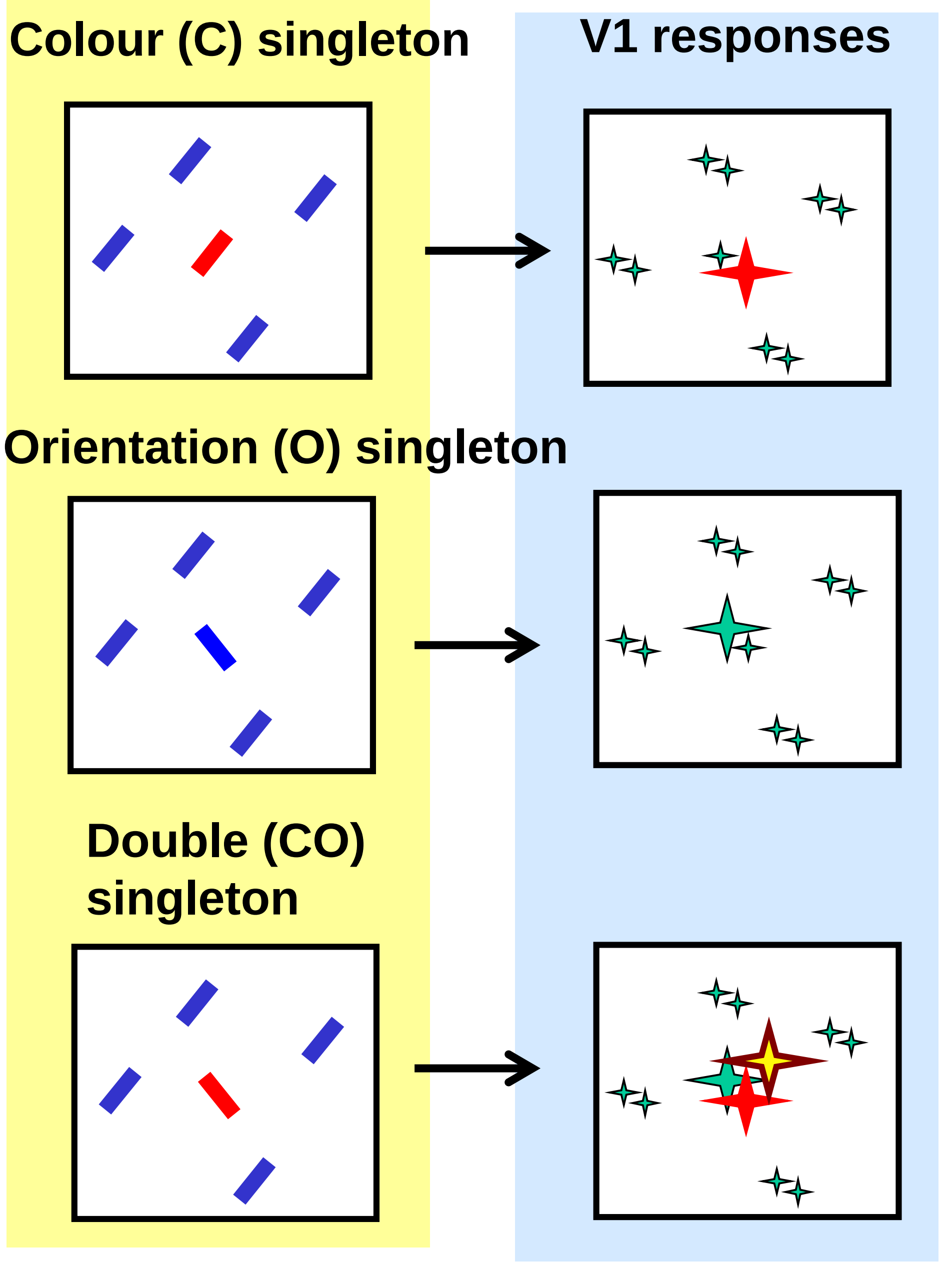


P value ~ 0.00

Predicted RT significantly longer than observed RT

Because --- real V1 has CO conjunction cells

Because --- real V1 has CO conjunction cells



$RT_{CO} <= \min(RT_C, RT_O)$

~~$RT_{CO} = \min(RT_C, RT_O)$~~

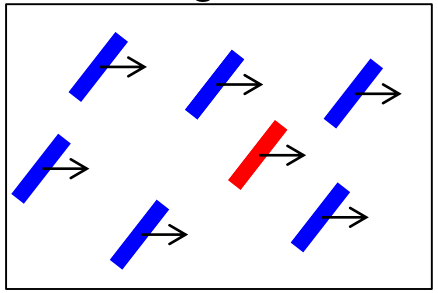
Singleton strongly evoke all three cell types: C, O, CO

V1 without CO cells, → $RT_{CO} = \min(RT_C, RT_O)$

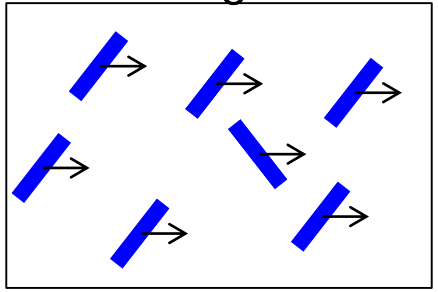
V1 with CO cells, → RT_{CO} £ $\min(RT_C, RT_O)$

Introduce another feature dimension: motion direction (M),

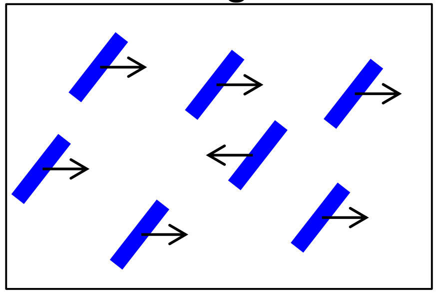
C singleton



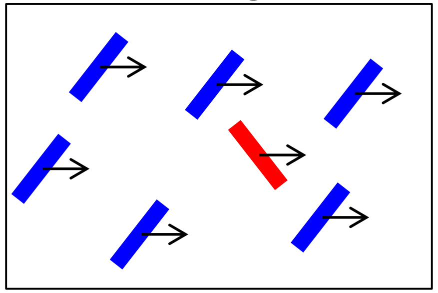
O singleton



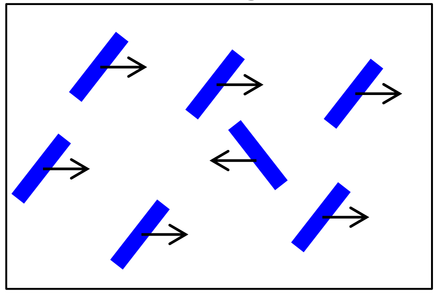
M singleton



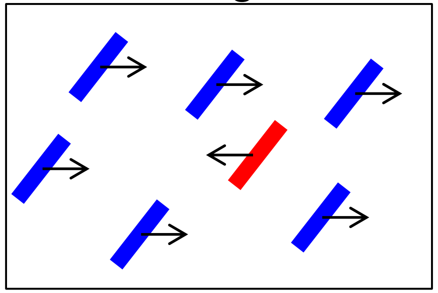
CO singleton



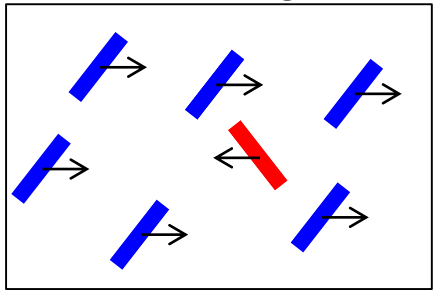
MO singleton



CM singleton



CMO singleton



V1 without CO cells, → $RT_{CO} = \min(RT_C, RT_O)$

V1 with CO cells, → $RT_{CO} \le \min(RT_C, RT_O)$

Analogously

Introduce another feature dimension: motion direction (M),

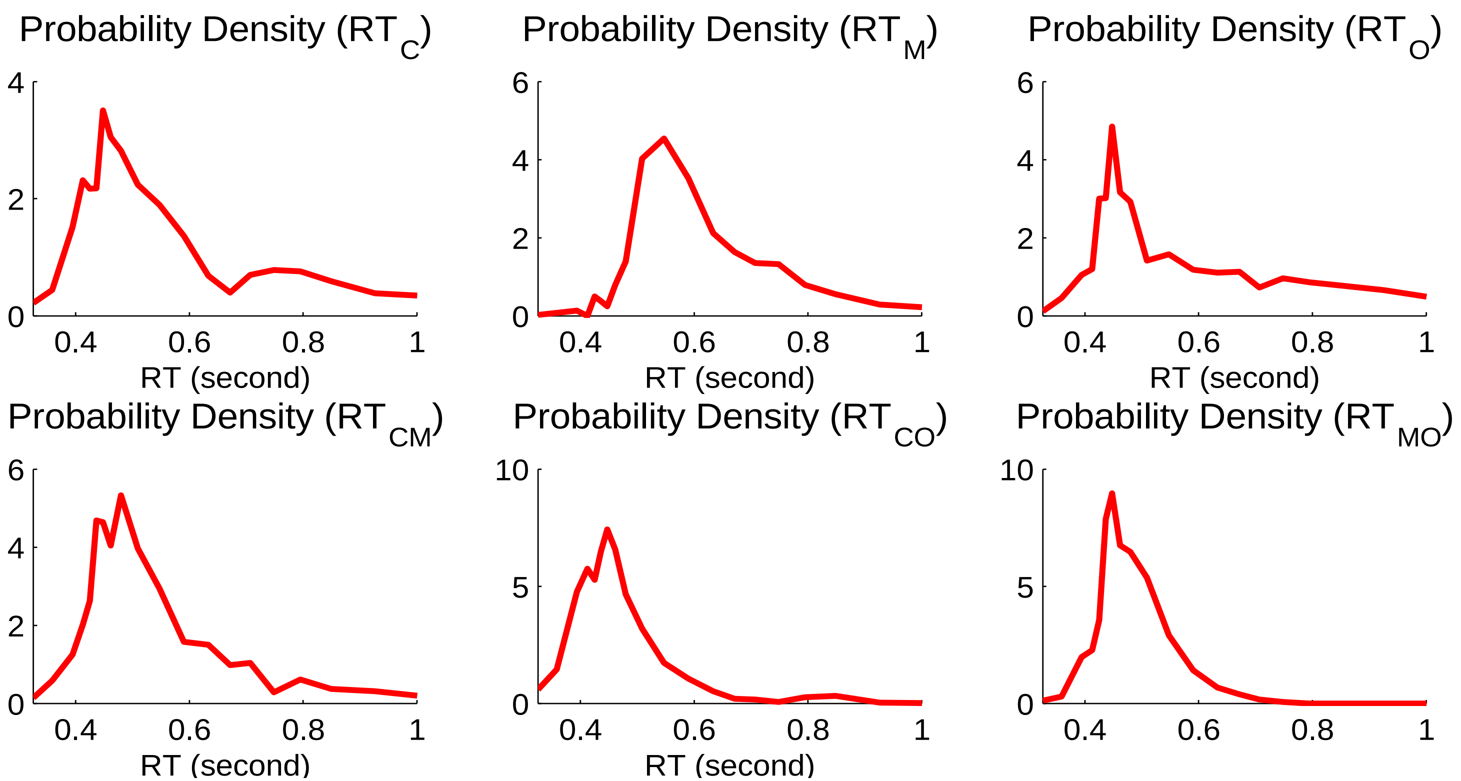
V1 has no CMO cells, →

$$\min(RT_{CMO}, RT_C, RT_M, RT_O) = \min(RT_{CM}, RT_{CO}, RT_{MO})$$

Predict

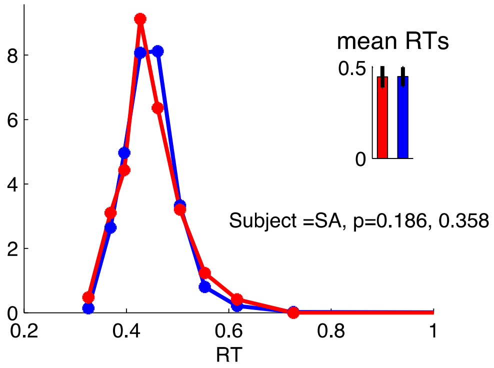
Distributions of RTs for a particular subject

