Feedback from higher to lower visual areas for visual recognition may be weaker in the periphery: glimpses from the perception of brief dichoptic stimuli

Li Zhaoping, University College London, z.li@ucl.ac.uk

Published in *Vision Research*, volume 136, pages 32–49, 2017

Abstract: Eye movements bring attended visual inputs to the center of vision for further processing. Thus, central and peripheral vision should have different functional roles. Here, we use observations of visual perception under dichoptic stimuli to infer that there is a difference in the top-down feedback from higher brain centers to primary visual cortex. Visual stimuli to the two eyes were designed such that the sum and difference of the binocular input from the two eyes have the form of two different gratings. These gratings differed in their motion direction, tilt direction, or color, and duly evoked ambiguous percepts for the corresponding feature. Observers were more likely to perceive the feature in the binocular summation rather than the difference channel. However, this perceptual bias towards the binocular summation signal was weaker or absent in peripheral vision, even when central and peripheral vision showed no difference in contrast sensitivity to the binocular summation signal relative to that to the binocular difference signal. We propose that this bias can arise from top-down feedback as part of an analysis-by-synthesis computation. The feedback is of the input predicted using prior information by the upper level perceptual hypothesis about the visual scene; the hypothesis is verified by comparing the feedback with the actual visual input. We illustrate this process using a conceptual circuit model. In this framework, a bias towards binocular summation can arise from the prior knowledge that inputs are usually correlated between the two eyes. Accordingly, a weaker bias in the periphery implies that the top-down feedback is weaker there. Testable experimental predictions are presented and discussed.

Keywords: Central vision, peripheral vision, dichoptic stimuli, visual decoding, top-down feedback, primary visual cortex (V1), analysis-by-synthesis.

1 Introduction

The most obvious difference between central and peripheral vision in humans is the spatial resolution of visual perception (Anstis, 1974). This difference is in line with the differential brain resources devoted to central and peripheral vision. Each unit solid angle in the central visual field has devoted to it more photoreceptors and retinal ganglion cells, and a larger surface area in primary visual cortex and most of the extrastriate visual cortical areas (Osterberg, 1935; van Essen and Anderson, 1995) than the same angle in the periphery. One might therefore wonder whether central and peripheral vision would be equal if the different spatial resolutions were compensated for by scaling the visual inputs spatially (Koenderink et al., 1978).

However, a more important difference between central and peripheral vision is the extra difficulty in recognizing visual shape or form in the latter even when visual inputs in the periphery are enlarged to compensate for the lower spatial resolution, see (Strasburger et al., 2011) for a review. This peripheral problem is twofold, one is the difficulty or impossibility of recognizing a single item or pattern, such as a numeral (Strasburger et al., 1994; Strasburger and Rentschler, 1996) or a spatial pattern for hyperacuity tasks (Westheimer, 1982; Fendick and Westheimer, 1983), even after input enlargement; the other is the additional difficulty when the input pattern to be recognized is surrounded by neighboring inputs (Anstis, 1974; Strasburger et al., 1991; Pelli et al., 2004; Strasburger, 2014), for review see (Levi, 2008; Strasburger et al., 2011). The latter is referred to as crowding, as if the contextual inputs inhibit recognition by making the visual scene more cluttered. The underlying neural causes for these peripheral difficulties are poorly understood; there are only some observations suggesting that visual cortical areas like V1, V2, and V4 might be involved (Flom et al., 1963; Strasburger et al., 1994; Strasburger and Rentschler, 1996; Tyler and Likova, 2007; Levi, 2008; Chen et al., 2014a).

A different perspective on the differences between central and peripheral vision comes from considering visual attention. The brain has limited resources; thus only a small fraction of visual inputs can be selected to be brought into the attentional spotlight for further processing. New destinations for selection typically start in the periphery; thus attending them happens in two stages. First, the attentional spotlight is shifted very briefly to the peripheral target. Second, a saccade brings that destination to the center of the visual field (Hoffman, 1998). We refer to this process of selecting the small fraction of inputs and bringing them into the attentional spotlight as the process of looking.

After looking is complete, another process, which we refer to as seeing, focuses on the selected inputs within the attentional spotlight. It recognizes or infers properties of visual scenes. While a single word "attention" is often used to refer to both looking and seeing, it is important to note that looking involves visual selection (attentional selection of visual inputs) while seeing involves visual decoding (of inputs within the attentional spotlight) (Zhaoping, 2014). In natural visual behavior, shifting one's gaze to attended inputs is mandatory (Deubel and Schneider, 1996; Hoffman, 1998). Thus, peripheral and central vision should differentially emphasize looking and seeing respectively (Zhaoping, 2014), with peripheral vision's role in looking compromising its ability to see. This inferiority is manifest in its lower spatial resolution (Aubert and Foerster, 1857; Weymouth, 1958; Koenderink et al., 1978), reduced performance in recognition tasks such as hyperacuity and character recognition (Fendick and Westheimer, 1983; Strasburger et al., 1991; Strasburger and Rentschler, 1996), and its affliction to visual crowding (Strasburger et al., 1991, 1994; Pelli and Tillman, 2008; Levi, 2008).

Here, we probe differences in the ways that central and peripheral vision perform seeing or visual recognition tasks. These differences go beyond mere spatial resolution or visual input sensitivity; and so we examine them using computationally motivated visual psychophysical methods. In particular, we consider different influences over ambiguous percepts coming from prior knowledge about the statistics of natural visual inputs.

We use specially designed dichoptic inputs that have ambiguous perceptual interpretations; the relevant prior knowledge in this case is that the inputs to the two eyes are normally correlated (Li and Atick, 1994). This prior expectation can influence the perceptual outcome in a case of ambiguous perception by favoring the interpretation associated with similar or identical inputs to the two eyes over that associated with dissimilar or opposite inputs to the two eyes. To anticipate, we report that this prior influence was indeed observed and was significantly stronger in central than in peripheral vision.

Our particular prior influence must use information about the eye-of-origin of visual inputs to tell whether inputs from the two eyes are similar or dissimilar. This information is abundant in the primary visual cortex (V1) but is absent or very scarce beyond V1 along the visual pathway (Hubel and Wiesel, 1968; Burkhalter and Van Essen, 1986). Therefore, we argue that this prior influence cannot act in higher visual areas beyond V1 in a purely feedforward manner, but has to involve feedback processes to V1. In particular, we argue that it acts through the following processes which compute analysis-by-synthesis: (1) feedforward visual inputs from V1 suggest initial perceptual hypotheses about properties (e.g., the motion direction of a drifting grating) of the visual scene; (2) the visual inputs expected under each hypothesis are generated or synthesized according to prior knowledge of the visual world, (3) these synthesized visual inputs are fed back from higher brain areas to be compared with the actual visual inputs in the primary visual cortex, allowing the match between hypothetical and actual input to be assessed (Carpenter and Grossberg, 2011); and (4) a perceptual hypothesis is strengthened or weakened for close or far matches, respectively. In these processes, the prior knowledge shapes the hypothesized inputs and, consequently, biases the final perceptual outcome. Accordingly, a weaker prior influence in the periphery should suggest that the feedback pathway in the periphery is weaker.

Note that although we suggested that peripheral and central vision differ according to their differential foci on looking and seeing, we investigated the difference by considering feedback influences in the seeing computation of analysis-by-synthesis. This paper makes no commitment to other computational or mechanistic differences. In particular, a weaker or absent feedback in peripheral vision for analysis-by-synthesis does not necessarily mean that visual attention does not operate in peripheral vision. After all, we argued above that peripheral vision has to play an important or leading role in attentional selection (looking). Some such selections are exogenous or stimulus driven, governed mainly by a bottom-up saliency map generated by feedforward and intra-cortical mechanisms (Li, 2002; Zhaoping, 2014) (if we exclude the mechanism of reading out this saliency map to guide gaze shifts through the midbrain). Other such selections are endogenous or driven by top-down factors such as in feature-based attention, by which visual inputs having a particular feature across the visual field are preferentially selected for further scrutiny (e.g., red-colored inputs across visual field are preferentially selected when one is searching for a red cup). Questions such as how feature-based attentional selection operates, e.g., whether it operates through feedback fibers from higher to lower brain areas or by weighting the feedforward fibers according to the goals of ongoing visual tasks (Wolfe et al., 1989; Müller et al., 2003), lie beyond the scope of this study.

Parts of the data in this study and their theoretical implications have been previously presented in abstract form (Zhaoping, 2013b,a, 2015) and discussed in Zhaoping (2014).

2 Ambiguous perception using dichoptic stimuli

Shadlen and Carney (1986) presented two brief flashing horizontal gratings, one each to the two eyes; the two gratings differed from each other in both spatial and temporal phase by 90^o, and the authors reported that observers perceived a drifting grating which was the sum of the two flashing gratings. More specifically, let S_L and S_R be the input stimuli to the left and right eye, respectively (for simplicity, the background

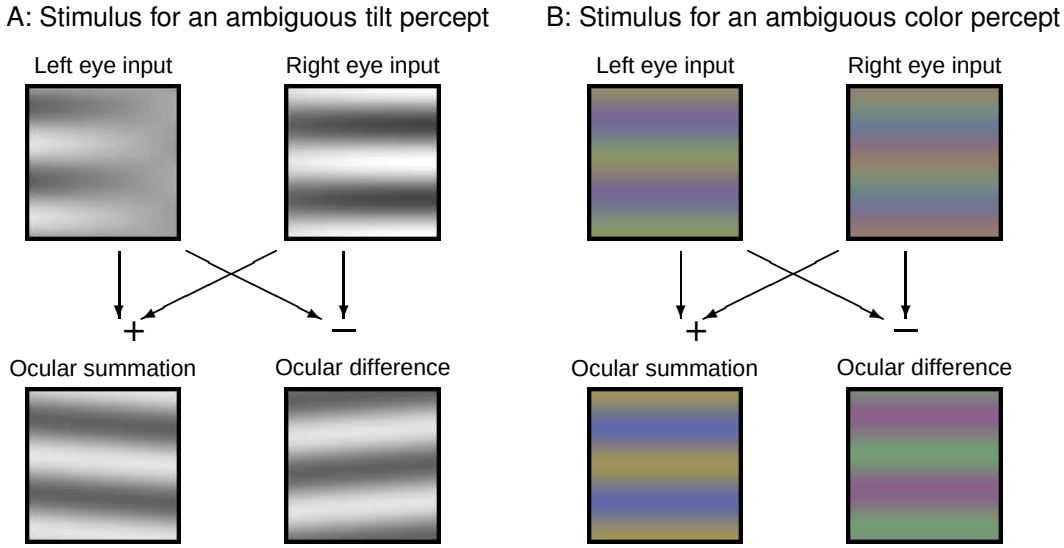

Figure 1: Examples of dichoptic stimuli for which the sum and difference of the monocular inputs lead to very distinct or opposite percepts. These percepts can be probed by asking observers to provide two-alternative-forced-choice (2FAC) answers about whether they see a grating tilted clockwise or anticlockwise from horizontal (for the stimulus in A), or to which of the two reference patterns made from the ocular sum or difference patterns the color percept is closer (for the stimulus in B). The stimulus in A can be generalized to dynamic gratings (by making the horizontal dimension represent temporal rather than spatial coordinates of visual inputs), so that each monocular input is a flashing grating and the ocular summation and difference are horizontal gratings drifting in opposite vertical directions (see equations (1) - (3)).

luminance used to ensure that $S_{L,R}$ is non-negative is omitted in the following expressions),

$$\begin{aligned} S_L &= c \cdot \cos(k \cdot y + \phi_y) \cos(2\pi\omega \cdot t + \phi_t), \\ S_R &= c \cdot \sin(k \cdot y + \phi_y) \sin(2\pi\omega \cdot t + \phi_t), \end{aligned} \tag{1}$$

where y denotes vertical spatial location and t denotes time, ϕ_y and ϕ_t are two arbitrary phase values; $c \leq 1$ is the contrast amplitude of the gratings, and k and w are their spatial and temporal frequencies, respectively. The binocular sum

$$\begin{aligned} S_+ &\equiv S_L + S_R \\ &= c \cdot \cos\left[k \cdot (y - 2\pi\frac{\omega}{k}t) + \phi_y - \phi_t\right] \end{aligned} \tag{2}$$

is the perceived drifting grating, drifting at speed $v = 2\pi\omega/k$.

However, it is known that, because natural inputs to the two eyes are correlated, the primary visual cortex (V1) represents these inputs more efficiently (Barlow, 1961; Li and Atick, 1994) by transforming them into two decorrelated channels: one for the sum of, and the other for the difference between, the inputs in the two eyes (see section 6.4 for physiological details and discussions). The sensitivities to the binocular summation and difference signals can independently adapt to the visual inputs statistics (May et al., 2012; May and Zhaoping, 2016). We thus predicted, and duly found, that the difference S_- between the two flashing gratings should also be observed, albeit more rarely. This difference signal S_- is another grating that drifts in the opposite direction to the binocular sum grating

$$\begin{aligned} S_- &\equiv S_L - S_R \\ &= c \cdot \cos\left[k \cdot (y + 2\pi\frac{\omega}{k}t) + \phi_y + \phi_t\right]. \end{aligned} \tag{3}$$

Hence, a dichoptic input such as that used by Shadlen and Carney (1986) evokes an ambiguous percept, of either the sum of, or the difference between, the inputs to the two eyes, with the percept of the sum dominating.

We later generalized this class of ambiguous percepts to static spatial gratings (May and Zhaoping, 2016), replacing a drifting grating in space-time by a tilted grating in two dimensional space, see Fig 1A. Specifically, we replaced time t by horizontal space coordinate x, replaced ϕ_t by ϕ_x, and replaced $2\pi\omega$ by its corresponding spatial frequency k_x. For notational clarity, the original k is now denoted as k_y. This leaves

$$\begin{aligned} S_L &= c \cdot \cos(k_y \cdot y + \phi_y)\cos(k_x \cdot x + \phi_x), \\ S_R &= c \cdot \sin(k_y \cdot y + \phi_y)\sin(k_x \cdot x + \phi_x), \end{aligned} \tag{4}$$

and

$$\begin{aligned} S_+ &= c \cdot \cos(k_y \cdot y - k_x \cdot x + \phi_y - \phi_x), \\ S_- &= c \cdot \cos(k_y \cdot y + k_x \cdot x + \phi_y + \phi_x), \end{aligned} \tag{5}$$

where ϕ_x and ϕ_y are arbitrary phase values. In this case, the binocular sum and difference gratings are tilted in opposite directions from the horizontal axis by a tilt angle of $\tan^{-1}(k_x/k_y)$; and either of these gratings could be perceived.

Hence, regardless of whether the stimuli are drifting or static, one can generate any desired binocular summation and difference stimuli S_+ and S_- by presenting $S_L = (S_+ + S_-)/2$ to the left eye and $S_R = (S_+ - S_-)/2$ to the right eye. A further extension is to endow S_+ and S_- with unequal contrasts c_+ and c_-, respectively, rather than having the same contrast. In this case, decreasing $|c_+/c_-|$ will decrease the bias towards perceiving the S_+ signal; in the limit of small $|c_+/c_-|$, S_- could even dominate the percept (see later in the paper).

