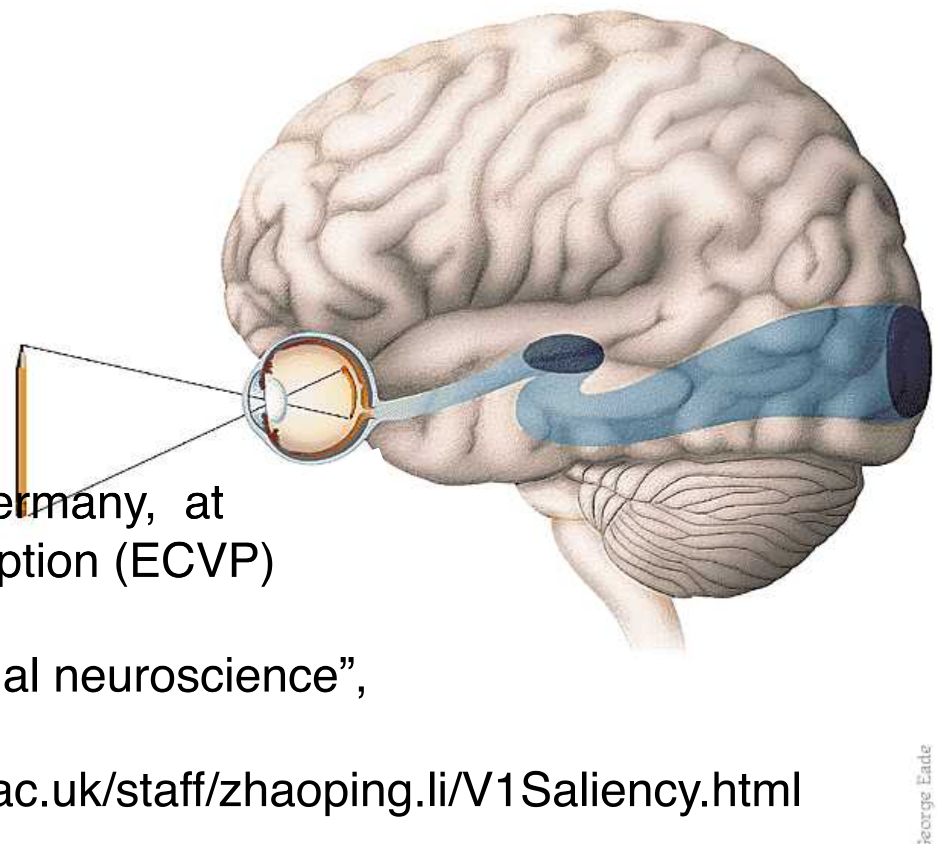
A theory of the primary visual cortex (V1):

predictions,

experimental tests,

and implications for future research



Li Zhaoping

University College London

Presented Aug. 27, 2013, Bremen, Germany, at
European conference on visual perception (ECVP)
in the plenary symposium
“Visual perception meets computational neuroscience”,

More information: http://www0.cs.ucl.ac.uk/staff/zhaoping.li/V1Saliency.html

Talk outline

Motivating questions:

Theory:

Predictions and tests:

Implications.

Talk outline

Motivating questions:

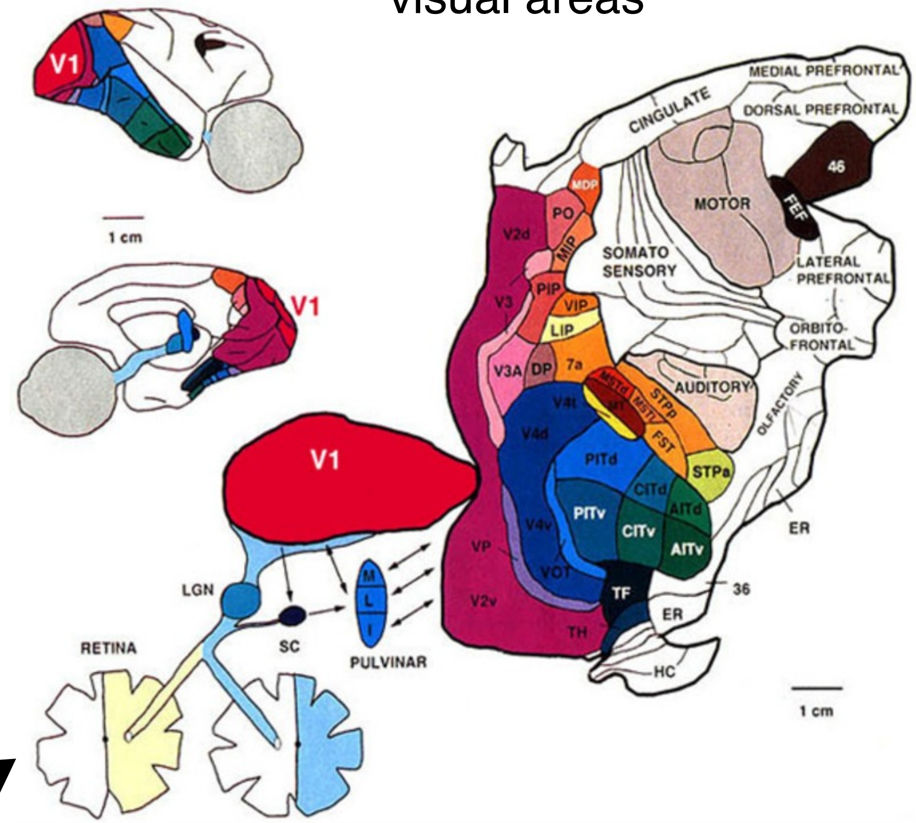
(1) What is V1 doing?

Theory:

Predictions and tests:

Implications.

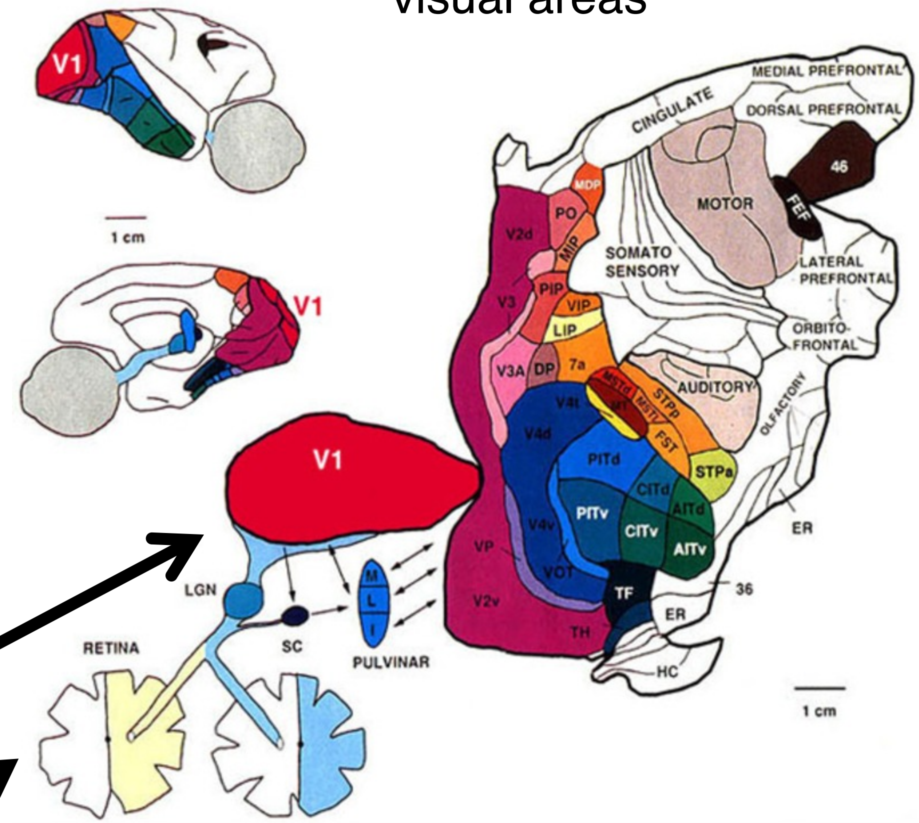
Organization of visual areas



1953, Stephen Kuffler

Center-surround feature detectors in retina

Organization of visual areas



1960s-70s, Hubel and Wiesel

Bar/edge feature detectors in V1

1953, Stephen Kuffler

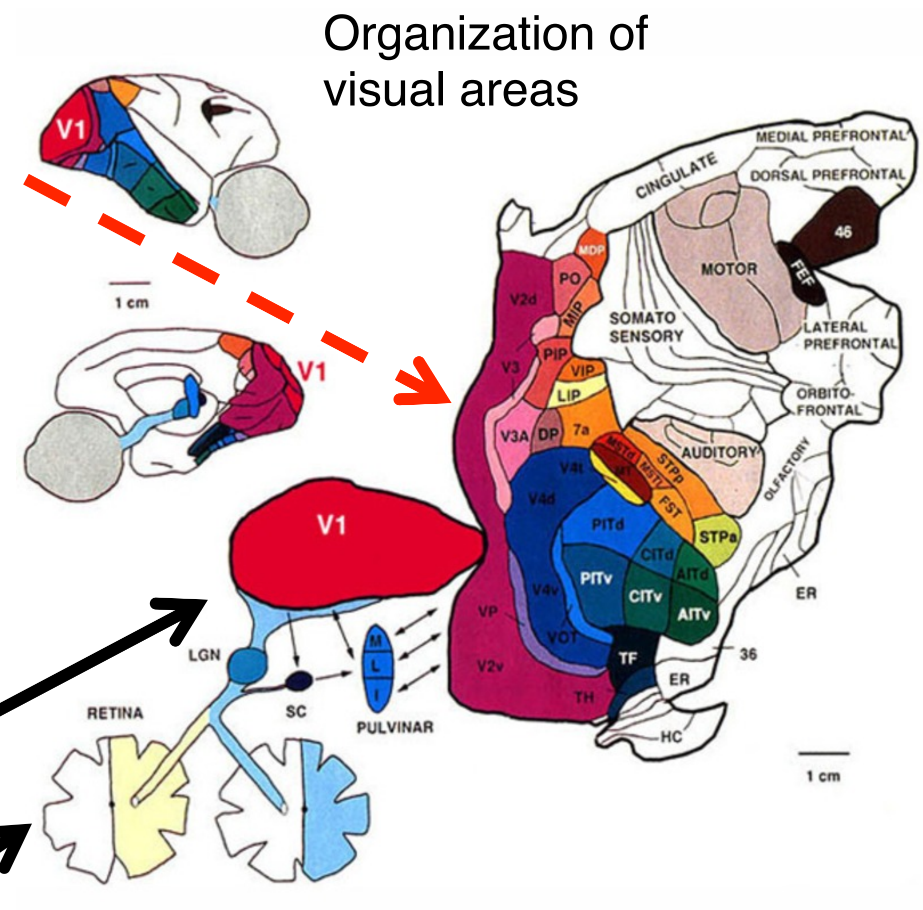
Center-surround feature detectors in retina

1980-200? --- V2, V3, V4 …
No more data as illuminating

The feature detection idea no longer applies around V1-V2?

Calling for new ideas!!!
--- **What is V1 doing?**

Organization of visual areas



1960s-70s, Hubel and Wiesel
Bar/edge feature detectors in V1

1953, Stephen Kuffler
Center-surround feature detectors in retina

Motivating question: what exactly is V1 doing?

Detect edges/bars ----for what?

Prepares information for later?

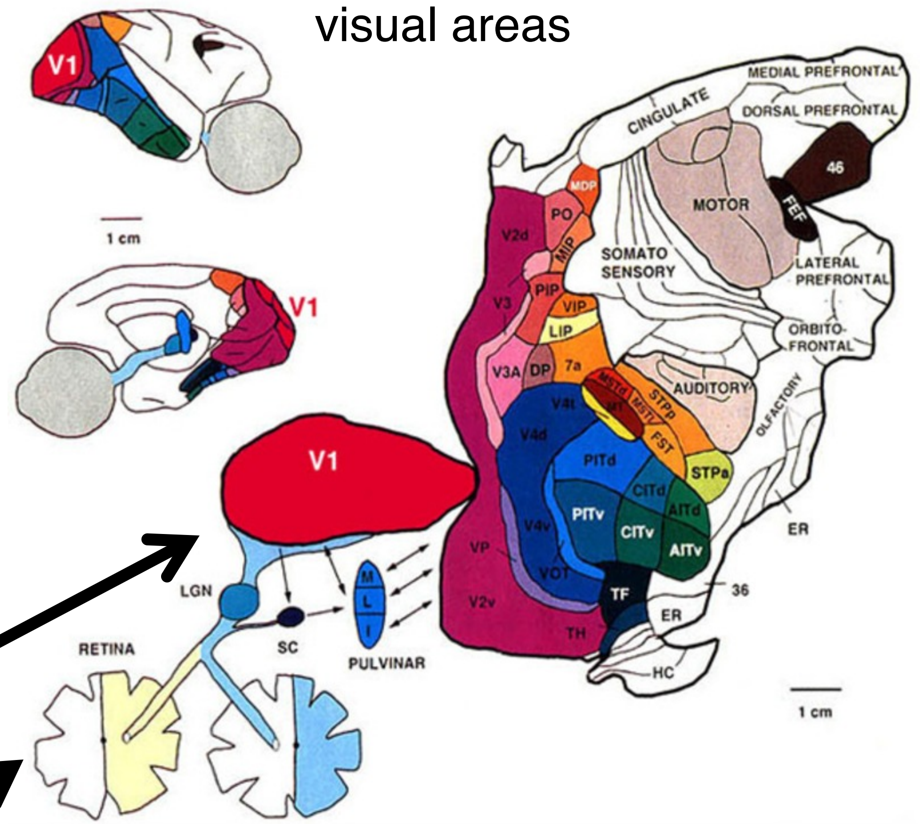
Verifies top-down information?

The back-office role?
But V1 is huge! (and thus expensive)

1960s-70s, Hubel and Wiesel
Bar/edge feature detectors in V1

1953, Stephen Kuffler
Center-surround feature detectors in retina

Organization of visual areas



Talk outline

Motivating questions:

(1) What is V1 doing?

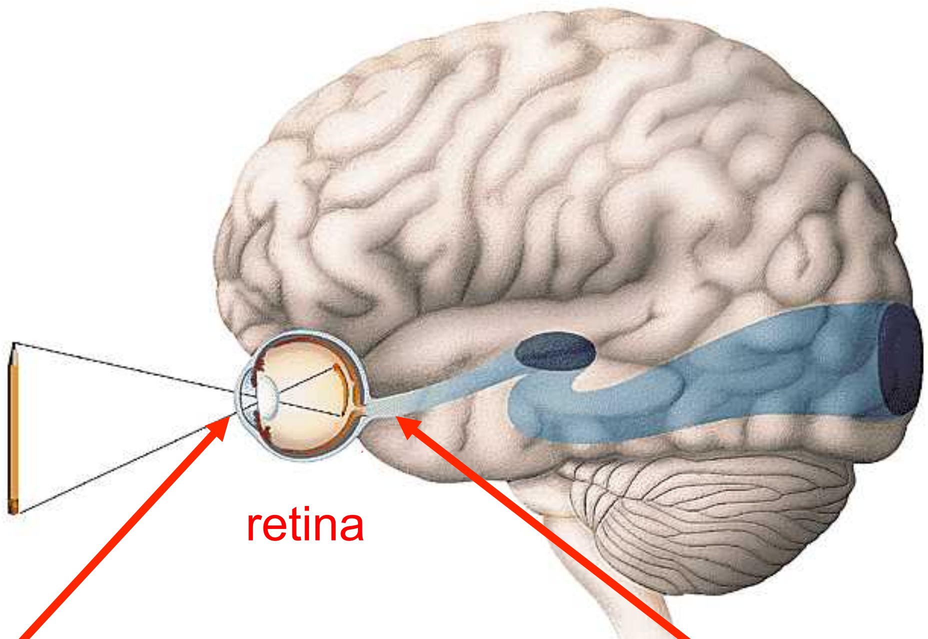
(2) Which brain areas control the direction of attention? ←

Theory:

Predictions and tests:

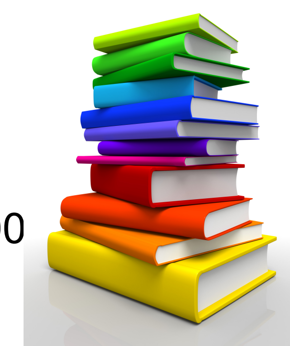
Implications.

Information bottlenecks in the visual pathway:



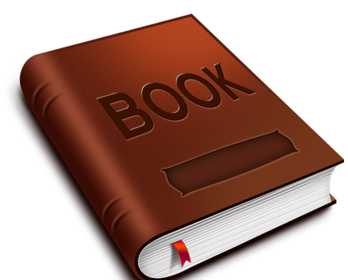
retina

Primary visual cortex
(V1)



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

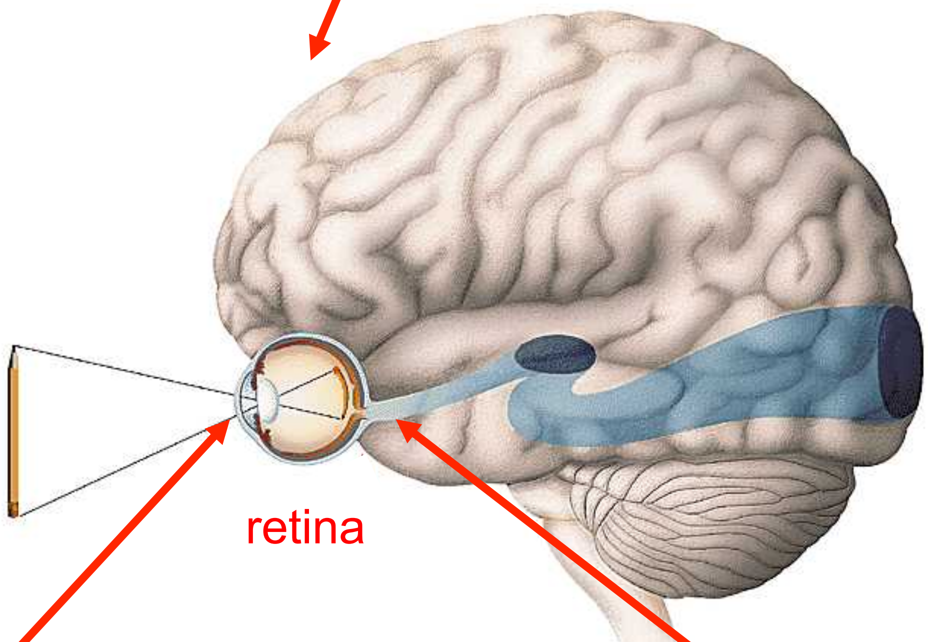
10^7 bits/second
~ 10^6 neurons ,
~10 spikes/neuron
~1 spike/second



Information bottlenecks in the visual pathway:

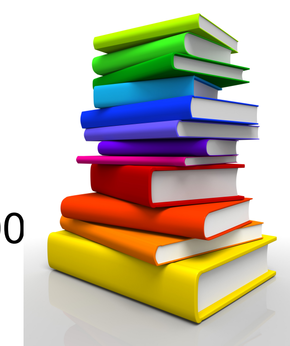
Attentional bottleneck ~ 40 bits/second

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

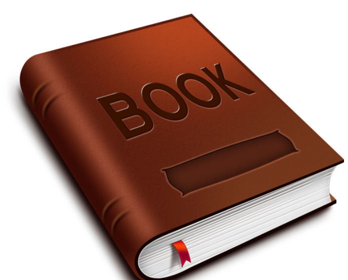


Primary visual cortex (V1)

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



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



Demo of information deletion --- change blindness

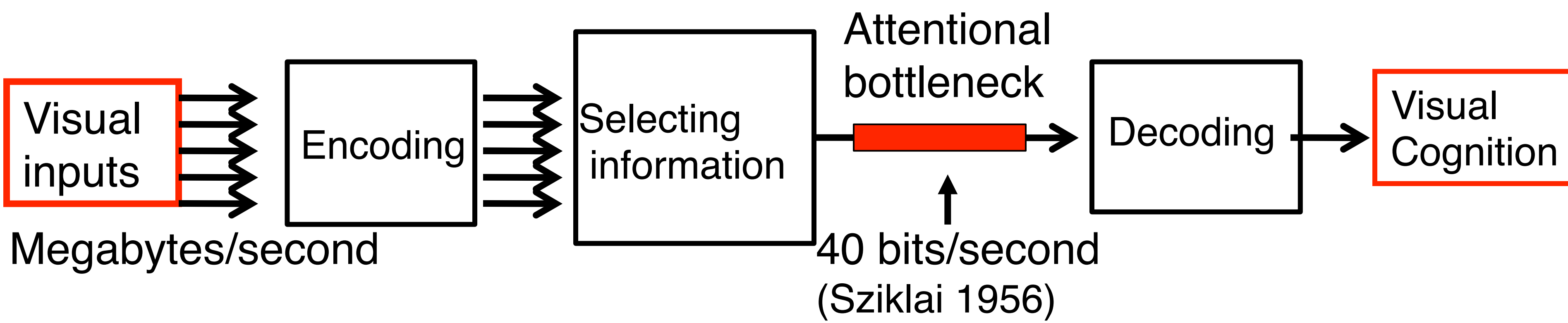
Inattentional blindness — spotting the difference between the two images

We are blind to almost everything except the tiny bit that we pay attention to!

