

# **Parallel advantage: further evidence for bottom-up saliency computation by human primary visual cortex**

In press 2021 for *Perception*

Li Zhaoping

University of Tübingen, Max Planck Institute for Biological Cybernetics, Tübingen, Germany

## **Abstract**

Finding a target among uniformly oriented non-targets is typically faster when this target is perpendicular, rather than parallel, to the non-targets. The V1 Saliency Hypothesis (V1SH), that neurons in primary visual cortex (V1) signal saliency for exogenous attentional attraction, predicts exactly the opposite in a special case: Each target or non-target comprises two equally sized disks displaced from each other by 1.2 disk diameters center-to-center along a line defining its orientation. A target has two white or two black disks. Each non-target has one white disk and one black disk, and thus, unlike the target, activates V1 neurons less when its orientation is parallel rather than perpendicular to the neurons' preferred orientations. When the target is parallel, rather than perpendicular, to the uniformly oriented non-targets, the target's evoked V1 response escapes V1's iso-orientation surround suppression, making the target more salient. I present behavioral observations confirming this prediction.

**Keywords:** Saliency, primary visual cortex, V1SH, visual search

## **1 Introduction**

Where we should direct our attention or gaze is determined by both top-down (also called goal-dependent and endogenous) and bottom-up (goal-independent, stimulus driven, and exogenous) factors (Treisman and Gelade, 1980; Theeuwes, 2010; Wolfe, 2021). For example, top-down factors guide our gaze to a page of a book we are reading, while bottom-up factors distract our gaze to an unexpected insect that suddenly appears beyond the book pages. Brain regions such as the frontal eye field and parietal cortical areas are believed to be the neural basis for the top-down control. However, the neural basis for bottom-up control is still hotly debated.

To limit the debate, define a visual location's saliency as its ability to attract attention exogenously. A location can be salient by, e.g., having a unique vertical bar among horizontal bars, a unique red item among green items, or a uniquely left-moving object among right-moving ones. Hence, saliency is general across many visual features (e.g., vertical orientation or red color) and feature dimensions (e.g., orientation, motion, and color). This has motivated the idea that a saliency map to guide attention exogenously should reside in brain areas such as the frontal eye field and parietal cortex, which contain retinotopic maps of the visual field and whose neurons are untuned to any specific visual features or feature dimensions (Koch and Ullman, 1985; Itti and Koch, 2001).

The V1 Saliency Hypothesis, V1SH, however, posits that saliency of a visual location is signalled by

the highest response it evokes from V1 neurons whose receptive fields(RFs) cover that location, relative to the highest V1 responses to other locations. This is without regard to the preferred features (such as color and orientation) of the neurons (Li, 1999, 2002). For example, an orientation singleton, such as a unique vertical bar among horizontal bars, is salient because its evoked V1 response is higher than responses evoked by the background bars. The underlying neural mechanism is iso-feature suppression, whereby a V1 neuron's response to visual input is typically suppressed by active neighboring neurons preferring similar input features (Knierim and Van Essen, 1992; Zhaoping, 2014). Hence, a neuron preferring horizontal orientation and responding to a horizontal background bar is, by iso-orientation suppression, suppressed by other horizontal-preferring neurons responding to nearby horizontal bars. In contrast, a vertical-preferring neuron responding to the unique vertical bar escapes the iso-orientation suppression, thereby giving a higher response. These V1 saliency signals are monosynaptically sent to the superior colliculus (a midbrain area heavily involved in eye movement (Schiller, 1984)), allowing execution of attentional shift to the most salient location. Iso-color suppression and iso-motion-direction suppression analogously make color singletons and motion-direction-singletons salient, respectively.

Various predictions of V1SH have been experimentally confirmed, see a review in (Zhaoping, 2014). This includes the surprising phenomenon that an eye-of-origin singleton, such as a horizontal bar presented to the left eye among other horizontal bars presented to the right eye, captures attention automatically even when the singleton is task-irrelevant and indistinguishable to the observers from the background bars (Zhaoping, 2008). This singleton, whose unique eye-of-origin is visible to V1 but not to higher visual areas, evokes a higher V1 response due to iso-eye-of-origin-suppression. V1SH has also been tested electrophysiologically in monkeys: V1 neurons' responses to an orientation singleton are inversely correlated with the behavioral latencies to saccade to the singleton (Yan et al., 2018).