Spatiotemporal gratings can be further generalized to static color gratings by replacing each luminance value in $S_\pm$ by a three-component vector with red, green, and blue components; see Fig 1B. Hence, each contrast, c_+ or c_-, now becomes a three-component vector for the corresponding contrast in the input color component. For example, if the contrast vector is (using bold-faced characters to denote vectors) $\mathbf{c}_+ = (0.5, -0.5, 0)$, then $\mathbf{S}_+$ is a grating that alternates between red and green colors; if the contrast vector is $\mathbf{c}_- = (0.5, 0.5, 0)$, then the $\mathbf{S}_-$ grating alternates between blue and yellow (after adding the luminance background). More specifically, we generated horizontal colored gratings using

$$\begin{aligned} \mathbf{S}_+ &= \mathbf{c}_+ \cdot \cos(ky + \phi_+), \\ \mathbf{S}_- &= \mathbf{c}_- \cdot \cos(ky + \phi_-), \end{aligned} \tag{6}$$

(where $\phi_\pm$ are arbitrary), and presented the corresponding monocular stimuli $\mathbf{S}_{L,R} = (\mathbf{S}_+ \pm \mathbf{S}_-)/2$ to the left and right eyes, respectively. Ambiguous color percepts can be examined by asking observers to report whether the color of their perceived grating more closely resembled $\mathbf{S}_+$ (e.g., a red-green grating) or $\mathbf{S}_-$ (e.g., a blue-yellow grating).

Here, we report drift direction, tilt direction, and color of the ambiguous percepts generated by these dichoptic stimuli. We show that in all three cases, the percept of S_+ arising from binocular summation is favored over the percept of S_- from the binocular difference, if the S_+ and S_- gratings have similar contrasts. However, this bias towards S_+ is weaker or even absent when the stimuli are viewed peripherally.

We then argue that the bias towards S_+ can arise from a perceptual prior for binocular inputs that are similar or identical between the two eyes, that this prior can be implemented in the perceptual feedback from higher visual cortical areas to V1, and that such feedback is weaker or absent in the periphery.

To exclude a potential confound from a response bias towards one of the two ambiguous perceptual reports, the S_+ stimulus is made to express either at random. Thus, using the drift gratings for example, let S_1 and S_2 be gratings that drift up and down, respectively, so that $S_{1,2} = \cos\left[k \cdot (y \mp 2\pi\frac{\omega}{k}t) + \phi_{1,2}\right]$ (with $\phi_{1,2}$ as random phases), then the stimulus in each trial, given input contrast $c_\pm$ for $S_\pm$, can be randomly either

$$\begin{array}{ll} S_+ = c_+S_1, & \\ S_- = c_-S_2; & \text{or} \end{array} \quad \begin{array}{l} S_+ = c_+S_2, \\ S_- = c_-S_1. \end{array} \tag{7}$$

Then, averaging over many trials, the perceptual bias towards S_+ would be measured by the fraction of trials in which the reported percept, S_1 or S_2, arises from the signal in the S_+ channel. Analogously, for static tilt gratings, equation (7) is applied to $S_{1,2} = \cos(k_y \cdot y \mp k_x x + \phi_{1,2})$.

Of course, if observers expressed a strong intrinsic bias towards reporting the feature from S_1 or S_2, then this would obscure (though not confound) the feature-independent bias towards $S_\pm$ that we sought to measure. Fortunately, few observers had such a strong bias towards any drift or tilt direction. This allows us to draw strong conclusions about the brain's visual decoding mechanisms and the neural feedback architecture associated with these features. By contrast, observers do often have an idiosyncratic intrinsic bias towards color (e.g., favoring red or blue). Hence, $\mathbf{S}_1$ and $\mathbf{S}_2$ color stimuli were chosen such that the intrinsic response bias towards S_1 or S_2 was not too strong across most observers. Intrinsic biases in color make our observations about the architecture of color processing less telling. Nevertheless, we include these results to illustrate the generality of our conclusions.

3 Experimental Methods

We performed several main experiments, one control experiment, and one auxiliary experiment. In each main experimental session, two types of trials were randomly interleaved; one type involved a centrally viewed test grating of a fixed size, and the other type involved a peripherally viewed test grating at a fixed peripheral visual field location and enlarged according to the eccentricity of the peripheral location. Each main experiment was defined by the following characteristics: (1) the stimulus dimension: motion, tilt, or color; (2) the location in the visual field for the peripherally viewed test grating; and (3) the collection of different stimulus conditions, such as input contrasts, randomly interleaved within an experimental session (see details next). All the details of the methods presented before section 3.5 are for the main experiments. Sections 3.5 and 3.6 detail the methods for the control and auxiliary experiments, respectively. The experimental work was carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Informed consent was obtained for experimentation with human subjects.

3.1 Procedure

Fig. 2 shows the sequence of events in a trial. The three main stimuli in a trial include a binocular start-image or fixation image, a dichoptic test stimulus, and a binocular end-image. In each trial, the observer pressed a button to start fixation on the fixation point in the binocular start-image, examples of which are shown in Fig. 3. The start-image was shown before this fixation period started, with instruction text "Press any button for the next trial" written near the fixation point to prompt the observer to press a button to start the trial. This button press triggered the start of the period intended for fixation. At and after 0.7 seconds into this fixation

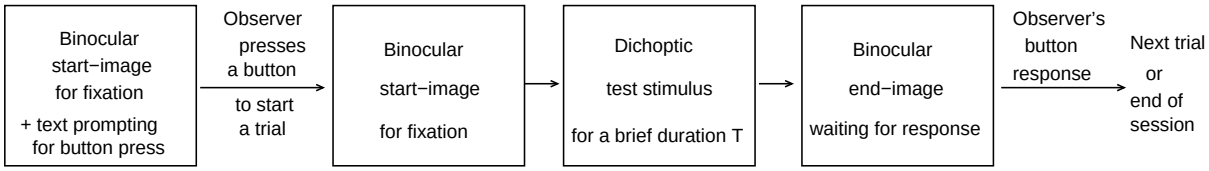

Figure 2: Sequence of events in a trial.

period, once the eye tracker verified that the observer had been properly fixating continuously for 100 ms (eye tracking was carried out on all observers bar the author), the dichoptic test stimulus containing the test grating was shown for a duration T, which was always less than half a second and depended on the stimulus condition. During the test stimulus, the observer was required to maintain fixation on the fixation point which was still present at the same location. The test stimulus was then replaced by a binocular end-image, upon which the observer was free to move their gaze. While the end-image was displayed, the observer could take as much time as required to complete the trial by pressing one of the two specified buttons to report which of the two possible percepts (as instructed) of the dichoptic test stimulus was perceived.

Before the start of an experimental session, the observer was instructed about the procedure and the task. In particular, they were informed about the two alternative percepts between which they had to choose by pressing a button following the disappearance of the test grating. Specifically, for an experiment involving drifting gratings, they were asked to report on each trial whether they perceived the test grating as moving up or down; for an experiment involving static tilted gratings, they had to report whether they perceived the tilt direction of the test grating as being clockwise or counter-clockwise from horizontal; for an experiment involving colored gratings, they had to report which of the two colored reference gratings (which were the $\mathbf{S}_1$ and $\mathbf{S}_2$ patterns) displayed in the binocular end-image (see details in Stimuli) the test grating more closely resembled.

Within each experimental session or block, there were only two possible binocular start-images or fixation images, one for all trials involving a centrally viewed test grating, and the other for all trials involving a peripherally viewed test grating (see Fig. 3 for an example). Centrally and peripherally viewed gratings were equally likely and randomly chosen in each trial. The observer knew to expect the test grating to be at a fixed location on the display screen throughout a session, regardless of whether the test grating should be viewed centrally or peripherally. This location was made to be central or peripheral by changing the fixation location in the display screen; and the test grating was then small or large accordingly. The text "Press any button for the next trial" was displayed near the fixation point at the start of each trial to cue the observers toward the fixation location. The fixation point and a square box which would frame the upcoming test grating were clearly shown in the start-image. Hence, before the observers pressed a button to start a trial, they already knew where to fixate, the location and size of the upcoming test grating, and whether the grating had to be centrally or peripherally viewed.

3.2 Subjects

There were altogether 84 subjects (26 male and 58 female) who participated in the various experiments reported in this paper. They were recruited from the university community, with minimum, median, and maximum ages 17, 22, and 52 years, respectively. They had normal or corrected to normal vision, and were tested for their ability to see depth. Four of the subjects could not see depth. One of these four subjects participated in the auxiliary experiment (on contrast detection thresholds, see Fig. 11). The other three of these four subjects participated in the experiment on ambiguous drift perception (and their data contributed

to the plots in Fig. 4B and Fig. 6B); their data appear qualitatively the same as those from other subjects.

Except for three subjects, each observer participated in only one experiment. Two of the three subjects each participated in two different experiments (one in color perception and one in drift direction perception); the third, the author (who also collected all the data), participated in three different experiments (one main experiment on tilt perception, the control experiment, and the auxiliary experiment). All subjects except the author were naive to the purpose of the experiments. All experimental sessions and blocks were designed such that knowledge of the purpose of the experiments could not influence a subject's responses (taking advantage of the observation that the ocular summation signal and the ocular difference signal did not appear distinguishable in our test gratings).

3.3 Stimuli

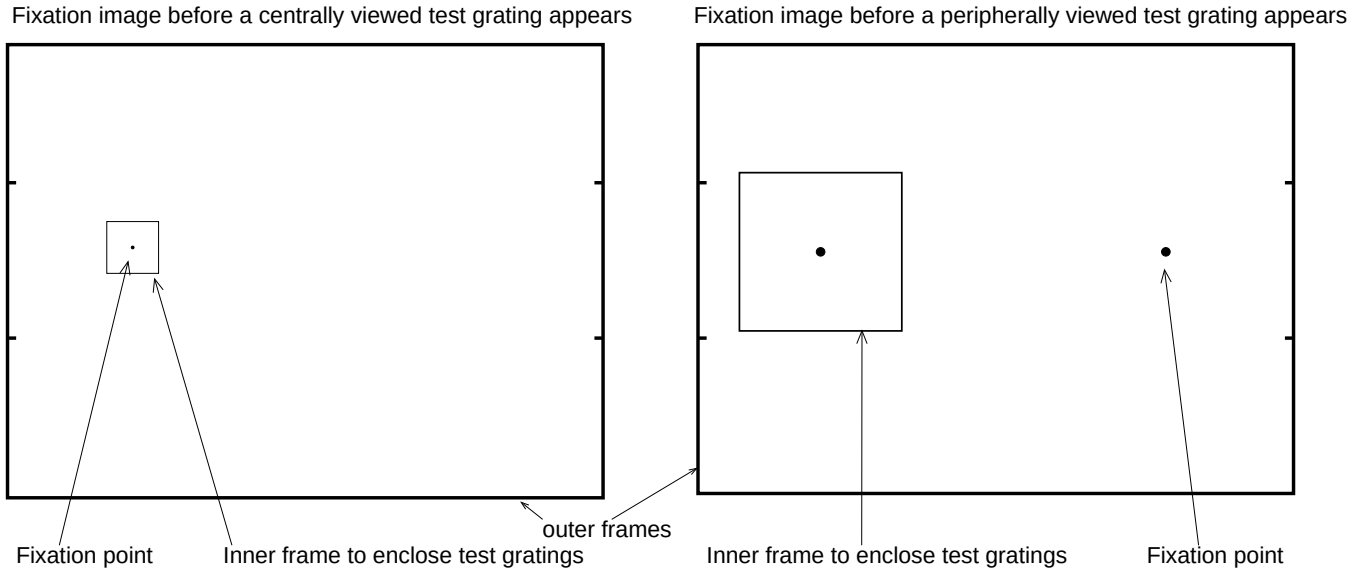

Figure 3: Schematics of two binocular fixation images for an example block of trials. The left and right schematics above were respectively for trials when the test grating was viewed centrally and peripherally, these two types of trials were randomly interleaved in a block. In this example block, the peripheral grating appeared in the left visual field. The outer frame in the fixation images was the same across the trials, and was displayed continuously at a fixed location on the display screen throughout an experimental session. The inner frame to enclose the test grating was centered at the same location on the screen across all trials in a session, regardless of whether the test grating was centrally or peripherally viewed. The location of the fixation point was inside or outside the inner frame, depending on whether the upcoming test grating was viewed centrally or peripherally, respectively. When a fixation image was displayed at the start of a trial, the text string "Press any button for the next trial" was shown near the fixation point to guide the observer to fixate at the correct location. The text string disappeared after the button press to start the fixation period for this trial. The black disk at the center of the inner frame had a radius that was 1/20 of the side length of the inner frame, and was present regardless of whether it was intended as the fixation point, so that the centrally and peripherally viewed stimuli within the inner frames were identical to each other except for their spatial scale.

Dichoptic stimuli were viewed by a mirror stereoscope as previously described (Zhaoping, 2012). The

inputs to the left and right eyes were displayed side by side on a Mitsubishi 21-inch cathode-ray tube (CRT) and viewed through a stereoscope with an assembly of four mirrors which directed each monocular input to the intended eye. The stimulus was displayed on a white background whose luminance was $73\mathrm{cd}/\mathrm{m}^2$ on the display screen. The actual background luminance $\bar{S}$ for visual input should be a fraction of that on the screen due to attenuation along the optical path (for each eye) from the display screen to the eye via two reflective mirrors for the stereoscope and one semi-reflective glass plate for the eye tracker, but $\bar{S}$ was safely within the photopic range.

Fig. 3 shows schematic binocular start-images intended for the fixation period of each trial of a session in which the test gratings in the peripheral view were presented in the left visual field. Each monocular input stimulus was contained within a rectangular black frame (marked as the outer frames in Fig. 3) that was 20^o in width, 16.7^o in height, and 0.22^o in thickness of the frame skeleton, and had four spikes on its two vertical sides to help anchor vergence. Each monocular test grating stimulus was framed within a black square (marked as the inner frames in Fig. 3) whose side had a length L that was larger for the peripheral presentation (see below). The thickness of the skeleton of the inner square frame was $L/25$. A small black disk of radius $L/20$ was displayed at the center of the square frame. This black disk was the fixation point when the test gratings were viewed centrally. When the test gratings were viewed peripherally at eccentricity e (in degrees of visual angle), another small black disk of the same size was additionally displayed at distance e away from the center of the square to serve as the fixation point.

The size of the test grating was normalized by the visual spatial resolution that changes with eccentricity e (degrees), with $e = 0$ for central viewing and $e > 0$ for peripheral viewing, so that the side of the inner frame was $L = 1.13^o \cdot (1 + e/e_2)$ with $e_2 = 3.3^o$. It should be noted that the values of e_2 in the literature are diverse and depend on the visual task employed (see Tables 4 and 5 of Strasburger et al. (2011)). The value $e_2 = 3.3^o$ is quantitatively similar to those employed for visual acuity tasks using gratings (Rovamo and Virsu, 1979) or letters (Higgins et al., 1996) as visual inputs. If we associate our tasks with stereo acuity tasks (Fendick and Westheimer, 1983), then e_2 should be $e_2 = 0.81^o$, and our grating size would have to be rescaled more steeply with eccentricity. Sections 3.5 and 5.2.2 describe a control experiment that verified that using a very different rescaling (i.e., no rescaling at all) did not substantially change our conclusions.

Each monocular (left or right eye) image of the dichoptic test stimulus was identical to the start-image of the trial except that the image pixels within the inner frame, excluding the pixels for the central black disk, were replaced by image pixels from the corresponding monocular (left or right eye) image of the dichoptic test grating. Hence, the monocular images within and including the inner frame are like those shown in Fig. 1, except that the central black disk is not displayed in Fig. 1. When the dichoptic test grating was drifting and so was dynamic rather than static, the monocular images were updated every frame at a frame rate of 100 Hz.

