이학박사 학위논문

Neuroimaging studies on the role of human primary visual cortex during perceptual decision-making

지각판단 중 인간 1차시각피질의 역할 규명을 위한 뇌영상연구

2015년 8월

서울대학교 대학원
자연과학대학 뇌인지과학과
최 경 환

Neuroimaging studies on the role of human primary visual cortex during perceptual decision-making

지도 교수 이 상 훈

이 논문을 이학박사 학위논문으로 제출함

2015년 8월

서울대학교 대학원

뇌인지과학과

최 경 환

최경환의 박사 학위논문을 인준함

2015년 5월

위 원 장 이 춘 길 (인)

부위원장 이 상 훈 (인)

위 원 **Randolph Blake** (인) RBlake

위 원 강 민 석 (인)

위 원 이 종 호 (인)

Abstract

Neuroimaging studies on the role of human primary visual cortex during perceptual decision-making

Kyoung whan Choe

Department of Brain and Cognitive Sciences

The Graduate School

Seoul National University

Primary visual cortex (V1) forms the initial cortical representation of objects and events in our visual environment, and it distributes information about that representation to higher cortical areas within the visual hierarchy. Decades of work have established tight linkages between neural activity occurring in V1 and features comprising the retinal image, but it remains debatable how that activity relates to perceptual decisions. An actively debated question is the extent to which V1 responses determine, on a trial-by-trial basis, perceptual choices made by observers. By inspecting the population activity of V1 from human observers engaged in a difficult visual discrimination task, we tested one essential prediction of the deterministic view: choice-related activity, if it exists in V1, and stimulus-related activity should occur in the same neural ensemble of neurons at the same time. Our findings do not support this prediction: while cortical activity signifying the variability in choice behavior

was indeed found in V1, that activity was dissociated from activity representing stimulus differences relevant to the task, being advanced in time and carried by a different neural ensemble. Moreover, realizing that small deviations in fixational eye movements could affect our fMRI measurements, we tested and confirmed that this pattern of results cannot be attributed to fixational eye movements. The spatiotemporal dynamics of population responses suggest that short-term priors, perhaps formed in higher cortical areas involved in perceptual inference, act to modulate V1 activity prior to stimulus onset without modifying subsequent activity that actually represents stimulus features within V1.

Keywords: visual perception, perceptual decision, primary visual cortex (V1), fMRI, eye-tracking, choice probability

Student Number: 2010-31300

Contents

LIST OF FIGURES

1. Introduction

An enduring challenge in visual neuroscience has been to understand how neural activity in visual cortex relates to what we see. During the decades immediately following Hubel and Wiesel's seminal discoveries (Hubel and Wiesel, 1962), research focused on predicting neural responses to simple visual features in the primary visual cortex (V1), where the first cortical representations of visual information are formed. More recently the challenge has expanded to neural activity in response to dynamic visual stimuli embedded in more complex contexts, leading to neural models of visual cortex that incorporate non-linear neural operations such as gain control and normalization (Carandini and Heeger, 2012). Moreover, the scope of work on this challenge has expanded to human brain imaging studies that seek to identify (Kay et al., 2008) and even reconstruct (Nishimoto et al., 2011) natural scenes by decoding cortical responses.

While establishing a tight linkage between V1 activity and stimulus conditions setting off those neural chain reactions, research has also explored V1 activity's impact on the final outcome of neural processes ensuing from it. A question at the core of this exploration is whether or how V1 neurons' responses contribute to perceptual judgments. A fruitful strategy for

addressing this question is to compute a trial-to-trial correlation between single neurons' responses to physically identical stimuli and perceptual choices made by observers performing a difficult perceptual decision task on those stimuli, dubbed 'choice probability (CP)'. Whereas above-chance-level CPs have been consistently found in high-tier sensory areas (for reviews, see Nienborg and Cumming, 2010; Nienborg et al., 2012), non-sensory associative areas (Hernandez et al., 2010) and subcortical areas (Liu et al., 2013), the presence of CPs in V1 remains controversial (Grunewald et al., 2002; Nienborg and Cumming, 2006; Palmer et al., 2007). Moreover, mere demonstration of statistically significant CPs does not necessarily support a causal role of V1 in settling perceptual choices. Significant CPs in given neurons may arise when a fraction of activity of those neurons is modulated by other neurons that actually cause choices, either via feedback (Nienborg and Cumming, 2009) or via horizontal connections (Cohen and Newsome, 2009; Law and Gold, 2009; Nienborg and Cumming, 2010). Another caveat to previous attempts at relating V1 activity to its perceptual consequences is that CPs have not been estimated for population responses, despite growing evidence for the importance of population activity in neural representation of sensory signals (Hol and Treue, 2001; Chen et al., 2006; Jazayeri and Movshon, 2007; Graf et al., 2011; Berens et al., 2012).

By acquiring functional magnetic resonance imaging (fMRI) measurements of V1 population activity while human observers performed a difficult ring-size discrimination task, we identified choice-correlated responses and compared them to stimulus-correlated responses. We reasoned that if V1 activity causally contributes to choices, choice-related V1 responses should match stimulus-related responses both in timing and in neural origin. Our fMRI measurements of V1 population activity, however, run counter to this prediction: stimulus- and choice-related components arise at different points in time and in different cortical subpopulations. Moreover, realizing that small deviations in fixational eye movements could affect our fMRI measurements, we tested and confirmed that this pattern of results cannot be attributed to fixational eye movements.

2. Experiment 1: Neural signatures of stimulus and choice are dissociated in population activity of human V1 during perceptual decision-making

2.1. Fine ring-size classification task

We devised a difficult one-interval two-alternative forced-choice task wherein observers viewed one of three different-sized rings that were always symmetrically centered around a central fixation mark and classified the ring as either ‘small’ or ‘large’ (Fig. 1A). To detect cortical signatures of stimulus and choice simultaneously in trial-to-trial population fMRI activity of V1, we optimized the task and stimulus parameters as follows.

As stimuli whose subtle differences are best resolvable with population fMRI measurements, we opted for concentric rings whose size was the feature dimension of relevance for perceptual decision (Fig. 1B). Owing to its configuration, a concentric ring engages a large ensemble of neurons whose peak activities within the V1 retinotopic map will vary with the eccentricity of the ring’s circumference (i.e., ring size). Because of this feature of concentric ring stimuli, we could exploit the fact that the retinotopic architecture of V1 is resolvable with a mesoscopic-scale analysis

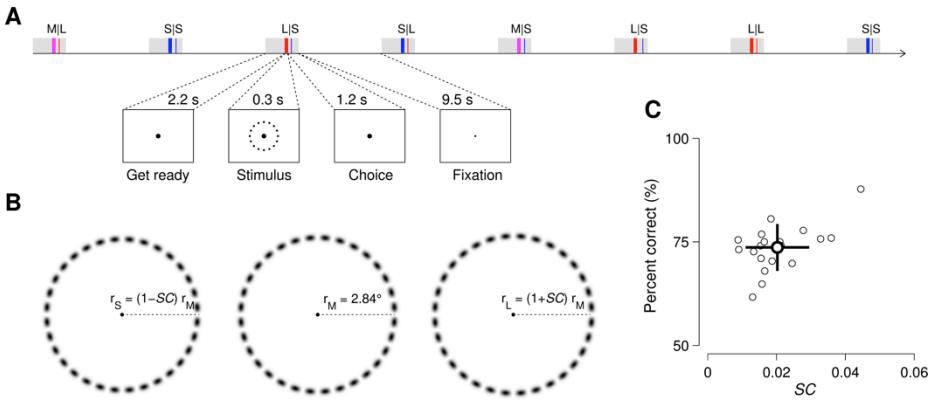

Figure 1. Task, stimuli and behavioral performance.

A, An example sequence of trials and phases constituting a trial. Eight exemplar trials are shown, each belonging to one of the six possible classes (labeled by letter symbols at the top, 'stimulus|choice'). The gray rectangles represent sparse, brief periods during which observers were warned of stimulus onset, viewed a ring stimulus (colored thick vertical bars) and made a choice at a particular point in time (colored thin vertical bars). ***B,*** Examples of the three, different-sized rings. The luminance polarity is reversed here for illustrative purpose. ***C,*** Distribution of threshold *SC* values (on the horizontal axis) and actual performances in the main fMRI experiment (on the vertical axis). The small circles represent individual observers, and the large circle and error bars are their population average and standard deviation (SD), respectively.

of fMRI (Lee et al., 2005; Dumoulin and Wandell, 2008; Kay et al., 2008; Park et al., 2013), allowing us to take advantage of population coding of subtle stimulus differences (Paradiso, 1988; Pouget et al., 2000; Jazayeri and Movshon, 2006). Also, these ring stimuli provide the additional benefit of encouraging observers to maintain central fixation, for this insures optimal retinal stimulation for performance of the task: shifting fixation toward any selected portion of a ring inevitably images the remaining portions of the ring at even more eccentric areas of the retina with poorer spatial resolution.

