What are the functions of primary visual cortex (V1)?

Li Zhaoping
University of Tübingen, Max Planck Institute for Biological Cybernetics, Tübingen, Germany
email: li.zhaoping@tuebingen.mpg.de

Abstract:

Although Hubel and Wiesel established decades ago how individual V1 neurons transform retinal inputs, functions of V1 as a whole are being discovered only recently. First, V1 acts as a motor cortex for exogenously guiding saccades by constructing a bottom-up saliency map of the visual field. Second, V1 initiates a processing bottleneck: a massive reduction of visual information begins at its output to downstream areas. Third, downstream recognition is limited by impoverished information, V1 supports ongoing recognition by providing additional information queried by top-down feedback from downstream areas, directed predominantly to central visual field representations. These V1 functions underpin a framework in which vision is mainly looking and seeing through the bottleneck. Looking selects a fraction of visual information into the bottleneck, largely by saccades that center selected contents at gaze. Seeing recognizes the selected contents. Looking and seeing rely mainly on processing in the peripheral and central visual fields.

Highlights

- Primate V1 serves vision's main function of looking and seeing through a bottleneck.
- It creates a bottom-up saliency map to exogenously guide looking via gaze shifts.
- It initiates an information bottleneck at V1's output to downstream brain areas
- It supports feedback queries mainly directed to the central visual field
- The queried information aids ongoing visual recognition in light of the bottleneck

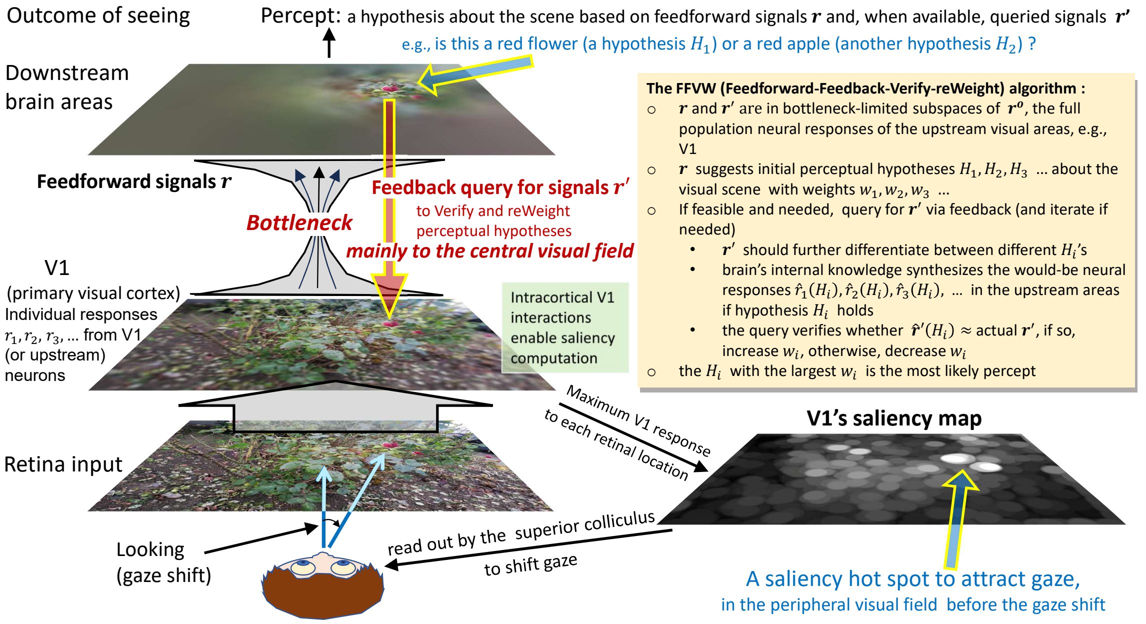
Keywords

primary visual cortex, gaze shifts, saliency map, bottleneck, top-down feedback, selection, recognition, the looking-and-seeing framework.

Introduction

V1 is not alone in that we remain ignorant about its function decades after we have known its neural properties[1, 2, 3]. Before patient HM, the hippocampus's critical role in memory remained unknown long after its anatomy was characterized. V1 is clearly essential for vision, but so is the retina. What, then, is V1's special function?

The emerging answer is tied to the question of what vision is. Our brain faces an information processing bottleneck, constrained by metabolic energy [4], space for neurons and wiring, and time for (e.g.,) complex sensory inference [5]. Often termed the attentional bottleneck, it allows only a tiny fraction of visual inputs, largely near the center of gaze, to reach recognition. Vision is therefore not only seeing, but also looking [5]: looking selects the fraction allowed to enter the bottleneck (cf. attentional selection), typically by gaze shifts that center the selected inputs on



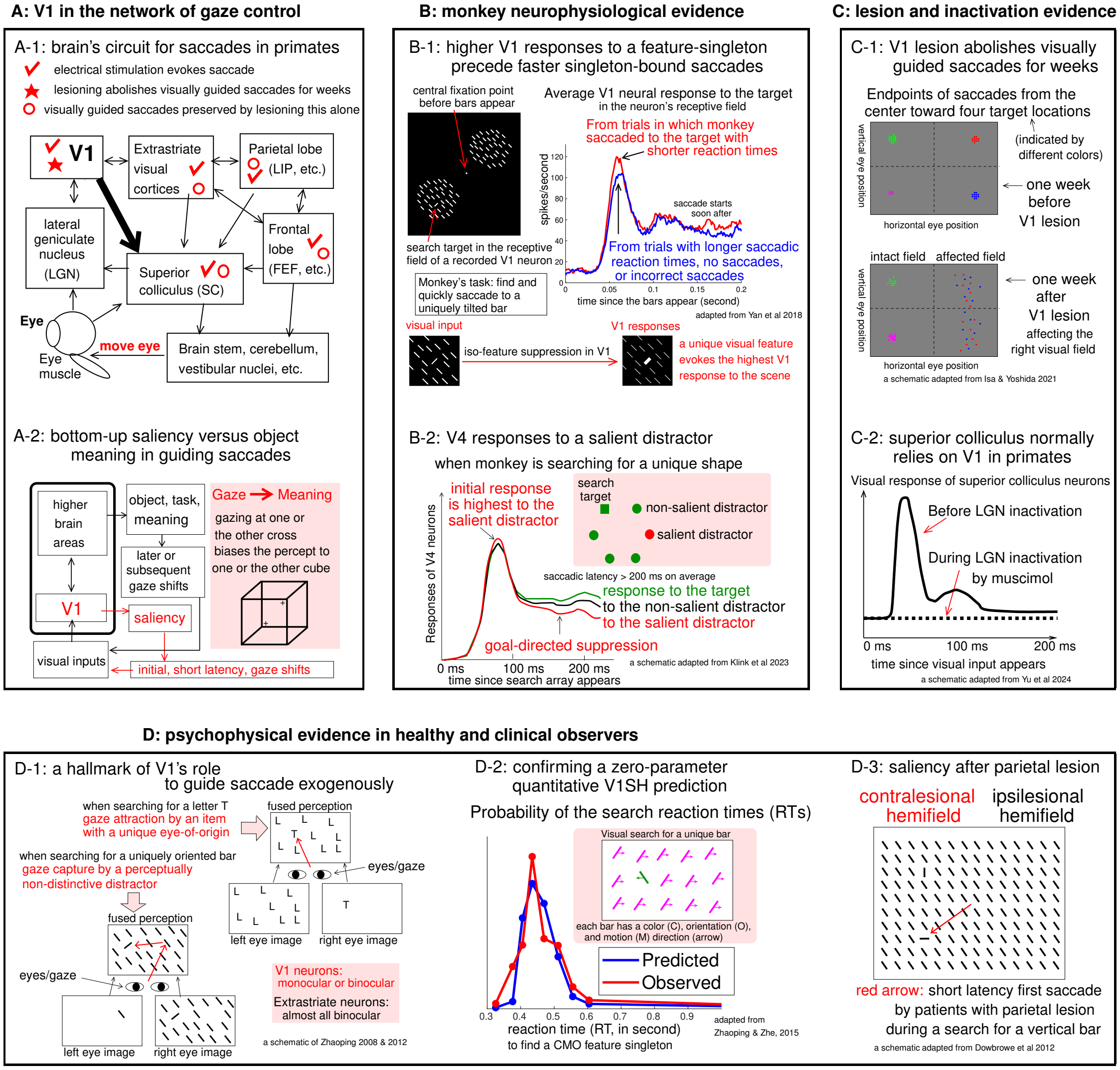
The FFVW (Feedforward-Feedback-Verify-reWeight) algorithm :

