A bottom up visual saliency map in the primary visual cortex --- theory and its experimental tests.

Li Zhaoping

University College London

Adapted and updated from the invited presentation at COSYNE (computational and systems Neuroscience) conference Salt Lake City, Utah, February 24, 2007

Last changed Jan, 2012

Outline

Saliency --- for visual selection and visual attention

Hypothesis --- of a bottom-up saliency map in the primary visual cortex (V1) theory

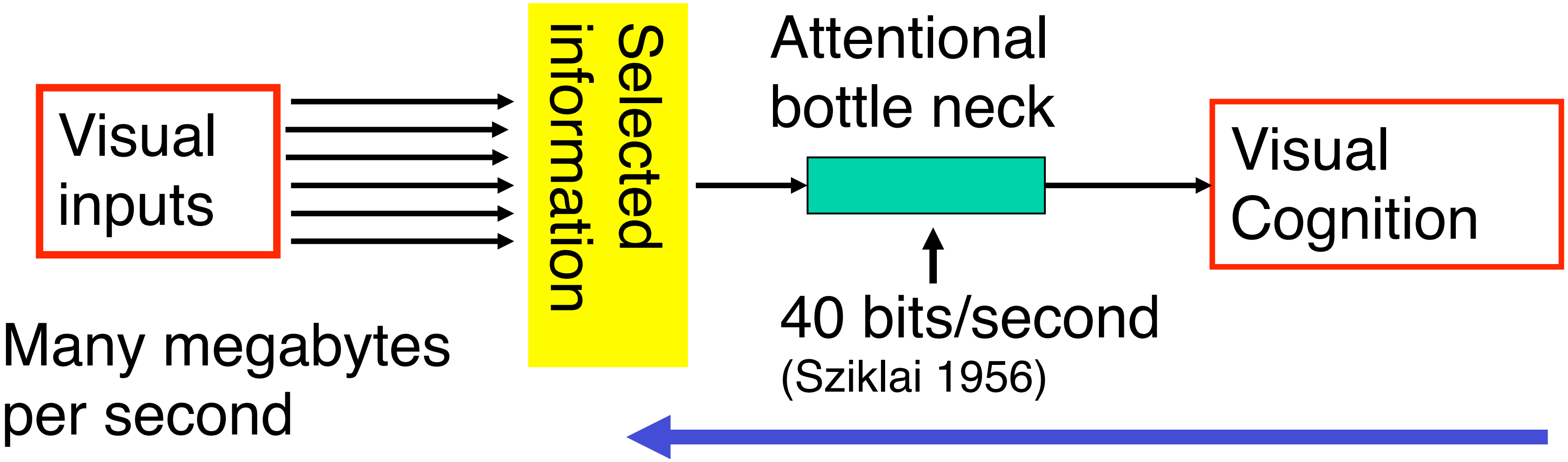
Test 1:

V1 mechanisms (simulated in a model) explain the known behavioral data on visual saliency

Test 2:

Psychophysical/fMRI/ERP tests of the predictions of the V1 theory

Visual selection



Top-down selection: goal directed

(Desimone & Duncan 1995, Treisman (1980), Tsotsos (1991), Duncan & Humphreys (1989), etc.)

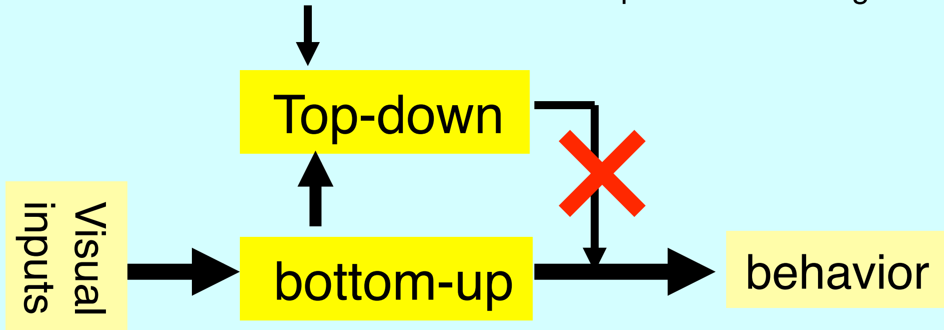
Bottom-up selection: input stimulus driven

Faster and more potent

(Jonides 1981, Nakayama & Mackeben 1989)

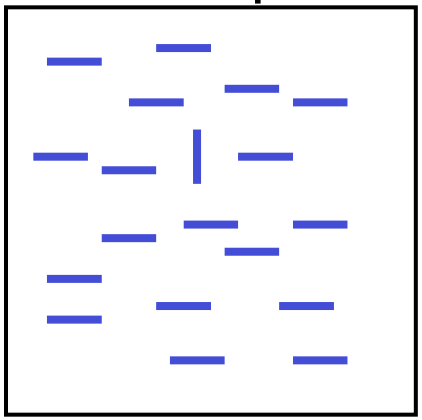
Focus of this talk

Studying bottom-up, by a reduction-ist approach, in an open loop condition when the top-down factors are negligible, e.g., soon after stimulus onset and when there is no top-down knowledge



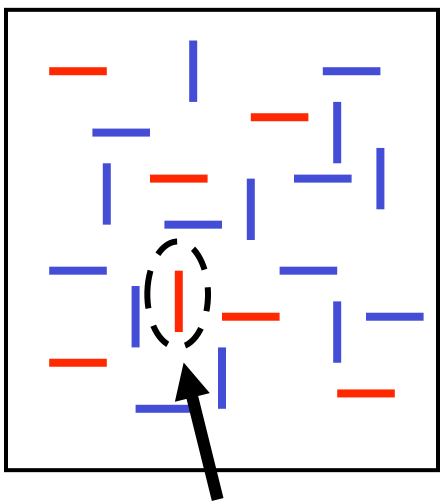
Bottom up visual selection and visual saliency

Visual inputs



Feature search

The vertical bar pops out automatically --- very fast, parallel, pre-attentive, effortless.



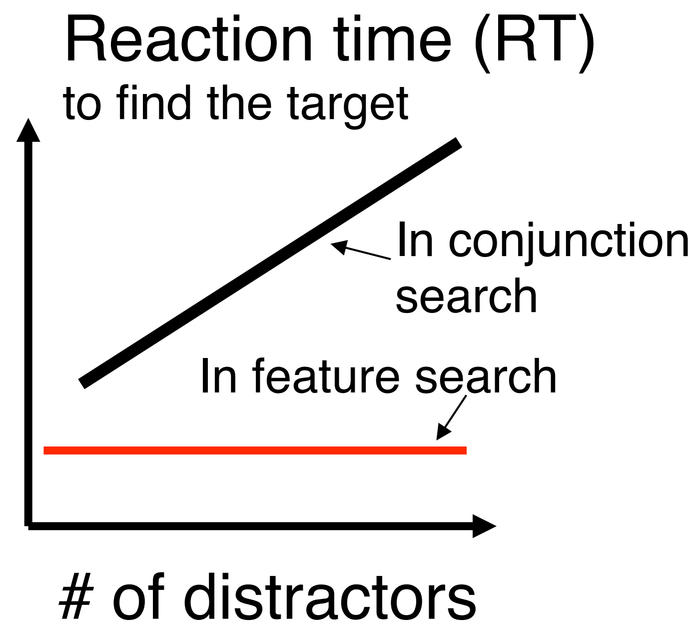
Conjunction search

slow & effortful

Unique conjunction of red color and vertical orientation

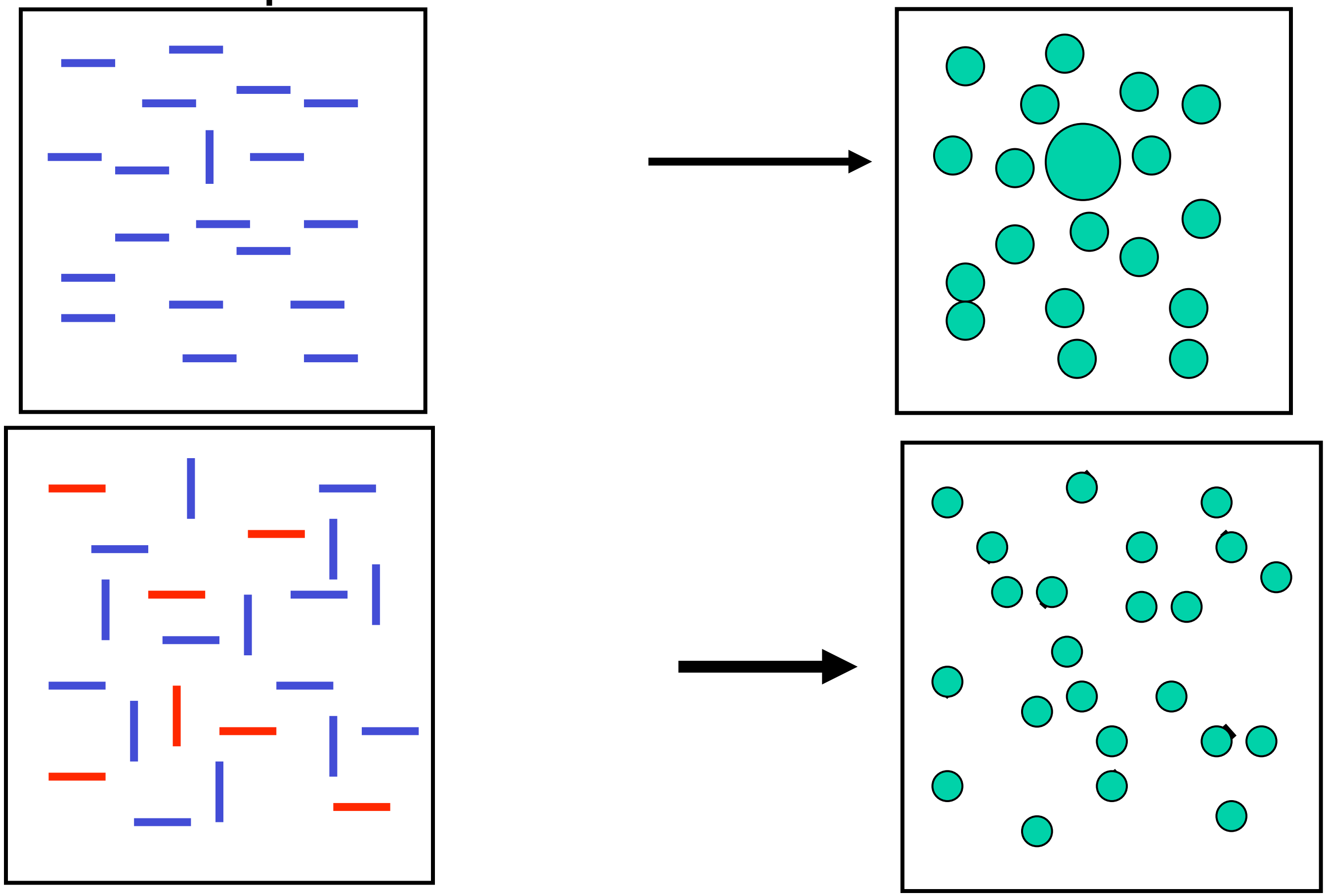
Studied in visual search

(Treisman & Gelade 1980, Julesz 1981, Wolfe et al 1989, Duncan & Humphreys 1989 etc)



Bottom up visual selection and visual saliency

Visual inputs | Saliency map of the visual space



To guide attentional selection.

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

Question: where is the saliency map in the brain?

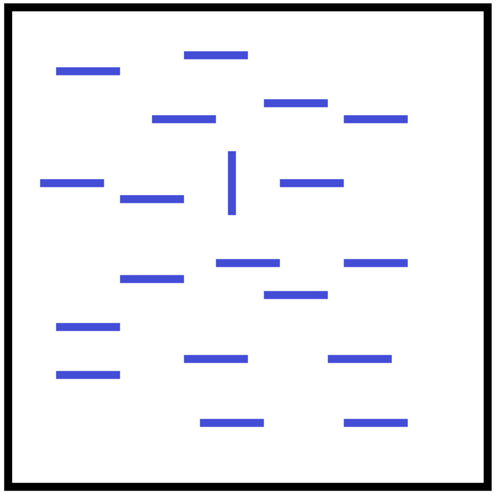
Hint: selection must be very fast, the map must have sufficient spatial resolution

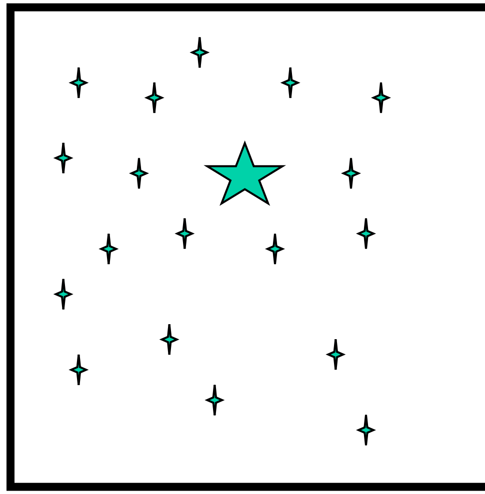
Additionally: let us find an answer that is as simple as possible

Hypothesis: (Li, Z . PNAS 1999, Trends in Cognitive Sciences, 2002)

The primary visual cortex (V1) creates a saliency map

Retina inputs → V1 neural firing rates → Higher visual areas for other functions **(after selection)**

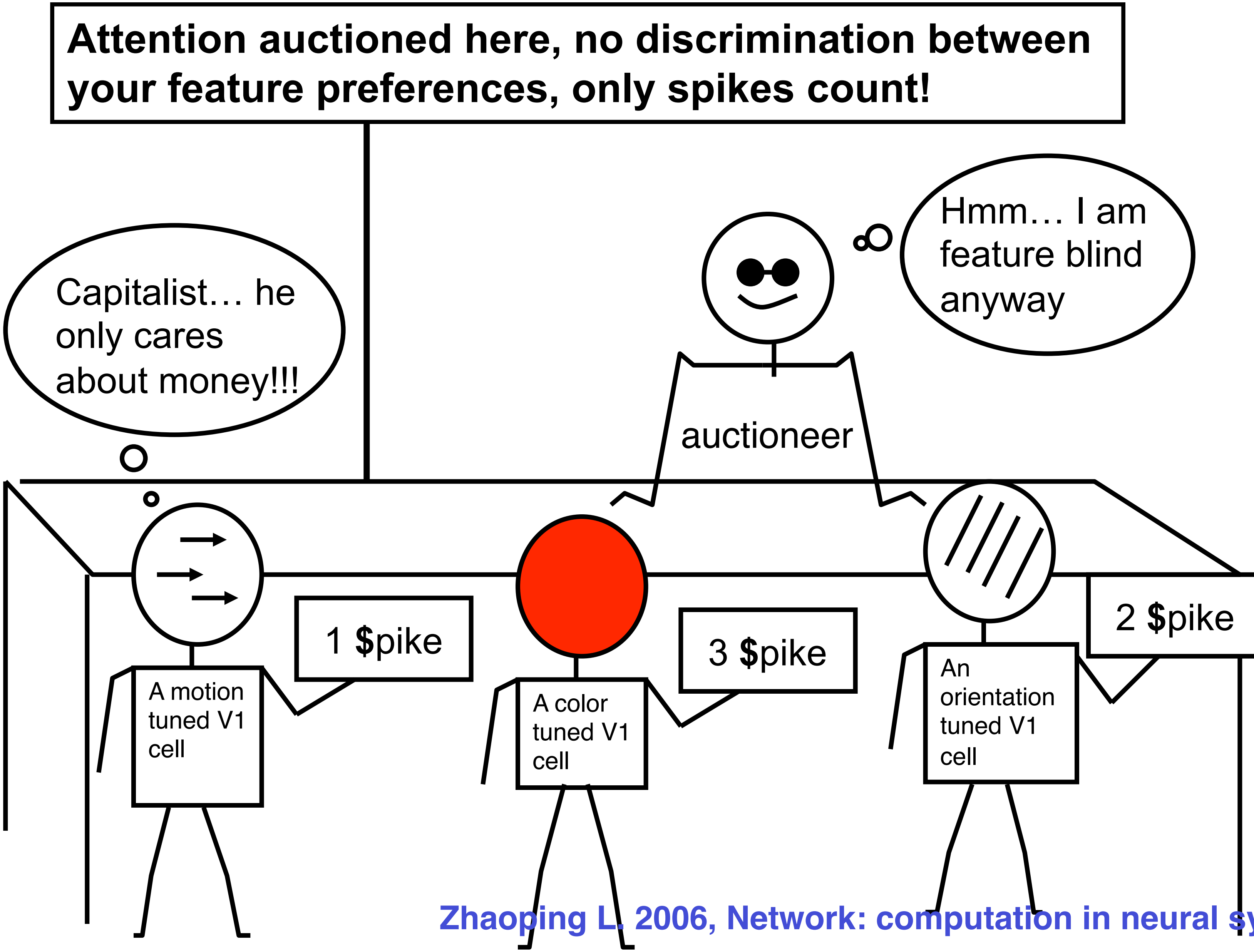




→ Superior Colliculus to drive gaze shift and thus selection

How does V1 do it?
(explained in a moment)
But V1 cells are tuned to image features like orientation, etc, how come they signal saliency? --- see next page

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



Zhaoping L. 2006, Network: computation in neural systems

Attention does not have a fixed price!

