A saliency map in primary visual cortex

Li Zhaoping
University College London
www.gatsby.ucl.ac.uk/~zhaoping

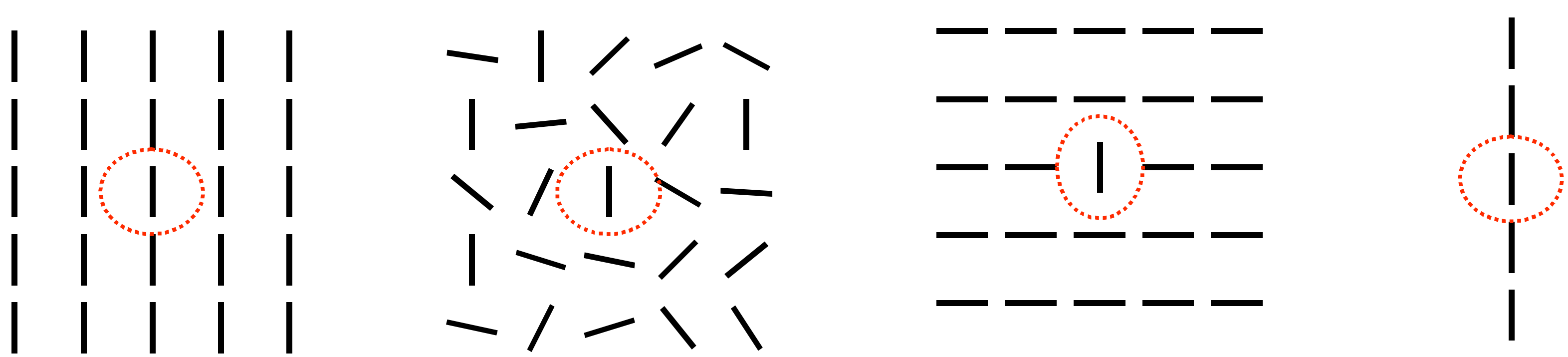
"A saliency map in primary visual cortex", in
Trends in Cognitive Sciences Vol 6, No. 1,
page 9-16, 2002,

What is in V1?

Classical receptive fields: --- bar and edge detectors or filters, too small for global visual tasks.

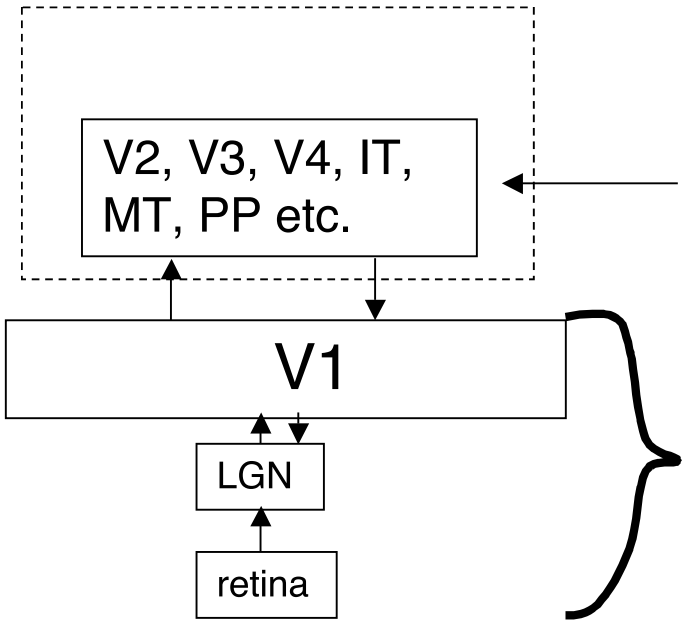
Contextual influences, and their confusing role:

Strong suppression | suppression | Weak suppression | facilitation



Horizontal intra-cortical connections observed as neural substrates

Where is V1 in visual processing?



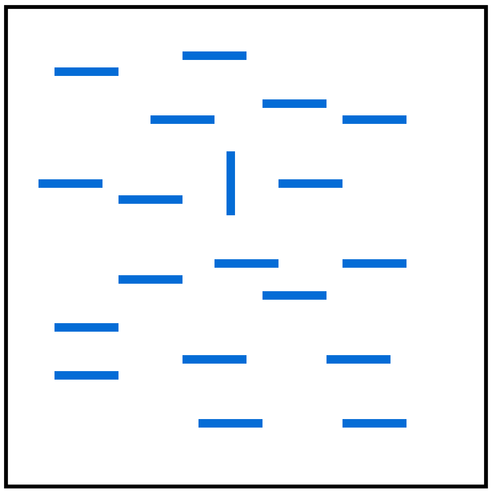
Larger receptive fields, much affected by visual attention.

Small receptive fields, limited effects from visual attention

V1 had traditionally been viewed as a lower visual area, as contributing little to complex visual phenomena.

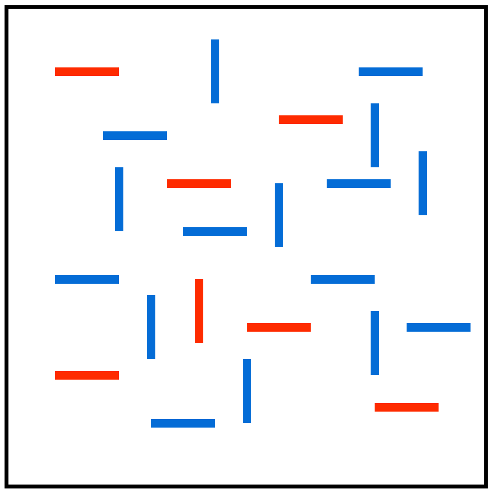
But, V1 is the largest visual area in the brain --- a hint towards my proposed role for V1.

Feature search



Fast, parallel, pre-attentive, effortless, pops out

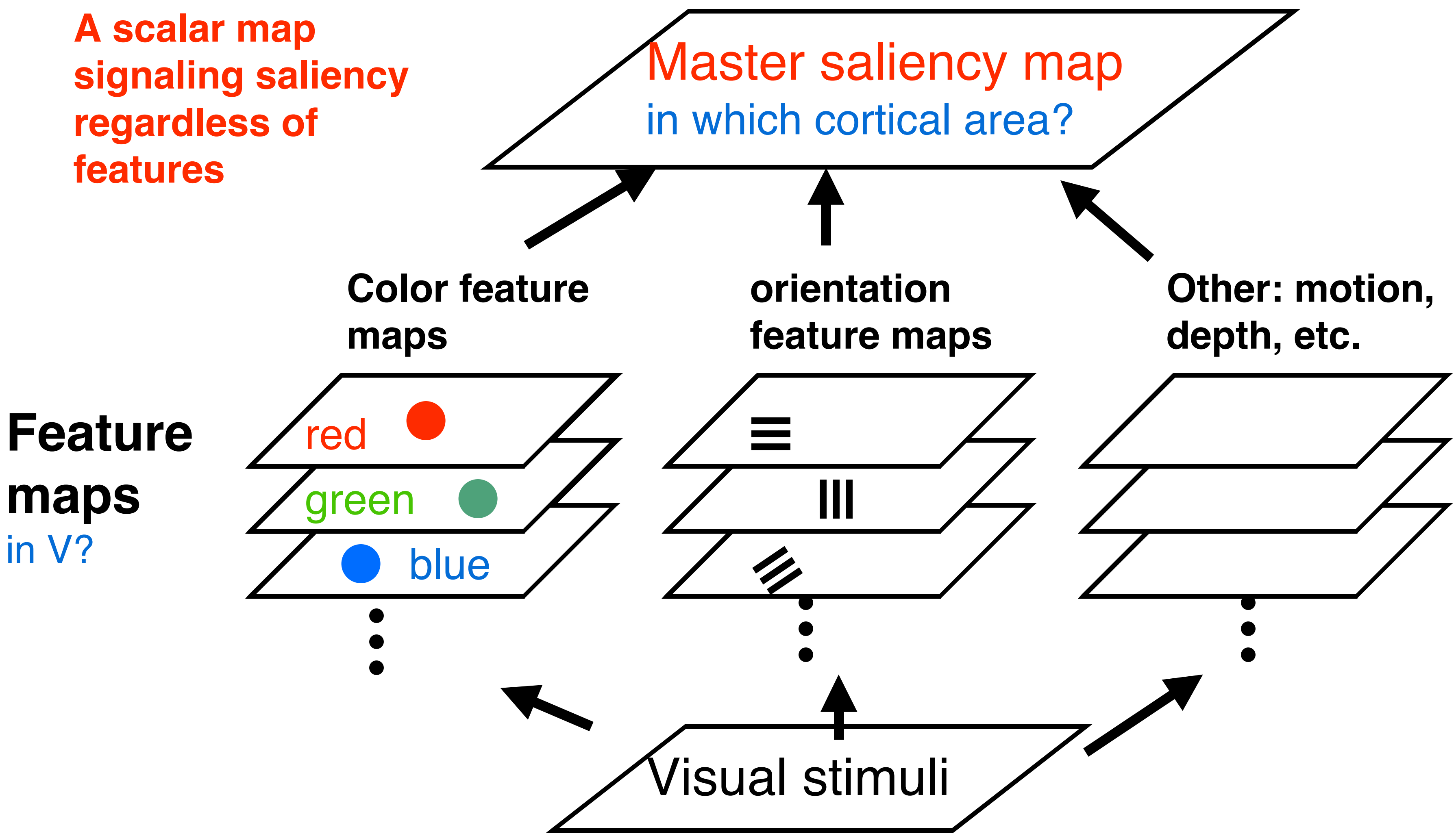
Conjunction search



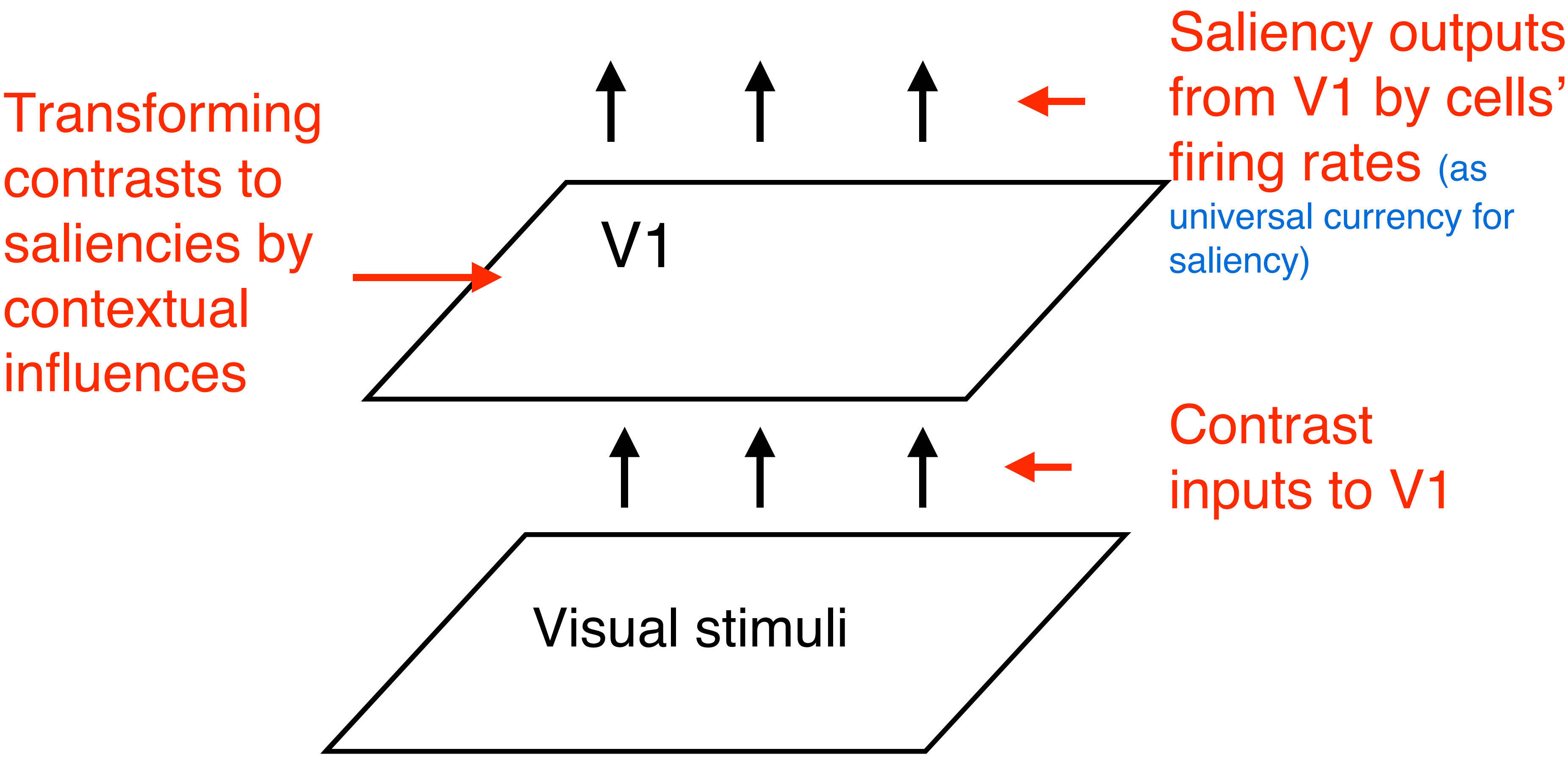
Slow, serial, effortful, needs attention, does not pop out

A saliency map serves to select stimuli for further processing

Previous views of saliency map (Treisman, Koch, Ullman, Itti, Wolfe etc)