- $\boldsymbol{r}$ and $\boldsymbol{r}'$ are in bottleneck-limited subspaces of $\boldsymbol{r}^{\boldsymbol{o}}$, the full population neural responses of the upstream visual areas, e.g., V1
- $\boldsymbol{r}$ suggests initial perceptual hypotheses H_1, H_2, H_3 ... about the visual scene with weights w_1, w_2, w_3 ...
- If feasible and needed, query for $\boldsymbol{r}'$ via feedback (and iterate if needed)
  - $\boldsymbol{r}'$ should further differentiate between different H_i's
  - brain's internal knowledge synthesizes the would-be neural responses $\hat{r}_1(H_i), \hat{r}_2(H_i), \hat{r}_3(H_i)$, ... in the upstream areas if hypothesis H_i holds
  - the query verifies whether $\hat{\boldsymbol{r}}'(H_i) \approx$ actual $\boldsymbol{r}'$, if so, increase w_i, otherwise, decrease w_i
- the H_i with the largest w_i is the most likely percept

Figure 1: Functional roles of the primary visual cortex V1. V1 first creates a bottom-up saliency map that exogenously guides gaze shifts (looking), selecting a fraction of visual input for the brain's processing bottleneck and centering it at the point of gaze. V1 also initiates this bottleneck, limiting downstream information and thereby making recognition (seeing) nontrivial. To support ongoing recognition, V1 supplies additional information when queried by feedback from downstream areas along the visual pathway. This feedback predominantly targets the central visual field, which is specialized for seeing. In contrast, the peripheral visual field is specialized for looking, determining where to shift gaze.

the fovea; and seeing (recognizing) decodes and discriminates this fraction. This looking-and-seeing framework has developed alongside investigations of V1's function, beginning with the V1 Saliency Hypothesis (V1SH) in the late 1990s [6, 7], and progressing to the central-peripheral dichotomy (CPD) theory from the 2010s [8, 9]. This framework is supported by empirical findings, particularly the most recent ones and various experimental confirmations of falsifiable predictions.

First, V1 serves looking by guiding gaze shifts exogenously. According to V1SH, it creates a map of salience, defined as the strength of a visual location to attract attention by external visual inputs (Figure 1). This map is read out by the superior colliculus (SC), which specifies saccades to salient locations. Second, to reduce cost, the bottleneck begins immediately at V1's output to downstream stages. Third, to aid ongoing recognition (seeing) from the limited information admitted into the bottleneck, feedback from downstream stages queries upstream stages such as V1 for additional relevant information; according to CPD, this feedback is largely restricted to the central visual field to reduce cost [9]. Accordingly, saccades to any scene location enable a query to that location. Apparently, evolutionary optimization had led to the CPD in this looking-and-seeing framework to balance computation and costs.



adapted from Yan et al 2018

a schematic adapted from Klink et al 2023

a schematic adapted from Isa & Yoshida 2021

a schematic adapted from Yu et al 2024

a schematic of Zhaoping 2008 & 2012

adapted from Zhaoping & Zhe, 2015

a schematic adapted from Dowbrowe et al 2012

Figure 2: V1 exogenously guides saccades. A: V1 in a network of brain areas for guiding gaze and attention. A-1: in primates, electric stimulation of retinotopic locations in V1, extrastriate cortices, parietal cortex, superior colliculus (SC), or frontal eye field (FEF) evokes saccades, but only V1 lesions abolish visually guided saccade for weeks. A-2: a view on how bottom-up saliency (computed by V1) and object meaning jointly guide saccades. B: neurophysiological evidence. B-1: during a monkey's search for an orientation singleton bar, faster saccades to the target are typically preceded by higher initial responses of V1 neurons to the target. Iso-feature suppression makes nearby V1 cells tuned to similar features suppress each other. In this example, iso-orientation suppression makes the V1 neural response to the target typically the strongest response to this image. B-2: in a visual search for a unique shape (randomly a square among disks or vice versa), V4 neural response to a salient distractor (randomly, red among green or vice versa) is initially significantly higher (somewhat exaggerated schematically for better visualization) than the responses to the target and other nontargets, before it is suppressed [10]. C: lesion evidence. C-1: after monkeys had learned to saccade to a light spot in any of the four corners of a display for rewards, V1 lesions abolished visually guided — but spared memory-guided — saccades in the affected hemifield[11]. C-2: in monkeys, muscimol inactivation of LGN abolishes visual responses in SC neurons[12]. D: behavioral evidence. D-1: V1 is the only cortical area to represent eye-of-origin accurately. Uniqueness in eye-of-origin is perceptually invisible but strongly attracts gaze [13, 14]. D-2: with no free parameters, V1SH accurately predicts a distribution of reaction times (RTs) to find a bar unique in color (C), orientation (O), and motion direction (M) from RTs to singleton bars unique in one or two of these three feature dimensions [15]. D-3: patients with parietal lesions retain saliency effects in the affected hemifield in their short latency first saccades, which are more likely directed to a salient distractor in a visual search [16].

V1 as a motor cortex to guide saccades exogenously for visual input selection

By CPD, V1's role in looking is by its representation of the peripheral visual field, which we focus on in this section. Like the conventional motor cortices, which project to muscles indirectly via spinal cord and brain stem, V1 projects indirectly to the eye movement muscles through SC, which is part of the brain stem. In primates, electrical stimulation of V1, SC, frontal eye fields (FEF), extrastriate cortical areas (e.g.,V2 and V4), and parietal areas (e.g., LIP) can elicit eye movements (Fig. 2A)[5]. V1 is unique, however, in that lesioning it alone abolishes visually guided saccades for weeks[11]. The fastest set of output axons from visual cortex are those projecting to SC [17, 18], and V1 contributes most of these axons [19].