So saliency depends on relative rather than absolute responses between neurons, multi-unit recording from many cells required to determine saliency in physiological experiments.

Questions one may ask **(answered in Zhaoping 2006, Network, Computational in Neural Systems)**

Haven't the others said that V1 is only a low-level area, and the saliency map is in LIP (Gottlieb & Goldberg 1998), FEF, or higher cortical areas?

--- short answer, "yes", but the bottom-up components of saliency signals in these higher areas maybe relayed from V1

Didn't you say more than a decade ago that V1 does efficient (sparse) coding which also serves object invariance?

--- short answer, "yes"

(but data compression is not enough to fit all data in the attentional bottle neck)

Do you mean that cortical areas beyond V1 could not contribute to saliency additionally? --- short answer "no".

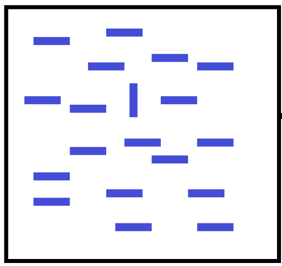
(empirical studies needed to find the contributions from other areas)

Do you mean that V1 does not also play a role in learning, object recognition, and other goals? --- short answer "no"

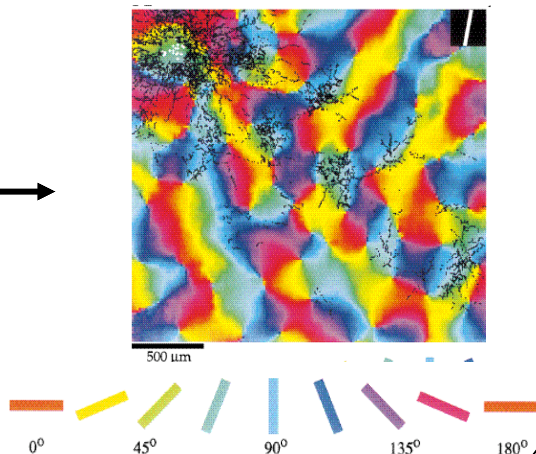
How does V1 do it ? (after all saliency depends on context)

Many cells, with overlapping receptive fields, tuned to orientation, color, or both, can all respond to a single item

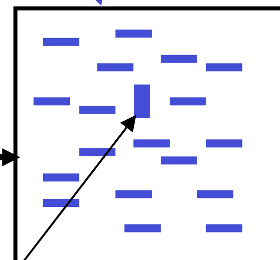
Visual input



V1

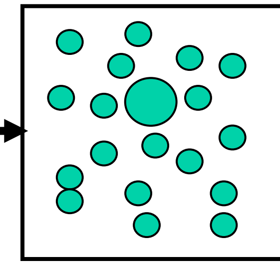


Bosking et al 1997



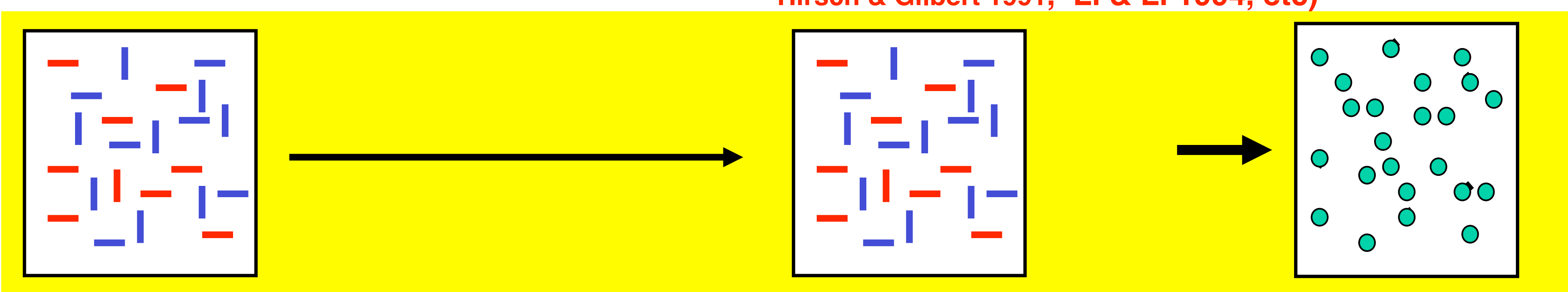
Maximum response at each location

Saliency map



Neuron tuned to vertical orientation responding to the vertical bar is the only one not suffering from iso-orientation (iso-feature) suppression, thus gives the highest response.

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)

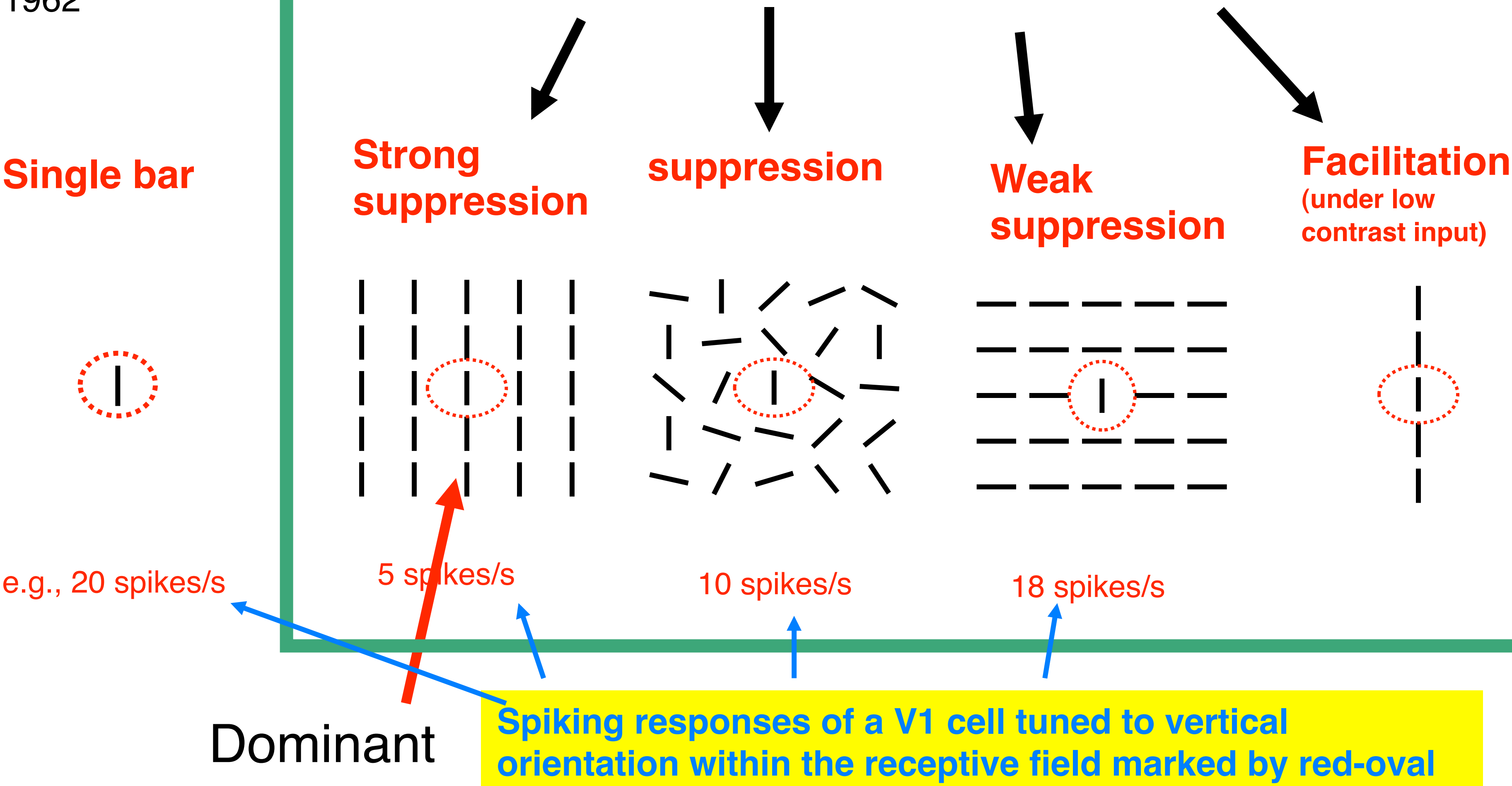


Physiologically observed in V1:

Classical receptive fields

Hubel & Wiesel 1962

Contextual influences (since 1970s, Allman et al 1985, Knierim van Essen 1992, Hirsch & Gilbert 1991, Li & Li 1994, Kapadia et al 1995, Nothdurft et al 1999 etc) --- **nuisance for Hubel & Wiesel's receptive fields, but useful for saliency computation**



Testing the V1 saliency map --- 1

V1 outputs —Explain→ Saliencies in visual search and segmentation

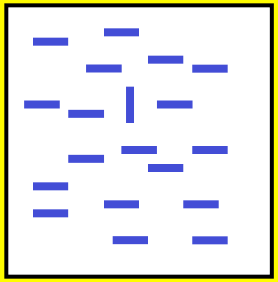
Few physiological data

difficult experiments to do multiunit recording ….

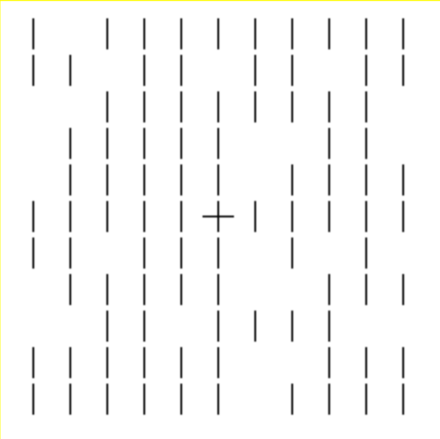
Solution: build a V1 model

multi-unit recording on the model
(Li, 1998, 1999, 2000, 2002, etc)

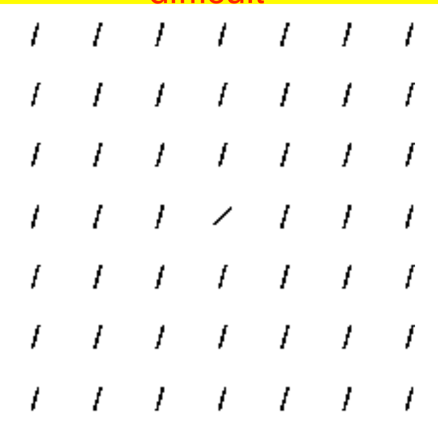
Feature search
--- easy



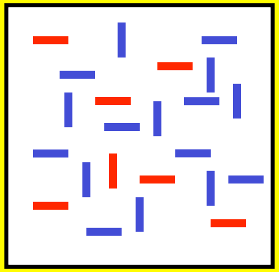
'+' among 'l's
--- easy



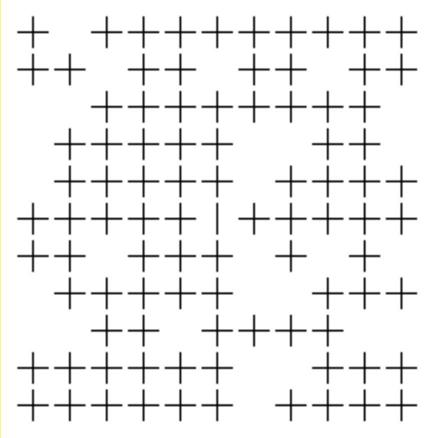
'/' among 'l's
regular background
---difficult