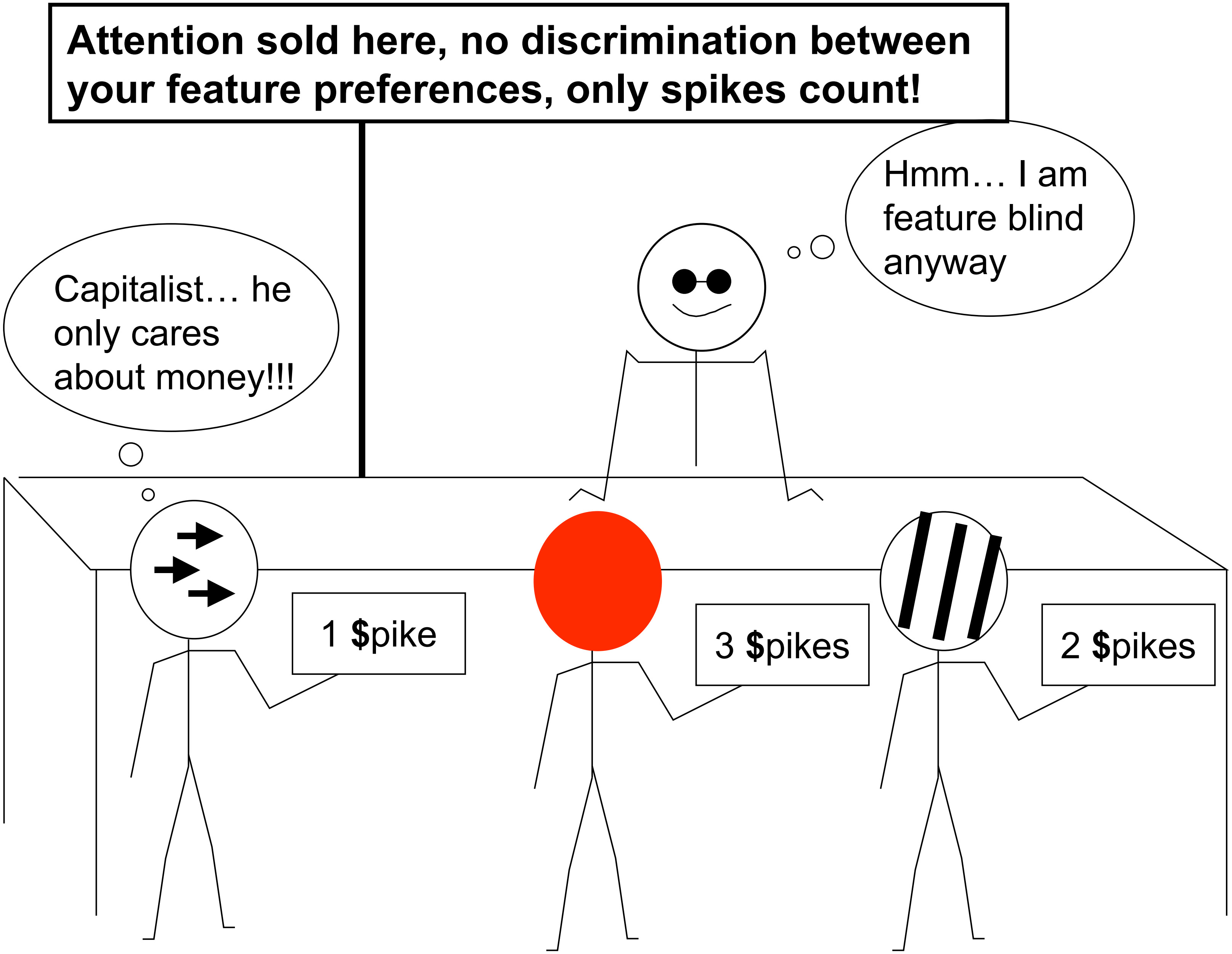
My proposal of V1 that produces a saliency map



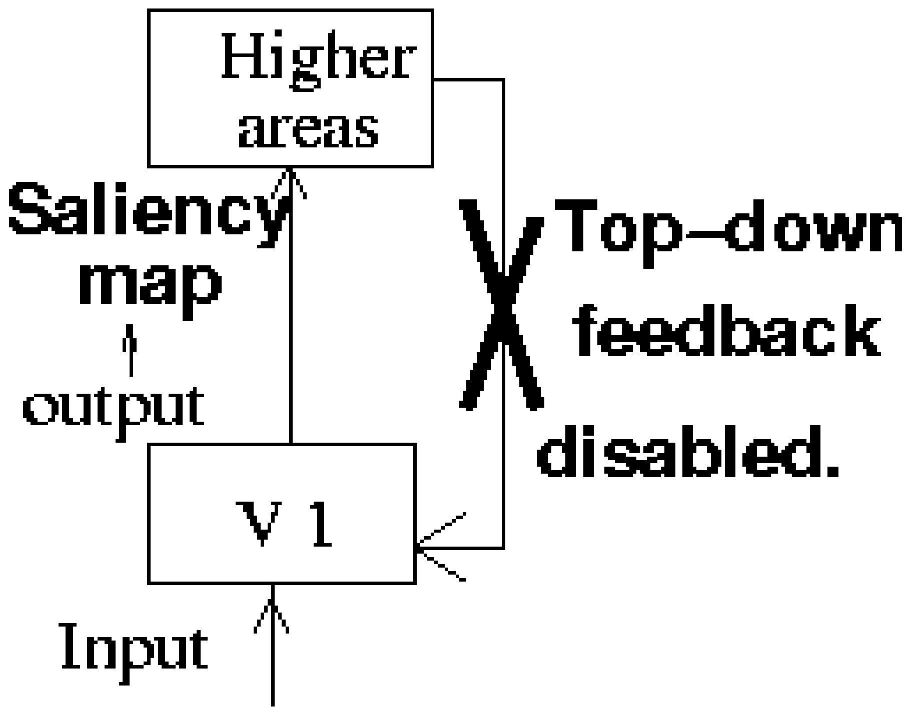
No separate feature maps, nor any combination of them

V1 cells' firing rates signal saliencies, despite their feature tuning

Strongest response to any visual location signals its saliency



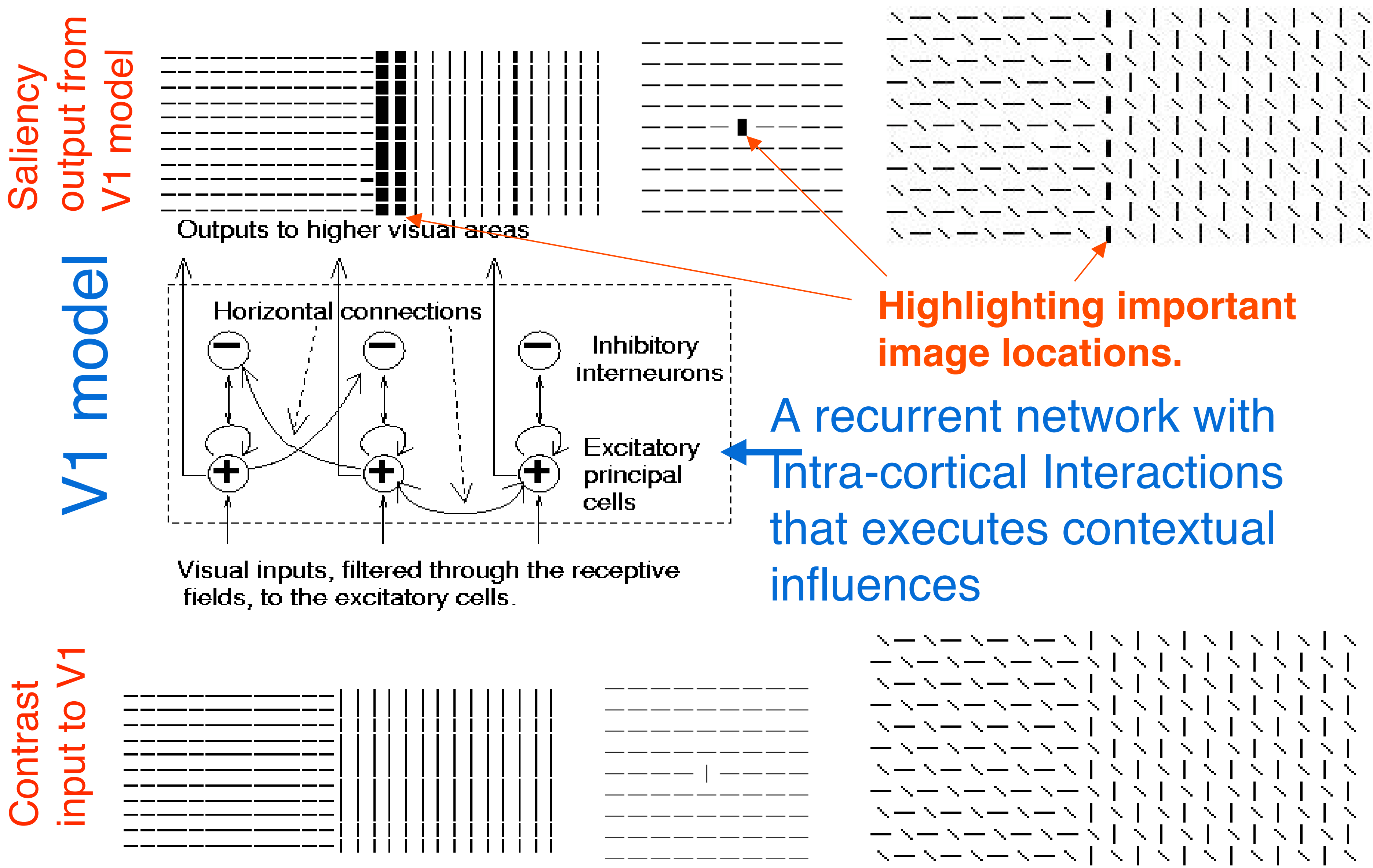
Attention sold here, no discrimination between your feature preferences, only spikes count!
Hmm… I am feature blind anyway
Capitalist… he only cares about money!!!
1 $pike
3 $pikes
2 $pikes



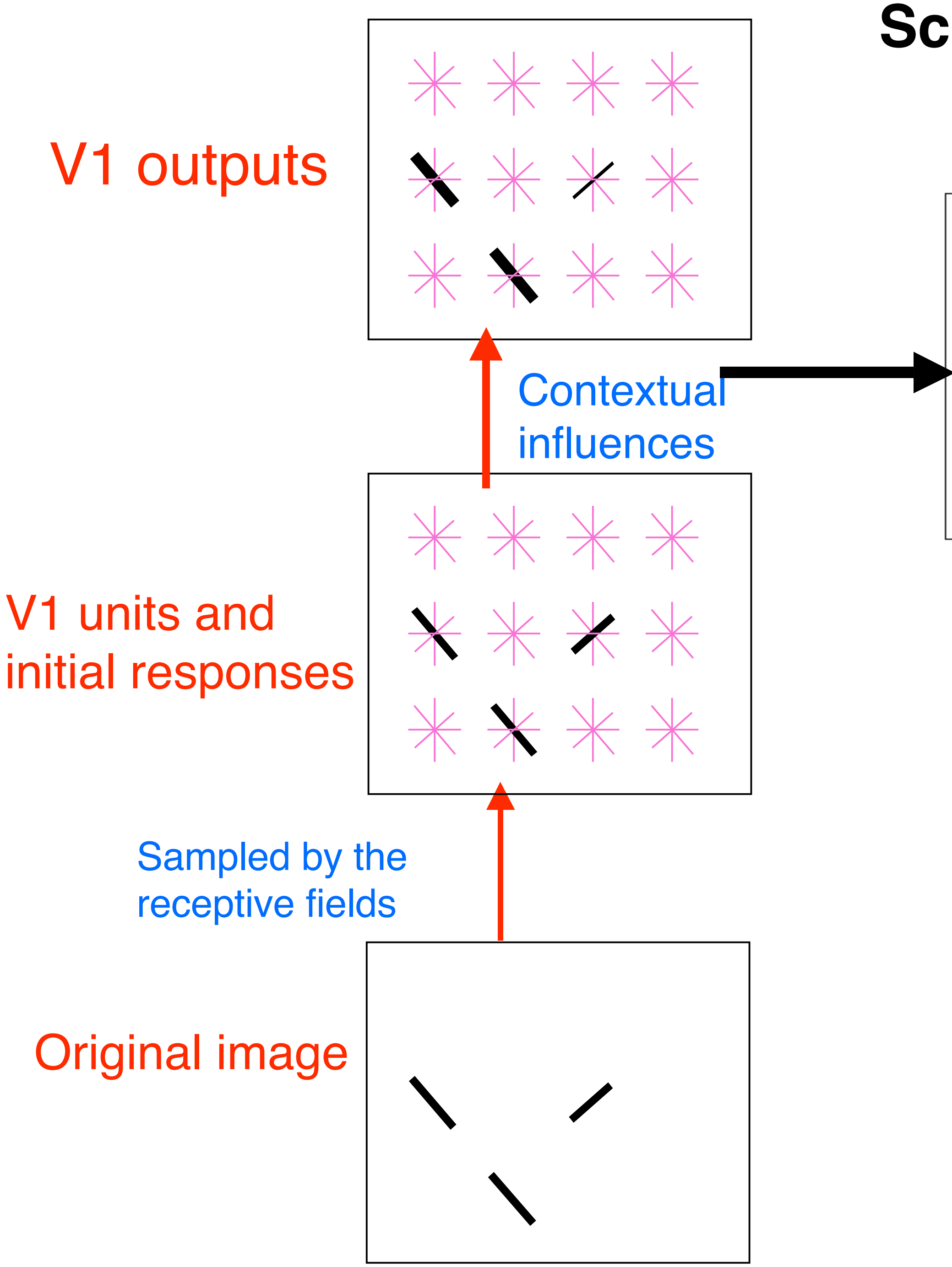
Saliency from bottom-up factors only.