3.3.1 Test gratings and the end-image for experiments on ambiguous drift direction

Given luminance $\bar{S}$ of the white background and length L of each side of the square inner frame, the monocular test gratings S_L and S_R were made as follows. For motion percepts, the test grating stimulus in each trial was completely determined by the following five stimulus parameters, k, ω, c_+, c_-, and T, plus three numbers, ϕ_1, ϕ_2, and q, that were randomly generated for each trial. This resulted in two underlying percepts of oppositely drifting gratings S_1 and S_2, which are $S_{1,2} = \cos\left[k \cdot (y \mp 2\pi\frac{\omega}{k}t) + \phi_{1,2}\right]$, in which k is the spatial frequency of the grating and was always $4\pi/L$ (giving two cycles of the grating inside the inner frame), ω is the temporal frequency of the grating (e.g., $\omega = 2.5$, 5, or 10 Hz), y is the vertical coordinate location within the inner square frame, t is time (in second) from $t = 0$ to $t = T$ in discrete steps of 10 ms (given the frame rate of 100 Hz for the CRT), and ϕ_1 and ϕ_2 are two independent phase values (uniformly

distributed within $[0, 2\pi)$) randomly generated for each trial. Meanwhile the random number q takes value $q = 1$ and $q = 2$ with equal probability, so that $S_+ = c_+ S_q$ and $S_- = c_- S_{q'}$, with $q' = 2$ or $q' = 1$ when $q = 1$ or $q = 2$, respectively. Therefore, the monocular gratings are

$$\begin{aligned} S_L &= \bar{S}\left[1 + (c_+ S_q + c_- S_{q'})/2\right] \\ S_R &= \bar{S}\left[1 + (c_+ S_q - c_- S_{q'})/2\right]. \end{aligned} \tag{8}$$

Note that these stimulus values include the background luminance $\bar{S}$ which was hitherto omitted for simplicity. Furthermore, this equation defines for this paper the meanings of c_+ and c_-, which are somewhat modified from the conventional meaning for "contrast".

The binocular end-image displayed after the dichoptic test gratings was the same as the binocular start-image displayed during the fixation period.

3.3.2 Test gratings and the end-image for experiments on ambiguous tilt perception

Analogously, equation (8) constructs the monocular test gratings for the experiments on tilt perception from $S_{1,2}$, with the fundamental drift gratings $S_{1,2}$ replaced by the tilted gratings. The parameters that determine the gratings are k, ψ, c_+, c_-, and T, plus the same three random numbers ϕ_1, ϕ_2, and q. These parameters have the same meanings as those for the drift gratings, except for replacing parameter ω by parameter ψ, which is the absolute angle (in radian), of the tilted gratings $S_{1,2}$ from horizontal. This paper's results were obtained from $\psi = 0.0698$ arising from a 4^o tilt angle. In particular, $S_{1,2} = \cos\left[k \cdot y_{1,2} + \phi_{1,2}\right]$, in which $y_{1,2} \equiv y\cos(\psi) \mp x\sin(\psi)$, with x and y as the horizontal and vertical coordinate locations of the image pixel within the inner frame that encloses the test gratings. Hence, $S_{1,2}$ are two gratings tilted in opposite directions by angle ψ from horizontal.

The binocular end-image after the dichoptic testing stimulus was the same as the start-image for the fixation of the trial.

3.3.3 Test gratings and the end-image for experiments on ambiguous color perception

For static color gratings inside the inner frame, each image pixel is described by a three-component vector for the red, green, and blue (RGB) part of the color. Meanwhile, $\mathbf{S}_1$ and $\mathbf{S}_2$ are no longer related by simple obvious symmetry. Explicitly,

$$\mathbf{S}_1 = \mathbf{c}_1 \cos(ky + \phi_1), \quad \mathbf{S}_2 = \mathbf{c}_2 \cos(ky + \phi_2), \tag{9}$$

where $\mathbf{c}_1$ and $\mathbf{c}_2$ are three-component vectors. In addition, the magnitudes, but not the directions, of the vectors $\mathbf{c}_1$ and $\mathbf{c}_2$ may depend on whether the grating was viewed centrally or peripherally. This is because $\mathbf{c}_{1,2}$ were chosen empirically to make sure that, averaged across observers, the response bias towards or against $\mathbf{S}_1$ was sufficiently weak as not to obscure the bias towards $\mathbf{S}_+$ that we aimed to measure. One way to balance the response bias is by adjusting the magnitude of $\mathbf{c}_1$ relative to that of $\mathbf{c}_2$. That the sensitivities to color in central and peripheral vision differ (Mullen and Kingdom, 1996) implies that this adjustment could depend on eccentricity. The results in this paper used the following two sets of choices for $\mathbf{c}_1$ and $\mathbf{c}_2$. In the first set, the relative magnitudes of $\mathbf{c}_1$ and $\mathbf{c}_2$ depend on eccentricity. Specifically,

$$\begin{aligned} \mathbf{c}_1 &= (1, \quad 1, \quad 0) \cdot 27/50, \\ \mathbf{c}_2 &= (1, \quad -1, \quad 0) \cdot 27/50 \end{aligned} \tag{10}$$

centrally, and

$$\begin{array}{l} \mathbf{c}_1 = (1, \quad 1, \quad 0) \cdot 207/555, \\ \mathbf{c}_2 = (1, \quad -1, \quad 0) \cdot 333/555 \end{array} \tag{11}$$

peripherally. After adding a constant white luminance to $\mathbf{S}_1$ and $\mathbf{S}_2$, stimulus $\mathbf{S}_1$ appears as a blue-yellow grating and $\mathbf{S}_2$ appears as a red-green grating, regardless whether they are for central or peripheral viewing. It is known that human sensitivity to red-green gratings decays with eccentricity relative to the sensitivity to the luminance-like blue-yellow gratings (Mullen and Kingdom, 1996). Therefore, the magnitudes of $\mathbf{c}_1$ and $\mathbf{c}_2$ were made to depend on eccentricity here to compensate for the changes in the sensitivities.

In the second set, $\mathbf{c}_{1,2}$ does not depend on eccentricity. Specifically

$$\begin{array}{l} \mathbf{c}_1 = (\quad 21, \quad 14, \quad -28 \quad)/75, \\ \mathbf{c}_2 = (\quad -7, \quad 21, \quad -7 \quad)/75 \end{array} \tag{12}$$

both centrally and peripherally. After adding a constant white luminance to $\mathbf{S}_1$ and $\mathbf{S}_2$, stimuli $\mathbf{S}_1$ and $\mathbf{S}_2$ resemble the ocular summation and difference images in Fig. 1 (so they appear like brown-blue and purple-green gratings, respectively).

Again, a random number $q = 1$ or $q = 2$ with equal probability determines whether $\mathbf{S}_{+,-}$ are associated with $\mathbf{S}_{1,2}$ or $\mathbf{S}_{2,1}$. Let the background white luminance be $\bar{S}(1, 1, 1)$ which is now a three-component vector with equal red, green, and blue components. Then the monocular gratings $\mathbf{S}_L$ and $\mathbf{S}_R$ to the left and right eyes are

$$\begin{array}{l} \mathbf{S}_L = \bar{S}\left[(1, 1, 1) + (c_+\mathbf{S}_q + c_-\mathbf{S}_{q'})/2\right], \\ \mathbf{S}_R = \bar{S}\left[(1, 1, 1) + (c_+\mathbf{S}_q - c_-\mathbf{S}_{q'})/2\right] \end{array} \tag{13}$$

or

$$\begin{array}{l} \mathbf{S}_L = \bar{S}\left[(1, 1, 1) + (c_+\mathbf{S}_q - c_-\mathbf{S}_{q'})/2\right], \\ \mathbf{S}_R = \bar{S}\left[(1, 1, 1) + (c_+\mathbf{S}_q + c_-\mathbf{S}_{q'})/2\right] \end{array} \tag{14}$$

with equal probability for each trial. In the experiments, (scalar) c_+ in the above equations for the color gratings are fixed to $c_+ = 1$, since its scale is already absorbed in the overall scales for $\mathbf{c}_1$ and $\mathbf{c}_2$, while (scalar) c_- could vary.

In each trial, we helped observers respond as to which of the two percepts corresponding to $\mathbf{S}_1$ and $\mathbf{S}_2$ the dichoptic test grating more closely resembled by displaying color grating images $\mathbf{R}_1$ and $\mathbf{R}_2$ for $\mathbf{S}_1$ and $\mathbf{S}_2$, i.e.,

$$\mathbf{R}_{1,2} \equiv \bar{S}\left[(1, 1, 1) + \mathbf{S}_{1,2}\right], \tag{15}$$

binocularly in the end-image presented upon the disappearance of the dichoptic stimulus, so that the observers could view these color grating references. The end-image was just like the start-image except for also including the reference color grating images $\mathbf{R}_1$ and $\mathbf{R}_2$. In all our color grating experiments, the start-images were just like those in Fig. 3, with the inner frame in the left half of the outer frame and the fixation point for peripheral viewing in the right half of the outer frame. Then, the center of $\mathbf{R}_1$ and $\mathbf{R}_2$ were placed 3.5^o below and above the center of the outer frame, respectively. Observers were instructed to press one or the other of the two buttons in their response if the color pattern in the test grating appeared more like the lower or the upper reference grating, respectively. They were instructed to ignore the phases of the test or reference gratings and focus on the perceived color in their decisions. In each session, the qualitative nature of the two color reference patterns were fixed across trials, whether the trial involved central or peripheral viewing of the test grating. For example, throughout one session, the lower reference grating always appeared yellow-blue while the upper reference grating always appeared red-green, although the phases of these gratings were random across trials.

3.4 Experimental conditions within a session and data analysis

Each experimental session contained a set of two-alternative forced choice trials in a single task domain: motion direction, tilt direction, or color. Within this single task domain, each test grating's stimulus condition was defined by two or three sets of parameters denoted as $\Gamma_{\rm location}$, $\Gamma_{\rm space-time}$, and, if the experiment involved color gratings, $\Gamma_{\rm color}$. The $\Gamma_{\rm location} \equiv (e, \theta)$ set indicates the test grating's location, with eccentricity e and its direction θ from fovea when $e > 0$. This $\Gamma_{\rm location}$ also determines the test grating size which depends only on e. The $\Gamma_{\rm space-time}$ set describes the other spatiotemporal properties of the grating. For the experiments on motion perception, $\Gamma_{\rm space-time} \equiv (k, \omega, T, c_+, c_-)$. Analogously, in the experiments on tilt percepts $\Gamma_{\rm space-time} \equiv (k, \psi, T, c_+, c_-)$; for the experiments on color percepts, $\Gamma_{\rm space-time} \equiv (k, T, c_+, c_-)$. For colored gratings only, $\Gamma_{\rm color} \equiv (\mathbf{c}_1, \mathbf{c}_2)$ contains the two three-dimensional color vectors $\mathbf{c}_1$ and $\mathbf{c}_2$ defining the two reference color gratings. A given stimulus condition was determined by these various values of Γ; variations across trials for each condition were produced by sampling random numbers trial-by-trial such as the random phases of the gratings and the random number $q = 1$ or $q = 2$ to determine whether the first or second of the two standard percepts was used for the S_+ channel. In any session, randomly half of all the trials of a given condition had $q = 1$.

Each experimental session randomly interleaved trials from $2 \times N$ experimental conditions. These conditions involved two $\Gamma_{\rm location} = (e, \theta)$ sets, one with $e = 0$ (for central viewing) and another with $e > 0$ (for peripheral viewing), and each of these two $\Gamma_{\rm location}$ was associated with the same collection of N variations of $\Gamma_{\rm space-time}$ by the variations of $c_\pm$, ω, and/or T that controls the stimulus contrast, drifting speed, and/or presentation duration of the test grating. All sessions used a spatial frequency of $k = 2$ cycles per size of the side of the inner frame. All sessions on tilt perception used a ψ value that gave a tilt angle of 4^o. In each session on color perception, each of the two $\Gamma_{\rm location}$ was associated with its corresponding $\Gamma_{\rm color} = (\mathbf{c}_1, \mathbf{c}_2)$. Within a session, the test grating for both (e, θ) sets was always centered at the same location on the display screen, and the fixation point was at or away from it, depending on (e, θ) of the trial. Typically, for each subject, about 50 or 60 trials were employed for each condition. A block of trials typically took around ten to twenty minutes to complete, with short breaks between blocks.

To estimate the perceptual bias towards the S_+ signal for each stimulus condition and each observer, we obtained the fraction F_+ of the trials (for this stimulus condition) in which the observer reported the percept arising from signals in the S_+ channel, regardless of whether the signal in the S_+ channel came from the S_1 or the S_2 input stimulus. In experiments on color perception, since a non-negligible response bias was expected, this fraction F_+ was corrected for response bias as follows. For each $a = 1, 2$ and $b = 1, 2$, let $r_{a,b}$ be the number of trials in which the $\mathbf{S}_+$ signal arises from $\mathbf{S}_b$ while the observer reported the percept arising from $\mathbf{S}_a$. Then $n_a = r_{a,1} + r_{a,2}$ (with $a = 1$ or 2) is the total number of trials that the percept from $\mathbf{S}_a$ was reported regardless of what $\mathbf{S}_+$ was in each trial. If $n_1 > n_2$, the observer was biased to see $\mathbf{S}_1$ (in addition to any bias towards $\mathbf{S}_+$) so that $r_{1,b}$ for both $b = 1$ and 2 are larger than they should be if there was no response bias towards $\mathbf{S}_1$. This is corrected by $r_{1,b} \to r_{1,b} - (n_1 - n_2)/2$ for both $b = 1$ and 2. Similarly, if $n_2 > n_1$, then $r_{2,b}$ is corrected by $r_{2,b} \to r_{2,b} - (n_2 - n_1)/2$ for both $b = 1$ and 2. Then the response-bias-corrected fraction F_+ for the $\mathbf{S}_+$ percept is $F_+ = (r_{1,1} + r_{2,2}) / \sum_{a,b} r_{a,b}$.

3.5 The control experiment on possible effects of the stimulus sizes

A control experiment was carried out on ambiguous drift perception (results are shown in Fig. 10). The method for this experiment, including procedure, stimuli, and the experimental conditions, were exactly the same as that in one of the main experiments (whose results are shown in Fig. 4A, Fig. 6A, and Fig. 8A) on ambiguous drift perception, except that the size (L) of the centrally viewed gratings (and the associated

inner frame and fixation point) was enlarged to be the same as the size of the peripherally viewed gratings (and the associated inner frame and fixation point). This control experiment was designed to see whether the difference in our results for centrally and peripherally viewed stimuli was merely caused by the difference in the sizes of the stimuli. It also served to see whether our results depend sensitively on whether our stimuli were exactly normalized by the eccentricity-dependent visual resolution.

3.6 The auxiliary experiment to measure the contrast detection thresholds for ocular summation and ocular difference signals

We also conducted an auxiliary experiment to measure the detection thresholds for c_+ and c_-, using drifting gratings as examples. Each experimental session measured a particular pair of thresholds, $c_{th,+}$ and $c_{th,-}$, for c_+ and c_-, respectively, for a given and fixed set of stimulus parameter values that specified grating position (e, θ), grating size, and grating frequencies k and ω, and grating duration T. Each such set of parameter values was identical to those in one of the corresponding main or control experiments. The sequence of events in a trial resembled that in the main experiment, as indicated in Fig. 2, except for the following: between the start-image for fixation and the end-image waiting for response, the observers would hear two short beeps separated by a time gap (a random number uniformly distributed between 0.5 to 0.7 seconds) and a dichoptic test grating would appear randomly with the first or the second beep. The observer's task was to indicate, by a button press, whether the first or the second beep coincided with the appearance of the test grating. The drifting direction and initial phase of the grating in each trial were independently and randomly chosen and were task irrelevant. The test grating in each trial was randomly chosen to be either an ocular summation (i.e., $c_+ > 0$ and $c_- = 0$) or an ocular difference grating (i.e., $c_+ = 0$ and $c_- > 0$). This means that, at every instant, for a purely ocular summation grating of contrast c_+, the test gratings in the two eyes were exactly identical to each other and had the same contrast $c_+/2$; while for a purely ocular difference grating of contrast c_-, the test gratings in the two eyes were exactly photo negatives of each other and had the same contrast $c_-/2$. In each session, the values of c_+ and c_- at the initial trials were sufficiently high (at $c_\pm = 0.05$) to make the gratings most likely visible to the observer, and then adjusted across trials by two independent, three-down-one-up, staircases, one for c_+ and one for c_- (so that, before a trial started, $c_\pm \to 0.7c_\pm$ or $c_\pm \to c_\pm/0.7$, respectively, if the last three consecutive observer responses for a fixed $c_\pm$ were all correct or if the last response was incorrect; otherwise, $c_\pm$ stayed unchanged). The experimental session terminated when each of the staircases had no less than 20 reversals in the progression of the $c_\pm$ values.