Demo of information deletion --- change blindness

Inattentional blindness — spotting the difference between the two images

We are blind to almost everything except the tiny bit that we pay attention to!



Attention is guided by bottom-up or top-down factors.

Talk outline

Motivating questions:

(1) What is V1 doing?

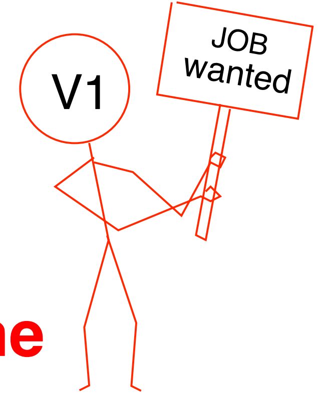
(2) Which brain areas control the direction of attention?

Theory:

Predictions and tests:

Implications.

Talk outline



Motivating questions:

(1) What is V1 doing?

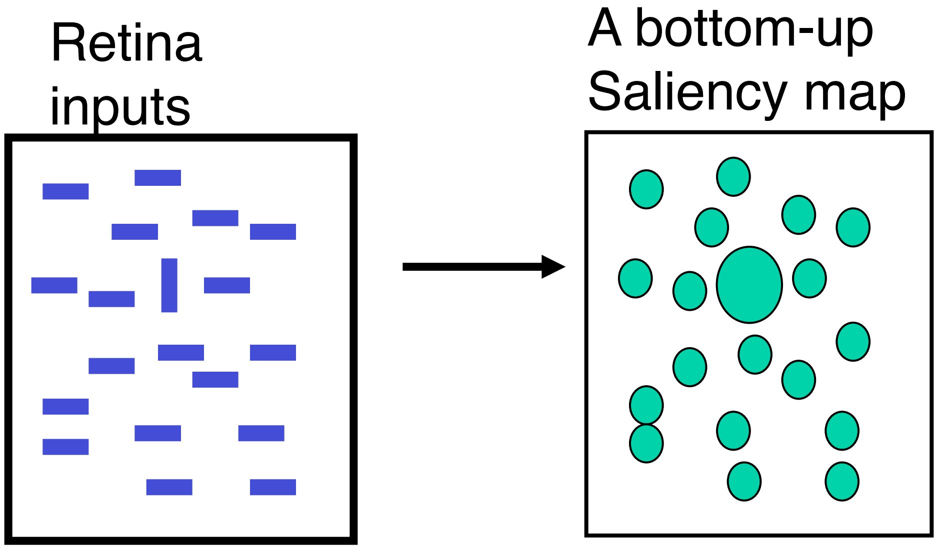
(2) Which brain areas control the direction of attention?

Theory:

A bottom-up saliency map in V1

Predictions and tests:

Implications.

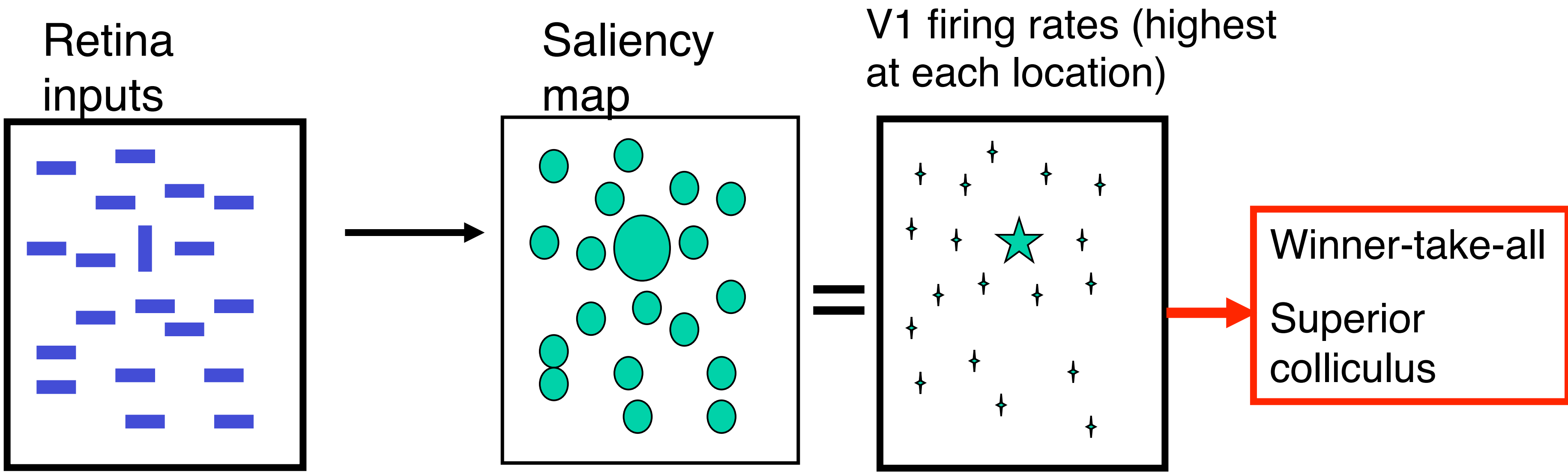


To guide attentional selection.

(Koch & Ullman 1985, Wolfe et al 1989, Itti & Koch 2000, etc.)

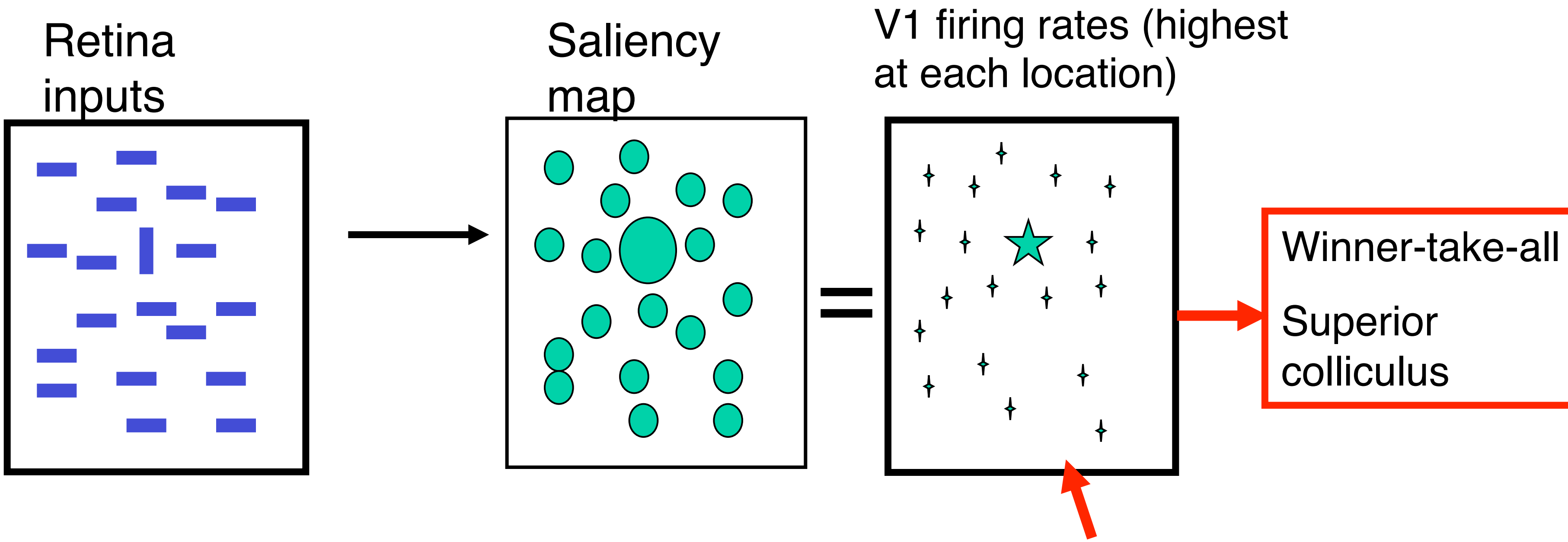
Shorter reaction time (RT) of a saccade to a target, or to find a target → higher saliency.

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

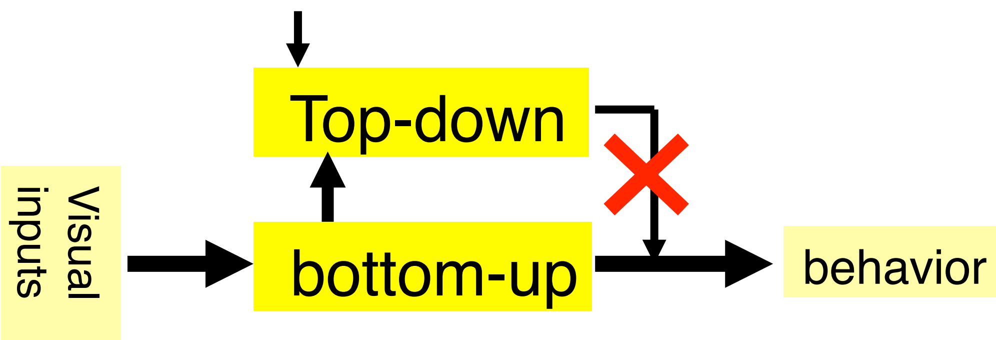


The V1 Saliency Hypothesis:

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

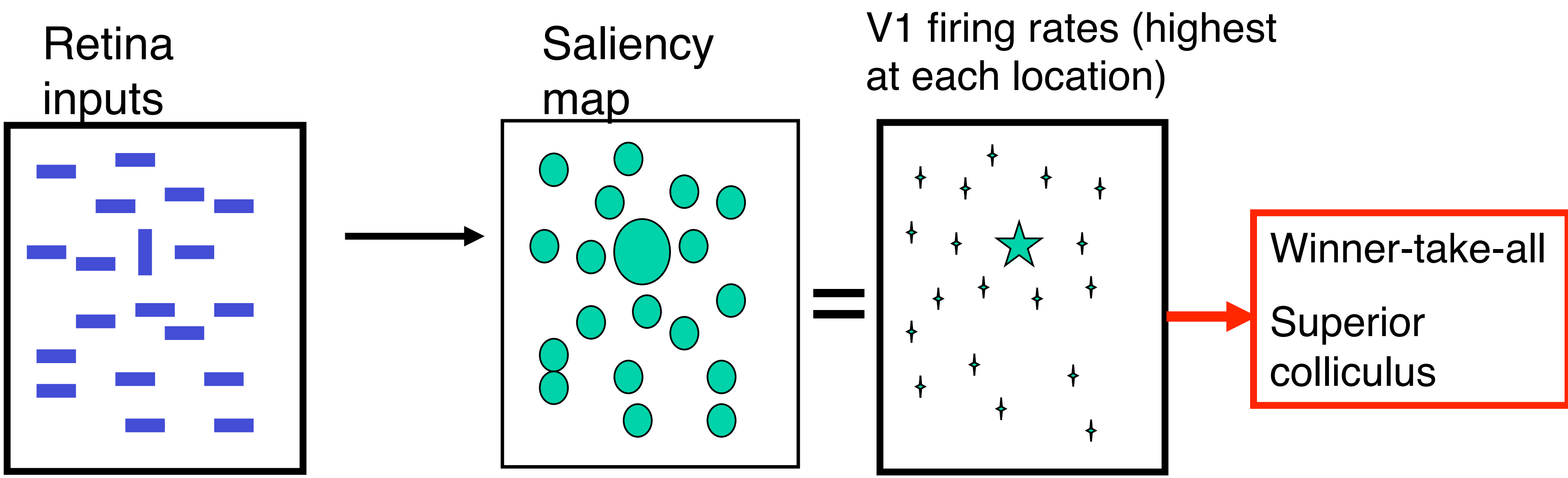


Neural activities as universal currency to bid for visual selection.
The receptive field of the most active V1 cells is selected

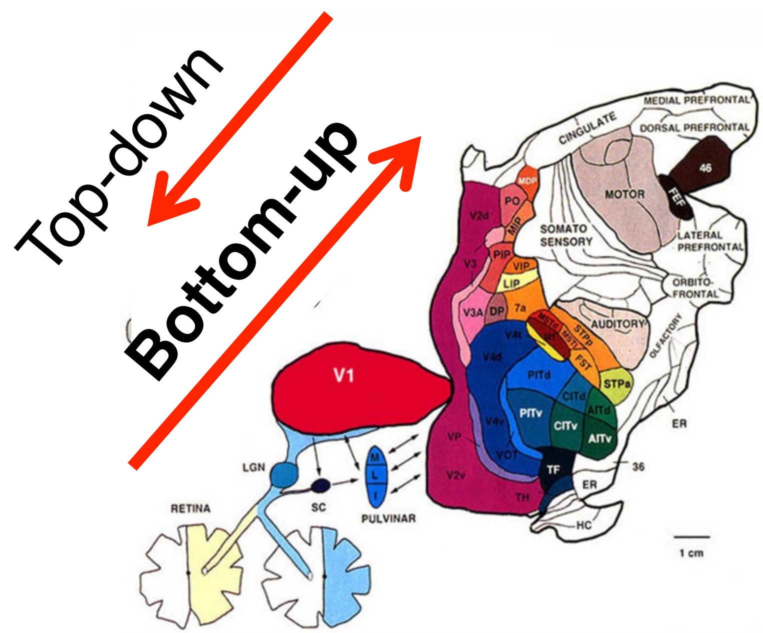


The V1 Saliency Hypothesis:

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

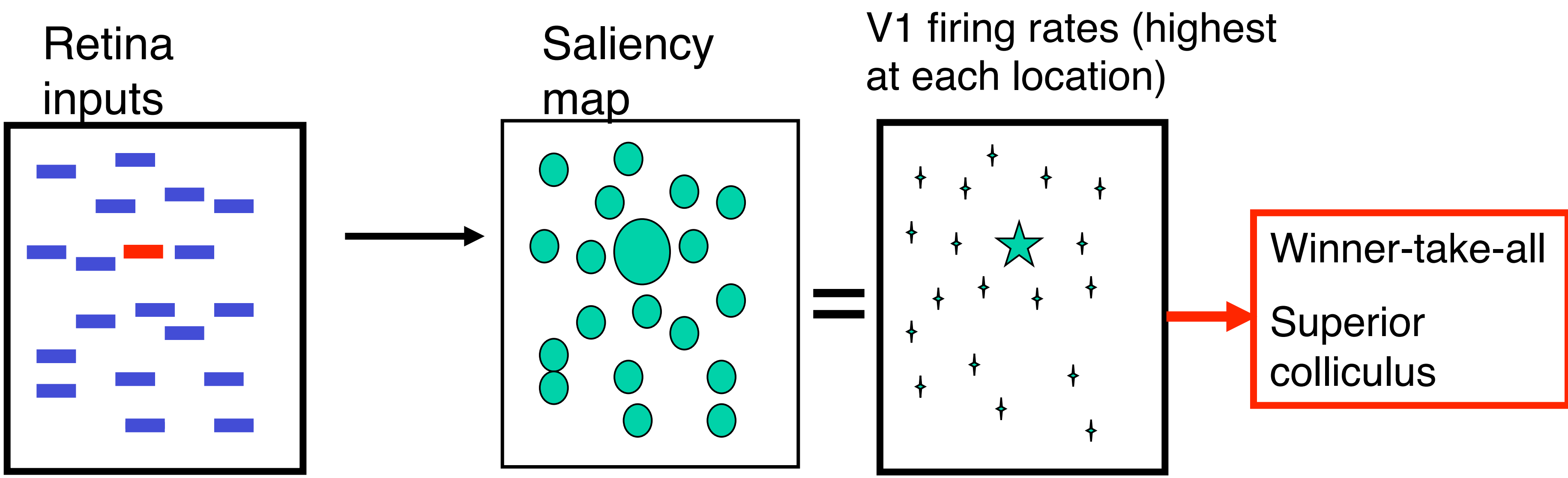


This hypothesis is against traditional wisdoms which presume that higher cortical areas guide the attentional selection (Treisman, Koch, Desimone, itti , etc).

