A zero parameter quantitative prediction from the V1 Saliency Hypothesis and its match with experimental data

#56.407

Li Zhaoping[1,2] *, Li Zhe[2],

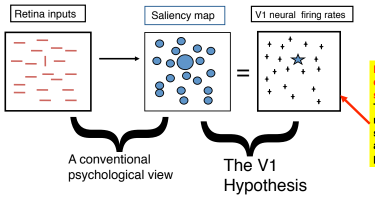
[1]Department of Computer Science, University College London
[2]Computational Neuroscience Lab, School of Medicine, Tsinghua University, Beijing, China

This is a presentation at Vision Science Society annual meeting May 11-16, 2012, in Naples, Florida, U.S.A.

Background:

The V1 Saliency Hypothesis (Li 1999, 2002)

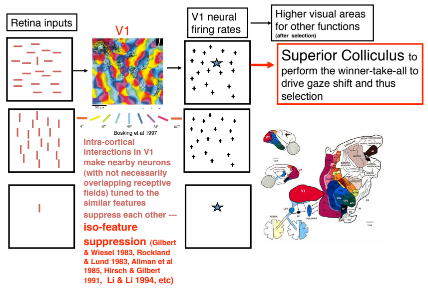
A bottom-up saliency map in the primary visual cortex



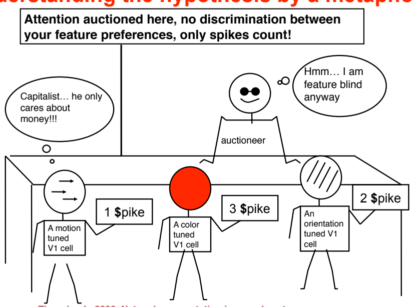
Neural activities as universal currency to bid for visual selection.
The receptive field of the most active V1 cells is selected by exogenous attention for detailed processing

Physiological mechanisms

The reason that the vertical bar evokes highest response is because of Iso-feature suppression between V1 neurons tuned to same or similar features!

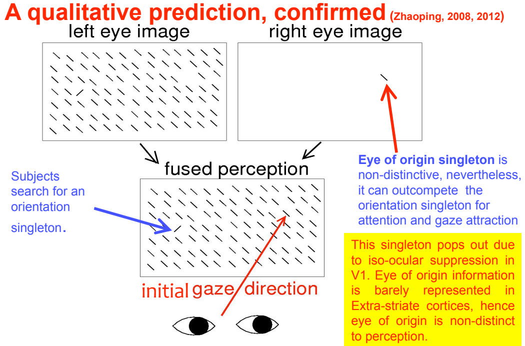


Understanding the hypothesis by a metaphor ---



Zhaoping L. 2006, Network: computation in neural systems

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



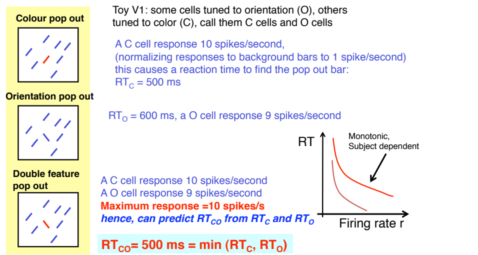
Eye of origin singleton is non-distinctive, nevertheless, it can outcompete the orientation singleton for attention and gaze attraction

This singleton pops out due to iso-ocular suppression in V1. Eye of origin information is barely represented in Extra-striate cortices, hence eye of origin is non-distinct to perception.

Other qualitative predictions which have been experimentally tested and confirmed: Zhaoping & Snowden 2006, Zhaoping & May 2007, Koene & Zhaoping 2007, Jingling & Zhaoping 2008, Zhang et al 2012, etc.

A quantitative prediction: an illustration from a toy V1

Toy V1: some cells tuned to orientation (O), others tuned to color (C), call them C cells and O cells



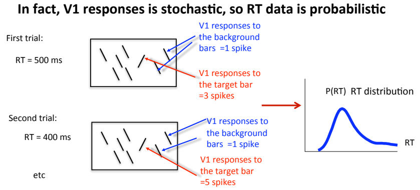
A C cell response 10 spikes/second, (normalizing responses to background bars to 1 spike/second) this causes a reaction time to find the pop out bar: RT_C = 500 ms

RT_O = 600 ms, a O cell response 9 spikes/second

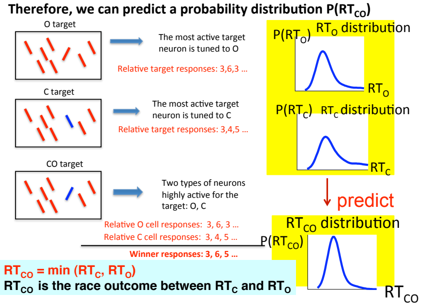
A C cell response 10 spikes/second
A O cell response 9 spikes/second
Maximum response =10 spikes/s
hence, can predict RT_{CO} from RT_C and RT_O

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

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

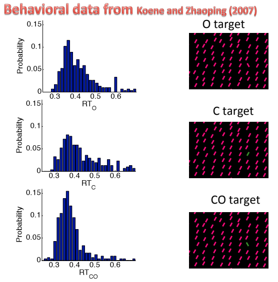


Therefore, we can predict a probability distribution $P(RT_{CO})$



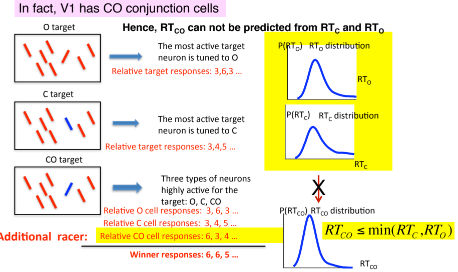
RT_{CO} = min (RT_C, RT_O)
RT_{CO} is the race outcome between RT_C and RT_O

Behavioral data from Koene and Zhaoping (2007)



Visual search task: to press a left or right button as soon as possible to report whether an odd bar is in the left or right half of the image. Button press RT was measured. Each trial started with a central fixation by the subject, then search stimulus appeared and stayed till subject response.