Let each staircase yield c_1, c_2, ..., as the contrast values tested, each c_i (for $i = 1, 2, ...$) had n_i trials of which $m_i \leq n_i$ trials had correct responses. The contrast detection threshold c_{th} is then obtained by the following maximum-likelihood method (Wichmann and Hill, 2001). Let the response data be generated by an underlying psychometric function $p(c, a, b, \lambda)$ such that the probability of giving a correct response at contrast value c follows this Weibull function (with parameters a, b, and λ)

$$p(c, a, b, \lambda) = 0.5 + (0.5 - \lambda)\left[1 - \exp\left(-\left(\frac{c}{a}\right)^b\right)\right]. \tag{16}$$

Let $p_i \equiv p(c = c_i, a, b, \lambda)$, then the probability P of getting m_i correct responses out of n_i trials across various contrasts c_i is

$$P = \Pi_i \frac{n_i!}{m_i!(n_i - m_i)!} p_i^{m_i}(1 - p_i)^{n_i - m_i}. \tag{17}$$

The set of parameters a, b, and λ in $p(c, a, b, \lambda)$ that maximizes the logarithm of this probability P is then

obtained by an optimization procedure (using the *fmincon* function from Matlab). The contrast detection threshold c_{th} is the resulting parameter a.

4 Results

4.1 Central vision is more biased than peripheral vision to see the summation signal S_+

Figs. 4 and 5, respectively, show the probability F_+ of perceiving the motion direction and tilt direction arising from the stimulus in the S_+ channel when the S_+ and S_- channels have the same input contrast, i.e., $c_+ = c_-$. For the motion percept (Fig 4), F_+ is about 80% when the test grating was viewed centrally but is significantly smaller for the peripheral viewing, with F_+ around 70% in the left peripheral visual field at $e = 7.2^o$ eccentricity, and F_+ around 65% in the lower peripheral visual field at $e = 10^o$ eccentricity. For the tilt percept, central viewing gave an F_+ of around 65%, whereas peripheral viewing showed no bias, with F_+ around 50% at both the lower ($e = 10^o$) and left ($e = 7.2^o$) peripheral locations.

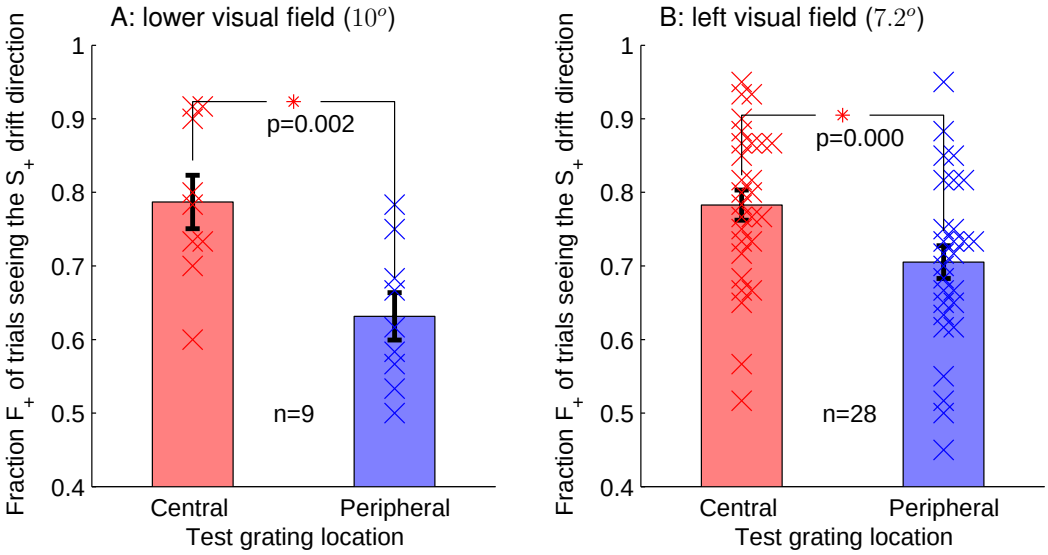

Figure 4: Higher probabilities F_+ of seeing the S_+ drift directions in gratings presented in the central compared to the peripheral visual field, when S_+ and S_- gratings had equal input contrasts $c_+ = c_- = 0.3$. The gratings were presented for $T = 0.2$ seconds, drifting at $\omega = 5$ Hz. A: the peripheral presentation was in the lower visual field at 10^o eccentricity; data bars represent results averaged over $n = 9$ subjects. B: the peripheral presentation was in the left visual field at 7.2^o eccentricity; data bars represent results averaged over $n = 28$ subjects. The peripheral stimuli were enlarged so that they roughly scale with the visual resolution which changes with eccentricity. The individual crosses are F_+'s from individual subjects. Each red asterisk, with its accompanying p value, indicates a significant difference (by a matched sample t-test) between the F_+'s (averaged across subjects) for the central and peripheral locations. The error bars in all the figures of this paper denote the standard errors of the means.

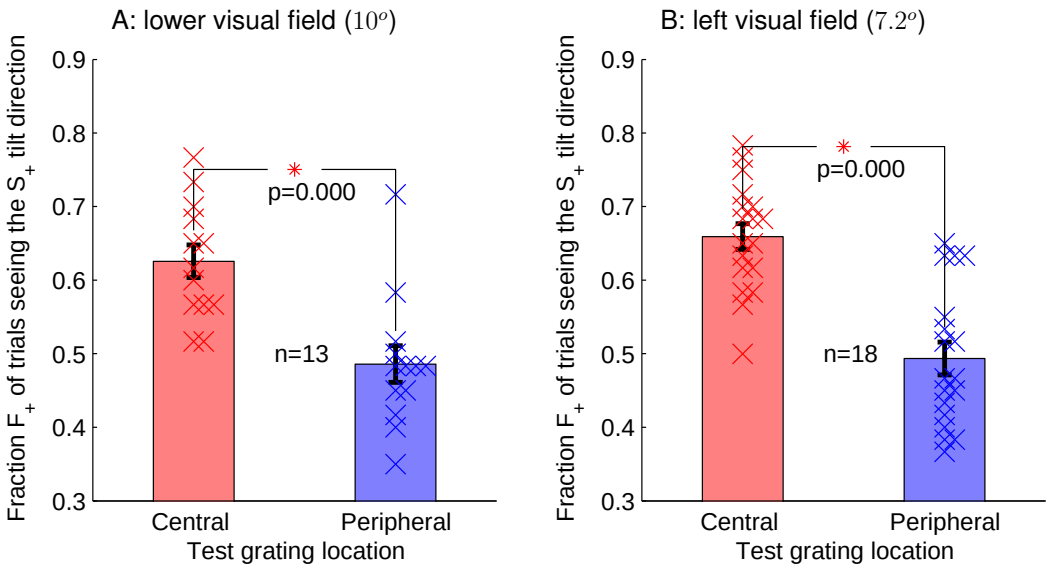

Figure 5: Higher probabilities F_+ of seeing the S_+ tilt directions in static gratings presented in the central compared to the peripheral visual field when S_+ and S_- gratings had equal input contrasts $c_+ = c_- = 0.2$. The gratings were presented for $T = 0.1$ seconds and the absolute tilt angle from horizontal was $\psi = 4^o$. Same figure format as in Fig. 4.

4.2 The weaker bias towards S_+ in periphery was not due to a weaker discrimination sensitivity

The sizes of the test gratings in the periphery were enlarged so that the stimuli roughly scale with visual resolution, yet if in peripheral vision the observers still could not see as clearly as in central vision, they might have to use more random guesses in their forced choice reports. If this were the case, their F_+ would indeed approach the chance level of 50%. This was particularly so for the tilt perception task which showed no bias towards S_+ in periphery. To test whether this reduced discrimination sensitivity was a confounding cause, the F_+ was measured for various (c_+, c_-) combinations to examine whether F_+ depended sensitively on the relative weight $c_+/(c_+ + c_-)$ of the S_+ signal in the test grating. If observers could not see clearly at all at the peripheral location, their F_+ would still be close to 50% even when $c_+/(c_+ + c_-)$ is far from a value of 50%.

Fig. 6 shows how F_+ varied with $c_+/(c_+ + c_-)$ in motion and tilt perception. It is clear that near a value of $c_+/(c_+ + c_-) = 0.5$, i.e., when $c_+ = c_-$, central and peripheral viewing give comparable slopes of the functions F_+ versus $c_+/(c_+ + c_-)$. This suggests that a weaker or zero bias towards S_+ in peripheral viewing is not due to an insensitivity to stimulus features at the peripheral location.

Fig. 7 shows that analogous results can be obtained for color perception. The bias to perceive the color feature arising from the stimulus in the S_+ channel is stronger in central than peripheral vision (Fig. 7A), and this is not due to a weaker sensitivity in the periphery for color discrimination (Fig. 7B).

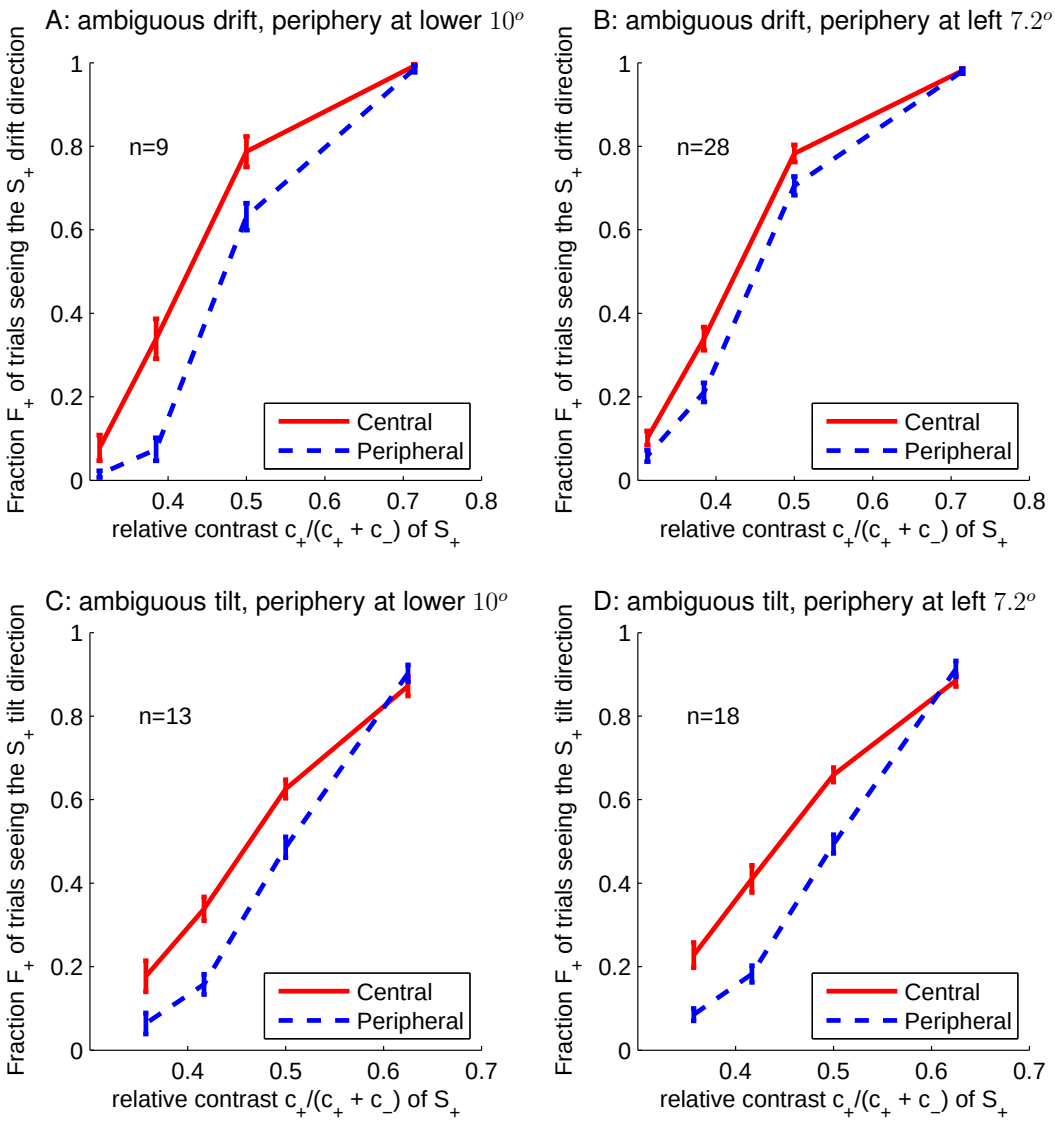

Figure 6: The probabilities of seeing the S_+ shown as psychometric functions of the relative contrast $c_+/(c_+ + c_-)$ of the S_+ signal. A and B are for motion percepts, and are from data collected in the same sessions as the data plotted in Fig. 4; $T = 0.2$ s and $\omega/(2\pi) = 5$ Hz. C and D are for the tilt percept, and are from data collected in the same sessions as the data plotted in Fig. 5; $T = 0.1$ s and $\psi = 4^o$. Across different stimulus conditions along each curve, c_- was varied while fixing c_+. ($c_+ = 0.3$ for A and B, and $c_+ = 0.2$ for C and D). All data points represent the averages across the respective n subjects.

