Eye movement evidence for V1 saliency hypothesis and central-peripheral dichotomy in an anomalous visual search task

Junhao Liang, Severin Maher and Li Zhaoping*

Eberhard Karls University of Tübingen and Max Planck Institute for Biological Cybernetics,
Tübingen, Germany
* li.zhaoping@tuebingen.mpg.de

In press in Vision Research (accepted 2 Aug 2023)

Abstract:
Typically, searching for a target among uniformly tilted non-targets is easier when this target is perpendicular, rather than parallel, to the non-targets. V1 Saliency Hypothesis (V1SH) – that V1 creates a saliency map to guide attention exogenously – predicts exactly the opposite in a special case: each target or non-target is a pair of equally-sized disks, a homo-pair of two disks of the same color, black or white, or a hetero-pair of two disks of the opposite color; the inter-disk displacement defines its orientation. This prediction – parallel advantage – was supported by the finding that parallel targets require shorter reaction times (RTs) to report targets' locations. Furthermore, it is stronger for targets further from the center of search images, as predicted by the central-peripheral dichotomy entailing that saliency effects are stronger in peripheral than in central vision. However, the parallel advantage could arise from a shorter time required to recognize – rather than to shift attention to – the parallel target. By gaze tracking, the present study confirms that the parallel advantage is solely due to the RTs for the gaze to reach the target. Furthermore, when the gaze is sufficiently far from the target during search, saccade to a parallel, rather than perpendicular, target is more likely, demonstrating the central-peripheral dichotomy more directly. Parallel advantage is stronger among observers encouraged to let their search be guided by spontaneous gaze shifts, which are presumably guided by bottom-up saliency rather than top-down factors.

Keywords: Saliency, Primary visual cortex, V1 saliency hypothesis (V1SH), Visual search, Eye movement, Central-peripheral dichotomy (CPD)

1. Introduction

The saliency of a visual location is defined as its ability to attract attention exogenously (Treisman & Gelade, 1980; Itti & Koch, 2001; Wolfe, 2021). V1 saliency hypothesis (V1SH) (Li, 1999a, 2002; Zhaoping, 2014) proposes that the saliency of a visual location is signaled by the highest V1 neural
response to this location, relative to the highest responses to other locations. Iso-feature suppression (Li, 1999a, 1999b) – whereby V1 neurons’ response to visual input is suppressed by neighboring V1 neurons preferring similar input features (Knierim & Van Essen, 1992) – is the neural mechanism underlying V1SH. For example, a vertical bar surrounded by other vertical bars is not salient since its evoked V1 response is under iso-orientation suppression by nearby V1 neurons responding to surrounding bars; this bar would be more salient when the surrounding bars are horizontal instead (by escaping such iso-orientation suppression).

V1SH is supported by behavioral and neurophysiological experiments confirming its predictions (see a review in Zhaoping, 2014). One surprising prediction is that an item uniquely shown to one eye among items shown to the other eye is salient (Zhaoping, 2008) through iso-eye-of-origin suppression (DeAngelis et al, 1994). For instance, during visual search for a non-horizontal bar among non-target horizontal bars, the observers’ gaze is often distracted by a non-target bar uniquely presented to the left eye whereas all the other bars are presented to the right eye (Zhaoping, 2012). This distraction is present even though this distractor’s unique eye-of-origin feature is typically invisible perceptually (Zhaoping, 2008, 2012, 2018). This invisibility arises because the eye-of-origin feature is not encoded by visual cortices beyond V1 (Hubel & Wiesel, 1968; Zeki, 1978; Burkhalter & Van Essen, 1986). This attentional capture demonstrates that attentional selection, typically by looking or gaze shifting, can occur before or without decoding, i.e., seeing or visual recognition (Zhaoping, 2014, 2019).

The current study is motivated by another V1SH prediction – parallel advantage – featuring a saliency mechanism characteristic of V1 (Zhaoping, 2022). Typically, finding a target among uniformly tilted non-targets is easier when the target is perpendicular, rather than parallel, to the non-targets (Fig. 1A) – this is the perpendicular advantage typically seen in visual search literature (Wolfe, 2021). V1SH predicts that parallel advantage occurs in a special case when a target or non-target is a pair of equally sized disks: a homo-pair of two disks of the same color, black or white, or a hetero-pair of two disks of opposite color (Fig. 1B). At the symbolic level, each disk-pair has an orientation defined by that of the displacement between the two disks and likely signaled in higher visual cortical areas. At V1 level, however, a homo-pair and a hetero-pair evoke markedly different responses, such that their most excited neurons prefer an orientation parallel and perpendicular to their symbolic orientations, respectively. This becomes apparent by considering that a black or a white disk activates an on- or off-subfield of a V1 simple cell (Fig. 1C), and is confirmed by neural data in monkey V1 (Smith et al., 2002). Consequently, by escaping iso-orientation suppression, a homo-pair is predicted by V1SH to be more salient when surrounded by hetero-pairs parallel rather than perpendicular to it (Fig. 1D, Zhaoping, 2022). If saliency is, as conventionally held (Gottlieb et al., 1998; Itti & Koch, 2001; Schall, 2004), computed in higher brain areas, which are presumably coding orientations at the symbolic level, perpendicular advantage would be predicted instead.

This predicted parallel advantage was tested by a visual search task, in which observers pressed a button to report the homo-pair target’s location as quickly as possible (Zhaoping, 2022). The reaction time (RT) for this button press, denoted as RT_report, was indeed shorter for the parallel target, supporting V1SH. However, since RT_report incorporates the time required for selection (attentional shift to the target), decoding (for target recognition), and any additional cognitive processes necessary such as decision-making, it is unclear whether the shorter RT_report for parallel targets originates from visual selection by saliency (to confirm the V1SH prediction) or from visual decoding (thereby not confirming the V1SH prediction).

Such a distinction between selection (looking) and decoding (seeing) has recently motivated the concept of a central-peripheral dichotomy (CPD, Zhaoping, 2017, 2019). Since typically looking selects saccade destination while seeing recognizes the object at the destination (typically in the center of gaze after the
saccade), CPD proposes that central and peripheral vision are specialized for decoding and selection, respectively. Therefore, visual saliency effects for selection should be stronger in the peripheral visual field. Indeed, the parallel advantage, as measured by $RT_{\text{report}}$, was stronger for targets further from the center of search images (Zhaoping, 2022). However, this finding assumes an equivalence between the display location and the retinal location of the target. This equivalence only holds approximately, since gaze shifts can easily bring the gaze closer to a target at a peripheral display location or further from a target at a central display location.