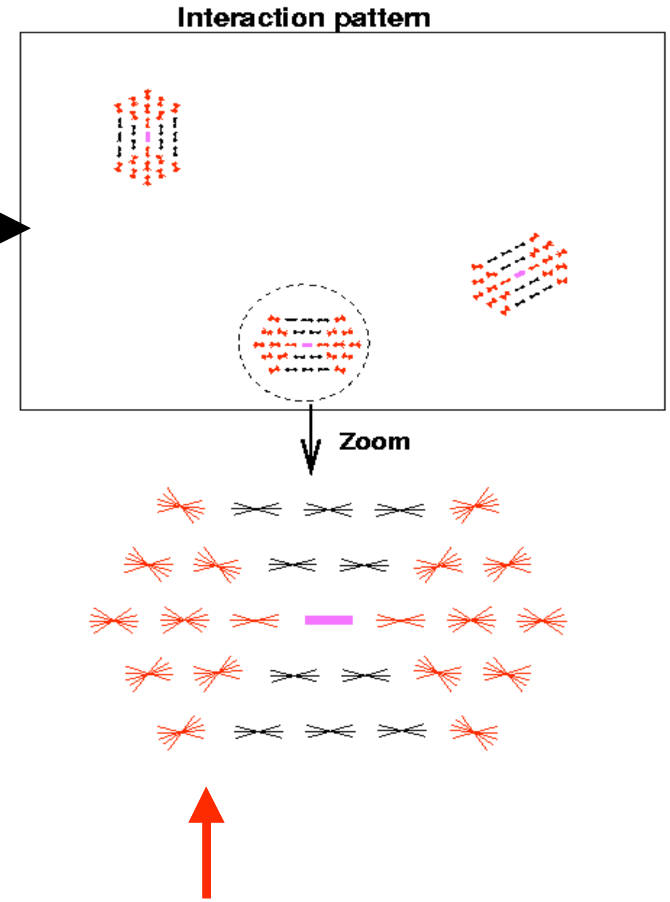
V1’s output as saliency map is viewed under the idealization of the top-down feedback to V1 being disabled, e.g., shortly after visual exposure or under anesthesia.

Implementing the saliency map in a V1 model



Schematics of how the model works



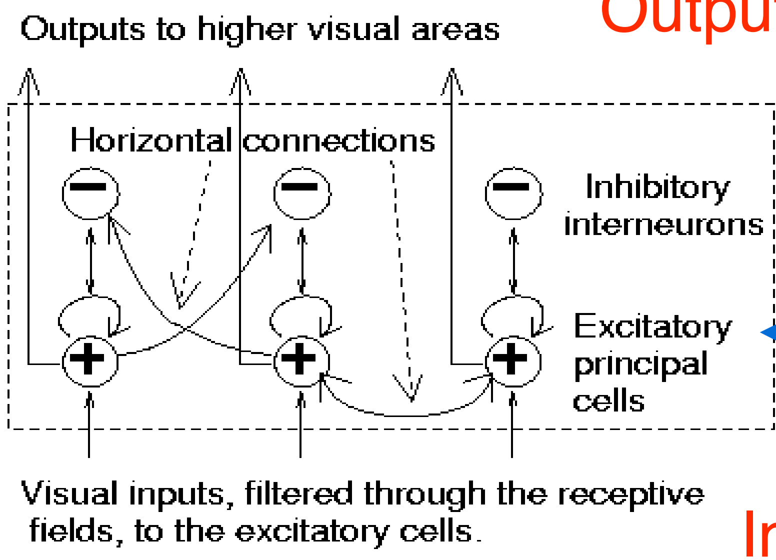


Designed such that the model agrees with physiology and produces desired behavior.

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

V1 model

Output $g_x(x)$



Intra-cortical Interactions

Input I

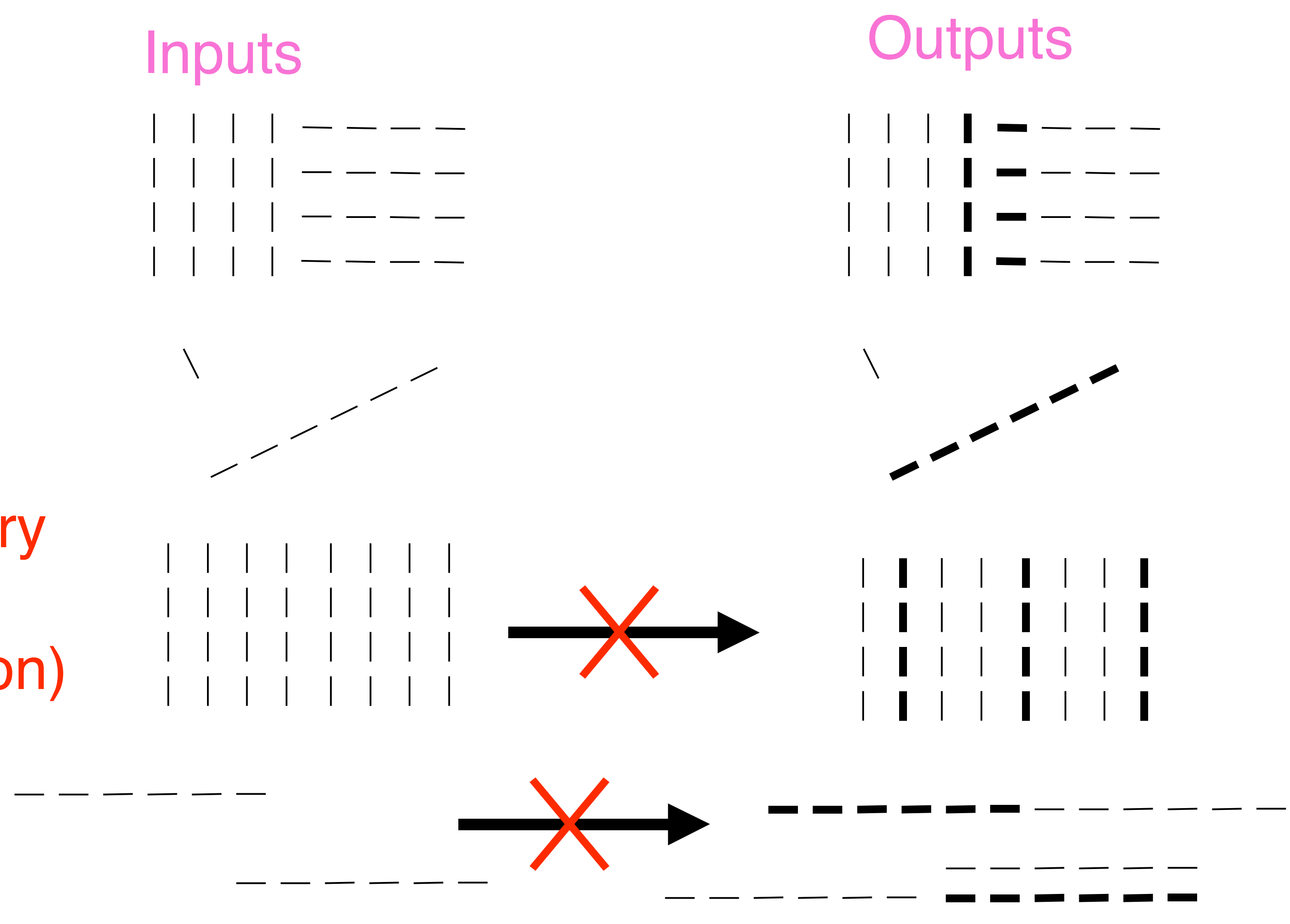
$$dx_i/dt = -x_i - g_y(y_i) + J_o\, g_x(x_i) + \sum_j J_{ij}\, g_x(x_j) + I_i$$

$$dy_i/dt = -y_i + \sum_j W_{ij}\, g_x(x_j) + I_c$$

The behavior of the network is ensured by computationally designing the recurrent connection weights, using dynamic system theory.

Conditions on the intra-cortical interactions.

Zhaoping Li (2001) Computational design and nonlinear dynamics of a recurrent network model of the primary visual cortex *Neural Computation 13/8, p. 1749-1780*



Highlight boundary

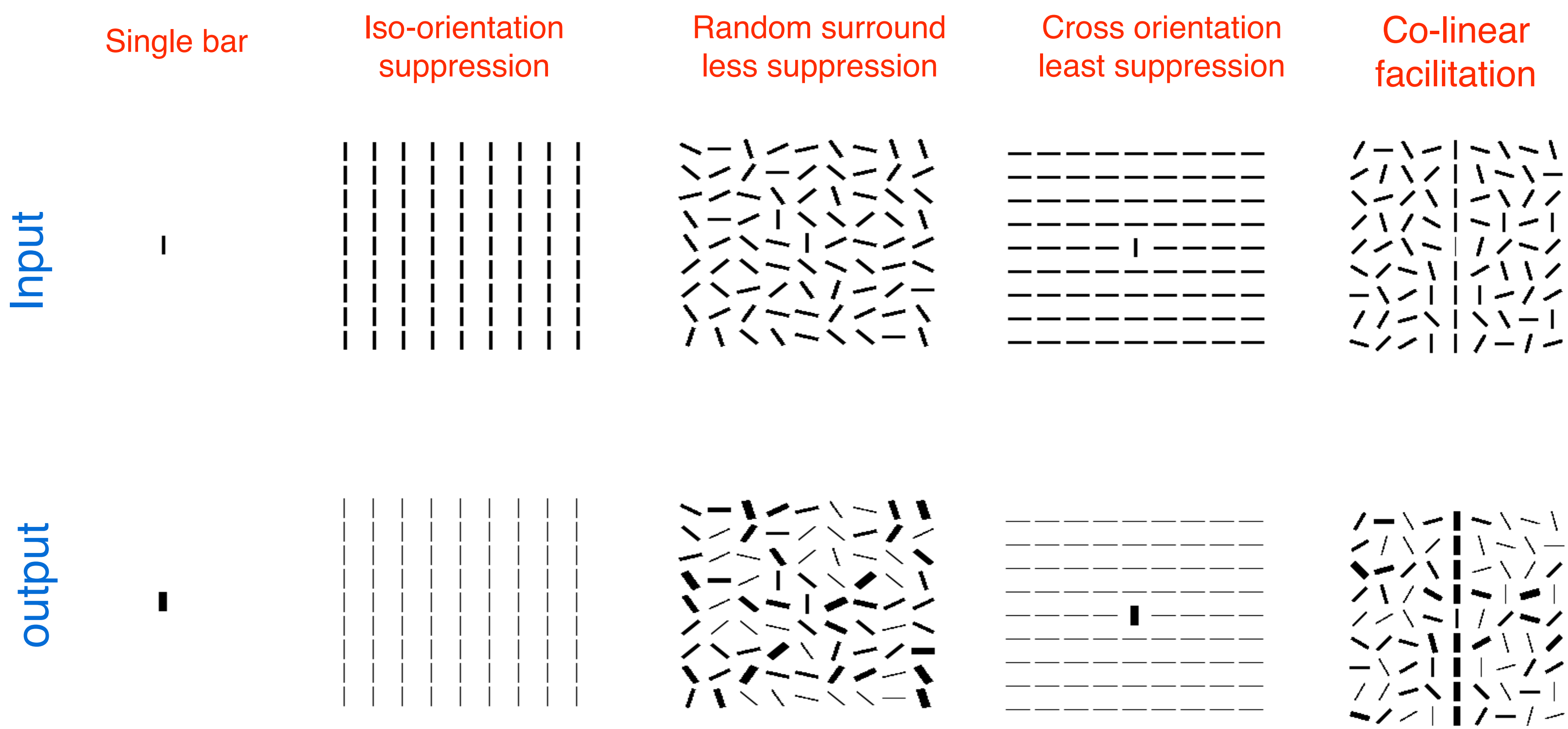
Enhance contour

No symmetry breaking (hallucination)

No gross extension

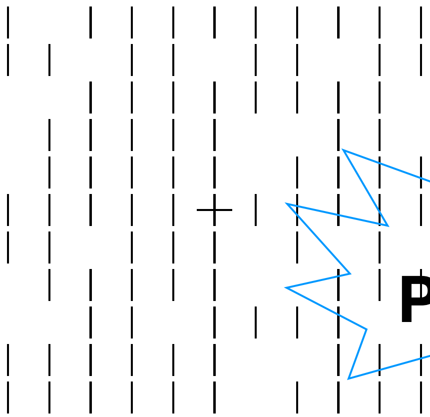
Design techniques: mean field analysis, stability analysis. Computation desires constraint the network architecture, connections, and dynamics. Network oscillation is one of the dynamic consequences.