4.3 Central and peripheral vision differ in their dependence of the ambiguous percept on viewing duration and drift speed

So far, we have looked at the quantitative difference between central and peripheral vision in their biases towards S_+. Fig 8 shows a qualitative difference, focusing on the case when S_+ and S_- had the same input contrast $c_+ = c_-$. For motion gratings, as the gratings moved faster, the fraction F_+ of seeing the S_+ percept decreased in central viewing. However, in the periphery, F_+ was not significantly affected. In static tilt perception, the fraction F_+ increased with increasing duration of viewing the test gratings for central

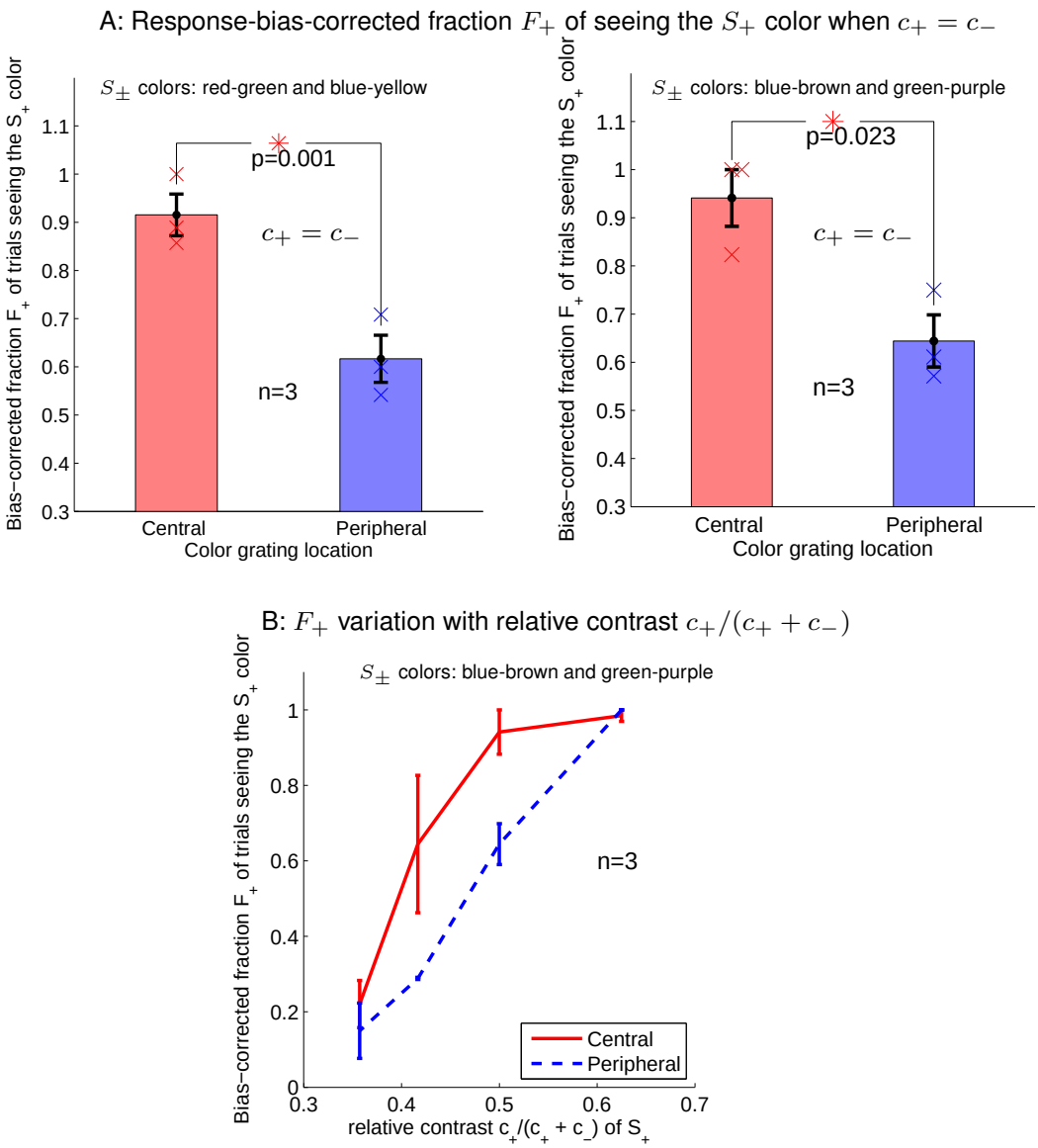

Figure 7: In ambiguous perception of color gratings, central view also had a stronger bias towards S_+ than peripheral view. Peripheral presentation was in the left visual field at eccentricity 7.2^o. In A is the response-bias-corrected fraction F_+ of reporting the S_+ percept when S_+ and S_- had the same input contrast $c_+ = c_- = 1$ (given the $\mathbf{c}_1$ and $\mathbf{c}_2$ for the input stimuli). Shown are two plots (in the same format as those in Fig. 4) for two different compositions for $S_\pm$ options; the left plot is for when S_+ and S_- were one each as red-green gratings and blue-yellow gratings indicated by equations (10–11), the right plot is for when they were blue-brown and green-purple gratings (in equation (12)) as shown by an example in Fig. 1B. B shows the variation of F_+ with $c_+/(c_+ + c_-)$, showing that peripheral perception is just as sensitive as central perception to $c_+/(c_+ + c_-)$. These two particular curves were obtained using blue-brown and green-purple gratings for the $S_\pm$ percepts as in the right plot in A, and all points on the same curve have the same $c_+ = 1$ but different c_-'s. Data from $n = 3$ subjects contributed to each plot.

vision, but for peripheral vision, F_+ did not follow this trend. (One may notice that, in each plot of Fig 8,

F_+ for peripheral vision exhibits a trend opposite to that exhibited in F_+ for central vision, but the difference between peripheral F_+ values for any two temporal frequencies or two presentation durations did not reach significance in our data.) These observations suggest that there may be a qualitative difference between central and peripheral vision in their underlying neural mechanisms for visual decoding.

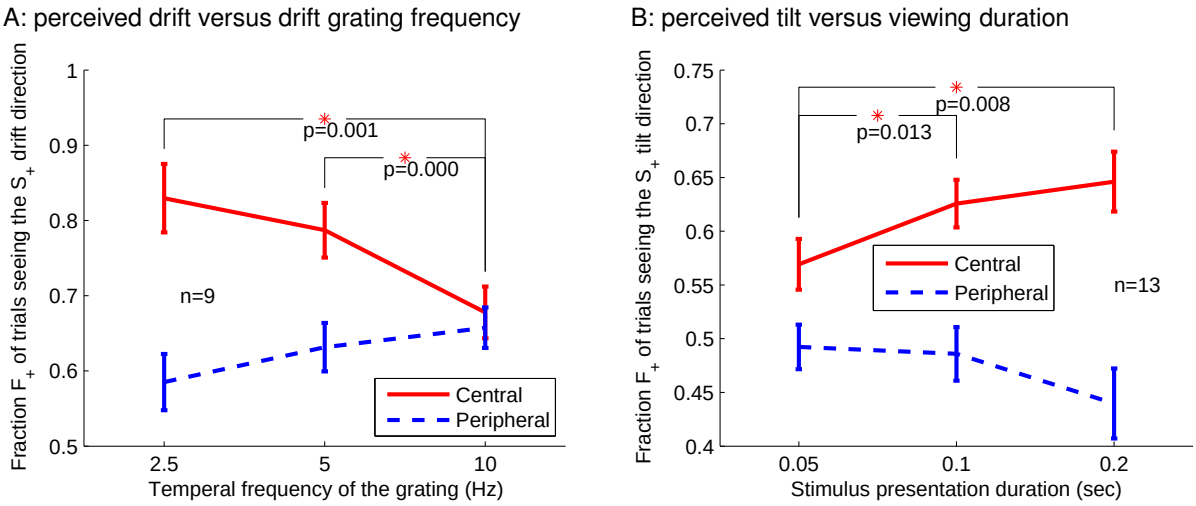

Figure 8: Central and peripheral viewing differ as to how stimulus properties affect ambiguous percepts. When stimuli S_+ and S_- had equal input contrast $c_+ = c_-$ ($c_\pm = 0.3$ in A and $c_\pm = 0.2$ in B), the fraction F_+ for S_+ in central vision decreased with increasing temporal frequency for the drifting grating (A) and increased with viewing duration for the static grating (B). However, in peripheral vision, the respective trends were either non-existent or in the opposite direction. In both A and B, peripheral viewing was at 10^o eccentricity in the lower visual field, $T = 0.2$ seconds in A. Each data point is the average across the respective n subjects. Each red asterisk, with its accompanying p value, indicates a significant difference between two linked data points by matched sample t-tests.

5 A proposal of eccentricity-dependent feedback mechanisms from higher brain areas to V1

5.1 Feedback from higher visual areas to V1 could explain the perceptual bias towards the S_+ signal

The bias to perceive the input signals S_+ in the ocular summation channel may arise from an analysis-by-synthesis process associated with the visual decoding of ambiguous inputs. In analysis-by-synthesis, higher visual stages synthesize or generate the visual input signals that are expected under the hypothesis that a particular visual scene property (e.g., upward drift) holds, and compare this synthesized input with the actual visual inputs to see how well they match. If they indeed match, the hypothesized scene property is more likely to be perceived than otherwise. For example, with ambiguously tilted input gratings, the higher visual stages decide between two alternative hypotheses, one stating that the input arises from a grating tilted clockwise from horizontal, the other stating that the input arises from a grating tilted counterclockwise. To do so, they synthesize the visual inputs associated with each hypothesis, and compare them with the actual

visual inputs. The hypothesis that gives a better match between the synthesized and actual inputs is more likely to be the perceptual outcome, at least if the two alternative hypotheses are equally likely before the degrees of the matches are assessed.

Prior knowledge about visual input properties is used to generate the synthesized visual inputs, leading to perceptual biases. For example, normal visual experience leads to the prior knowledge that the inputs to the two eyes are usually similar to each other (Li and Atick, 1994). Hence, if the higher stages hypothesize a clockwise-tilted grating in a fronto-parallel plane where the axes of two eyes converge, they expect a clockwise-tilted grating as inputs to both eyes and thus expect a signal for this grating to be substantial in V1's S_+ channel but much weaker or near-zero in V1's S_- channel. Based on this expectation, the synthesized visual input is a clockwise-tilted grating in the S_+ channel. Under normal, ecological, circumstances, with typical attended inputs to the two eyes, the prior expectation is duly true, leading to a close match between the synthesized input and the actual input in the S_+ channel [1].

The initial perceptual hypotheses are presumed to be suggested by external or feedforward sensory inputs. Fig. 9 shows a conceptual model of a hierarchical network that performs analysis-by-synthesis. For convenience, we use the acronym FFVW — for feed-forward, feedback, verify, and weight — to refer to this sequence of processes: feed the input signals from V1 forward to suggest an initial set of perceptual hypotheses to higher centers, feed back from the higher centers the synthesized visual inputs (generated according to prior knowledge) for the hypotheses, verify by quantifying how closely the feature values in the synthesized inputs match those in the actual inputs, and weight the relative influences of the perceptual hypotheses in the final perceptual outcome according to the relative degrees of match. In Fig. 9, features A and B refer to alternative feature values, perceptual hypotheses, or perceptual outcomes in general. There are two critical facets of the FFVW network: first, feedforward signals are sent from both the S_+ and S_- channels in V1, so that both these channels contribute to the initial hypotheses for perception; second, feedback signals are dominantly aimed at the S_+ channel for verification because prior knowledge assumes that the percept arises mainly from the S_+ signals.

Fig. 9 also illustrates an example of applying the FFVW network to the ambiguous input from Fig. 1A. This ambiguous input is not usually encountered in ecological scenes. In this example, 'Feature A' and 'Feature B' are clockwise-tilt and anti-clockwise-tilt, respectively. These two features are both present in the input to each eye, but the exact input composition differs between the two eyes such that the 'Feature A' inputs from the two eyes cancel each other in V1's S_- channel, and the 'Feature B' inputs from the two eyes cancel each other in V1's S_+ channel. V1 sends the information about these two features — one is active in the S_+ channel and the other is active in the S_- channel — to higher visual stages, thereby suggesting two different hypotheses about the tilt direction. For each hypothesis, clockwise or anti-clockwise tilt direction, the synthesized input signal is a signal of this same tilt direction and is assumed to be mainly in the S_+ channel according to the prior knowledge. This is regardless of whether the hypothesis has originated from (a feedforward signal from) V1's S_+ or S_- channel. Such synthesized inputs are generated in the higher centers and fed back to V1 to be compared with the actual signals in V1's S_+ channel. However, only the synthesized input for the hypothesis originating from V1's S_+ channel matches the actual input in S_+ channel, specifically, the feature value "clockwise tilt" in this synthesized input matches the feature value in the actual input in S_+ channel. Consequently, the clockwise tilt as a hypothesis is reinforced, but the

[1] Note that the S_- channel should contain a non-zero signal for a clockwise-tilted grating if such a grating in a visual scene is viewed at a non-zero disparity. However, regardless of this, generally, the S_+ signal should still show a clockwise-tilted grating, typically S_+ and S_- channels agree in the tilt direction of the grating, and the signal power in the S_+ channel is generally stronger for ecological scenes. If vision aims to infer the tilt rather than disparity of the grating, then matching the synthesized and actual inputs in the S_+ channel is sufficient and whether the same tilt feature is present in the S_- channel is irrelevant. See later discussions for more complex cases such as motion in depth.

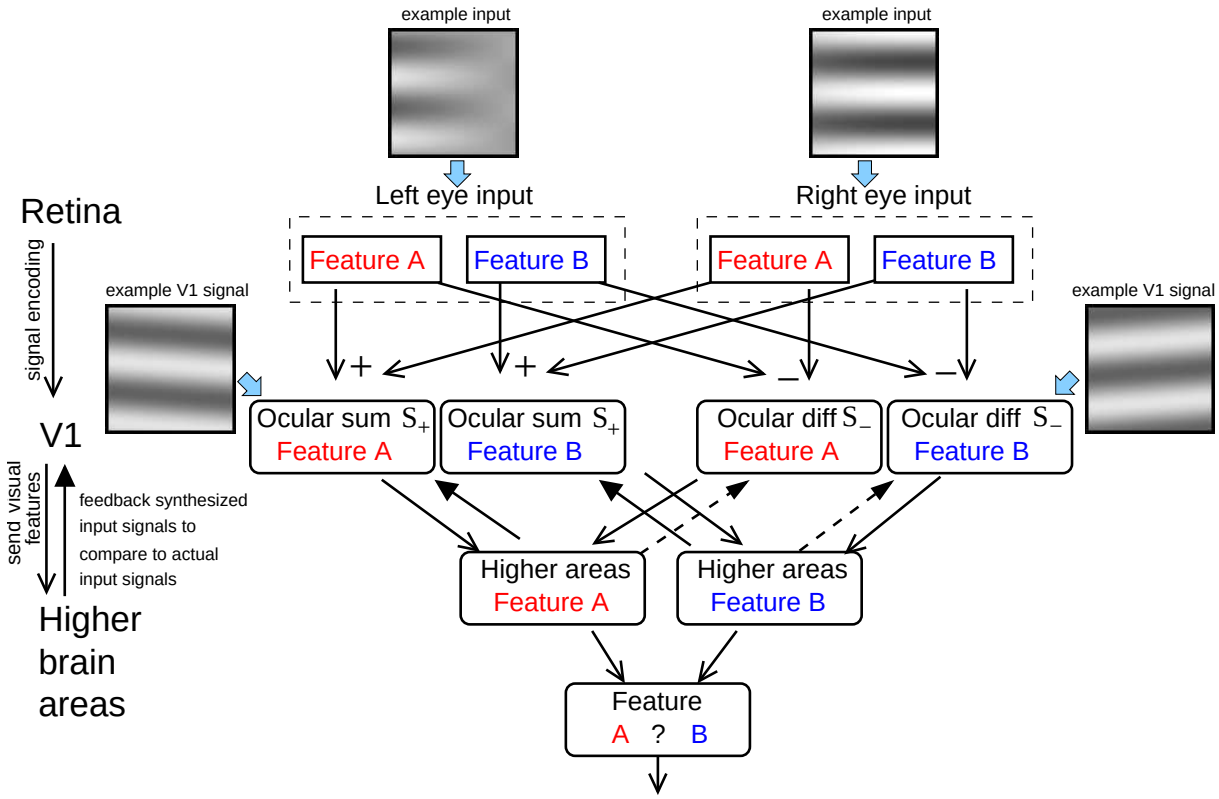

Figure 9: Conceptual model of a hierarchical network of the FFVW (feedforward, feedback, verify, and weight) circuit for visual decoding. The features, e.g., tilt orientations or motion directions, are fed forward for eventual perceptual recognition, regardless of their eye of origin or whether they arise from the ocular summation S_+ or difference S_- channel. However, for decoding most (non-depth) visual features, the synthesized signals are fed back mostly to the ocular summation S_+ channel for input verification. The dashed arrows indicate much weaker or absent feedback routes. Note that information about the eye-of-origin of visual inputs (or about whether the signal is from the S_+ or S_- channel) is absent after V1 in this network. The activation levels of the 'Feature A' and 'Feature B' units in V1's $S_\pm$ channels depend on the retinal input. For the example retinal input shown in the grayscale images (taken from Fig. 1A), the activation level of the 'Feature A' (clockwise tilt) unit is non-zero in V1's S_+ channel but not S_- channel, and that of the 'Feature B' (anti-clockwise tilt) unit is non-zero in V1's S_- channel but not S_+ channel.

alternative hypothesis, the anti-clockwise tilt, is not reinforced, making the former hypothesis the more likely perceptual outcome.

Therefore, the FFVW process, involving V1 and higher visual centers for the purpose of analysis-by-synthesis using prior knowledge, is hypothesized to contribute to the perceptual bias towards signals arising from S_+ in the ambiguous perception in the current study.