Here, we test a further surprising prediction of V1SH. We consider the search task of finding a target, a homo-pair of disks (two white disks or two black disks) in a background of hetero-pairs of disks (each made of one white disk and one black disk). The orientation of each disk-pair, target or non-target, is that of the displacement between the two disks. Figure 1 shows that a hetero-pair better excites a V1 neuron preferring an orientation perpendicular rather than parallel to itself, since this is when the black and white disks, respectively, fall into the off- and on-subfields of the neuron's RF. Smith et al. (2002) analyzed such a stimulus-response relationship in detail, showing that monkey's V1 neurons indeed respond accordingly. Hence, when the background hetero-pairs are perpendicular to the target (homo-pair) (e.g., in Figure (2B)), they mainly excite V1 neurons preferring the target's orientation (see Fig. (1D) right), leading to iso-orientation suppression of the neuron responding to the target. In contrast, when the background hetero-pairs are parallel to the target (e.g., in Figure (2A)), the V1 neuron responding to the target escapes this suppression (see Fig. (1D) left). Hence, V1SH predicts that this target is less salient when perpendicular rather than parallel to the non-targets. By contrast, higher brain areas are likely to define the orientation of the non-targets at the more symbolic level of the displacement between the two disks in each pair, and so predict that the target would be less salient when parallel rather than perpendicular to the non-target as would be conventional in the visual search literature (Wolfe, 2021).

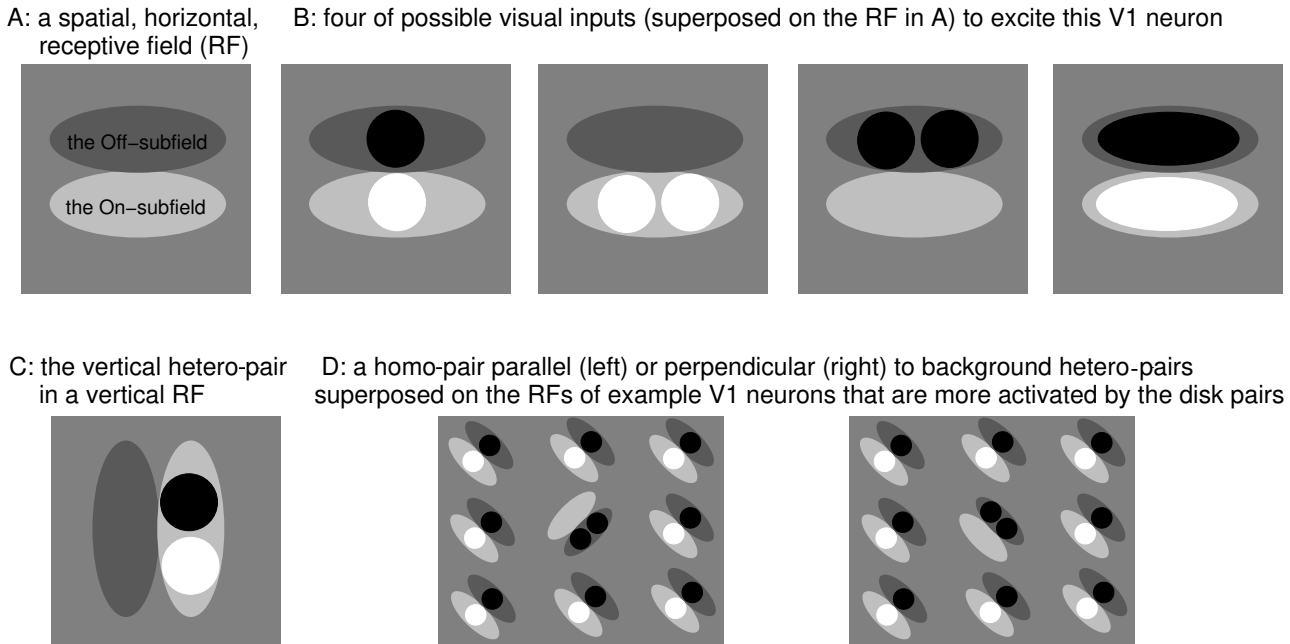


Figure 1: A: Schematic of the Gabor-like, horizontally oriented, RF of a V1 neuron, with on- and off-subfields. B: Activation of (A) neuron by a vertical hetero-pair; horizontal homo-pairs and a horizontal gabor. C: Vertical hetero-pair in a vertical RF. The vertical hetero-pair activates a horizontal-tuned cell in B better than a vertical-tuned cell in C. D: visual inputs resembling subparts of visual search stimuli in Fig.2 superposed on some example RFs of the V1 neurons that are more activated by the disk-pairs in these inputs. A-C are adapted from Figure 2 of (Zhaoping, 2020).