We optimized the spatiotemporal parameters of the ring stimuli to

generate an optimal level of uncertainty in perceived size of the ring, so as to observe cortical representations of choice information. The size contrast (*SC* in Fig. 1B) between the rings was calibrated to be at a threshold level for each individual, based on the performance in pre-scan practice trials carried out in the scanner (0.020 ± 0.009; mean ± standard deviation (SD) across observers; see Section 6.3 for details). In addition, we created trials in which observers' choices would not correlate with stimuli by introducing a middle-sized ring (M-ring) whose radius (r_M=2.84°) was halfway between the radii of the smallest (S-ring) and largest (L-ring) rings (Fig. 1B). Observers were not told there would be three different-sized rings; they were only told to classify each ring as 'small' or 'large'. The ring was shown for 0.3 s ('stimulus' period in Fig. 1A), a duration sufficiently long to produce reliable fMRI responses in V1 yet sufficiently brief to contribute to a degree of uncertainty in perceived size of the ring. This tailor-made calibration of stimulus size and duration succeeded at holding observers' performances during the fMRI scan sessions within a threshold range (73.7% ± 5.7%; mean ± SD across observers; Fig. 1C).

We adopted a sparse event-related design (Fig. 1A) to individuate trial-to-trial fluctuations of fMRI responses to repeated presentations of the rings. To encapsulate neural events associated with a single trial of

perceptual decision-making within a short period of time, we forced observers to make a perceptual choice within 1.2 s after stimulus onset, which resulted in actual response times with the mean of 0.66 s and SD of 0.13 s (2825 trials, pooled across observers). To minimize carryover effects in fMRI signal between consecutive trials due to hemodynamic delay, individual trials were separated by 13.2 s. To stabilize eye position and to regulate cortical and cognitive states during this inter-trial period, we required observers to maintain their gaze on the fixation dot (diameter 0.12°) and signaled an upcoming trial by increasing the size of the fixation dot slightly (diameter 0.18°) 2.2 s before stimulus onset. It is worth noting that stable, central fixation is essential for optimizing psychophysical performance, as mentioned above, and for successful measurement of high resolution fMRI responses to the ring stimuli. To prevent unwanted feedback-related events from contaminating trial-locked fMRI measurements, we did not provide trial-by-trial feedback. Instead, observers were updated about their overall performance at the end of each scan run containing 26 trials.

2.2. Definition of eccentricity-tuning for individual voxels

While observers performed the ring-size classification task with parameters optimized as described above, we acquired time-series of fMRI measurements from a population of unit gray-matter volumes (voxels) in the V1 cortical surface whose width (0.5°~7.5°; a region marked by color spectrum in Fig. 2A) was larger than the site directly stimulated by the rings (2.72°~2.96°, i.e., the smallest and largest rings, respectively, used in the experiment across observers; dotted circle in the inset of Fig. 2A). To inspect trial-to-trial patterns of population responses in a feature dimension relevant to the perceptual decision task, we first mapped the coordinates of those individual voxels in visual eccentricity space.

By applying the model-based population receptive field estimation method (Dumoulin and Wandell, 2008) to fMRI time-series responses to an expanding/contracting annulus, we defined the eccentricity-tuning curves with a Gaussian function for individual voxels in each observer's V1 (Fig. 2A; see Section 6.5 for details). The range of estimated widths of the tuning curves (1.08° ± 0.51°; mean ± SD across 6,379 V1 voxels with R^2>0.4, pooled across observers) and their positive correlation with preferred eccentricity (Pearson's r=0.32 ± 0.08, 10^{-17}<p<0.001; mean ± SD across observers; Fig. 2B) were consistent with previous studies

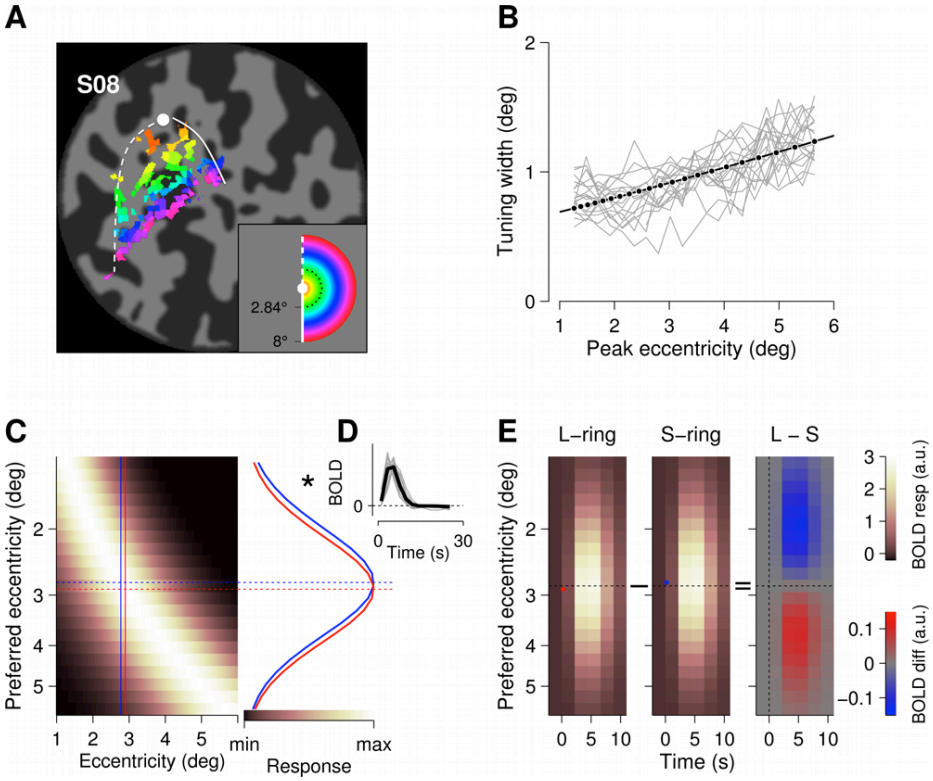

Figure 2. Eccentricity-tuning curves in V1.

A*,** Eccentricity map of V1 from observer S08 shown on the flattened left occipital cortex. The white dot, dashed and solid curves demarcate the V1 cortical sites representing the fovea, the upper vertical meridian and the lower vertical meridian, respectively, in visuotopic space. The colors indicate the eccentricities of the voxels with high goodness-of-fit by the tuning-curve model (R^2>0.4; see Section 6.5). The black dotted circle in the color legend represents the eccentricity of the M-ring stimulus. ***B, Relationship between preferred eccentricity and tuning width. The gray lines plot tuning widths (the vertical axis) as a function of preferred eccentricity (the horizontal axis) for individual observers, and the black line is a pseudo linear regression of the eccentricity to the tuning width (see Section 6.5), which was used to estimate the eccentricity-tuning curves at for the twenty-one cortical bins (the black dots) shown in ***C***. ***C,*** Population-averaged eccentricity-tuning curves. The horizontal axis specifies stimulus eccentricity, the vertical axis the estimated preferred eccentricity of a cortical bin, and the intensity corresponds to the normalized cortical activity. The solid blue and red vertical lines indicate the population averages of the eccentricities of the S- and L-rings, respectively. The dotted blue and red horizontal lines indicate the two cortical sites that show maximum responses to the S- and L-rings, respectively. The blue and red curves on the right are the predicted population responses to the S- and L-rings, respectively. ***D,*** Individual (thin gray) and averaged (thick black) HIRFs. ***E,*** Predicted fMRI responses to ring stimuli. In all of the three panels, the horizontal and vertical axes specify time relative to stimulus onset (indicated by the colored dots) and the preferred eccentricity of a cortical bin, respectively. The L-ring and S-ring panels show responses to the L- and S-ring stimuli, respectively, which were predicted by convoluting the red and blue curves in ***C*** with the averaged V1 HIRF in ***D***. The L-S panel shows the differences between the L- and S-ring panels, with hue and saturation representing the sign and magnitude of the differential responses, respectively.

(Dumoulin and Wandell, 2008; Kay et al., 2008; Harvey and Dumoulin, 2011; Park et al., 2013), supporting the validity of our estimation procedure. These tuning estimates from individual observers (thin gray lines in Fig. 2B) were merged and summarized by fitting a power function (black line in Fig. 2B; see Eq. 5; Duncan and Boynton, 2003) to obtain a reference eccentricity map (Fig. 2C) with twenty-one eccentricity bins (whose centers are marked by the filled circles in Fig. 2B; see Section 6.6 for details).