According to V1SH [7], the saliency map is represented by the highest V1 response to each visual location (without decoding scene properties). Intracortical V1 interactions cause nearby neurons tuned to similar features to suppress each other [5]. For example, when an image contains a uniquely oriented bar among uniformly oriented background bars, responses to the background bars suffer from iso-orientation suppression [20], whereas the response to the orientation singleton escapes this suppression. When monkeys searched for such a singleton (Fig. 2B-1), trial-by-trial fluctuations in the initial neural responses to the singleton (at a latency $l = 40$–60 ms after visual input onset) were inversely correlated with saccade latencies to the singleton [21]. Hence, these initial responses are saliency signals (read out by SC, Figure 1). Also, electrical or optogenetic activations of monkey V1 evoke saccades towards the retinotopically corresponding locations[22, 23]. In monkey V4, saliency signals likely from V1 emerge at $l \approx 60$ ms in layers not receiving feedback[24]. In FEF and parietal cortex, responses do not distinguish salient from non-salient items until $l \geq 100$ ms [25, 26].

Direct retinal inputs to SC appear insufficient for SC visual responses or for guiding saccades in normal monkeys. Muscimol inactivation of lateral geniculate nucleus (LGN), which relays retinal inputs to V1, abolishes any visually evoked SC responses[12] (Fig. 2C-2). Equally, in a study [11], V1 was lesioned unilaterally after monkeys had learned to earn reward by making a saccade to a light spot presented randomly at one of the four corners of a display (Fig. 2C-1). Afterwards, when the light spot was at either corner of the affected hemifield, saccades were aimed randomly into this hemifield, apparently memory guided (mediated by FEF) to bring the target into the intact visual field, or to earn reward by chance. Blindsight after V1 lesion emerges only after months or years of neural reorganization [11]. Note that another study [27] reported that saliency signals in monkeys emerged in SC ($l \approx 65$ ms) before they did in V1 ($l \approx 139$ ms). Replication studies should reconcile discrepancies. Meanwhile, SC in lower vertebrates does compute saliency[28, 29].

A hallmark of V1's role is the gaze capture by a singleton unique in the eye of origin of visual inputs [13, 14]. For example (Fig. 2D-1), the reaction time (RT) to find a unique letter 'T' among letter 'L's normally increases with the number N of 'L's; however, if the 'T' is shown to one eye and the 'L's are shown to the other eye, the 'T' pops out by a short RT independent of N [13]. V1 is unique among cortical areas in having a substantial percentage of monocular neurons (in the binocular visual field), encoding the eye-of-origin. Normal observers cannot discriminate eye-of-origin, suggesting that the eye-of-origin information is lost beyond V1 (Fig. 3A). Moreover, pop-out by eye-of-origin can dominate pop-out by orientation in a search for an orientation singleton bar. When a non-target bar is an eye-of-origin singleton (Fig. 2D-1), it tends to capture the first gaze shift during search when both singletons are sufficiently eccentric [13, 14]. This gaze distraction exemplifies looking without seeing (the eye-of-origin) [14].

Without free parameters, V1SH quantitatively predicted a distribution of the RTs to find a feature singleton unique in color (C), orientation (O), and motion direction (M) from distributions of the RTs for finding a singleton unique in only one or two of these feature dimension (Fig. 2D-2). The prediction was confirmed [15].

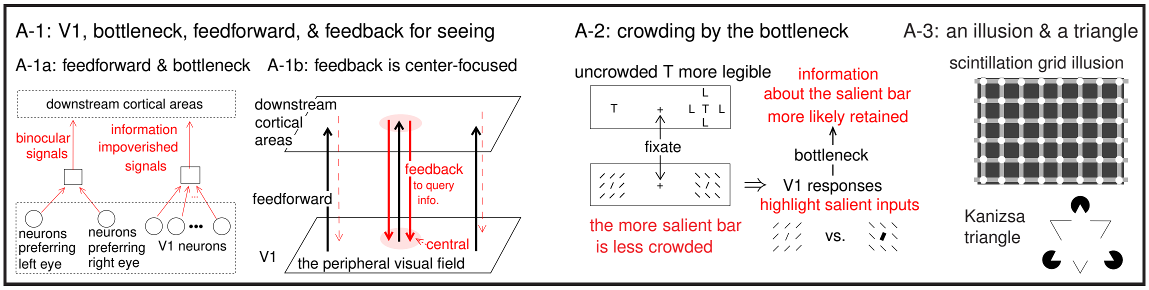
Although endogenous control can readily suppress salient distractors when their features or locations are predictable [30, 31]), suppression is delayed for unpredictable distractors. For example (Fig. 2B-2), when monkeys searched for a shape-defined target while a color singleton distractor was unpredictable in color, shape, and location [10], V4 neural responses to the distractor were initially higher than responses to the other items before they were suppressed at $l \gtrsim 100$ ms. These responses likely reflect early feedforward inputs from V1 followed by top-down suppression from, e.g., FEF and parietal cortex[25, 32, 33, 34]. Long-latency target enhancement and distractor suppression are also evident in magnetoencephalography imaging of human V1 responses[35].

Patients with parietal lesions have impaired attention [33]. However, during their search for a vertical bar among uniformly tilted oblique bars and a distracting horizontal bar (Fig. 2D-3), the distractor tends to capture their short-latency first saccade even when both the target and distractor are in the affected hemifield[16]. Therefore, parietal cortex does not control bottom-up saliency, whose effects are typically short-lived [36]. Indeed, data from high-field functional magnetic resonance imaging (fMRI) suggest that saliency signals in parietal cortex originate from V1 [37].

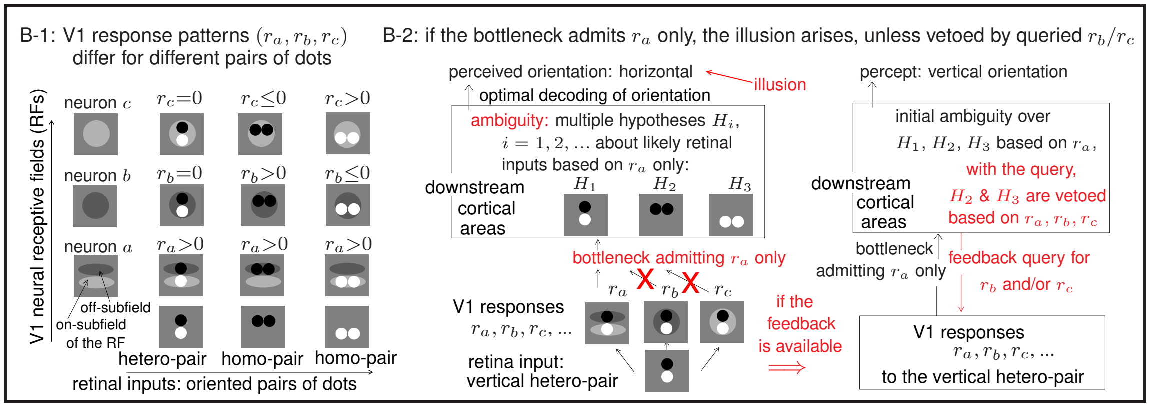
Subsequently, object meaning often predicts gaze shifts better than saliency [38], consistent with the short-lived nature of saliency effects [36]. After initial saliency-driven saccades extract scene structure, subsequent gaze shifts can be endogenously influenced by meanings. Thus, patients with visual object agnosia show normal early, but altered later, gaze patterns [39]. However, initial fixations set initial conditions for subsequent explorations, and gaze position can shape perceived meaning, as in viewing an ambiguous Necker cube in Fig. 2A-2 [40, 41]. Whether brief saliency exerts a lasting "butterfly effect" on subsequent explorations remains unclear.