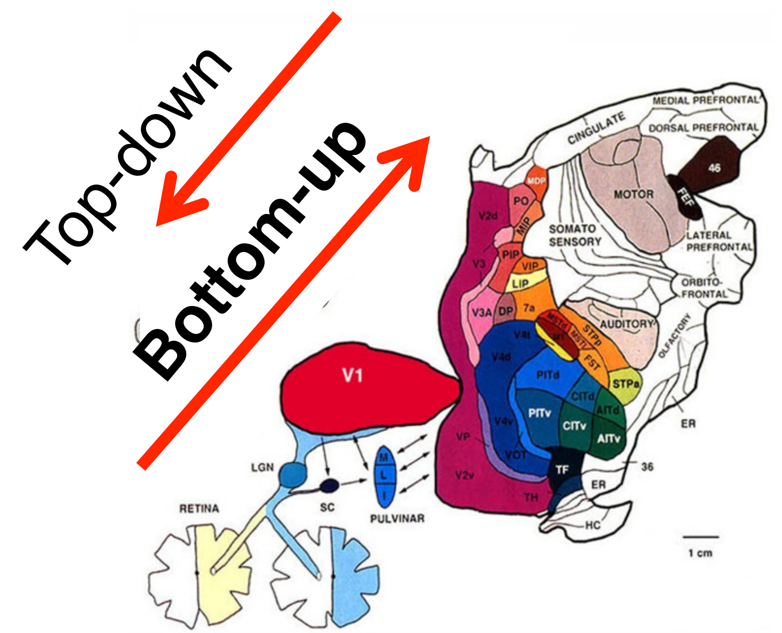


The V1 Saliency Hypothesis:

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

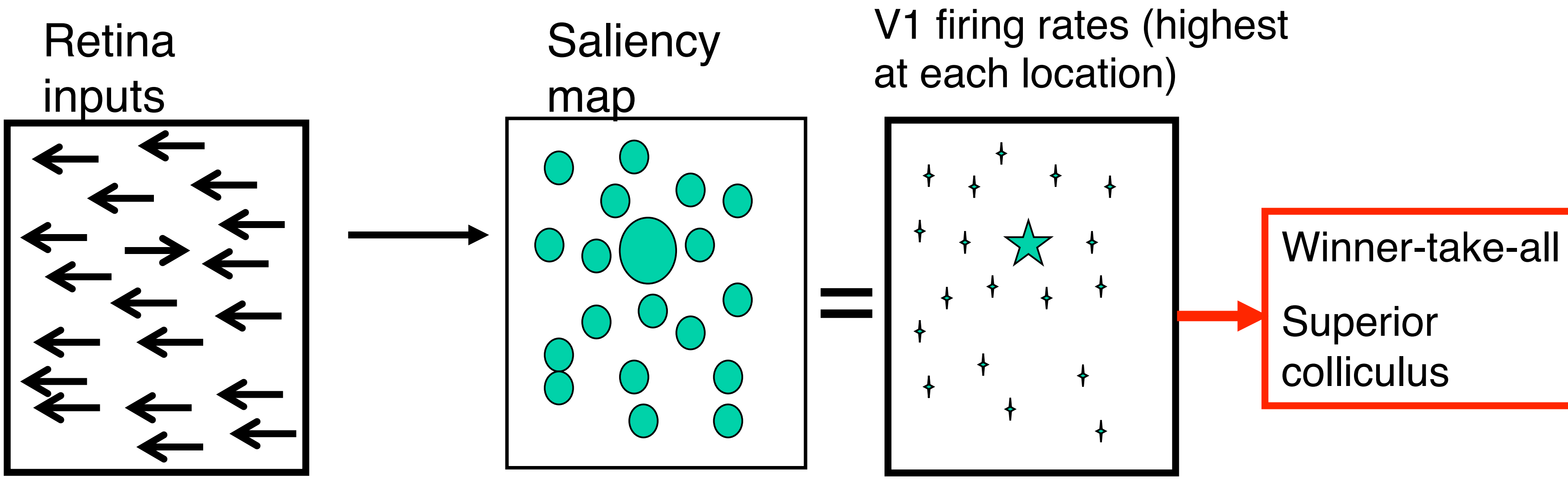


This hypothesis is against traditional wisdoms which presume that higher cortical areas guide the attentional selection (Treisman, Koch, Desimone, itti , etc).

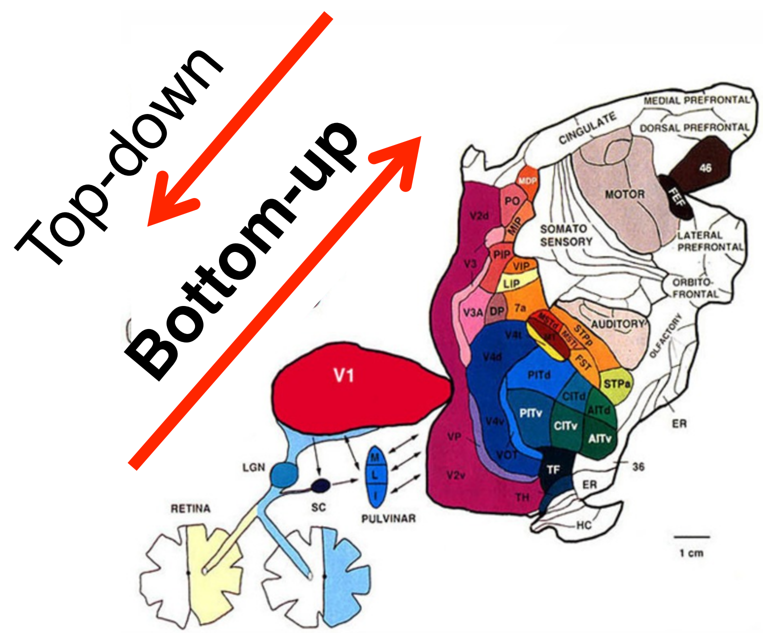


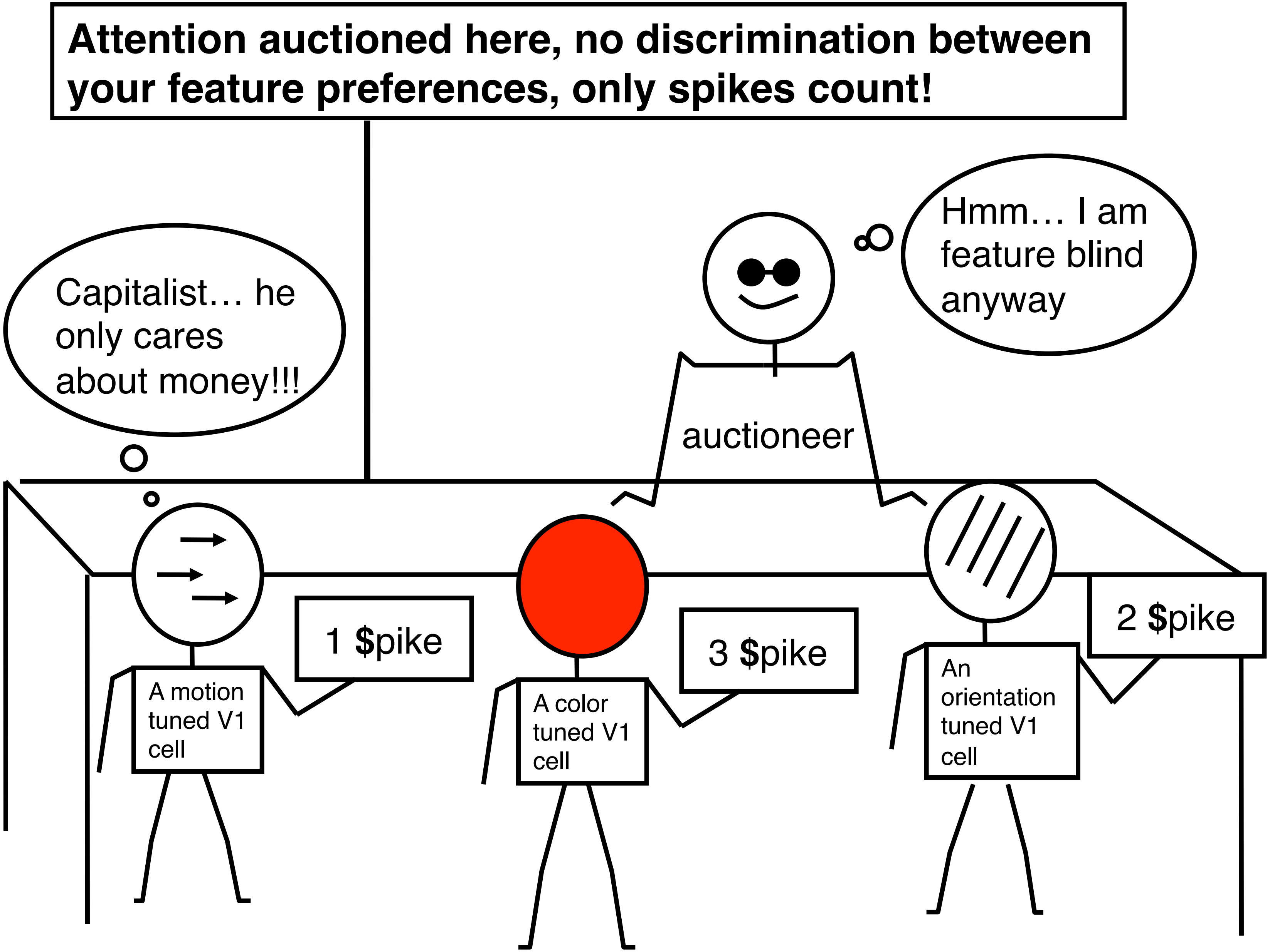
The V1 Saliency Hypothesis:

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



This hypothesis is against traditional wisdoms which presume that higher cortical areas guide the attentional selection (Treisman, Koch, Desimone, itti , etc).

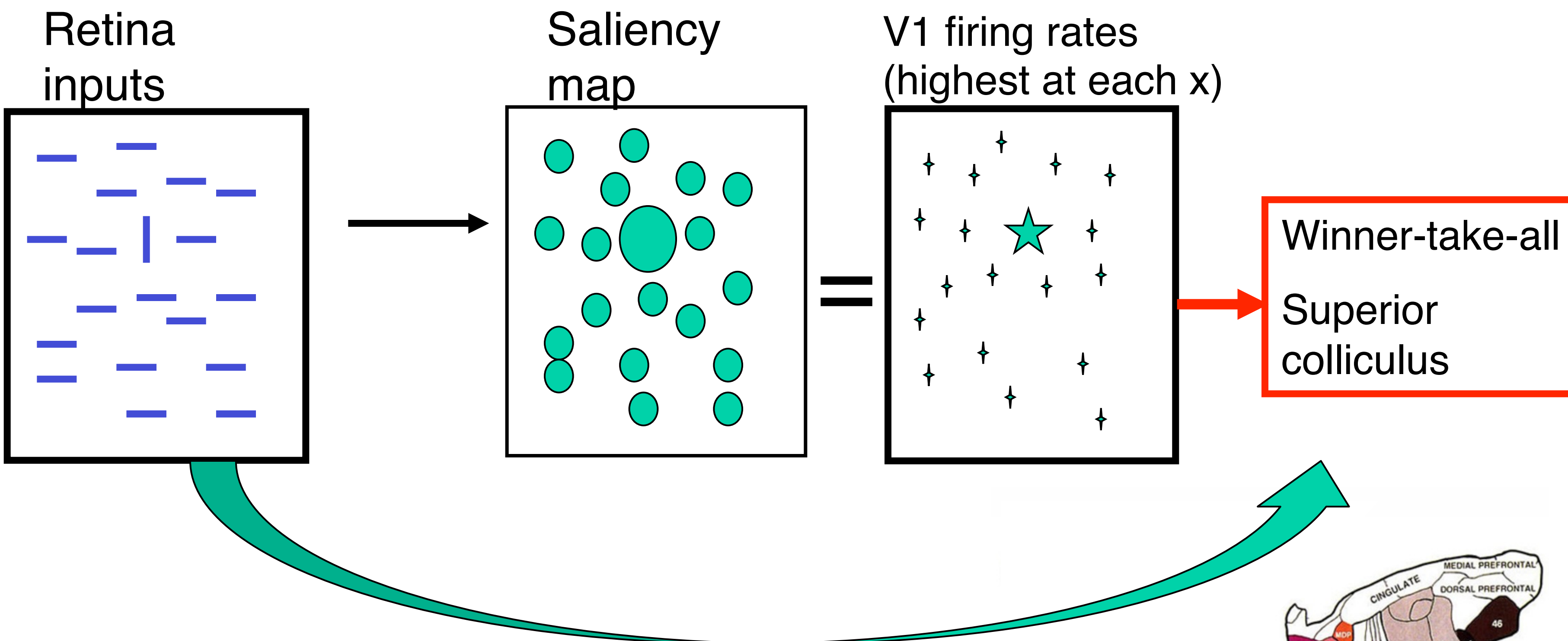


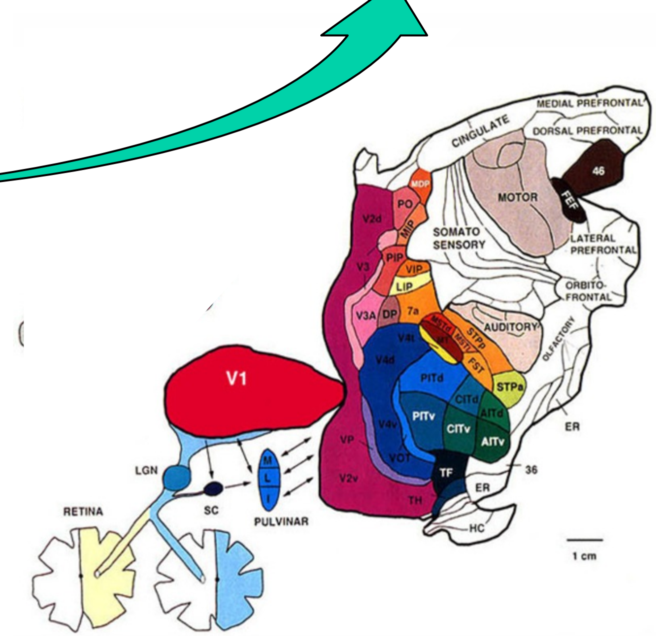


Zhaoping L. 2006, Network: computation in neural systems

The V1 Saliency Hypothesis:

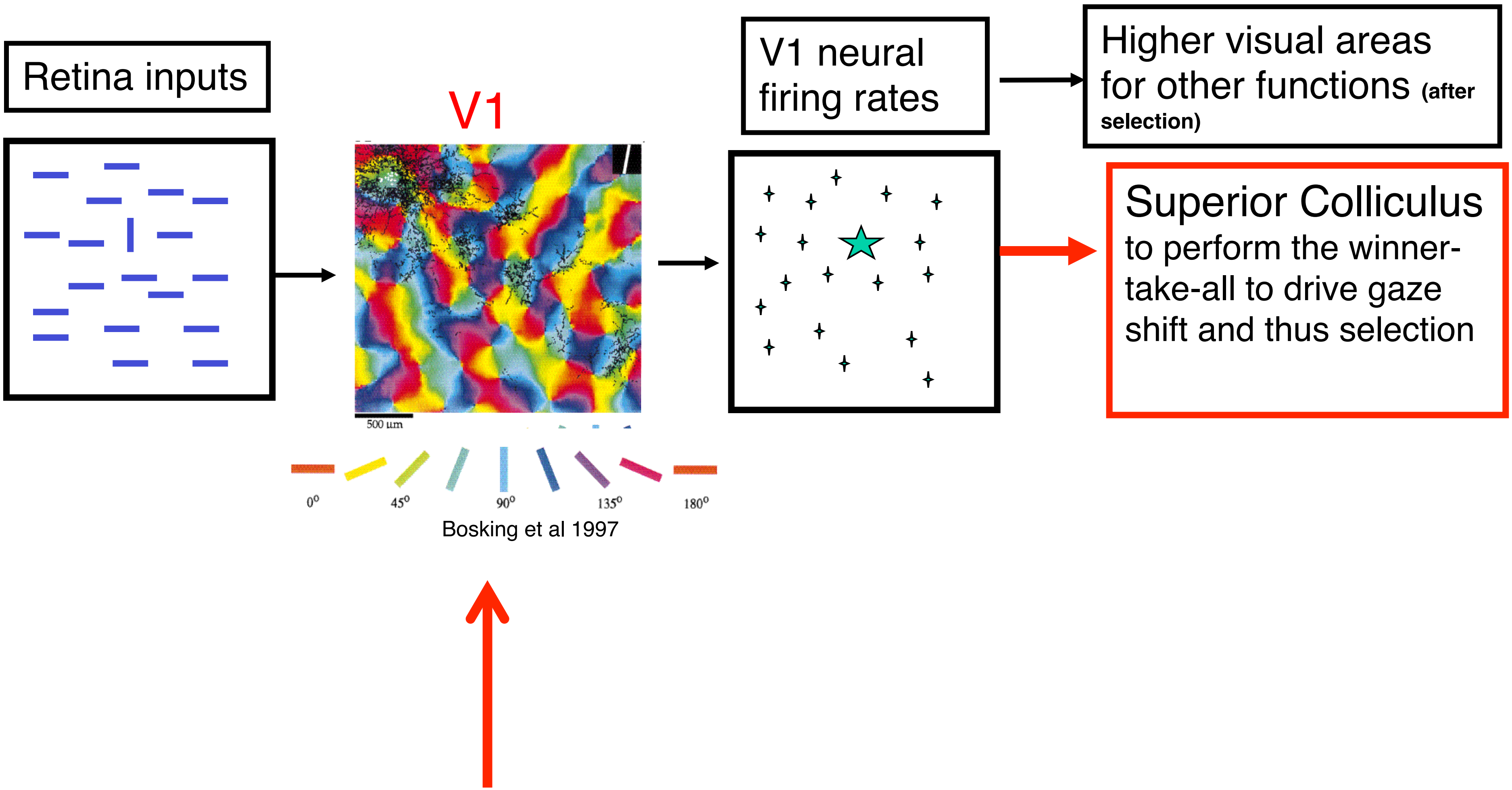
A bottom-up saliency map in the primary visual cortex (Li 1999, 2002)





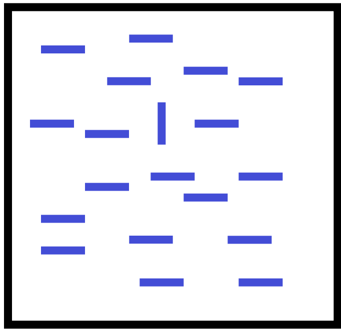
In monkeys, retina to superior colliculus connection is unable to drive visually guided saccades.