$$\min(RT_{CMO}, RT_C, RT_M, RT_O) = \min(RT_{CM}, RT_{CO}, RT_{MO})$$



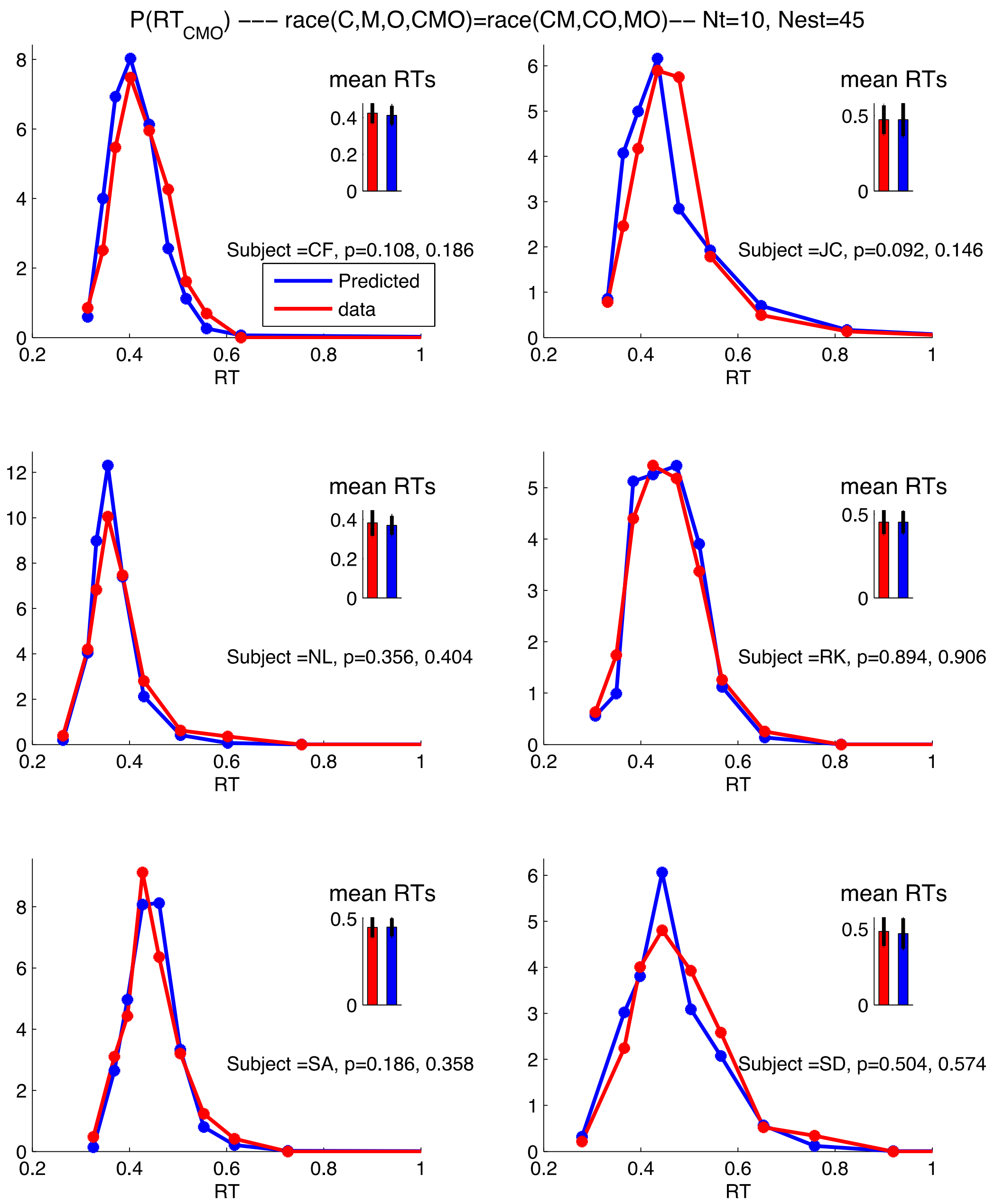
> 0.18,
cant

$$\min(RT_{CMO}, RT_C, RT_M, RT_O) = \min(RT_{CM}, RT_{CO}, RT_{MO})$$



Prediction=data for all six observers

$$\min(RT_{CMO}, RT_C, RT_M, RT_O) = \min(RT_{CM}, RT_{CO}, RT_{MO})$$



V1 without CO cells, → $RT_{CO} = \min(RT_C, RT_O)$

V1 with CO cells, → $RT_{CO} \le \min(RT_C, RT_O)$

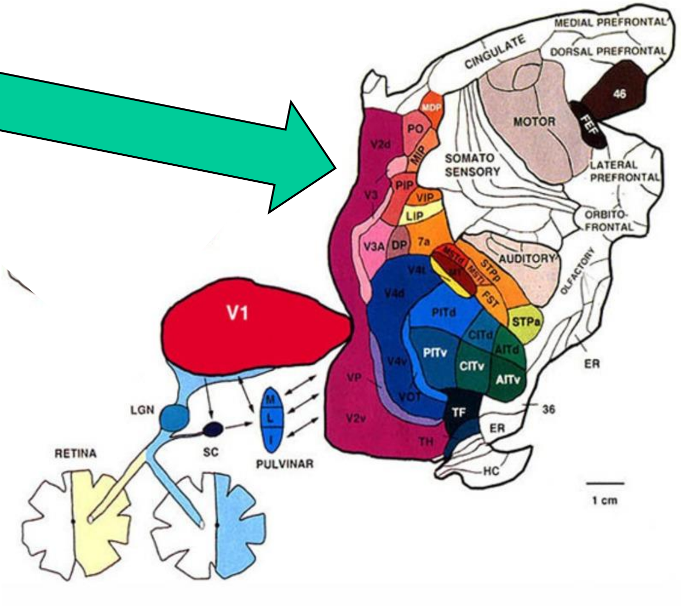
analogously

analogously

V1 has no CMO cells, → $\min(RT_{CMO}, RT_C, RT_M, RT_O) = \min(RT_{CM}, RT_{CO}, RT_{MO})$

Higher visual areas have CMO cells → $\min(RT_{CMO}, RT_C, RT_M, RT_O) \le \min(RT_{CM}, RT_{CO}, RT_{MO})$

Our data suggest that the higher visual areas are not involved in bottom-up saliency computation.



Talk outline

(1) The functional role of the primary visual cortex (V1) ← **Attentional selection**

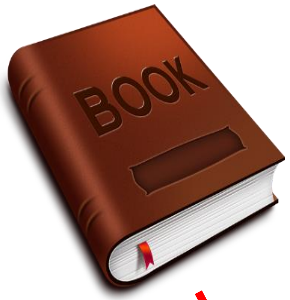
(2) In light of V1's role → a new plan to understanding vision

(3) A first example study in this new plan

Information bottlenecks in the visual pathway:

"To be or not to be,
This is the question .."

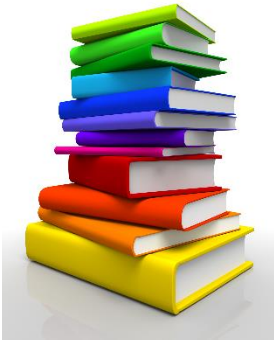
10^7 bits/second
~ 10^6 neurons ,
~10 spikes/neuron
~1 bit/spike

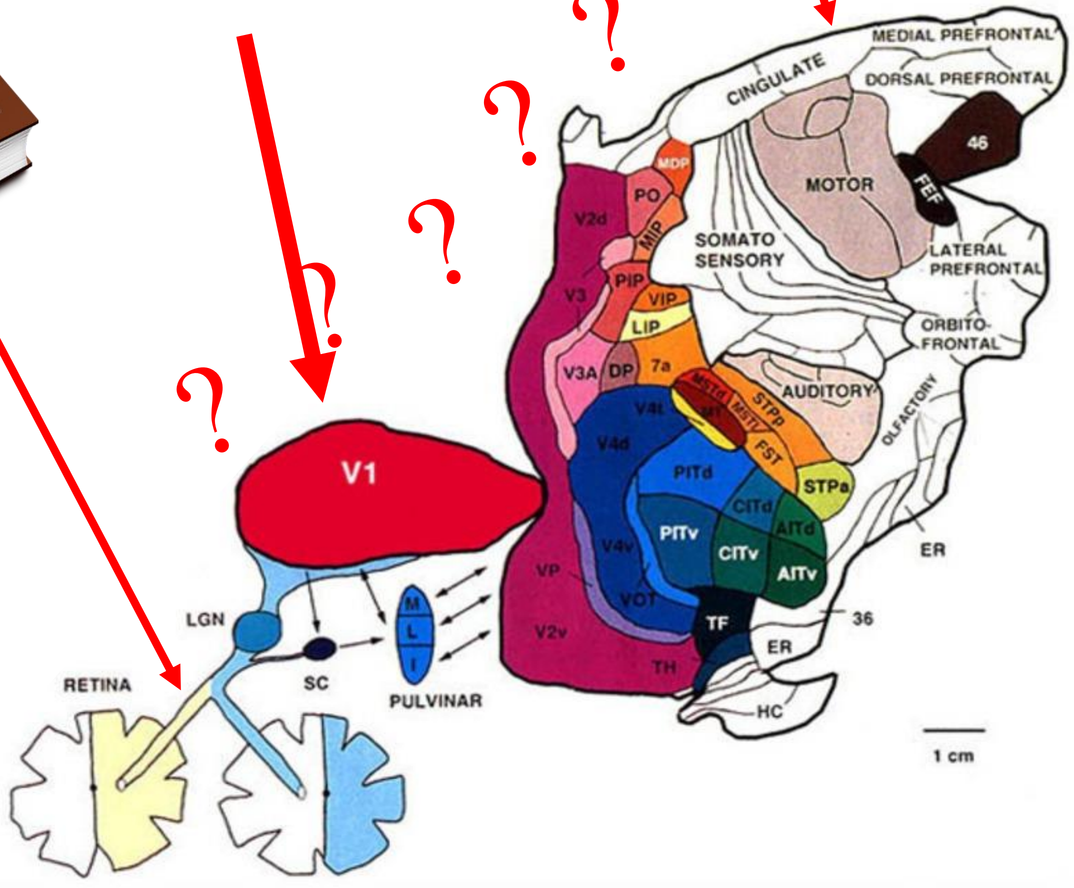


Bottom-up selection

40 bits/second
(Sziklai, 1956)

10^9 bits/second (Kelly 1962)
~ 25 frames/second,
2000x2000 pixels,
1 byte/pixel





A new path to understanding vision

Visual inputs → Encoding → Selection ⇄ Decoding → for motor action or cognitive decisions

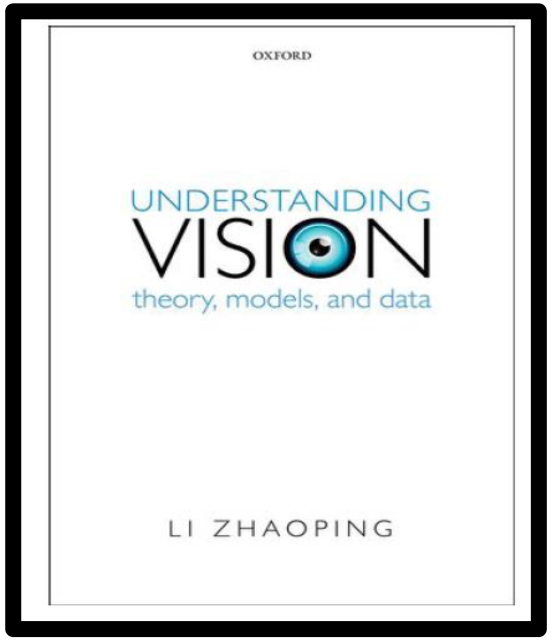
(e.g., by retinal neural activities)

(attentional s... n, often by sacc... to selected locatio...