Make sure that the model can produce the usual contextual influences



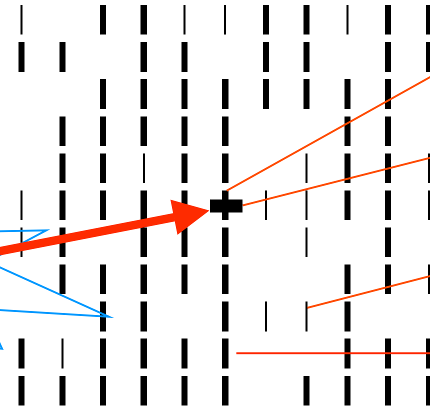
Proposal: V1 produces a saliency map

Original input



Pop-out

V1 response S



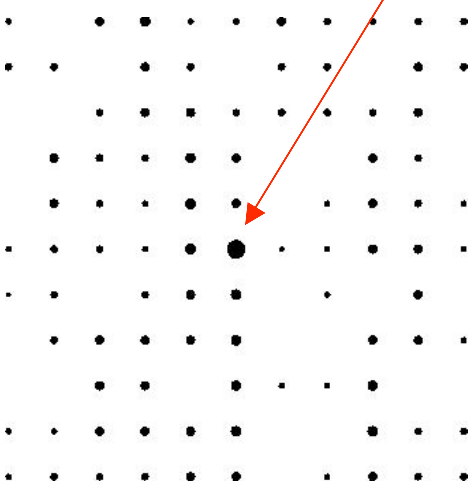
S=0.2, Z=1.0

S=0.4, **Z=7**

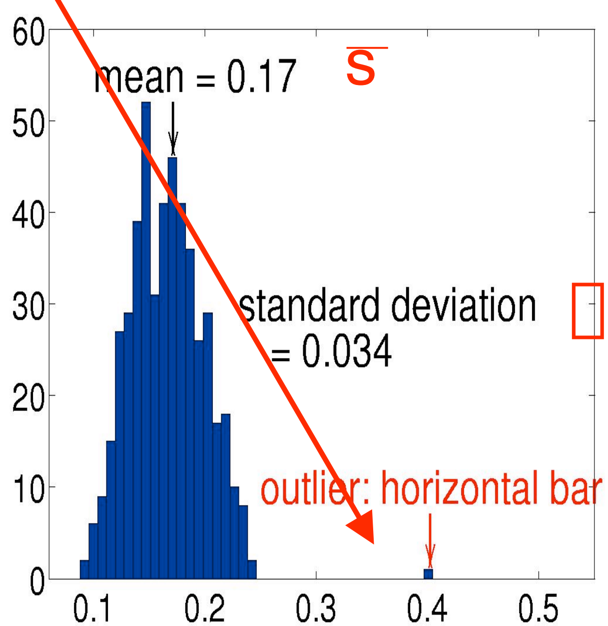
S=0.12, Z=-1.3

S=0.22 Z=1.7

Saliency map



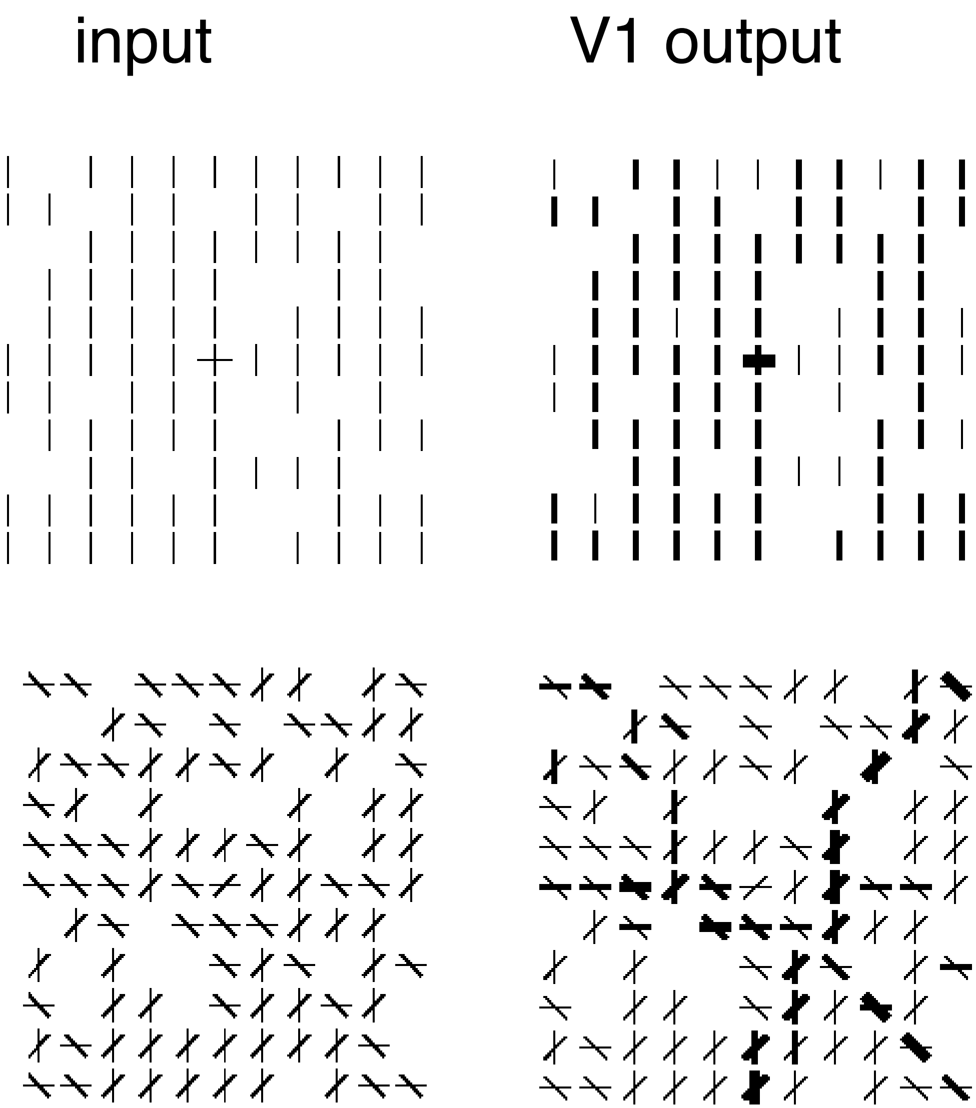
Histogram of all responses S regardless of features



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

--- z score, measuring saliencies of items

The V1 saliency map agrees with visual search behavior.



Target = +

Z=7

Feature search --- pop out

Target=

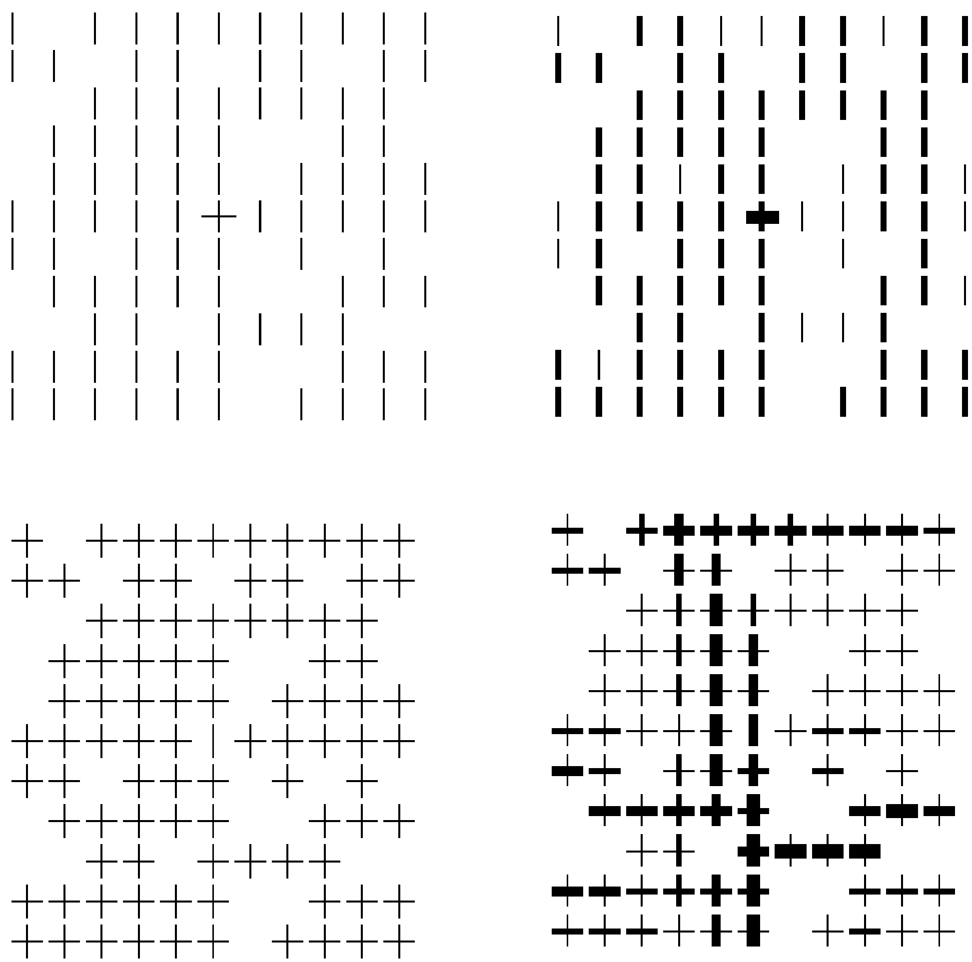
Z= - 0.9

Conjunction search --- serial search

A trivial example of search asymmetry

input

V1 output



Target = +

Z=7

Feature search --- pop out

Target = |

Z=0.8

Target lacking a feature

What defines a basic feature?

Psychophysical definition: enables pop-out ⟷ basic feature

Computational or mechanistic definition: two neural components or substrates required for basic features:

(1) Tuning of the cell receptive field to the feature

(2) Tuning of the horizontal connections to the feature --- the horizontal connections are selective to that optimal feature, e.g.,orientation, of the pre- and postsynaptic cells.

new

There should be a continuum from pop-out to serial searches

The ease of search is measured by a graded number : z score

Treisman's original Feature Integration Theory may be seen as the discrete idealization of the search process.

Influence of the background homogeneities

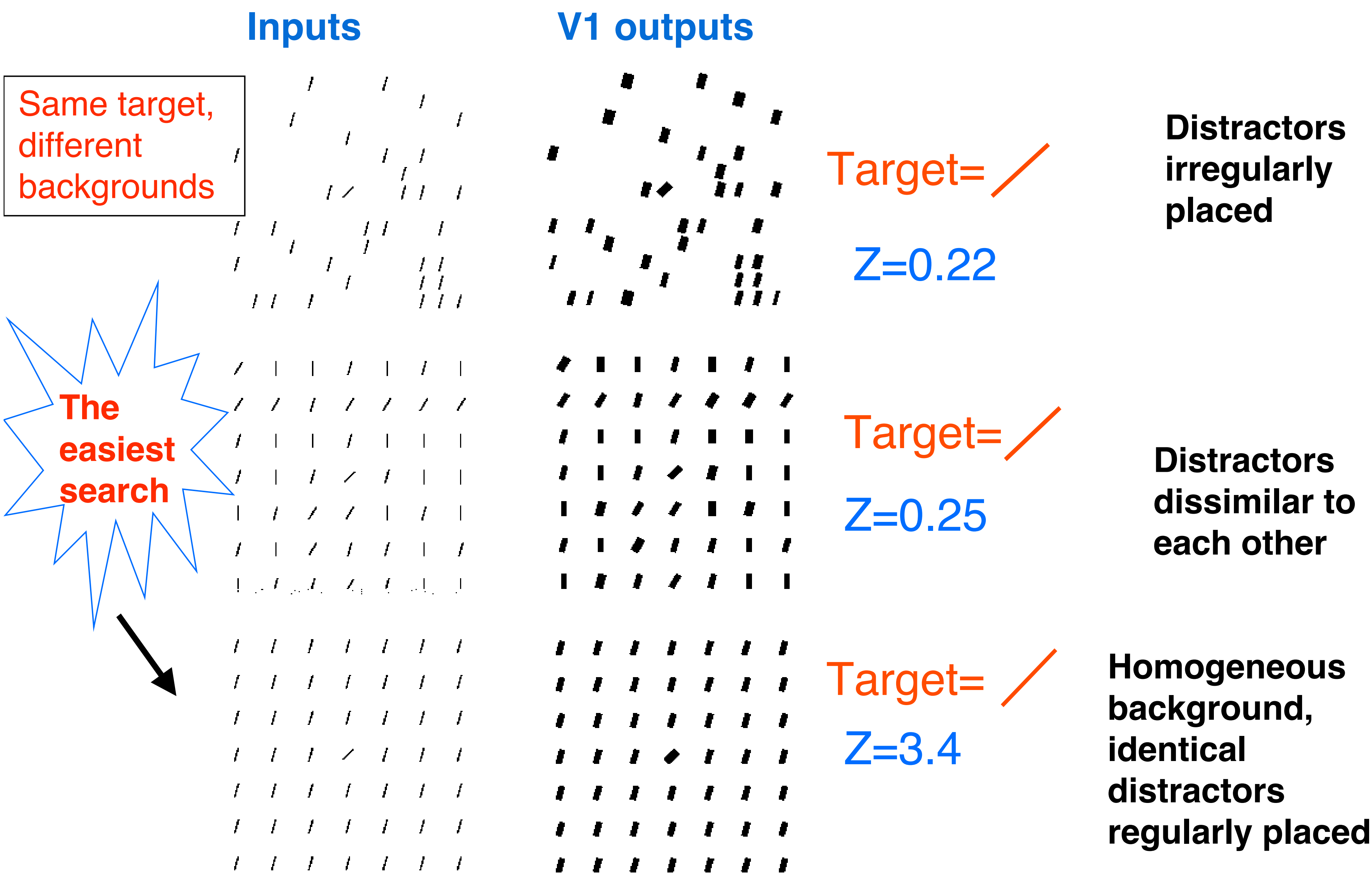
(cf. Duncan & Humphreys, and Rubinstein & Sagi.)

Saliency measure: $Z = (S - \bar{S})/$_

_ increases with the background in-homogeneity.

Hence, homogeneous background makes target more salient.

Explains spatial configuration and distractor effects.