Intra-cortical interactions lead to saliency signals

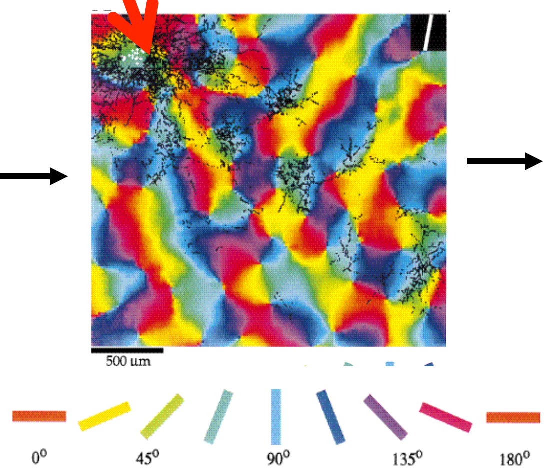


Intra-cortical interactions lead to saliency signals

Retina inputs

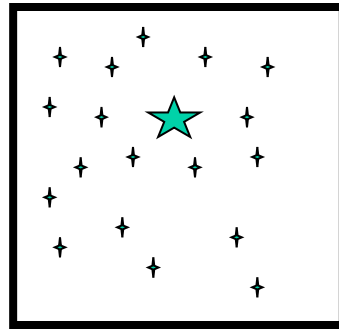


V1



Bosking et al 1997

V1 neural firing rates



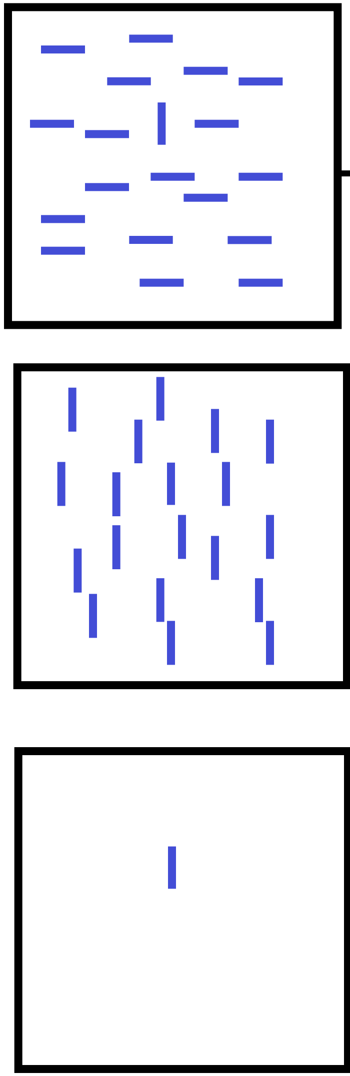
Higher visual areas for other functions **(after selection)**

Superior Colliculus
to perform the winner-take-all to drive gaze shift and thus selection

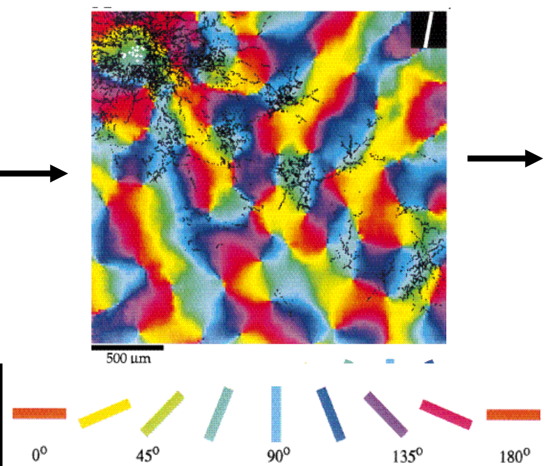
Intra-cortical interactions in V1 make nearby neurons (with not necessarily overlapping receptive fields) tuned to the similar features suppress each other --- iso-feature suppression (Gilbert & Wiesel 1983, Rockland & Lund 1983, Allman et al 1985, Hirsch & Gilbert 1991, Li & Li 1994, etc)

Intra-cortical interactions lead to saliency signals

Retina inputs



V1

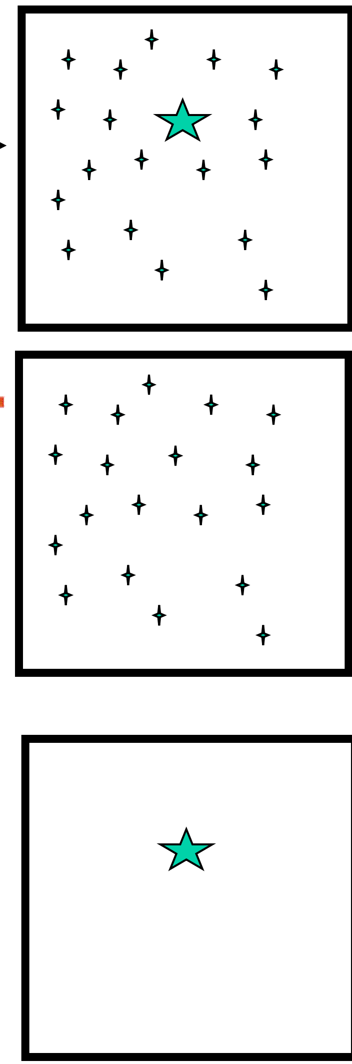


Bosking et al 1997

Intra-cortical interactions in V1 make nearby neurons (with not necessarily overlapping receptive fields) tuned to the similar features suppress each other ---

iso-feature suppression (Gilbert & Wiesel 1983, Rockland & Lund 1983, Allman et al 1985, Hirsch & Gilbert 1991, Li & Li 1994, etc)

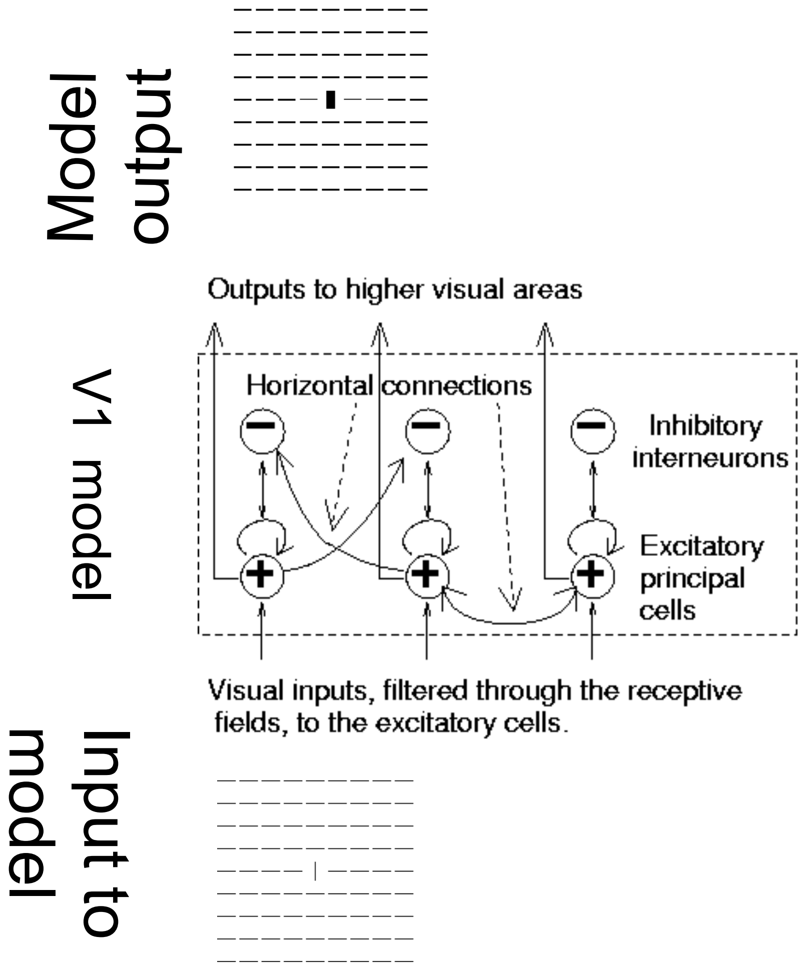
V1 neural firing rates



Higher visual areas for other functions (after selection)

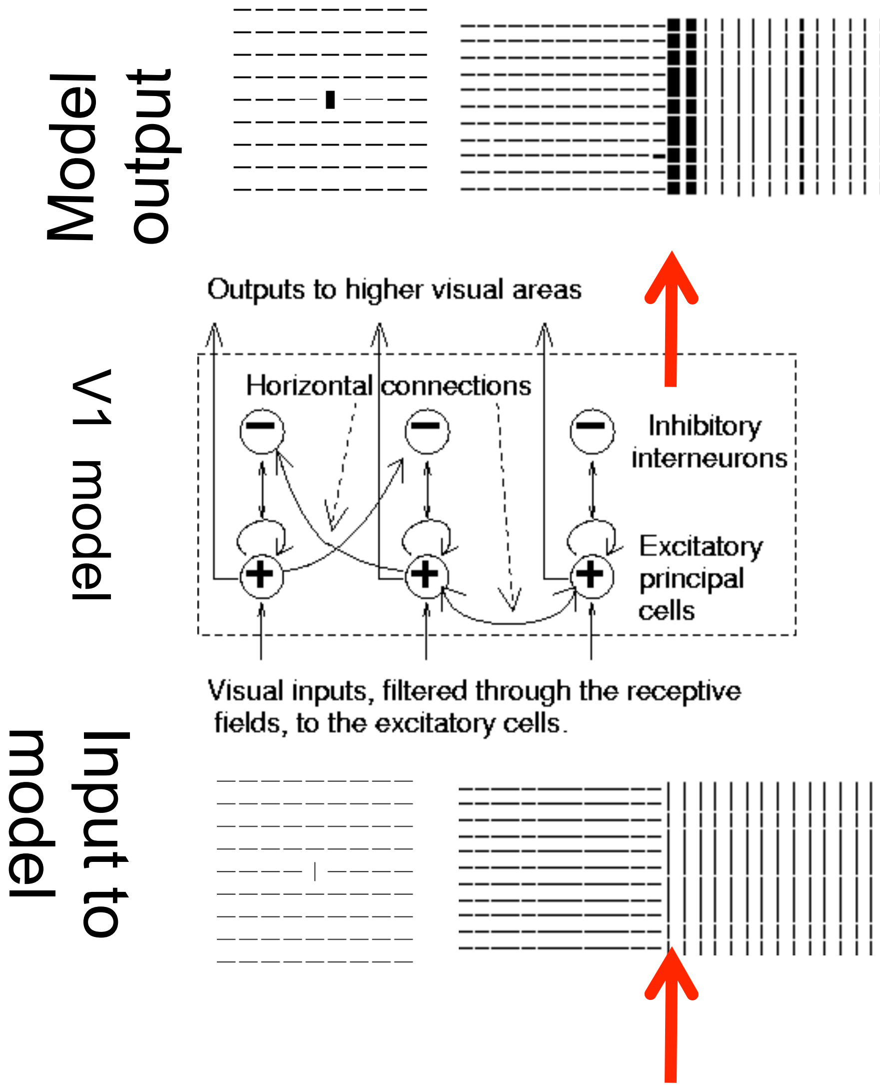
Superior Colliculus to perform the winner-take-all to drive gaze shift and thus selection

The V1 mechanism for saliency simulated by a V1 model



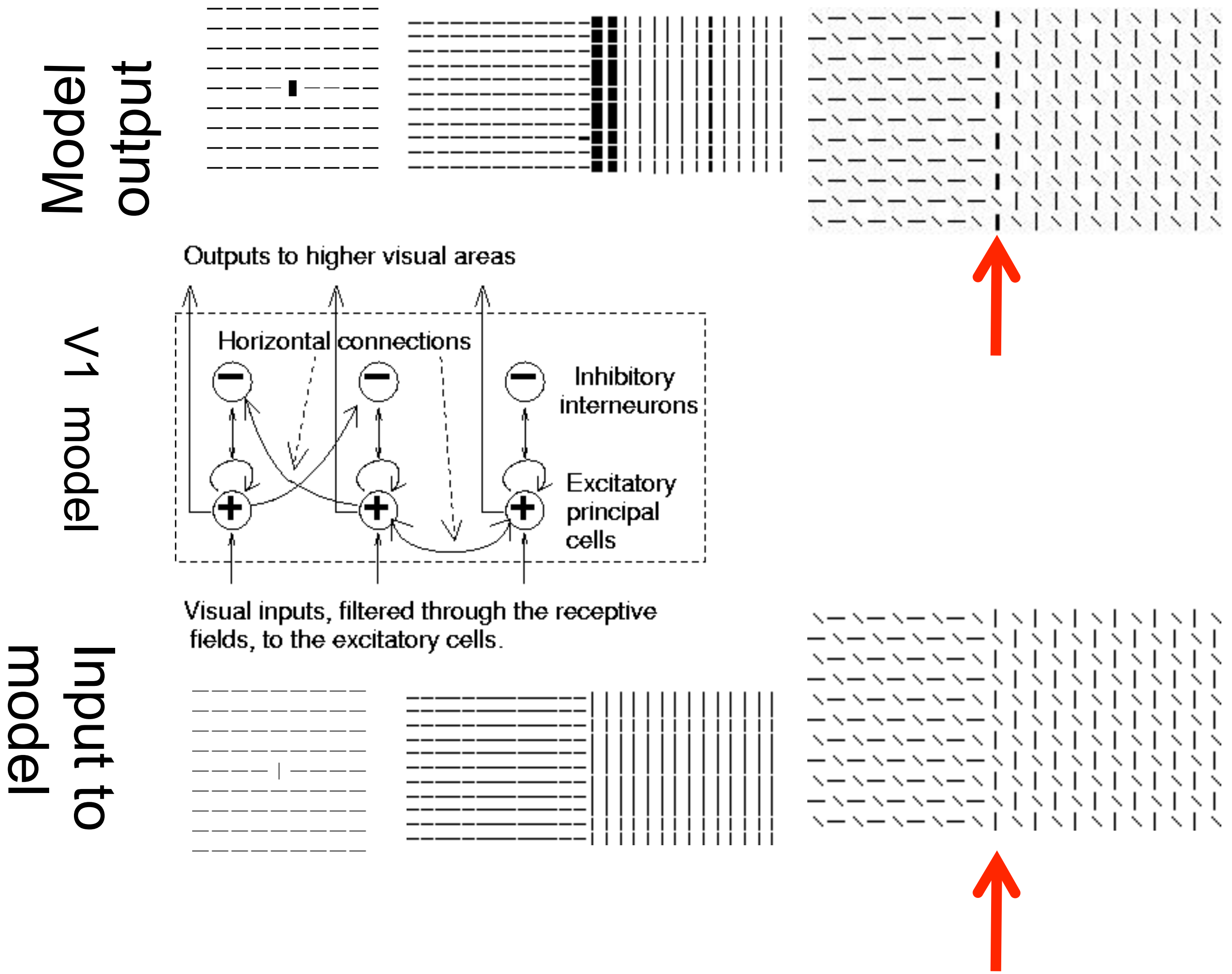
Li, 1998, 1999, 2000, 2002.

The V1 mechanism for saliency simulated by a V1 model



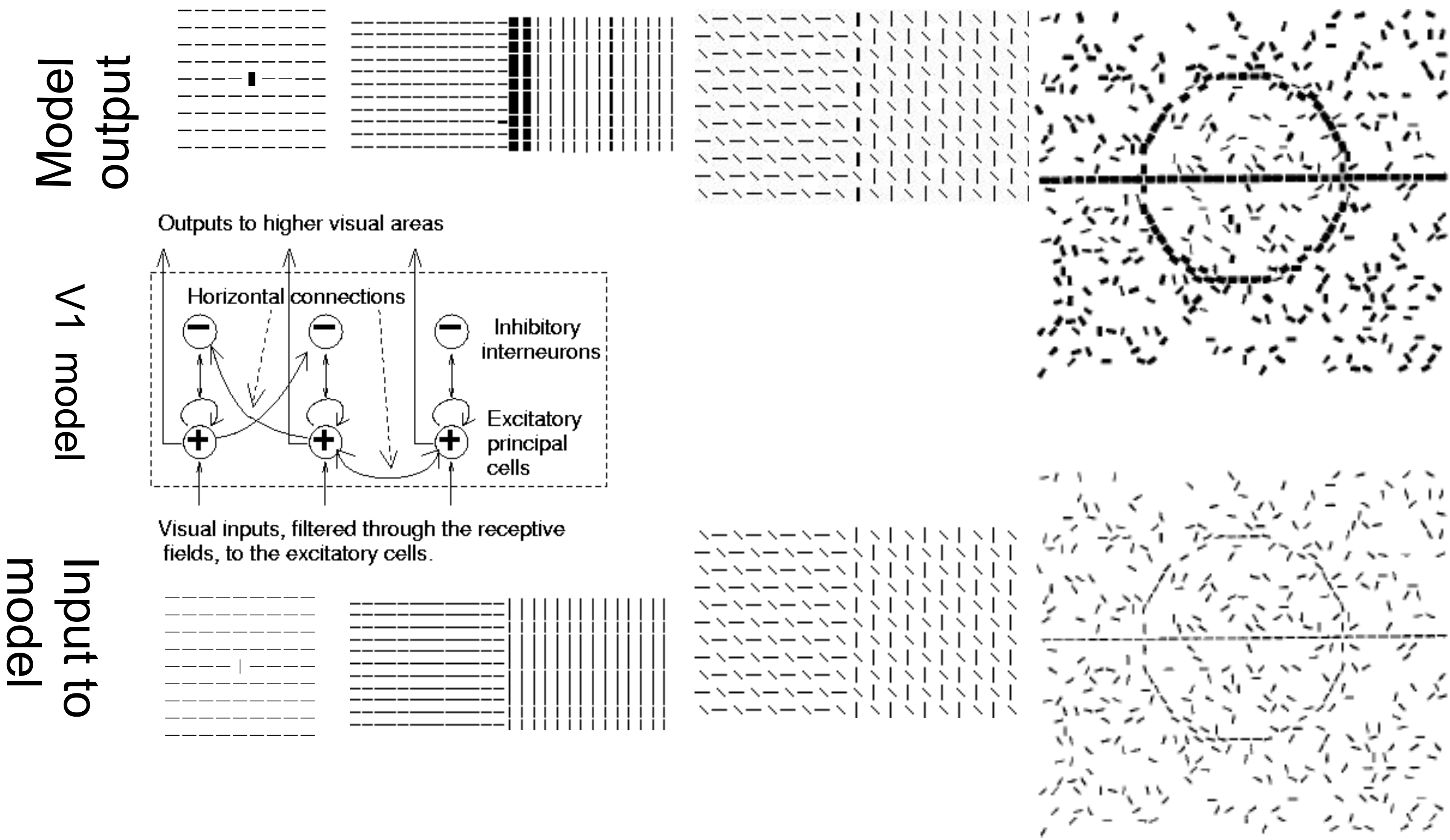
Li, 1998, 1999, 2000, 2002.

The V1 mechanism for saliency simulated by a V1 model



Li, 1998, 1999, 2000, 2002.

The V1 mechanism for saliency simulated by a V1 model



Li, 1998, 1999, 2000, 2002.

Talk outline

Motivating questions:

(1) What is V1 doing?

(2) Which brain areas control the direction of attention?