(e.g., face recognition)

20 frames, 20 megabytes/second

40 bits/second



Traditional paths to understanding vision

(1) Low level vision, mid-level vision, high-level vision

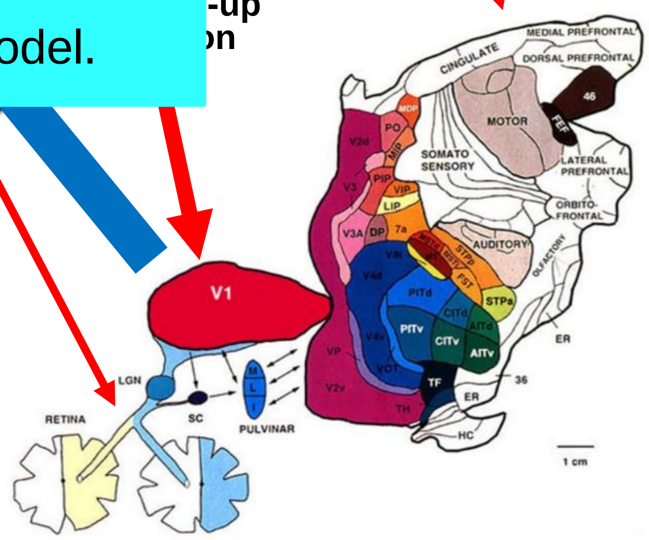
(2) David Marr: primal sketch, 2.5 d sketch, 3-d model.

"To be or not to be,
This is the question .."

40 bits/second
(Sziklai, 1956)

-up
on

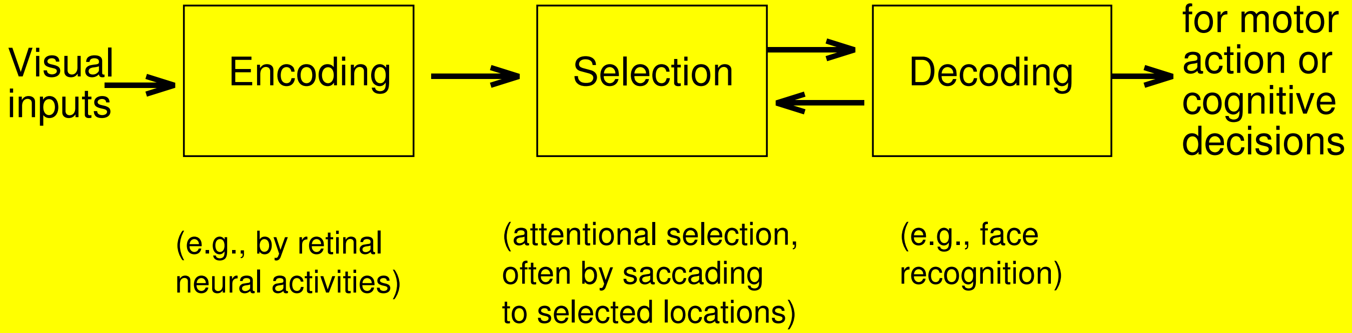
(Kelly 1962)
~ 25 frames/second,
2000x2000 pixels,
1 byte/pixel



Talk outline

(1) The functional role of the primary visual cortex (V1) ← **Attentional selection**

(2) In light of V1's role →
a new plan to understanding vision



(3) A first example study in this new plan

A new path to understanding vision

Visual inputs → Encoding → Selection ⇄ Decoding → for motor action or cognitive decisions

20 frames,
20 megabytes/second

(e.g., by retinal neural activities)

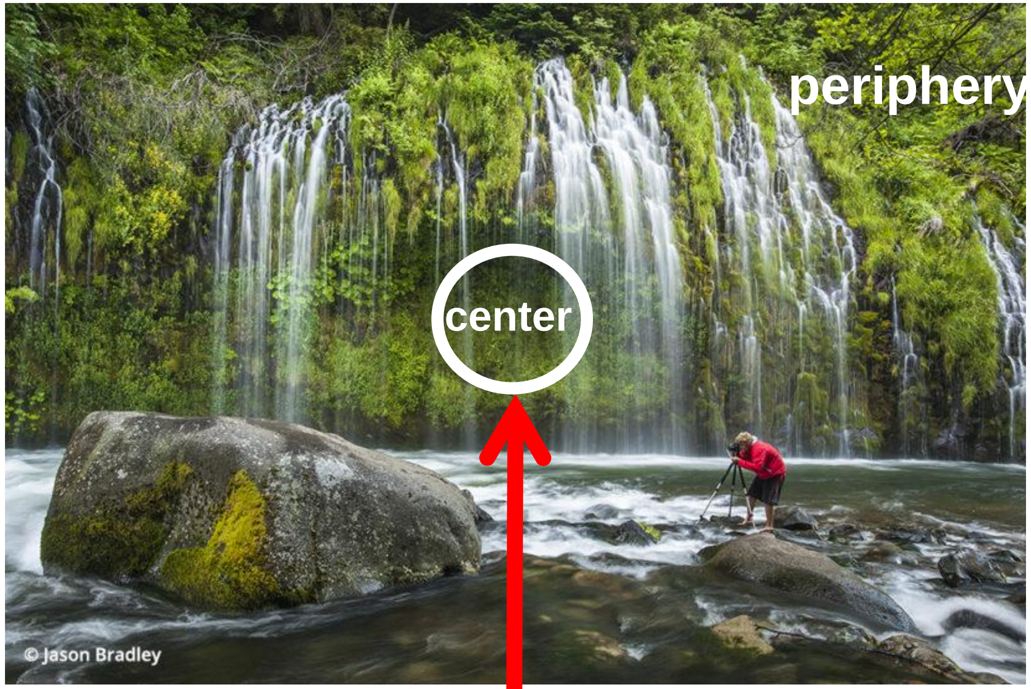
(attentional selection, often by saccading to selected locations)

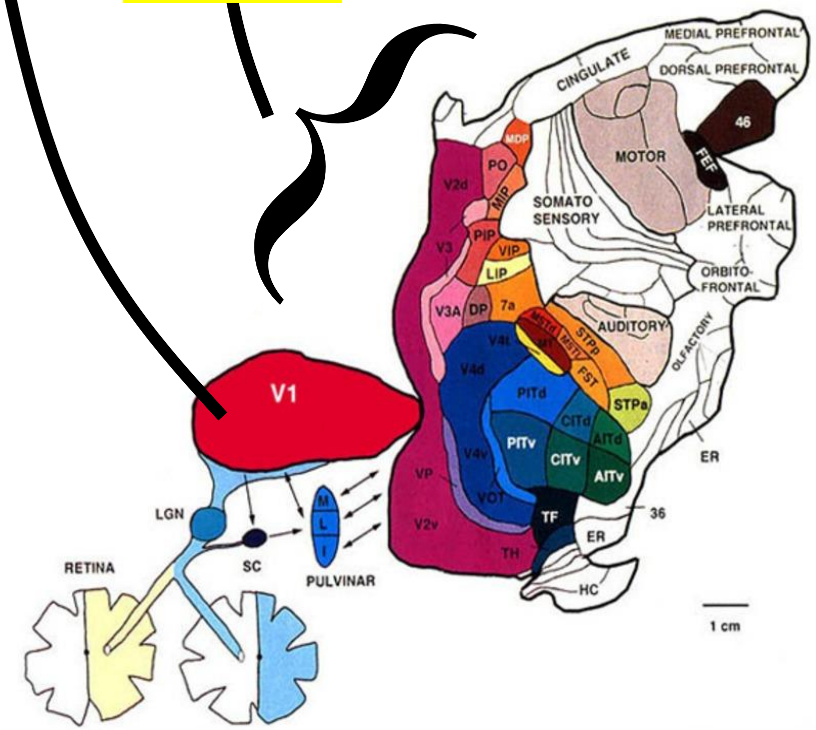
Looking

(e.g., face recognition)