Another example of background regularity effect

Input

Output

Target=

Z=-0.63, next to target, z=0.68

Distractors irregularly placed

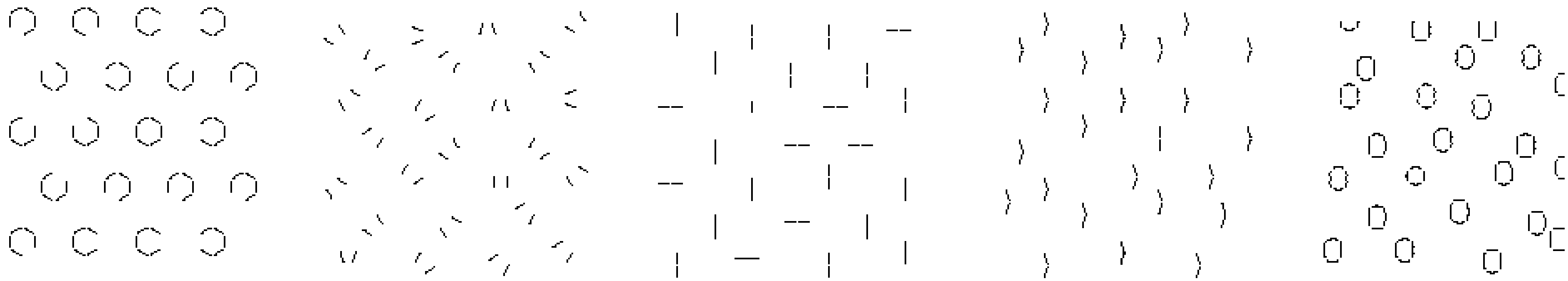
Target=

Z=-0.83, next to target, z=3.7

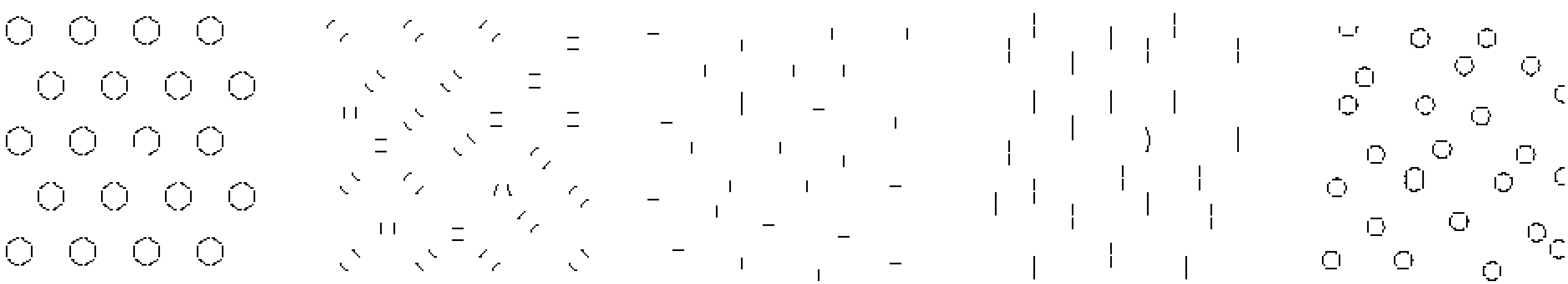
Homogeneous background, identical distractors regularly placed

More severe test of the saliency map theory by using subtler saliency phenomena --- **search asymmetries** (Treisman)

Open vs. closed | parallel vs. divergent | long vs. short | curved vs. straight | elipse vs. circle

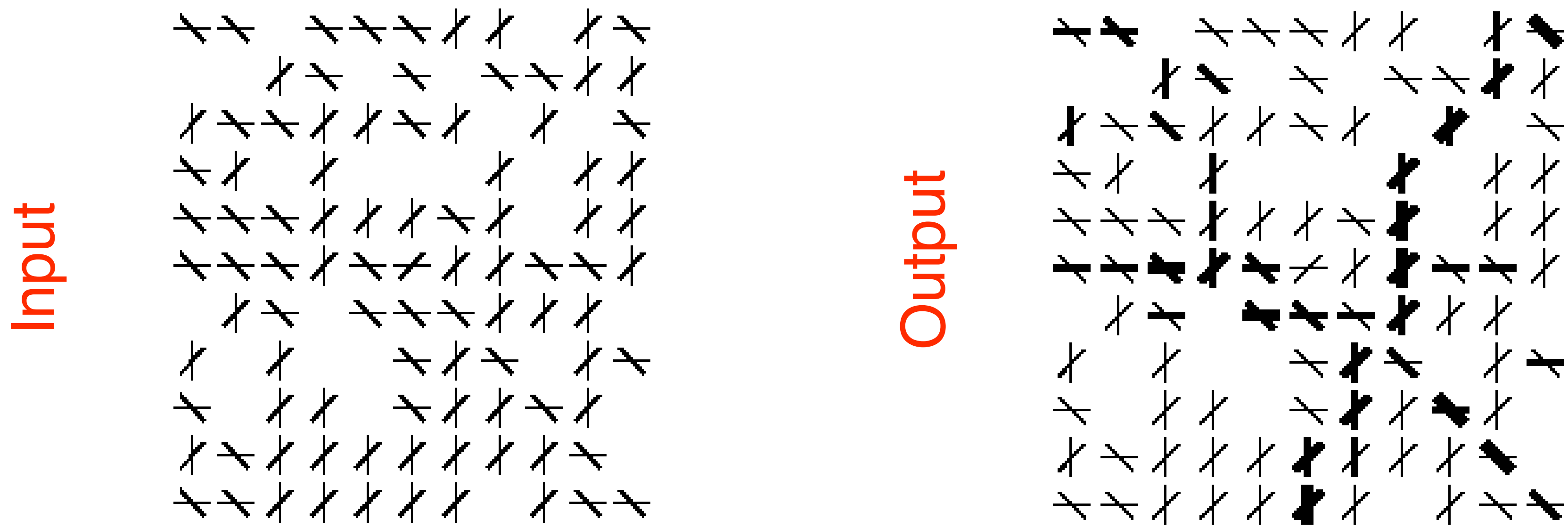


Z=0.41 | **Z= -1.4** | **Z= -0.06** | **Z= 0.3** | **Z= 0.7**



Z=9.7 | **Z= 1.8** | **Z= 1.07** | **Z= 1.12** | **Z= 2.8**

Conjunction search revisited



Some conjunction searches are easy

e.g.: *Conjunctions of motion and form* (orientation) --- McLeod, Driver, Crisp 1988)

e.g., *Conjunctions of depth and motion or color* --- Nakayama and Silverman 1986.

Why?

Recall the two neural components necessary for a basic feature

(1) Tuning of the receptive field (CRF)

(2) Tuning of the horizontal connections

For a conjunction to be basic and pop-out:

(1) Simultaneous or conjunctive tunings of the V1 cells to both feature dimensions (e.g., orientation & motion, orientation and depth, but not orientation and orientation)

(2) Simultaneous or conjunctive tunings of the horizontal connections to the optimal features in both feature dimensions of the pre- and post- synaptic cells

Predicting from psychophysics to V1 anatomy

Since conjunctions of motion and orientation, and depth and motion or color, pop-out

The horizontal connections must be selective simultaneously to both orientation & motion, and to both depth and motion (or color) --- can be tested

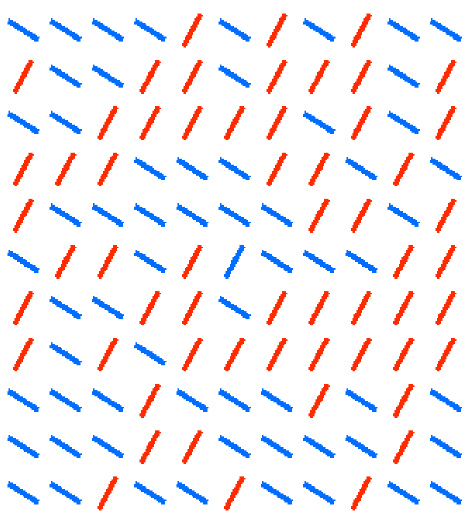
Note that it is already know that V1 cells can be simultaneously tuned to orientation, motion direction, depth (and even color sometimes)

Color-orientation conjunction?

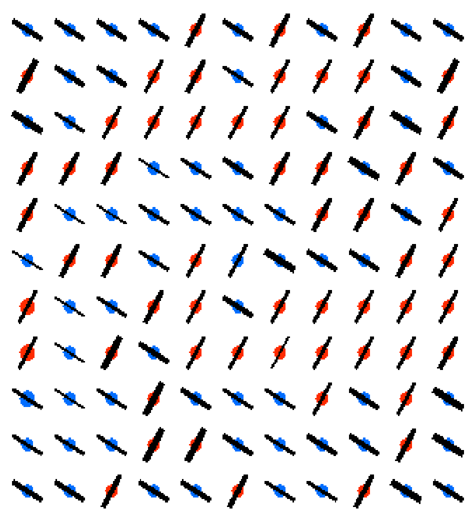
Prediction: **Color-orientation conjunction search can be made easier by adjusting the scale and/or density of the stimuli,**

since V1 cells conjunctively tuned to both orientation and color are mainly tuned to a specific spatial frequency band.

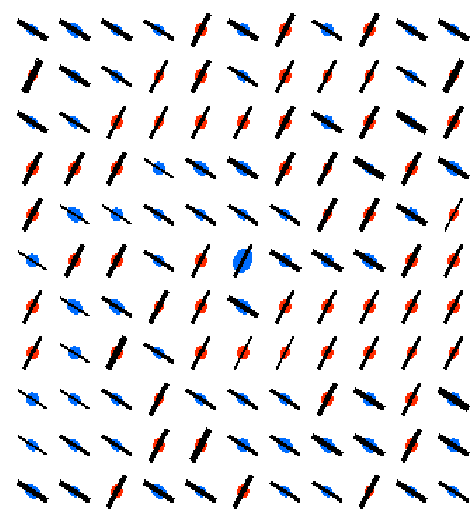
Stimuli for a conjunction search for target



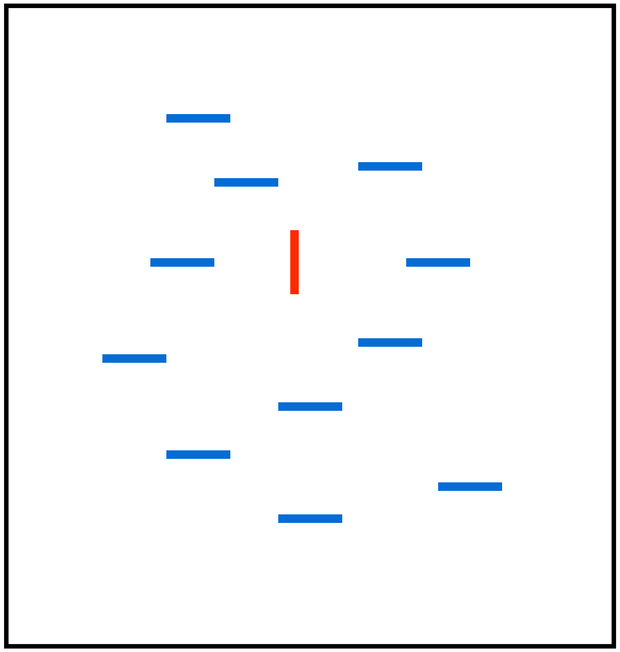
Response from a model without conjunction cells



Response from a model with conjunction cells



Double feature search --- opposite of conjunction search



Responses to target from 3 cell types:

(1) orientation tuned cells tuned to vertical

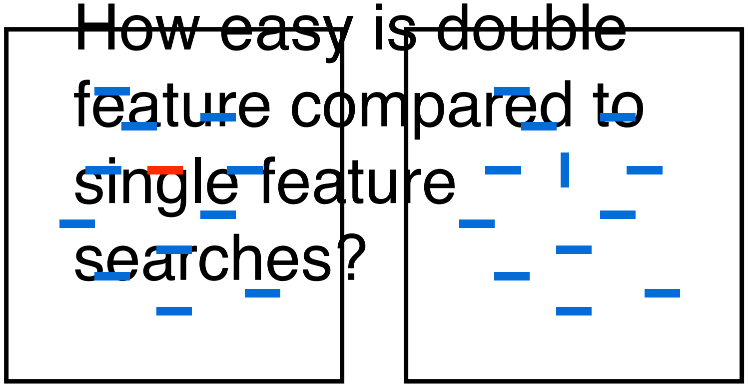
(2) color tuned cells tuned to red

(3) conjunctively tuned cells tuned to red-vertical

The most responsive of them should signal the target saliency.

Let (1) and (2) determine eases in single searches. Existence of (3) makes double feature search possibly easier.

Single feature searches

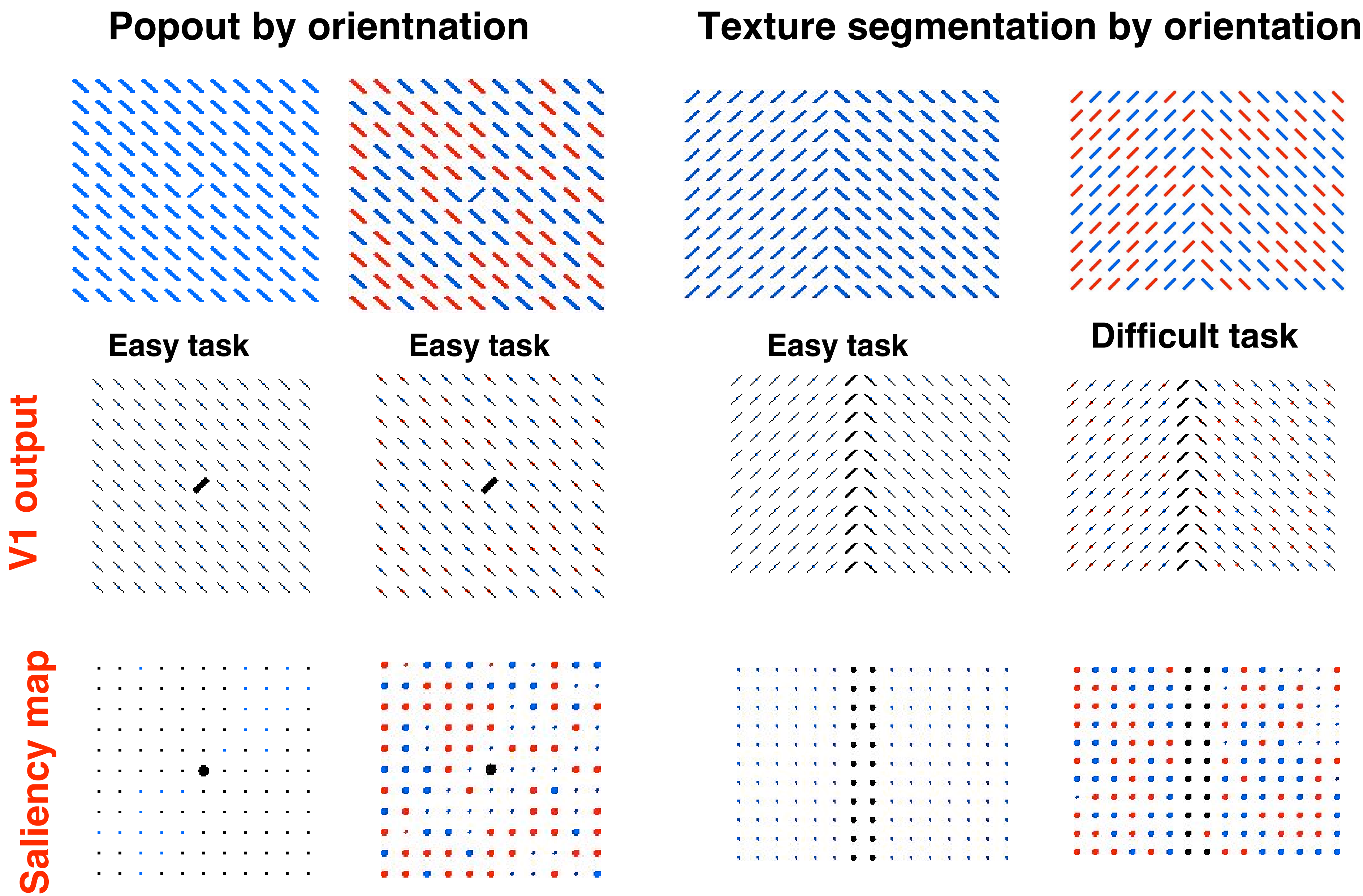


How easy is double feature compared to single feature searches?