To address these limitations, the present study tracks gaze position during the same visual search task (Zhaoping, 2022), in order to confirm the parallel advantage and provide additional evidence for V1SH and CPD. First, to focus on visual selection for the V1SH prediction, we measure the reaction time $RT_{\text{gaze}}$ for the gaze to reach the target during search. By testing the parallel advantage in $RT_{\text{gaze}}$, we confirm that selection rather than decoding causes this advantage. Second, including all fixation locations during search trials, we show that the more peripheral or more central targets on the display are more likely to be more peripheral or more central on the retina, explaining why the CPD still holds even when tested using the display locations of the targets. Third, from the fixation point at the display center, where the retinal and display locations of the target are equivalent, we reveal a parallel advantage in the probability of the first saccade to reach the target during search. Fourth, this parallel advantage by the first saccade is stronger for more peripheral targets (on the retina and on the display, equivalently), consistent with CPD. Fifth, this CPD-supporting fourth result can be extended to all saccades during the search. Specifically, when the target is more peripheral ($>10^\circ$ eccentricity) or more central ($<5^\circ$ eccentricity) on the retina, the parallel or perpendicular target is more likely to attract gaze, respectively. In our visual search image with crowded image elements, it is difficult to recognize a target at $>10^\circ$ from gaze position due to visual crowding (Levi, 2008; Whitney & Levi, 2011). Meanwhile, in central vision, the perpendicular target appears self-evidently as more distinctive than a parallel target (Fig. 1B). Therefore, parallel versus perpendicular advantage of gaze attraction in peripheral versus central vision are manifestations of the saliency versus recognition effects, supporting the CPD.

In addition, we use a model to comprehend how the perpendicular advantage for central vision and the parallel advantage for peripheral vision combine to produce gaze dynamics during the visual search task. Our model gives simulated $RT_{\text{gaze}}$ that, in agreement with our $RT_{\text{gaze}}$ data, exhibit parallel advantage and confirm that this advantage is stronger for targets at peripheral display locations. In the behavior of our model and of our observers, the retinal location of the target is typically sufficiently peripheral to enjoy the parallel advantage. This is even for targets at central display locations, and is more so for targets at peripheral display locations.

The bottom-up saliency nature of the parallel advantage is further supported by our finding that this advantage is stronger among observers encouraged to let their search be guided by spontaneous gaze shifts, rather than any top-down strategy such as systematic scanning.
Fig. 1. Theoretical background for the parallel advantage predicted by V1SH. (A) A black shoe is easier to find when it is perpendicular rather than parallel to the uniformly oriented, surrounding, shoes of another type. (B) A black homo-pair is parallel or perpendicular to the surrounding hetero-pairs -- V1SH predicts the parallel advantage that the homo-pair is easier to find when it is parallel to the hetero-pairs. (C) Schematics of a V1 neuron’s Gabor-like receptive field (RF) with its on- and off-subfields; this V1 neuron should be better excited by a hetero-pair or homo-pair, respectively, when this pair is perpendicular or parallel to its preferred orientation. (D) Example RFs of V1 neurons (superposed on the visual stimuli in (B)) most activated by the disk-pairs in the images. Since V1 neurons are more suppressed by neighboring neurons tuned to the same orientation, the neuron responding to the homo-pair is less suppressed when the homo-pair is parallel, predicting a parallel advantage via V1SH.

2. Method

2.1 Equipment

A CRS (Cambridge Research Systems Ltd, Rochester, UK) Display++ monitor (32 inches) was used to display the visual stimuli. The screen resolution was 1920 × 1080 pixels, and the refresh rate was 120 Hz. The visual stimuli were controlled by custom-written software in Matlab R2021a (Math Works) and Psychtoolbox 3 3.0.17 (Brainard & Vision, 1997; Pelli & Vision, 1997), running on a host PC with 64-bit Ubuntu 20.04.2 LTS. The display screen was calibrated to ensure a linear gamma correction, and the gray background luminance was 49.6 cd/m² at the screen center in a dim room. The viewing distance was 65 cm. Eye movements were recorded with an Eyelink 1000+ (SR Research Ltd., Ontario, Canada) eye tracker at a sampling rate of 2000 Hz. The button responses of the observers were collected via a MilliKey MH-5 response box (LabHackers Research Equipment, Halifax, Canada).
2.2 Participants

A total of 32 observers (7 males and 25 females, ages: 18 to 69 with a mean of 26) participated in the study. Eye tracking data was collected from 30 of these observers (6 males and 24 females, ages: 19 to 34, with a mean of 24.5). All observers provided informed consent and were naïve as to the purposes of the study. They received payments or course credit for participation. All observers had normal or corrected-to-normal vision. The study was conducted according to the Declaration of Helsinki (1964), and ethical approval was obtained from the University Clinic Tübingen ethics committee.

2.3 Stimuli

Each search image contained 27 columns and 15 rows of disk-pairs, spanning 71 cm × 39.5 cm or about 57.3° × 33.8° in visual angle at a 65 cm viewing distance. Each pixel spanned ~ 0.03° at the display center. The target was equally likely to be in any row within the 3rd–13th rows and in any column within the 3rd–11th and 17th–25th columns. From the display center, the target was therefore between 6.7° and 25.8° in display eccentricity, and was within 6.7°–23.7° horizontally and 0°–11.4° vertically. The target homo-pair was equally likely to consist of two white or two black disks, and was tilted 45 degrees clockwise or counter-clockwise from vertical. The non-targets were hetero-pairs, each consisting of one white and one black disk. They were uniformly tilted, parallel or perpendicular to the target with equal probability. The black disk in each hetero-pair was randomly the upper or lower one with equal probability, independently from other hetero-pairs. The location of each disk-pair was randomly and uniformly jittered (independently from other disk-pairs) within ±3 pixels (~ ±0.09°), horizontally and vertically, from the regular 27 × 15 grid, so that two neighboring disk-pairs were between 64 pixels (~ 1.92°) and 76 pixels (~ 2.28°), and on average 70 pixels (~ 2.1°), apart center-to-center horizontally and vertically. Each disk had a diameter $D = 20$ pixels, and two disks in a pair were $R = 24$ pixels apart center-to-center, making a disk-pair span 1.32° × 0.6° in the display center. Thus, a hetero-pair should best excite V1 neurons preferring a spatial frequency of approximately 0.69 cycles/degree, within the range of preferred spatial frequencies of monkey V1 neurons for each target's retinal eccentricity within about 2–50° (Gattass et al., 1987). A gray background with a luminance of $L_0 = 49.6 \text{cd/m}^2$ was constant throughout an experimental session. Within a single disk, each pixel at a distance $r$ from the disk center had a luminance higher or lower than the gray background by approximately $L_0 \exp(-r^2/2\sigma^2)$ cd/m², where $\sigma = 7.2$ pixels.