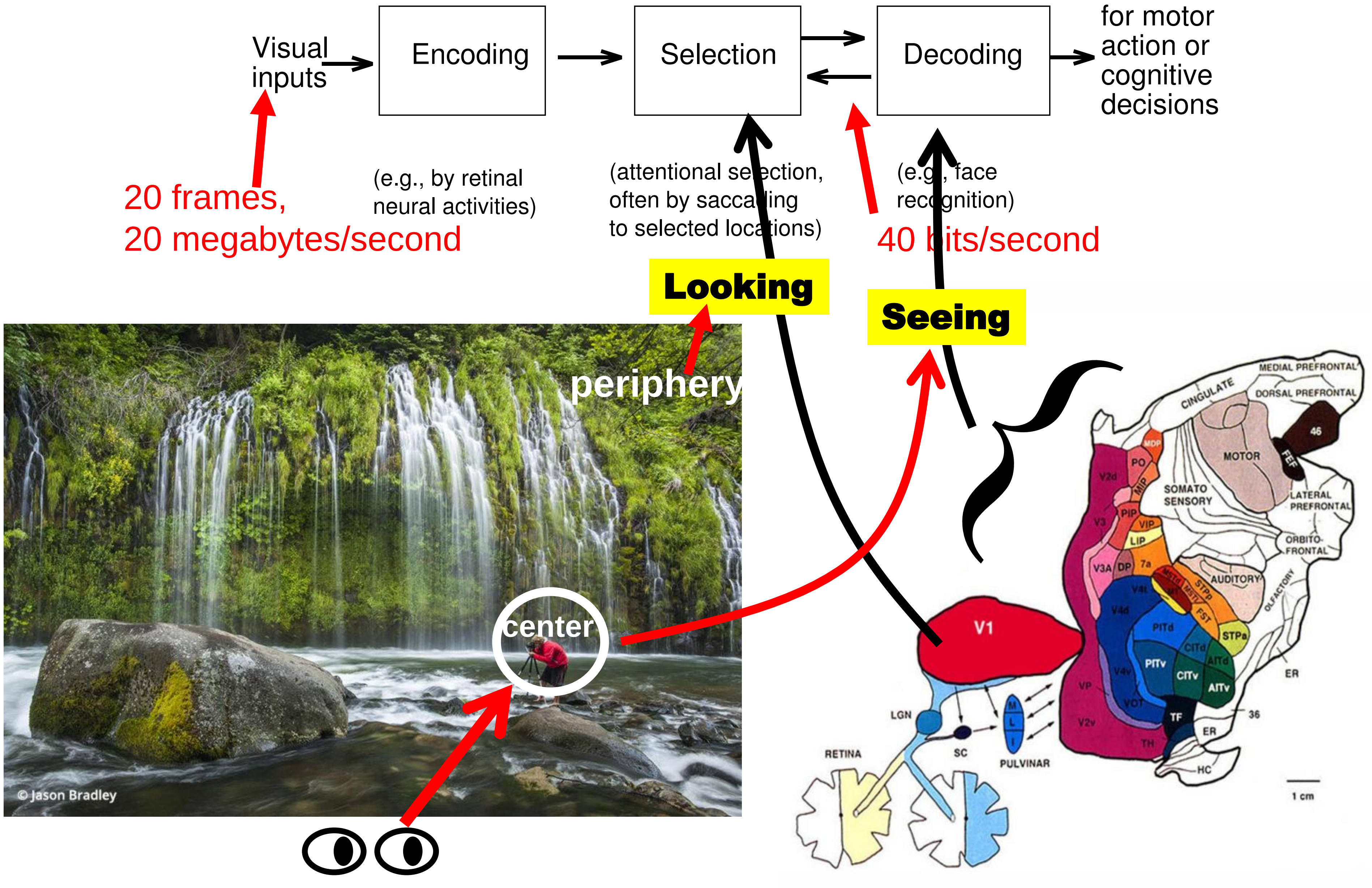
40 bits/second

Seeing





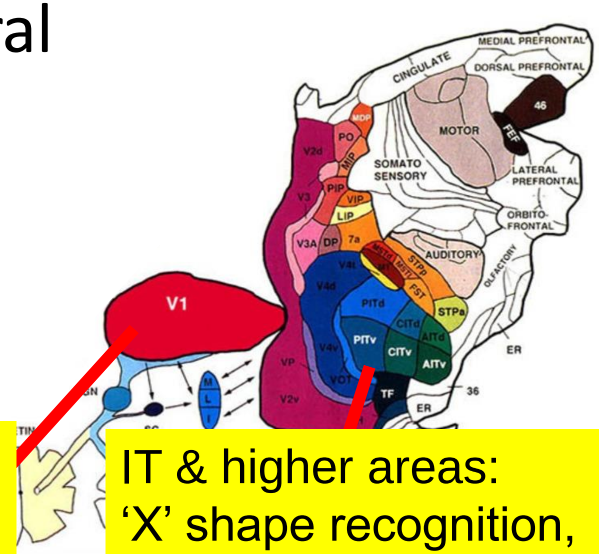
A new path to understanding vision



Looking and seeing --- peripheral and central

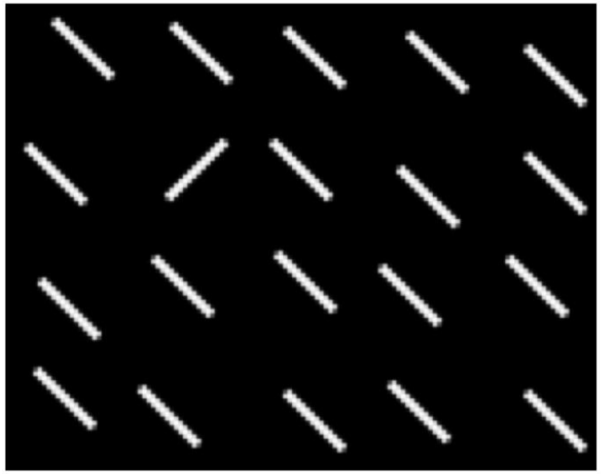
Two separate processes

Demo: (Zhaoping & Guyader 2007)

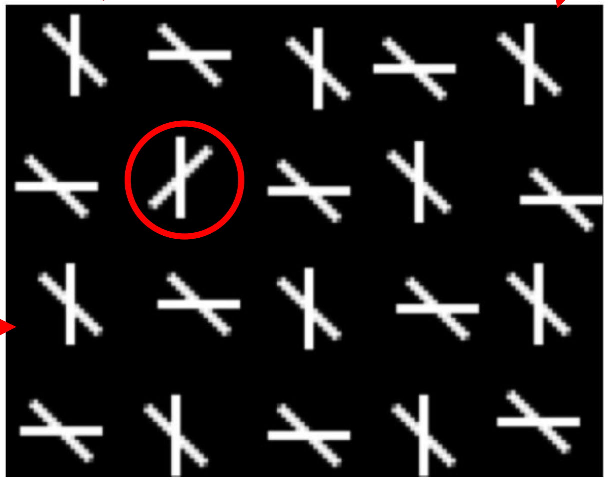


V1: tuned to primitive bars

IT & higher areas: 'X' shape recognition, rotationally invariant



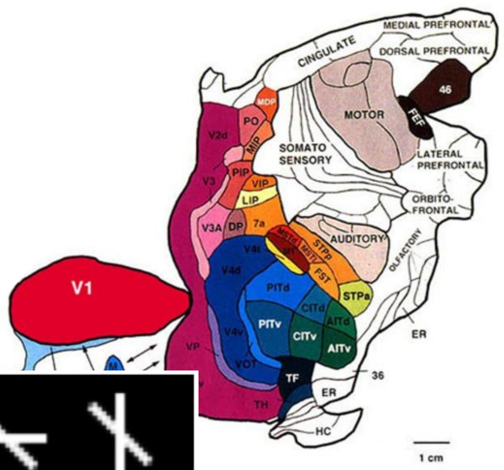
Add a horizontal bar or a vertical bar to each oblique bar

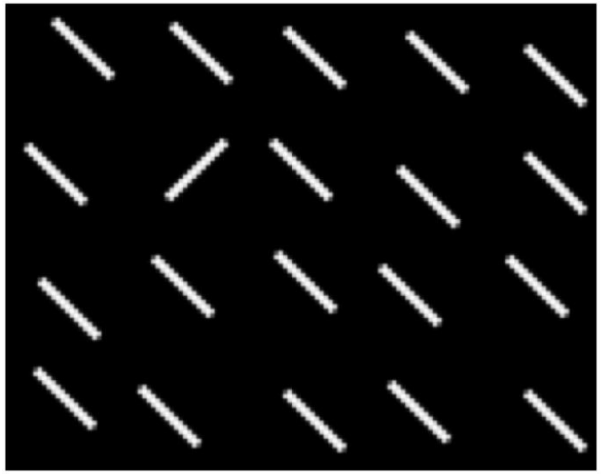


Looking and seeing --- peripheral and central

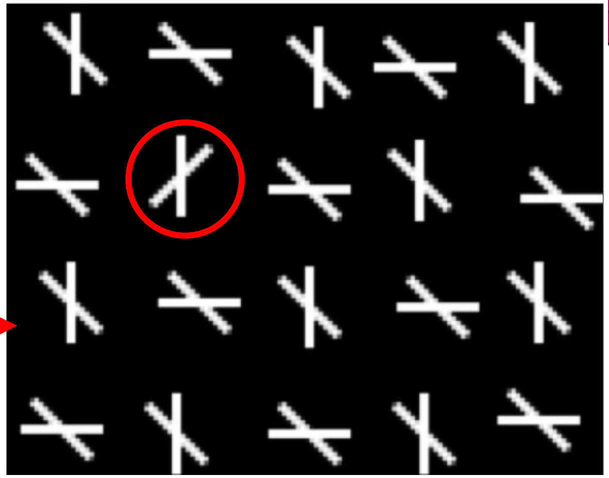
Two separate processes

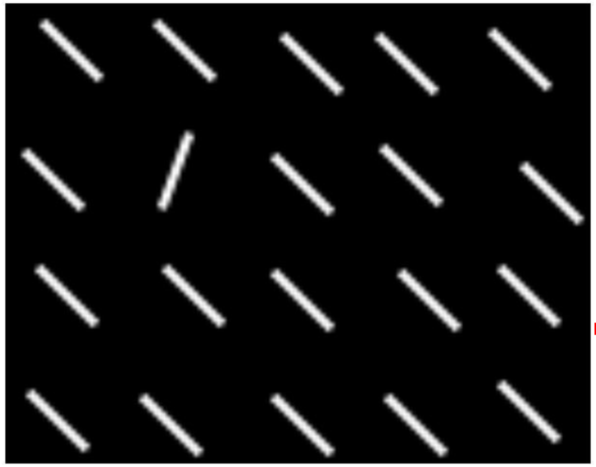
Demo: (Zhaoping & Guyader 2007)





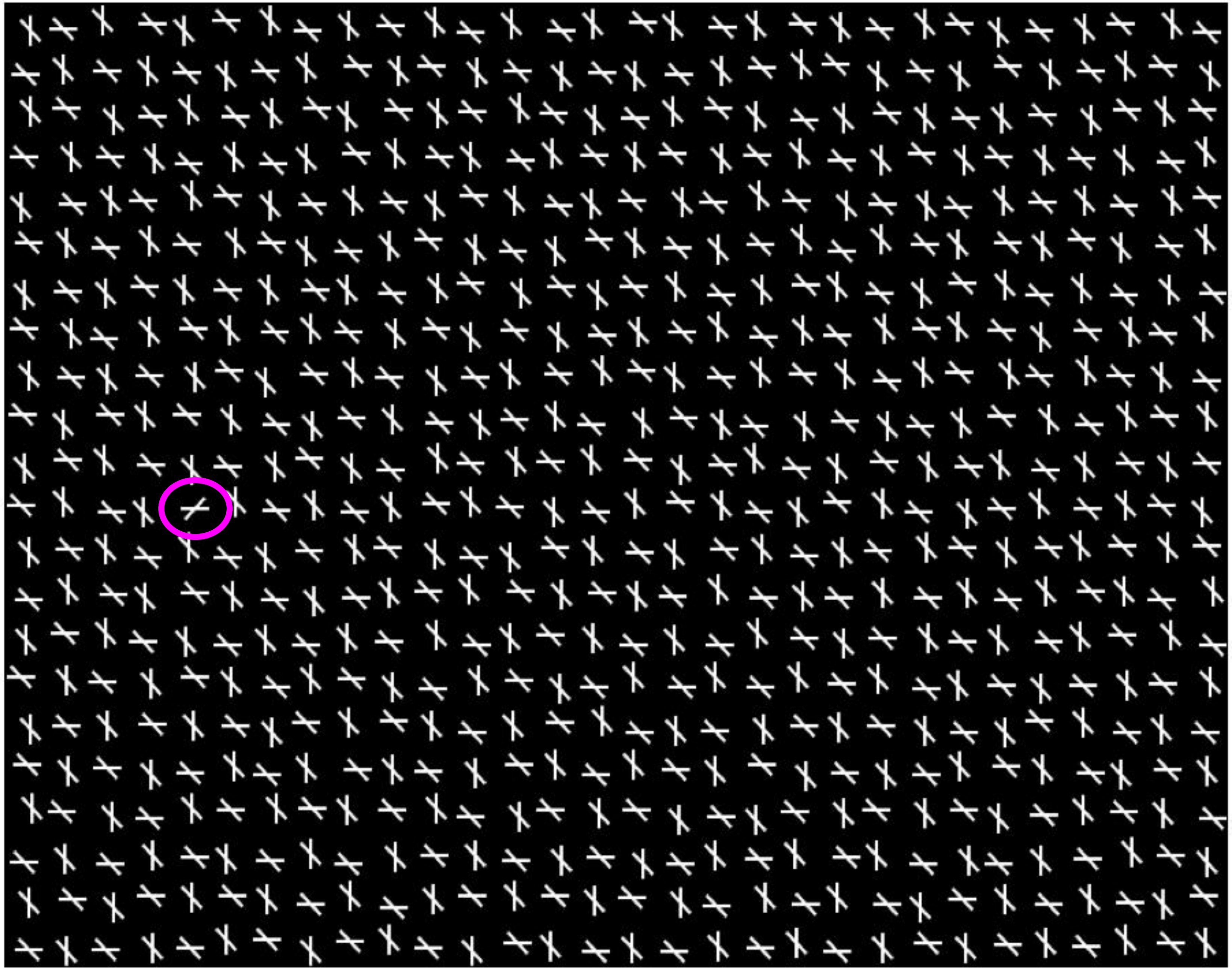
Add a horizontal bar or a vertical bar to each oblique bar