## 2 Experimental methods and results

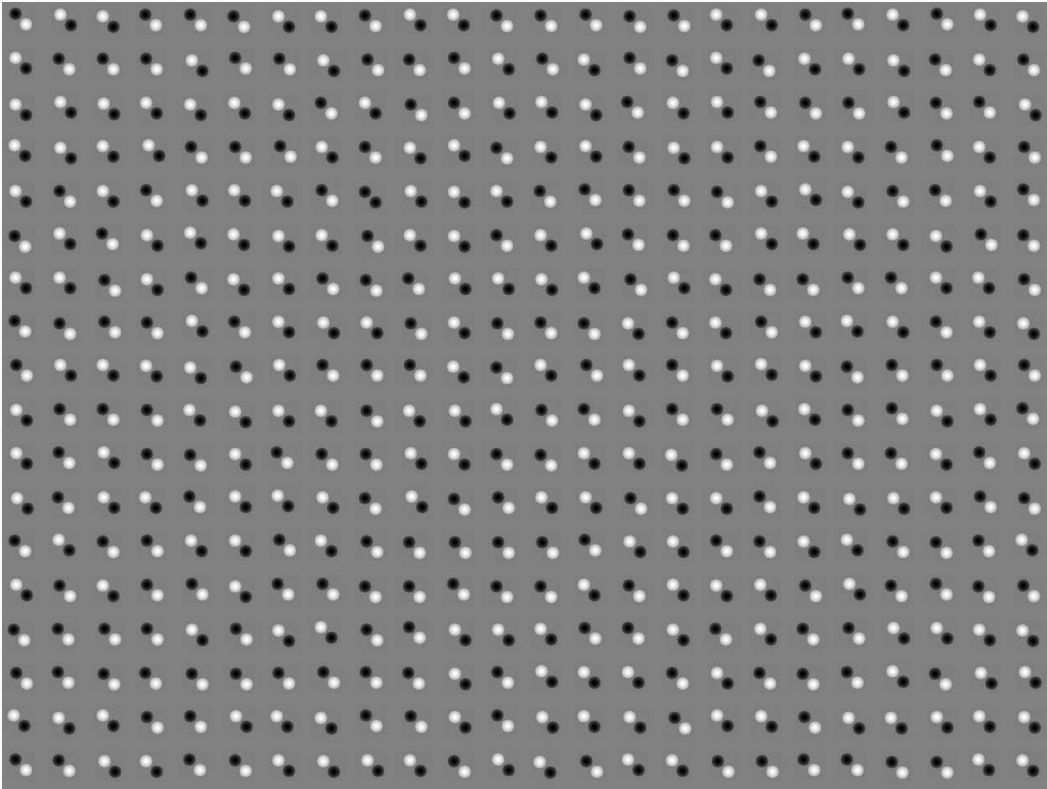
### 2.1 Methods

The search images (as in Figure (2)) contained 24 columns and 18 rows of disk pairs. They were displayed on a Sony cathode ray tube of resolution  $1264 \times 948$  pixels, spanning  $38.3 \text{ cm} \times 29.1 \text{ cm}$  or  $48.0^\circ \times 37.4^\circ$  in visual angle at a 43 cm viewing distance in a dim room.

Each trial started with a central fixation cross on a gray background for 0.7 second, followed by 0.2 second without the fixation cross before a search image appeared waiting for observer's response. Observers had to fixate on the cross until the search image appeared, then could freely move their gaze and press a left or right button as quickly as possible to report, respectively, a target in the left or right half of the display. They started the next trial by another button press. Right after several practice trials, each observer performed 400 testing trials and typically took one short break in the middle. The average durations ( $\pm$  standard deviation) to complete the first 200 testing trials, the second 200 testing trials, and all the 400 testing trials (including break times), were  $18.8 \pm 4.8$  minutes,  $13.9 \pm 1.9$  minutes, and  $36.2 \pm 8.0$  minutes, respectively.

Excluding the central four columns in the search array, the target was randomly and equally likely at any location within 9 columns and 7 rows from the display center, making it between about  $5.4^\circ$  and  $21.7^\circ$  from the display center. In each trial, the target was also randomly and equally likely to be a white or