Explains Nothdurft (2000) data: orientation-motion double feature search is easier than orientation-color double feature search.

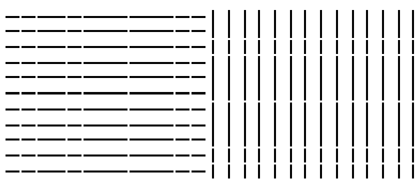
Interference from irrelevant feature dimensions

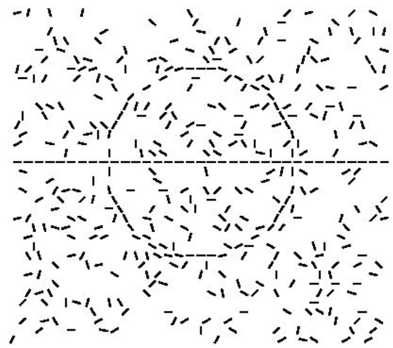
--- Rob Snowden’s data 1998

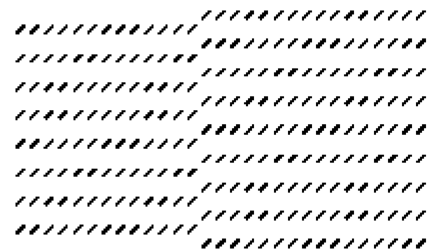


V1's saliency computation on other visual stimuli

model input

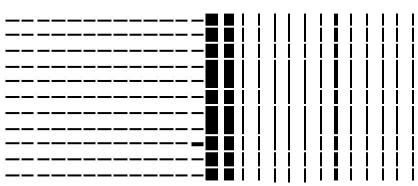


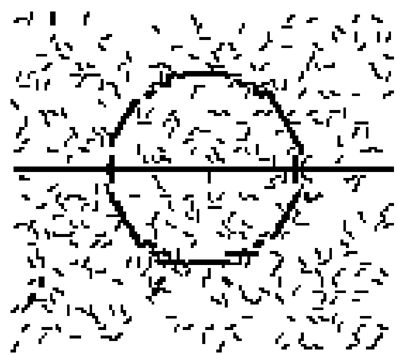


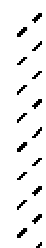


model output

Output highlights







Prediction: bias in the perceptual estimation of the location of the texture boundary (tests by Ariella Popple).

Summary:

Theory: V1--- saliency map
for pre-attentive segmentation.

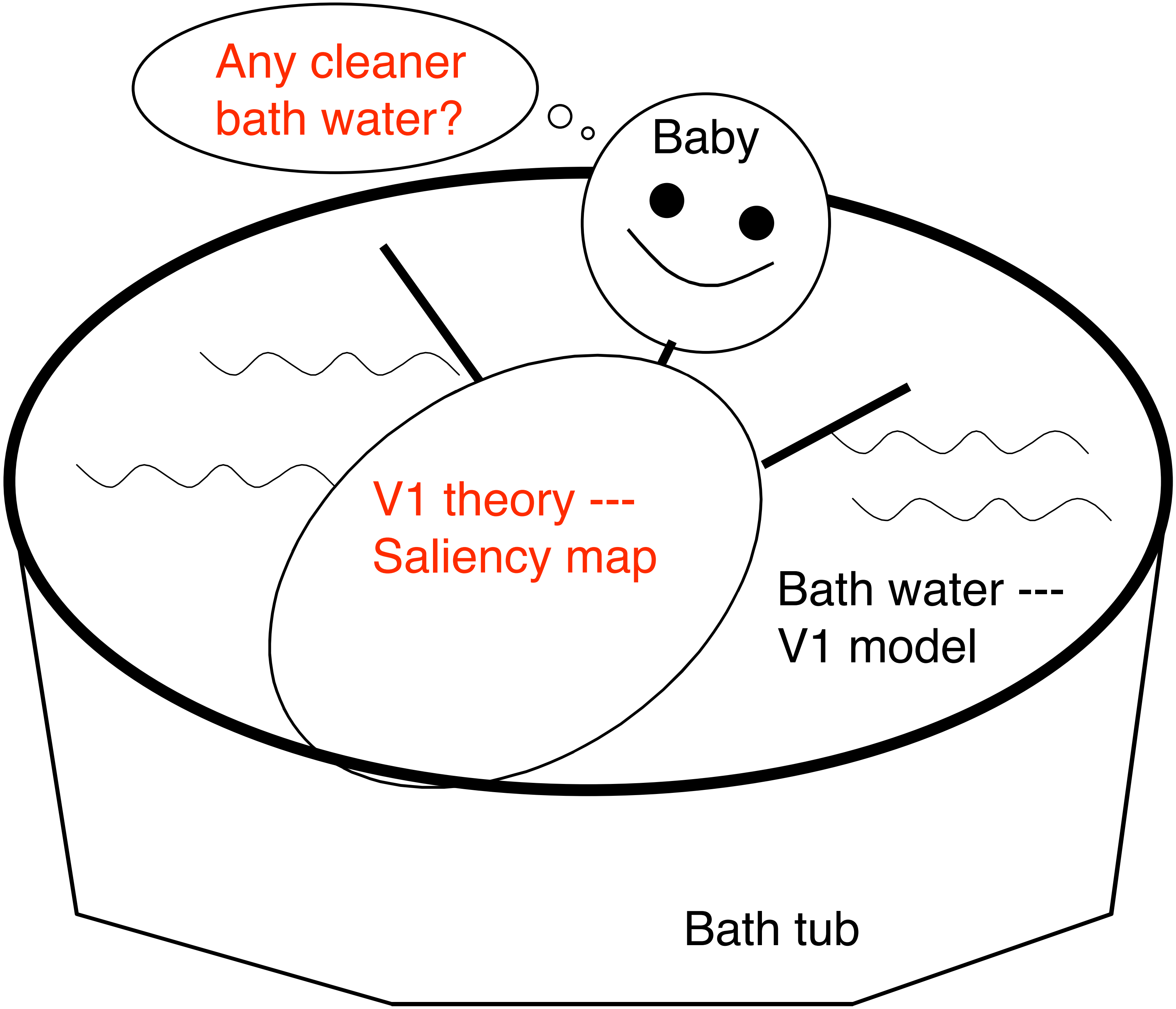
Linking physiology with psychophysics.

Theory “tested” or demonstrated on an imtation V1 (model) ---Recurrent network model: from local receptive fields to global behaviour for visual tasks.

Testable predictions, some confirmed, others to be tested.

“A saliency map in primary visual cortex” by Zhaoping Li, published in *Trends in Cognitive Sciences* Vol 6, No. 1, page 9-16, 2002,

see http://www.gatsby.ucl.ac.uk/~zhaoping/ for more information.


Any cleaner
bath water?
Baby
V1 theory ---
Saliency map
Bath water ---
V1 model
Bath tub

Theory - From hypothesis to predictions

Computational role of V1

Pre-attentive segmentation,

segmentation without classification,

V1 as a saliency map.

Neuroscience

Neural implementation and manifestation in V1

receptive fields, contextual influences, intra-cortical horizontal connections, cell tuning to local and global features, figure-ground effects.

Cognitive Science

Behavioral and perceptual manifestation

texture segmentation, contour integration (enhancement), popout, illusions, various eases in visual search task and search asymmetric.

Network model

computational mechanisms

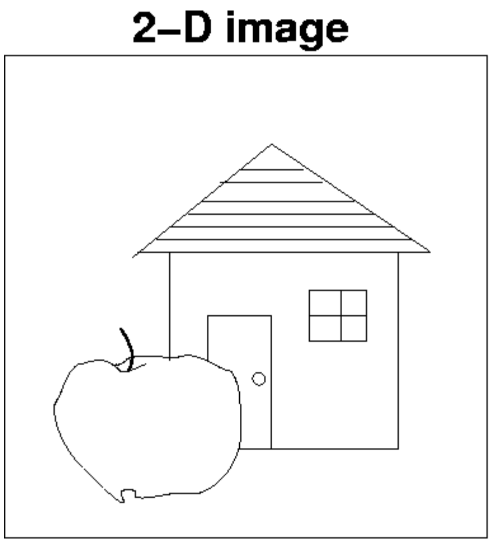
A recurrent model of V1 from local interactions to global behavior, algorithm/design/stability of the recurrent network.

V1, perhaps the largest cortical area in neocortex (12% of the macaque money's neocortex), with most experimental data.

a theorist's goldmine.

The segmentation problem (must be addressed for object recognition)

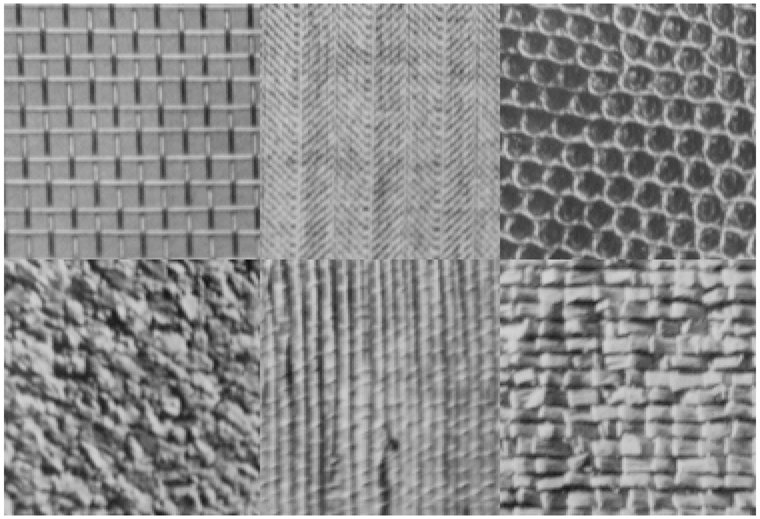
To group image pixels belonging to one object



Dilemma:

Segmentation presumes recognition
recognition presumes segmentation.

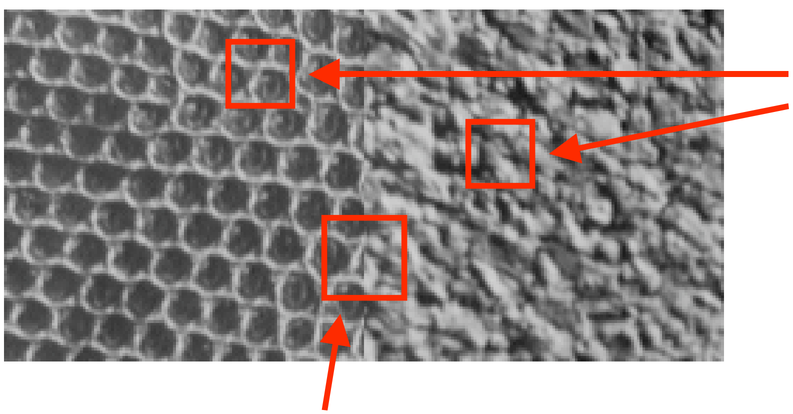
To start: focusing on region segmentation



A region can be characterized by its smoothness regularities, average luminance, and many more descriptions.

Define segmentation as locating the border between regions.