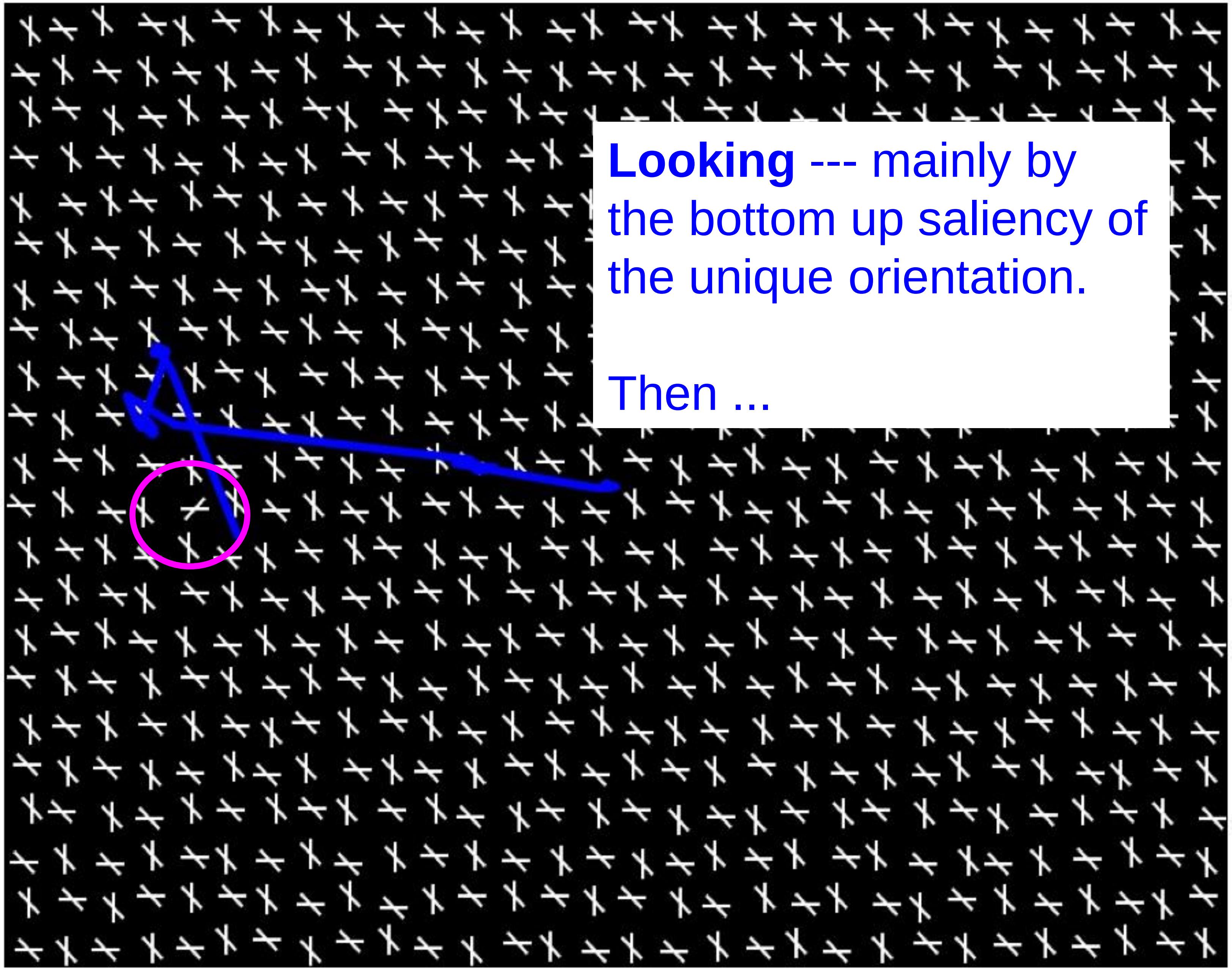


Add a horizontal bar or a vertical bar to each oblique bar

Display span 46x32 degrees in visual angle --- condition A



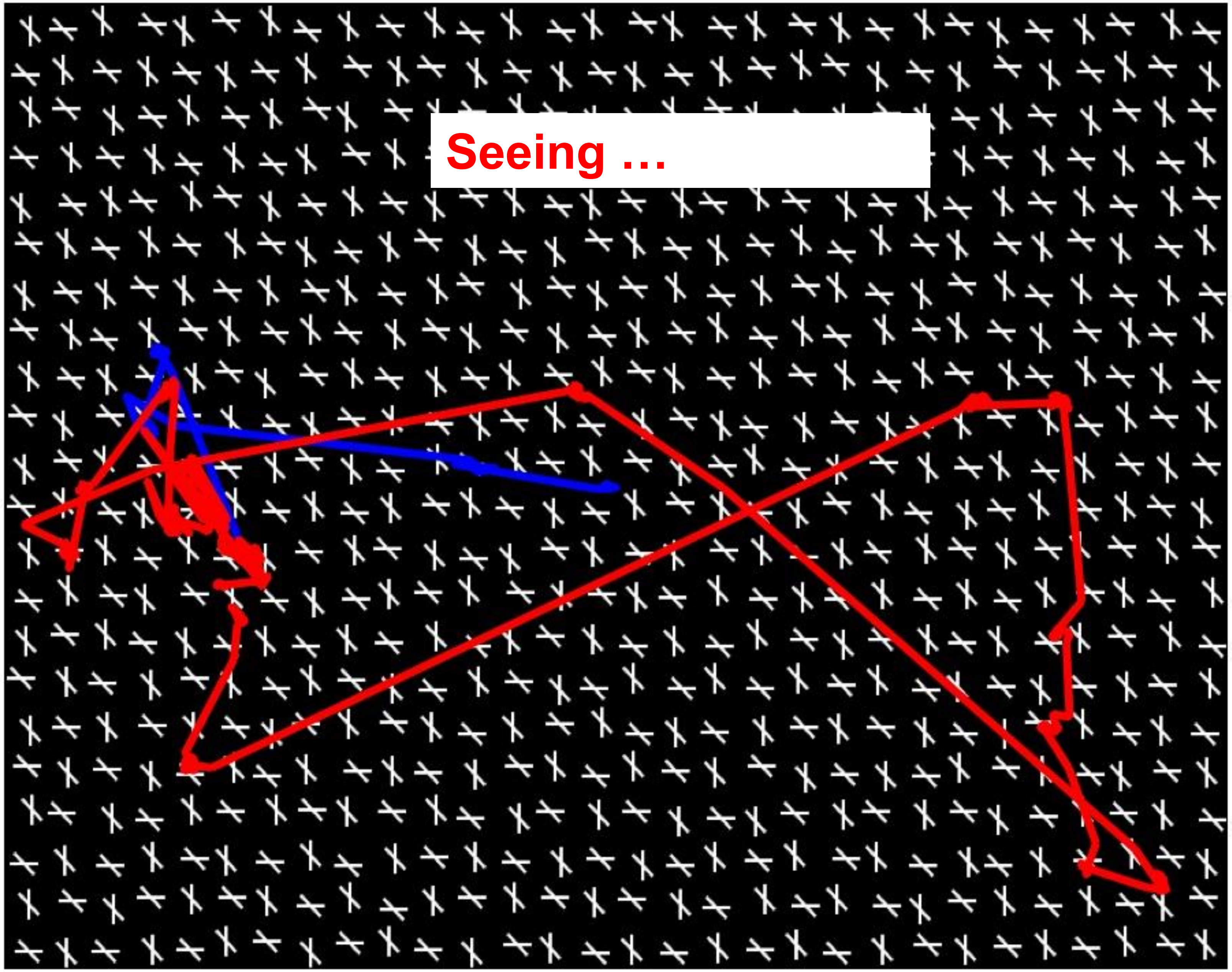
Gaze arrives at target after a few saccades



Looking --- mainly by the bottom up saliency of the unique orientation.

Then ...

Gaze dawdled around the target, then abandoned and returned.



A new path to understanding vision

Visual inputs → Encoding → Selection ⇄ Decoding → for motor action or cognitive decisions

20 frames,
20 megabytes/second

(e.g., by retinal neural activities)

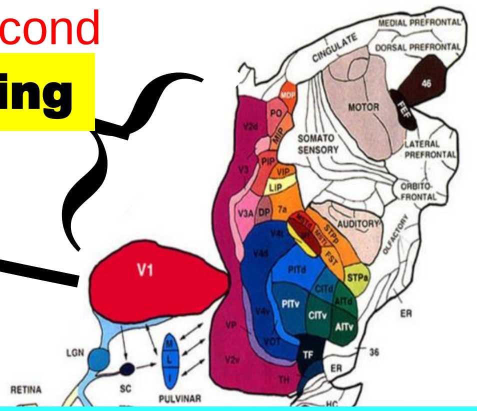
(attentional selection, often by saccading to selected locations)

(e.g., face recognition)

40 bits/second

Looking

Seeing



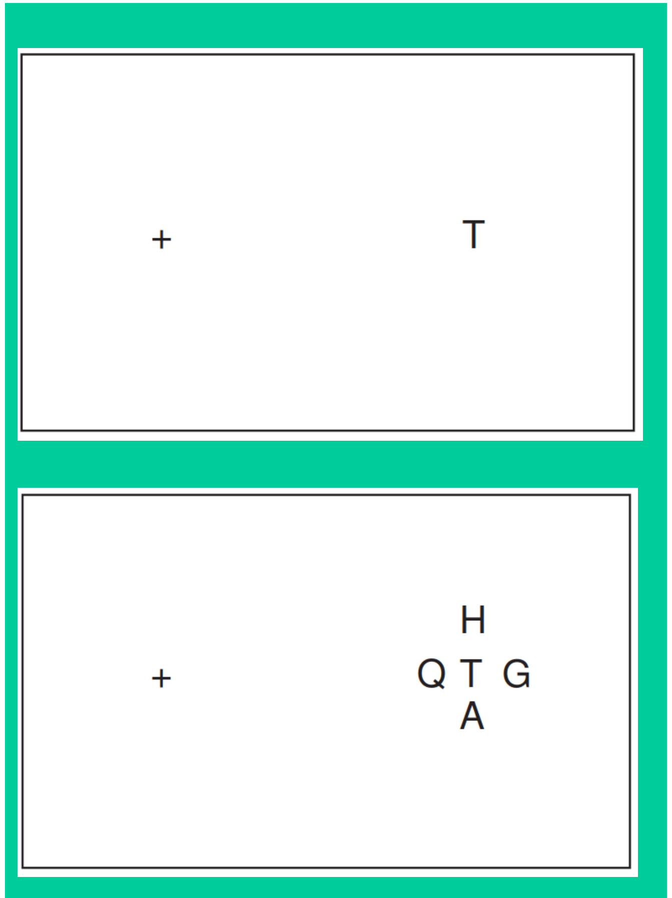
Peripheral vision

Must qualitatively differ in Seein

Central vision

Computational algorithms?