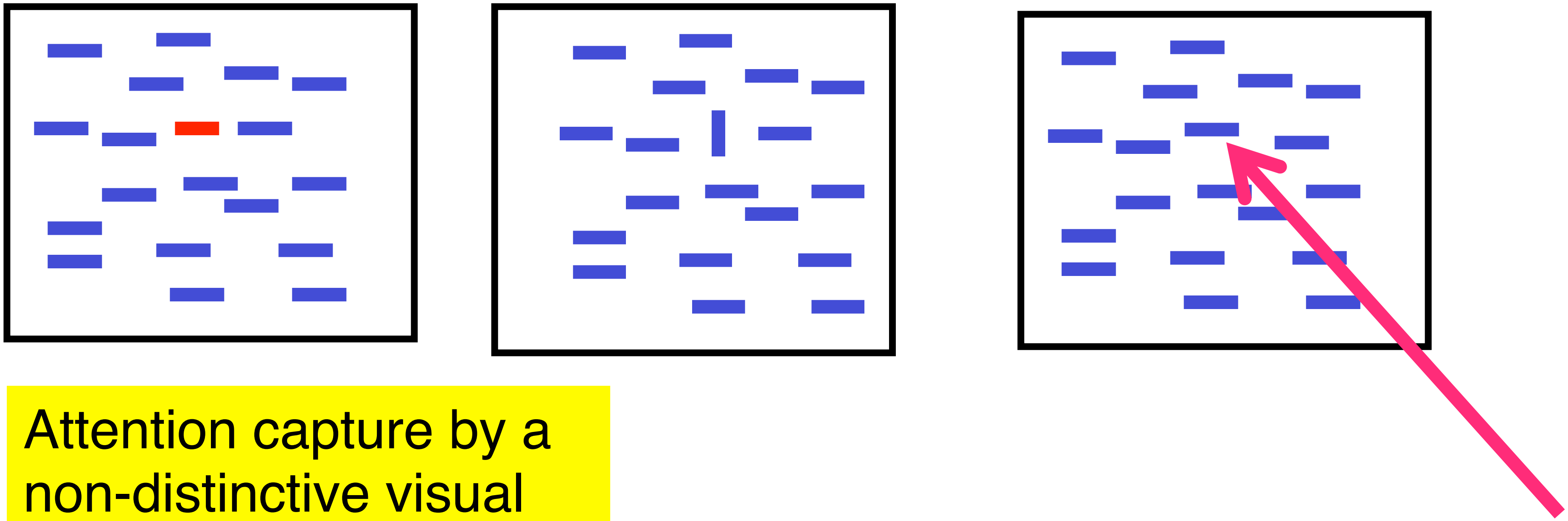
Theory:

A bottom-up saliency map in V1

Predictions and tests: ←

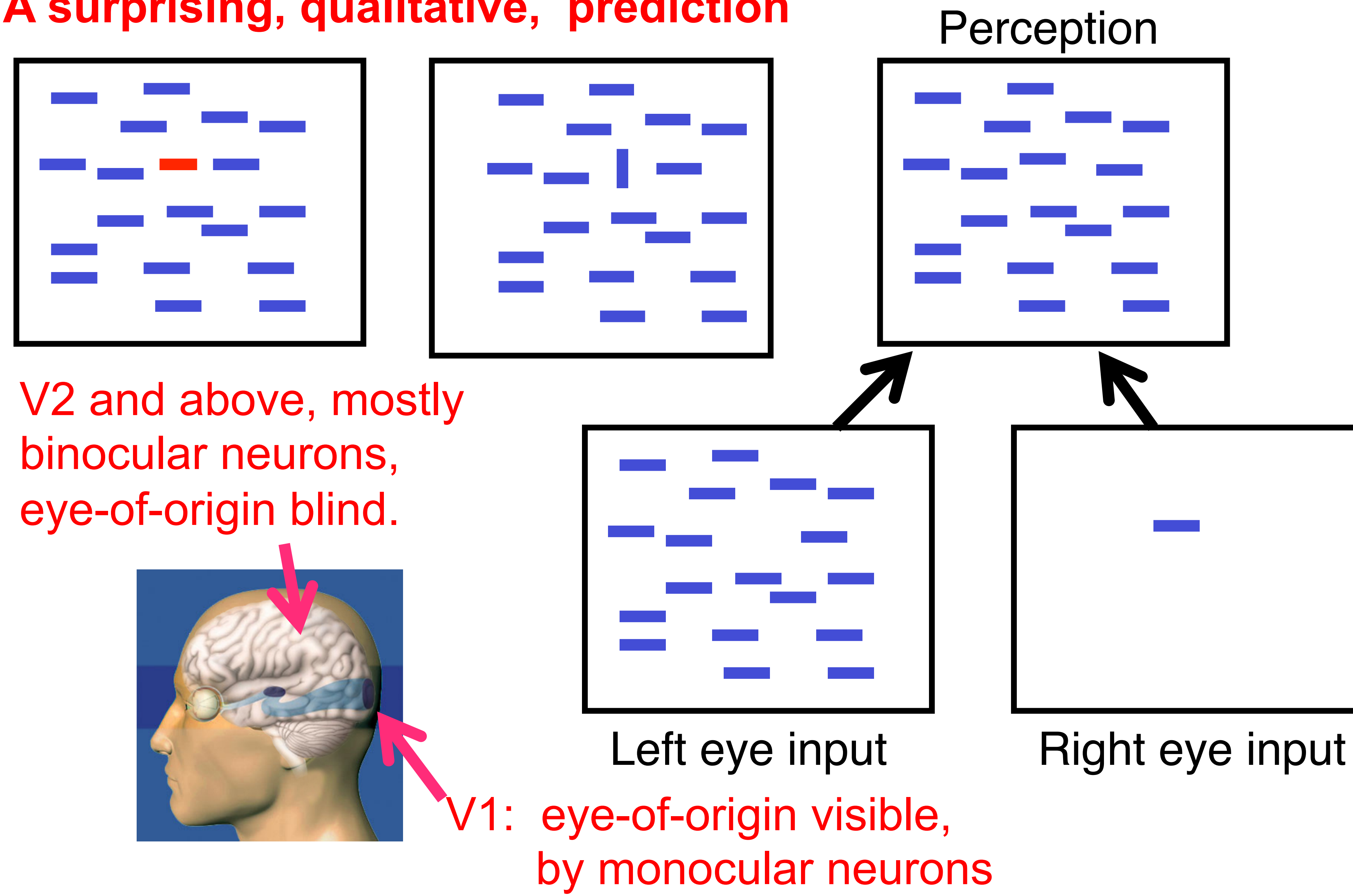
Implications.

A surprising, qualitative, prediction,

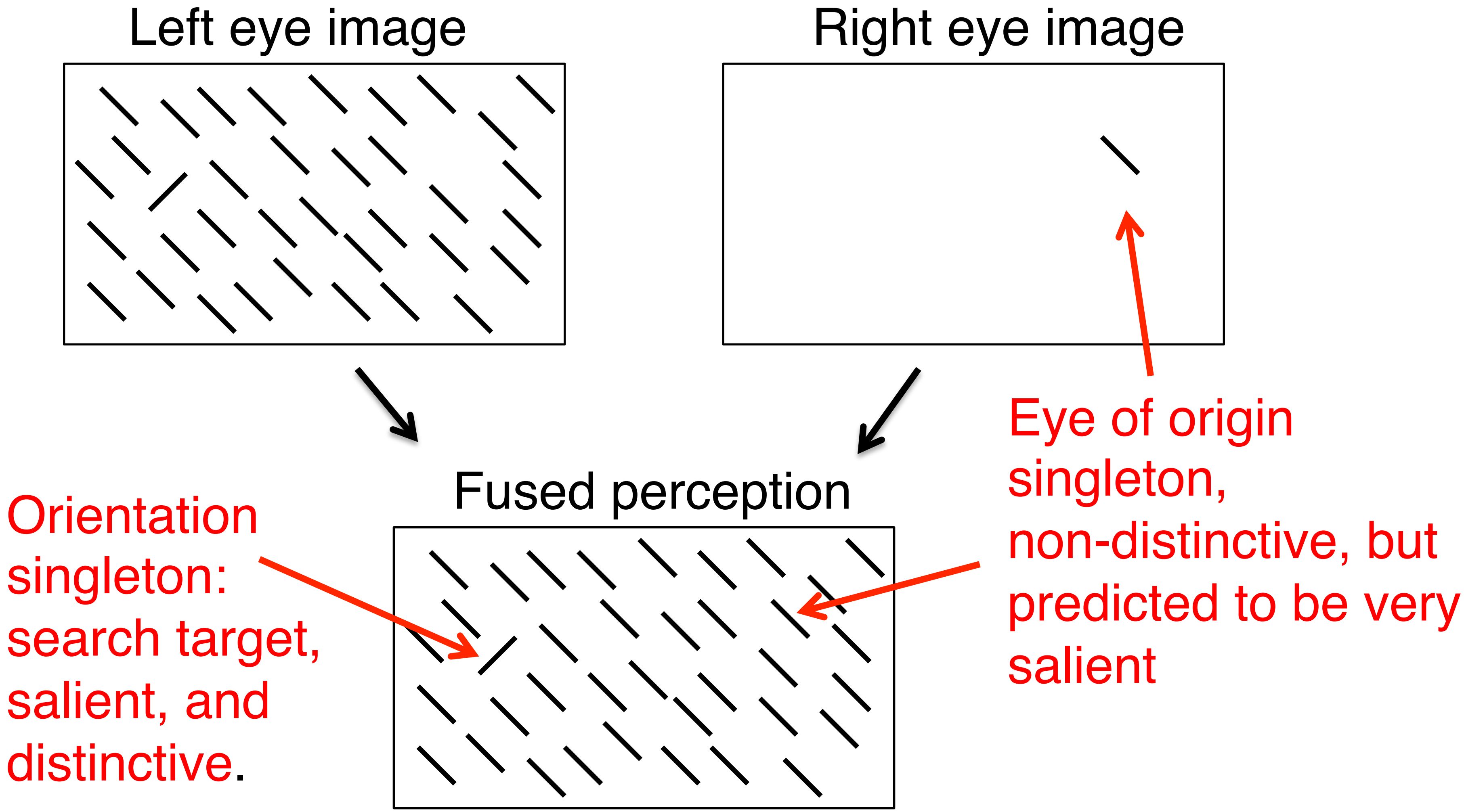


Attention capture by a non-distinctive visual item

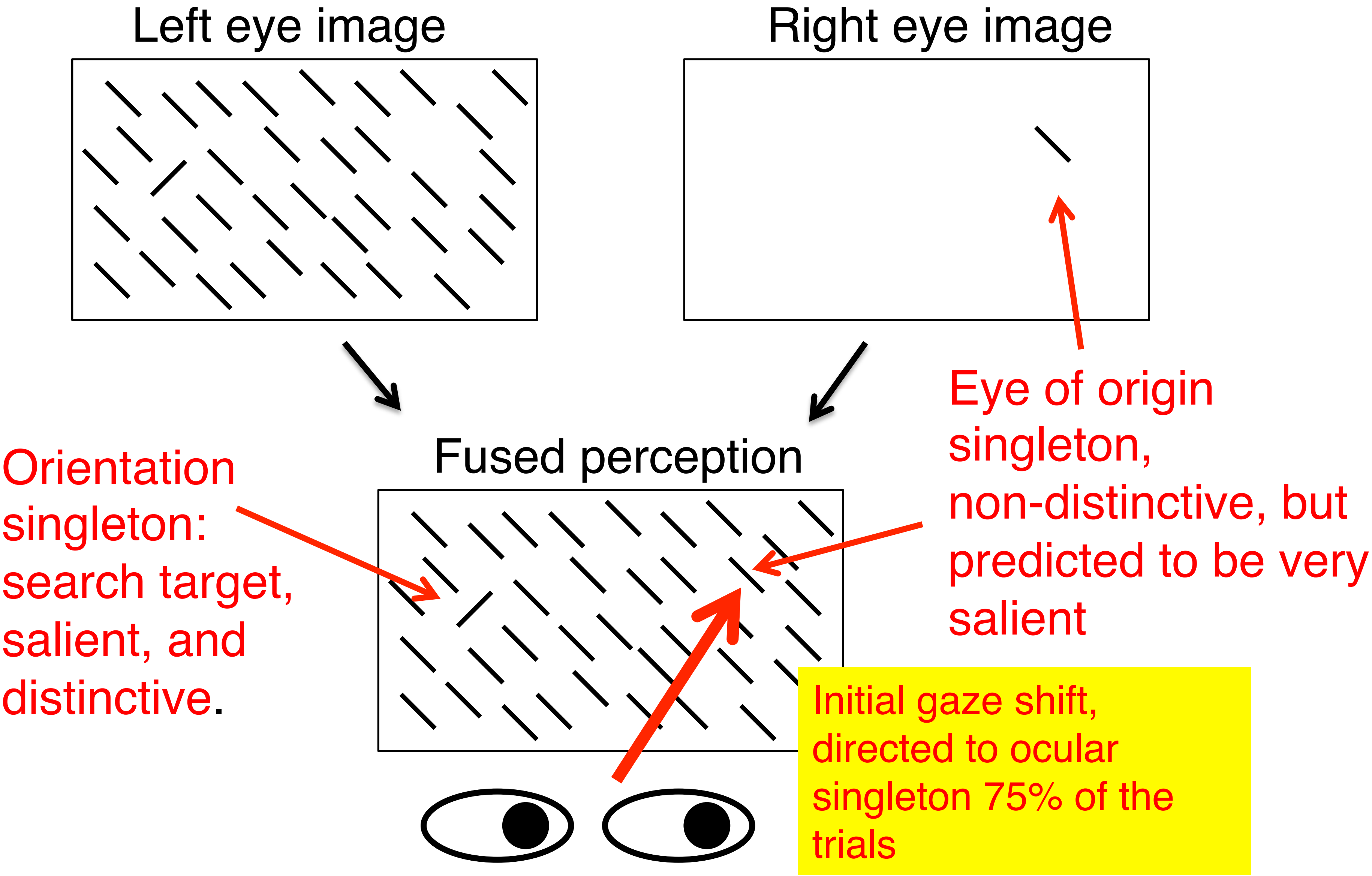
A surprising, qualitative, prediction



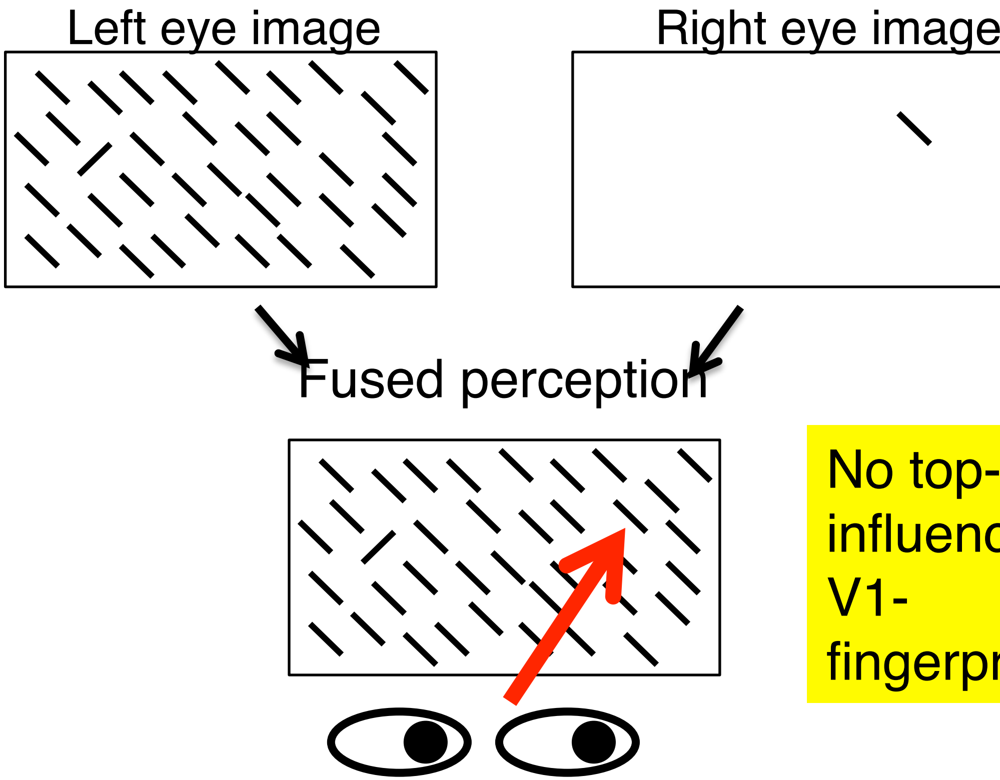
A qualitative prediction, confirmed (Zhaoping, 2008, 2012)

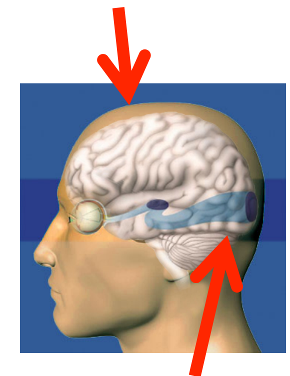


A qualitative prediction, confirmed (Zhaoping, 2008, 2012)



A qualitative prediction, confirmed (Zhaoping, 2008, 2012)





No top-down influence! V1-fingerprint!

Dissociation between perceptual distinction and saliency

Other qualitative predictions confirmed: Zhaoping & Snowden 2006, Zhaoping & May 2007, Koene & Zhaoping 2007, Jingling & Zhaoping 2008, Zhang et al 2012, etc.

Talk outline

Motivating questions:

(1) What is V1 doing?

(2) Which brain areas control the direction of attention?

Theory:

A bottom-up saliency map in V1

Predictions and tests:

(1) Qualitative: ocular singleton pop-out ⟵

Implications.

Talk outline

Motivating questions:

(1) What is V1 doing?
(2) Which brain areas control the direction of attention?

Theory:

A bottom-up saliency map in V1

Predictions and tests:

(1) Qualitative: ocular singleton pop-out
(2) Quantitative: reaction times for pop-out ←

Implications.

Examples of quantitative predictions from other theories/models

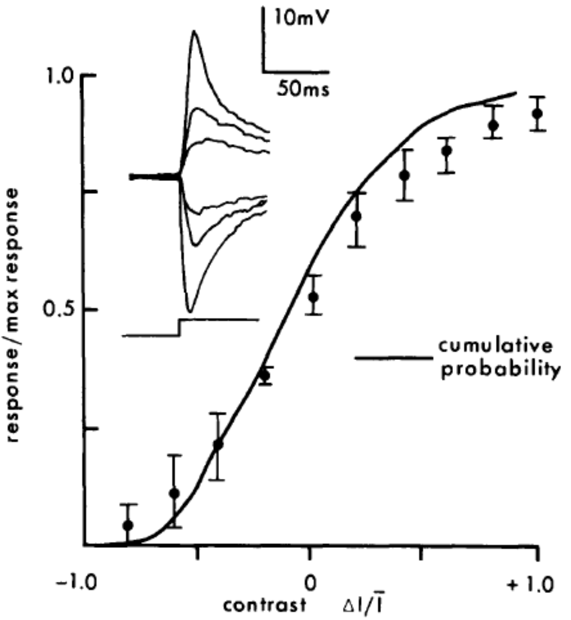


Fig. 2. The contrast-response function of light adapted
LMC's compared to the cumulative probability functi

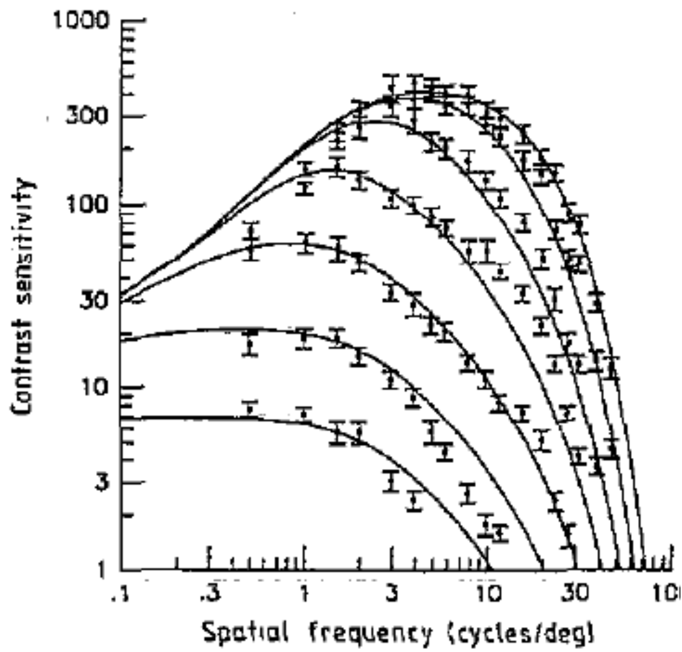
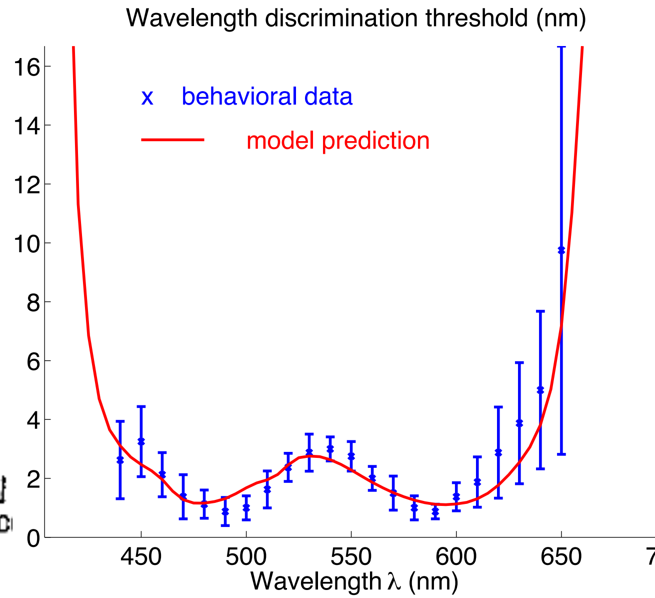


Figure 10. Predicted retinal filters, (5.14), at dif-



Laughlin, 1978
predicting neural contrast response function from input distribution

Zero parameters

Atick and Redlich's (1990) predicting human psychophysical contrast sensitivity from Barlow's efficient coding theory, with a few free parameters.