2.4 Procedure

Before data taking for each observer, eye tracking calibration was performed to ensure that the average gaze tracking error was less than 1° at all target display locations.

The time course of a trial is as follows (Fig. 2A). After a button press by the observer to start a trial, a red fixation cross (spanning a diameter of 1.25°) appeared at the display center. Observers were instructed to look at the fixation cross in order to trigger the onset of the search image, which appeared once the eye tracker verified that the observer’s gaze had been within 2° from the center of the fixation cross for at least 0.4 seconds continuously. For the two observers who participated without eye tracking, the fixation cross
was displayed for a fixed duration of 0.7 seconds. After the fixation verification (or 0.7 seconds of fixation for the two observers without eye tracking), the fixation cross disappeared for 0.2 seconds before the search image appeared. Observers were instructed to freely move their eyes to search, and to press a left or right button as quickly as possible to report a target in the left or right half of the search display, respectively. Upon the button press, the trial was completed, and the search image was replaced by a text message, which signaled either a prompt for the observer to press a button for the next trial, a break in the trials, or the end of search trials.

After 4 ~ 10 practice trials, each observer performed 400 testing trials, with a short break between the 200th and the 201st trial. A standard drift check procedure provided by the Eyelink eye tracker was performed before the 201st trial to ensure that the tracking error was no more than 1° at the display center, otherwise a recalibration of the eye tracker was performed before proceeding.

Our thirty-two observers were divided into three groups according to the types of additional instructions for task strategies before the first practice trial. The first group of ten observers were not provided with any additional task strategy. The second group of eleven observers (one of them did not have their gaze tracked) were instructed to let their spontaneous gaze shifts guide their search. The third group of eleven observers (one of them did not have their gaze tracked) were not only instructed to let their spontaneous gaze shifts guide their search, but also given additional strategy instructions as follows. Firstly, they were told that having ‘spontaneous gaze shifts’ meant to not purposely control their gaze shifts, and were shown examples of purposeful control such as scanning through the search array line by line as if reading a text. Secondly, they were asked to be relaxed and to perform the search task as though playing a shooting game (to treat the target as an adversary, so that the observer should shoot at the target before the target shoots back). We refer to these three types of task strategy instructions as “no encouragement”, “some encouragement”, and “more encouragement” to rely on bottom-up saliency signals for their search task.

2.5 Saccade and fixation detection

The saccades and fixations of the eye movements in each trial were determined by an adaptive algorithm (Nyström & Holmqvist, 2010). This algorithm uses the statistics of the fluctuations of the recorded gaze positions during the duration of one trial, from the fixation cross onset to the button press, to obtain the gaze velocity thresholds used for identifying the saccades and fixations in each trial. Using the built-in Eyelink 1000+ algorithm (under ‘Psychophysical configuration’ parameters) to define saccades and fixations qualitatively gives the same results reported in this paper.
Fig. 2. The visual search experiment: design and quantities measured. (A) The procedure and time course of a trial. After a button press by the observer, a red central fixation cross appears. After the eye tracker verifies continuous (for 0.4 seconds) gaze fixating at the cross, a blank display is shown for 0.2 seconds before the search image appears and stays on until the subject's button press. The texts and the fixation cross are enlarged in this panel for clarity. (B) An example trial. Gaze locations later in time are visualized by darker brown colors (see panel C). The numbers 0, 1, …, 5 mark 0th, 1st, 2nd, …, 5th fixations, each marked by a blue or red square. The yellow dashed circles mark the fixation window to verify that the observer is fixating at the cross (which continues into the 0th fixation after the stimulus onset) before the stimulus onset and the target window for defining the target fixations (the 4th and 5th in this example). The 3rd fixation is the pre-target fixation, and the 4th fixation (red square) is the target landing fixation. (C) Time of the start and end of each fixation and button press in the example trial in (B) to define $RT_{\text{report}}$, $RT_{\text{gaze}}$, and $RT_{\text{lapse}}$.

2.6 Performance and eye movement analysis

We define a target window as within 2.5° from the target’s center, and fixations within this window are defined as target fixations. For example, the 4th and 5th fixations in Fig. 2B are target fixations. Results reported in this paper are qualitatively the same by choosing a target window with radius 1.5° – 3°. If observers completed a trial by a button press to report target’s location without any fixation inside the target window, then there is no target fixation in this trial.
We define a target landing fixation as a target fixation not preceded by another target fixation. For example, the 4th fixation in Fig. 2B is the target landing fixation. The fixation immediately before a target landing fixation, e.g., the 3rd fixation in Fig. 2B, is defined as a pre-target fixation. We define $RT_{gaze}$ as time between the stimulus onset and the first target landing fixation, $RT_{report}$ as the time between the stimulus onset and the button press, and $RT_{lapse} = RT_{report} - RT_{gaze}$ (Fig. 2C). Trials with incorrect button press (when observers pressed a left/right button for a target in the right/left half of the display) or $RT_{report} < 0.2$ seconds are excluded from the analysis of the reaction times. They comprise 0-5% of the trials among all observers.

For each fixation, we measured its distance to the target’s center as the target’s retinal eccentricity $e$ relative to the gaze position. Across all the trials of each observer for a given target type (parallel or perpendicular), we obtain

\[ n(e) = \text{number of pre-target fixations of retinal eccentricity } e \text{ across all the trials,} \]

\[ N(e) = \text{number of all fixations of retinal eccentricity } e \text{ across all the trials,} \]

and then the probability $P_{\text{saccade-to-target}}(e)$ that the next fixation, via a saccade, reaches the target (i.e., enters the target window) as a function of $e$, is

\[ P_{\text{saccade-to-target}}(e) = \frac{n(e)}{N(e)} \quad (1) \]

In practice, $P_{\text{saccade-to-target}}(e)$ is calculated using $e$ within a range, e.g., 5°– 10°, to gather sufficient statistics for $n(e)$ and $N(e)$.

When we presume that in trials without target fixations, a target fixation would occur after the button press, then the probability in equation (1) is revised to

\[ P_{\text{saccade-to-target}}(e) = \frac{n(e) + m(e)}{N(e)} \quad (2) \]

in which $m(e)$ is the number of trials (of the given target type) without target fixations where the last fixation had retinal eccentricity $e$.

2.7 Statistical analysis

To test for a significant difference between two quantities, we used two-tailed paired permutation tests across observers on their across-trial statistics (median RT across trials, the probability of saccades towards targets, or percentage of trials without target fixations); hence, our tests require no assumptions regarding the distribution of statistics of the observed quantities.