Conjunction search
--- difficult



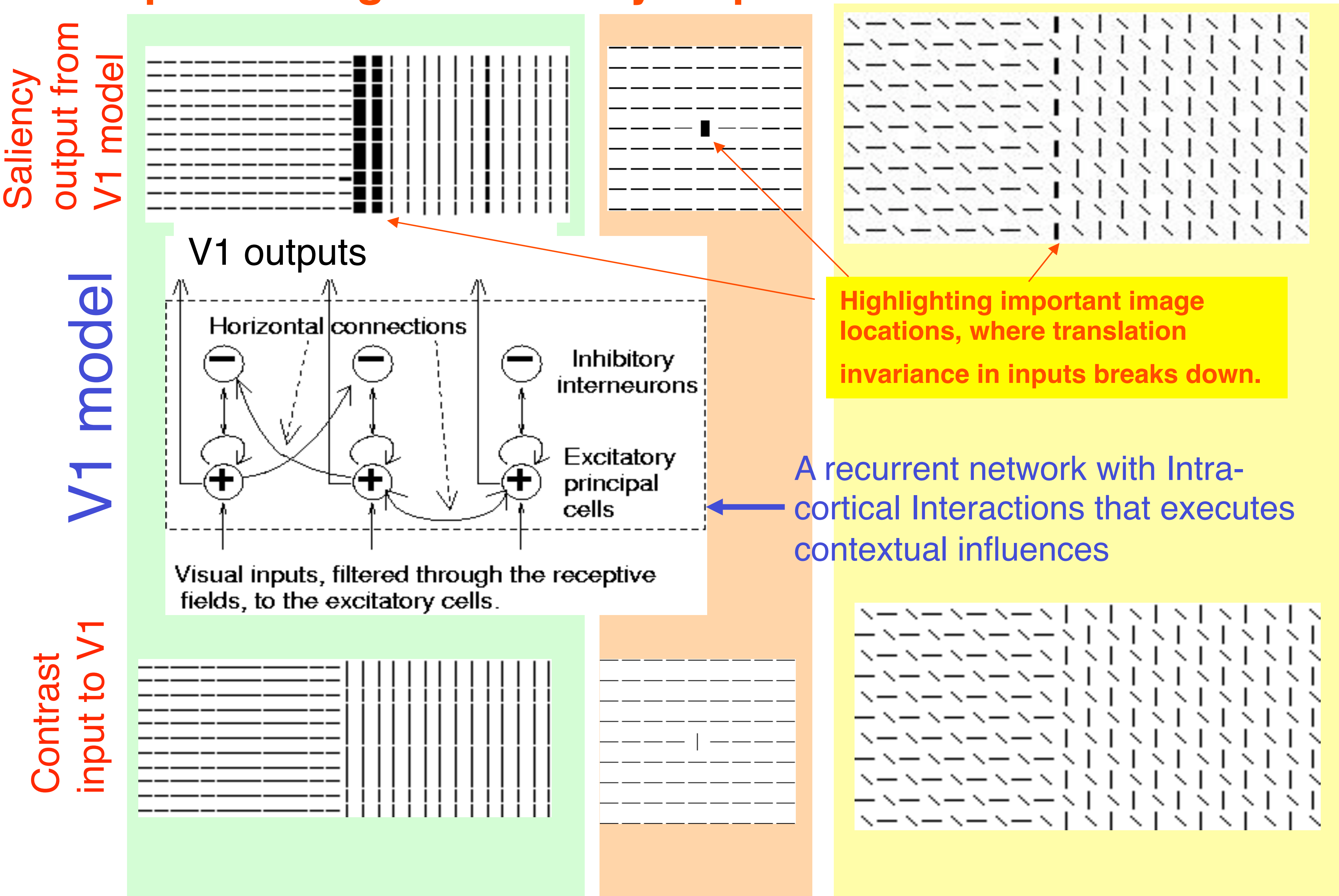
'l' among '+'s
--- difficult



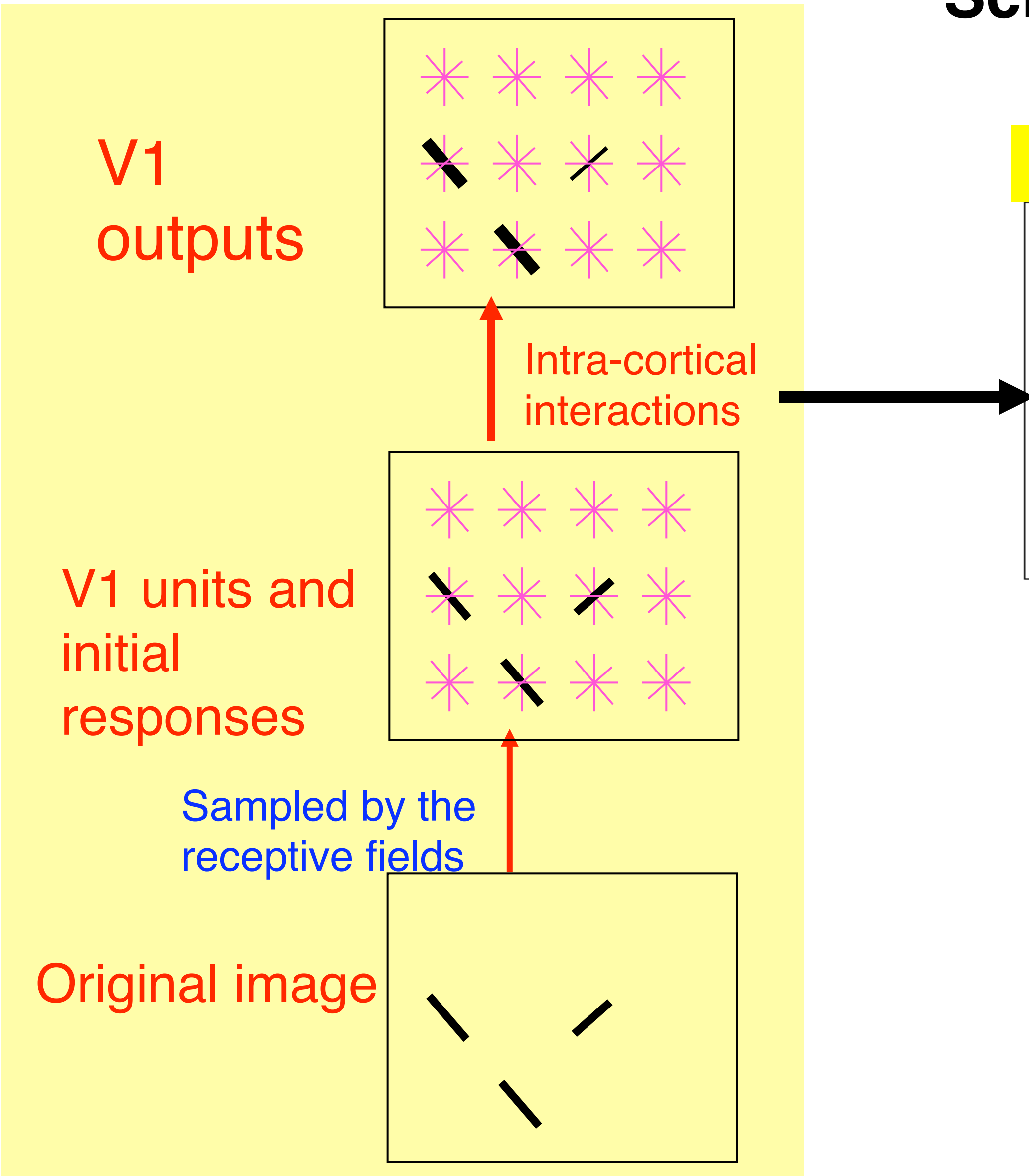
'/' among 'l's
irregular background
---difficult

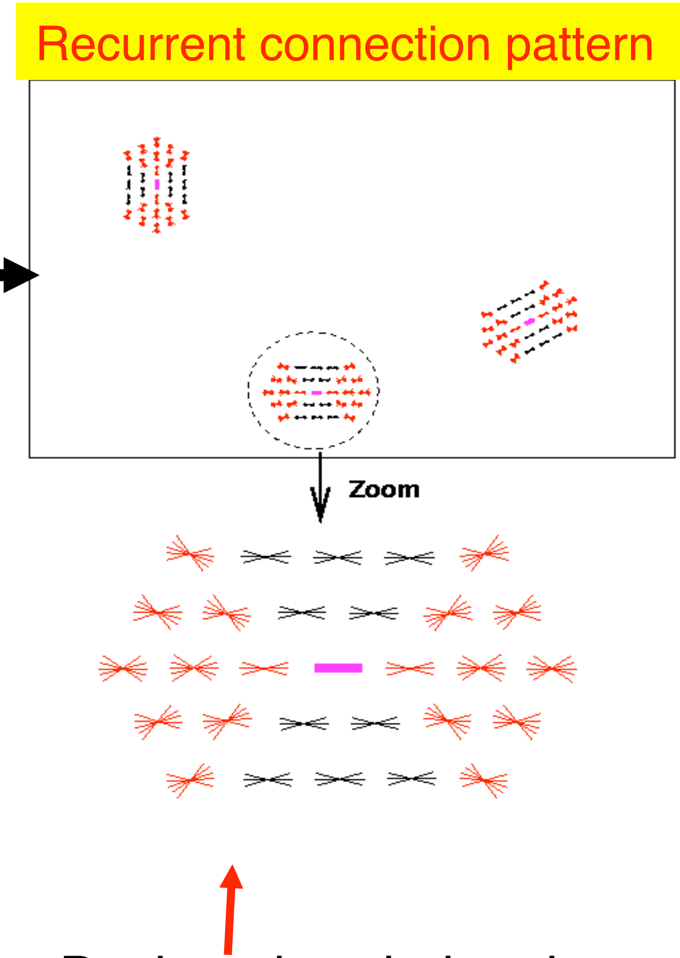
More examples in literature, **e.g., Treisman & Gelade 1980, Julesz 1981, Duncan & Humphreys 1989, Wolfe et al 1989, etc.**

Implementing the saliency map in a V1 model



Schematics of how the model works





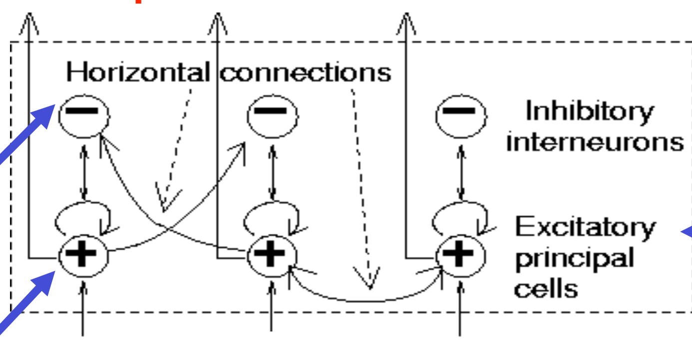
Designed such that the model agrees with physiology results on contextual influences.

Recurrent dynamics -- differential equations of firing rate neurons interacting with each other with sigmoid like nonlinearity.

See Li (1998, 1999, 2001), Li & Dayan (1999) for the mathematical analysis and computational design of the nonlinear dynamics.

Output $g_x(x_{i\theta})$

V1 model



Intra-cortical Interactions

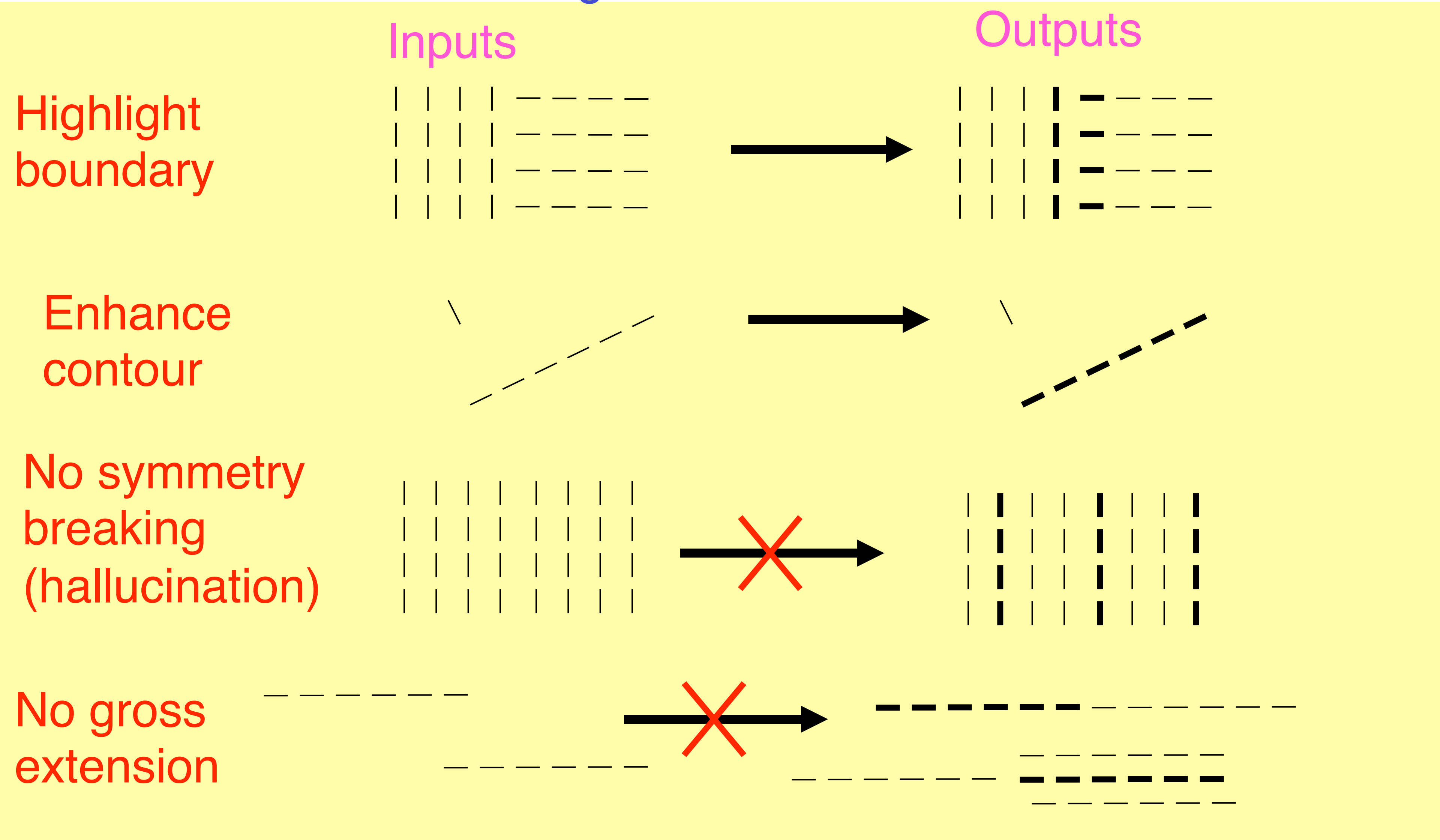
Input $I_{i\theta}$, after filtering through classical receptive fields

$$dx_{i\theta}/dt = -x_{i\theta} - g_y(y_{i\theta}) + \sum_{j\theta'} J_{i\theta,j\theta'} g_x(x_{j\theta'}) + I_{i\theta}$$

$$dy_{i\theta}/dt = -y_{i\theta} + \sum_{j\theta'} W_{i\theta,j\theta'} g_x(x_{j\theta'}) + I_o$$

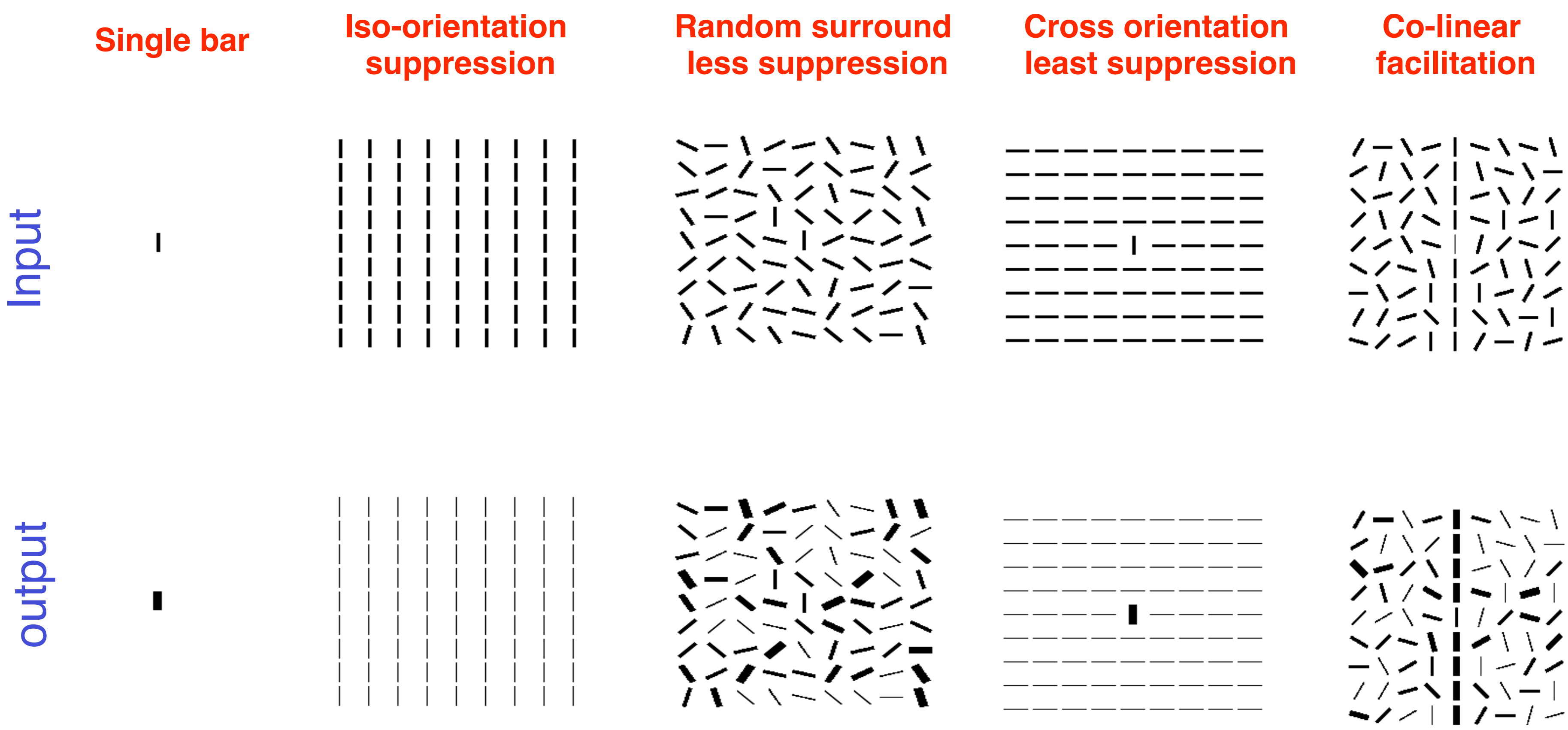
Recurrent connections

Constraints used to design the intra-cortical interactions.