The eccentricity map (Fig. 2C) allowed us to preview the potential population fMRI responses to the different-sized rings. This map predicts that the S- and L-rings (whose eccentricities are marked by the vertical blue and red solid lines, respectively, in Fig. 2C; the group average values, r_S=2.78° and r_L=2.90°, were used) produce profiles of activity that are broad across-eccentricity but that are nonetheless slightly offset with respect to one another (blue and red bell-shape curves with dotted lines at center, plotted on the right-hand vertical axis in Fig. 2C). These spatial profiles of predicted cortical responses were then convolved with the V1 hemodynamic impulse response functions (HIRFs; Fig. 2D), which were estimated from the retinotopy-mapping scan runs, to produce matrices of noise-free fMRI blood-oxygenation-level dependent (BOLD) responses to the L-ring and S-ring stimuli (L-ring and S-ring panels, respectively, in Fig. 2E; see Section 6.8 for

details). By subtracting the matrix predicting responses to the S-ring from that predicting responses to the L-ring, we obtained the matrix predicting differential fMRI responses to the L- and S-rings (L-S panel in Fig. 2E). The predicted differential responses peaked in time at 3.3~5.5 seconds after stimulus onset due to hemodynamic delay, and in space at two flanking banks of eccentricity bins (blue and red pixels in L-S panel, Fig. 2E), representing the foveal and the peripheral sides of the rings.

2.3. Definition of trial-related matrices of population responses

With eccentricity-tuning curves defined for individual voxels, we expressed the V1 population responses of individual observers performing the ring-size classification task in a matrix with two dimensions, one defined by voxels' preferred eccentricities and the other defined by time frames at which fMRI measurements were acquired (Fig. 3A). To increase signal-to-noise ratio at the cost of resolution, and to be able to merge data across individuals, fMRI responses of neighboring voxels were summed with windows of weights centered at discrete-step eccentricity values (Fig. 3B; see Section 6.6 for details), as done similarly in previous fMRI studies (Brefczynski-Lewis et al., 2009; Kolster et al., 2010; Park et al., 2013). This smoothing procedure

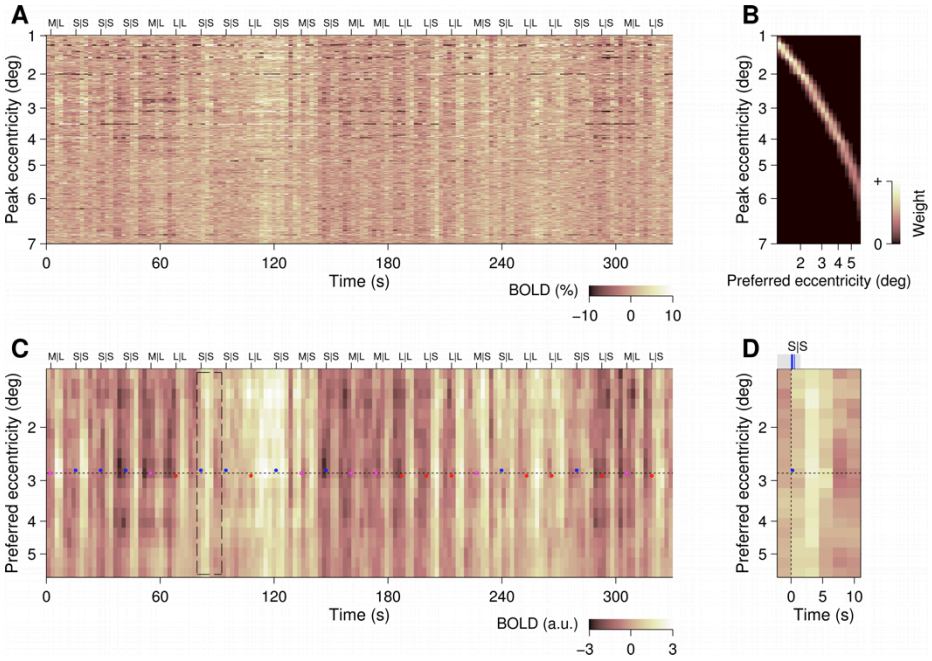

Figure 3. Definition of trial-related matrices of population responses.

A, Eccentricity-sorted individual voxels' time-series of fMRI measurements during a single scan run from S08. The horizontal and vertical axes specify the time bin of measurement and the peak eccentricity of voxels, respectively, with image intensity corresponding to level of fMRI activity. The letters at the top represent the stimulus shown and the choice made by the observer in a given trial ('stimulus|choice'). ***B,*** Kernels used for spatial smoothing over eccentricity. The horizontal axis specifies the preferred eccentricity of a target eccentricity bin in cortical scale. The vertical axis specifies the center of a given voxel's eccentricity-tuning curve in visuotopic scale. The image intensities correspond to the weights of the smoothing kernels, whose widths were constant over the eccentricity in cortical scale. ***C,*** Responses at eccentricity bins. The format is identical to the one in ***A***, except that the vertical axis specifying the preferred eccentricity is scaled in cortical distance. The blue, magenta and red circles indicate the spatiotemporal locations of S-, M-, and L-rings, respectively, presented over trials. ***D,*** An example matrix of trial-related population responses. Each trial-related matrix spanned 13.2 s (2.2 s per frame) in time and about 5.5 degrees in space. The responses within the dashed black box in ***C*** are re-plotted here, with time axis magnified. The shaded rectangle and the bars within it represent the same events depicted in Fig. ***1A***. The dotted vertical and horizontal lines indicate the stimulus onset and the eccentricity of the M-ring stimulus, respectively.

resulted in a 21 (the number of eccentricity bins)-by-150 (the number of time frames per scan run) matrix of responses for each scan run (Fig. 3C). As a final step of preparatory analysis of fMRI measurements, we dissected the response matrix for each scan run into 'trial-related' matrices of responses

with twenty-one rows representing the eccentricity bins and six columns representing the time frames defined relative to stimulus onset (an example shown by the matrix demarcated by the dotted box in Fig. 3C). Then, with these trial-related matrices (Fig. 3D), we searched for cortical signatures of stimulus and choice by examining whether responses in each of those 126 (=21 eccentricity bins x 6 time frames) individual spatiotemporal cells co-vary with stimuli shown or choices made over trials.

2.4. Neither stimuli nor choices significantly correlated with raw responses

For each of the spatiotemporal cells of the trial-related matrix, we computed 'stimulus probabilities' (*SP*s; Tolhurst et al., 1983; Newsome et al., 1989) and 'choice probabilities' (*CP*s; Celebrini and Newsome, 1994; Britten et al., 1996) by assessing how well the trial-to-trial distributions of the raw responses predicted the stimulus actually presented and the choices that were made by individual observers. We sorted the individual trials into the six possible classes jointly defined by a stimulus shown and a choice made in a given trial − 'S|S', 'S|L', 'M|S', 'M|L', 'L|S' or 'L|L' ('stimulus|choice'; Fig. 4A). The *SP*s were estimated by comparing the distributions of the raw responses belonging to two 'stimulus-contrast' pairs (S-ring vs. L-ring;

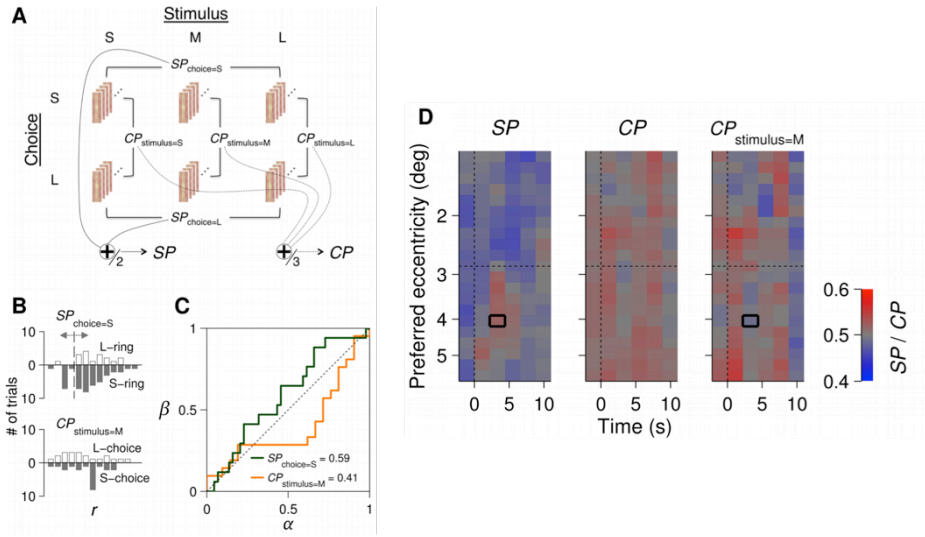

Figure 4. Computation of *SP*s and *CP*s.