We used median RT ($RT_{\text{report}}, RT_{gaze}, RT_{\text{lapse}}$) values across trials for each observer before averaging across observers for the results in our figures. By using the median RTs, we avoid removal of RT outliers (as RT distributions have long tails) using arbitrary thresholds while preserving the statistical nature of our RT data.
For results in our figures, the total number of possible permutations is very large ($2^{32}$ for $RT_{report}$ results for $N = 32$ observers, or $2^{30}$ for other eye movement results for $N = 30$ observers); thus we randomly selected $10^5$ permutations out of all possible permutations to compute the $p$ value in each test.

### 2.8 Data-driven simulated experiments by a Monte Carlo model

To understand how the gaze dynamics leads to $RT_{gaze}$, we carried out data-driven Monte Carlo simulated experiments for $RT_{gaze}$ based on statistical properties of gaze dynamics in our $N = 30$ observers whose eye tracking data were collected. The particular statistical properties used for the simulation are: the distribution $P_{fix}(T)$ of fixation duration $T$ during search, the distribution $P_{sac}(A, \theta)$ of the saccadic amplitude $A$ and saccadic direction in terms of the polar angle $\theta$, and the probability $P_{saccade-to-target}(e)$ for a saccade that reaches the target at retinal eccentricity $e$ given by equation (1).

Each simulated search trial contains the following steps in sequence:

1. A target location is chosen as from our target locations in the experiment. The target is randomly assigned as parallel or perpendicular. The starting fixation location is at the center of the display.

2. Given a fixation location outside the target window, generate the fixation duration $T$ as a random sample from $P_{fix}(T)$, and calculate the target’s retinal eccentricity $e$.

3. Following 2), generate the next saccade, such that with probability $p = P_{saccade-to-target}(e)$ it arrives inside the target window, with a saccadic amplitude $A = e$. With a probability $1 - p$, the saccade arrives outside the target window, with saccadic amplitude $A$ and direction $\theta$ randomly generated from $P_{sac}(A, \theta)$; and this random generation process is repeated unless the saccadic destination is within the search display and outside the target window. Saccadic duration is calculated as $t = 2.2 \times A + 27$ ms, where the saccadic amplitude $A$ is in degree of visual angle, following an empirical formula (Collewijn et al., 1988). The saccadic destination is assigned as the next fixation location.

4. Following 3), if this fixation is inside the target window, accumulate all the $T$ and $t$ across the previous fixations and saccades to obtain the simulated $RT_{gaze}$ denoted as $RT_{gaze}^{simul}$; if this fixation is outside the target window, go back to 2).

Note that $P_{saccade-to-target}(e)$ in 3) depends on whether the target is parallel or perpendicular, according to our data. Our simulation used a simple model of $P_{sac}(A, \theta) = P_{sac}(A) \times P_{sac}(\theta)$, where $P_{sac}(A)$ is according to our data and $P_{sac}(\theta)$ is a uniform distribution across $\theta$. Our qualitative results still hold (while reducing quantitatively the overall $RT_{gaze}^{simul}$ by about 300 ms) when $P_{sac}(\theta)$ is such that $\theta$ has a probability $q \approx 0.55$ (as in our data, $q = 0.55 \pm 0.029$ and $0.549 \pm 0.026$ (mean $\pm$ std.), respectively, for parallel and perpendicular targets) to be within 90 degree from the polar angle of the target relative to the current fixation location.

A total of $N_{trial} = 400$ simulated trials were simulated for a simulated observer. This was then repeated for $N = 30$ simulated observers to give a simulated experiment, with $N_{trial}$ and $N$ values exactly as in our real experiment. A total of $N_{sim} = 1000$ simulated experiments were carried out to obtain the distributions of the $p$ values.
3. Results

The saliency of the target can be quantified by the behavior reaction time and the eye movement patterns of the observers during the search.

3.1 Parallel advantage is due to visual selection rather than visual recognition processes

We found a shorter reaction time $RT_{gaze}$ in visual selection (Fig. 3A, $p < 10^{-4}$) when the target was parallel rather than perpendicular to non-targets. Furthermore, this parallel advantage was stronger for targets at peripheral display locations (Fig. 3C, $p < 10^{-3}$), compared to targets at central display locations (Fig. 3B, $p = .06$). A target is defined to be at a central or peripheral display location, respectively, when its distance to the image center was smaller or larger than the average of such distances for all target locations. The reaction time of the button press to report the target, $RT_{report}$, shows a slightly weaker (than $RT_{gaze}$) but significant parallel advantage ($p < 10^{-4}$, $p = .40$, $p < 10^{-2}$ for the $RT_{report}$ comparison in Fig. 3A, B, C). In contrast, the latency from the gaze reaching the target to the button press, $RT_{lapse}$, was shorter for perpendicular targets (Fig. 3A, $p < 10^{-2}$). This perpendicular advantage was more significant for targets at central display locations (Fig. 3B, $p < 10^{-2}$) than those at peripheral display locations (Fig. 3C, $p = .051$).

Since $RT_{lapse}$ is a manifestation of visual recognition rather than selection, our data demonstrate that the parallel advantage in our task reaction time $RT_{report}$ is due to visual selection manifested by $RT_{gaze}$. The stronger parallel advantage for the more peripheral targets on the display suggests a central-peripheral dichotomy. Fig. 3D shows a statistical tendency that the more peripheral targets on the display are more likely to be more peripheral on the retina during search before the gaze reaches the target. This explains why the central-peripheral dichotomy, which applies strictly to the retinal locations, still holds even when tested using the display locations. However, Fig. 3D also reveals that, in a minority of trials, peripheral targets on the display could have more central retinal locations and vice versa. Hence, we examine next the central-peripheral dichotomy based strictly on the retinal locations of the target. To clearly distinguish display versus retinal positions of the target, we use the terms display versus retinal eccentricities of the target. Furthermore, when the fixation is at the central display location at the start of a search trial, display and retinal locations (or eccentricities) are equivalent to each other.
Fig. 3. Reaction times $RT_{\text{report}}$, $RT_{\text{gaze}}$ and $RT_{\text{lapse}}$ (for a button press to report the target, gaze arrival to the target, and their difference $RT_{\text{lapse}} = RT_{\text{report}} - RT_{\text{gaze}}$, in the visual search task) and target’s retinal eccentricity $e$ in support of the parallel advantage and central-peripheral dichotomy. These reaction times are compared between parallel and perpendicular targets at all (A), central (B), or peripheral (C) display locations. Median reaction times across trials for individual observers are averaged across $N = 30$ observers for $RT_{\text{gaze}}, RT_{\text{lapse}}$ and $N = 32$ observers for $RT_{\text{report}}$ for these results. (D) The probability distribution of the target’s retinal eccentricity $e$ averaged across fixations before the gaze reaches the target in a trial, for targets at central and peripheral display locations respectively. The curves are the average distributions over $N = 30$ observers, and the shaded areas represent the standard deviation. The inset plots analogously the distributions of the number of saccades needed to reach the target in a trial. In this figure, a ‘n.s.’, ‘*’, ‘**’, or ‘***’ above a black line linking two data quantities indicates the statistical test for equivalence between the two quantities yields a $p$ value of $p \geq 0.05$, $10^{-2} \leq p < 0.05$, $10^{-3} \leq p < 10^{-2}$, $p < 10^{-3}$, respectively. Error bars represent the standard errors.
3.2 Central-peripheral dichotomy in target propensity to attract gaze

Next, we show a dichotomy between central and peripheral vision in the gaze attraction of the target: the parallel target is more likely to attract gaze than a perpendicular target in peripheral vision when the gaze is far from the target; in contrast, the perpendicular target is more likely to attract gaze than the parallel target in central vision when the gaze is close to the target.