A: A target parallel to the non-targets



B: A target perpendicular to the non-targets

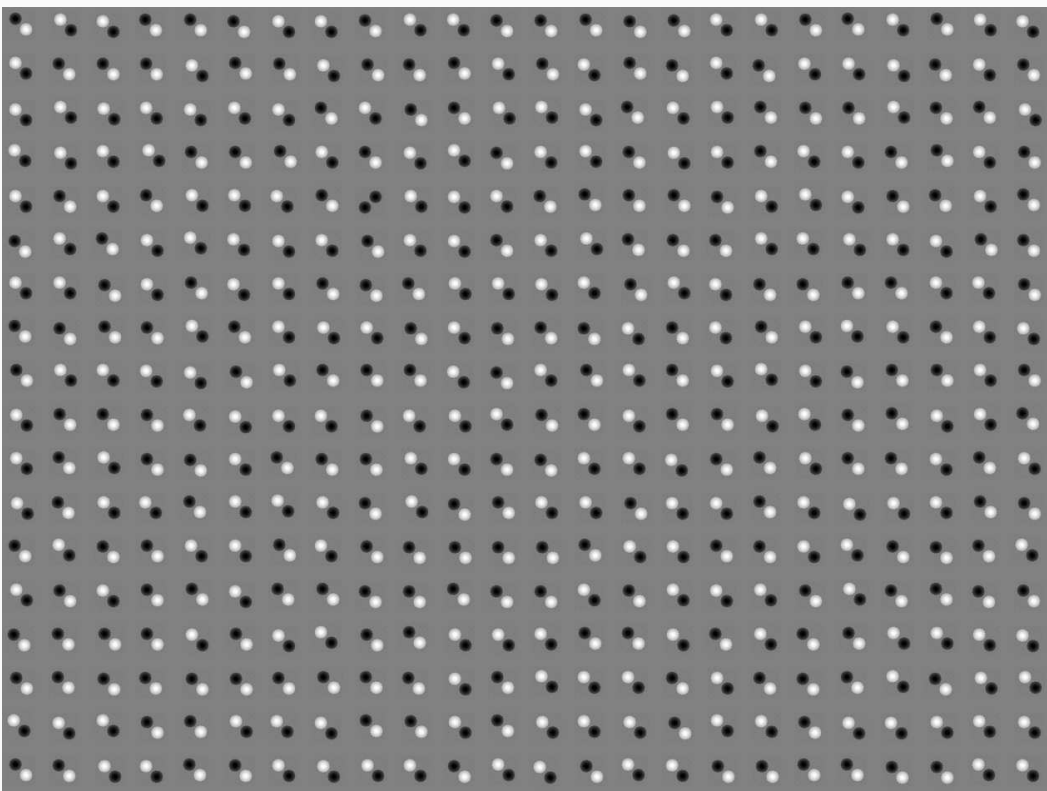


Figure 2: Two example search images. They differ only in the orientation of the target, a homo-pair of black disks, in the fifth row from the top and ninth column from the left. The non-targets are hetero-pairs of disks.