The usual approach: **segmentation with (by) classification**



(1) image feature classification

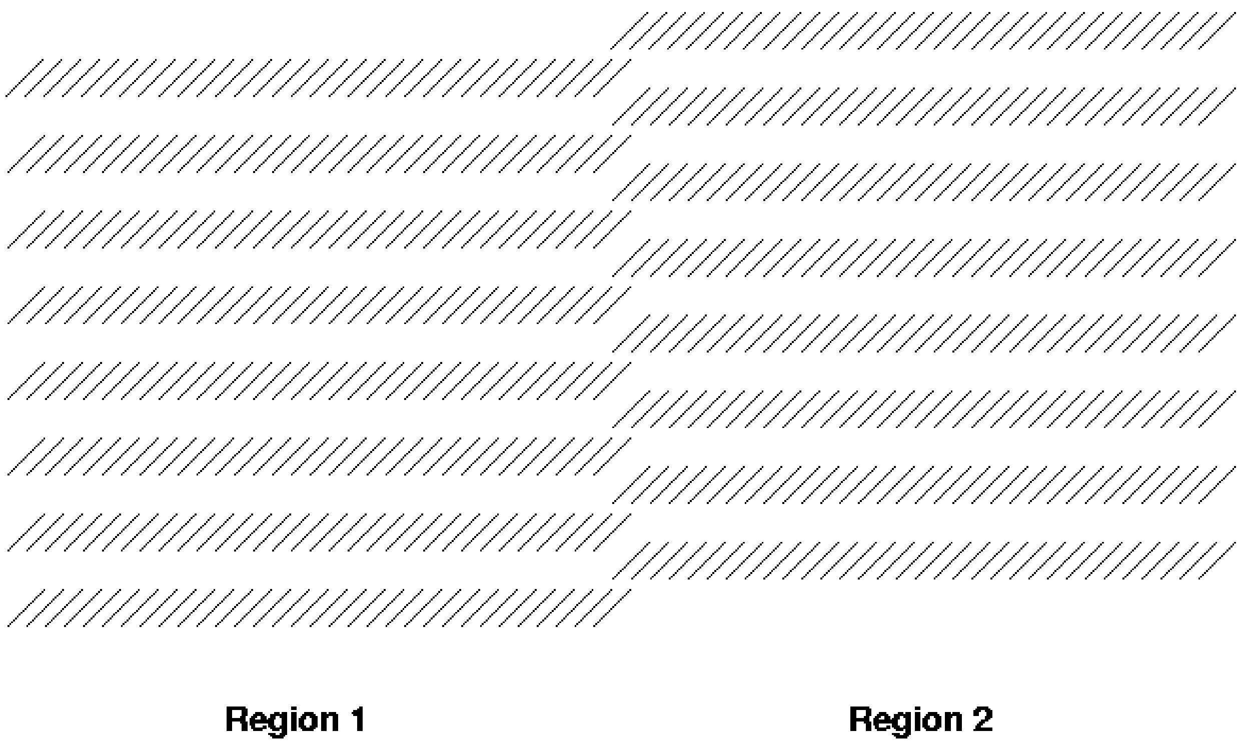
(2) Compare features to segment

Problem: boundary precision vs. feature precision.

Dilemma: segmentation vs. classification

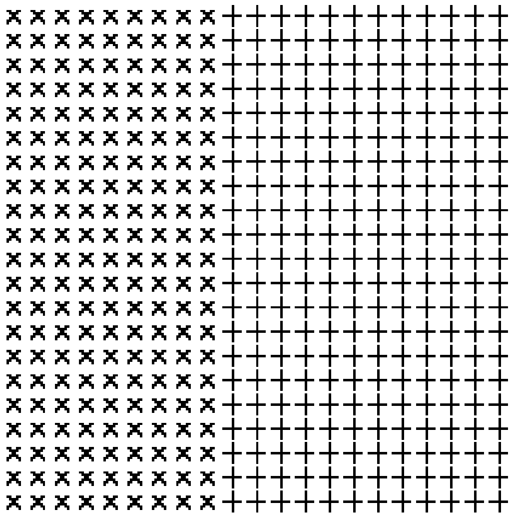
In biological vision:

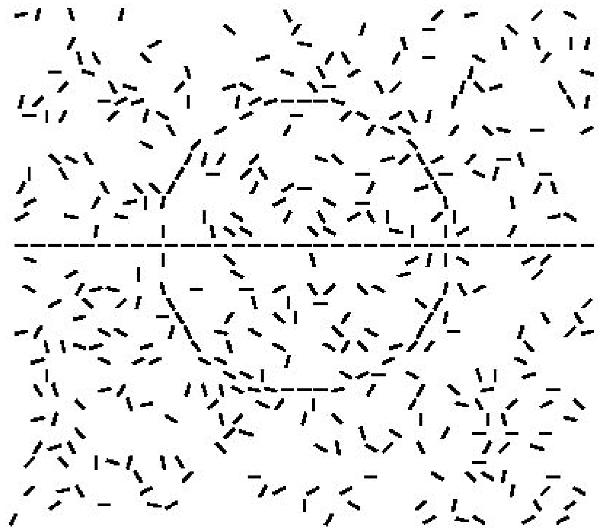
recognition (classification)
is neither necessary nor
sufficient for segmentation

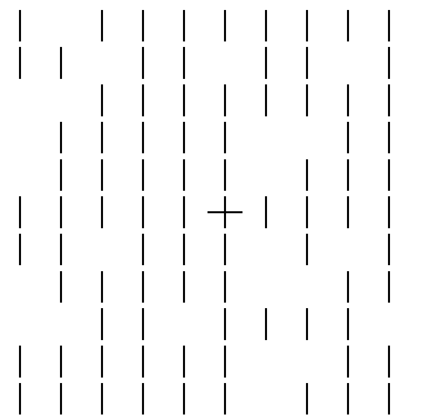


Pre-attentive and attentive segmentations -- very different.

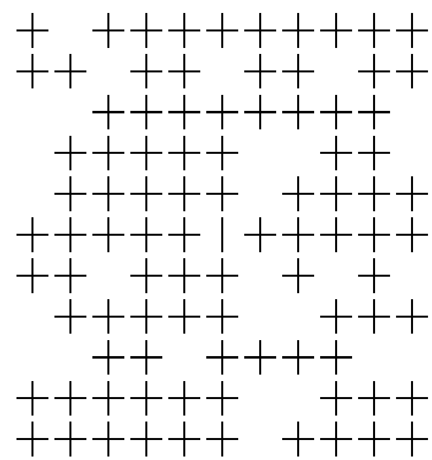
Pre-attentive: **effortless, popout**







Attentive: **effortful, slow**



My proposal:

Pre-attentive segmentation without classification

•Detecting the boundaries by detecting translation invariance breaking in inputs via V1 mechanism.

I show a model of V1 on how this can be done by neural mechanisms to highlight boundaries or conspicuous areas, creating saliency maps from images.

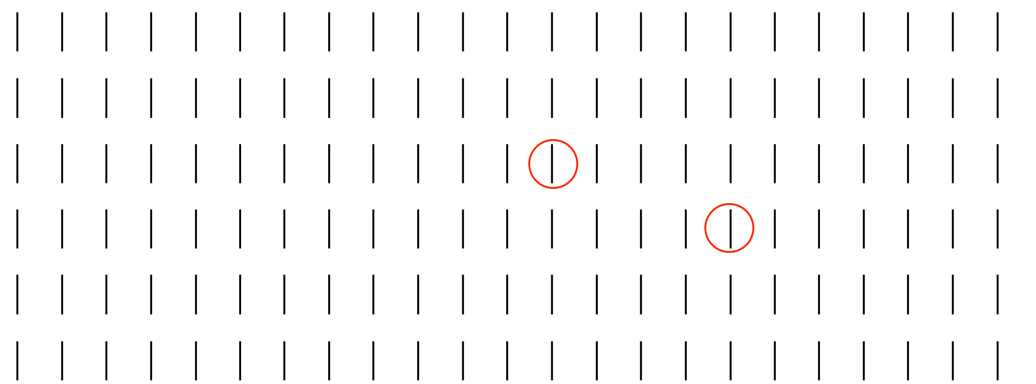
•Individual V1 neurons are like edge detectors.

•Different V1 neurons interact with each other (cf. Markov random field)

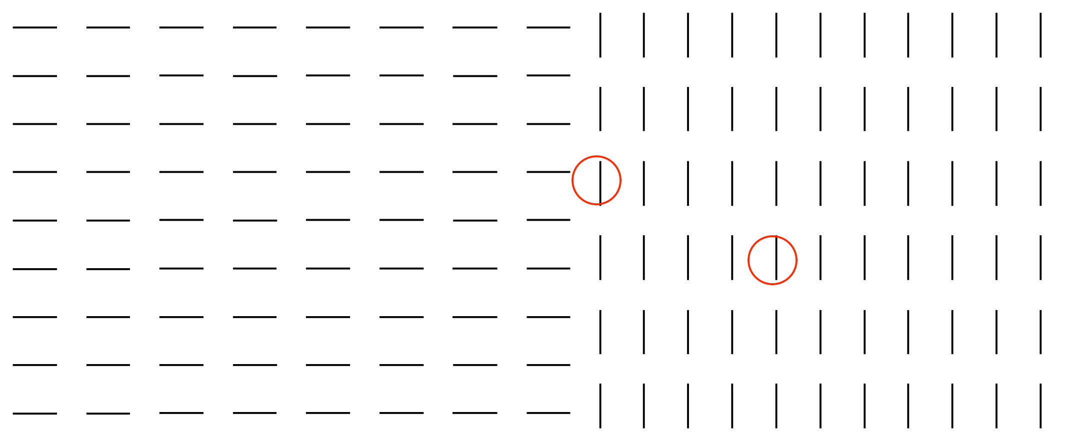
•The interactions creates saliency map.

Principles in my framework: Detecting region boundaries by detecting the breakdown of homogeneity or translation invariance input using contextual influences.

Homogeneous input (one region)

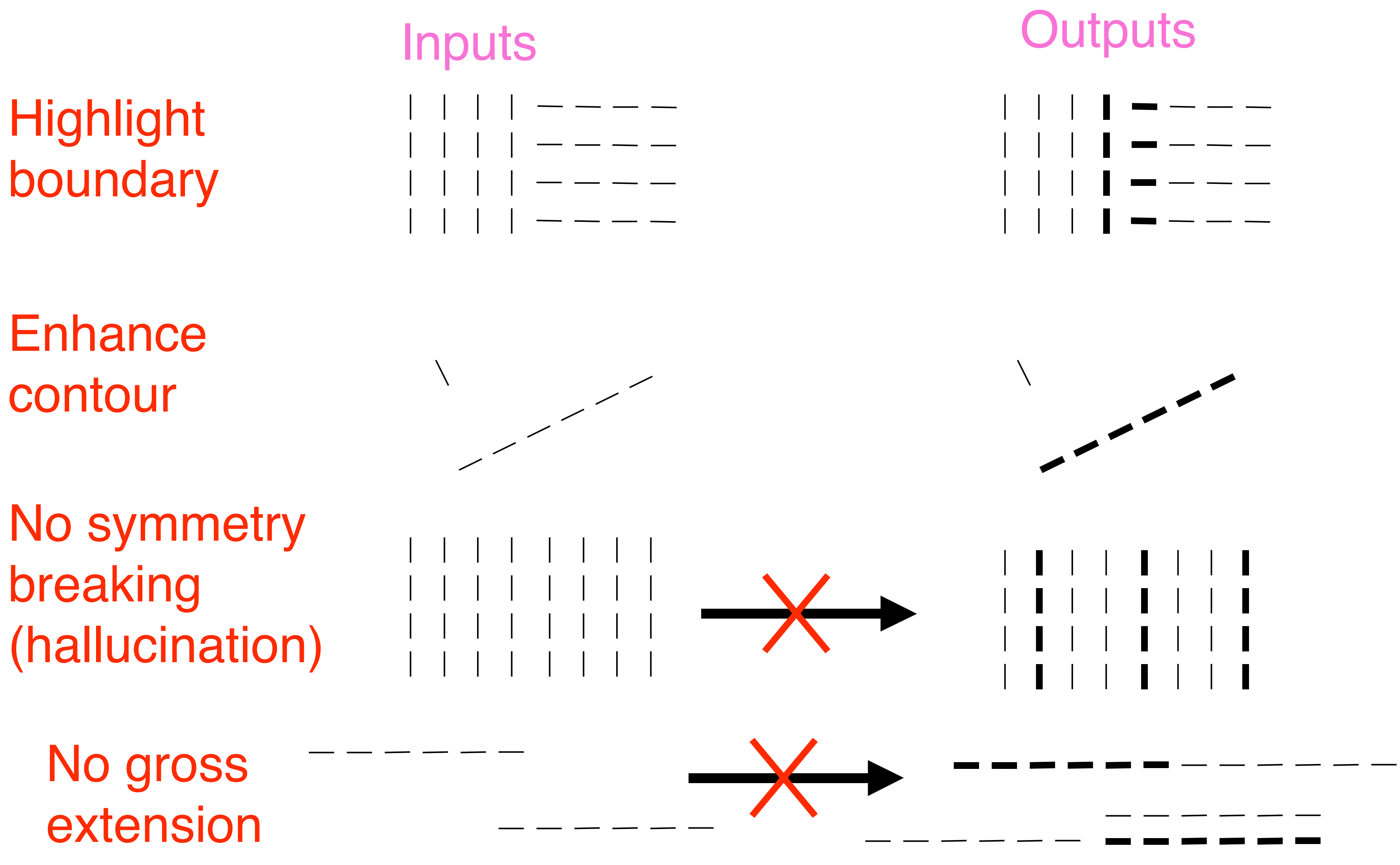


Inhomogeneous input (Two regions)



Separating A from B without knowing what A and B are

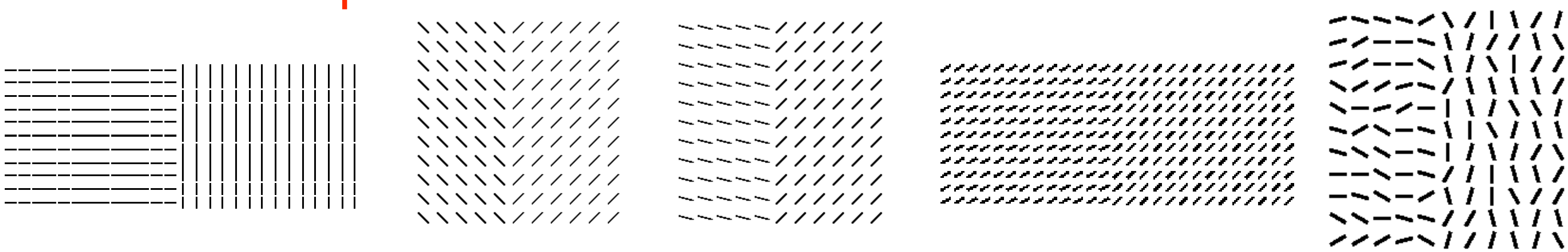
Conditions on the intra-cortical interactions.