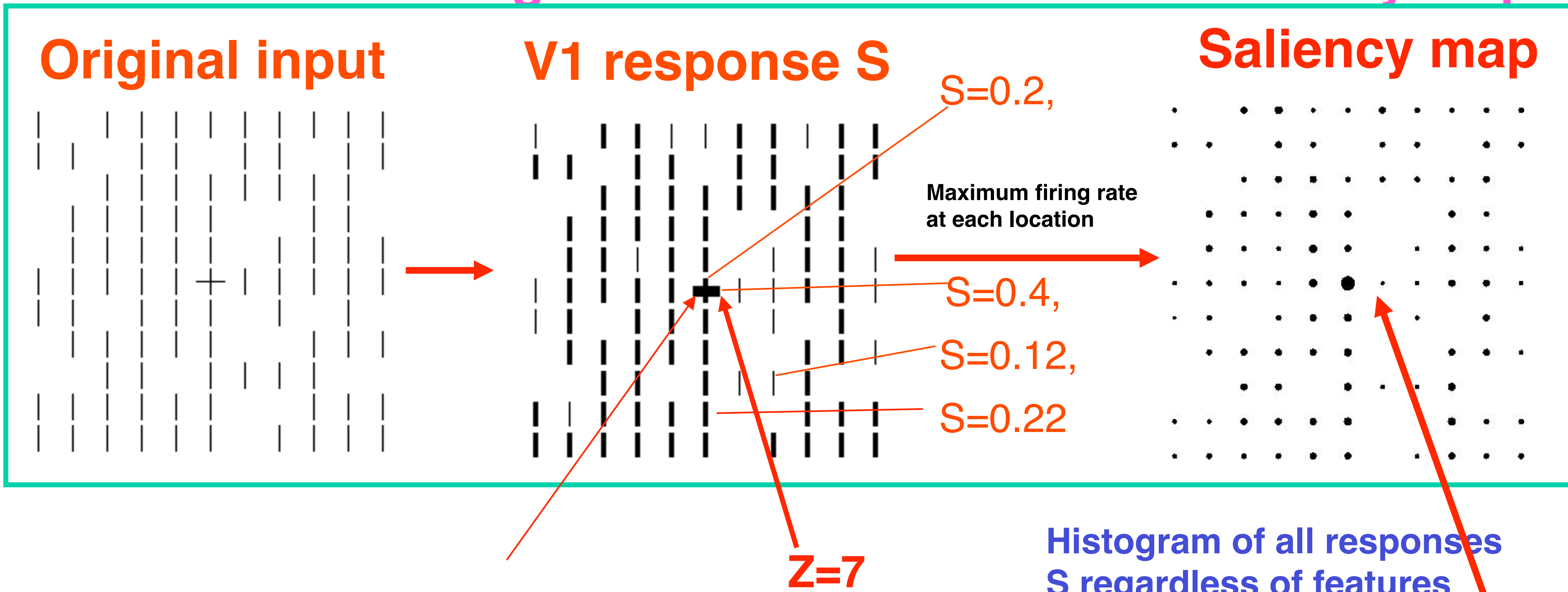
Design techniques: mean field analysis, stability analysis. Computational design constraints the network architecture, connections, and dynamics. Network oscillation and synchrony between neurons to the same contour is one of the dynamic consequences (Li, 2001, Neural Computation).

Make sure that the model can reproduce the usual physiologically observed contextual influences



Once the V1 model is calibrated by the real V1 using this procedure, all model parameters are fixed and we can proceed to examine the model behavior when presented with visual inputs.

Multi-unit recording on the model to view the saliency map



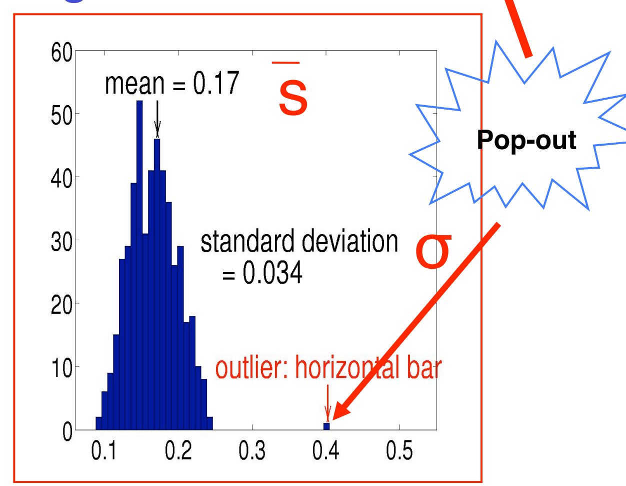
The horizontal bar evokes the highest response since it is the only one without any iso-orientation neighbors, thus the neuron responding to it does not suffer from iso-orientation suppression.

Note that the cross pops out of the bars even though V1 does not have any neuron tuned to the shape of a cross.

$Z = (S-\bar{S})/\sigma$

--- z score, measuring saliencies of items

Histogram of all responses S regardless of features



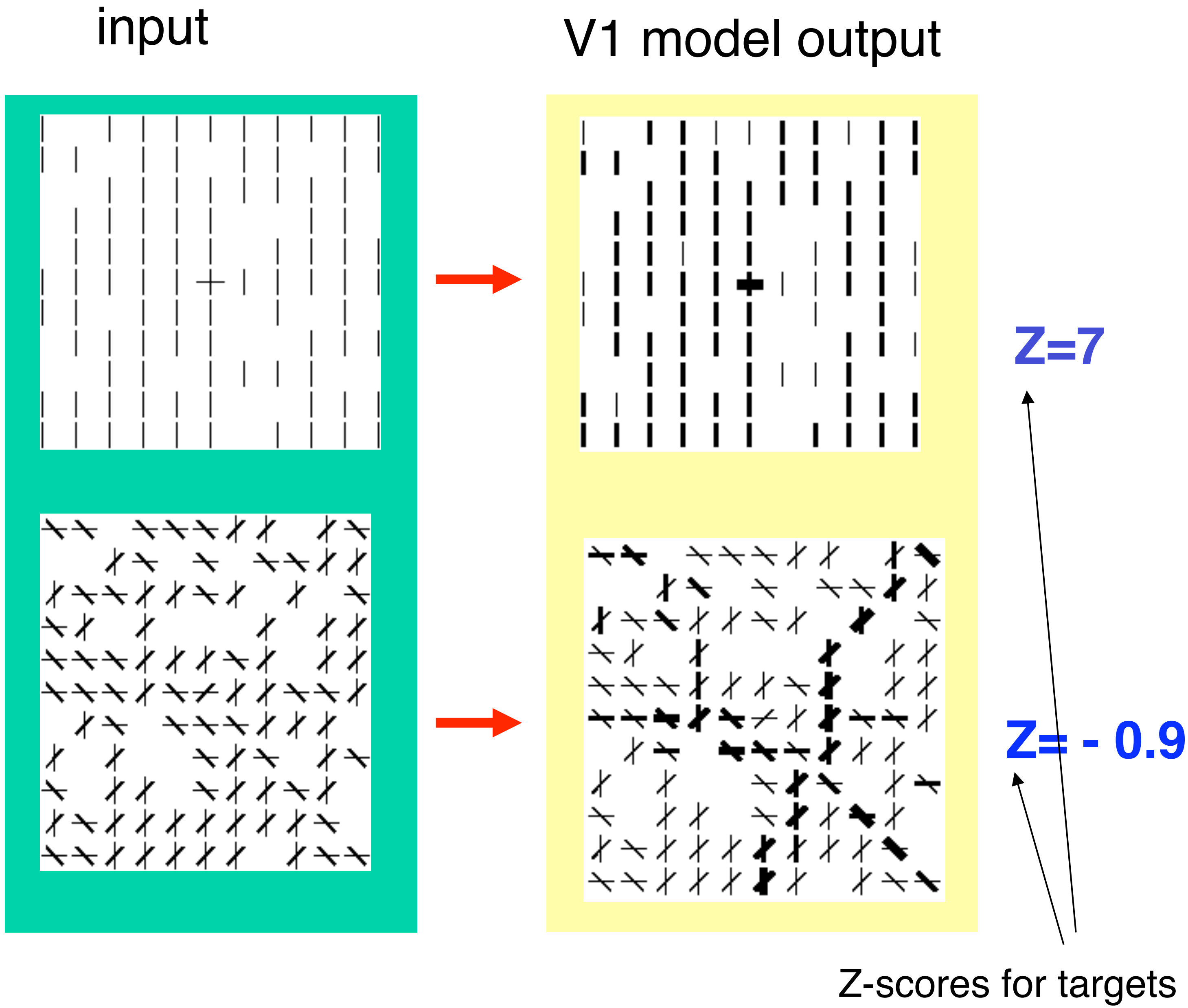
The V1 saliency map agrees with visual search behavior.

Feature search --- pop out

Target = '+'

Conjunction search --- serial search

Target=



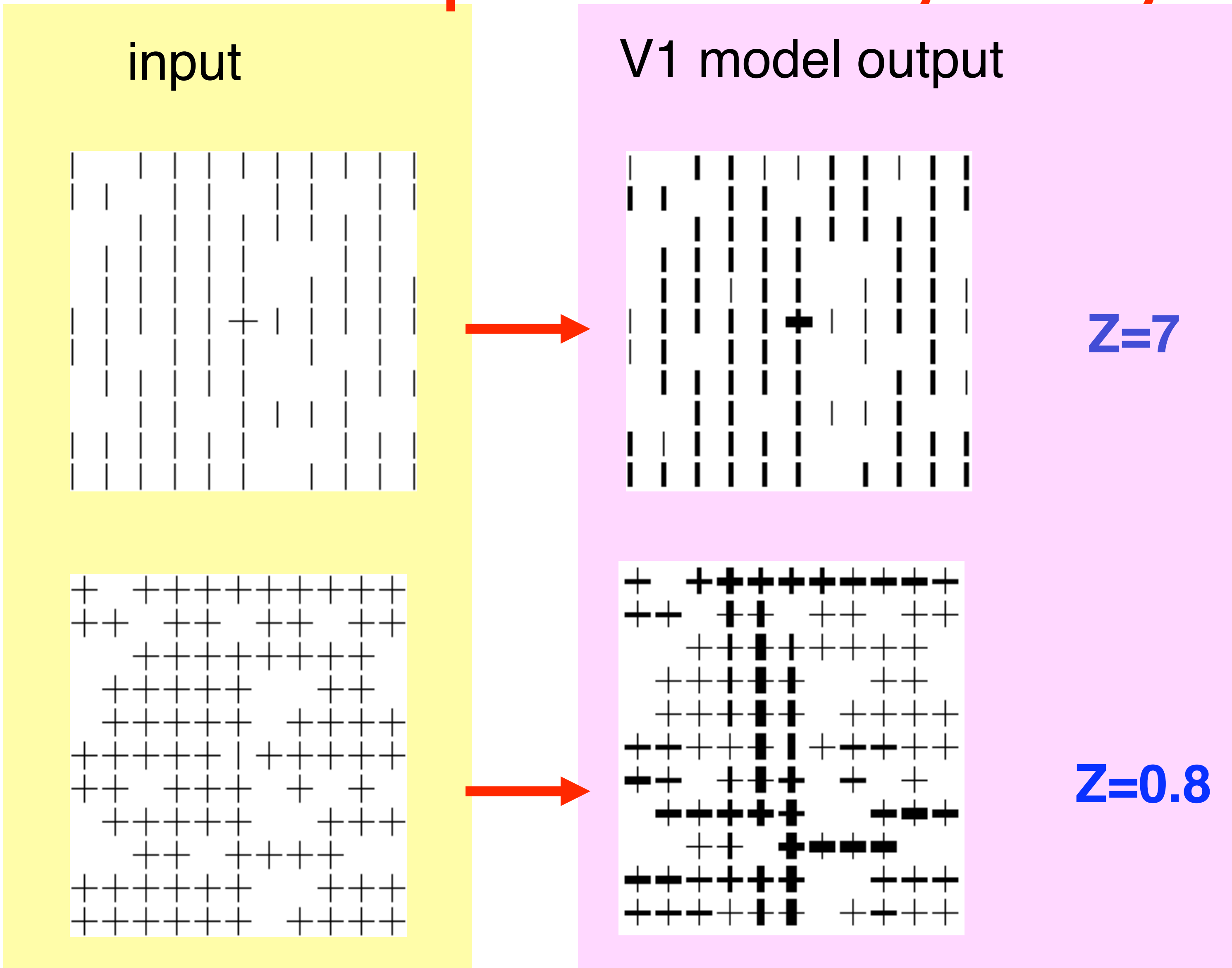
Explains a trivial example of search asymmetry

Feature search --- pop out

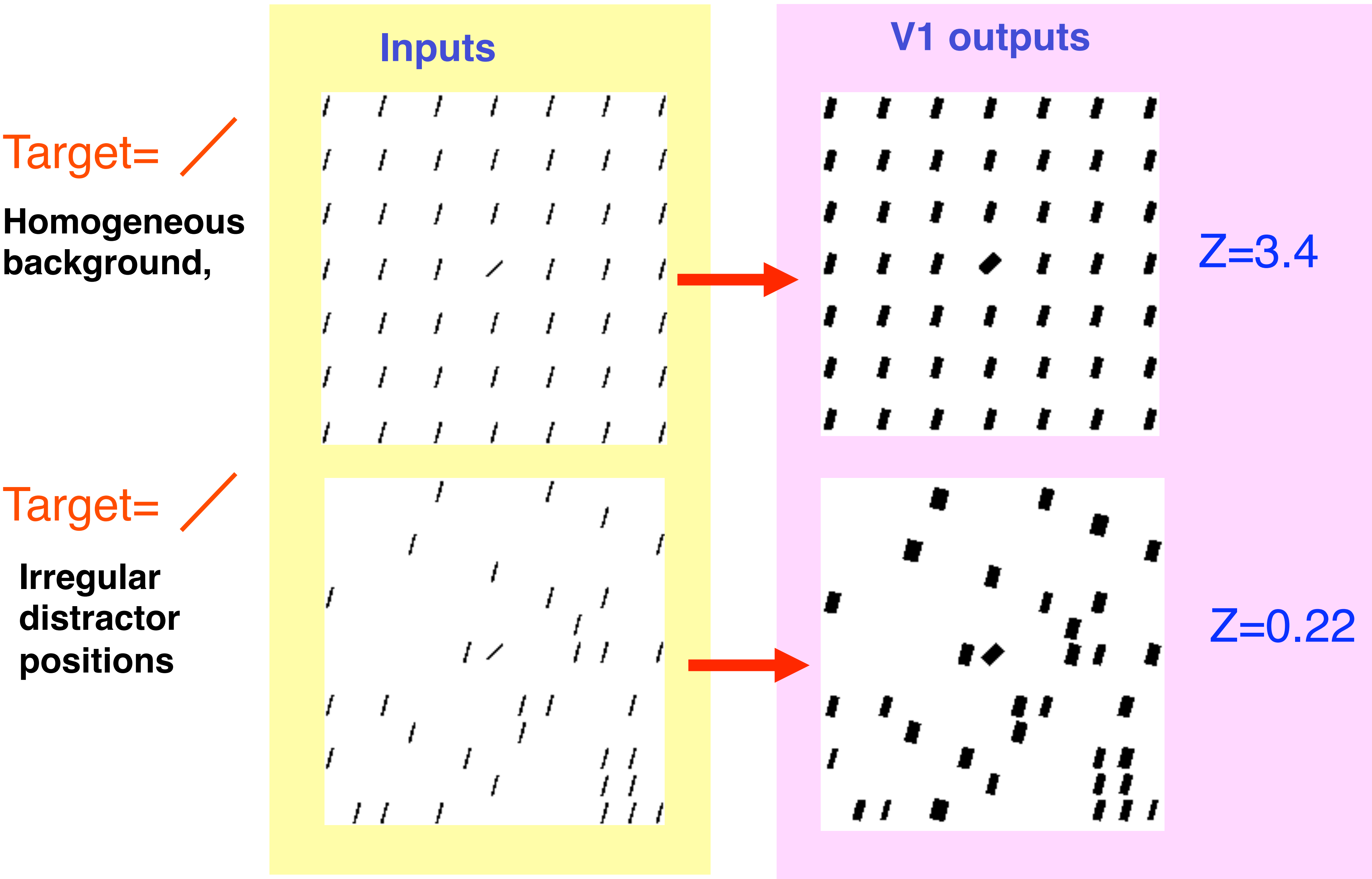
Target = +

Target lacking a feature

Target = |

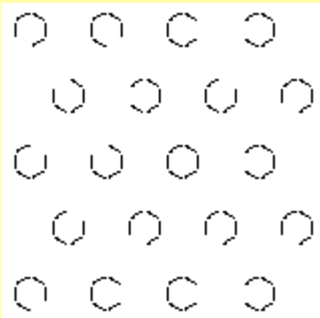


Explains background regularity effect

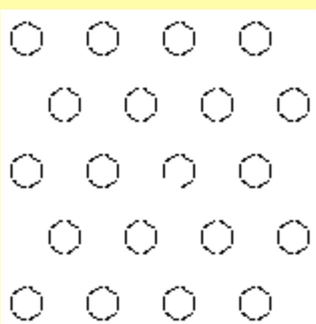


More severe test of the saliency map theory by using subtler saliency phenomena --- search asymmetries (when ease of visual search changes upon target-distractor identity swap, Treisman and Gormican 1988)