Visual recognition through a bottleneck via feedforward from V1 and feedback to V1

A: V1's role in seeing according to the CPD theory: crowding, illusions, & perceptual completion



B: deriving the predictions of the flip tilt illusion and its veto by the feedback to query for more information



C: CPD predicted illusions in the presence or absence of the top-down feedback query

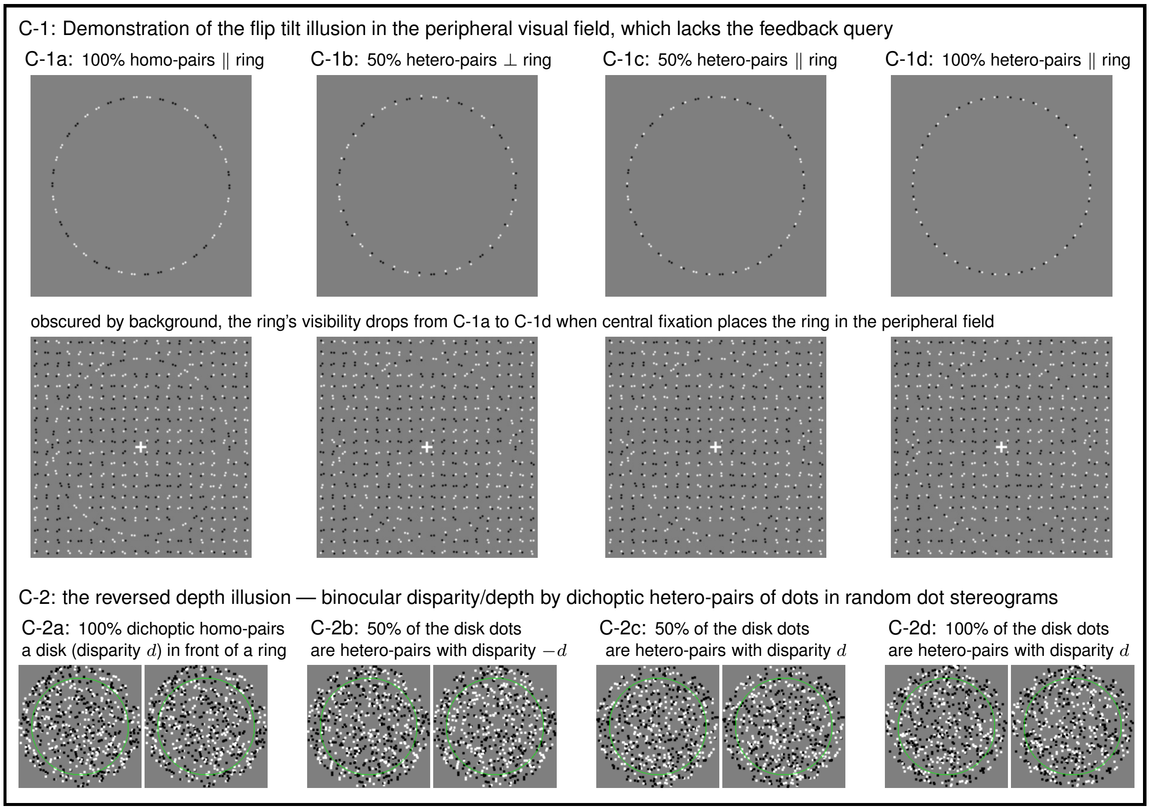


Figure 3: , caption in the next page

Figure 3: Caption of figure 3 in the preceding page: V1 works with higher brain areas to realize recognition under a processing bottleneck. A: overview of the CPD theory and its account of some perceptual phenomena. A-1: the bottleneck starts at V1's output to downstream areas, discarding information such as eye-of-origin. Feedback, directed mainly to the central field representation, queries for additional information to support ongoing recognition. A-2: visual crowding arises from a lack of the feedback query in the peripheral visual field to aid recognition using impoverished information through the bottleneck. Salient items evoke stronger V1 responses and are less susceptible to information loss [42]. A-3: the lack of the query also makes the peripheral field vulnerable to illusions as errors in recognition, e.g., the illusory scintillation is visible only outside the central gaze. The white Kanizsa triangle is more vivid in the central than the peripheral field: the feedback query aids perceptual completion by analysis-by-synthesis. B: the flip tilt illusion is predicted in the peripheral field, if the bottleneck blocks the V1 signals, e.g., neural responses r_b and r_c, necessary to disambiguate a hetero-pair of dots from homo-pairs of dots orthogonal to the hetero-pair. This illusion makes a hetero-pair of dots appear tilted orthogonal to its actual tilt (of the axis of alignment between the two dots). This illusion can be vetoed by feedback to V1 querying for the relevant information. C: demonstration of CPD predicted flip tilt illusion and the reversed depth illusion. C-1: The hetero-pairs on each ring are in the peripheral visual field when gaze is at the center of the ring, and should appear — by the illusion — tilted orthogonal to their true orientation. This makes the ring's visibility in noise drop from C-1b to C-1c to C-1d. All the homo-pairs on the rings are parallel to the tangent of the ring. C-2: C-2a to C-2d are stereograms (a left-eye image and a right-eye image each) analogous to C-1a to C-1d, homo- and hetero-pairs are dichoptic pairs of dots (one dot each eye) to signal depth, veridically and illusorily (as reversed-depth), by V1 neurons tuned to binocular disparity. Each stereogram contains a disk (made of homo- and/or hetero-pairs, all the homo-pairs have disparity d) and a surrounding ring (made of zero-disparity homo-pairs). Free-fuse each stereogram in the central visual field to view. C-2a reveals a disk (disparity d) veridically in front of the ring. (The zero-disparity green circles mark the inner borders of the rings for illustration only.) In C-2d, the disk's depth is indiscernible, since the feedback query vetoes the illusory signals from the hetero-pairs. The disk appears in front of the ring more clearly in C-2b than in C-2c, demonstrating that in C-2b, our brain constructively uses V1's illusory depth signals (from the hetero-pairs) that agree with the veridical and verifiable depth signals from the homo-pairs [43].

V1's role in looking motivates the idea [9] that the information bottleneck starts immediately at V1's output to downstream stages [9] (Figure 3A). For example, a binocular downstream neuron preferring vertical orientation could pool inputs from two monocular V1 neurons of different eyes, discarding eye-of-origin information. More generally, downstream combination of V1 signals can lose substantial details; when mainly summary statistics of visual inputs are preserved, phenomena like visual crowding emerge [44, 45, 14].

By the Central-Peripheral Dichotomy (CPD) theory [9], peripheral and central visual fields are mainly for looking and seeing [5, 46], respectively. For seeing in light of the bottleneck, the CPD theory proposes that downstream stages query for additional relevant information for the ongoing processing by sending feedback to upstream areas like V1. This feedback focuses on the central field representation.

Lacking this feedback, the peripheral field suffers from visual crowding [9, 14], see Figure 3A-2. Thus a peripheral bar that is surrounded by other bars is more legible when it has a larger orientation contrast from the other bars and so is salient. Such a salient bar evokes a higher V1 response by escaping iso-orientation suppression, making its information better preserved through the bottleneck. Indeed, salient inputs are more faithfully represented in V4 responses [42].

Besides crowding, information impoverishment can lead to visual illusions. For example, in Figure 3A-3, illusory scintillations are only visible in white disks outside the central gaze; they are vetoed in the central visual field by the feedback query.