5.2 Difference between central and peripheral vision in the feedback

5.2.1 Proposal: a weaker feedback for analysis-by-synthesis in peripheral vision

In the previous subsection, we proposed that feedback from higher to lower visual areas is a likely basis for the perceptual bias towards the S_+ signal. Since this bias is weaker or even absent in the peripheral locations, we propose here in addition that this feedback is weaker or absent in the peripheral visual field. This proposal is consistent with the observation in Fig. 8 that central and peripheral vision differ in their dependence on input stimulus properties that can affect the effectiveness of feedback signals. Assuming that it takes time for feedback and verify to develop and act effectively, we can expect that longer viewing durations should give rise to greater feedback-based biases towards S_+. Fig. 8B shows that this is indeed the case for central, but not for peripheral, vision. For drifting gratings, if feedback involves generating dynamically changing synthesized input signals for matching with bottom-up input, or if feedback matching involves visually tracking the drifting gratings, it would be more effective when inputs drift slowly. Indeed, Fig. 8A shows that, for the drift-direction percept, the bias towards S_+ decreases with higher drift frequency for central but not peripheral vision.

5.2.2 Ruling out a potential confound in the sizes of the central and peripheral test stimuli

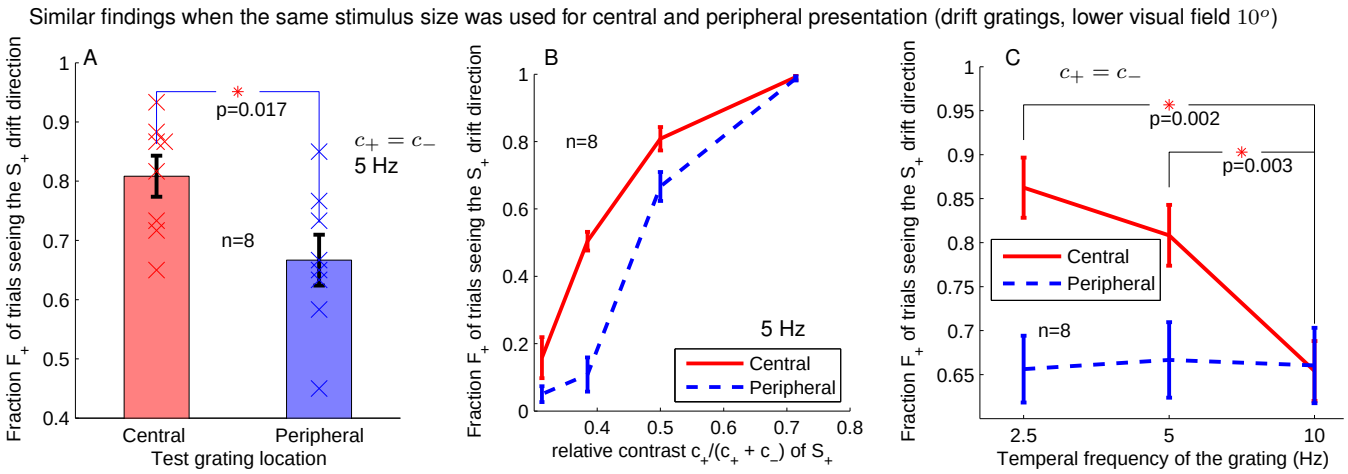

Figure 10: The difference between the central and peripheral biases towards S_+ persisted in the control experiment (involving $n = 8$ subjects) when the centrally and peripherally viewed stimuli were of the same size. Results in A, B, and C are the counterparts of those shown in Fig. 4A, Fig. 6A, and Fig. 8A, respectively, from the main experiment involving the same design and conditions except for the size of the centrally viewed stimuli.

In the main experiment, the central stimuli were smaller than the peripheral ones; this difference in size might be an alternative cause for the different perceptual biases. To rule this out, we conducted a control experiment in which the centrally viewed stimuli were enlarged to be of the same size as those in the periphery. Other than this change in size, the control experiment had exactly the same design and details as the version of the main experiment that tested ambiguous perception of drifting directions involving the lower visual field at 10° eccentricity (the result of this main experiment are shown in Fig. 4A, Fig. 6A, and Fig. 8A). Fig. 10 shows that the relative difference between central and peripheral vision in terms of their biases towards S_+ is little affected by this change in stimulus size for the central presentation. This also

implies that our conclusion is not sensitive to whether the sizes of the stimuli are exactly normalized by the visual spatial resolution at each eccentricity.

5.2.3 Ruling out a potential confound in the visual sensitivities to central and peripheral inputs

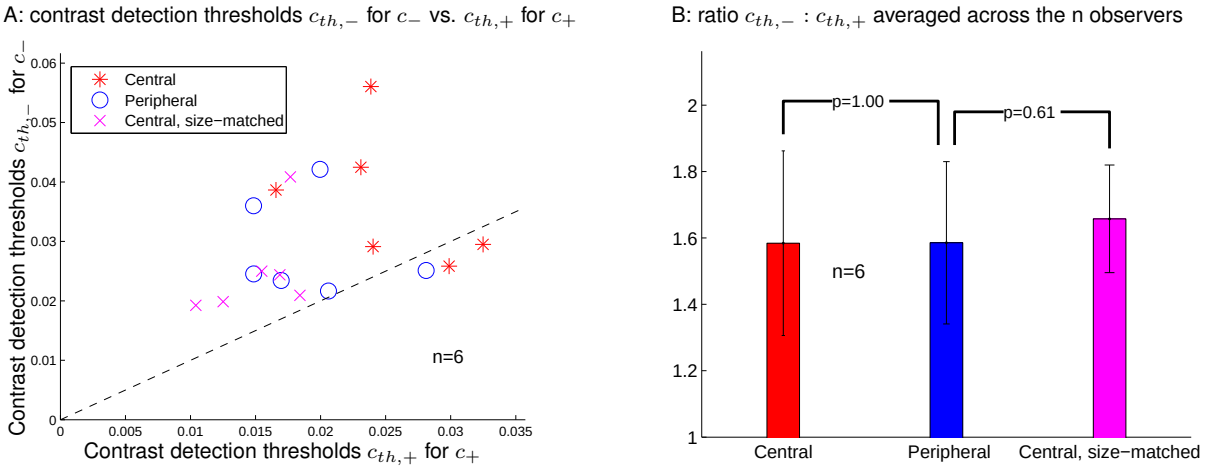

Figure 11: Thresholds $c_{th,+}$ and $c_{th,-}$ to detect input contrasts c_+ and c_- for S_+ and S_- inputs, respectively, obtained in the auxiliary experiment involving $n = 6$ subjects (see Methods) using gratings drifting at 2.5 Hz for $T = 0.2$ seconds. The peripheral location was in the lower visual field at 10° eccentricity. A: the individual thresholds for the six subjects, when the gratings were presented centrally at the size used in the main experiment for the central location (red asterisk), peripherally at the size used in the main experiment for the peripheral location (blue circle), or centrally at the size used in the main experiment for the peripheral location, i.e., as in the control experiment (magenta 'x'). The dashed line indicates $c_{th,+} = c_{th,-}$. B: the ratios between $c_{th,-}$ and $c_{th,+}$ averaged across the six subjects, at the three conditions of stimulus locations and/or sizes. The p values marked in the plot came from the matched-sample t-test for any difference between the indicated ratios.

A second potential confound is that central and peripheral V1 may differ in the relative contrast sensitivities g_+ and g_- to the sum contrast c_+ and the difference contrast c_-, respectively. If the ratio g_+/g_- is larger for central than for peripheral vision, it could cause a larger perceptual bias towards S_+ without implying anything about feedback mechanisms.

We thus designed an auxiliary experiment to assess g_+ and g_- by measuring the contrast detection thresholds $c_{th,+}$ and $c_{th,-}$, respectively, for input contrasts c_+ and c_- of the S_+ and S_- signals. In this experiment, we used as S_+ and S_- drifting gratings like those in the corresponding main and control experiments except that the grating contrasts were near the detection thresholds and the test grating in each trial was either a purely S_+ or a purely S_- grating (see Method). Fig. 11 shows the results for six observers. The gratings were viewed in three different ways: (1) centrally at a size that was the same as for the central location in the main experiment; (2) peripherally at a size that was the same as for the peripheral location in the main experiment; (3) centrally at a size that was the same as for the peripheral location in the main experiment, i.e., as in the control experiment. Overall, the threshold $c_{th,+}$ was smaller than the threshold $c_{th,-}$. However, the ratio of thresholds $c_{th,-}$: $c_{th,+}$, which reflects the ratio of contrast sensitivities g_+/g_-, was insensitive to whether the inputs were presented centrally or peripherally, and whether the central inputs were smaller than the peripheral inputs. Note that, in these test gratings of contrast $c_\pm$, the contrasts of the

monocular grating contrast were $c_{\pm}/2$, identical between the two eyes; however, g_+ was about 50% larger than g_- in our data. If the observers had adopted a strategy to detect the grating by detecting either of the two monocular gratings, this would give $g_+ = g_-$.

The findings above suggest that the difference in perceptual bias between central and peripheral vision in our data cannot arise from a difference in the relative sensitivity to the ocular summation over the ocular difference signals. Note that this conclusion was reached using gratings drifting at a frequency of 2.5 Hz, which among our measured frequencies demonstrated the largest difference between central and peripheral vision in the perceptual bias towards the S_+ signal (see Fig. 8A and Fig. 10C). These findings support our proposal that the difference between central and peripheral vision in their biases towards the S_+ signal arises from a difference in their feedback mechanisms.

6 Summary and Discussion

6.1 Summary of the experimental findings and the computational proposal

In this study, we used dichoptic visual input stimuli made as the superposition of two gratings, one in the binocular sum channel S_+ and the other in the binocular difference channel S_-. These two gratings have different motion directions, tilt orientations, or colors, respectively, hence the dichoptic inputs evoke ambiguous percepts in the corresponding feature dimension (motion, tilt, or color). In central vision, we found that the percept is biased towards the grating in the S_+ rather than the S_- channel. However, in peripheral vision, this perceptual bias was weaker or absent.

We propose that, at least for central vision, the perceptual bias towards the signals in the S_+ channel arises partly from top-down feedback from higher visual centers to the primary visual cortex (V1), and that this feedback is involved in the computation of analysis-by-synthesis for visual recognition. Further, we propose that the weaker perceptual bias in the periphery reflects a weaker or absent feedback. This proposal is consistent with the additional observations that, in central but not peripheral vision, the perceptual bias towards the S_+ signal increases when visual inputs are viewed for longer, or change more slowly, as these input conditions should make feedback more effective. This proposal is also consistent with the observation that the difference between the central and peripheral biases is present even when central vision and peripheral vision do not differ in the contrast sensitivity to the ocular summation signal relative to the contrast sensitivity to the ocular difference signal.

6.2 The FFVW network for feedforward, feedback, verify, and weight

We use a conceptual network in Fig. 9 to explain how perceptual decisions are reached in the face of ambiguity. This involves the FFVW process — feedforward, feedback, verify, and weight — that performs analysis using the synthesis. In this FFVW process, a perceptual hypothesis suggested by the feedforward visual inputs is weighted more heavily in the ultimate percept when its synthesized visual input, generated by the higher centers and fed back to V1 for verification, better matches the actual visual input in V1.

Accordingly, the perceptual bias towards S_+ arises because the synthesized visual inputs are generated according to the prior knowledge that visual inputs are normally binocularly correlated. In particular, the synthesized visual inputs are presumed by the prior knowledge to arise mainly or exclusively from the ocular summation S_+ channel. Therefore, the synthesized visual inputs are fed back to V1's S_+ channel for verification, regardless of whether these synthesized inputs arise from a perceptual hypothesis suggested by the feedforward visual input from the S_+ or the S_- channel. Hence, in the FFVW model, the feedforward

signals for input properties such as motion, tilt, and color, arise from both the S_+ and S_- channels in V1, while the feedback signals are mainly directed to the S_+ channel for verification; see Fig 9. This model explains the perceptual biases in Figs. 4 –7. Meanwhile, the proposal that the feedback is weaker in periphery is consistent with the central-peripheral differences shown in Figs . 4 –8, 10.

In the FFVW network shown in Fig. 9, information about the eye-of-origin of visual inputs is absent beyond V1 along the visual pathway, as suggested by physiological data (Hubel and Wiesel, 1968; Burkhalter and Van Essen, 1986). However, this information is needed in order to apply the prior bias towards percepts arising from binocularly correlated rather than anticorrelated visual inputs. This implies that mechanisms to apply the prior bias cannot arise purely by feedforward processes beyond V1, and that they must involve feedback towards V1 to access eye-of-origin information assuming that perceptual decisions are ultimately reached beyond V1.

This FFVW network is meant for perception of natural inputs, and unnatural inputs such as our dichoptic test gratings or inputs that evoke binocular rivalry are tools that can be used to diagnose its mechanisms. In natural vision, typical inputs to the two eyes are such that the inputs to V1's S_+ channel and the inputs to V1's S_- channel convey the same feature values. For example, a clockwise-tilted grating in both eyes leads to the clockwise-tilt feature value (let it be 'Feature A') in the $S_+ = S_L + S_R$ channel, activating V1's binocular cells tuned to this tilt orientation. Typically, depending on whether $S_L = S_R$ holds exactly (when the two eyes' axes converge precisely on the attended visual location) or not, the $S_- = S_L - S_R$ will be zero or otherwise activating a V1 cell tuned to this clockwise-tilt and the corresponding binocular disparity. Hence, 'Feature A' but not 'Feature B' units in V1 are activated, in S_+ channel alone or in both S_+ and S_- channels. This feature value is fed forward from both S_+ and S_- channels as the initial perceptual hypothesis, whose synthesized input naturally matches the actual input in the S_+ channel in their feature value ('Feature A') to give an unambiguous perceptual outcome of clockwise-tilt.

However, if (as in studies on binocular rivalry) the input to the left eye S_L is a clockwise-tilted ('Feature A') grating but that to the right eye S_R is an anti-clockwise-tilted ('Feature B') grating of the same contrast, then both the 'Feature A' and 'Feature B' units will be activated in V1's S_+ channel *and* in V1's S_- channel (these S_- units could manifest themselves as monocular neurons or neurons that are tuned-inhibitory (Poggio and Fischer, 1977)). Both 'Feature A' and 'Feature B' are fed forward as initial perceptual hypotheses with the same strength; their respective synthesized inputs would have the same degree of match with the actual inputs in V1's S_+ channel, giving no bias towards either of them in the ambiguous perceptual outcome.

The circuit model in Fig 9 is purely conceptual. It specifies the functional goals and routes of the feedforward and feedback pathways, but does not specify the implementation details of the FFVW process. For example, it does not specify whether the feedback should excite or inhibit V1 or leave V1 activities unchanged, or indicate the number of separate neural populations within each network hierarchy for the respective computational process, or the areas and or even the number of hierarchical levels above V1 in the FFVW circuit. These details are currently poorly constrained; there are experimental data and computational arguments for and against various of these possibilities, depending on perceptual tasks, visual inputs, and context (Murray et al., 2002; Hsieh et al., 2010; Carpenter and Grossberg, 2011; Cardin et al., 2011; Kok and de Lange, 2014; Chen et al., 2014b; Qiu et al., 2016). For our purposes, the implementation details are not essential, as long as the FFVW processes achieve the functional goal of analysis-by-synthesis, and the synthesized input signals are generated according to prior knowledge reflecting the statistical properties, notably the binocular correlation of natural visual inputs.

6.3 Theoretical and anatomical motivations for the central-peripheral difference in feedback

To explain the difference between central and peripheral vision in their perceptual biases towards the S_+ signal, a difference in feedback mechanism for analysis-by-synthesis is proposed. This proposal is consistent with the following theoretical and anatomical considerations.