Open vs. closed circles



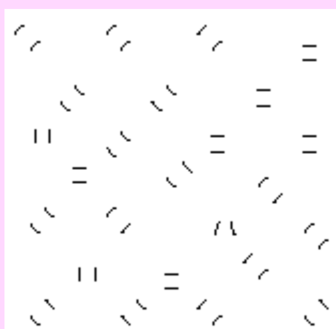
Z=0.41



Z=9.7

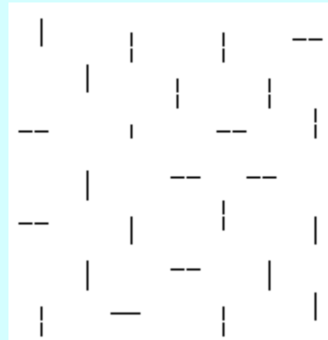
parallel vs. divergent pairs of bars

Z= -1.4

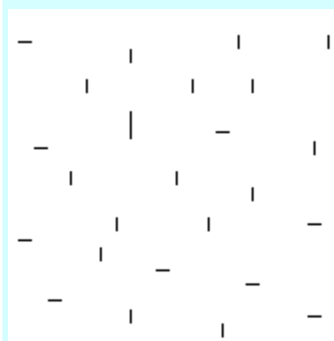


Z= 1.8

long vs. short bars



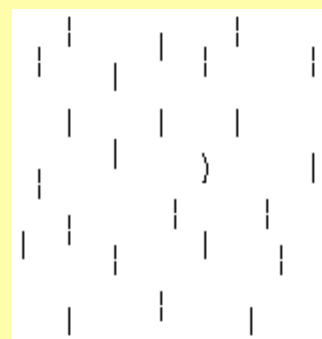
Z= -0.06



Z= 1.07

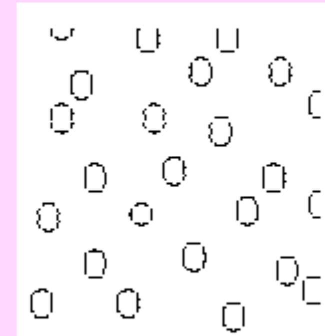
curved vs. straight

Z= 0.3

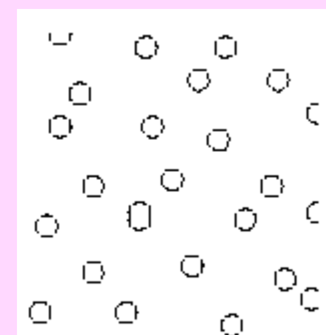


Z= 1.12

ellipse vs. circle



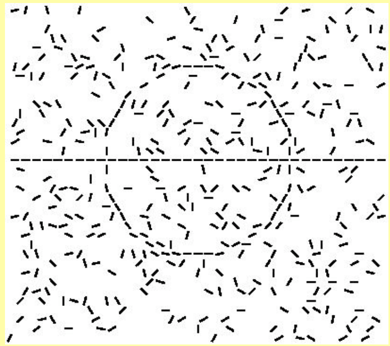
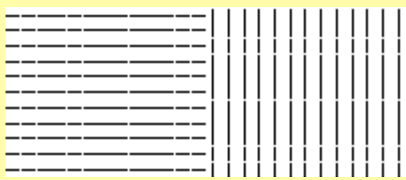
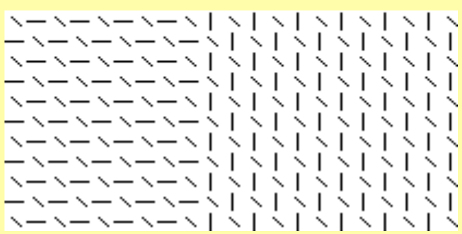
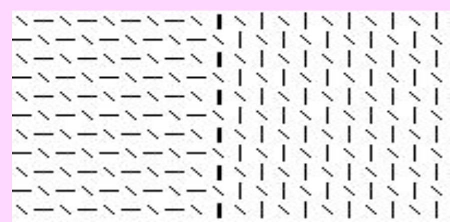
Z= 0.7



Z= 2.8

Model behavior agrees with the directions of asymmetry in all five examples, with zero parameter tuning. Note that V1 cells are not tuned to circles etc, but respond to oriented bar/curve segments in inputs. Highest response to segments of the target is used to compute the Z-score for the target.

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 according to the V1 model response

See Li 1998, 1999, 2000 and Zhaoping 2003 for more examples of the model’s accounts of previous behavioral data

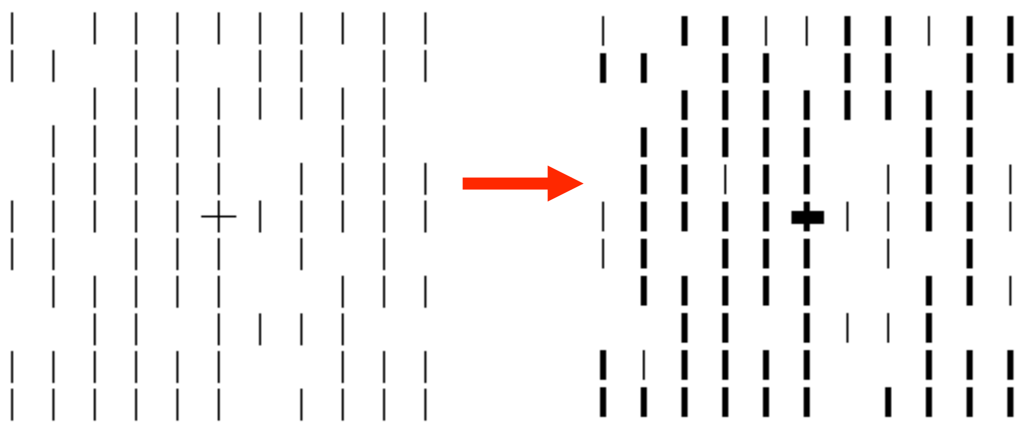
Testing the V1 saliency map --- 2

Predicting previously unknown behavior: psychophysical test

Theory statement:

the **strongest** response at a location signals saliency.

e.g., input → V1 output



The cross is salient due to the horizontal bar alone ---

the less salient vertical bar in the cross is invisible to saliency

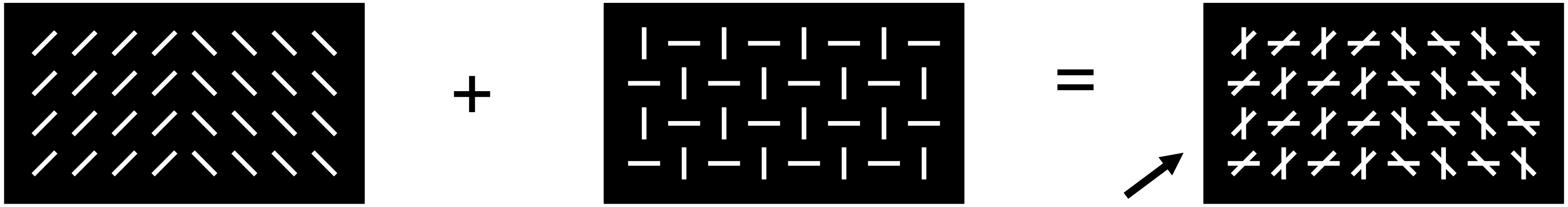
V1 theory prediction 1: A task becomes difficult when the most salient feature (at some locations) is task irrelevant.

Test stimuli

Prediction: segmenting this composite texture is much more difficult

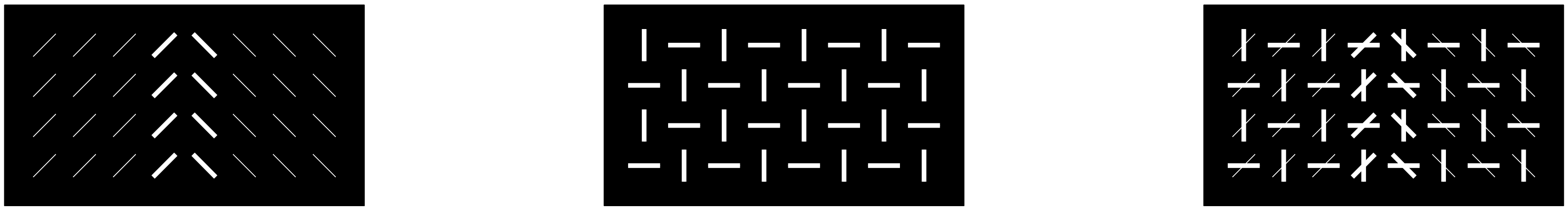
a task relevant **b** **c** composite = a+b

Schematic stimuli



Component b is task irrelevant for segmenting the texture

V1 responses

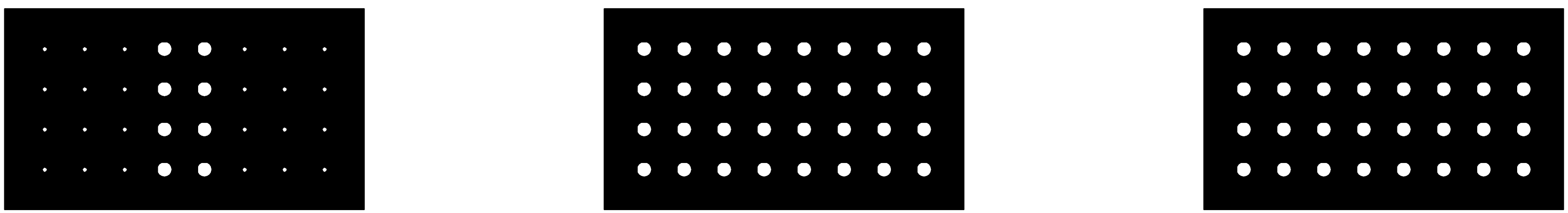


Higher responses to the texture border bars, each of which has fewer iso-orientation neighbors

Each bar, parallel to half of its neighbors, evokes a response of comparable level to that by a texture border bar in a

Responses to task irrelevant bars dictate saliency at many locations

Saliency maps, or the highest response to each location



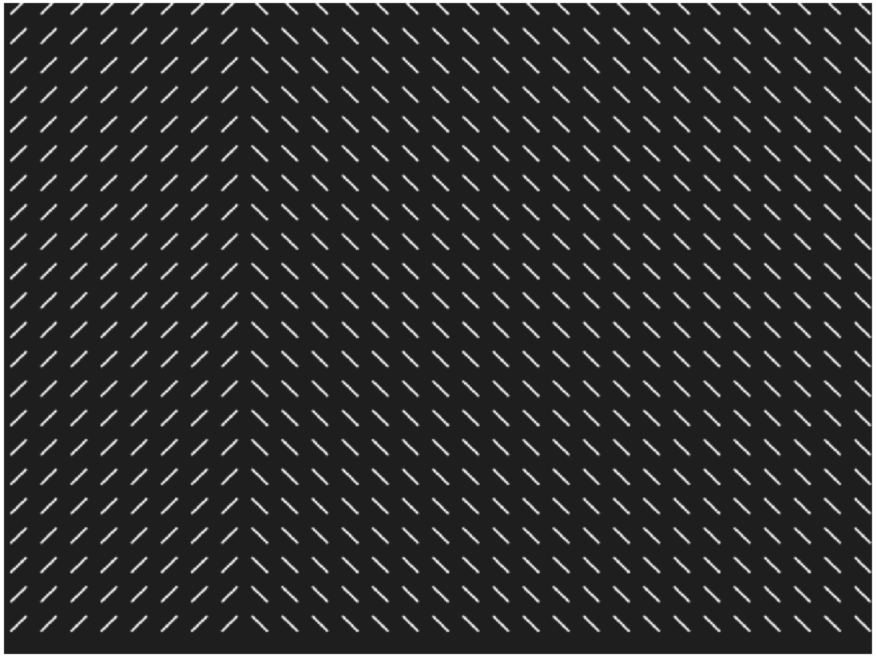
Saliency highlights at the border makes segmentation easy

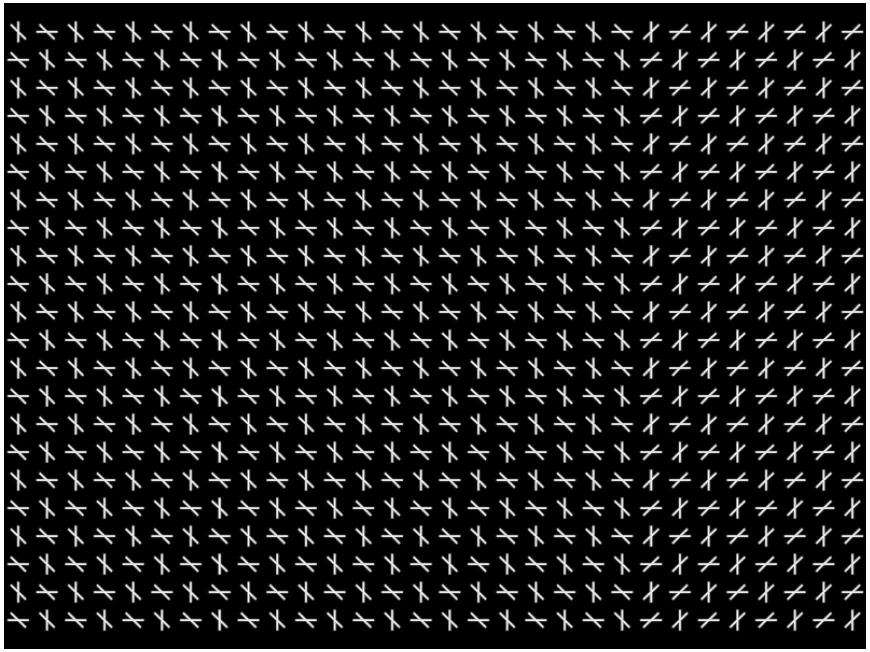
No saliency highlight at the border.