CPD predicts the flip tilt illusion in the peripheral visual field (Figure 3B). Three different pairs of dots — a vertical hetero-pair of dots (one black, one white), a horizontal homo-pair of black dots, or a horizontal homo-pair of white dots — evoke three different patterns of responses (r_a, r_b, r_c) from V1 neurons a, b, and c. Neuron a prefers horizontal orientations, but is excited by all three pairs, including the vertical hetero-pair since this pair's black and white dots fall into this neuron's off- and on-subfields.

During recognition, downstream areas entertain hypotheses H_1, H_2, and H_3 that the retinal input is the hetero-pair, the black homo-pair, or the white homo-pair, based on V1 responses admitted into the bottleneck. If the vertical hetero-pair is shown but only r_a is admitted, perception is ambiguous: all three hypotheses can produce $r_a > 0$ (Figure 3B-2). Perceived orientation is likewise ambiguous, with H_2 and H_3 voting horizontal and H_1 vertical, so a forced-choice report is likely horizontal from the majority voting. Thus, when the bottleneck admits only r_a, a retinal hetero-pair appears tilted orthogonally to the actual tilt — the flip tilt illusion. Formally, substantial likelihood $p(r_a|H_i)$ of r_a given H_i and prior probability $p(H_i) > 0$ for H_i yield comparable posterior $p(H_i|r_a)$ between i's, and orientation inference follows $p(\text{orientation}|r_a) = \sum_i p(\text{orientation}|H_i)p(H_i|r_a)$.

If downstream areas send feedback to V1 to query for r_b and/or r_c, responses $r_b = 0$ and $r_c = 0$ to the hetero-pair veto hypotheses H_2 and H_3, leaving only H_1. This removes the ambiguity and vetoes the illusion. Such a query requires brain's internal knowledge ($p(r_b, r_c|H_i)$) of expected (r_b, r_c) for each H_i, and thus the knowledge that (r_b, r_c) can discriminate between the hypotheses.

The Feedforward-Feedback-Verify-reWeight (FFVW) algorithm has been proposed to perform recognition [9] (Figure 1). The bottleneck first admits feedforward responses $\mathbf{r}$, generating hypotheses H_i's with weights w_i (which should grow with $p(H_i|\mathbf{r})$). Then, feedback queries for additional responses $\mathbf{r}'$. In our example, $\mathbf{r} = r_a$ proposes the hypotheses H_i's, and $\mathbf{r}' = (r_b, r_c)$ disambiguate them. Using learned expectations of would-be responses $\hat{\mathbf{r}}'(H_i)$ for each H_i, the brain compares the synthesized $\hat{\mathbf{r}}'(H_i)$ with the actual $\mathbf{r}'$ to reweight each hypothesis: w_i increases if $\hat{\mathbf{r}}'(H_i) \approx \mathbf{r}'$ and decreases otherwise. Here $(r_b, r_c) = (0, 0)$ boosts w_1 and suppresses w_2 and w_3. This FFVW algorithm thus implements analysis-by-synthesis.

Figure 3C-1 demonstrates the flip tilt illusion. In noise, the C-1d ring is the least conspicuous in our peripheral field. This is because it only contains hetero-pairs which appear illusorily orthogonal to the tangent. This illusion also makes the C-1b ring more conspicuous than the C-1c ring in the peripheral field. However, when gaze traces the circumferences of the rings, the feedback query in the central visual field vetoes the illusion, making the C-1b ring appear less smooth than the C-1c ring.

The effects of central-field feedback are also apparent in random dot stereograms in Figure 3C-2. Dichoptic homo-pairs arise from physically plausible three-dimensional (3D) dots. However, a dichoptic hetero-pair lacks a realistic 3D interpretation (since a black dot in one eye is paired with a white dot in the other). It excites a V1 neuron preferring a binocular disparity opposite in sign to the true interocular disparity [47], analogous to the excitation of horizontal-preferring neurons by vertical hetero-pairs in Figure 3B. This results in the reversed depth illusion [48], an analog of the flip tilt illusion.

In Figure 3C-2, C-2a is a standard random dot stereogram depicting a disk of binocular disparity d in front of a zero-disparity surrounding ring. In C-2d, all the disk dots are hetero-pairs of disparity d. Their V1 responses signal a disk behind the ring. However, free fusion of C-2d in the central visual field yields no clear depth ordering, because the feedback query (for, e.g., the monocular signals in V1) vetoes the misleading V1 signals. (V4 neurons are indeed less sensitive to the reversed depth signals [49].) In the peripheral visual field, which lacks the feedback query, the C-2d disk appears illusorily behind the ring as predicted [48]. As predicted [47], in the central visual field, interrupting the feedback by backward masking renders the illusion visible (and similarly makes the flip tilt illusion visible).

In Figure 3C-2bc, the hetero-pairs make up 50% of the disk dots. Their disparity is opposite to (in C-2b) or the same as (in C-2c) the disparity of the homo-pairs. Hence, their evoked V1 responses reinforce or oppose the veridical depth signals from the homo-pairs. Consequently, in the central visual field, the disk appears more clearly in front in C-2b than in C-2c. Hence, our brain constructively exploits illusory or ambiguous feedforward signals which align with verifiable veridical cues [43]. This likely explains why the white Kanizsa triangle[50] in Figure 3A-3 appears more vivid in the central than the peripheral visual field [51]: this triangle is (part of) a perceptual hypothesis suggested by fragmentary retinal inputs, the feedback query confirms it by verified occlusions of the black disks and another triangle.

Retrograde tracing in marmoset V1 shows that feedback axons are denser, by an order of magnitude, in V1's representation of the central, than the peripheral, visual field [52], confirming the critical CPD prediction. Consistent with this, frontal areas exhibit stronger functional connectivity with central than peripheral V1 [53], and feedback from V4 to V1 inferred from Granger causality analysis of neural activities is concentrated in foveal V1[54]. Equally, CPD predicts that indications of top-down feedback under ambiguous or difficult perceptual conditions [55, 56, 57] should decrease with increasing visual field eccentricity.

By CPD, a visual percept or phenomenon that is stronger in the central versus peripheral visual field is associated with a rich versus absent feedback query [9]. Accordingly, surface percepts arising from surface completion, 3D shape from 2D geometry, surface transparency, or closure — which are stronger in the central than the peripheral visual field [51] — are predicted to rely on feedback queries. Crowding is predicted to appear in the central visual field if the feedback is disrupted (e.g., by backward masking like in [47]). Likewise, foveal feedback is predicted to be weaker for amblyopic individuals (since they exhibit foveal crowding) [58, 59].

Whereas feedback queries adapt to ongoing recognition, I conjecture that the feedforward connections driving the initial sweep, particular from V1, are less flexible, especially in the peripheral visual field. These connections apparently favor letting salient inputs through the bottleneck. They are likely shaped during development and evolution to optimize initial recognition across scenes, and are likely modifiable to a limited extent in adulthood by perceptual learning[60] or by task and context[61].

Discussion and outlook

V1's functions are tied to the looking-and-seeing framework [5], in which percepts are integrated across fixations via brisk saccades [62, 63, 64] while the visual world remains relatively stable [65]. Because saccades bring peripheral objects to the fovea, feedback queries can link visual signals $\mathbf{r}$ at a peripheral location to the foveally queried signals $\mathbf{r}'$ at the next moment through ongoing perceptual hypotheses H_i's about the scene. This perspective aligns with evidence of feedback to fovea V1 from peripheral stimuli prior to saccades and of impaired peripheral recognition by foveal masking [66, 67, 68, 69]. Likewise, fixational eye movements can be understood as part of the feedback query process rather than random gaze fluctuations [70].