Focus on feedforward-feedback / V1



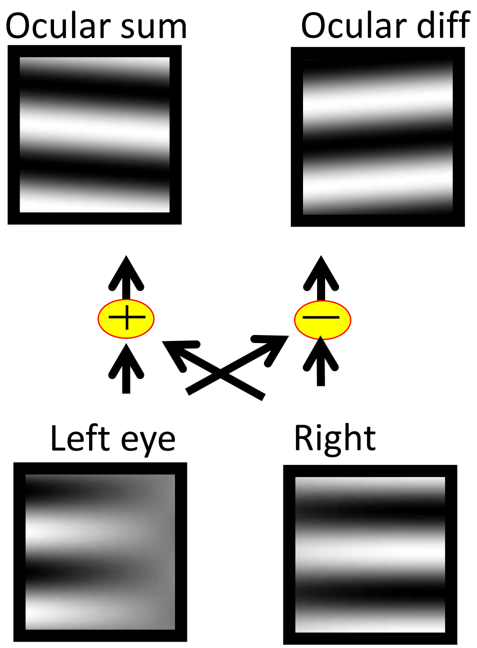
A demo of crowding in the periphery

Summation percept dominant

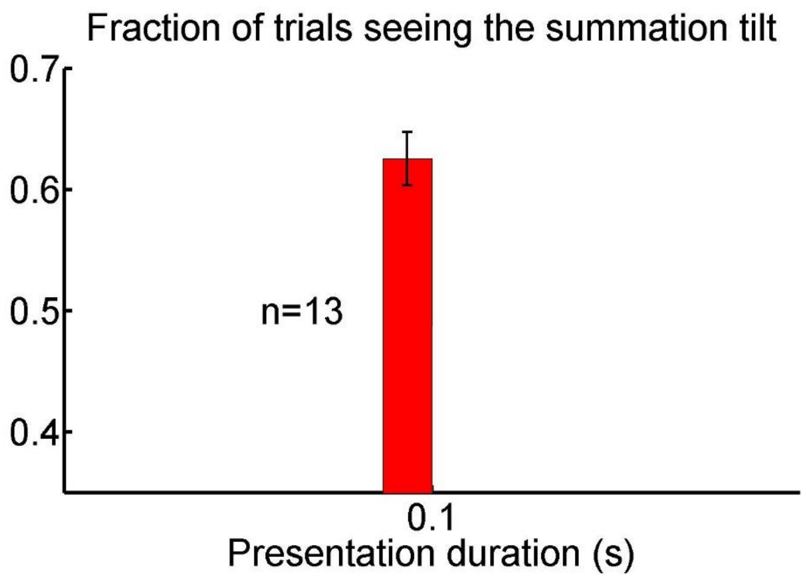
Perception = ?

In V1, signals are efficiently encoded by these two de-correlated channels
(Li & Atick 1994)

$$C_+ = C_-$$



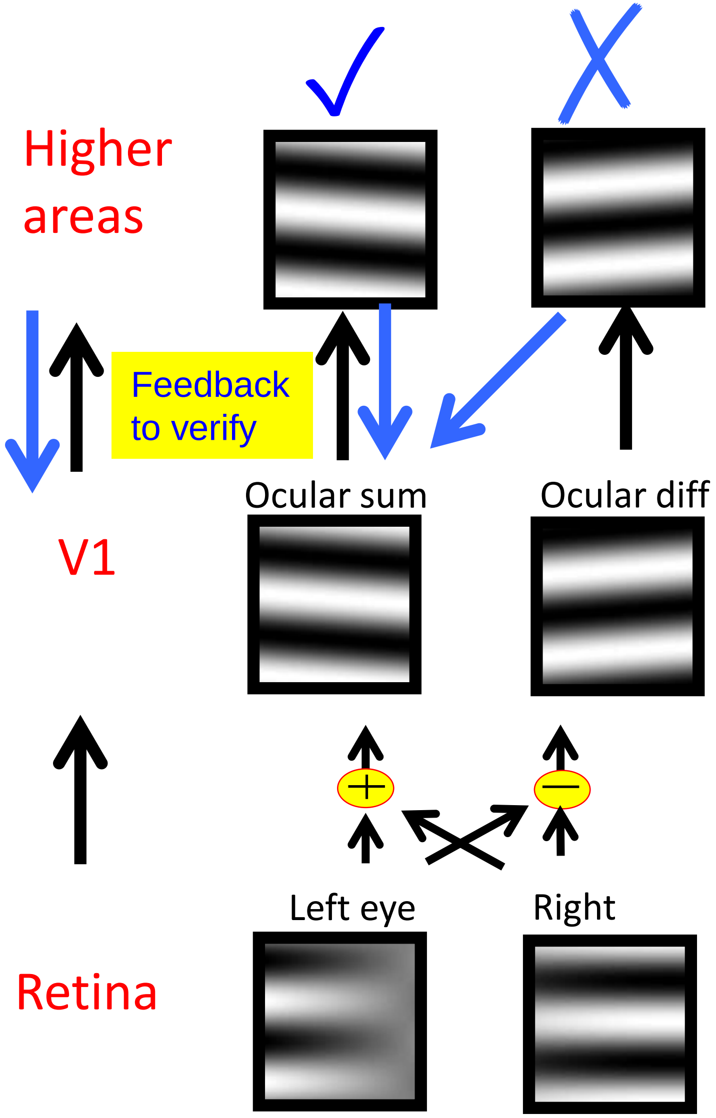
Subject task: report the perceived tilt.



Zhaoping 2017

Why does perception prefer ocular summation? (Zhaoping 2017)

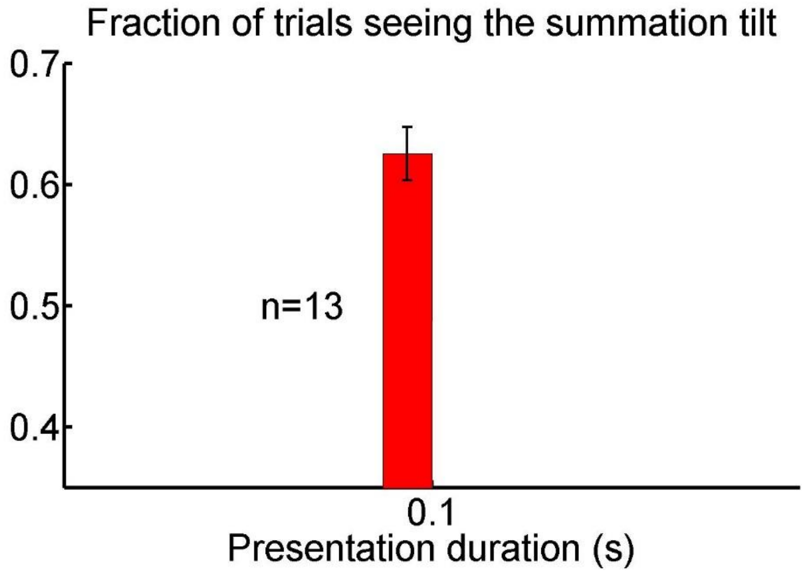
for analysis-by-synthesis (c.f. predictive coding)



Feedforward, feedback, verify, and re-weight (FFVW)

If I perceive it, it is likely (prior) shown to both my left and my right eyes,
so it should resemble the input in the sum channel!!!

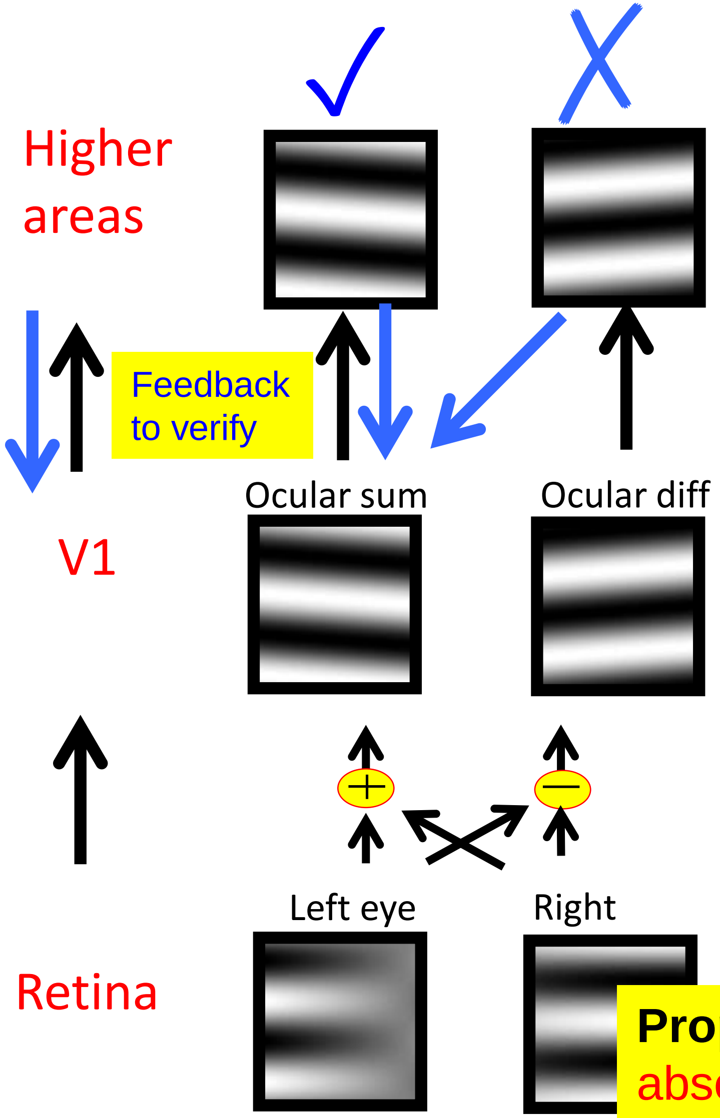
The Bayesian(?) monster



Why does perception prefer ocular summation? (Zhaoping 2017)

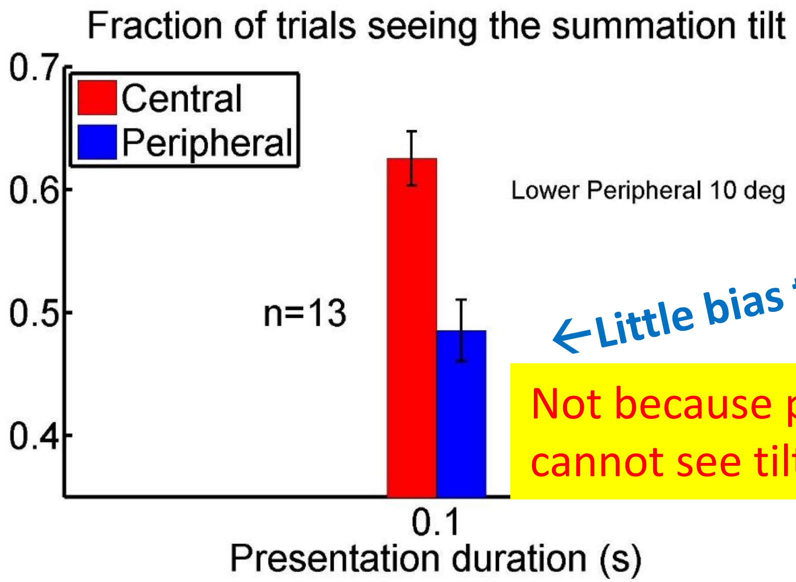
for analysis-by-synthesis (c.f. predictive coding)

Feedforward, feedback, verify, and re-weight (FFVW)



If I perceive it, it is likely (prior) shown to both my left and my right eyes, so it should resemble the input in the sum channel!!!