Note: if saliency at each location is determined by the sum of the neural activities at each location, the prediction would not hold.

Test: measure reaction times for segmentation:

Task: subject answer as soon as possible by button press whether the texture border is at left or right half of each image, a shorter reaction time (RT) is used to indicate a higher saliency of the texture border.

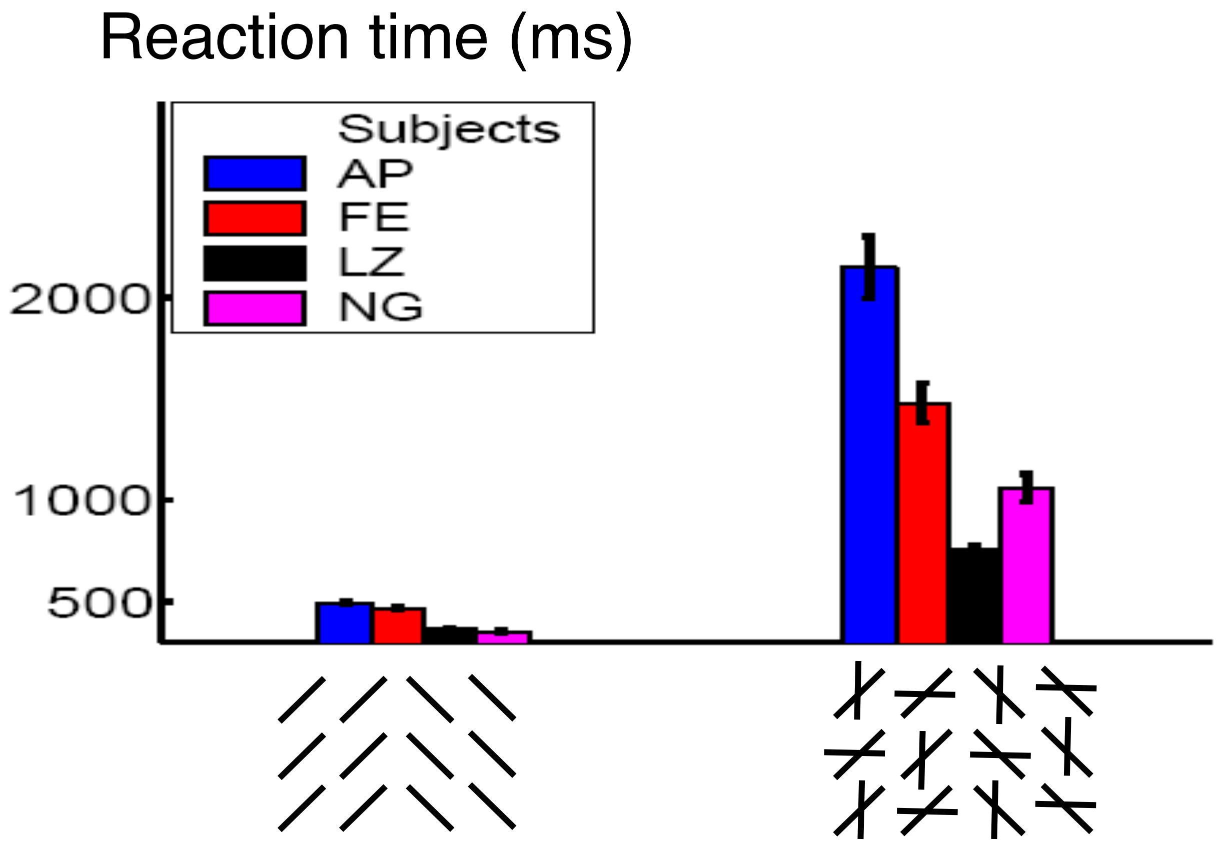




Two examples of the test stimuli

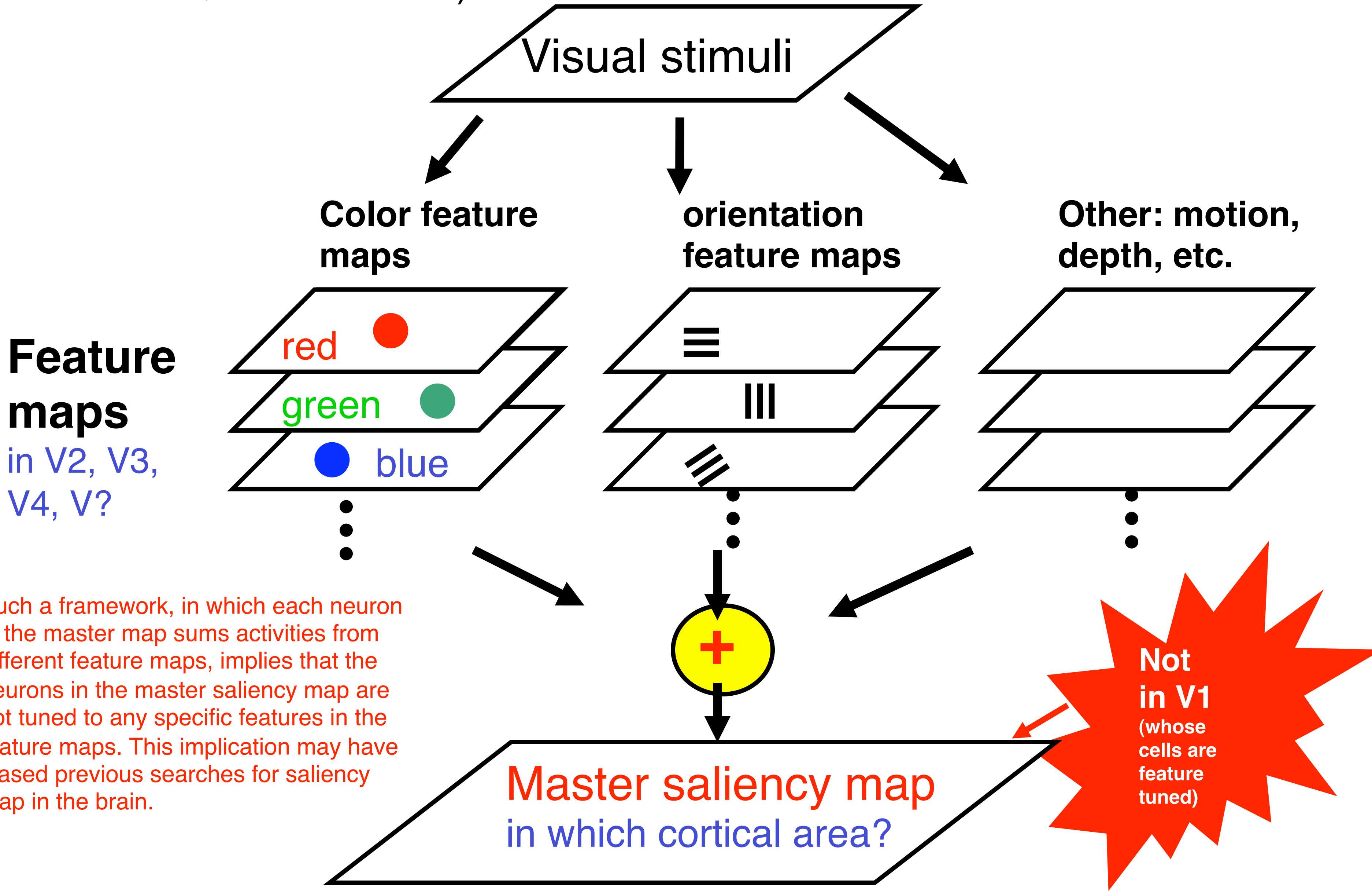
Test: measure reaction times in the segmentation task:

(Zhaoping and May, 2007, PLoS Computational Biology)



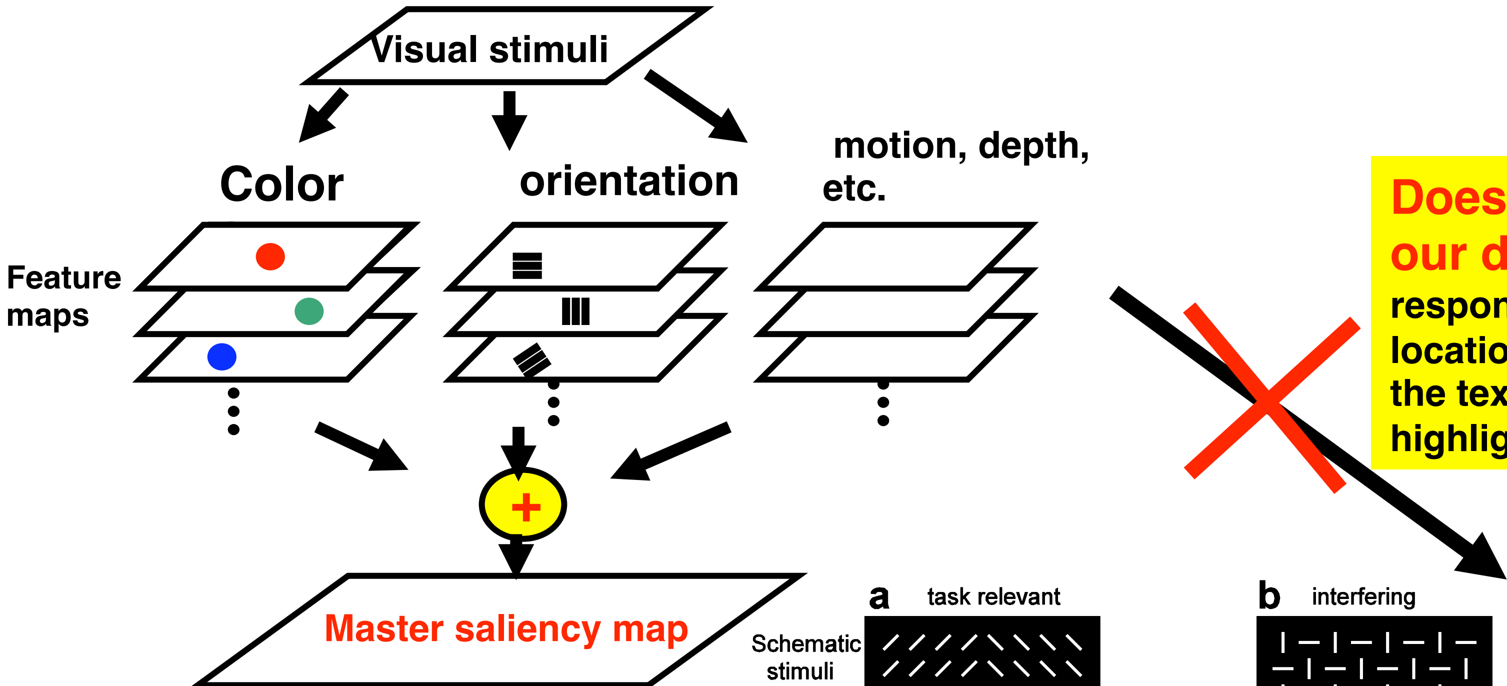
Supporting V1 theory prediction !

Previous views on saliency map (Koch & Ullman 1985, Wolfe et al 1989, Itti & Koch 2000 etc)

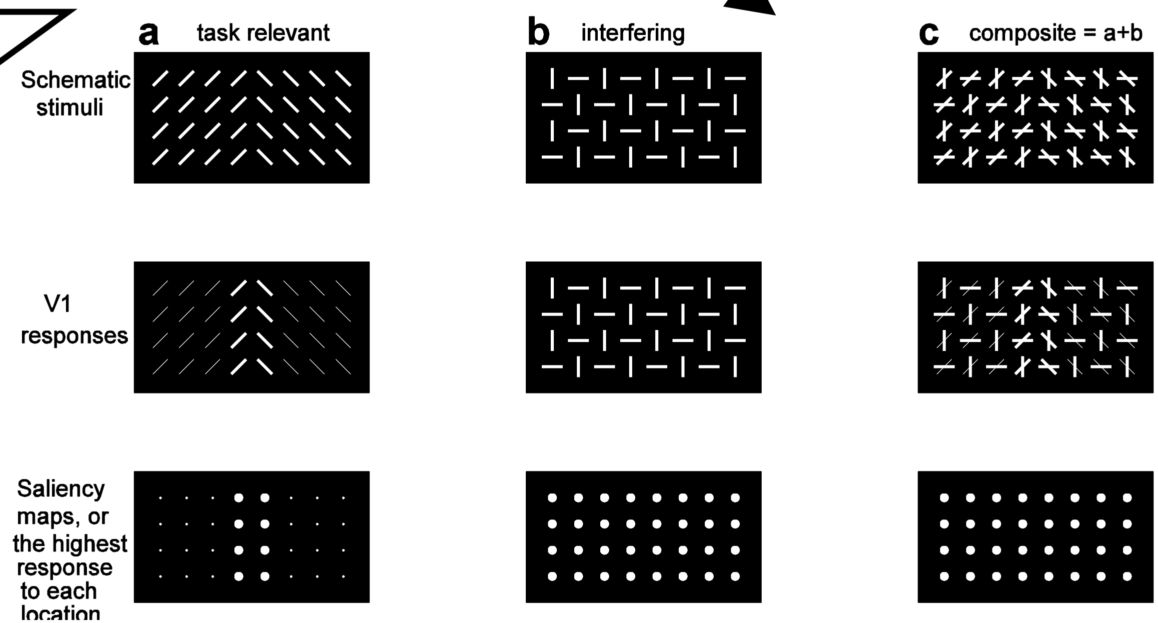


Such a framework, in which each neuron in the master map sums activities from different feature maps, implies that the neurons in the master saliency map are not tuned to any specific features in the feature maps. This implication may have biased previous searches for saliency map in the brain.

Previous views on saliency map (Koch & Ullman 1985, Wolfe et al 1989, Itti & Koch 2000 etc)



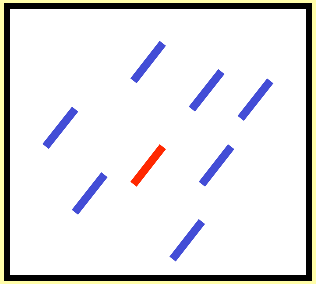
Does not predict our data since summing responses at each location would preserve the texture border highlight



V1 theory prediction 2: --- double-feature advantage

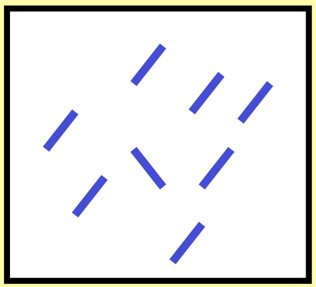
in reaction time (RT) to find singleton target

Colour pop out



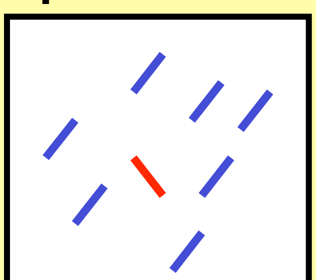
RT_1 = 500 ms → **Color** tuned cell dictates saliency

Orientation pop out



RT_2 = 600 ms → **Orientation** tuned cell dictates saliency

Double feature pop out



RT = ?

As in a race model

RT =min(RT_1, RT_2) =500 ms
or
RT< 500 ms

Color, or, **Orientation**, or
Color +Orientation conjunctive tuned cell
dictates saliency, depending on which cell is the most responsive.