First, the first saccade during a search trial was more likely to reach the target when the target is parallel rather than perpendicular (Fig. 4A, $p = 0.042$ including all targets). This saccade starts from the central fixation point when the target’s retinal positions are equivalent to their display positions. This parallel advantage in the first saccade was stronger for targets at larger retinal (= display) eccentricities. The significance of this advantage has a $p$-value $p < 0.01$ for targets at retinal (= display) eccentricities $> 10^\circ$, which constitute 89.9% of all trials. This confirms CPD’s prediction that the parallel advantage is stronger for more peripheral targets on the retina.

Extending our analysis from the first to all saccades in a trial, we examine the probability that the next fixation reaches the target via a saccade, $P_{\text{saccade-to-target}}$ (equation (1)), at various target retinal eccentricities $e$. This $P_{\text{saccade-to-target}}$ is higher for parallel than perpendicular targets when $e > 10^\circ$ (Fig. 4B, $p < 10^{-2}$ for $e$ within $10^\circ - 20^\circ$ and $p < 10^{-3}$ for $e > 20^\circ$). In contrast, this $P_{\text{saccade-to-target}}$ is higher for perpendicular than parallel targets when $e < 5^\circ$ (Fig. 4B, $p = .027$ for $e$ within $2.5^\circ - 5^\circ$).

This higher probability $P_{\text{saccade-to-target}}$ of making large saccades for parallel targets is one reason for a shorter reaction time $RT_{\text{gaze}}$ for the gaze to arrive at the target (Fig. 3). One reason for a shorter $RT_{\text{report}}$ is the shorter $RT_{\text{gaze}}$, the main component of $RT_{\text{report}}$. Another reason is that observers were marginally more likely to complete the trial without any target fixations when the target is parallel rather than perpendicular ($p = .048$, Fig. 4C). In these no-target-fixation trials, observers apparently made a fast decision on the target using peripheral vision (i.e. the gaze is outside the target window). This parallel advantage in such fast decisions appears to be stronger for more peripheral targets on the retina (Fig. 4C).

In no-target-fixation trials, if we presume a target fixation after the button press, then the last fixations in these trials should be included as pre-target fixations so that $P_{\text{saccade-to-target}}$ should be revised to follow equation (2). This revised $P_{\text{saccade-to-target}}$ strengthens the parallel advantage for target to attract gaze at large retinal eccentricities $e$ and weakens the perpendicular advantage for the target to attract gaze at small retinal eccentricities $e$ (see the $p$ values in Fig. 4B). Fewer gaze attractions by the parallel than perpendicular targets at small retinal eccentricities (Fig. 4B, the data points for current fixations with $e$ within $2.5^\circ - 5^\circ$ from the target) is partly due to a stronger tendency (although statistically insignificant, Fig. 4C, the data point in the same range of $e$ within $2.5^\circ - 5^\circ$) for parallel target trials to be no-target-fixation trials.
Fig. 4. Central-peripheral dichotomy in gaze attraction by targets as a function of target’s retinal eccentricity. (A) The probability that the first saccade in each trial reaches the target versus target’s retinal eccentricity (display eccentricity); and the percentage (orange bars) of such targets among all targets. (B) The probability that the next saccade reaches the target versus target’s retinal eccentricity at the current fixation, for all fixations during search, using equation (1). The inset shows the revised probability values by including trials without target fixations using equation (2), based on the presumption that a saccade following the button press would reach the target. (C) The percentage of trials without target fixations (among all parallel or perpendicular target trials) versus the target’s retinal eccentricity at the last fixation in such trials. Results are averaged across $N = 30$ observers. Note that the y-axis is logarithmic. Here, a ‘n.s.’, ‘*’, ‘**’, or ‘***’ (with its accompanying $p$ value) indicates that the statistical test for
equivalence between a red and a blue data points under this symbol yields a $p$ value $p \geq 0.05$, $10^{-2} \leq p < 0.05$, $10^{-3} \leq p < 10^{-2}$, $p < 10^{-3}$, respectively. Error bars represent the standard errors.

### 3.3 Observed parallel advantage and central-peripheral dichotomy in reaction times can be understood by stochastic gaze dynamics guided by saliency and task: a model simulation

On the one hand, there is a parallel advantage in reaction times, particularly for peripheral display targets (Fig. 3). On the other hand, parallel advantage by gaze attraction is only for peripheral retinal targets, which have a tiny probability to attract gaze, compared to much stronger gaze attraction by more central retinal targets (Fig. 4B). To comprehend this, we use a simulation model (Fig. 5B) to generate the gaze trajectories on search stimuli, based on the stochastic properties of gaze dynamic, including the gaze attraction by the targets (Fig. 4B), according to our data.

Specifically, the simulation model (see section 2.8) is formulated according to empirical data as follows. First, each gaze shift on the search display is stochastic, guided by target’s gaze attraction by saliency and task goal (Wolfe, 2021) (as $P_{\text{saccade-to-target}}$ in Fig. 4B) and driven by probability distributions of fixation durations and of saccadic amplitudes in our data (as in $P_{\text{fix}}(T)$ and $P_{\text{sac}}(A)$ in Fig. 5A). Secondly, the direction and amplitude of each gaze shift are approximated as independent of those of the previous gaze shifts, since visual search appears to have little memory from one saccade to another (Horowitz & Wolfe, 1998). Other than these approximations, we employ no free model parameters. The simulated $RT_{gaze}$, denoted as $RT_{gaze}^{\text{simul}}$, is simply the time taken by the simulated gaze trajectory, starting from the stimulus onset to the moment when the simulated gaze position enters the target window, using the same target window size as in our real data analysis.