The Bayesian(?) monster



←Little bias to Sum!

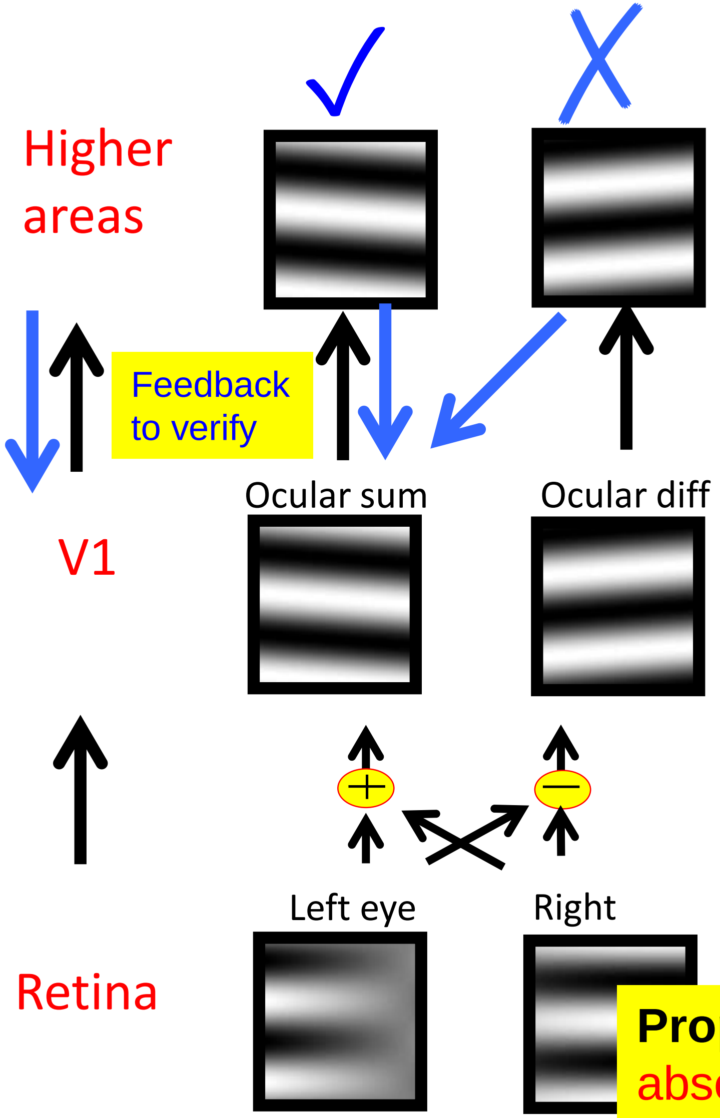
Not because periphery cannot see tilt!

Proposal: Top-down feedback to V1 is weaker or absent in peripheral vision for analysis-by-synthesis!

Why does perception prefer ocular summation? (Zhaoping 2017)

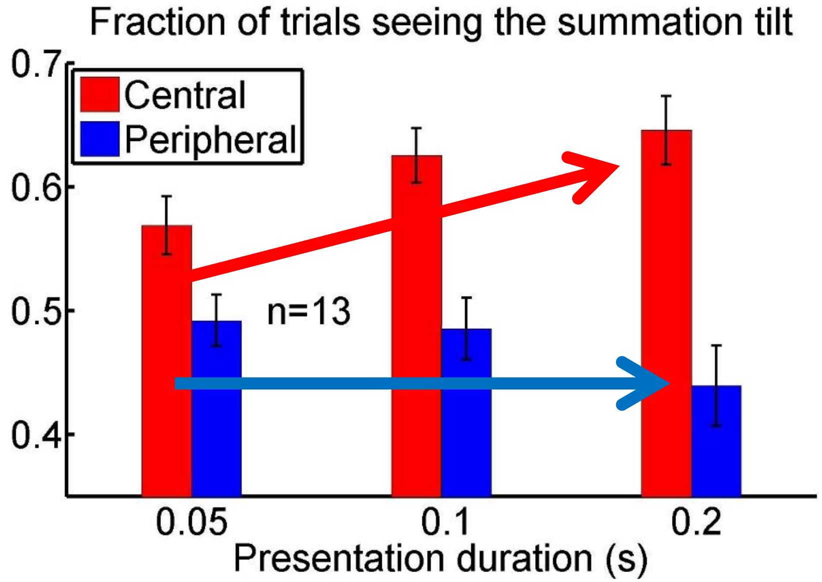
for analysis-by-synthesis
(c.f. predictive coding)

Feedforward, feedback, verify, and re-weight (FFVW)



If I perceive it, it is likely (prior) shown to both my left and my right eyes, so it should resemble the input in the sum channel!!!

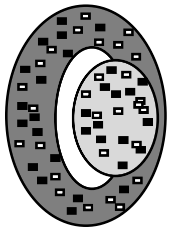
The Bayesian(?) monster



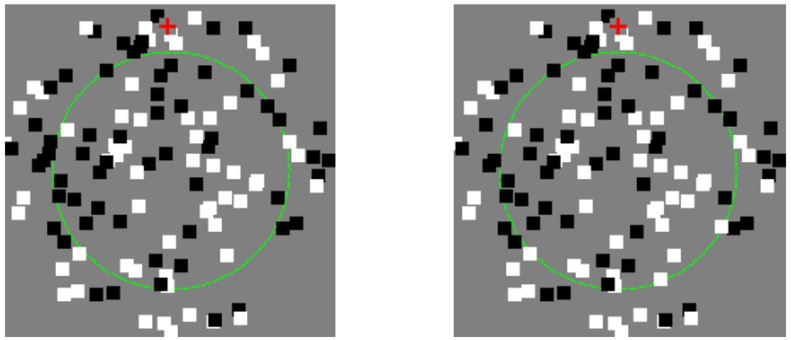
Proposal: Top-down feedback to V1 is weaker or absent in peripheral vision for analysis-by-synthesis!

Testing it in depth perception

Zhaoping & Ackermann, 2018

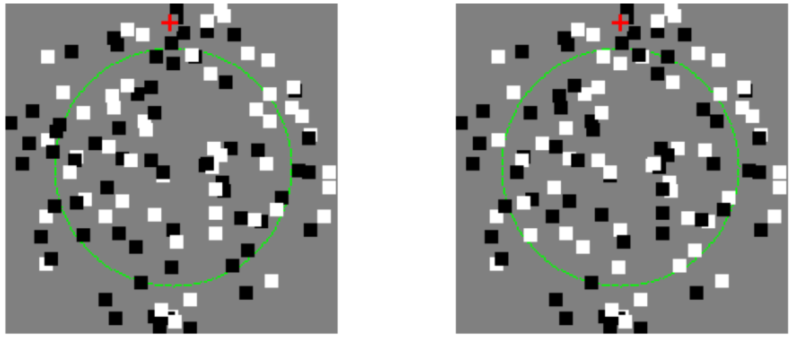


A central disk (non-zero disparity)
and a surrounding ring (zero disparity)



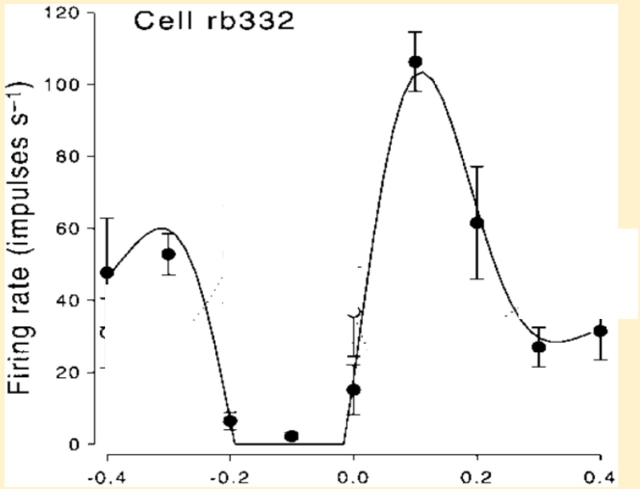
(Green circle not part of the stimuli)

Dots for the
central disk
Anti-correlated



Doi et al 2011

A V1 neuron's response to disparity in random dot stereograms



Disparity (degrees)

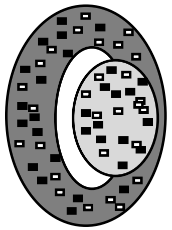
(Cumming and Parker 1997)

Proposal: Top-down feedback to V1 is weaker or absent in peripheral vision for analysis-by-synthesis!

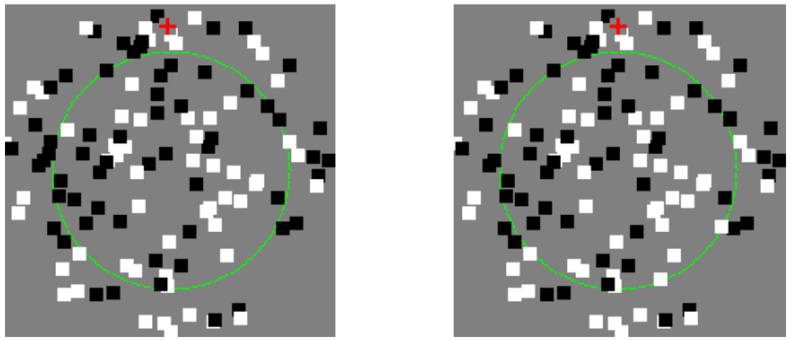
Proposal: Top-down feedback to V1 is weaker or absent in peripheral vision for analysis-by-synthesis!