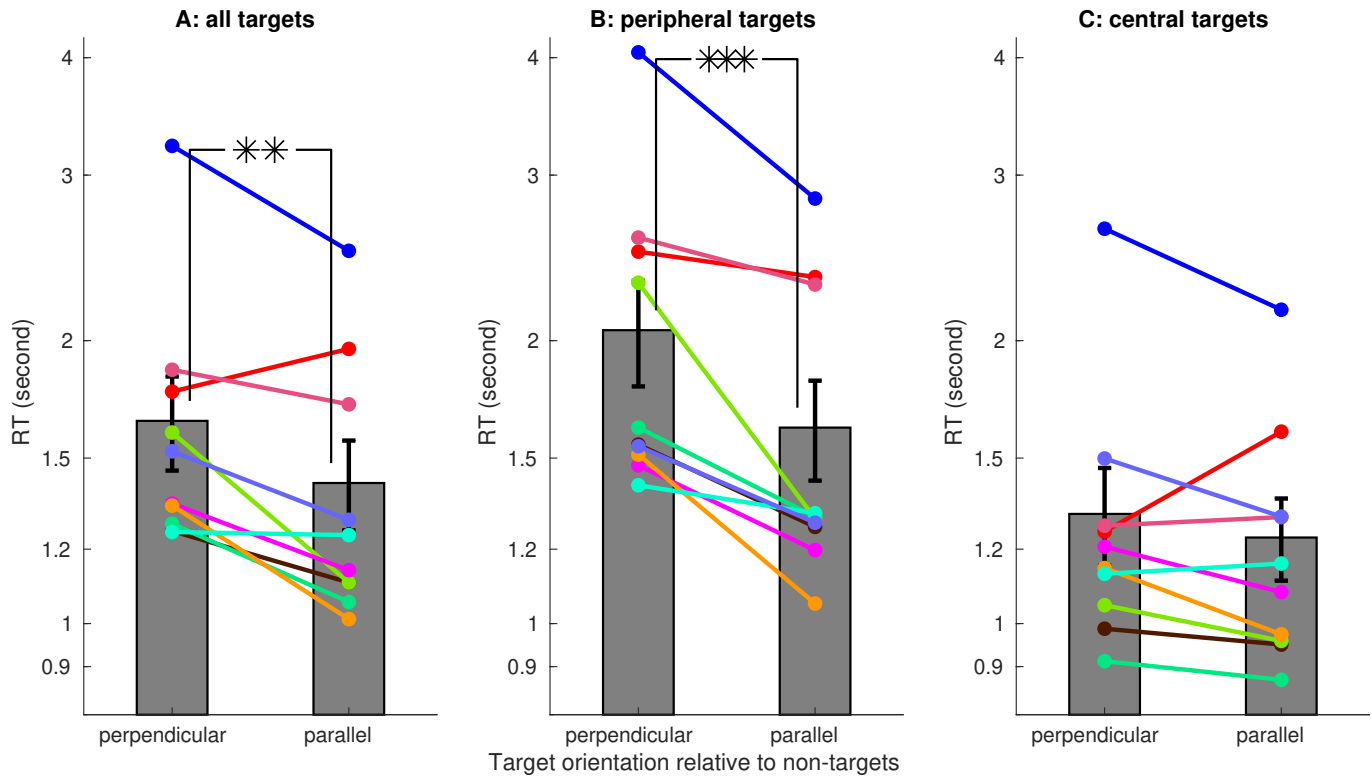


Figure 3: Reaction time (RT) to find a target was shorter for targets parallel rather than perpendicular to the non-targets. A target that is more or less than its average distance ( $14^\circ$ ) from the display center is termed “peripheral” or “central”, respectively. In the current and next figures, data bars mark the average RTs across  $n = 10$  observers (error bars are the standard errors); RTs of individual observers are visualized by individually colored  $\times$ 's linked by the correspondingly colored lines; two data bars linked by one, two, or three '\*'s indicate that the RT difference is significant with  $p$  values  $0.05 < p \leq 0.01$ ,  $0.001 < p \leq 0.01$ , or  $p \leq 0.001$ , respectively.

black homo-pair, and oriented  $45^\circ$  clockwise or counter-clockwise from vertical. Meanwhile, the uniformly oriented non-targets were randomly and equally likely to be parallel or perpendicular to the target. The black disk in each non-target was randomly and independently the upper or lower one with equal chance. The gray background had luminance  $L_o = 44 \text{ cd/m}^2$ . Each disk's diameter was  $D = 15$  pixels, two disks in a pair were  $R = 18$  pixels apart center-to-center, a disk-pair spanned  $1.33^\circ \times 0.61^\circ$  at the display center. With  $D \approx R$ , the most activated V1 neurons by each homo- or hetero-pair should prefer spatial frequency (SF)  $\sim 0.69$  cycles/degree. This SF is within the range of preferred SFs of monkey V1 neurons for each target eccentricity (Gattass et al., 1987) from the display center. Within a disk, each pixel at distance  $r$  from the center had a luminance deviation from the gray background by about  $L_o \cdot \exp(-r^2/(2\sigma^2))$  with  $\sigma = 5.4$  pixels. The location of each disk pair was randomly jittered, horizontally and vertically, from the regular  $24 \times 18$  grid such that two neighboring pairs were between 49 and 57 pixels (mean 52.6 pixels) apart center-to-center horizontally and vertically.

To test whether reaction times (RTs) for different target conditions differed significantly from each other, paired permutation tests (rather than t-tests) across observers on their trial-averaged quantities are