A, Classification of trials and definition of stimulus- and choice-contrast pairs associated with *SP*s and *CP*s. A trial-related matrix for each trial was classified according to the 'stimulus|choice' class. *SP*s and *CP*s were computed by averaging the stimulus-contrast (horizontal brackets) and choice-contrast (vertical bracket) pairs of the 'stimulus|choice' classes, as indicated (Section 6.7). ***B,*** Example distributions of responses at a representative spatiotemporal bin (black squares in ***D***) from S08. The top panel contrasts the histograms of raw fMRI responses between the 'L-ring|S-choice' trials (open) and the 'S-ring|S-choice' trials (filled). The bottom panel contrasts the histograms of raw fMRI responses between the 'M-ring|L-choice' trials (open) and the 'M-ring|S-choice' trials (filled). The dashed vertical line is a classification criterion that is slid to generate ROC curves. ***C,*** Example ROC curves. The horizontal and vertical axes specify the false alarm and hit rates, respectively (see Section 6.7 for definitions of α and β). The green and orange curves were derived from the top and bottom, respectively, distributions in ***B***. ***D,*** Across-observer averages of *SP*, *CP* and $CP_{stimulus=M}$ computed for raw fMRI responses. The format for axes is identical to that in Fig. ***3D***. Hue and saturation represent stimulus or choice probability values, as indicated.

horizontal brackets in Fig. 4A) of these classes, wherein the choice factor was held constant, '**S**|S' vs. '**L**|S' ($SP_{choice=S}$) and '**S**|L' vs. '**L**|L' ($SP_{choice=L}$). By varying the location of the discrimination criterion over those distributions (top panel of Fig. 4B), we constructed a 'receiver operating characteristic (ROC)' curve (green curve in Fig. 4C) and computed a *SP* by summing the area under the ROC curve. We defined the grand *SP* by taking

the average of the two *SP*s associated with different choices, $SP_{choice=S}$ and $SP_{choice=L}$ (as indicated by the operations at the bottom of Fig. 4A). The *CP*s were estimated similarly, first computing the three individual *CP*s for the three 'choice-contrast' pairs (vertical brackets in Fig. 4A) of the distributions ('S|**S**' vs. 'S|**L**' ($CP_{stimulus=S}$), 'M|**S**' vs. 'M|**L**' ($CP_{stimulus=M}$), and 'L|**S**' vs. 'L|**L**' ($CP_{stimulus=L}$)) and then averaging those three *CP*s. As a reminder, our definition of *CP*s is different from that used in single-cell studies (see Section 6.7 for details).

We estimated the *SP*s and *CP*s exhaustively over the entire trial-related matrix of the raw responses (Fig. 4D), but none of those values reached statistical significance (minimum TFCE-corrected p=0.41, 0.87, and 0.60 among 126 spatiotemporal bins, respectively, for *SP*, *CP* and $CP_{\text{stimulus=M}}$). Only the overall pattern of the across-observer averages of *SP*s (*SP* panel in Fig. 4D) exhibited a somewhat systematic distribution, which appeared similar to the pattern of the model-predicted BOLD differential responses (L-S panel in Fig 2E).

We wondered that these weak probability values in the raw responses might have been caused by the interference from large non-specific fluctuations in background cortical activity. Recent optical imaging and fMRI studies, wherein large-size population neural activities were

monitored simultaneously in early visual cortex, observed large-scale co-fluctuations over an entire population of neurons under observation regardless of whether or not individual neurons' stimulus preferences match incoming visual input (Sharon and Grinvald, 2002; Fiser et al., 2004; Chen et al., 2006; Jack et al., 2006; Donner et al., 2008; Sirotin and Das, 2009; Sirotin et al., 2012). In line with these findings, the raw responses in our study waxed and waned in synchrony over the entire array of eccentricity bins, which is readily appreciated by visual inspection of the sample matrix of the raw population responses (Figs. 3A and 3C; top panel in Fig. 5A). The presence of these so called 'untuned responses' was supported by significant widespread correlations among the eccentricity bins (Pearson's r=0.62 (mean) ± 0.17 (SD across observers), the mean correlation of non-overlapping eccentricity bin pairs, which are demarcated with the dotted boundary in Fig. 5B). These significant positive correlations were not confined to the pairs of nearby eccentricity bins (Fig. 5C), but were also found between ones representing the directly stimulated visual region and the ones representing either the foveal or peripheral regions (e.g., Pearson's r=0.54 ± 0.17, 0.67 ± 0.13; the mean ± SD correlations of the foveal and peripheral pairs marked by the gray boxes in Fig. 5B, respectively) and even between the ones representing the foveal and peripheral regions (e.g., Pearson's r=0.57 ± 0.16

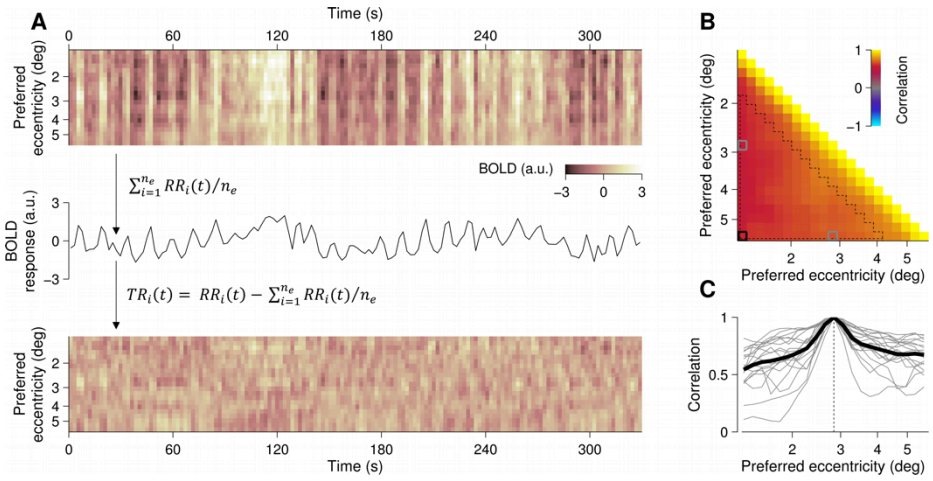

Figure 5. Filtering out a non-specific component from raw responses.

A, Definition of TRs via averaging and subtracting operations on RRs. The image in Fig. ***3C*** is re-plotted in the top panel. The middle panel shows the averages of RRs over the eccentricity bins at individual time frames. The TRs shown in the bottom panel was obtained by subtracting those averages from the RRs (see Section 6.9 for details). ***B,*** A matrix of population average correlations in RRs among eccentricity bins. The dotted line boundary demarcates the bin pairs whose fMRI measurements are not blended via spatial smoothing. The squares highlight the values of correlations for the three possible pairs from the time-series of RRs at the three eccentricity bins, representing 1.26°, 2.84°, and 5.65°, respectively. ***C,*** Correlations in RRs between all the eccentricity bins and the seed (2.84°) bin. The dotted vertical line indicates the preferred eccentricity of the seed bin, the visuotopic representation of which covers the locations of the ring stimuli. The gray thin lines are the correlations from individual observers, and the black thick line is their population average.

for the pair marked by the black box in Fig. 5B).

Having confirmed the non-specific nature of the moment-to-moment background fluctuations, we filtered out those correlated responses throughout the entire region of V1 under observation by averaging the raw responses across the entire set of eccentricity bins (middle panel in Fig. 5A) and subtracting that average from the raw responses at each time frame (bottom panel, Fig. 5A). These averaging and subtracting operations were validated by the additive nature of 'tuned' and 'untuned' responses

(Bianciardi et al., 2009; Cardoso et al., 2012; Schölvinck et al., 2012) and have been routinely employed in previous studies using optical imaging (Shtoyerman et al., 2000; Sharon and Grinvald, 2002; Benucci et al., 2009) and fMRI (Fox et al., 2006; Larsson et al., 2006; Donner et al., 2008; Pestilli et al., 2011; Schölvinck et al., 2012; Donner et al., 2013). Hereafter we will refer to the unfiltered raw responses as 'RRs (raw responses)' and the filtered responses as 'TRs (tuned responses)'.

2.5. Dissociated signatures of stimulus and choice in tuned responses

The subtraction of the untuned component from the RRs revealed clear signatures of stimulus and choice. To compute the *SP*s and *CP*s for the TRs, we followed the same procedure used for the RRs. Unlike the RRs, the TRs exhibited significant *SP*s and *CP*s, respectively, at different sets of spatiotemporal cells of the trial-related matrix (significant cells (corrected $p<0.05$) are marked with * in Fig. 6A; corrected for multiple comparisons across the 126 spatiotemporal cells using the 'threshold-free cluster enhancement' method (TFCE; Smith and Nichols, 2009)).