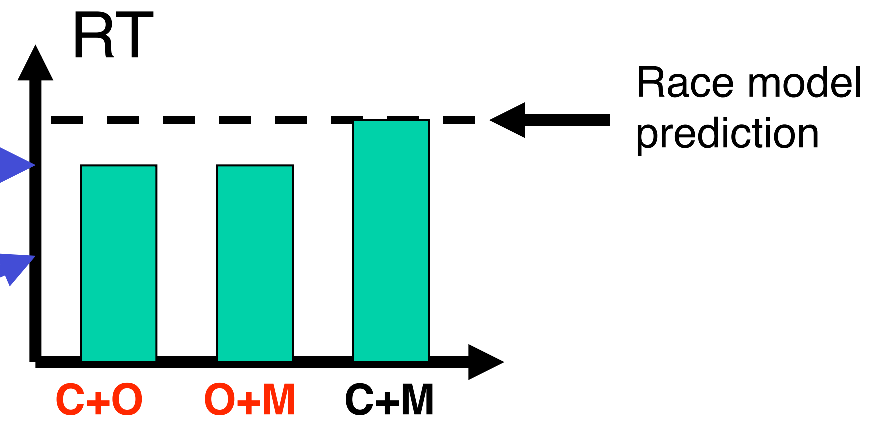
Double-feature advantage when RT is shorter than predicted by the race model

Prediction: given the conjunctive tuned cells,
RT <= 500 ms
double-feature advantage when averaged over many trials.

V1 theory prediction 2: --- double-feature advantage

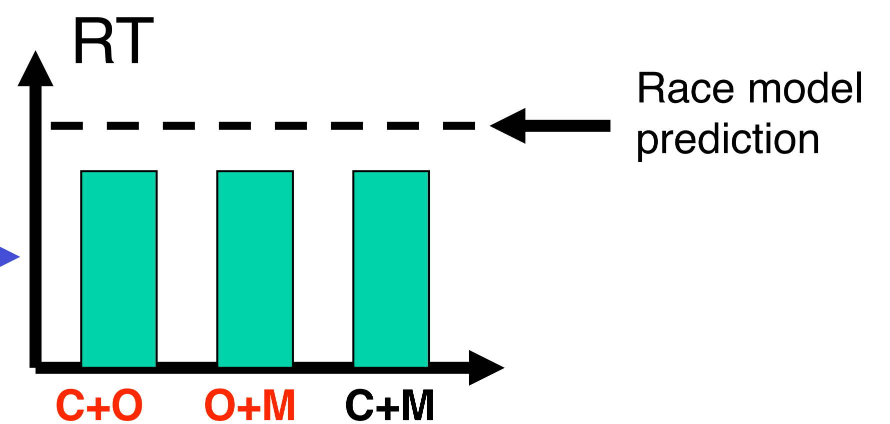
In V1, conjunctive cells exist for color and orientation (C+O), orientation and motion direction (O+M), but not for color and motion direction (C+M) (Livingstone & Hubel 1984, Horwitz & Albright 2005)

V1 saliency Prediction --- double- feature advantage for C+O, O+M, but not C+M



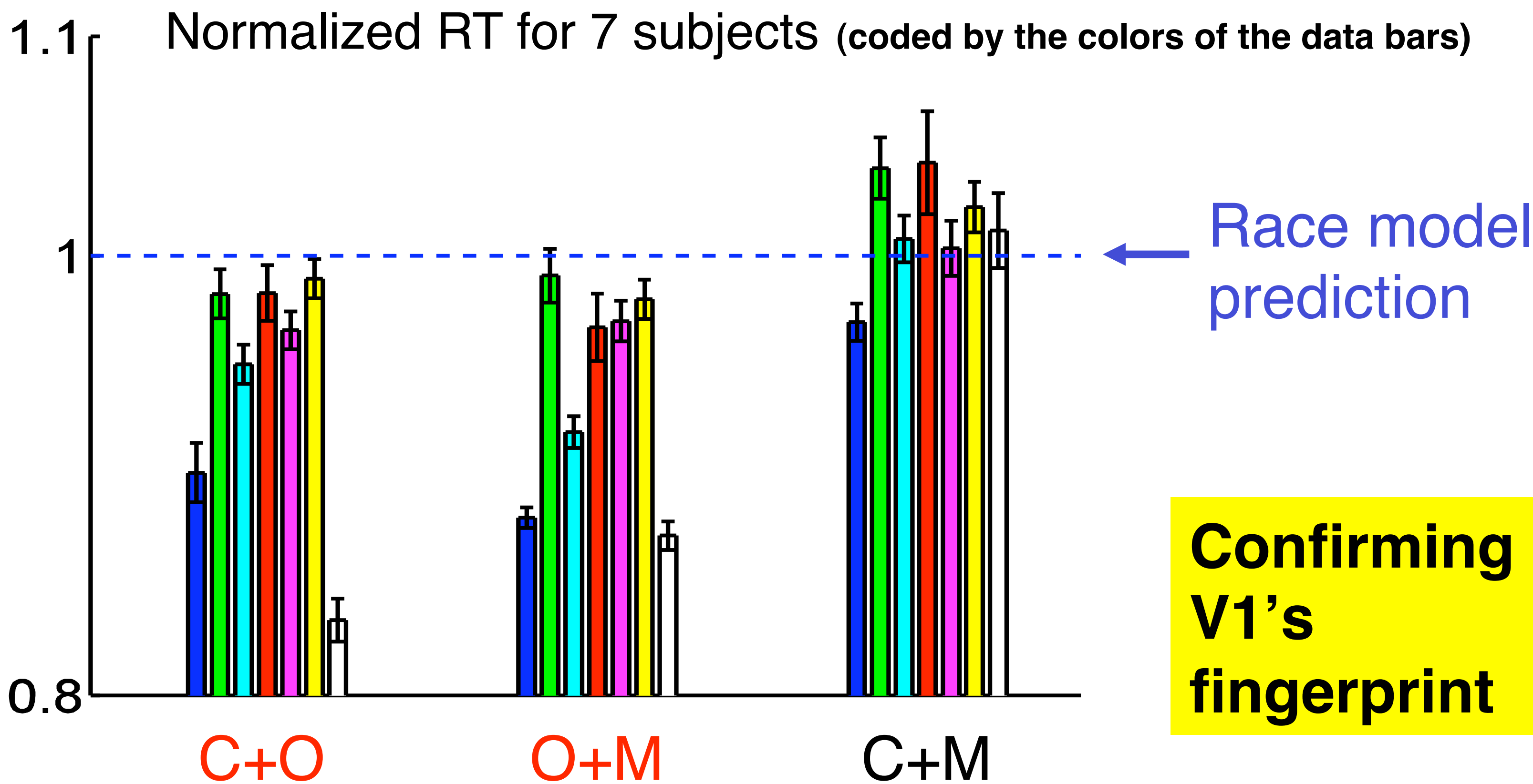
Fingerprint of V1: It is known that V2 has cells tuned to all types of conjunctive features, including C+M (Gegenfurtner et al 1996).

If V2 or higher cortical areas are responsible for saliency, then double-feature advantage should occur for all feature combinations C+O, O+M and C+M.



V1 theory prediction 2: --- double-feature advantage for C+O, O+M, but not for C+M

Test: compare the RT for double-feature search with that predicted by the race model **(Koene & Zhaoping 2007, Journal of Vision)**



Confirming V1's fingerprint

Method: subjects press button ASAP for odd-one-out target's location (left or right half of the display), target features are randomly interleaved in trials and unpredictable to subjects before each trial. RTs for single feature targets were used to derive the race model predictions for the double feature target using Monte Carlo simulations. Each subject's RT for a double-feature target is normalized by the corresponding race model prediction in the plot above.

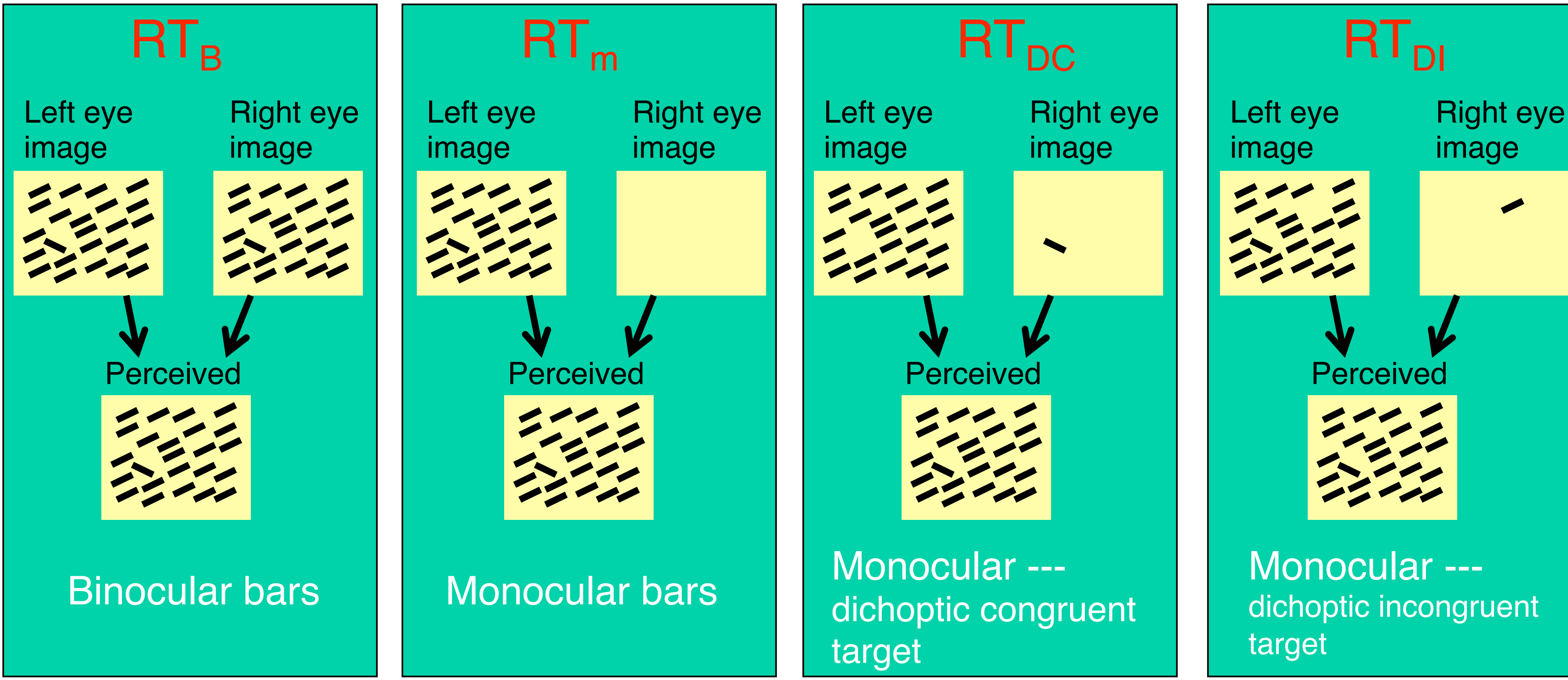
V1 theory prediction 3 --- ocular singleton pop out

(Zhaoping 2008, Journal of Vision)

Unique eye of origin

Another fingerprint of V1 since only V1 is the only cortical area with monocular cells and thus the eye origin information

Visual search for orientation singleton with various dichoptic designs



Task: --- report ASAP whether the orientation singleton is in the left or right half of the perceived image

Prediction: report reaction times $RT_{DI} > RT_m > RT_{DC}$,

V1 theory prediction 3 --- ocular singleton pop out

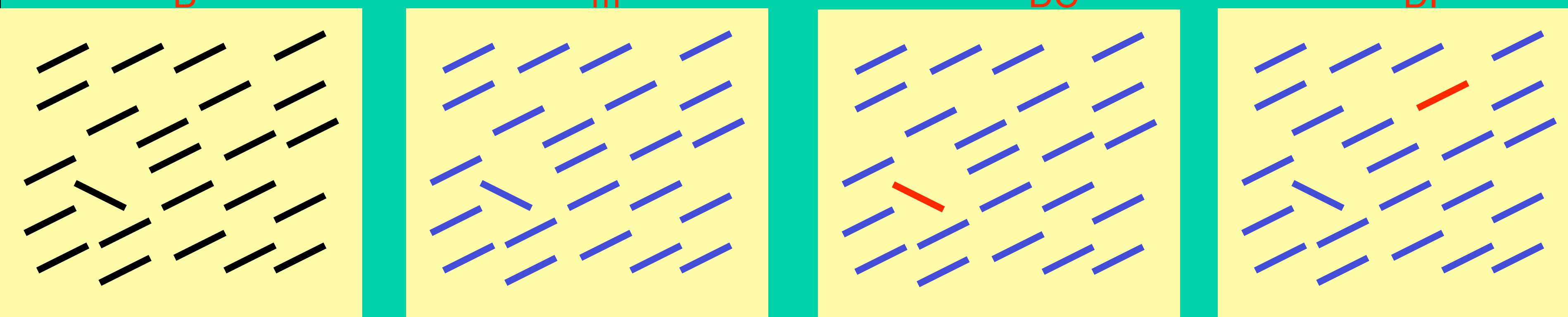
(Zhaoping 2008, Journal of Vision)

↑ Unique eye origin

Another fingerprint of V1 since only V1 is the only cortical area with monocular cells and thus the eye origin information

Visual search for orientation singleton with various dichoptic designs

RT_B | RT_m | RT_{DC} | RT_{DI}

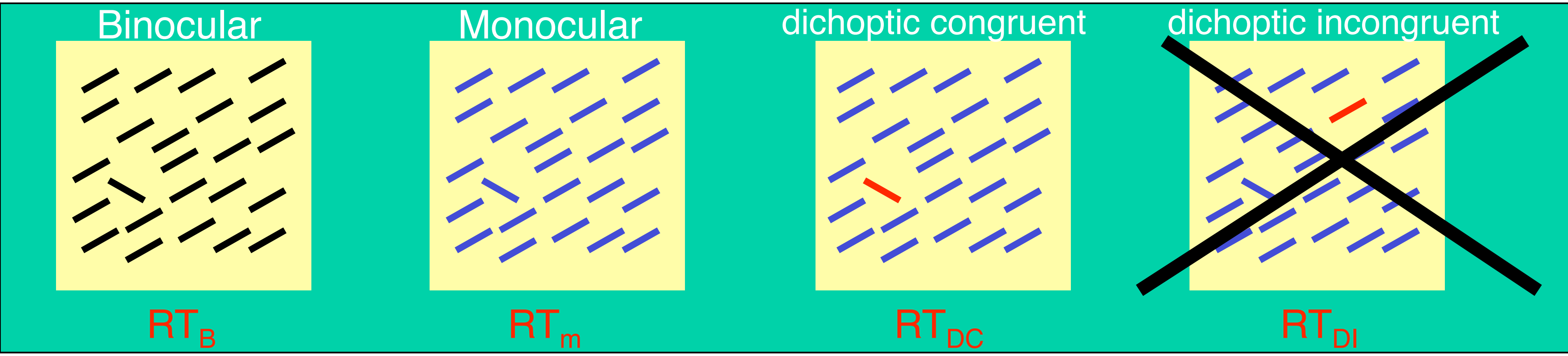


Binocular bars | Monocular bars | Monocular --- dichoptic congruent target | Monocular --- dichoptic incongruent target

For visualization, bar are color coded such that black, blue, and red bars denote bars presented binocularly, to left eye only, and to right eye only, respectively. The actual bars were not presented in color.