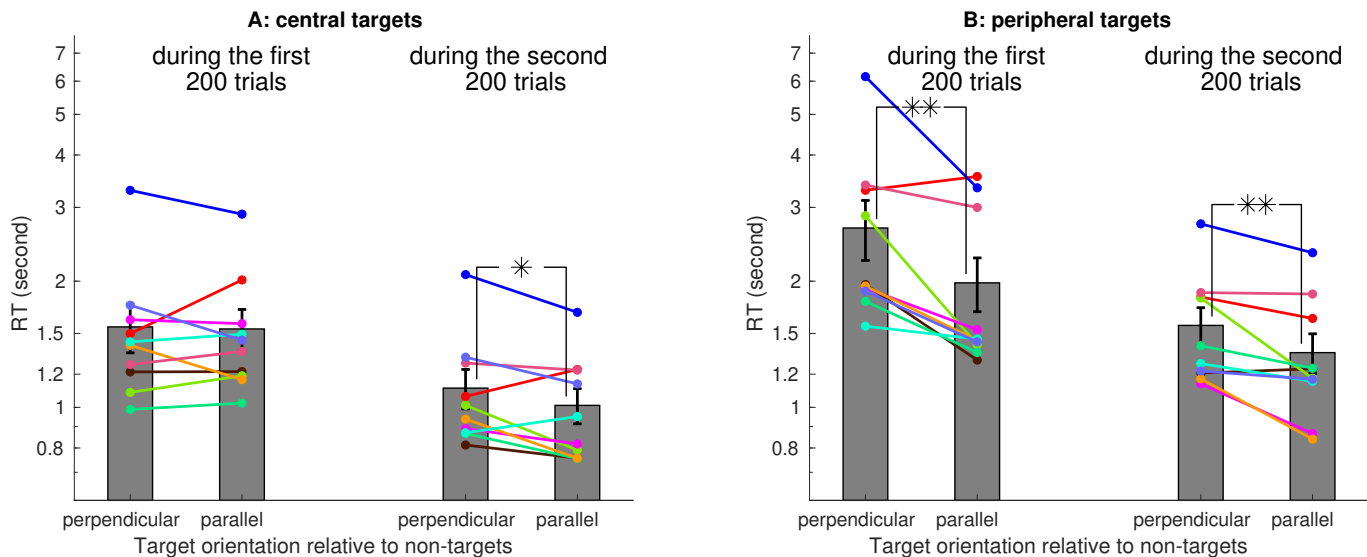


Figure 4: Learning effects. Each observer had several practice trials right before the 400 testing trials. The saliency effect, manifested as a shorter RT for the parallel than the perpendicular target, was significant for the peripheral but not the central targets in the first 200 testing trials, and became significant for the central targets in the second 200 testing trials.

used, so that no assumptions need to be made regarding whether the RTs are normally distributed. Trials with RT shorter than 0.2 second or a wrong button press constituted no more than 2% of the trials for each observer, and are excluded from analysis. The qualitative results are insensitive to whether the RT outliers are removed, and are also insensitive to whether the trial RTs are transformed to their logarithms (to make the distribution of the data closer to a normal distribution) for the data analysis.

## 2.2 Results

Averaged across 10 observers (8 of them were naive; and one of them was the author), the RT was significantly shorter when the target was parallel rather than perpendicular to the non-targets (Figure (3A)), confirming the V1SH prediction. This RT difference was mainly due to the peripheral target trials, in which the target is further from the center of the visual display (Figure (3B)). However, focusing on the central-target trials (i.e., non-peripheral target trials), this RT difference is insignificant (Figure (3C)). As observers progressed towards the second 200 of their 400 trials, RTs decreased, and the RT difference became significant for the central-target trials (Figure (4A)). Meanwhile, for the peripheral-target trials, this RT difference was significant already in the first 200 trials (Figure (4B)).

## 3 Summary and Discussion

Whether a target was parallel or perpendicular to the non-targets was unpredictable in each trial and irrelevant to the search task in the current study. Hence, the parallel target was found faster purely by bottom-up, or exogenous, factors. Indeed, we have defined saliency as the strength of attentional attraction by exoge-

nous factors (Zhaoping, 2014). Since the target and non-targets, respectively, excite V1 neurons preferring orientations parallel and perpendicular to themselves (Smith et al., 2002), the preferred orientations of the activated V1 neurons, rather than the figure orientations (i.e., the main axis) of symbolic objects (the disk-pairs), dictate saliency.

Among primate brain areas such as frontal eye field (FEF), lateral intraparietal area (LIP), superior colliculus (SC), and V1 that have been considered for saliency computation, V1 is most selective to the orientation feature, particularly for the sort of small disk-pairs we used in this study. Neurons in FEF and LIP are typically untuned to visual features unless the features are relevant for a task on which the animal has been extensively trained (Bichot et al., 1996; Freedman and Assad, 2006). In monkey SC neurons, traditional studies suggest that tunings to visual features is absent or very weak except in extensively trained monkeys (Schiller, 1984; Horwitz and Newsome, 2001). More recent works (Chen and Hafed, 2018; Chen et al., 2018) reported modest feature tunings to orientation and spatial frequency in monkey SC neurons, although these forms of tuning are much weaker than those in V1 neurons.

To read out a V1 saliency map for attentional shift to the most salient location by the winner-take-all computation, any feature signals (which are irrelevant for implementing a gaze or attentional shift) in V1 responses should be degraded and eliminated progressively along the read-out route (Zhaoping, 2014, 2016). The weaker feature selectivity in SC compared to V1 is consistent with the idea that SC, which has a longer response latency than V1 (White et al., 2017), inherits and degrades feature selectivity from V1 when reading out the saliency map. This selectivity degradation could even start in V1's layer 5 neurons which project to SC (see (Zhaoping, 2014, 2016) for discussions). Indeed, in mice (in which neural types are more easily differentiated experimentally than in monkeys), such layer 5 neurons have weaker feature tunings than other layer 5 neurons (Lur et al., 2016). Parietal and frontal areas can also inherit saliency signals from V1 to combine them with top-down factors for controlling attention (Bisley and Goldberg, 2011; Shomstein, 2012).