The *SP*s in the TRs (*SP* panel, Fig. 6A) were signed properly and clustered systematically both in space and time. At the time frames matched

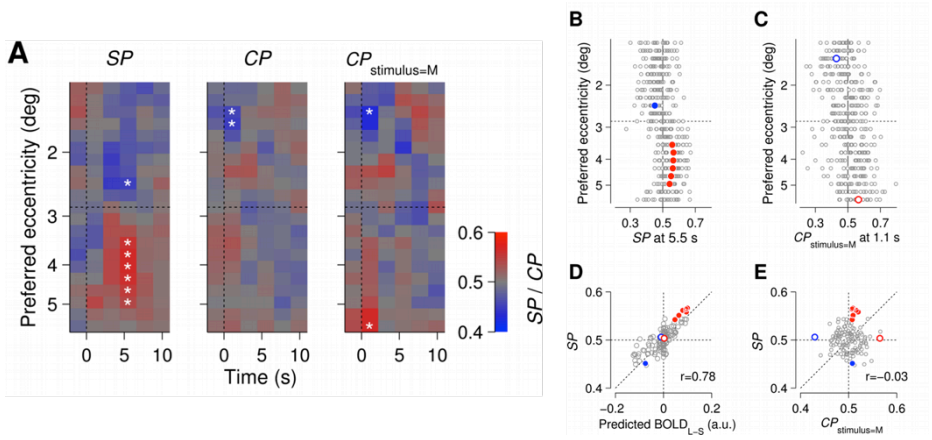

Figure 6. Stimulus and choice probabilities in tuned responses.

A, Across-observer averages of *SP*, *CP* and $CP_{stimulus=M}$ in tuned responses. The format is identical to that in Fig. ***4D***. The white asterisks mark the significant bins (TFCE-corrected p<0.05; see Section 6.7). ***B,*** *SP* values at 5.5 s after stimulus onset from individual observers. The *SP*s are plotted against the eccentricity bins. Gray circles represent individual observers, and the blue- and red-filled circles the *SP*s averaged across observers at eccentricity bins, at which they were significant, as indicated by the white asterisks in the *SP* panel in **A**. ***C,*** $CP_{stimulus=M}$ values at 1.1 s after stimulus onset from individual observers. The axis format and symbols are identical to those in **B**, except for the empty blue and red circles, which are the *CP*s averaged across observers at eccentricity bins, at which they were significant, as indicated by the white asterisks in the $CP_{stimulus=M}$ panel in **A**. ***D,*** Significant positive correlation between the model-predicted differential responses (L-S panel in Fig. ***2E***) and the observed *SP*s (*SP* panel in ***A***). Gray circles represent individual spatiotemporal cells, and the colored circles represent the cells wherein either significant *SP*s (as indicated by the corresponding filled circles in ***B***) or significant *CP*s (as indicated by the corresponding open circles in ***C***) were found. ***E,*** No correlation between the $CP_{stimulus=M}$s ($CP_{stimulus=M}$ panel in ***A***) and the *SP*s (*SP* panel in ***A***). The representation of spatiotemporal cells by the symbols is the same as in ***D***. Note that the filled and open symbols are located far away from one another, illustrating the spatiotemporal dissociation between the significant *SP*s and *CP*s.

to the typical hemodynamic delay (3.3 s and 5.5 s) from stimulus onset, the responses to the S-ring were greater than those to the L-ring within the cortical subregion representing the side of the rings nearer to the fovea (blue pixels with *SP*<0.5 in the *SP* panel of Fig. 6A), and the opposite was true within the cortical subregion representing the peripheral side of the rings (red pixels with *SP*>0.5 in the *SP* panel of Fig. 6A). This emergence of the sinusoidal-shape spatial profile of *SP*s centered around the stimulation site at

the time points a few seconds delayed from stimulus onset (Fig. 6B) is exactly what was previewed by our model prediction of differential cortical responses based on the eccentricity-tuning curves (Fig. 2E): the cortical sites that generate the largest differential responses to the ring stimuli with subtle differences are those with eccentricity preferences slightly deviated from the eccentricity of the stimuli. This strong resemblance between the spatiotemporal maps of the *SP*s and the model prediction of differential responses, as evidenced by the high cell-to-cell correlation between them (Fig. 6D; Pearson's $r=0.78$, $p<10^{-9}$, across 126 spatiotemporal cells), assures that our fMRI measurements, once corrected for non-specific background fluctuations, are reliable enough to delineate the cortical sites that encode the fine stimulus differences with high fidelity on a trial-to-trial basis.

Having characterized the V1 signature of stimulus by specifying which cortical sites carry that signature and when that signature is formed in relative to the stimulus onset, we set out to characterize the signature of choice in the same manner with the aim of examining whether the two signatures originate from the same neural population at the same time. This examination puts to a critical test the deterministic view of sensory neurons' role in perceptual decision, which posits that perceptual judgments on otherwise identical stimuli are caused by trial-to-trial fluctuations in

responses of the neuronal ensemble that participates in encoding sensory features of relevance to a given perceptual task (Newsome et al., 1989; Salzman et al., 1990; Celebrini and Newsome, 1994; Britten et al., 1996; Shadlen et al., 1996). Hence, we reasoned, as have previous single-cell studies (Celebrini and Newsome, 1994; Britten et al., 1996; Parker et al., 2002; Romo et al., 2002; Uka and DeAngelis, 2004; Purushothaman and Bradley, 2005; Gu et al., 2007; Gu et al., 2008; Law and Gold, 2008; Ghose and Harrison, 2009; Law and Gold, 2009; Price and Born, 2010; Smith et al., 2011; Liu et al., 2013), that, if the causal view is correct, significant *CP*s should be found in the vicinity of the spatiotemporal cells at which the significant *SP*s were identified.

The observed pattern of *CP*s, in fact, was inconsistent with this prediction in four important ways. First, we failed to observe significant *CP*s at any of those seven cells that housed significant *SP*s (marked with * in *SP* panel, Fig. 6A; blue or red filled circles in Fig. 6B) in the trial-related matrix of the TRs (*CP*= 0.51±0.06, 0.49±0.05, 0.49±0.05, 0.50±0.05, 0.51±0.06, 0.51±0.06, and 0.50±0.05; $CP_{\text{stimulus=M}}$= 0.51±0.06, 0.52±0.08, 0.51±0.08, 0.51±0.11, 0.52±0.10, 0.51±0.07, and 0.51±0.06; from foveal to periphery, respectively; mean±SD across observers; *CP* and $CP_{\text{stimulus=M}}$ panels in Fig. 6A). Second, instead, the significant *CP*s were found at the cells wherein the

insignificant *SP*s were found. The cells with the significant *CP*s were quite advanced in time and far away from the site with direct stimulation in space (marked with * in *CP* and $CP_{\text{stimulus=M}}$ panels of Fig. 6A; TFCE-corrected $p<0.05$). Only 1.1 s after the stimulus onset (fMRI activity at which probably reflects neural activity occurring before the stimulus onset), the responses at the cortical site representing a region very close to the fovea (1.51°; blue pixels with * in *CP* and $CP_{\text{stimulus=M}}$ panels of Fig. 6A; blue open circle in Fig. 6C) were greater in the S-choice trials than in the L-choice trials whereas the responses at the cortical site representing the far periphery (5.65°; red pixel with * in the $CP_{\text{stimulus=M}}$ panel of Fig. 6A; red open circle in Fig. 6C) were greater in the L-choice trials than in the S-choice trials. Third, in search of any hints of the meaningful relationship between the *SP*s and the *CP*s, we also considered the possibility that, despite the mismatch in statistical significance, the *SP*s and the *CP*s might have been weakly correlated with one another. However, the correlation analysis, which was conducted over the entire ensemble of cells constituting the trial-related matrix of the TRs, showed that the spatiotemporal distribution of *CP*s was not correlated (Fig. 6E; Pearson's $r=-0.03$, $p=0.74$, for $CP_{\text{stimulus=M}}$s) or even anti-correlated (Pearson's $r=-0.38$, $p<10^{-9}$ for *CP*s) with that of *SP*s. Fourth and finally, we further checked the possible involvement of the stimulus-encoding cortical

sites in representing choice-associated information by comparing the spatiotemporal pattern of the *CP*s to that of the model prediction of differential responses to the stimuli (L-S panel in Fig. 2E). Again, we failed to observe any significant correlations between the *CP*s and the model predictions (Pearson's r=-0.15 and 0.02, p=0.09 and 0.84, for *CP*s and $CP_{stimulus=M}$s, respectively).