Zhaoping, Geisler, May 2011, predicting wavelength discrimination threshold from cone spectrum sensitivities, one free parameter

Examples of quantitative predictions from other theories/models

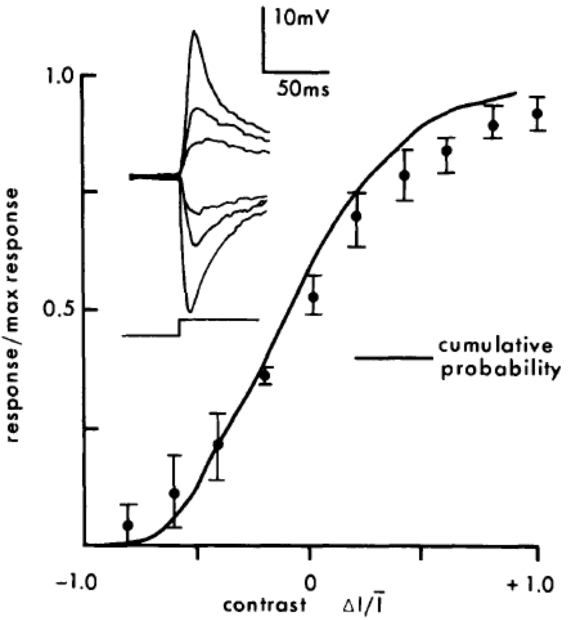


Fig. 2. The contrast-response function of light adapted LMC's compared to the

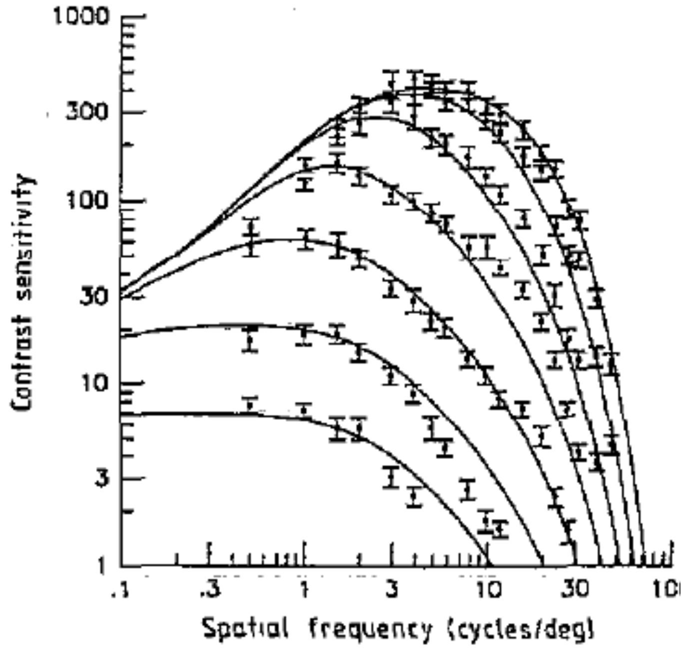
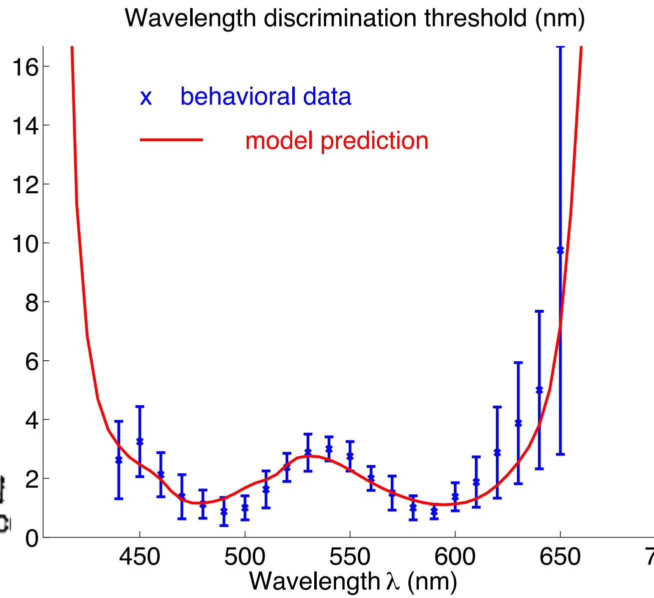


Figure 10. Predicted retinal filters (5.14) at dif



2011,

Laughli

predicting

contrast r

function fr

distribution

from

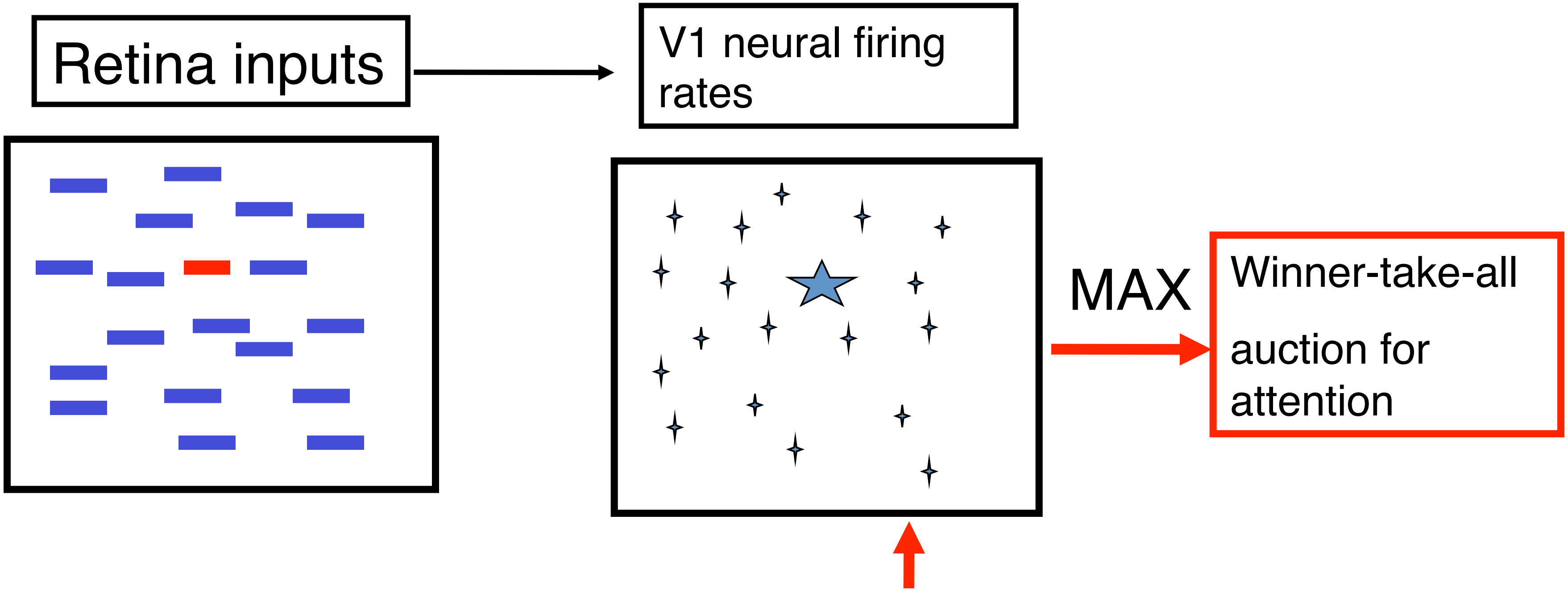
ies,

Zero parameters

Let us try to predict from the V1 Saliency Hypothesis, with zero parameters

(work done with Li Zhe)

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

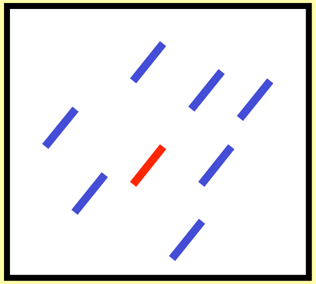


Neural activities as universal currency to bid for visual selection.
The receptive field of the most active V1 cells is selected

First, illustrate using a toy V1:

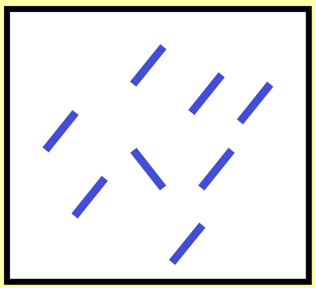
Toy V1: some cells tuned to orientation, others tuned to color

Colour (C) singleton

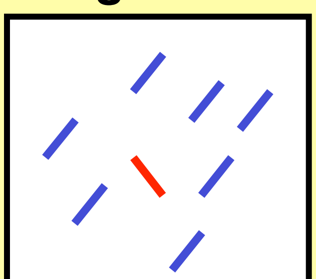


RT_C = 500 ms , Color (C) cell response 10 spikes/second

Orientation (O) singleton



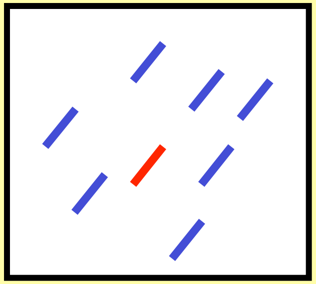
Double (CO) singleton



First, illustrate using a toy V1:

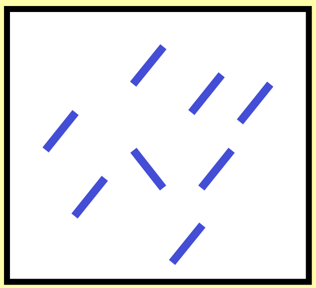
Toy V1: some cells tuned to orientation, others tuned to color

Colour (C) singleton



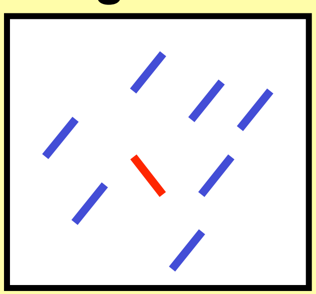
RT_C = 500 ms , Color (C) cell response 10 spikes/second

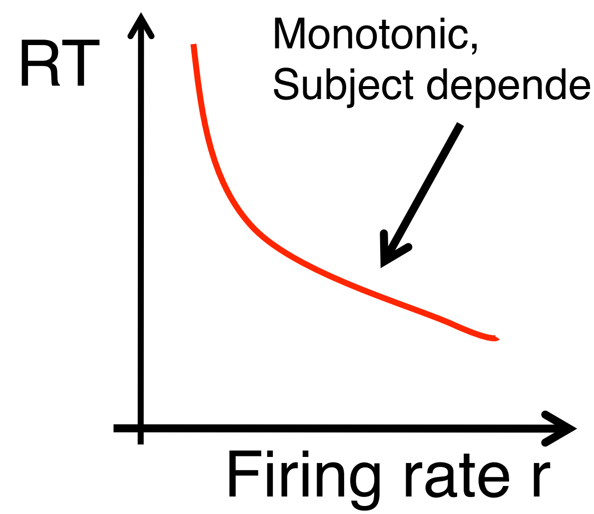
Orientation (O) singleton



RT_O = 600 ms, Orientation (O) cell response 9 spikes/second

Double (CO) singleton

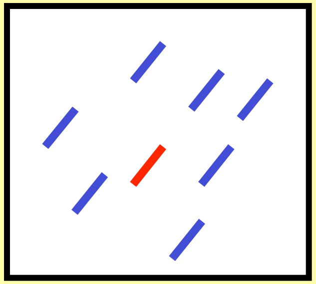




First, illustrate using a toy V1:

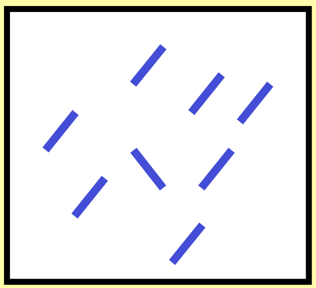
Toy V1: some cells tuned to orientation, others tuned to color

Colour (C) singleton



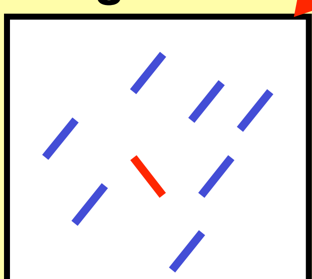
RT_C = 500 ms , Color (C) cell response 10 spikes/second

Orientation (O) singleton

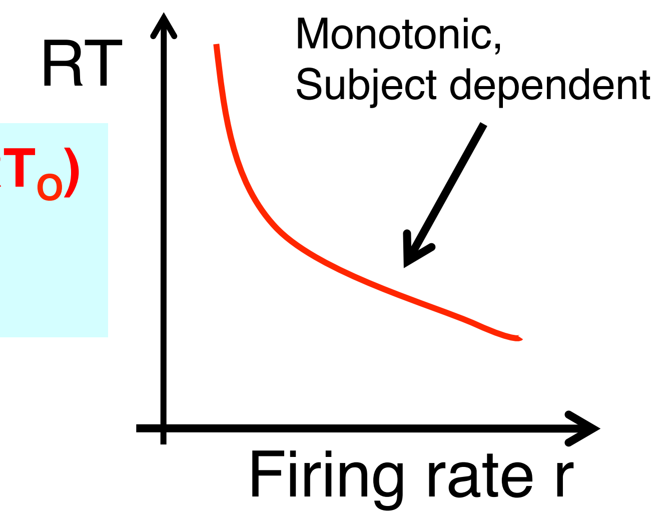


RT_O = 600 ms, Orientation (O) cell response 9 spikes/second

Double (CO) singleton

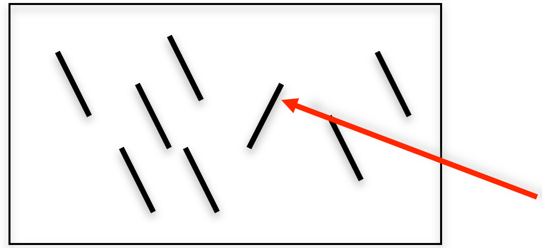


RT_{CO}= 500 ms = min (RT_C, RT_O)
C neuron: 10 spikes/second
O neuron: 9 spikes/second



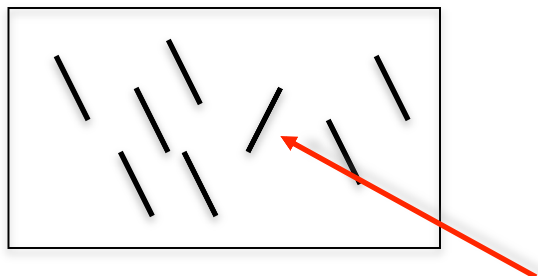
In fact, V1 responses are stochastic, so RT data is probabilistic

RT = 500 ms



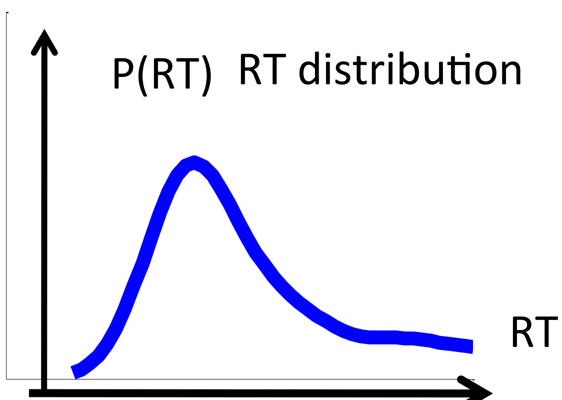
V1 responses to
the target bar
=10 spikes

RT = 400 ms



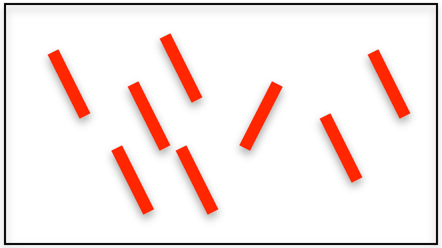
V1 responses to
the target bar
=12 spikes

etc



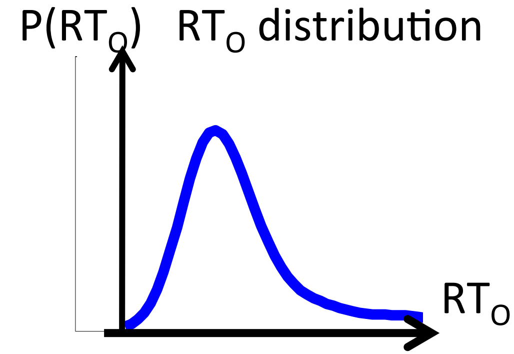
Therefore, we can predict a probability distribution $P(RT_{CO})$