This study adds to previous behavioral, electrophysiological, and neural imaging investigations that confirm V1SH predictions (Zhaoping, 2014; Yan et al., 2018). For our current stimuli, V1SH does not precisely predict whether the RTs (for a parallel or perpendicular target) should be insensitive to the number of non-targets, i.e., the search set size. Hence, measuring the set size effect, interesting though it is for research in visual search, would neither support, nor falsify, V1SH. One may also ask, if the target is a hetero-pair while non-targets are uniformly oriented and randomly mixed black homo-pairs and white homo-pairs (analogous to our current design), whether a parallel target should yield a shorter RT than a perpendicular target by the same V1 mechanisms. The answer is unclear since, unlike a hetero-pair whose average luminance (locally) matches that of the gray background, a homo-pair is brighter or darker. Consequently, a random array of black and white homo-pairs will create a luminance interference that, analogous to the color interference of orientation target search (Snowden, 1998; Zhaoping and Snowden, 2006; Zhaoping and May, 2007), should overwhelm any saliency effects by the target's orientation. Therefore, such a study would also be unable to support or falsify V1SH.

However, future studies could test other relevant predictions. For example, consider an orientation-tuned V1 neuron activated by an optimally oriented homo-pair in its RF. Surround suppression from uniformly oriented hetero-pairs outside the RF should be stronger when these hetero-pairs are perpendicular

rather than parallel to the homo-pair. For another example, if the two disks in our hetero-pair are both sufficiently lighter or both sufficiently darker than the gray background, then the preferred orientation of the most activated V1 neurons should be parallel rather than perpendicular to the hetero-pair. Using such hetero-pairs for our search task, and making sure that all disks (from the target and non-targets) are all lighter or all darker than the gray background to prevent the kind of luminance interference mentioned above, V1SH predicts that a perpendicular (homo-pair) target should be found more quickly than a parallel target (Jeremy Wolfe, private communications).

Our observers reported that the search was effortful in the initial trials and that it became increasingly easier and reflexive in later trials. This learning may be related to similar learning effects in an orientation singleton search task in monkeys observed in Yan et al. (2018), although the number of trials for the monkeys was much larger. The monkeys' learning was reflected in the correlation between shorter behavioral RTs (to saccade to the target) and higher V1 responses to the orientation contrast between the target and (uniformly oriented) non-targets (Yan et al., 2018). During the initial trials, this correlation was mainly with the V1 responses 100-200 ms after the search stimulus onset (and before the saccade); and the correlation with the initial V1 responses (40-60 ms after the stimulus onset) was insignificant except for the more salient targets. As monkeys' RTs decreased through practice, the correlation with the later V1 responses decreased, whereas the correlation with the initial V1 responses increased and became significant even for the less salient targets. The substantial presence of the initial V1 responses (to the orientation contrast) was regardless of whether the monkey was doing the search task, whereas that of the later responses was only during the task (Yan et al., 2018), suggesting that the early and later V1 responses reflect bottom-up and top-down mechanisms, respectively. The learning, by the monkeys and by our observers, perhaps involves coming to rely more on the bottom-up saliency signals and less on the top-down control for the task. The 40-ms latency (of the initial V1 responses to the orientation contrast of the target) is too short for the V1 saliency signals to arise from feedback from downstream brain areas such as FEF, LIP, or SC.

Like previous studies (Zhaoping and Guyader, 2007; Nuthmann et al., 2021), this study also illustrates the central-peripheral dichotomy that peripheral and central vision are mainly for looking (to select where to shift attention to) and seeing (object recognition), respectively (Zhaoping, 2019). The contrast between the saliency effects (i.e., shorter RTs to find a parallel target) in our central versus peripheral targets is consistent with the idea that saliency processes are stronger for more peripheral visual locations (Zhaoping, 2014). A target closer to the fovea is more likely to lie within the attentional spotlight and so to be subject to top-down factors (Zhaoping, 2019) beyond bottom-up saliency. A target in the peripheral visual field is more vulnerable to visual crowding, so that it attracts gaze mainly by its saliency rather than by a recognition of its shape. Since saliency is the strength to attract attention exogenously, saliency-guided gaze shifts should in principle occur without recognizing the object at the saccade destination. In practice this occurs at least in the case of gaze and attentional capture by the eye-of-origin singleton even though observers could not distinguish this singleton from other items in the visual field (Zhaoping, 2008, 2012). Our observed learning, which particularly increased the saliency effects in our central-target trials, enabled our observers to employ more looking rather than seeing processes for this task.



## Acknowledgement

Work supported by fundings from the University of Tübingen and Max Planck Society. I am grateful to Jeremy Wolfe and two anonymous reviewers for very helpful comments on the manuscript.