The left part of Fig. 5C plots the results of a simulated experiment containing $N = 30$ simulated observers, each having 400 trials. These plots are in the same format as that of Fig. 3 which plots the results from our real experimental data. They show a shorter reaction time $RT_{gaze}^{\text{simul}}$ for parallel than perpendicular targets (Fig. 5C, top panel, $p < 10^{-2}$ for over 90% of the simulations). This parallel advantage in RT arises solely from the dependence of $P_{\text{saccade-to-target}}$ on whether the target is parallel or perpendicular. Furthermore, our simulation shows a CPD: this parallel advantage is stronger for targets at peripheral display locations (Fig. 5C, bottom panel, $p < 10^{-3}$ for over 50% of the simulations and $p > 0.05$ for less than 10% of the simulations), compared to targets at central display locations (Fig. 5C, middle panel, $p < 10^{-3}$ for less than 10% of the simulations and $p > 0.05$ for over 50% of the simulations). This CPD character arises from the dependence of $P_{\text{saccade-to-target}}$ on the target’s retinal eccentricity, and from the tendency for more saccades being required to reach a more peripheral display target. A tiny parallel advantage in $P_{\text{saccade-to-target}}$ for targets with retinal eccentricity $e > 10^\circ$ can be amplified through multiple saccades to achieve the parallel advantage in $RT_{gaze}$, and targets requiring more saccades to reach them enables more chances to accumulate this amplification. For targets at smaller retinal eccentricities $e < 10^\circ$, closer to the fixation location, $P_{\text{saccade-to-target}}$ has a parallel disadvantage (significantly for $e < 5^\circ$, Fig. 4B). Apparently, this disadvantage is not effectively amplified since fixation locations are pre-dominantly at retinal eccentricity $e > 10^\circ$ in typical search trials. This is the case in our data (Fig. 3D) as well as our simulation model (Fig. 5C).
The simulated results in Fig. 5 were based on a simplified model in which each saccade, if not reaching target, is randomly directed regardless of the target location. In our real data, such saccades that do not reach the target have a small bias to decrease rather than increase the target’s retinal eccentricity. This bias is not significantly dependent on whether the target is parallel or perpendicular, but, when included in our simulation, does help to shorten the reaction time quantitatively by about 300 ms (see section 2.8). This improvement in quantitative agreement between the simulated $RT_{gaze}^{simul}$ and the actual $RT_{gaze}$ does not change the qualitative agreement between the model and data. Nevertheless, it suggests that the key mechanisms for the parallel advantage and CPD lie within the attentional guidance $P_{saccade-to-target}$ by the target. This search guidance (Wolfe, 2021) includes bottom-up saliency mechanisms (particularly in the parallel advantage for larger $e$) and top-down task-driven factors (particularly for small $e$), and manifests our CPD.

3.4 Encouraging observers to use bottom-up strategies enhances the parallel advantage

We compare the results of three groups of observers that received task strategy instructions with ‘no encouragement’, ‘some encouragement’, and ‘more encouragement’ toward using a bottom-up strategy for the search task (see details in section 2.4 Procedure). The shorter $RT_{gaze}$ for parallel than perpendicular targets (Fig. 6A) is significant for observers with ‘some encouragement’ ($p = 0.01$) and ‘more encouragement’ ($p = 0.01$) but not for those with ‘no encouragement’ ($p = .08$). The shorter $RT_{lapse}$ for perpendicular than parallel targets (Fig. 6B) is significant for observers with ‘more encouragement’ ($p = 0.02$) but not for those with ‘no encouragement’ ($p = .71$) or ‘some encouragement’ ($p = 0.13$). The shorter $RT_{report}$ for parallel than perpendicular targets (Fig. 6C) is significant for observers with ‘more encouragement’ ($p < 10^{-2}$) but not for those with ‘no encouragement’ ($p = .19$) or ‘some encouragement’ ($p = 0.11$). Moreover, the higher probability $P_{saccade-to-target}$ of gaze attraction by parallel targets at larger retinal eccentricity $e$ (Fig. 6D) is significant for observers with ‘some encouragement’ ($p = .023$ for $e$ within $10^\circ - 20^\circ$) and ‘more encouragement’ ($p < 10^{-2}$ for $e > 20^\circ$) but not for observers with ‘no encouragement’. Finally, the probability for a trial without target fixations is significantly higher for parallel than perpendicular targets (Fig. 6E) for observers with ‘more encouragement’ ($p = .039$), but not for observers with ‘no encouragement’ ($p = .17$) or ‘some encouragement’ ($p = .710$).

Regardless of whether the target is parallel or perpendicular, different task instructions do not cause any significant difference in any RT ($RT_{gaze}$, $RT_{lapse}$, or $RT_{report}$, $p > 0.2$ by two-sample t-tests).
Fig. 5. Guided search model simulation of the reaction time $R_{\text{gaze}}^{\text{simul}}$ for the gaze to arrive at the target, in support of the parallel advantage and central-peripheral dichotomy. (A) The statistical characteristics of gaze dynamics (from our data) used in the simulation: the distribution $P_{\text{fix}}(T)$ of fixation durations $T$ during search, the distribution $P_{\text{sacc}}(A)$ of saccadic amplitude $A$ during search (the black curves show the average over $N = 30$ observers and the shaded areas represent the standard deviation), and the probability that the next fixation (via a saccade) reaches the target $P_{\text{sacc-to-target}}(e)$ versus the target’s retinal eccentricity $e$, using equation (1) (same as in Fig. 4B). (B) Steps in the data-driven simulation of the gaze trajectory to arrive at $R_{\text{gaze}}^{\text{simul}}$ in a simulated trial (see section 2.8). (C) Simulated $R_{\text{gaze}}^{\text{simul}}$ are compared between parallel and perpendicular targets at all, central, or peripheral display locations. The left part shows the result of an example simulated experiment with $N_{\text{trial}} = 400$ trials repeated for $N = 30$. 

$\text{(A)}$ Statistical properties of gaze dynamics in experiment data

$\text{(B)}$ Simulation of a trial using $P_{\text{fix}}(T)$, $P_{\text{sacc}}(A)$, $P_{\text{sacc-to-target}}(e)$ in $\text{A}$

$\text{(C)}$ Simulated reaction time $R_{\text{gaze}}^{\text{simul}}$ in $\text{B}$ for the gaze to reach the target.
simulated observers. The $p$ values in this example simulated experiment are marked by the arrows pointing to the right part, where the cumulative distribution of the $p$ values across $N_{\text{sim}} = 1000$ simulated experiments are shown. The dotted lines mark the cumulative probability values 0.1 and 0.9, and the grey shaded areas mark the ranges for $p < 0.05$. In this and the next figure, the meanings of the symbols ‘n.s.’, ‘*’, ‘**’, and ‘***’ follow the same conventions as those in previous figures. Error bars represent the standard errors.