The perceptual hypotheses $H_i = \{\theta_1, \theta_2, ...\}$ are parameterized by $\theta_1, \theta_2...$ that specify scene properties, e.g., the shape and color of objects in H_i. One parameter θ_j specifies object viewpoints. In trans-saccadic recognition, θ_j should be updated to generate the would-be visual signals $\hat{\mathbf{r}}'$ after the saccade (including for relevant or salient objects that are not the saccadic target), using an efference copy of the saccadic command [71]. This aligns the expected $\hat{\mathbf{r}}'$ with the actual $\mathbf{r}'$ after the saccade, supporting perceptual stability. This perspective is related to works on neural receptive field remapping across saccades (e.g., [72, 73]). The bottleneck imposes a massive information reduction from retinal inputs to percepts (H_i's), allowing perceptual information accumulated across fixations to be maintained in capacity-limited working memory [74] before it is lost or consolidated into longer-term memory. Precise eye-tracking and free viewing behavior [75] are often needed to better understand natural vision, in which visual inputs are largely selected overly rather than covertly.

Some falsifiable theoretical predictions (Table 1) remain to be tested. Our framework raises further questions, including how the FFVW algorithm is implemented neurally; whether and how this algorithm and brain's internal states influence downstream activities (e.g., in V1); how exogenous and endogenous control of gaze and attention are prioritized and integrated in the brain; how some of the information loss through the bottleneck helps achieve object invariance in recognition; and what additional roles V1 plays in tasks such as navigation, balance, avoidance, and escape — behaviors that extend beyond looking-and-seeing since they require little or no detailed recognition [76, 77, 14]. Addressing these questions connects our theories with other lines of research and can generate further falsifiable predictions. Furthermore, understanding primate vision should inform multisensory processing across species — from primate V1 to the superior colliculus (optic tectum) in lower vertebrates [28] and from looking-and-seeing to multisensory orienting-and-recognition [78] — as all animals face processing bottlenecks.

Table 1: Some falsifiable predictions of the theories concerning primate V1's functional roles

theory of V1's functions	**theoretical prediction**	**experimental status**
V1 Saliency Hypothesis (V1SH)	ocular singleton pops out to attract gaze and attention	confirmed[13], Fig. 2D-1
	Higher V1 responses to a search target tend to precede a shorter latency saccade to the target	confirmed [21] Fig. 2B-1
	zero-parameter quantitative distribution of RTs for the feature singleton search task described in[15]	confirmed [15] Fig. 2D-2
	a collection of other predictions	confirmed and reviewed in [5].
Central-Peripheral Dichotomy (CPD) theory	the flip tilt illusion and the reversed depth illusion in the peripheral visual field	confirmed[48, 14], Fig. 3C
	the reversed depth illusion becomes visible in the central visual field by backward masking	confirmed [47]
	feedback in the ventral visual pathway (for recognition) to V1 is mainly directed to the representation of the central visual field	confirmed [52, 54]
	cortical neurons (e.g., V2, V4, perhaps also V1) are more sensitive to the reversed feature signals by hetero-pairs (in, e.g., orientation, depth) at shorter latencies after stimulus onset before the feedback query becomes effective, or at a more peripheral visual location where the feedback query is weaker	to be tested
	in upstream areas such as V1 and V2, the percentage of simple cells is higher nearer to the foveal representation, since simple cell responses have more sensory details for the feedback query	to be tested
	feedback to V1 along the ventral pathway is more likely directed to the binocular rather than monocular cells [9]	to be tested
	crowding appears in the central visual field when the feedback query is impaired, e.g., by backward masking	to be tested
	amblyopic eyes, which exhibit crowding in the central visual field, are associated with weaker or absent feedback query	to be tested
	perceptual phenomena, such as perceptual organization for surface completion [51], that are stronger in the central than peripheral visual field regardless of input spatial resolution should involve the feedback query, and can be weakened by impairing the feedback (e.g., by backward masking).	to be tested
	perceptual phenomena, such as crowding and some illusions, that are stronger in the peripheral than the central visual field regardless of input spatial resolution are associated with a weaker or a lack of the top-down feedback query.	to be tested

Acknowledgement

This work is supported in part by University of Tübingen and the Max Planck Society. I thank Peter Dayan for helpful comments to the manuscript.

References

[1] Hubel DH, Wiesel TN: **Receptive fields, binocular interaction and functional architecture in the cat's visual cortex**. *The Journal of physiology* 1962, **160**:106–154.

[2] Carandini M, Demb JB, Mante V, Tolhurst DJ, Dan Y, Olshausen BA, Gallant JL, Rust NC: **Do we know what the early visual system does?** *Journal of Neuroscience* 2005, **25**:10577–10597.

[3] Olshausen BA, Field DJ: **What is the other 85 percent of v1 doing**. *L van Hemmen, & T Sejnowski (Eds)* 2006, **23**:182–211.

[4] Attwell D, Laughlin SB: **An energy budget for signaling in the grey matter of the brain**. *Journal of Cerebral Blood Flow & Metabolism* 2001, **21**:1133–1145.

[5] Zhaoping L: **Understanding vision: theory, models, and data**. *Oxford University Press* 2014, .

[6] Li Z: **Contextual influences in V1 as a basis for pop out and asymmetry in visual search**. *Proceedings of the National Academy of Sciences of the USA* 1999, **96**:10530–10535.

[7] Li Z: **A saliency map in primary visual cortex**. *Trends in Cognitive Sciences* 2002, **6**:9–16.

[8] Zhaoping L: **Feedback from higher to lower visual areas for visual recognition may be weaker in the periphery: Glimpses from the perception of brief dichoptic stimuli**. *Vision Research* 2017, **136**:32–49.

[9] Zhaoping L: **A new framework for understanding vision from the perspective of the primary visual cortex**. *Current Opinion in Neurobiology* 2019, **58**:1–10.

[10] Klink PC, Teeuwen RR, Lorteije JA, Roelfsema PR: **Inversion of pop-out for a distracting feature dimension in monkey visual cortex**. *Proceedings of the National Academy of Sciences* 2023, **120**:e2210839120.

[11] Isa T, Yoshida M: **Neural mechanism of blindsight in a macaque model**. *Neuroscience* 2021, **469**:138–161.

• [12] Yu G, Katz L, Quaia C, Messinger A, Krauzlis R: **Face-related activity in superior colliculus and temporal cortex of primates.** *Neuron* 2024, **112**:2814–22.

> Showed that visually responsive neurons in the superior colliculus stop responding to visual inputs when LGN neurons are inactivated by muscimol, regardless of visual input preferences of the neurons. Since SC receives inputs from retina and V1 but not from LGN, and since LGN relays retinal inputs to V1, this finding suggests that, without V1's input to SC, direct retinal inputs to SC is insufficient to drive visually guided saccade mediated by SC. This supports V1 rather than SC as the source of the saliency map for guiding saccades exogenously.

[13] Zhaoping L: **Attention capture by eye of origin singletons even without awareness—a hallmark of a bottom-up saliency map in the primary visual cortex**. *Journal of Vision* 2008, **8**:article 1.