Theoretically, it has been argued (see Zhaoping (2014), section 5.7) from a perspective that vision is composed, roughly, of (1) input *selection* and (2) *decoding*, that central and peripheral vision should have different functional emphases. This argument starts from the observation that the brain has limited resources, and so needs to employ an attentional bottleneck (Sziklai, 1956) that admits only a small fraction of visual inputs for deeper processing. It is then proposed that this attentional bottleneck operates by (1) selecting a visual field location and bringing it to the attentional spotlight, and then (2) decoding or inferring properties (e.g., tilt direction, face identity) of the visual scene from the selected visual inputs within the spotlight. Finally, based on the observation that in natural behavior, gaze position cannot help but follow the position of attentional spotlight very quickly (Deubel and Schneider, 1996; Hoffman, 1998), we conclude that before a visual input is selected, it typically lies outside the central visual field, but that once selected, it lies inside the central visual field for most of the duration of visual decoding. It then naturally follows that peripheral vision should take the lead in the role of *selecting* the small fraction of visual input to be attended and bringing it to the attentional spotlight, and that, in contrast, central vision should take the lead in the role of *decoding* (i.e., recognizing) properties of the visual scene from the selected (i.e., attended) visual input. Note that this distinction between the roles of central and peripheral vision is not inferred from performance data about central-versus-peripheral vision (as reviewed, e.g., in Strasburger et al. 2011) but is a necessary consequence of the perspective that vision involves separate stages of selection and decoding. Analysis-by-synthesis via feedback is a mechanism for decoding or recognizing scene properties (e.g., input tilt direction) from the selected (i.e., attended) inputs; it therefore makes sense that such a mechanism should be stronger in central vision.

Anatomically, while V1 represents (Gattass et al., 1987) the full visual field, which extends to an eccentricity of up to 107° laterally (Rönne, 1915; Traquair, 1938), there is a trend (Zhaoping, 2011) that the fraction of resources devoted to representing or responding to peripheral input progressively decreases in higher cortical areas along the visual pathway, especially those in the ventral stream. For example, V3/V4 represents up to eccentricity $35 - 40^\circ$, even when considering the full spatial extent of the neural receptive fields collectively (Gattass et al., 1988). Also, medial temporal cortex (MT/V5) represents up to perhaps no more than 60° (Fiorani et al., 1989), and the inferotemporal cortex (IT) mainly represents central vision (Baizer et al., 1991). These observations come from monkeys; human data on far eccentricities are scarce since they mostly come from experiments using functional magnetic resonance imaging (fMRI) which typically present visual inputs within a limited eccentricity (e.g., $< 15^\circ$, Sereno et al. (1995)). Nevertheless, the same trend is suggested in human fMRI data: as a visual input centered on the fovea increases in size from 3° to 16° in radius, the responding cortical surface area expands considerably in V1 but either very little, or not at all, in higher areas such as the fourth human visual field map (hV4) and the ventral occipital cortex (VO) (Brewer et al., 2005). If a higher visual area does not represent a peripheral location, it is more constrained in providing feedback associated with this peripheral location, and thus we may expect that feedback to V1 from peripheral locations outside the fovea becomes weaker and is eventually absent further downstream along the visual pathway, particularly along the ventral stream. In our experiments, the eccentricity of the peripheral location is 10° or less and is most probably still represented in the higher visual areas, since humans can recognize characters at an eccentricity of up to at least 40° (Aubert and Foerster, 1857; Strasburger et al., 1994; Strasburger and Rentschler, 1996).

6.4 The relevant feedforward process in V1

The feedforward process which integrates visual inputs from the two eyes first occurs in the primary visual cortex V1. This is achieved by V1 neurons with a diversity of receptive field properties. Some neurons are binocular, others are monocular. One group of neurons, called tuned-excitatory cells, are excited by near-zero binocular disparities; a second group of neurons, called tuned-inhibitory cells, are inhibited by near-zero disparties; and a third group of neurons are tuned to near or far disparties (Poggio and Fischer, 1977; Ohzawa et al., 1990; Parker, 2007). Although the binocular tuned-excitatory cells and tuned-inhibitory cells mainly respond to the S_+ and S_- inputs, respectively, in general, a V1 neuron reports a weighted sum of the S_+ and S_- inputs (Li and Atick, 1994). Hence, while the S_+ and S_- signals are typically decorrelated with separately adaptable input sensitivities according to the efficient coding principle (Barlow, 1961; Li and Atick, 1994), they are multiplexed in their representation in the population responses from the V1 neurons. Hence, individual V1 neurons typically respond both to binocularly correlated inputs and to binocularly anti-correlated inputs with different sensitivities (Parker, 2007).

The behavioral contrast sensitivities for $S_\pm$ shown in Fig. 11 should reflect the feedforward input sensitivities for $S_\pm$ in the underlying neural population, in the same way that behavioral contrast sensitivities for the luminance and chromatic signals reflect neural contrast sensitivities (Mullen, 1985; Zhaoping, 2014). At least for the test stimuli used for the results in Fig. 11, central vision and peripheral vision do not differ in their behavioral sensitivities to the S_+ signal relative to their sensitivities to the S_- signals. This suggests that our observed difference between central and peripheral vision in the perceptual bias towards S_+ arises from feedback processes.

While we have argued that prior knowledge about the statistics of visual inputs is incorporated into the feedback process associated with analysis-by-synthesis, such priors also affect feedforward processes. In particular, the inputs to the two eyes are in general correlated; the correlation coefficient is close to unity for inputs of large spatial scale and decays with input spatial frequency (Li and Atick, 1994). According to the prescription of efficient coding (Barlow, 1961; Li and Atick, 1994; Zhaoping, 2014), these input statistics should directly affect the feedforward sensitivities to S_+ and S_-. Indeed, the input sensitivities to $S_\pm$ should quickly adapt to changes in the degree of binocular input correlation. Such adaptation has indeed been observed psychophysically (May et al., 2012; May and Zhaoping, 2016), and is a special case of the general adaptation of input contrast sensitivities.

Since our behavioral data suggest that feedforward input sensitivity g_+ for S_+ is generally higher than g_- for S_- for the testing gratings used in this study (Fig. 11), some of the perceptual bias towards S_+ in our test gratings could arise from this feedforward preference for S_+. The auxiliary experiment showed that this could not explain the difference between central and peripheral biases. However, our data do not specify the respective contributions of feedforward and feedback processes to the residual bias towards S_+ in the periphery. This should be investigated in future studies.

Although the ratio (g_+/g_-) between the contrast sensitivities to the S_+ and S_- gratings for central vision is the same as that for peripheral vision (for the data in Fig. 11), in general, central and peripheral vision do differ greatly even in the purely feedforward processes from retina to V1 (in addition to their difference in spatial resolution). For example, color vision in fovea is very different from that in periphery in terms of input sampling on retina (Packer and Williams, 2003) and in behavioral sensitivity (Mullen and Kingdom, 1996).

More relevantly, Durand et al. (2007) reported that central and peripheral vision have different distributions of preferred (binocular) vertical disparities of V1 neurons. In particular, they found that, among neurons tuned to near or far disparities (i.e., excluding neurons excited or inhibited by zero or near-zero disparities), the range of (absolute) preferred disparities is larger in periphery. We can ask whether their observations contradict our finding in Fig. 11 that there is no difference between central and peripheral vision

in the behavioral sensitivity to S_+ relative to that to S_- signals. By construction of their receptive fields (Zhaoping, 2014), neurons tuned to large (absolute) disparities are sensitive to both S_+ and S_- inputs, and their sensitivity to S_- tends to be larger when their preferred (absolute) disparity is larger. This implies that, among V1 neurons that are not tuned to or not inhibited by zero-disparity (or near-zero-disparity) inputs, sensitivity to S_- is likely higher in peripheral vision than in central vision. However, the tuned-inhibitory cells, which are inhibited by zero or near-zero disparity inputs, are mainly sensitive to S_- inputs by virtue of their receptive field constructions (Li and Atick, 1994; Zhaoping, 2014). This suggests that V1 neurons for central vision have the capacity to be sensitive to S_- as a population. It could therefore happen that, even though central and peripheral V1 neurons have different distributions of preferred (vertical) disparities, at a neural population level the sensitivity to S_+ relative to the sensitivity to S_- within the central V1 neurons is the same as that within the peripheral V1 neurons. This is consistent with our data on behavioral sensitivities (Fig. 11).

6.5 Temporal dynamics in the FFVW process for analysis-by-synthesis

In reporting the data in Fig. 8, we assumed that the FFVW process works more effectively when visual inputs can be viewed for longer or when they evolve more slowly. This assumption may seem to contradict the idea that feedback signals between cortical areas can be transmitted quickly (based on the observations that feedforward and feedback fibers between cortical areas have similar conduction speeds (Nowak et al., 1997), and that the feedforward conduction speed is very large (Movshon and Newsome, 1996)). Furthermore, human visual perception can occur so quickly that observers can make complex judgements after a viewing duration of 20ms, such as whether or not an animal is present in an image (Thorpe et al., 1996). However, we argue that these findings do not present a conclusive challenge.

Firstly, for simple visual recognition with clear visual input, the feedforward signals in the FFVW circuit in Fig. 9 can already unambiguously favor one percept over the alternatives. Thus, a perceptual decision can be reached before the completion of the feedback, verify, and weight processes. This is presumably why some simple visual recognition tasks are feasible with very brief visual presentations – the feedforward signals in V1 need not even be present for long enough to verify a perceptual hypothesis. However, even if visual recognition is feasible by a 20-ms viewing duration, it takes more than 100 ms for signals in higher brain areas, e.g., inferotemporal cortex, to distinguish between alternative perceptual hypotheses (Thorpe et al., 1996; Hung et al., 2005) , i.e., brief viewing duration does not mean short reaction time to report percepts. This is so despite the fast feedforward fibers, presumably because information processing requires intra-cortical computation in each brain area along the visual pathway in addition to the feedforward signal transmission from one brain area to another.

Secondly, when inputs are ambiguous, perception takes much longer. For example, with typical dichoptic stimuli used for evoking binocular rivalry, a percept of one monocular image does not dominate over that of the other monocular image until the rivaling inputs have been viewed for at least several hundred milliseconds (Blake, 2001). Presumably, in ambiguous perception, the feedforward component in the FFVW process is typically insufficient for reaching a clear percept, and apparently, the subsequent feedback, verify, and weight components in the full FFVW process take time. It is also likely that the full FFVW process can take multiple cycles of feedforward, feedback, verify, and weight, like an iterative computational algorithm (Dayan, 1998), before the perceptual decision process decides on a particular perceptual hypothesis.

Thirdly, physiological observations are consistent with the presence of feedback, verify, and weight processes in difficult visual recognition tasks. For example, transcranial magnetic stimulations (TMS) of extrastriate cortex MT and striate cortex V1, respectively, evoke qualitatively different phosphene percepts. However TMS on V1 25 ms after, rather than before, TMS on MT evokes a percept consistent with that

from stimulating V1 alone (Pascual-Leone and Walsh, 2001), as if the MT-evoked perceptual hypothesis is vetoed by the V1-evoked hypothesis during the feedback-and-verify process.

In another example, a small figure in a background texture in a briefly presented image is not easy to detect. It can be made invisible when the input image is followed immediately by a masking image. Event-related potentials (ERPs) measured by electroencephalography reveal that the figure evokes an early ERP (78-108 ms after stimulus onset) in the occipito-temporal areas with or without the masking input, but masking eliminates a subsequent ERP (109-141 ms after onset) in the occipital area and a later ERP (180–305 ms after onset) in the occipito-temporal areas (Fahrenfort et al., 2007). The early, subsequent, and later ERPs, respectively, can be interpreted as reflecting the initial feedforward signals in the higher cortical areas, the feedback-and-verify process in the lower cortical areas, and perceptual confirmation from the weighting process again in the higher cortical areas. Masking apparently interfered in the feedback-verify-weight component of the FFVW process.

Third, successful recognition of the underlying visual object when viewing a two-tone Mooney image made the fMRI signals in early retinotopic cortices (V1, V2, V3) resemble those evoked by the grayscale photo version of the two-tone image (Hsieh et al., 2010). This is as if the synthesized input (which should represent the grayscale photo version of the Mooney image), generated from the perceptual hypothesis and fed back from higher brain areas, modified the neural activations in the lower areas evoked by the actual two-tone input images.

Fourth, when monkeys performed a demanding task of detecting a contour in a noisy background, the contour's presence was difficult to discern in measured V1 responses until about 30 ms after it could be identified in responses measured from extrastriate cortex V4 (Chen et al., 2014b). Feedback from V4 to V1 during the feedback-and-verification processes in this difficult task would be consistent with this observation.

It was notable that the perceptual biases in the periphery did not follow the same trend as in the center when drift frequency and stimulus duration of the test grating changed (Fig. 8). One could speculate that there may be no FFVW process in the periphery at all. Note, though, that each trend in Fig. 8 reflects at least a retina-to-V1 feedforward process as well as a FFVW process (if any). The retina-to-V1 process is manifest in input contrast sensitivity. Typically, as the temporal frequency of visual inputs increases (or, roughly, as input duration decreases), input contrast sensitivity increases for the sum channel and decreases for the opponency channel. Under efficient coding principles, this is required to smooth out input noise (see Zhaoping (2014)). For example, input sensitivity at higher temporal frequencies tends to favor luminance rather than chromatic input channels, which are the cone-summation and cone-difference channels, respectively. Hence, the contributions of retina-to-V1 and FFVW processes to the trend in each curve in Fig. 8 are mutually opposing. Apparently, from data in Fig. 8, the FFVW contribution dominates in central vision, whereas the retina-to-V1 contribution dominates for peripheral vision. Finally, central and peripheral vision differ in their temporal contrast sensitivity functions in the retina-to-V1 process: peripheral vision is sensitive to higher temporal frequencies than central vision (Hartmann et al., 1979; Carrasco et al., 2003) perhaps due to a larger proportion of the magnocellular cells in periphery (Azzopardi et al., 1999). Note that the higher temporal sensitivities help peripheral vision in its role in input selection to favor the selection of locations where inputs have changed.

6.6 Testable predictions from the computational framework

One important and testable prediction from our study is that top-down feedback to the primary visual cortex for the purpose of visual recognition should be stronger in central than peripheral visual field locations. In testing this prediction, one should note that, because the feedback that we have addressed is mainly involved

in recognition (answering about 'what', rather than 'where', a visual input is), this prediction should be more applicable to the ventral rather than the dorsal visual pathway (Ungerleider and Mishkin, 1982).

While we await experimental tests of this prediction, there are some other related experimental data. First, regions across different visual cortical areas responding to the fovea converge to the same or nearby anatomical locations in the brain, while those responding to the periphery are spatially much more distant from each other (Sereno et al., 1995). Since the anatomical organization should respect a need to minimize the volume of the brain devoted to neural connections, the closer vicinity between the foveal representations across different visual cortical areas suggests denser neural connections between the areas devoted to central than peripheral vision. The feedback connections are likely responsible for much of the difference between central and peripheral vision in terms of their inter-area connection densities, since feedforward connections between areas must be present whenever the periphery is represented in any higher visual area.

Another testable prediction from the current study is that feedforward connections from V1 should arise from both the S_+ and S_- channels, whereas the feedback connections to V1 mainly target the S_+ channel, at least for visual inputs in a fronto-parallel plane near the fixation. To test this prediction, one needs to distinguish between V1 neurons that code S_+ from V1 neurons that code S_- signals. From section 6.4, a V1 neuron mainly responds to S_+ signals if it is binocular, has similar sensitivities to the left-eye and right-eye input, and is tuned to near-zero disparity. For example, it can be a binocular and tuned-excitatory neuron (Poggio and Fischer, 1977) and, anatomically, is likely to be far from the center of any ocular dominance column in V1. One should examine whether the feedback preferentially targets one group of V1 neurons versus another according to the prediction. Meanwhile, the prediction that the S_+ channel is preferentially targeted by feedback may not apply to situations involving substantial depth percepts, such as percepts of motion in depth (Harris et al., 2008) which is more associated with the dorsal rather than the ventral visual pathway. This is because, for these percepts, the synthesized inputs should expect there to be a substantial difference between the two monocular inputs, so that the feedback for input verification should target the S_- channel.