**Fig. 6.** Stronger parallel advantage by stronger encouragement to observers for using bottom-up strategy in the search task. Comparison between parallel and perpendicular targets in $RT_{\text{gaze}}$ (A), $RT_{\text{lapse}}$ (B), $RT_{\text{report}}$ (C), $P_{\text{saccade-to-target}}$ (D), and the percentage of trials without target fixations among all parallel or perpendicular target trials (E) are plotted separately for three groups of observers (E) with ‘no encouragement’, ‘some encouragement’, and ‘more encouragement’ for using bottom-up strategy in the search task (see 2.4 Procedure). The results are averaged across $N = 10$ observers in each group. Note that the y-axis is logarithmic (D). Error bars represent the standard errors.
4. Discussion

In summary, we strengthen the previous evidence for the V1 saliency hypothesis (V1SH) and the central-peripheral dichotomy (CPD) by tracking gaze in a visual search task using stimuli diagnostic of bottom-up neural substrates in V1. Specifically, we confirm the parallel advantage predicted by V1SH – that a homo-pair target is more salient when it is parallel rather than perpendicular to the hetero-pair non-targets – by showing a shorter reaction time for the gaze to reach the parallel target during the visual search. This extends the previous finding by a shorter reaction time to report the parallel target (Zhaoping, 2022). In addition, we find that the parallel and perpendicular target is more likely to attract gaze during the search when the target is in the peripheral (> 10° retinal eccentricity) and central (< 5° retinal eccentricity) visual field, respectively, manifesting CPD. Through a simple model, our data can be understood by attentional shifts guided by bottom-up V1 saliency mechanisms (to make a parallel target more salient particularly for peripheral targets) and non-saliency factors such as top-down task goal and target recognition (particularly for central vision where a perpendicular target is self-evidently more distinctive than a parallel target). Furthermore, the parallel advantage can be enhanced by instructing observers to use bottom-up strategies of letting their spontaneous gaze shifts guide their search, consistent with saliency mechanisms as the cause of this advantage.

4.1 Looking versus seeing

Gaze shifts guided by bottom-up saliency should, in principle, occur without recognizing the object at the saccade destination. Indeed, visual search data (Zhaoping & Guyader, 2007; Zhaoping & Frith, 2011; Zhaoping, 2012; Nuthmann, 2014; Nuthmann et al., 2021) indicate that gaze shifts to a target – looking – and the subsequent target recognition – seeing – are dissociable. Therefore, an experimental test of a prediction for bottom-up saliency should focus on $RT_{gaze}$, the reaction time for looking, rather than $RT_{report}$, the reaction time to achieve both looking and seeing. Our finding of a shorter $RT_{gaze}$ for parallel targets (Fig. 3A) provides a more precise confirmation of the predicted parallel advantage.

In addition, focusing on seeing, $RT_{lapse}$, the latency from arrival of gaze at the target to the button press report, is longer for parallel targets (Fig. 3A). Further exploration (data not shown) indicates that this is largely due to trials in which the gaze abandoned the target to continue searching elsewhere after reaching the target. This target abandonment is more likely for parallel targets and may be due to a failed target recognition. More details on target recognition and eye movements in this task will be explored in another study.

The distinction between looking and seeing, by peripheral vision and central vision, respectively, is perhaps more evident in a previous visual search study (Zhaoping & Guyader 2007). In that study, the target was difficult to recognize in peripheral vision due to crowding by non-targets, and its shape recognition by central vision after the gaze arrives at the target often triggers a target veto, manifested as target abandonment, since the target is a rotated or reflected version of the non-targets and is therefore likely confused as a non-target due to the rotational or reflection invariance of shape recognition (see also Zhaoping & Frith, 2011; Becker et al., 2023).
4.2 Bottom-up versus top-down attentional guidance and their neural substrates

The orientation of the target relative to the non-targets is task-irrelevant and random in each trial. Therefore, the parallel advantage is a bottom-up effect.

Bottom-up effects are transient (Donk & Van Zoest, 2008) and often masked in behavioral manifestations by top-down factors (Einhäuser et al., 2008) such that gaze positions are often better predicted by object and scene properties rather than saliency. In our search image, disk-pairs were arranged according to a regular grid; this scene property is likely to encourage searching by top-down strategy such as row-by-row searches. As observers are guided by both bottom-up and top-down factors during their search, our strategy instruction aims to reduce the top-down factors, thereby enhancing the relative weight by the bottom-up guidance. Indeed, bottom-up effects as manifested in parallel advantage are better unmasked with instructions encouraging observers to avoid using top-down strategies (Fig. 6). Monkey neurophysiological data also suggest that bottom-up saliency is manifested briefly in early responses of V1, V4, or LIP neurons before top-down or task-dependent factors dominate the neural responses (Bisley & Goldberg, 2010; Yan et al., 2018; Klink et al., 2023).

Our data suggest that the preferred orientations of the activated V1 neurons, rather than the target orientation at the symbolic level (i.e., defined by the displacement between the two disks), dictate saliency. Adding to previous studies of human psychophysics (Zhaoping & Snowden, 2006; Koene & Zhaoping, 2007; Zhaoping & May, 2007; Jingling & Zhaoping, 2008; Zhaoping, 2008; Zhaoping & Zhe, 2015), human functional magnetic resonance imaging and electroencephalograms (Zhang et al., 2012), and monkey neurophysiology (Hupé et al., 2001; Self et al., 2013; Yan et al., 2018), our study provides additional evidence from human eye movement behavior supporting V1SH.

Traditionally, a saliency map to guide attention exogenously is thought to be generated in higher brain areas (Itti & Koch, 2001; Schall, 2004), such as the frontal eye field (FEF) (Thompson & Bichot, 2005) and lateral intraparietal cortex (LIP) (Gottlieb et al., 1998). Monkey studies suggest that neurons in FEF (Bichot et al., 1996) and LIP (Freedman & Assad, 2006) are only tuned to task-relevant visual features when the animals have been extensively trained. Those neurons are typically untuned to task-irrelevant visual features. Their receptive fields are also much larger than the targets and non-targets used in this study.

The superior colliculus (SC) is a midbrain area heavily involved in eye movement (Gandhi & Katnani, 2011) and attentional control (Kustov & Lee Robinson, 1996; McPeek & Keller, 2004). It receives the V1 saliency signals and can utilize them in executing attentional shifts to the most salient visual locations (Zhaoping, 2014). Visual feature tuning for SC neurons is thought to be very weak (Horwitz & Newsome, 2001) (but see Chen et al., 2018 for a recent finding of modest tunings to orientation and spatial frequency in SC). White et al. (2017) reported that saliency signals by an orientation contrast appeared first in SC before V1, even though the initial visually evoked responses in V1 neurons have a shorter latency than those in SC neurons (White et al., 2017, Yu et al., 2023). Meanwhile, in disagreement with White et al (2017), Yan et al (2018) observed saliency signals in V1 neurons during the initial peak responses, suggesting that the saliency signals in V1 has a shorter latency than that reported by White et al (2017). This inconsistency between different studies needs to be resolved by future studies.