Prediction: report reaction times $RT_{DI} > RT_m > RT_{DC},$

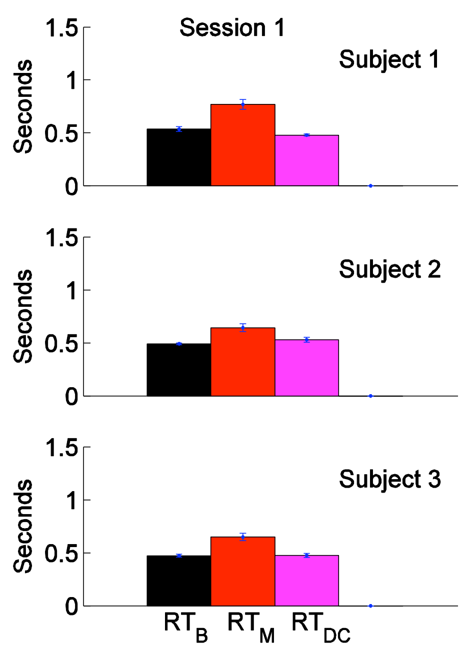
V1 theory prediction 3 --- ocular singleton or contrast pop out



Task: --- report ASAP whether the orientation singleton is in the left or right half of the display

Prediction: $RT_m > RT_{DC}$,

In Session 1:
only the first 3 conditions presented, randomly interleaved, subjects not informed about different presentation conditions, nor did they become aware of them.

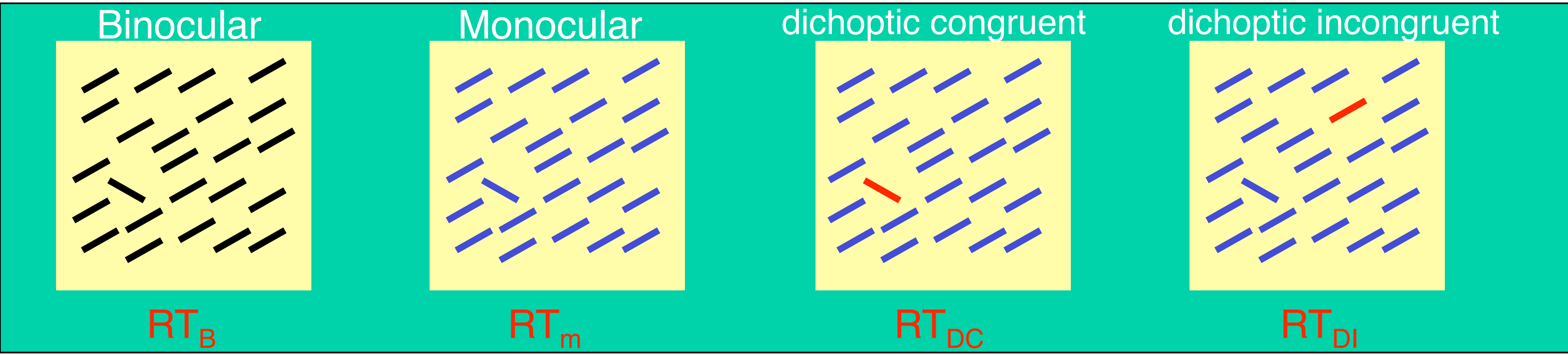


Results:
$RT_{DC} < RT_m$

Confirming the prediction

Zhaoping, SFN2007 Submitted

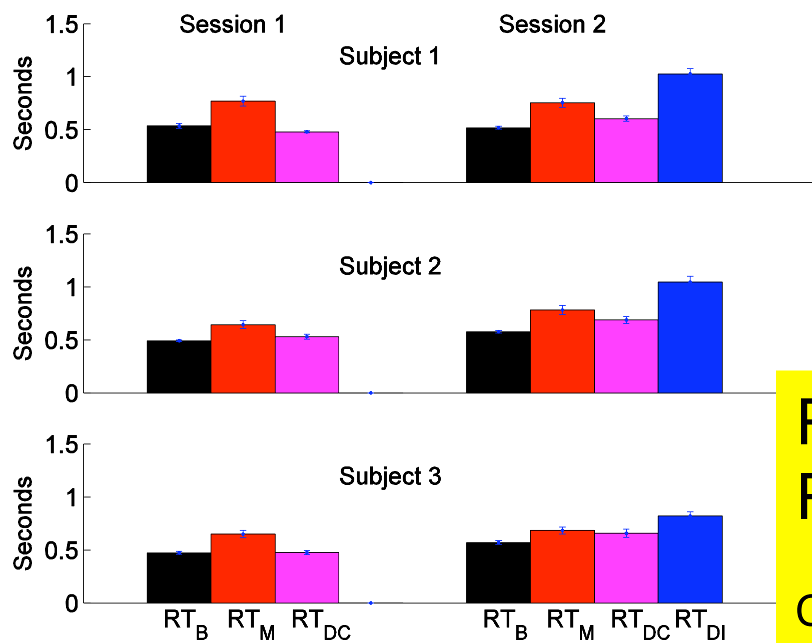
V1 theory prediction 3 --- ocular singleton or contrast pop out



Task: --- report ASAP whether the orientation singleton is in the left or right half of the display

Prediction: $RT_{DI} > RT_M$

In Session 2: All four conditions were randomly interleaved, subjects informed not to be distracted by any non-orientation singleton that might attract their attention.



Results:

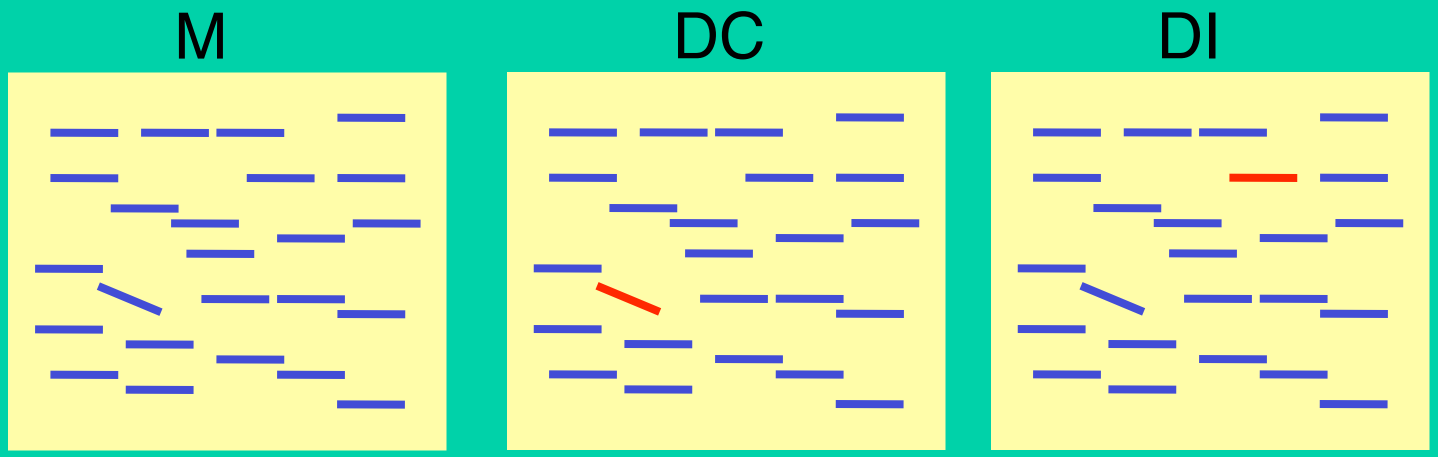
$RT_M < RT_{DI}$

Confirming the prediction

Zhaoping, 2008

Another experiment: when the search stimulus was masked after only 200 ms display, distractors are all horizontal, and subject had to identify the tilt direction of the orientation singleton target. Performance had lowest error in the DC condition, when the ocular singleton exogeneously cued the attention to target. This is so even when subjects could not answer by forced choice whether an ocular singleton existed in a trial (Zhaoping 2008) --- dissociation between awareness and attentional attraction

Visual search for orientation singleton with various dichoptic designs



Monocular bars

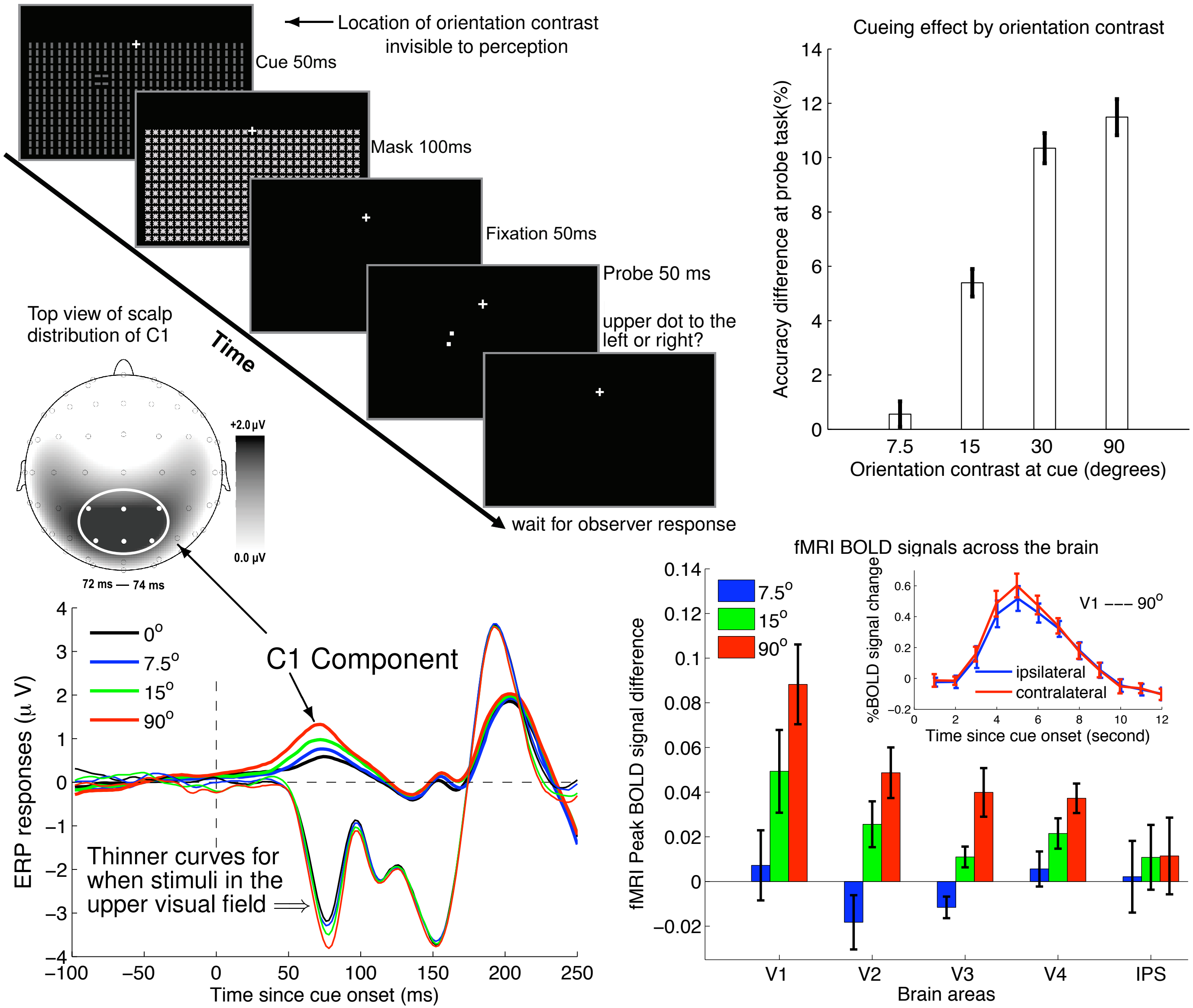
Dichoptic congruent target (DC)

Dichoptic incongruent (DI)

For visualization, bar are color coded such that black, blue, and red bars denote bars presented binocularly, to left eye only, and to right eye only, respectively. The actual bars were not presented in color.

Prediction: Error rate lowest in DC condition, confirmed

fMRI and ERP evidence of a saliency map in V1 (Zhang, Zhaoping, Zhou, and Fang, 2012)



We find brain substrates for saliency using stimuli that observers could not perceive (to minimize contributions from top-down factors and confound from awareness), but that nevertheless, through orientation contrast between foreground and background regions, attracted attention to improve a localized visual discrimination. When orientation contrast increased, so did the degree of attraction, and two physiological measures: the amplitude of the earliest (C1) component of the ERP, which is associated with V1, and fMRI BOLD signals in areas V1-V4 (but not the intra-parietal sulcus). Significantly, across observers, the degree of attraction correlated with the C1 amplitude and just the V1 BOLD signal.

Summary: A theory of a bottom up saliency map in V1

Tested by

(1) V1 outputs account for previous saliency data

(2) New behavioral data confirm the theory's predictions

The theory links physiology with behavior,
And challenges the previous views about the role of V1 and about the psychophysical saliency map.

Since top-down attention has to work with or against the bottom up saliency, V1 as the bottom up saliency map has important implications about top-down attentional mechanisms.

Note:

(1) This theory applies to cases when the effects of the top-down inputs to V1 are negligible and not dominant. These cases are, e.g., very immediately after changes in visual inputs or when prior knowledge/expectations of inputs are absent.

(2) Neural correlates of saliency signals in higher cortical areas (e.g., LIP) may be partly due to inputs from V1, plus other contributions such as top-down control and possibly (how much? an empirical question) additional bottom up contributions from beyond V1.

(3) This theory does not imply that cortical areas beyond V1 does not contribute additional bottom-up saliency signals. It is an empirical question to find out how much additional bottom-up saliency signals are contributed by areas beyond V1, including retina.

References:

Duncan J., Humphreys G.W. (1989) Visual search and stimulus similarity *Psychological Rev.* 96, 1-26.

Itti L., Koch C. (2000) A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Res.* 40(10-12):1489-506.

Koch C., Ullman S. (1985) Shifts in selective visual attention: towards the underlying neural circuitry. *Hum. Neurobiol.* 4(4): 219-27.

Li. Z. (1998) A neural model of contour integration in the primary visual cortex. *Neural Computation 10(4):903-940.*

Li Z. (1999) Visual segmentation by contextual influences via intracortical interactions in primary visual cortex. In *Network: Computation in Neural Systems* Volumn 10, Number 2, May 1999. Page 187-212.

Li Z. (2001) Computational design and nonlinear dynamics of a recurrent network model of the primary visual cortex. *Neural Computation* 13/8, p1749-1780.

Li Z. (2002) A saliency map in primary visual cortex, Published in *Trends in Cognitive Sciences Vol 6. No.1.page9-16*

Treisman A. M., Gelade G. (1980) A feature-integration theory of attention. *Cognit Psychol.* 12(1), 97-136.

Wolfe J.M., Cave K.R., Franzel S. L. (1989) Guided search: an alternative to the feature integration model for visual search. *J. Experimental Psychol.* 15, 419-433.

Wolfe J.M. (1998) Visual Search, a review. in *Attention* p. 13-74. H. Pashler (Editor), Hove, East Sussex, UK, Psychology, Press. Ltd.

Zhaoping L. and May K.A. (2007), Psychophysical tests of the hypothesis of a bottom-up saliency map in primary visual cortex, *Public Library of Science, Computational Biology*. 3(4):e62. doi:10.1371/journal.pcbi.0030062

Koene AR and Zhaoping L. (2007) Feature-specific interactions in salience from combined feature contrasts: Evidence for a bottom-up saliency map in V1. , Journal of Vision, 7(7):6, 1-14, http://journalofvision.org/7/7/6/, doi:10.1167/7.7.6

Zhaoping L. (2006) Theoretical understanding of the early visual processes by data compress and data selection, in *Network: Computation in neural systems* 17(4):301-334.

Zhaoping L. (2003) V1 mechanisms and some figure-ground and border effects, In Journal of Physiology Paris, 97(4-6): 503-515.

Zhaoping L (2008) Attention capture by eye of origin singletons even without awareness --- a hallmark of a bottom-up saliency map in the primary Visual cortex. *Journal of Vision, 8(5):1, 1-18, http://journalofvision.org/8/5/1/*

Zhang, Zhaoping, Zhou, Fang (2012) Neural activities in V1 create a bottom-up saliency map. NEURON, 73: 183-192

References by Li or Zhaoping (same person with different publication names in different time periods) can be downloaded from www.cs.ucl.ac.uk/staff/Zhaoping.Li/