O singleton

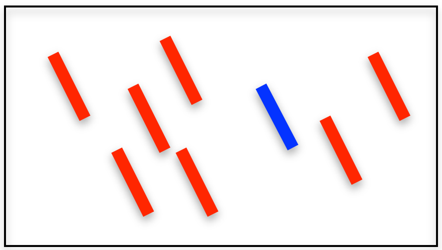


The most active target neuron is tuned to O

Relative target responses: 10,12,9 …

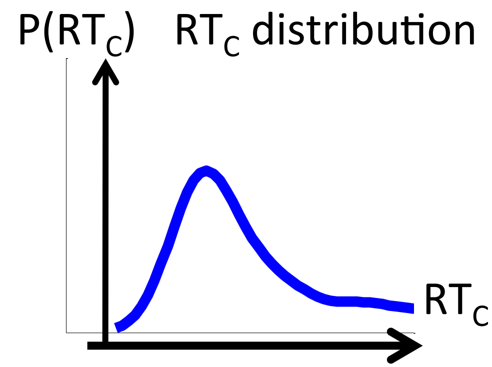


C singleton

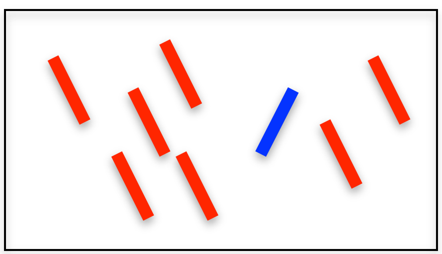


The most active target neuron is tuned to C

Relative target responses: 9,10,11 …



CO singleton



Two types of neurons highly active for the target: O, C

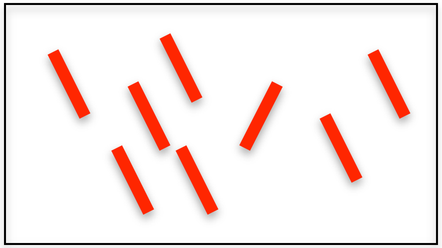
Relative O cell responses: 10, 12, 9 …

Relative C cell responses: 9, 10, 11 …

Winner responses: 10, 12, 11 …

Therefore, we can predict a probability distribution $P(RT_{CO})$

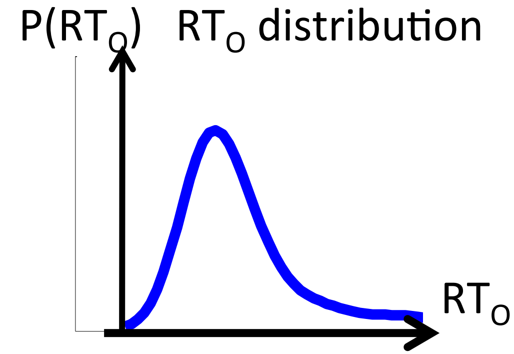
O singleton



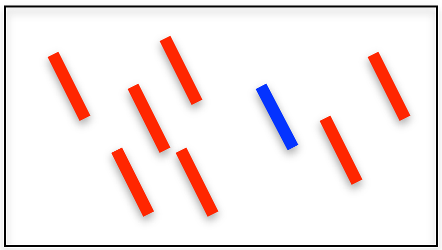
The most active target neuron is tuned to O

Relative target responses: 10,12,9 …

RT_O distribution



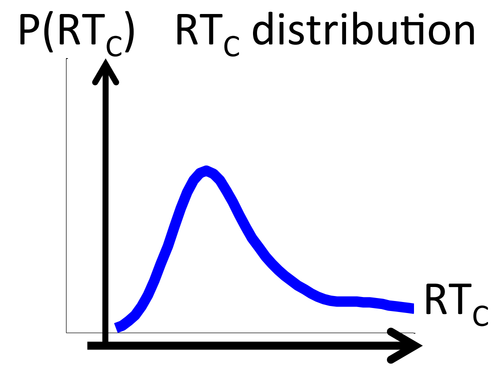
C singleton



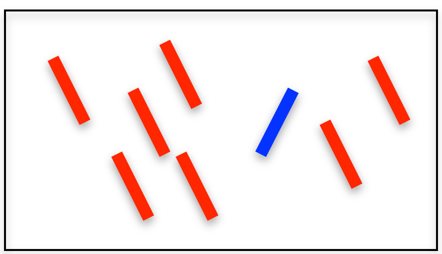
The most active target neuron is tuned to C

Relative target responses: 9,10,11 …

RT_C distribution



CO singleton



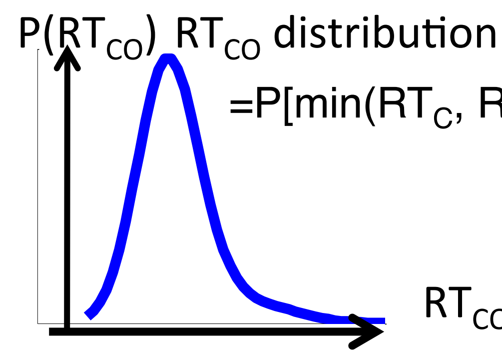
Two types of neurons highly active for the target: O, C

Relative O cell responses: 10, 12, 9 …

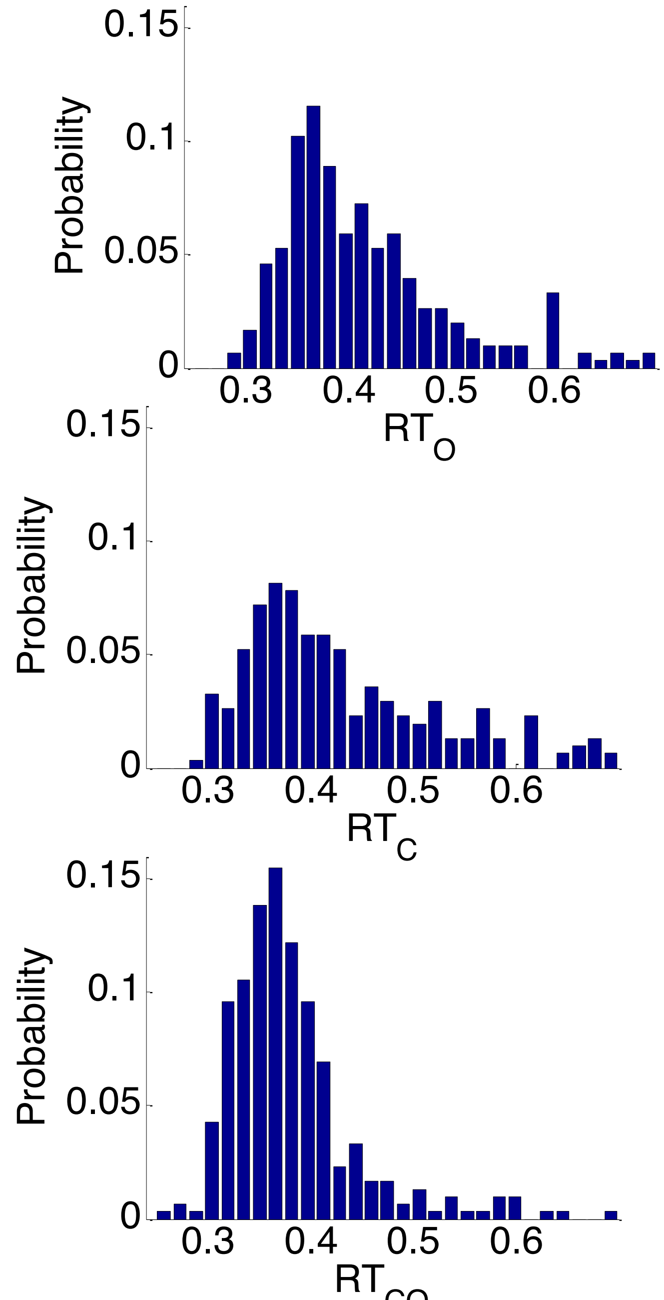
Relative C cell responses: 9, 10, 11 …

Winner responses: 10, 12, 11 …

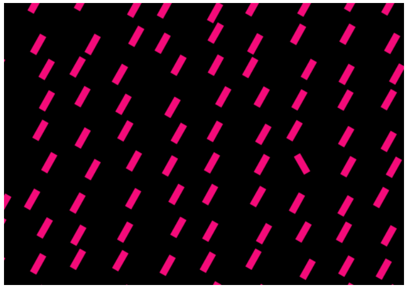
RT_{CO} distribution $=P[\min(RT_C, RT_O)]$



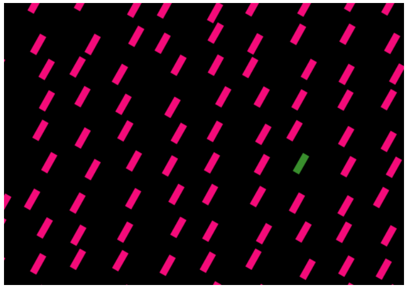
Behavioral data from Koene and Zhaoping (2007)



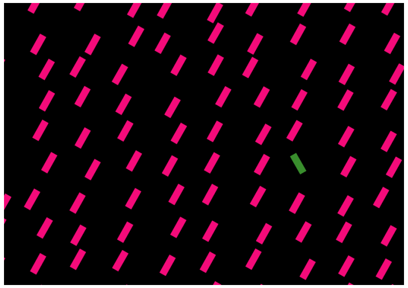
O target



C target

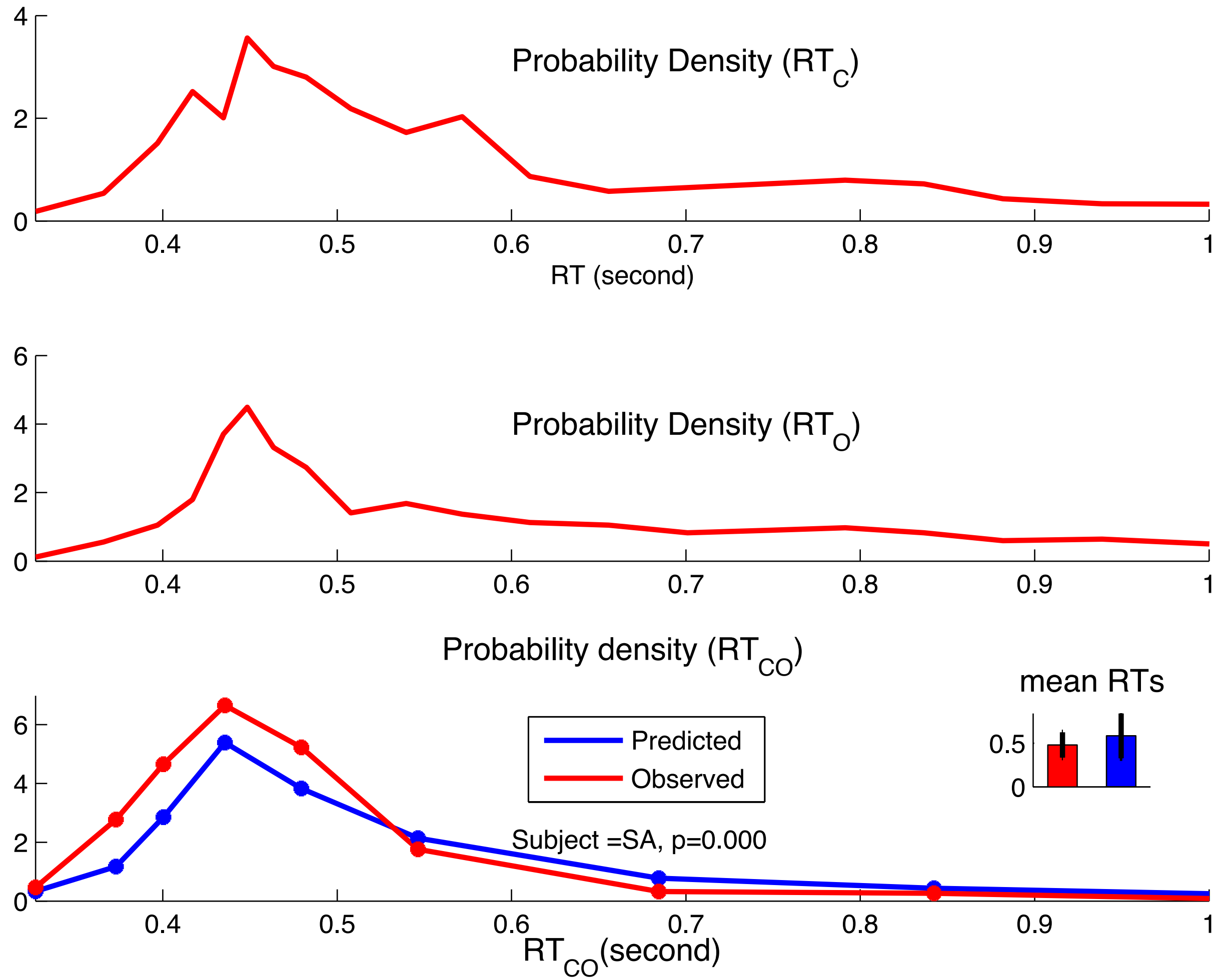


CO target



About 300 trials per condition

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

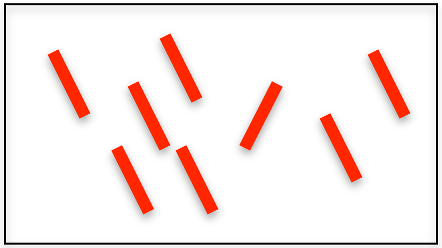


P value ~ 0.00

Predicted RT significantly longer than observed RT

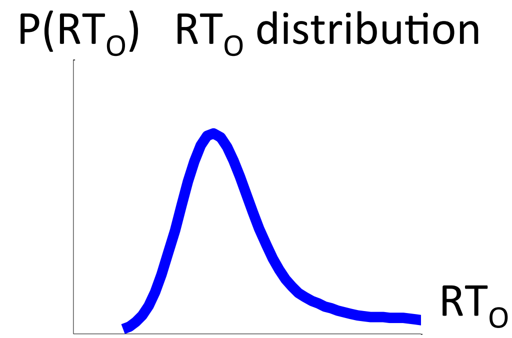
Because --- real V1 has CO conjunction cells

O target

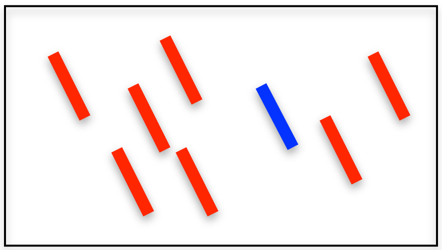


The most active target neuron is tuned to O

Relative target responses: 10,12,9 …

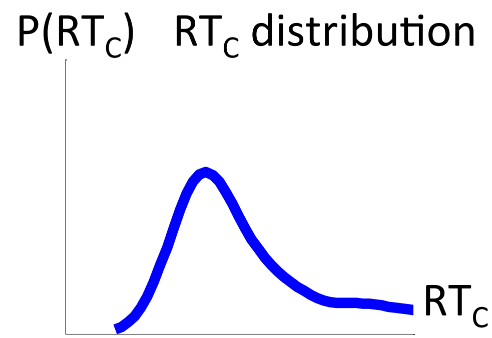


C target

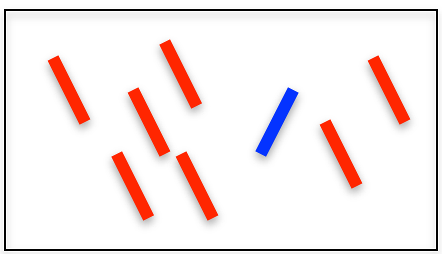


The most active target neuron is tuned to C

Relative target responses: 9,10,11 …



CO target



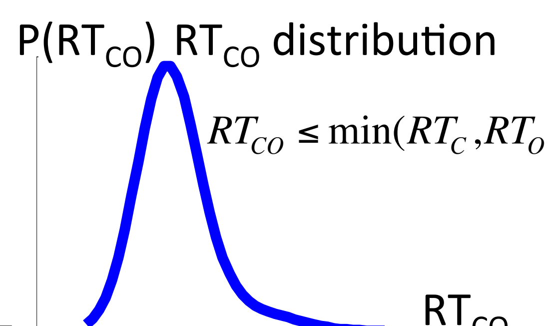
Three types of neurons highly active for the target: O, C, CO

Relative O cell responses: 10, 12, 9 …

Relative C cell responses: 9, 10, 11 …

Relative CO cell responses: 11, 10, 10 …

Winner responses: 11, 12, 11 …



Hence, RT_{CO} can not be predicted from RT_C and RT_O

If V1 did not have CO cells, $RT_{CO} = \min(RT_C, RT_O)$

However, V1 does have CO cells,

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

By analogy

V1 has C,O, M, CO, MO, and few CM cells (Hubel and Wiesel 1959, Livingstone and Hubel 1984, Horwitz and Albright 2005)

We assume that V1 has no CMO cells, then

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

Hence, Prob(RT_{CMO}) predictable from probability distributions of RT_C, RT_M, RT_O, RT_{CM}, RT_{CO}, and RT_{MO}.

If this prediction is confirmed --- extra-striate cortex excluded

V2 has CMO cells (Shipp, private communication 2011) would give

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

V1 has C,O, M, CO, MO, and few CM cells (Hubel and Wiesel 1959, Livingstone and Hubel 1984, Horwitz and Albright 2005)

We assume that V1 has no CMO cells, then

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

Hence, Prob(RT_{CMO}) predictable from probability distributions of RT_C, RT_M, RT_O, RT_{CM}, RT_{CO}, and RT_{MO}.

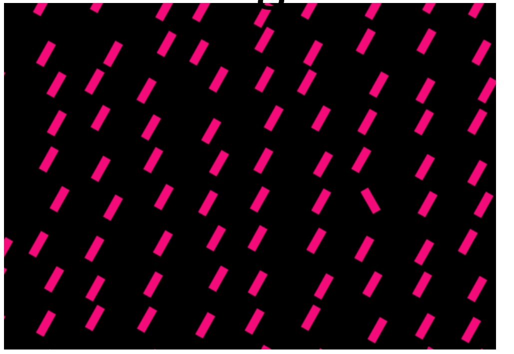
Behavioral data from Koene and Zhaoping (2007)

Target is different from distractors in *orientation* (**O**), *color* (**C**), *motion direction* (**M**), or combinations them.