Compared with the brain areas FEF, LIP, or SC, V1 is most selective to the orientation feature formed by the small disk-pairs in our search images. Note that the saliency signal in V1 needs to be read out in order to conduct eye movements, and the feature tuning in V1 responses could be degraded progressively along the read-out pathway. Thus, the weaker feature selectivity in SC compared to V1 may be due to inheriting
and degrading feature selectivity from V1 when reading out the saliency map. Moreover, V4 (Reynolds & Desimone, 2003; Burrows & Moore, 2009; Klink et al., 2023), parietal (Bisley & Goldberg, 2010; Shomstein, 2012) and frontal (Thompson & Bichot, 2005) areas could also inherit saliency signals from V1 to combine them with top-down factors for controlling attention. These observations suggest that V1 is the neural substrate for the counter-intuitive saliency effect observed in our results.

The discussions above are restricted to primates. In other animal species, neural substrates for bottom-up visual saliency may be in SC, which is called optic tectum in non-mammals (Zhaoping, 2016), particularly since non-mammals such as fish lack neocortex (containing V1) while SC or optic tectum is relatively conserved through evolution. Indeed, in rats, visual response latencies in SC are about the same as those in V1 if not shorter (Li et al, 2015). In monkeys, lesioning V1 eliminates visually guided saccades for about two months on average (Isa & Yoshida, 2021). This suggests that direct retinal inputs to SC is insufficient for exogenous attentional selection in normal situations in primates. In comparison, rodents are much less affected by V1 lesions (Isa & Yoshida, 2021). In V1 lesioned monkeys, recovery of visually guided saccades after two months, by training, is perhaps due to a return to an evolutionarily old neural circuit that relies more on the retina-SC pathway than the retina-LGN-V1 pathway for exogenous attentional guidance (Zhaoping, 2016; Isa & Yoshida, 2021). Indeed, from lower to higher mammals, the percentage of retinal ganglion cells that project to SC decreases from almost 100% in rabbits and rats, and at least 70% in mouse, to about 59% in cats and about 10% in primates (see Zhaoping, 2016 for a review of the evolutionary trend).

### 4.3 Central-peripheral dichotomy (CPD)

The task reaction time $RT_{\text{report}} = RT_{\text{gaze}} + RT_{\text{lapse}}$ contains the duration $RT_{\text{gaze}}$ and $RT_{\text{lapse}}$ before and after the gaze reaches the target. $RT_{\text{gaze}}$ is shorter for parallel targets, whereas $RT_{\text{lapse}}$ is shorter for perpendicular targets (Fig. 3A), suggesting that looking and seeing are distinctive processes. Importantly, the parallel advantage in $RT_{\text{gaze}}$ is stronger when the target is at peripheral display locations, whereas the perpendicular advantage in $RT_{\text{lapse}}$ is stronger when target is at central display locations (Fig. 3B, C). Furthermore, gaze attraction is relatively stronger or weaker from parallel, rather than perpendicular, targets, at large or small retinal eccentricities, respectively (Fig. 4B).

These central-peripheral differences cannot be explained by the increase of the average size of V1 neural receptive fields (RFs) with the retinal eccentricity. For multiscale coding (Zhaoping, 2014), the radius of individual RFs can vary by a factor of 4 or more at each retinal location (Gattass et al., 1987). The size of our search items is such that they can effectively activate some V1 neurons covering each retinal location between 2 to 50 degrees of retinal eccentricity.

Instead, our experimental findings support the central-peripheral dichotomy that central and peripheral vision are specialized for seeing and looking, respectively (Zhaoping, 2014, 2017, 2019; Zhaoping & Ackermann, 2018; Nuthmann, 2014; Nuthmann et al., 2021), and that top-down feedback to aid seeing is mainly directed to central vision (Zhaoping, 2017, 2019; Zhaoping & Ackermann, 2018). A target closer to the fovea is more likely to lie within the attentional spotlight, and thus is more subject to top-down factors than saliency. In contrast, a target in peripheral vision is more vulnerable to visual crowding (Levi, 2008; Whitney & Levi, 2011) impeding its recognition, but can attract gaze by its saliency (Zhaoping & Guyader, 2007). The stronger parallel advantage at more peripheral areas of the visual field is consistent with the idea that saliency mechanisms are stronger in peripheral vision (Zhaoping, 2014). Indeed, gaze...
attraction by an ocular singleton that is distinctive only by a perceptually invisible feature – eye of origin of visual inputs – is stronger at larger retinal eccentricities (Zhaoping, 2012).

Central-peripheral dichotomy (CPD) in primate vision can be extended to multisensory processing, so that, for example, the peripheral visual field is generalized to peripheral sensory field that includes not only the peripheral visual field but also the other sensory field for sound, touch, small, etc. For example, a sound source can direct looking, i.e., attracting gaze exogenously or endogenously, placing an initially peripheral sensory location to the central sensory location, the fovea, for seeing. CPD can also be extended to other animal species across the animal kingdom (Zhaoping, 2019, 2023). Looking and seeing in primates are thus generalized to orienting and recognition in general by the corresponding peripheral and central senses, respectively. Common across species is an attentional bottleneck in each brain limiting processing resources shared across senses, such that an animal cannot process all sensory input information and must select the most important sensory input for deeper processing. Accordingly, sensory selection can also involve orienting, for example, head, limbs, whiskers, tentacles, snout, and/or ears, instead of gaze. Recognition after the selection can involve, for example, microvibrissae for mice, nose for dogs, lips and tongue for human infants, and/or acoustic fovea for echolocating bats, instead of retinal fovea which is lacking in various species such as mice. Orienting (looking) before or even without recognition (seeing) by the central sense is therefore comprehensible natural behavior. Therefore, studying looking and seeing in human vision provides a highly accessible window to examining CPD as a common framework that links ecological, behavioral, neurophysiological, and anatomical data across senses and species.

5. Conclusion

Using eye tracking during observers’ visual search on stimuli diagnostic of V1 mechanisms, this study provides further evidence for the V1 Saliency Hypothesis (V1SH) and the Central-peripheral dichotomy (CPD). V1SH states that V1 is the neural basis for saliency to guide attention exogenously; CPD asserts that central and peripheral vision are specialized for seeing and looking, respectively.

Acknowledgements

This work was supported by funding from the University of Tübingen, the Max Planck Society, and the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG) with Projektnummer 465358224.

Declaration of commercial interests and conflicts: none

References