2.6. Decoding stimulus and choice information from raw responses with population read-out weights

So far, reliable signatures of stimulus or choice were available only in the TRs derived by removing non-specific background fluctuations from the RRs in which those TRs were embedded. Does this imply that the large-scale, moment-to-moment fluctuations in 'untuned' activity impose a fundamental limitation on V1's capacity to carry stimulus- or choice-related information? That implication is not necessarily correct. Instead, the failure to find reliable signatures in the RRs could reflect the limitation of our 'local coding' strategy, which evaluated the stimulus- or choice-related variability in neural responses confined to local cortical sites separately. Indeed, if a decision stage in the brain relies on population coding to interpret sensory signals within V1 (Paradiso, 1988; Pouget et al., 2000), fluctuations in untuned

activity, a substantial fraction of which is shared by an entire population of encoding neurons, can be efficiently canceled at the decision stage regardless how large those fluctuations are.

To test this hypothesis, we revisited the RRs, this time decoding stimulus signals and choice signals from the RRs over the entire extent of the eccentricity matrix and computing the stimulus probabilities and choice probabilities by comparing the trial-to-trial distributions of those decoded signals at the population level. We will refer to these probabilities as 'population *SP*s' and 'population *CP*s' to distinguish them from the probabilities estimated at the individual cells of the trial-related matrix. For population decoding, we developed three different read-out weight profiles, implementing the three major decoding schemes proposed by previous studies (Gold and Shadlen, 2001; Jazayeri and Movshon, 2006; Graf et al., 2011; Berens et al., 2012; Haefner et al., 2013). The simplest form of read-out weights was a uniform read-out, in which the eccentricity bins are divided into either the 'fovea' pool or the 'periphery' pool, with uniform weights assigned to the bins within each pool (Fig. 7A). In the remaining two weight profiles, the eccentricity bins had non-uniform weights that were derived from the eccentricity-tuning curves according to two different task-optimal decoding schemes. In one scheme, the read-out weights were

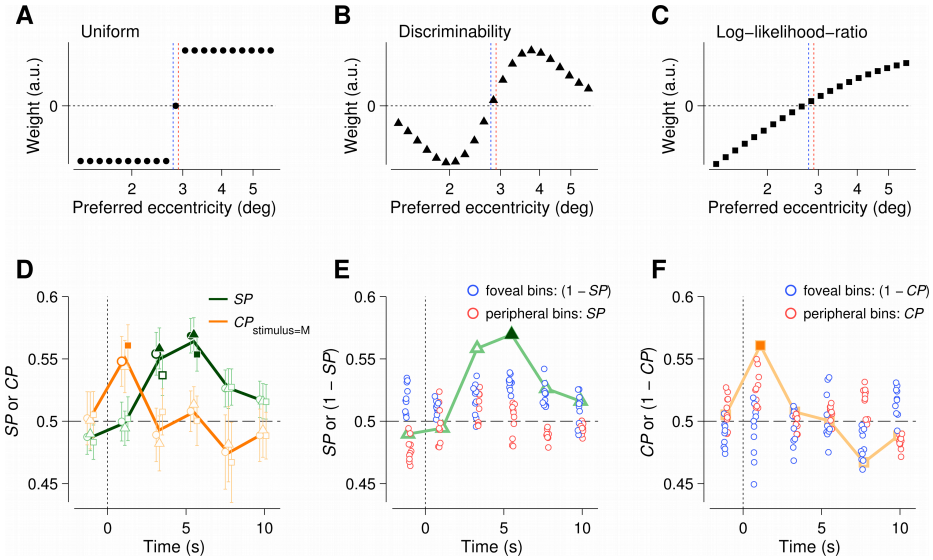

Figure 7. Population decoding of stimulus and choice information in raw responses.

A-C, Weight profiles defined by three different population decoding schemes. Individual symbols represent arbitrary-unit weight values assigned to eccentricity bins. ***D,*** Time-courses of across-observer averages of population *SP*s (green line and symbols) and population $CP_{stimulus=M}$s (orange line and symbols). The circles, triangles and squares represent the uniform (***A***), discriminability (***B***), and log-likelihood-ratio (***C***) weights, respectively. The salient open and filled symbols are the probability values significant (uncorrected) at $p<0.05$ and $p<0.005$, respectively. Error bars are SEM across observers. ***E,*** Comparison of population and individual *SP* values for the RRs. The blue open circles indicate the *SP* values from the 10 foveal bins, which were adjusted for preference by (1 - *SP*). The red open circles indicate the *SP* values from the 10 peripheral bins. The eccentricity bin corresponding to the M-ring (2.84°) is not shown. The pale green line indicates the population *SP*s with the discriminability weight. ***F,*** Comparison of population and individual $CP_{stimulus=M}$ values for the RRs. Blue open circles indicate the *CP* values from the 10 foveal bins, which were adjusted for preference by (1 - *CP*). Red open circles indicate the *CP* values from the 10 peripheral bins. The pale orange line indicates the population *CP*s with the log-likelihood-ratio weight.

proportional to stimulus discriminability of given cortical sites (Fig. 7B), which are similar to the profile of model-predicted differential responses at the time frame with hemodynamic peak. The other scheme determined read-out weights by evaluating the contributions of given cortical sites to probabilistic inference of differences between the S-ring and L-ring stimuli (Fig. 7C), estimated by log-likelihood-ratios between tuning responses to

those two stimuli (see Section 6.10 for detailed definitions of the three weight profiles).

All three decoding schemes resulted in similar outcomes, each revealing clear-cut signatures of stimulus and choice in the RRs at the time points where the significant *SP*s and *CP*s were found for the TRs (Fig. 7D). The population *SP*s were significant at 3.3 s and 5.5 s after stimulus onset (green-filled markers in Fig. 7D; uncorrected $p<0.005$). These signatures were seen using all the three read-out schemes, but they were most conspicuous in the 'discriminability' read-out (green-filled triangles in Fig. 7D). In contrast, the population *CP*s were significant only 1.1 s after stimulus onset and were strongest when derived using the 'log-likelihood-ratio' scheme (yellow-filled squares in Fig. 7D; uncorrected $p<0.005$). This unmistakable dissociation between the population *SP*s and *CP*s estimated in the RRs, both in time and in profile shape, neatly dovetails with the results from the local *SP*s and *CP*s estimated in the TRs, further corroborating our conclusion that V1 carries stimulus and choice signatures that are embodied in different neural ensembles at different points in time.

The advantage of the population coding strategy over the local coding strategy was substantial in RRs: the best population probabilities (filled markers in Figs. 7E and 7F) surpassed all of the individual

probabilities estimated at the local cells of the trial-related matrix of the RRs (open circles in Figs. 7E and 7F; for the 10 bins located in the foveal bank, their individual probabilities were adjusted for preference by taking 1-*SP* or 1-*CP*, so that they can be directly compared to the population *SP*s and *CP*s). This analysis verifies that the RRs, despite including a substantial untuned component, retain sufficient information for supporting perceptual judgments at a subsequent decision stage.

Given the advantageous effect of population coding in the RR signals (see Figs. 7E and F), why does population coding not do better when applied to TRs (Figs. 8A and B)? Why, in other words, are population *SP*s and population *CP*s no larger than the best *SP*s and *CP*s exhibited by local cells of the trial-related matrix? One obvious possibility is that the beneficial effect of pooling signals from neurons with similar preferences is limited when those neurons' responses are highly correlated across trials (Averbeck et al., 2006). To check that possibility in the case of our TRs, we calculated pairwise temporal correlations among those TRs and found that the responses from nearby eccentricity bins are indeed highly correlated even after removal of global fluctuations (Fig. 8C). These correlations might reflect moment-to-moment co-fluctuations among neurons with similar stimulus preferences, but they might have arisen owing to our method for

combining voxel signals (Fig. 3B) and/or owing to the spatially correlated nature of the fMRI signal.

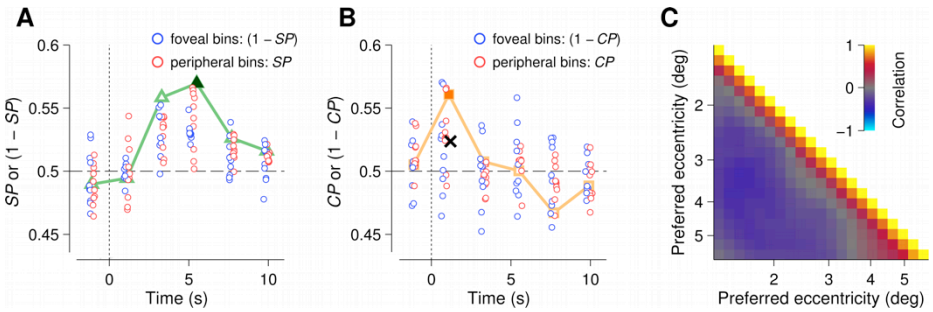

Figure 8. Population versus individual probabilities in tuned responses.