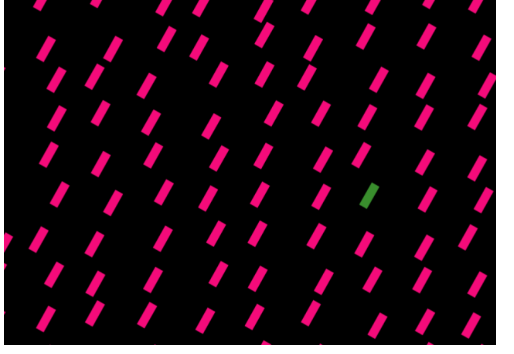
7 kinds of targets in total:
C, O, M,
CO, MO, CM,
CMO

Each about 300 trials / subject/ condition,
6 subjects in total

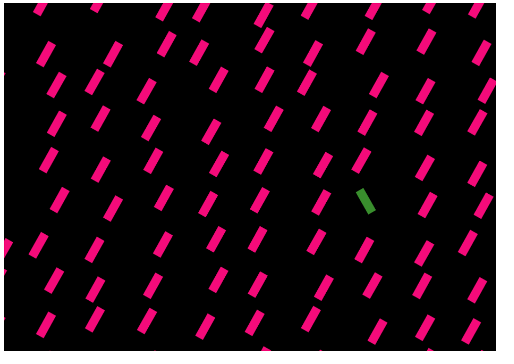
O target



C target

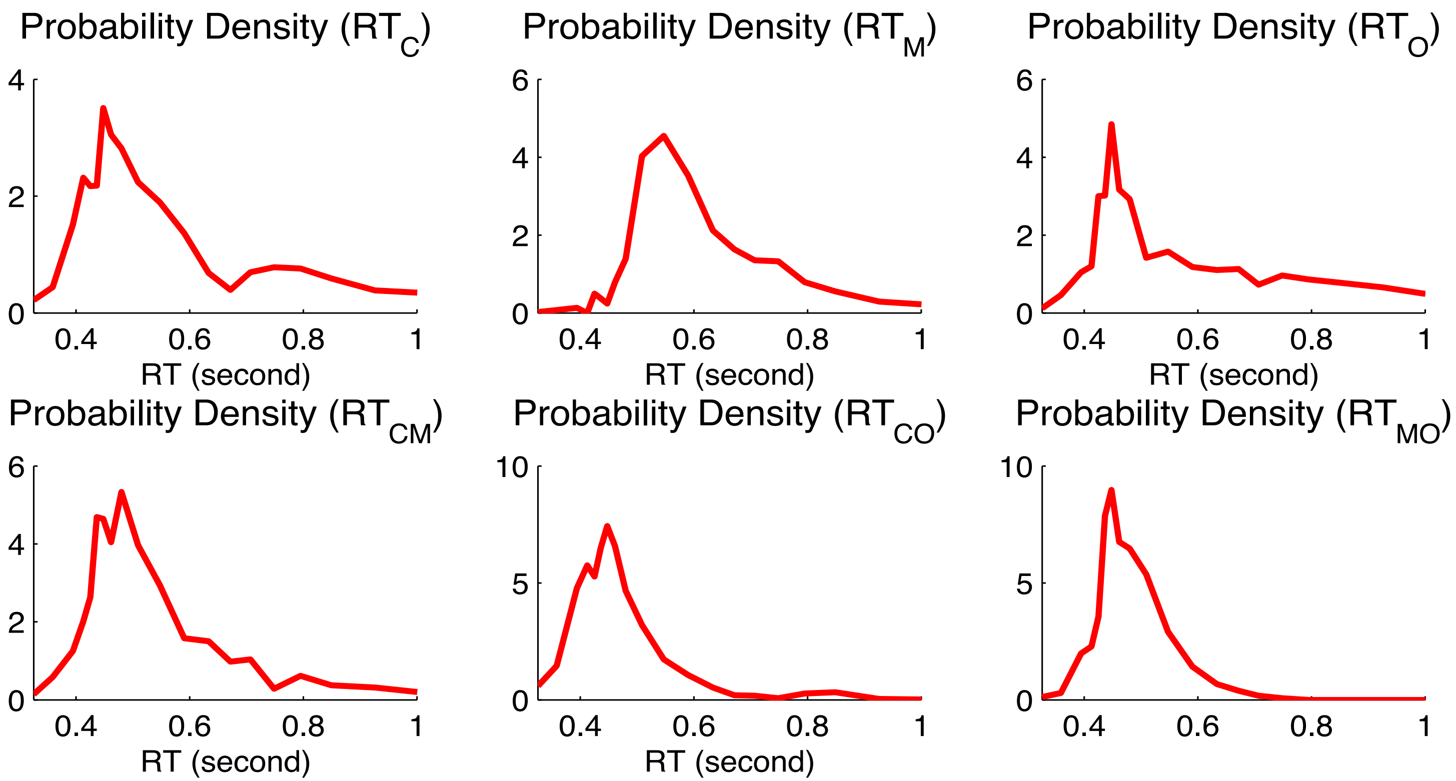


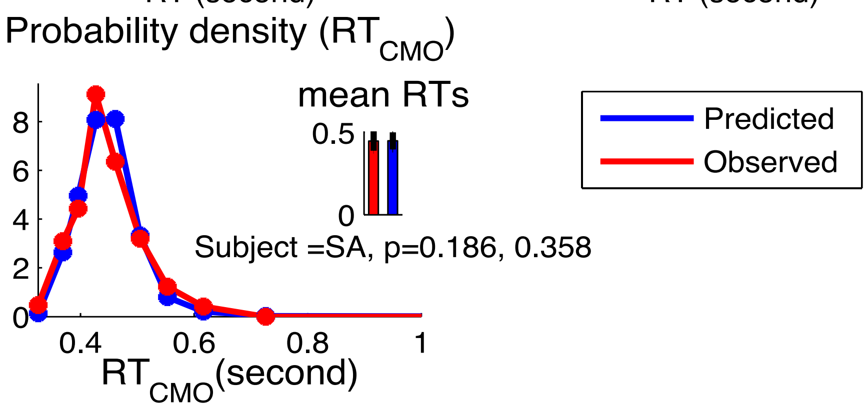
CO target



Distributions of RTs for a particular subject:

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

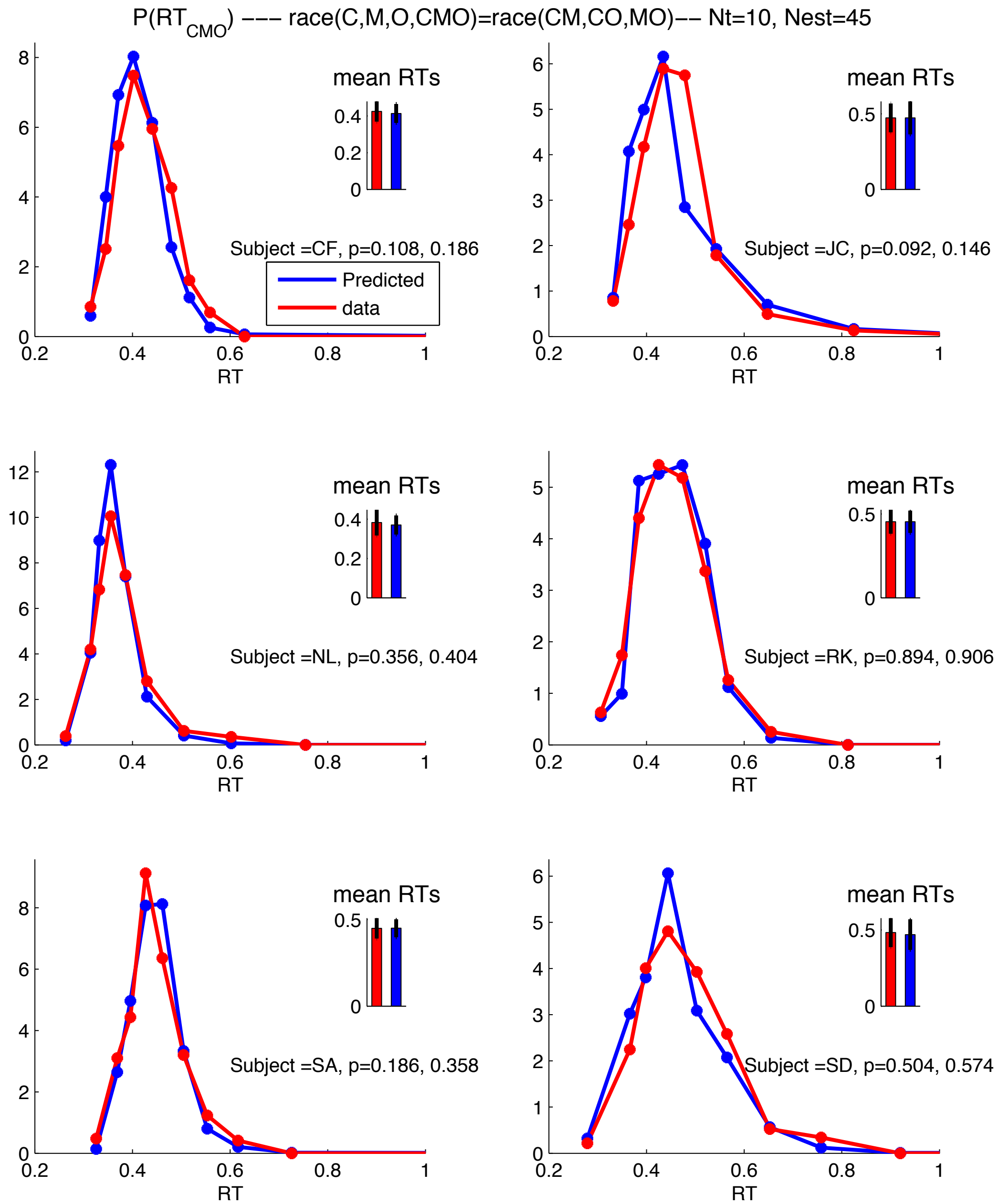




P values > 0.18,
No significant difference

For all six observers

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

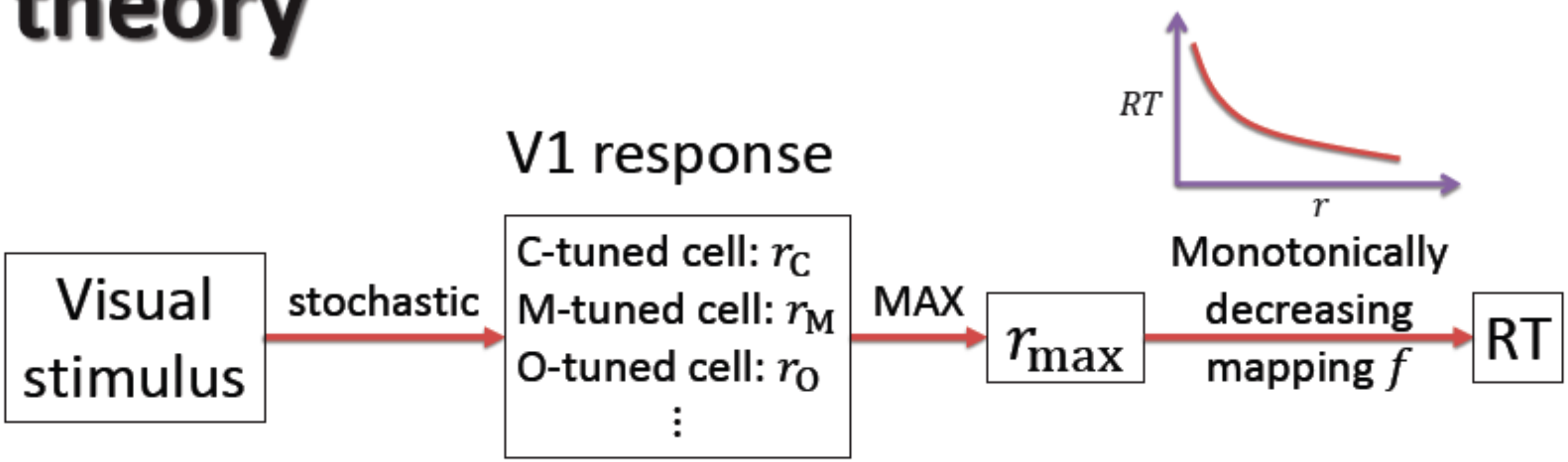


Note: the requirements for our prediction are:

(1)V1 theory: the highest firing neuron signals saliency of the most salient item
(2)A monotonic relationship between saliency and RT.
(3)Physiological knowledge that V1 has no CMO cells.

Hence, no parameters are required for our prediction!

V1 theory



$$RT = f(r_{max}) = f(\max\{r_C, r_M, \ldots\})$$
$$= \min\{f(r_C), f(r_M), \ldots\}$$

Talk outline

Motivating questions:

(1) What is V1 doing?
(2) Which brain areas control the direction of attention?

Theory:

A bottom-up saliency map in V1

Predictions and tests:

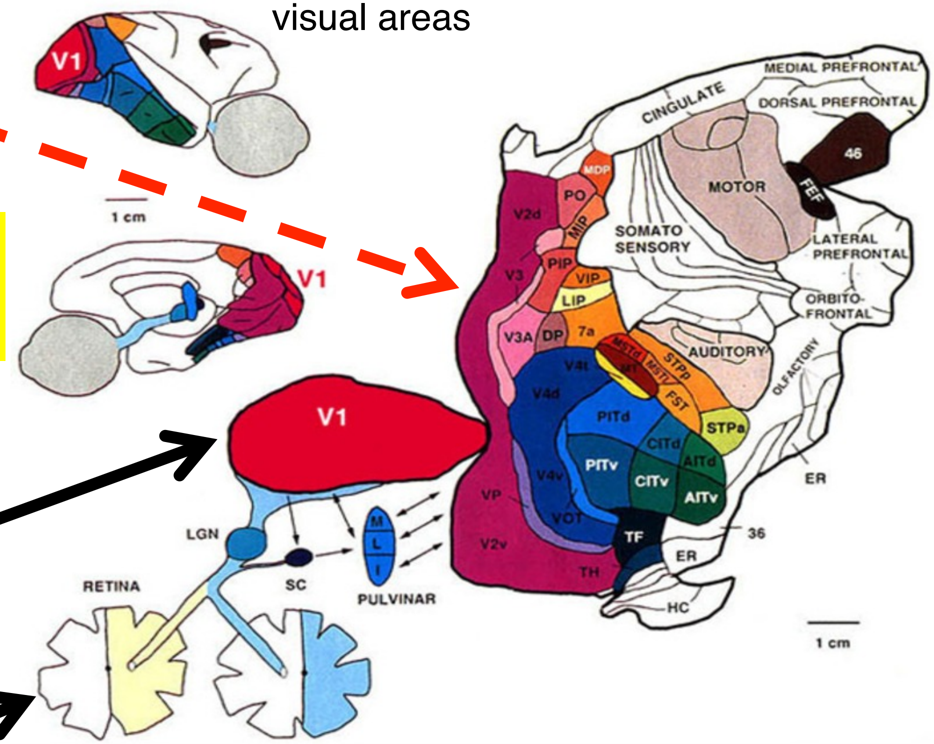
(1) Qualitative: ocular singleton pop-out
(2) Quantitative: reaction times for pop-out

Implications. ←

1980-200? --- V2, V3, V4 …
No more data as illuminating

Organization of visual areas

Calling for new ideas!!!
--- **What is V1 doing?**



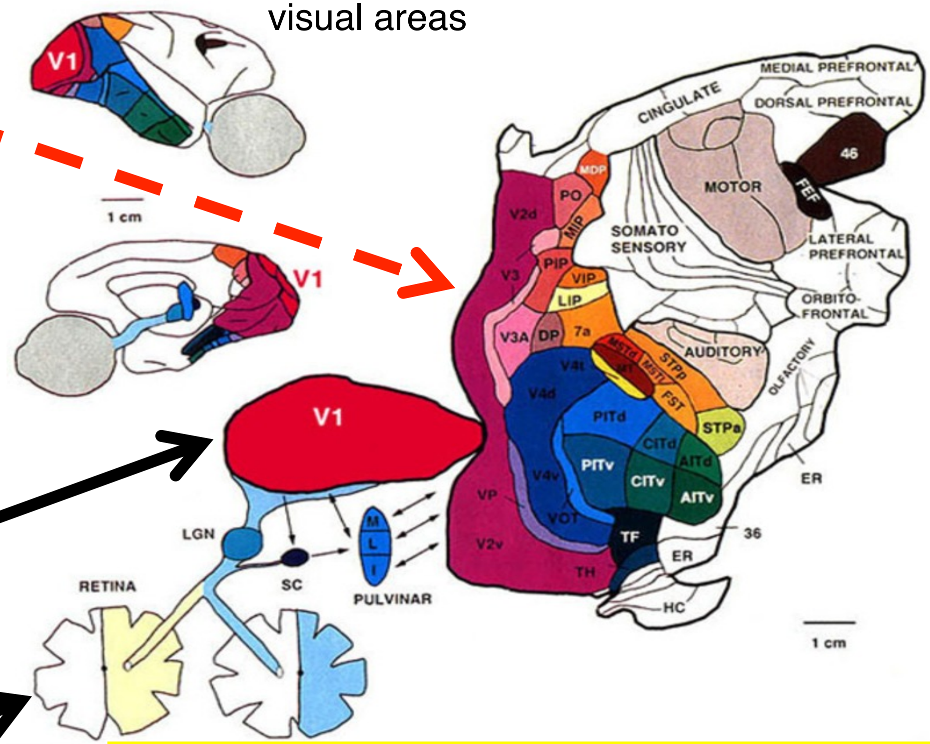
1960s-70s, Hubel and Wiesel
Bar/edge feature detectors in V1

1953, Stephen Kuffler
Center-surround feature detectors in retina

1980-200? --- V2, V3, V4 …
No more data as illuminating

V1 for bottom-up selection

Organization of visual areas



1960s-70s, Hubel and Wiesel
Bar/edge feature detectors in V1

1953, Stephen Kuffler
Center-surround feature detectors in retina

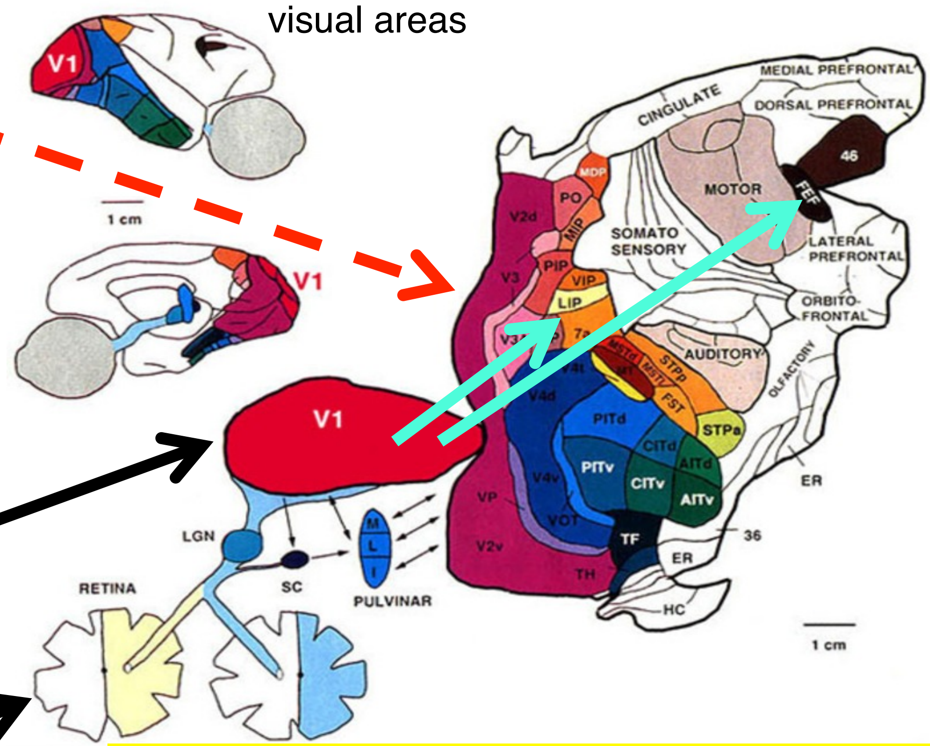
Probe V2 and higher areas for:

(1) top-down selection
(2) post-selectional decoding

1980-200? --- V2, V3, V4 …
No more data as illuminating

V1 for bottom-up selection

Organization of visual areas



1960s-70s, Hubel and Wiesel
Bar/edge feature detectors in V1

1953, Stephen Kuffler
Center-surround feature detectors in retina

Probe V2 and higher areas for:

(1) top-down selection
(2) post-selectional decoding

Saliency signals
in LIP, FEF
inherited from V1

Talk outline

Motivating questions:

(1) What is V1 doing?
(2) Which brain areas control the direction of attention?

Theory:

A bottom-up saliency map in V1

Predictions and tests:

(1) Qualitative: ocular singleton pop-out
(2) Quantitative: reaction times for pop-out

Implications.:

(1) V2 and higher areas for top-down selection and post selectional decoding
(2) Saliency signals in higher areas inherited from V1