## References

- Bichot, N. P., Schall, J. D., and Thompson, K. G. (1996). Visual feature selectivity in frontal eye fields induced by experience in mature macaques. *Nature*, 381(6584):697–699.
- Bisley, J. and Goldberg, M. (2011). Attention, intention, and priority in the parietal lobe. *Annual Review of Neuroscience*, 33:1–21.
- Chen, C.-Y. and Hafed, Z. M. (2018). Orientation and contrast tuning properties and temporal flicker fusion characteristics of primate superior colliculus neurons. *Frontiers in neural circuits*, 12:58.
- Chen, C.-Y., Sonnenberg, L., Weller, S., Witschel, T., and Hafed, Z. M. (2018). Spatial frequency sensitivity in macaque midbrain. *Nature Communications*, 9:Article number 2852.
- Freedman, D. J. and Assad, J. A. (2006). Experience-dependent representation of visual categories in parietal cortex. *Nature*, 443(7107):85–88.
- Gattass, R., Sousa, A., and Rosa, M. (1987). Visual topography of V1 in the cebus monkey. *The Journal of Comparative Neurology*, 259(4):529–548.
- Horwitz, G. D. and Newsome, W. T. (2001). Target selection for saccadic eye movements: direction-selective visual responses in the superior colliculus. *Journal of Neurophysiology*, 86(5):2527–2542.
- Itti, L. and Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203.
- Knierim, J. and Van Essen, D. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology*, 67(4):961–80.
- Koch, C. and Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 4(4):219–27.
- Li, Z. (1999). Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proceedings of the National Academy of Sciences of the USA*, 96(18):10530–10535.
- Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, 6(1):9–16.
- Lur, G., Vinck, M. A., Tang, L., Cardin, J. A., and Higley, M. J. (2016). Projection-specific visual feature encoding by layer 5 cortical subnetworks. *Cell reports*, 14(11):2538–2545.

- Nuthmann, A., Clayden, A. C., and Fisher, R. B. (2021). The effect of target salience and size in visual search within naturalistic scenes under degraded vision. *Journal of Vision*, 21(4):article 2.
- Schiller, P. H. (1984). The superior colliculus and visual function. In Brookhart, J., Mountcastle, V., Darian-Smith, I., and Geiger, S., editors, *Handbook of Physiology, Section 1: The Nervous System; Vol. III, Sensory Processes, Part 1*, DOI:10.1002/cphy.cp010311, pages 457–505. Wiley Online Library.
- Shomstein, S. (2012). Cognitive functions of the posterior parietal cortex: top-down and bottom-up attentional control. *Frontiers in integrative neuroscience*, 6:38.
- Smith, M. A., Bair, W., and Movshon, J. A. (2002). Signals in macaque striate cortical neurons that support the perception of glass patterns. *Journal of neuroscience*, 22(18):8334–8345.
- Snowden, R. J. (1998). Texture segregation and visual search: A comparison of the effects of random variations along irrelevant dimensions. *Journal of Experimental Psychology: Human perception and performance*, 24(5):1354–1367.
- Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta psychologica*, 135(2):77–99.
- Treisman, A. M. and Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1):97–136.
- White, B. J., Kan, J. Y., Levy, R., Itti, L., and Munoz, D. P. (2017). Superior colliculus encodes visual saliency before the primary visual cortex. *Proceedings of the National Academy of Sciences*, 114(35):9451–9456.
- Wolfe, J. M. (2021). Guided search 6.0: An updated model of visual search. *Psychonomic Bulletin & Review*, 28:1060–1092.
- Yan, Y., Zhaoping, L., and Li, W. (2018). Bottom-up saliency and top-down learning in the primary visual cortex of monkeys. *Proceedings of the National Academy of Sciences*, 115(21):10499–10504.
- 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):article 1.
- Zhaoping, L. (2012). Gaze capture by eye-of-origin singletons: Interdependence with awareness. *Journal of Vision*, 12(2):article 17.
- Zhaoping, L. (2014). Understanding vision: theory, models, and data. *Oxford University Press*.
- Zhaoping, L. (2016). From the optic tectum to the primary visual cortex: migration through evolution of the saliency map for exogenous attentional guidance. *Current opinion in neurobiology*, 40:94–102.
- Zhaoping, L. (2019). A new framework for understanding vision from the perspective of the primary visual cortex. *Current Opinion in Neurobiology*, 58:1–10.

- Zhaoping, L. (2020). The flip tilt illusion: Visible in peripheral vision as predicted by the central-peripheral dichotomy. *i-Perception*, 11.
- Zhaoping, L. and Guyader, N. (2007). Interference with bottom-up feature detection by higher-level object recognition. *Current Biology*, 17(1):26–31.
- Zhaoping, L. and May, K. (2007). Psychophysical tests of the hypothesis of a bottom-up saliency map in primary visual cortex. *PLoS Computational Biology*, 3(4):e62.
- Zhaoping, L. and Snowden, R. (2006). A theory of a saliency map in primary visual cortex (V1) tested by psychophysics of color-orientation interference in texture segmentation. *Visual Cognition*, 14(4-8):911–933.