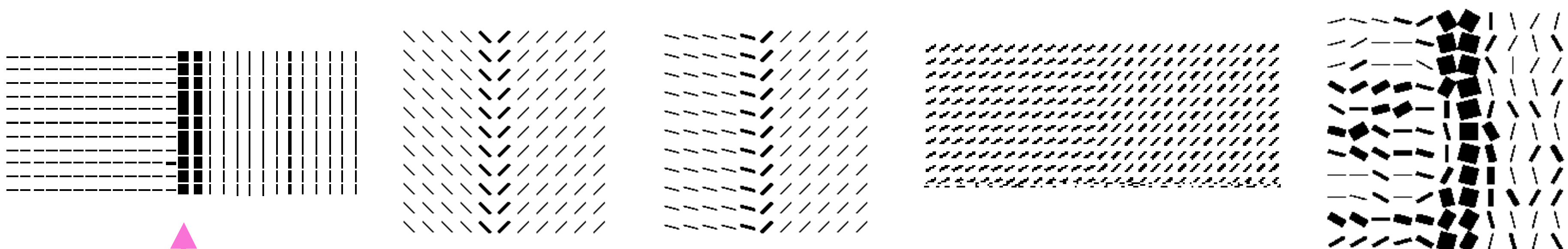
Design techniques: mean field analysis, stability analysis. Computation desired constraints the network architecture, connections, and dynamics. Network oscillation is one of the dynamic consequences.

Texture segmentation simulation results ---
quantitative agreement with psychophysics
(Nothdurft's data)

V1 model input

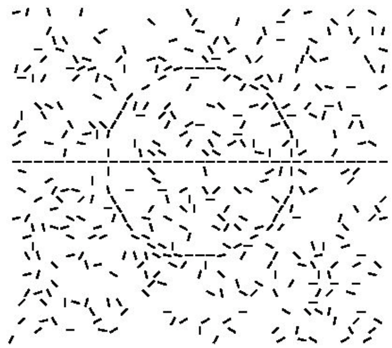


V1 model output

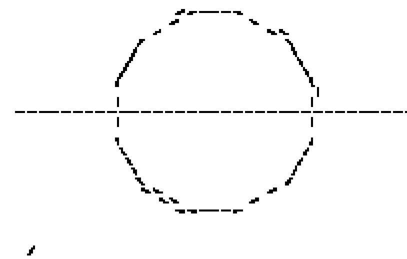


Prediction: bias in the perceptual estimation of the location of the texture boundary.

Input image

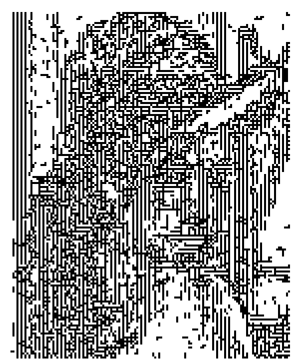


Output highlights

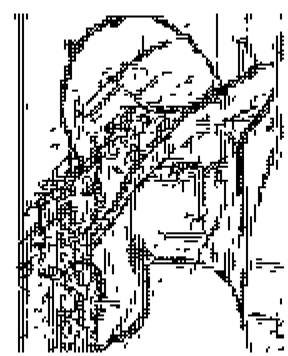


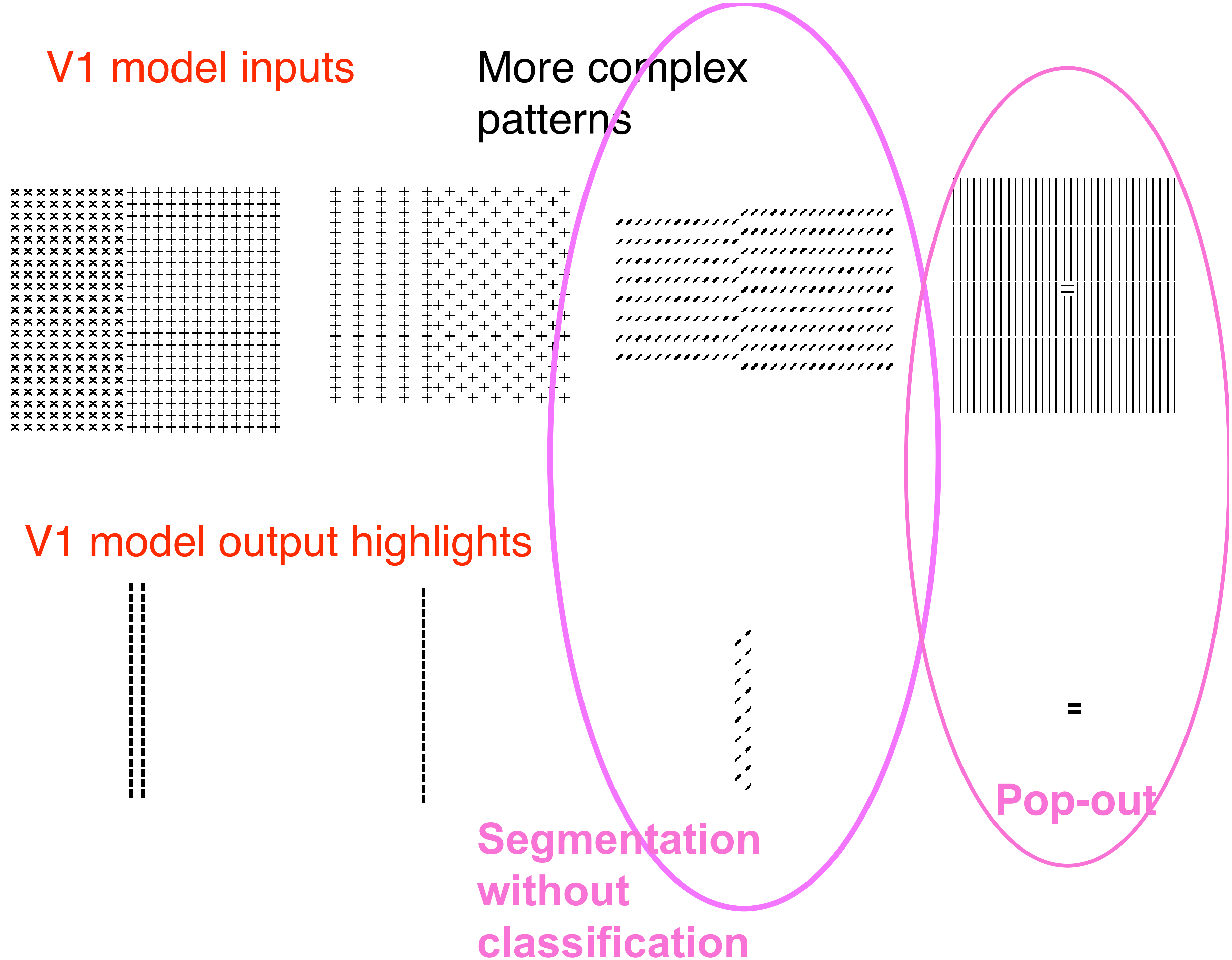
Original image

Sampled image



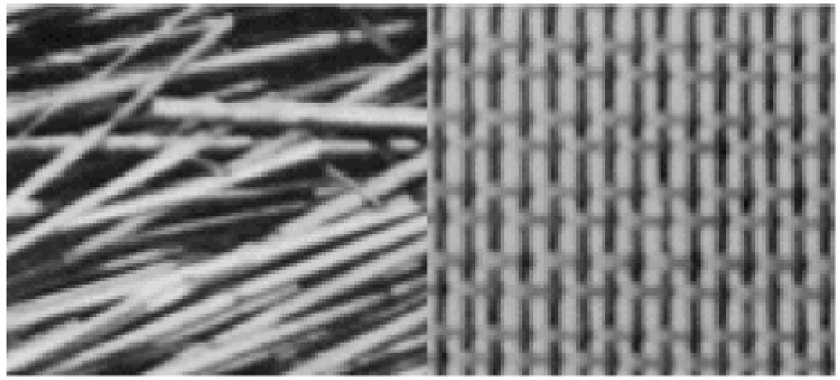
Output image



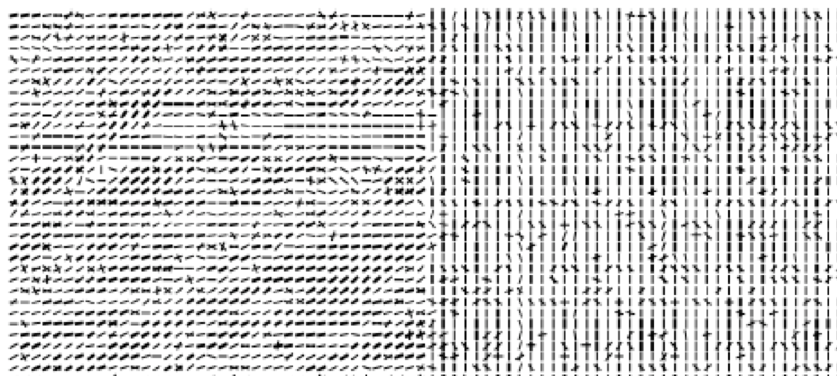

V1 model inputs
More complex patterns
V1 model output highlights
Segmentation without classification
Pop-out

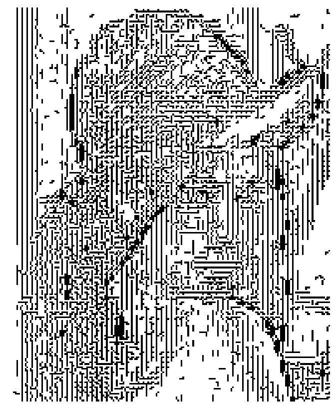
Use natural images

Image

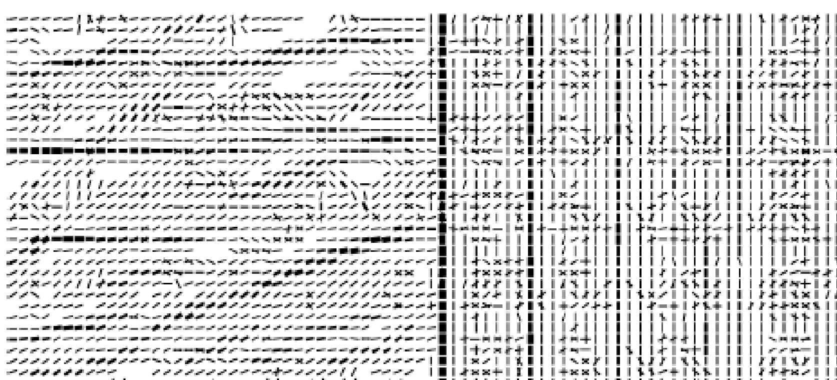


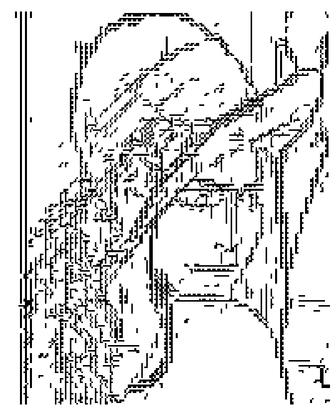
V1 model
inputs





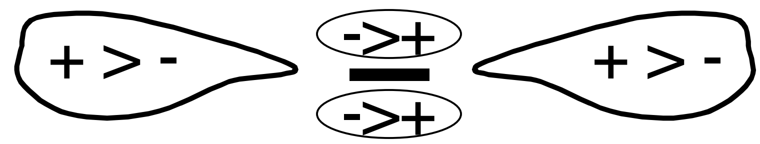
V1 model
outputs





Testable, falsifiable, predictions:

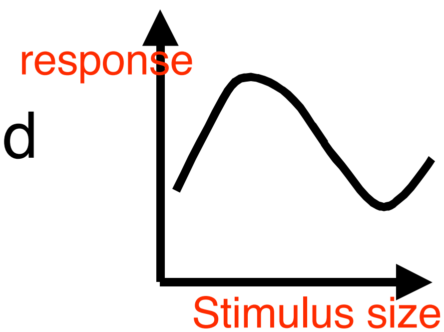
1. Intra-cortical connections:



2. Intra-cortical connections should link cells tuned to same orientation and same motion direction

3. Cells responses should be tuned to orientation of the global texture border.

4. Receptive field summation curves should rebound.

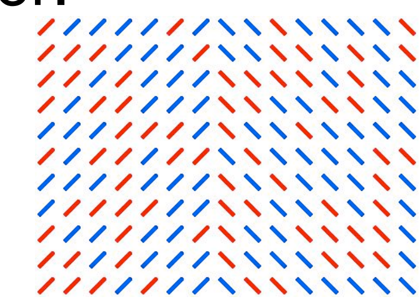


Tested, confirmed, predictions:

1. Lammi, Zipser et al Figure-ground effects diminish for larger figure sizes.

2. Perceptual bias in localizing texture border.

3. Color-orientation interference in texture segmentation increases with color categories.



Comparison with other models

1. Somers, Dragoi, Stemmler, et al. --- of one hypercolumn
   we are trying to get larger, denser spatial sampling.

2. With Grossberg et al. --- my model is V1 only, no top down, reproducible (already reproduced), layer 2-3 only. Does contour integration, texture segmentation, popout, etc. in the same circuit.

3. My model is to supplement a theory --- V1 saliency map

What my model does not do or fail:

1. No top-down, does not say how the saliency map is read or by which cortical areas

2. Current implementation, although 100x100 big, 1 million neurons (diff. Eqs.), is too sparsely sampled, lack multiscale, for natural images, and not yet including motion, depth, etc. (Please give me faster computers!!!)

3. No end-stop cells, no layer 5-6, etc., my model is a minimal model striped down to essentials just to account for saliency effects and related.