Testing it in depth perception

Zhaoping & Ackermann, 2018

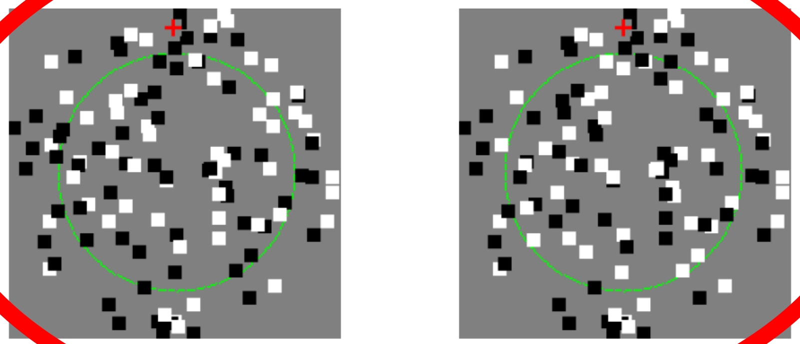


A central disk (non-zero disparity)
and a surrounding ring (zero disparity)



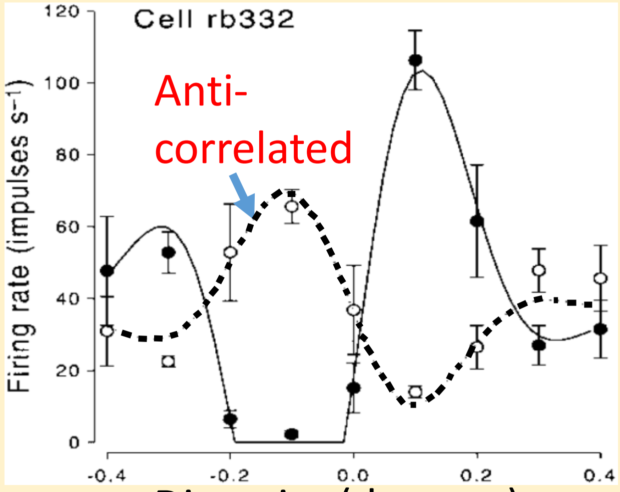
(Green circle not part of the stimuli)

Dots for the central disk Anti-correlated



Doi et al 2011

A V1 neuron's response to disparity in random dot stereograms



Disparity (degrees)

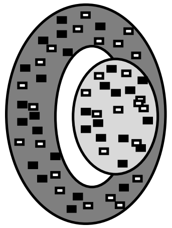
(Cumming and Parker 1997)

Humans cannot see depth in such stereograms in central vision

Proposal: Top-down feedback to V1 is weaker or absent in peripheral vision for analysis-by-synthesis!

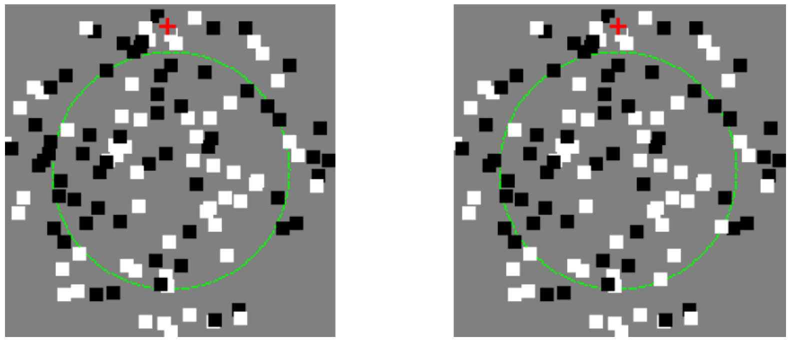
Testing it in depth perception

Zhaoping & Ackermann, 2018

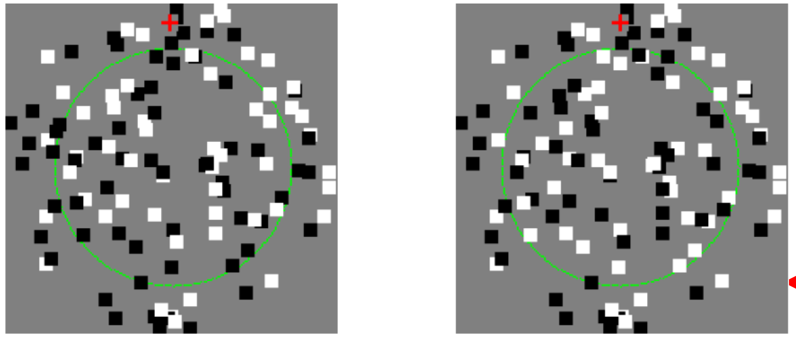


A central disk (non-zero disparity) and a surrounding ring (zero disparity)

Dots for the central disk correlated



Dots for the central disk Anti-correlated



No reversed depth percept

V1's report vetoed

Expected disparity and binocular correlation not found in V1

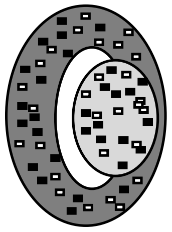
Top-down feedback to verify V1's report

V1 feeds forward reverse depth to higher brain areas!

Proposal: Top-down feedback to V1 is weaker or absent in peripheral vision for analysis-by-synthesis!

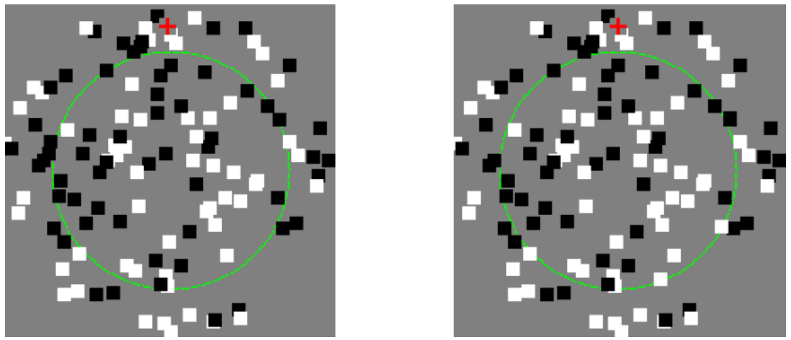
Testing it in depth perception

Zhaoping & Ackermann, 2018

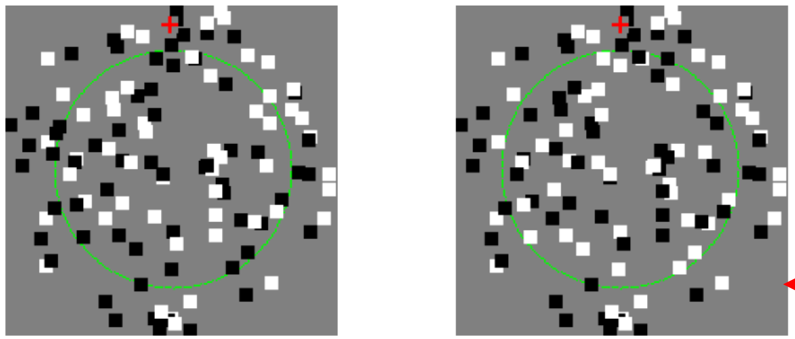


A central disk (non-zero disparity) and a surrounding ring (zero disparity)

Dots for the central disk correlated



Dots for the central disk Anti-correlated



Prediction

No reversed depth percept

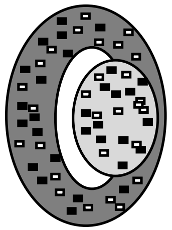
If peripheral vision has no feedback

V1 feeds forward reverse depth to higher brain areas!

Proposal: Top-down feedback to V1 is weaker or absent in peripheral vision for analysis-by-synthesis!

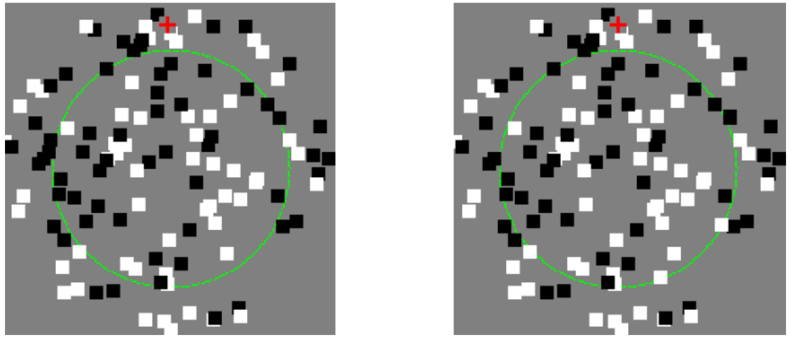
Testing it in depth perception

Zhaoping & Ackermann, 2018

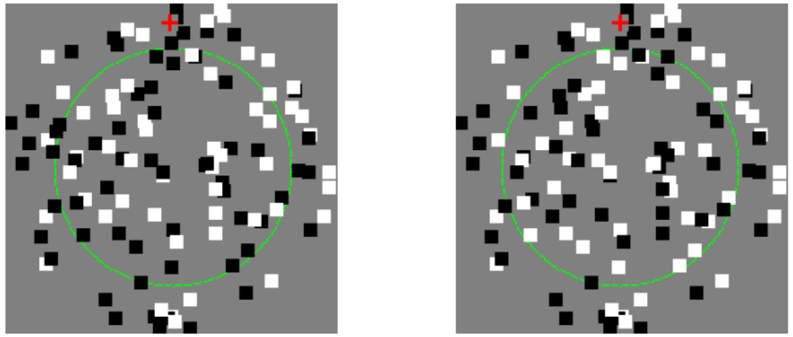


A central disk (non-zero disparity) and a surrounding ring (zero disparity)

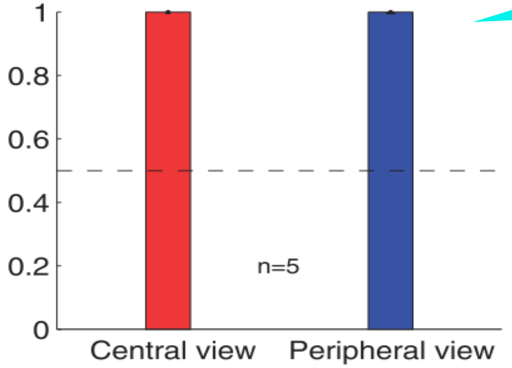
Dots for the central disk correlated



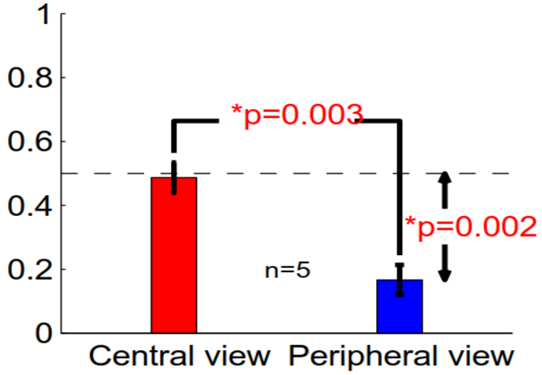
Dots for the central disk Anti-correlated



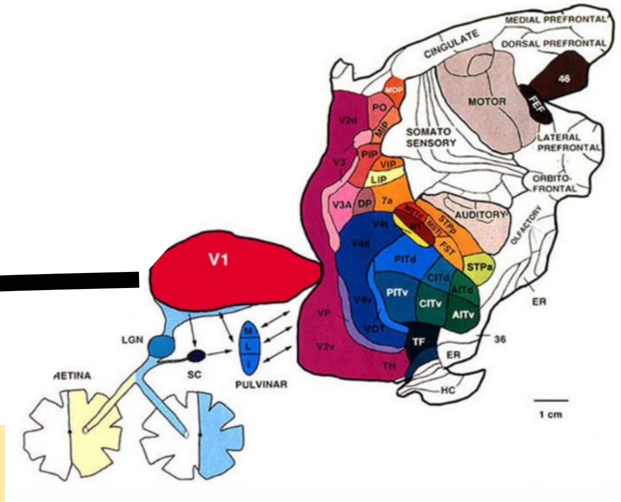
Prediction



Accuracy reporting disparity-defined depth



Thanks to
Funding source: Gatsby Foundation, EPSRC, BBSRC.
Collaborators: Wu Li, Yin Yan, Ansgar Koene, Li Zhe, Joealle Ackermann, etc



A new path to understanding vision

Visual inputs → Encoding → Selection ⇄ Decoding → for motor action or cognitive decisions

(e.g., by retinal neural activities)

(attentional selection, often by saccading to selected locations)

(e.g., face recognition)

20 frames,
20 megabytes/second

40 bits/second

Central-peripheral dichotomy

Falsifiable

Opening the window to our brain