A*,** Comparison of population and individual *SP* values for the TRs. The format is identical to that in Fig. ***7E. The pale green line indicates the population *SP*s with the discriminability weight applied on the TRs. Note that this line is identical to the line in Fig. ***7E*** because those weights have removed the global fluctuations that distinguish TRs from RRs. ***B*,** Comparison of population and individual $CP_{\text{stimulus=M}}$ values for the TRs. The format is identical to that in Fig. ***7F*,** and again the population values are identical to those in Fig. ***7F*** for the same reason mentioned above. The black X denotes the average (0.52) of the individual *CP* values at 1.1 s after stimulus onset. ***C*,** A matrix of across-observer average correlations in TRs between the eccentricity bins.

3. Experiment 2: Pupil size dynamics during fixation impact the accuracy and precision of video-based gaze estimation

Our primary goal was to examine the 'causal' hypothesis regarding the role of V1 activity in trial-to-trial variability of perceptual choice. Although our results are inconsistent with the causal hypothesis, we were puzzled about why the choice-related cortical activity, which does not match stimulus-related responses either in timing or in neural origin, appeared in V1. One possible explanation for that puzzle attributes the seemingly errant activity to tiny eye movements, which are known to affect V1 neural activity (Gur et al., 1997; Martinez-Conde et al., 2000; Snodderly et al., 2001; Tse et al., 2010). So, we decided to conduct an eye-tracking experiment to test the eye movement hypothesis using a video-based eye-tracker. Knowing that the video-based eye-tracking technique can generate spurious gaze position shifts up to several degrees in visual angle (Wyatt, 2010; Kimmel et al., 2012), we first assessed the accuracy and reliability (precision) of our eye-tracking setup (Fig. 9; see Section 6.11 for details) before testing the eye movement hypothesis.

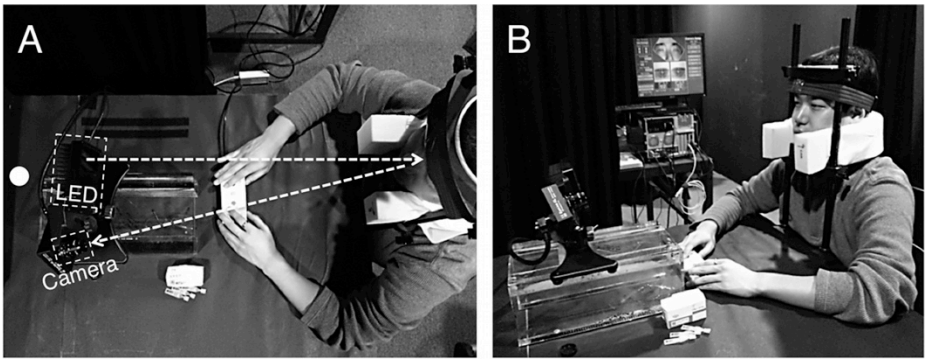

Figure 9. Eye-tracking experimental setup with an observer.

A, Locations of the LED illuminator and camera relative to the head and the display. The white dot demarcates the center of the screen. The dashed line arrows indicate the mean trajectory of the LED light projected onto and reflected from the eyes. ***B,*** Setup for minimizing body and head motion. The setup included height-adjustable chair and table, a forehead and chin rest, a memory-foam cushion around the neck, and a buckled cotton strap over the forehead (see Section 6.11 for details).

3.1. Spurious eye movement signals in video-based gaze estimation

Video-based eye trackers estimate gaze positions by inferring the center of the pupil from sampled video images of the eye (Merchant et al., 1974; Young and Sheena, 1975). Because they are noninvasive, easy to use, and robust, particularly compared to the alternative method relying on magnetic search coils (Robinson, 1963; Collewijn et al., 1975), video-based eye trackers are widely used for monitoring eye movements and for enforcing strict fixation in behavioral and neuroimaging experiments on humans. Despite its merits and popularity, however, this technique has a potentially serious drawback: it can generate spurious eye movement signals up to

several degrees in visual angle (Wyatt, 2010; Ivanov and Blanche, 2011; Drewes et al., 2012; Kimmel et al., 2012) mainly because pupil centration changes as pupil size changes (Walsh, 1988; Wilson et al., 1992; Charlier et al., 1994; Wyatt, 1995; Yang et al., 2002; Wildenmann and Schaeffel, 2013). Although this drawback was recognized by the inventors themselves when they first described the video-based eye tracking method (p.314 of Merchant et al., 1974), relatively little attention was paid to the problem until very recently, when it was highlighted in a series of papers by Wyatt (1995, 2010).

Recent studies have characterized the basic relationship between pupil size and gaze position measurements by explicitly evoking changes in pupil size by variations in light intensity (Wyatt, 2010; Ivanov and Blanche, 2011; Drewes et al., 2012; Kimmel et al., 2012), a reasonable strategy since the pupillary reflex is highly predictable with minimal variation across individual observers. But these studies capture only part of the problem arising from pupil size changes, for modulations in pupil size can also arise from endogenous factors, including arousal (Hess and Polt, 1960; Bradshaw, 1967; Henson and Emuh, 2010) and task-related cognitive demands (Hess and Polt, 1964; Kahneman and Beatty, 1966; Nassar et al., 2012; de Gee et al., 2014), that are bound to occur in studies using even simple tasks. Thus, it is important to learn the relationship between pupil size and gaze position

measurements under situations where endogenous factors may be influencing pupil size dynamics and, hence, measurements of gaze control. This motivated the current study, which aims to characterize endogenously driven changes in pupil size and to examine the relationship of those changes with video-based gaze position measurements from a relatively large sample of observers while they performed two different tasks each with its own unique demands.

3.2. Pupil size variation and gaze position estimates during fixation

In Experiment 2, we assessed the impact of variations in pupil size on the quality of video-based gaze position signals during sustained fixation. Observers were instructed to maintain strict fixation on a 0.12° diameter fixation target (FT; denoted schematically by the thick black bar on the right of Fig. 10A) that either remained in the center of the monitor screen (0°) for 16 s (‘prolonged-fixation (PR)’ condition; labeled ‘PR’ in Fig. 10A) or jumped unpredictably between two lateral (±0.12°) positions during a 31 s period of time (‘short-lived fixation (SL)’ condition; labeled ‘SL’ in Fig. 10A); throughout the entire 251 s experimental run, pupil area (PA) and gaze positions (GP; see Section 6.13 for their definitions and preprocessing) were

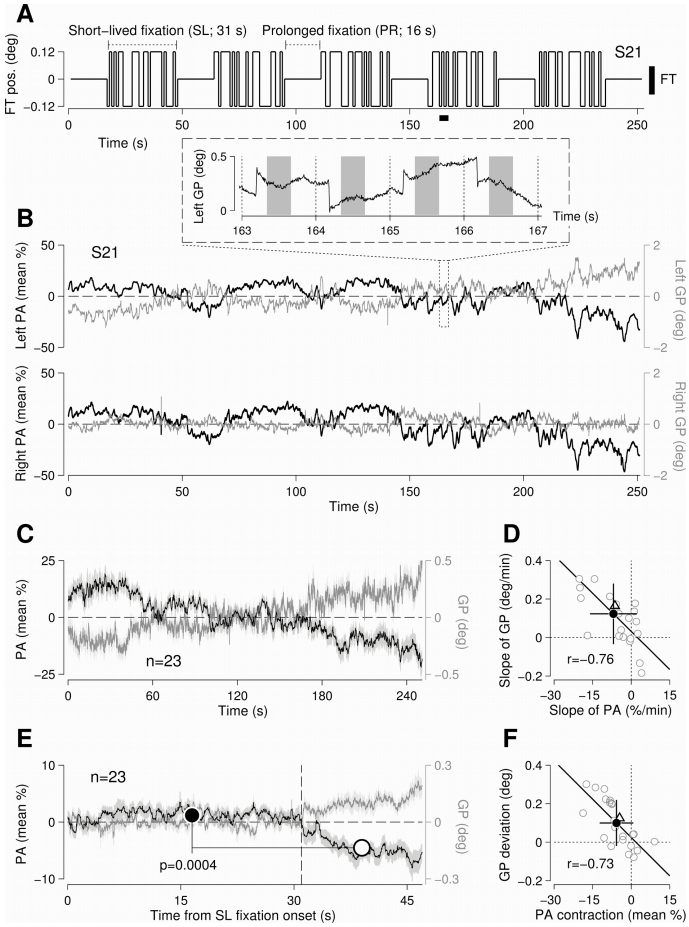

Figure 10. Changes in pupil size and gaze position during visually guided saccades.