In fact, V1 has CO conjunction cells



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

A quantitative prediction from the actual V1

If V1 did not have CO cells, we could have

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

Unfortunately, V1 has CO cells, so the above cannot be used to predict $\mathrm{Prob}(RT_{CO})$ from $\mathrm{Prob}(RT_C)$ and $\mathrm{Prob}(RT_O)$.

Fortunately, V1 has no CMO cells, we can then analogously show that

$$RT_A \equiv \min(RT_C, RT_M, RT_O, RT_{CMO})$$
$$RT_B \equiv \min(RT_{CM}, RT_{CO}, RT_{MO})$$
then
$$\mathrm{Prob}(RT_A) = \mathrm{Prob}(RT_B)$$

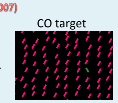
V1 has C,M,O, CO, MO, and few CM cells (Hubel and Wiesel 1959, Livingstone and Hubel 1984, Horwitz and Albright 2005)

Hence, $\mathrm{Prob}(RT_{CMO})$ can be analogously predicted from probability distributions of RT_C, RT_M, RT_O, RT_{CM}, RT_{CO}, and RT_{MO}.

Furthermore, V2 has CMO cells (Shipp, private communication 2011)! Hence if our prediction matches the data, then V2 and above are not needed for visual saliency (at least for such stimuli/tasks).

Behavioral data from Koene and Zhaoping (2007)

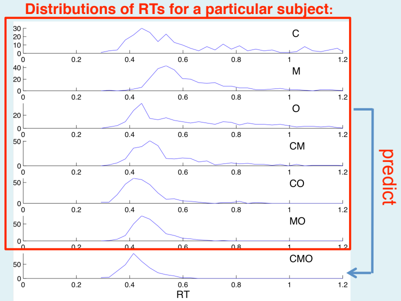
Target is different from distractors in *orientation* (**O**), *color* (**C**), *motion direction* (**M**), or combinations them.



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

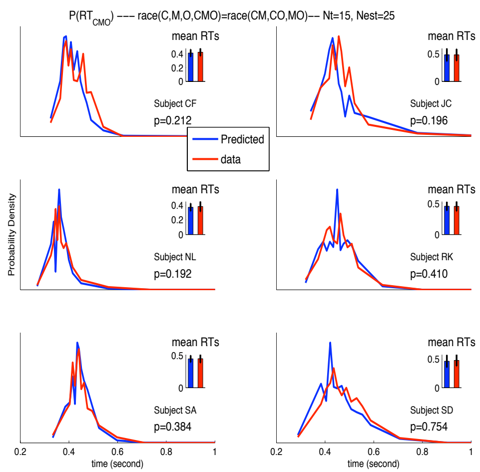
Each about 320 trials / subject, 6 subjects in total

Distributions of RTs for a particular subject:



Approach: predict $P(RT_{CMO})$ from the distributions of the other RTs, and compare the prediction with behavioral data to see if they match each other

Result: predicted $P(RT_{CMO})$ and behavioral $P(RT_{CMO})$ are not significantly different from each other (p>0.05) for each subject.



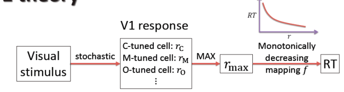
No free parameters in our quantitative prediction

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!

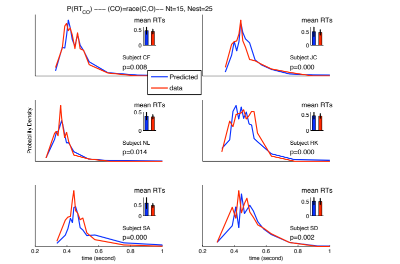
V1 theory



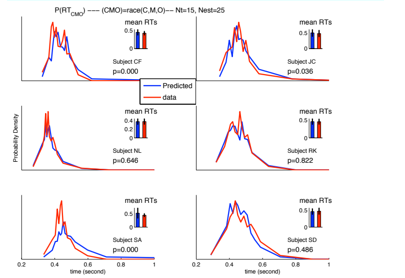
$$RT(r_{max}) = RT(\max(r_C, r_M, ...)) = \min(RT(r_C), RT(r_M)...)$$

Our data are sufficient to distinguish incorrect predictions from data

Incorrect prediction 1: RT_{CO} = min (RT_C, RT_O)



Incorrect prediction 2: RT_{CMO} = min (RT_C, RT_O, RT_M)



Summary:
(1)A theoretical hypothesis that V1 creates a bottom-up salient map can lead to a quantitative prediction of the RTs in visual search task without any free parameters.
(2) The prediction matches quantitatively to the behavioral data.

Reference:

Li Z. (2002) A saliency map in primary visual cortex TICS,*6: 9-16*
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. *JOV, 8(5):1, 1-18,*
Koene AR and Zhaoping L. (2007) Feature-specific interactions in salience from combined feature contrasts: Evidence for a bottom-up saliency map in V1. , *JOV, 7(7): 6, 1-14.*
Zhang X., Zhaoping L. , Zhou T. and Fang F. (2012) Neural activities in V1 create a bottom-up saliency map. Neuron, 73,183-192.