6.7 Dichoptic inputs as tools rather than restrictions

While some previous studies aimed to investigate phenomena or mechanisms specific for dichoptic inputs (Shadlen and Carney, 1986; May et al., 2012; May and Zhaoping, 2016; Kingdom and Wang, 2015), we have used dichoptic stimuli as a window through which to observe the difference between central and peripheral vision in terms of their visual decoding mechanisms, particularly about the properties of top-down feedback for visual decoding. Hence, our predictions regarding the properties of feedforward and feedback pathways from and to V1 should apply to general visual inputs beyond the dichoptic ones used here.

Dichoptic inputs are merely convenient tools for our purpose since they can be easily made to evoke ambiguous percepts, which can hopefully exercise suitable parts of the system for visual recognition, most particularly feedback. Our kind of dichoptic stimuli are an alternative to those that evoke binocular rivalry (i.e., rivalry between percepts from the two monocular inputs (Blake, 2001)), which also provide insights into the mechanisms of visual decoding. In this study, we took advantage of the different statistical properties associated with the S_+ and S_- signals to search for new insights into visual-perceptual mechanisms, rather than using the similar or identical statistical properties associated with the left and right eye inputs. In particular, we probed the feedback mechanism for the computation of analysis-by-synthesis based on the observation that this relies on prior knowledge, which in turn, favors binocularly correlated inputs.

6.8 Concluding remarks

There has been more than a century of research on the differences between central and peripheral vision; see the review by Strasburger et al. (2011). The differences investigated so far include those manifested in a wide range of psychological or perceptual phenomena, input sampling in the retina, neural receptive field properties in the retina, lateral geniculate nucleus, and primary visual cortex (Levi, 2008; Strasburger et al., 2011; Yu et al., 2015). In addition, differences between central and peripheral vision have been observed in various brain networks, mainly in terms of feedforward connections along the visual pathway (Baizer et al., 1991; Schall et al., 1995) or functional connections between various brain regions (Griffis et al., 2016). We added to these past studies by asking questions about the the top-down feedback circuit from higher brain areas to V1, motivated by the functional difference between central and peripheral vision in visual selection and visual decoding. Experimental tests of the predictions arising from this study should hopefully guide us further in our quests.

Acknowledgements: this work is supported in part by the Gatsby Charitable Foundation. I am very grateful to Hans Strasburger and an anonymous reviewer for their very helpful comments on the manuscript, and to Hans Strasburger and Peter Dayan for help with English editing.

References

Anstis, S. M. (1974). A chart demonstrating variations in acuity with retinal position. *Vision Research*, 14:589–592.

Aubert, H. R. and Foerster, C. F. R. (1857). Beitrge zur kenntniss des indirecten sehens. (i). untersuchungen ber den raumsinn der retina. *Archiv für Ophthalmologie*, 3:1–37.

Azzopardi, P., Jones, K. E., and Cowey, A. (1999). Uneven mapping of magnocellular and parvocellular projections from the lateral geniculate nucleus to the striate cortex in the macaque monkey. *Vision Research*, 39(13):2179–2189.

Baizer, J. S., Ungerleider, L. G., and Desimone, R. (1991). Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. *The Journal of Neuroscience*, 11(1):168–190.

Barlow, H. (1961). Possible principles underlying the transformations of sensory messages. In Rosenblith, W. A., editor, *Sensory Communication*, pages 217–234. MIT Press.

Blake, R. (2001). A primer on binocular rivalry, including current controversies. *Brain and Mind*, 2(1):5–38.

Brewer, A. A., Liu, J., Wade, A. R., and Wandell, B. A. (2005). Visual field maps and stimulus selectivity in human ventral occipital cortex. *Nature Neuroscience*, 8(8):1102–1109.

Burkhalter, A. and Van Essen, D. C. (1986). Processing of color, form and disparity information in visual areas vp and V2 of ventral extrastriate cortex in the macaque monkey. *The Journal of Neuroscience*, 6:2327–2351.

Cardin, V., Friston, K. J., and Zeki, S. (2011). Top-down modulations in the visual form pathway revealed with dynamic causal modeling. *Cerebral Cortex*, 21(3):550–562.

Carpenter, G. A. and Grossberg, S. (2011). *Adaptive Resonance Theory*. Springer.

Carrasco, M., McElree, B., Denisova, K., and Giordano, A. M. (2003). Speed of visual processing increases with eccentricity. *Nature Neuroscience*, 6(7):699–700.

Chen, J., He, Y., Zhu, Z., Zhou, T., Peng, Y., Zhang, X., and Fang, F. (2014a). Attention-dependent early cortical suppression contributes to crowding. *The Journal of Neuroscience*, 34(32):10465–10474.

Chen, M., Yan, Y., Gong, X., Gilbert, C. D., Liang, H., and Li, W. (2014b). Incremental integration of global contours through interplay between visual cortical areas. *Neuron*, 82(3):682–694.

Dayan, P. (1998). A hierarchical model of binocular rivalry. *Neural Computation*, 10(5):1119–1135.

Deubel, H. and Schneider, W. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, 36(12):1827–1837.

Durand, J.-B., Celebrini, S., and Trotter, Y. (2007). Neural bases of stereopsis across visual field of the alert macaque monkey. *Cerebral Cortex*, 17(6):1260–1273.

Fahrenfort, J. J., Scholte, H. S., and Lamme, V. A. (2007). Masking disrupts reentrant processing in human visual cortex. *Journal of Cognitive Neuroscience*, 19(9):1488–1497.

Fendick, M. and Westheimer, G. (1983). Effects of practice and the separation of test targets on foveal and peripheral stereoacuity. *Vision Research*, 23(2):145–150.

Fiorani, M., Gattass, R., Rosa, M., and Sousa, A. (1989). Visual area MT in the cebus monkey: location, visuotopic organization, and variability. *The Journal of Comparative Neurology*, 287(1):98–118.

Flom, M. C., Heath, G. G., and Takahashi, E. (1963). Contour interaction and visual resolution: Contralateral effects. *Science*, 142(3594):979–980.

Gattass, R., Sousa, A., and Gross, C. (1988). Visuotopic organization and extent of V3 and V4 of the macaque. *The Journal of Neuroscience*, 8(6):1831–1845.

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.

Griffis, J. C., Elkhetali, A. S., Burge, W. K., Chen, R. H., Bowman, A. D., Szaflarski, J. P., and Visscher, K. M. (2016). Retinotopic patterns of functional connectivity between V1 and large-scale brain networks during resting fixation. *NeuroImage*.

Harris, J. M., Nefs, H. T., and Grafton, C. E. (2008). Binocular vision and motion-in-depth. *Spatial Vision*, 21(6):531–547.

Hartmann, E., Lachenmayr, B., and Brettel, H. (1979). The peripheral critical flicker frequency. *Vision Research*, 19(9):1019–1023.

Higgins, K., Arditi, A., and Knoblauch, K. (1996). Detection and identification of mirror-image letter pairs in central and peripheral vision. *Vision Research*, 36:331–337.

Hoffman, J. (1998). Visual attention and eye movements. In Pashler, H., editor, *Attention*, pages 119–153. Psychology Press.

Hsieh, P.-J., Vul, E., and Kanwisher, N. (2010). Recognition alters the spatial pattern of fmri activation in early retinotopic cortex. *Journal of Neurophysiology*, 103(3):1501–1507.

Hubel, D. and Wiesel, T. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology*, 195(1):215–43.

Hung, C., Kreiman, G., Poggio, T., and DiCarlo, J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science*, 310(5749):863–866.

Kingdom, F. A. and Wang, D. (2015). Dichoptic colour-saturation masking is unmasked by binocular luminance contrast. *Vision Research*, 116:45–52.

Koenderink, J. J., Bouman, M. A., de Mesquita, A. E. B., and Slappendel, S. (1978). Perimetry of contrast detection thresholds of moving spatial sine wave patterns. iii. the target extent as a sensitivity controlling parameter. *JOSA*, 68(6):854–860.

Kok, P. and de Lange, F. P. (2014). Shape perception simultaneously up-and downregulates neural activity in the primary visual cortex. *Current Biology*, 24(13):1531–1535.

Levi, D. (2008). Crowding—an essential bottleneck for object recognition: a mini-review. *Vision Research*, 48:635–654.

Li, Z. (2002). A saliency map in primary visual cortex. *Trends in Cognitive Sciences*, 6(1):9–16.

Li, Z. and Atick, J. J. (1994). Efficient stereo coding in the multiscale representation. *Network: Computation in Neural Systems*, 5(2):157–174.

May, K., Zhaoping, L., and Hibbard, P. (2012). Perceived direction of motion determined by adaptation to static binocular images. *Current Biology*, 22:28–32.

May, K. A. and Zhaoping, L. (2016). Efficient coding theory predicts a tilt aftereffect from viewing untilted patterns. *Current Biology*, 26(12):1571–1576.

Movshon, J. A. and Newsome, W. T. (1996). Visual response properties of striate cortical neurons projecting to area mt in macaque monkeys. *Journal of Neuroscience*, 16(23):7733–7741.

Mullen, K. (1985). The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of Physiology*, 359(1):381–400.

Mullen, K. T. and Kingdom, F. A. (1996). Losses in peripheral colour sensitivity predicted from hit and miss post-receptoral cone connections. *Vision Research*, 36(13):1995–2000.

Müller, H. J., Reimann, B., and Krummenacher, J. (2003). Visual search for singleton feature targets across dimensions: Stimulus- and expectancy-driven effects in dimensional weighting. *Journal of Experimental Psychology: Human Perception and Performance*, 29(5):1021–1035.

Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., and Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences*, 99(23):15164–15169.

Nowak, L., James, A., and Bullier, J. (1997). Corticocortical connections between visual areas 17 and 18a of the rat studied in vitro: spatial and temporal organisation of functional synaptic responses. *Experimental Brain Research*, 117(2):219–241.

Ohzawa, I., DeAngelis, G., and Freeman, R. (1990). Stereoscopic depth discrimination in the visual cortex: neurons ideally suited as disparity detectors. *Science*, 249(4972):1037–1041.

Osterberg, G. (1935). Topography of the layer of rods and cones in the human retina. *Acta Ophthalmology*, 6, supplement 13:1–102.

Packer, O. and Williams, D. (2003). Light, and retinal image, and photoreceptors. In Shevell, S., editor, *The Science of Color*, pages 41–102. Optical Socieity of America, 2 edition.

Parker, A. J. (2007). Binocular depth perception and the cerebral cortex. *Nature Reviews Neuroscience*, 8(5):379–391.

Pascual-Leone, A. and Walsh, V. (2001). Fast backprojections from the motion to the primary visual area necessary for visual awareness. *Science*, 292(5516):510–512.

Pelli, D. G., Palomares, M., and Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of vision*, 4(12):article 12.

Pelli, D. G. and Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience*, 11(10):1129–1135.

Poggio, G. and Fischer, B. (1977). Binocular interaction and depth sensitivity in striate and prestriate cortex of behaving rhesus monkey. *Journal of Neurophysiology*, 40(6):1392–1405.

Qiu, C., Burton, P. C., Kersten, D., and Olman, C. A. (2016). Responses in early visual areas to contour integration are context dependent. *Journal of Vision*, 16(8):19–19.

Rönne, H. (1915). Zur theorie und technik der bjerrumschen gesichtsfelduntersuchung. *Arch Augenheilkd*, 78(4):284–301.

Rovamo, J. and Virsu, V. (1979). An estimation and application of the human cortical magnification factor. *Experimental Brain Research*, 37(3):495–510.

Schall, J. D., Morel, A., King, D. J., and Bullier, J. (1995). Topography of visual cortex connections with frontal eye field in macaque: convergence and segregation of processing streams. *The Journal of Neuroscience*, 15(6):4464–4487.

Sereno, M. I., Dale, A., Reppas, J., Kwong, K., Belliveau, J., Brady, T. J., Rosen, B. R., and Tootell, R. H. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, 268(5212):889–893.

Shadlen, M. and Carney, T. (1986). Mechanisms of human motion perception revealed by a new cyclopean illusion. *Science*, 232(4746):95–97.

Strasburger, H. (2014). Dancing letters and ticks that buzz around aimlessly: On the origin of crowding. *Perception*, 43(9):963–976.

Strasburger, H., Harvey, L. O., and Rentschler, I. (1991). Contrast thresholds for identification of numeric characters in direct and eccentric view. *Perception, & Psychophysics*, 49(6):495–508.

Strasburger, H. and Rentschler, I. (1996). Contrast-dependent dissociation of visual recognition and detection fields. *European Journal of Neuroscience*, 8(8):1787–1791.

Strasburger, H., Rentschler, I., and Harvey, L. O. (1994). Cortical magnification theory fails to predict visual recognition. *European Journal of Neuroscience*, 6(10):1583–1588.

Strasburger, H., Rentschler, I., and Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, 11(5):13–13.

Sziklai, G. (1956). Some studies in the speed of visual perception. *IRE Transactions on Information Theory*, 2(3):125–8.

Thorpe, S., Fize, D., and Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, 381(6582):520–522.

Traquair, H. (1938). *An Introduction to Clinical Perimetry*. London: Kimpton.

Tyler, C. W. and Likova, L. T. (2007). Crowding: a neuroanalytic approach. *Journal of vision*, 7(2):article 16.

Ungerleider, L. G. and Mishkin, M. (1982). Two cortical visual systems. In Ingle, D., Goodale, M. A., and Mansfield, R. W., editors, *Analysis of Visual Behavior*, pages 549–586. MIT Press, Cambridge, MA, USA.

van Essen, D. and Anderson, C. (1995). Information processing strategies and pathways in the primate visual system. In Zornetzer, S., Davis, J., Lau, C., and McKenna, T., editors, *An Introduction to Neural and Electronic Networks*, pages 45–76. Academic Press, Florida, USA, 2 edition.

Westheimer, G. (1982). The spatial grain of the perifoveal visual field. *Vision Research*, 22(1):157–162.

Weymouth, F. W. (1958). Visual sensory units and the minimal angle of resolution. *American journal of ophthalmology*, 46(1):102–113.

Wichmann, F. A. and Hill, N. J. (2001). The psychometric function: I. fitting, sampling, and goodness of fit. *Attention, Perception, & Psychophysics*, 63(8):1293–1313.

Wolfe, J., Cave, K., and Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 15:419–433.

Yu, H.-H., Chaplin, T., and Rosa, M. (2015). Representation of central and peripheral vision in the primate cerebral cortex: Insights from studies of the marmoset brain. *Neuroscience Research*, 93:47–61.

Zhaoping, L. (2011). A saliency map in cortex: Implications and inference from the representation of visual space. *Perception*, 40:ECVP Abstract Supplement, page 162. Presented at European Conference on Visual Perception, August, 2011, Toulouse, France.

Zhaoping, L. (2012). Gaze capture by eye-of-origin singletons: Interdependence with awareness. *Journal of Vision*, 12(2):article 17.

Zhaoping, L. (2013a). Dichoptic orientation stimuli show that ocular summation bests ocular opponency in central but not peripheral vision. *PsyCh Journal*, 2(Supplement 1) for the 9th Asian Pacific Conference on Vision (APCV2013):48.

Zhaoping, L. (2013b). Different perceptual decoding architectures for the central and peripheral vision revealed by dichoptic motion stimuli. *Perception*, 42:ECVP Abstract Supplement, page 21.

Zhaoping, L. (2014). *Understanding Vision: Theory, Models, and Data*. Oxford University Press.

Zhaoping, L. (2015). Dichoptic color gratings reveal a perceptual bias for binocular summation over binocular difference, which is stronger in central than peripheral vision. *Perception*, 44(S1) for ECVP 2015:286–287.