•• [14] Zhaoping L: **Peripheral vision is mainly for looking rather than seeing**. *Neuroscience Research* 2024, **201**:18–26.

> An understanding of a wide array of phenomena in the peripheral visual field, e.g., crowding, illusions, uncanny ability to guide gaze, by the CPD proposed function of vision in the peripheral field to select destinations for gaze shift or select inputs into a processing bottleneck, rather than to recognize objects (the CPD proposed function of vision in the central visual field).

[15] Zhaoping L, Zhe L: **Primary visual cortex as a saliency map: A parameter-free prediction and its test by behavioral data**. *PLoS Comput Biol* 2015, **11**:e1004375.

[16] Dombrowe I, Donk M, Wright H, Olivers CN, Humphreys GW: **The contribution of stimulus-driven and goal-driven mechanisms to feature-based selection in patients with spatial attention deficits**. *Cognitive neuropsychology* 2012, **29**:249–274.

[17] Finlay B, Schiller P, Volman S: **Quantitative studies of single-cell properties in monkey striate cortex. iv. corticotectal cells**. *Journal of Neurophysiology* 1976, **39**:1352–1361.

[18] Nowak L, Bullier J: **The timing of information transfer in the visual system**. In *Cerebral Cortex: Extrastriate Cortex in Primate*, Edited by Rockland K, Kaas J, Peters A, New York, Plenum Publishing Corporation; 1997, 205–242. 1997.

[19] Cerkevich CM, Lyon DC, Balaram P, Kaas JH: **Distribution of cortical neurons projecting to the superior colliculus in macaque monkeys**. *Eye Brain* 2014, **2014**:121–137.

[20] Knierim J, Van Essen D: **Neuronal responses to static texture patterns in area V1 of the alert macaque monkey**. *Journal of Neurophysiology* 1992, **67**:961–80.

[21] Yan Y, Zhaoping L, Li W: **Bottom-up saliency and top-down learning in the primary visual cortex of monkeys**. *Proceedings of the National Academy of Sciences* 2018, **115**:10499–10504.

[22] Tehovnik E, Slocum W, Schiller P: **Saccadic eye movements evoked by microstimulation of striate cortex**. *The European Journal of Neuroscience* 2003, **17**:870–8.

[23] Jazayeri M, Lindbloom-Brown Z, Horwitz GD: **Saccadic eye movements evoked by optogenetic activation of primate v1**. *Nature neuroscience* 2012, **15**:1368–1370.

[24] Westerberg JA, Schall JD, Woodman GF, Maier A: **Feedforward attentional selection in sensory cortex**. *Nature Communications* 2023, **14**:article number: 5993.

[25] Thompson KG, Hanes DP, Bichot NP, Schall JD: **Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search**. *Journal of neurophysiology* 1996, **76**:4040–4055.

[26] Constantinidis C, Steinmetz MA: **Neuronal responses in area 7a to multiple-stimulus displays: I. neurons encode the location of the salient stimulus**. *Cerebral Cortex* 2001, **11**:581–591.

[27] White BJ, Kan JY, Levy R, Itti L, Munoz DP: **Superior colliculus encodes visual saliency before the primary visual cortex**. *Proceedings of the National Academy of Sciences* 2017, **114**:9451–9456.

[28] Zhaoping L: **From the optic tectum to the primary visual cortex: migration through evolution of the saliency map for exogenous attentional guidance**. *Current opinion in neurobiology* 2016, **40**:94–102.

[29] Wu R, Xu J, Li C, Zhang Z, Lin S, Li Ly, Li Yt: **Preference-independent saliency map in the mouse superior colliculus**. *Communications biology* 2025, **8**:article number: 565.

[30] Ipata AE, Gee AL, Gottlieb J, Bisley JW, Goldberg ME: **Lip responses to a popout stimulus are reduced if it is overtly ignored**. *Nature neuroscience* 2006, **9**:1071–1076.

[31] Gaspelin N, Ma X, Luck SJ: **Signal suppression 2.0: An updated account of attentional capture and suppression**. *Psychonomic bulletin & review* 2025, **32**:2648–2668.

[32] Buschman TJ, Miller EK: **Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices**. *science* 2007, **315**:1860–1862.

[33] Bisley J, Goldberg M: **Attention, intention, and priority in the parietal lobe**. *Annual Review of Neuroscience* 2010, **33**:1–21.

[34] Zhou H, Desimone R: **Feature-based attention in the frontal eye field and area V4 during visual search**. *Neuron* 2011, **70**:1205–1217.

[35] Duecker K, Shapiro KL, Hanslmayr S, Griffiths BJ, Pan Y, Wolfe JM, Jensen O: **Guided visual search is associated with target boosting and distractor suppression in early visual cortex**. *Communications Biology* 2025, **8**:article number: 912.

[36] Donk M, Van Zoest W: **Effects of salience are short-lived**. *Psychological Science* 2008, **19**:733–739.

[37] Liu C, Liu C, Huber L, Zhaoping L, Zhang P: **The superficial layers of the primary visual cortex create a saliency map that feeds forward to the parietal cortex**. *Plos Biology* 2025, **23**:e3003159.

[38] Stoll J, Thrun M, Nuthmann A, Einhäuser W: **Overt attention in natural scenes: Objects dominate features**. *Vision research* 2015, **107**:36–48.

[39] Mannan S, Kennard C, Husain M: **The role of visual salience in directing eye movements in visual object agnosia**. *Current Biology* 2009, **19**:R247–8.

[40] Meng M, Tong F: **Can attention selectively bias bistable perception? differences between binocular rivalry and ambiguous figures**. *Journal of vision* 2004, **4**:article 2.

[41] Hsu YH, Chen CC: **Eye-movement patterns for perceiving bistable figures**. *Journal of Vision* 2025, **25**:article 3.

•• [42] Kim T, Pasupathy A: **Neural correlates of crowding in macaque area V4**. *Journal of Neuroscience* 2024, **44**:e2260232024.

Examined neural responses in monkey V4 to an object presented with or without surrounding objects, and found that shape selectivity in V4 neurons is reduced by surrounding objects and that this reduction is less when the surrounded object is salient. This finding suggests that visual information loss between V1 and downstream V4 is less severe when visual inputs are salient, consistent with V1SH since salient visual inputs evoke higher V1 responses (which are therefore less vulnerable to subsequent degradation by an information bottleneck).

[43] Zhaoping L: **Contrast-reversed binocular dot-pairs in random-dot stereograms for depth perception in central visual field: Probing the dynamics of feedforward-feedback processes in visual inference**. *Vision Research* 2021, **186**:124–139.

[44] Whitney D, Levi DM: **Visual crowding: A fundamental limit on conscious perception and object recognition**. *Trends in cognitive sciences* 2011, **15**:160–168.

[45] Rosenholtz R, Yu D, Keshvari S: **Challenges to pooling models of crowding: Implications for visual mechanisms**. *Journal of vision* 2019, **19**:aritcle 15.

[46] Nuthmann A: **How do the regions of the visual field contribute to object search in real-world scenes? evidence from eye movements.** *Journal of Experimental Psychology: Human Perception and Performance* 2014, **40**:342–360.

•• [47] Zhaoping L: **Testing the top-down feedback in the central visual field using the reversed depth illusion**. *iScience* 2025, **28**:112223.

Showed that the reversed depth illusion in contrast-reversed random dot stereograms, normally only visible in the peripheral visual field, becomes visible in the central visual field when visual inputs are backward masked to interrupt top-down feedback query to aid recognition, as predicted by the central-peripheral dichotomy theory.

[48] Zhaoping L, Ackermann J: **Reversed depth in anticorrelated random-dot stereograms and the central-peripheral difference in visual inference**. *Perception* 2018, **47**:531–539.

[49] Tanabe S, Umeda K, Fujita I: **Rejection of false matches for binocular correspondence in macaque visual cortical area v4**. *Journal of Neuroscience* 2004, **24**:8170–8180.

[50] Kanizsa G: **Subjective contours**. *Scientific American* 1976, **234**:48–53.

• [51] Moore CM, Zheng Q, Semizer Y: **Perceptual organization is limited in peripheral vision: Evidence from configural superiority**. *Journal of Vision* 2025, **25**:article number: 16.

Showed that percepts of surfaces in 3D world arising from surface completion, 3D shape from 2D images, transparency/surface scission, and shape from closure are less visible in the peripheral visual field even when difference in visual acuity is compensated for. Based on these findings, the CPD theory predicts that such perceptual effects rely on top-down feedback queries which are lacking in the representations of the peripheral visual field.

[52] Majka L P Zhaoping, Rosa M: **A central-field focus in ventral-stream feedback to V1 in primates: theoretical prediction confirmed**. *Oral presentation, Vision Sciences Society Annual Meeting, May 2026* 2026, .

[53] Sims SA, Demirayak P, Cedotal S, Visscher KM: **Frontal cortical regions associated with attention connect more strongly to central than peripheral V1**. *NeuroImage* 2021, **238**:article:118246.

•• [54] Morales-Gregorio A, Kurth AC, Ito J, Kleinjohann A, Barthélemy FV, Brochier T, Grün S, van Albada SJ: **Neural manifolds in V1 change with top-down signals from V4 targeting the foveal region**. *Cell Reports* 2024, **43**:114371.

Inferred top-down feedback from V4 to V1 in monkeys by a Granger causality analysis of simultaneous neural activities in V1 and V4, and found that the inferred feedback is concentrated at and around fovea. This finding supports the CPD theory, which predicts that top-down feedback to V1 along the ventral visual pathway is focused on V1's central field representation.

[55] Kar K, DiCarlo JJ: **Fast recurrent processing via ventrolateral prefrontal cortex is needed by the primate ventral stream for robust core visual object recognition**. *Neuron* 2021, **109**:164–176.

[56] Xin Y, Yan Y, Li W: **A central and unified role of corticocortical feedback in parsing visual scenes**. *Nature Communications* 2025, **16**:article number: 6930.

[57] Pizzuti A, Gulban OF, Huber L, Peters JC, Goebel R: **In the brain of the beholder: bistable motion reveals mesoscopic-scale feedback modulation in v1**. *Brain Structure and Function* 2025, **230**:47.

[58] Levi DM: **Rethinking amblyopia 2020**. *Vision research* 2020, **176**:118–129.

[59] Hess RF: **Towards a principled and efficacious approach to the treatment of amblyopia. a review**. *Vision Research* 2025, **226**:108503.

[60] Lu ZL, Dosher BA: **Current directions in visual perceptual learning**. *Nature reviews psychology* 2022, **1**:654–668.

[61] Li W, Piëch V, Gilbert C: **Perceptual learning and top-down influences in primary visual cortex**. *Nature Neuroscience* 2004, **7**:651–657.

[62] Melcher D, Colby CL: **Trans-saccadic perception**. *Trends in cognitive sciences* 2008, **12**:466–473.

• [63] Liang J, Zhaoping L: **Trans-saccadic integration for object recognition peters out with pre-saccadic object eccentricity as target-directed saccades become more saliency-driven**. *Vision Research* 2025, **226**:number 108500.

By measuring locations and durations of the fixations during visual search for a moderately salient target in a cluttered image, this paper shows that even when the target is too peripheral to be recognized at all, saccades are still directed toward it at rates many times greater than chance; when the target is less peripheral, recognition results from integrating pre- and post-saccadic views. These findings demonstrate that looking can occur without seeing or recognition, and that seeing can integrate visual inputs across saccades via looking.

[64] Kroell LM, Rolfs M: **The magnitude and time course of pre-saccadic foveal prediction depend on the conspicuity of the saccade target**. *Elife* 2025, **12**:RP91236.

[65] Golomb JD, Mazer JA: **Visual remapping**. *Annual review of vision science* 2021, **7**:257–277.

[66] Williams MA, Baker CI, De Beeck HPO, Shim WM, Dang S, Triantafyllou C, Kanwisher N: **Feedback of visual object information to foveal retinotopic cortex**. *Nature neuroscience* 2008, **11**:1439–1445.

[67] Knapen T, Swisher JD, Tong F, Cavanagh P: **Oculomotor remapping of visual information to foveal retinotopic cortex**. *Frontiers in systems neuroscience* 2016, **10**:54.

[68] Fan X, Wang L, Shao H, Kersten D, He S: **Temporally flexible feedback signal to foveal cortex for peripheral object recognition**. *Proceedings of the National Academy of Sciences* 2016, **113**:11627–11632.

[69] Oletto CM, Contemori G, Cessa R, Battaglini L, Bertamini M: **Foveal masking impairs orientation discrimination of peripheral low-level stimuli**. *Heliyon* 2026, **12**:e44279.

[70] Rucci M, Poletti M: **Control and functions of fixational eye movements**. *Annual review of vision science* 2015, **1**:499–518.

[71] Wurtz RH: **Corollary discharge contributions to perceptual continuity across saccades**. *Annual review of vision science* 2018, **4**:215–237.

• [72] Wang X, Zhang C, Yang L, Jin M, Goldberg ME, Zhang M, Qian N: **Perisaccadic and attentional remapping of receptive fields in lateral intraparietal area and frontal eye fields**. *Cell Reports* 2024, **43**:113820.

> Showed that remapping directions in the LIP and FED change with time from attentionally driven convergent remapping to corollary-discharge-drive forward remapping. This suggests a temporal separation of a looking phase (during the convergent remapping for a shift of spatial attentional focus) and a seeing phase (during the forward remapping for perceptual stability) around the time of a saccade.

[73] Cavanagh P, Melcher D: **Steerable autoencoders underlying remapping, spatiotopy, and visual stability**. *PsyArXiv,2026, osfio/preprints/psyarxiv/5cku8_v2* 2025, .

[74] Luck SJ, Vogel EK: **Visual working memory capacity: from psychophysics and neurobiology to individual differences**. *Trends in cognitive sciences* 2013, **17**:391–400.

[75] Yates JL, Coop SH, Sarch GH, Wu RJ, Butts DA, Rucci M, Mitchell JF: **Detailed characterization of neural selectivity in free viewing primates**. *Nature Communications* 2023, **14**:article number 3656.

[76] Sereno MI, Huang RS: **Multisensory maps in parietal cortex**. *Current opinion in neurobiology* 2014, **24**:39–46.

[77] Vater C, Wolfe B, Rosenholtz R: **Peripheral vision in real-world tasks: A systematic review**. *Psychonomic bulletin & review* 2022, **29**:1531–1557.

[78] Zhaoping L: **Peripheral and central sensation: multisensory orienting and recognition across species**. *Trends in Cognitive Sciences* 2023, **27**:539–552.