A, Time-course of fixation target location. The small FT (thick vertical bar in the right; diameter 0.12°) stayed at the center during the prolonged (PR) fixation period and alternated between the two slightly de-centered positions during the short-lived (SL) fixation period. ***B,*** Pupil area (PA, black curves) and gaze position (GP, gray curves) measurements during an experimental run from the left and right eyes of a representative observer, S21. The time-course of GPs around 165 s (dotted box) are magnified at the top, where the shaded regions indicate the middle, one-third portions of the 1 s intervals that were used in down-sampling (see Section 6.15 for details). ***C,*** PA (black curve) and GP (gray curve) time-courses during the entire run, averaged across observers. The shaded curves represent ±SEM across observers. ***D,*** Inter-observer comparison of the rates of change in PA and GP for an entire run. The gray open circles represent individual observers, the black open triangle indicates the representative observer shown in ***B*** and the black filled circle with error bars are the population average with SDs. ***E,*** PA (black curve) and GP (gray curve) time-courses during short-lived (SL; 0-31 s) and prolonged (PR; 31-47 s) fixation, averaged across observers. The legends for the symbols are identical to those in ***C***. The large filled and open circles are the mean PA of SL and PR fixations, respectively. ***F,*** Comparison of the difference of PAs and GPs between SL and PR conditions. See ***D*** for format.

measured binocularly (black and gray curves, respectively, in Fig. 10B) at 500 Hz. Unlike previous studies, where the pupil size was forced to vary by manipulating the intensity of light (Wyatt, 2010; Drewes et al., 2012; Kimmel et al., 2012), we performed the experiment under constant luminance so as to monitor natural pupil behavior occurring while human observers simply maintain fixation. We specifically wondered whether pupil size would vary over time differently in the PR and the SL conditions because those two conditions differ in the level of arousal or vigilance that are known to endogenously affect pupil dilation (Hess and Polt, 1960; Bradshaw, 1967; Henson and Emuh, 2010). We also wanted to learn whether pupillary responses to those endogenous factors, unlike responses to changes in light intensity, might differ substantially across individuals. This was indeed the case, thereby allowing us to capitalize on these individual differences in pupil behavior for our analysis of related variations in inferred gaze position.

Unlike natural viewing conditions, wherein about three saccades occur every second typically toward visually salient features (Ibbotson and Krekelberg, 2011), observers in our experiment had to gaze continuously on the same, small FT appearing within a limited (0.24°) region of the visual field for an extended period of time (251 s). One could imagine that this task

would bore observers, thereby decreasing their general level of arousal and, concomitantly, constricting their pupils as time passes within an experimental run. Indeed, PA decreased steadily and substantially (black curve in Fig. 10C) with time. To quantify the rate of reduction in PA over time, we calculated for each observer the linear trend slope of the PA time-course during the 251 s run (which comprised multiple PR and SL epochs). The slopes of the regression lines (horizontal axis in Fig. 10D; -6.9 ± 9.1 %/min, mean ± SD across observers) deviated significantly from zero in the negative direction (t test across observers, $p=0.002$). Likewise, the PA significantly decreased during the periods of prolonged fixation, during which the FT stayed at the center (black curve in Fig. 10E). During the SL condition (0-31 s from SL fixation onset), compared to the PR condition (31-47 s), observers were more likely to be alert because the FT shifted position unpredictably. Conforming to this reasoning, the pupil was more enlarged (paired t test across observers, $p=0.0004$) during the SL condition (filled circle in Fig. 10E) than during the PR condition (open circle in Fig. 10E). In summary, the pupils dilated when saccades had to be made frequently, and contracted steadily as fixation was prolonged at both small (zoomed in, <16 s) and large (zoomed out, ~4 min) time scales.

As reported by previous studies (Merchant et al., 1974; Wyatt, 2010;

Drewes et al., 2012; Kimmel et al., 2012), these pupillary responses were accompanied by systematic errors in gaze estimation. In our study, we had no pupil-independent reference signals to compare directly against video-based gaze measurements (e.g., gaze measurements from scleral search coils). But we did see clear evidence for this pupil size artifact in the inter-observer correlation between PA and GP, which was significantly high at both large and small time scales. At the large time scale, the rates of changes in PA and in GP over the entire (251 s) run were significantly correlated across observers (Fig. 10D; Pearson's $r=-0.76$, $p<0.0001$): observers exhibiting larger pupillary contraction tended to show greater shifts in gaze position signal. At the small time scale, the averaged amounts of pupil contraction during the PR periods (16 s), which were computed against the average pupil size during the SL periods (Fig. 10E), were also significantly correlated with the amounts of deviation in GPs during the PR periods (Fig. 10F; Pearson's $r=-0.73$, $p=0.0001$). Note that the amounts of gaze shift associated with unit pupil size change were similar between the large (-0.0178 °/%; Fig. 10D) and small (-0.0175 °/%; Fig. 10F) time scales, indicating a scale-invariant relationship between pupil size and gaze position measurements.

3.3. Correction of gaze position measurements for pupil size artifact

Having confirmed the inter-observer correlation between PA and GP, we further inspected the relationship between the two measurements within single eyes of each of the observers, in an attempt to correct GP measurements for the errors confounded with pupil size.

Consistent with Wyatt (2010), the degree of PA-GP correlation varied greatly across observers and between eyes in a given individual as well. Consider, for example, the two observers S21 and S01, who showed rather different results (Fig. 11A). In S21, the 1 Hz down-sampled values of PA and GP (dsPA and dsGP, respectively; see Section 6.15 and Fig. 10B for the down-sampling procedure) showed high negative correlations in both of the eyes, but substantially higher in the left eye (Pearson's r=-0.86; upper left panel in Fig. 11A) than in the right eye (Pearson's r=-0.38; upper right panel in Fig. 11A). In contrast, S01 showed small degrees of PA-GP correlation in both of the eyes (Pearson's r=0.11 and -0.17 for the left and right eye, respectively; bottom panels in Fig. 11A). When inspected across all observers, the PA-GP correlation values (Fig. 11B) ranged from -0.96 to 0.11 in the left eye (-0.58 ± 0.32; mean ± SD across observers) and from -0.92 to 0.71 in the right eye (-0.27 ± 0.39), indicating large degrees of

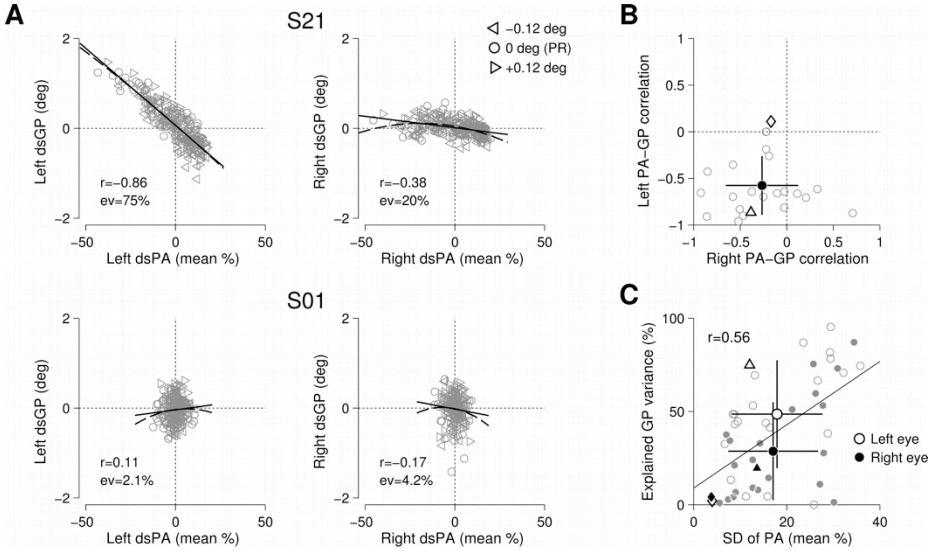

Figure 11. Inter-observer and between-eye differences in pupil size artifact.

A, Relationship between the down-sampled (1 Hz) PAs and GPs in two representative observers. The shape of the symbols indicates the position of the FT. The solid line and dashed curves are the best-fitting first and second order polynomials, respectively. ***B,*** Comparison of PA-GP correlations between the eyes. The gray open circles represent individual observers, and the triangle and diamond represent S21 and S01, respectively. The black filled circle with error bars is the population average with SDs. ***C,*** Comparison of PA variance (SD) and the variance of GPs explained by the second order polynomial of PA. Open and filled symbols indicate the left and right eye, respectively, of individual observers, and the triangles and diamonds represent S21 and S01, respectively. The large symbols with error bars represent the population average with SDs.

idiosyncrasy for both of the eyes. In addition, the PA-GP correlation values seemed independent between the eyes, as evidenced by the low, statistically non-significant correlation between-eyes (Pearson's r=0.02, p=0.92). Overall, the PA and GP were more strongly correlated in the left eye than in the right eye (paired t test across observers, p=0.006).

Given the eye difference and large idiosyncrasy in the PA-GP relationship, the correction for the pupil size artifact was done separately for

the two eyes and for individual observers by regressing out PA-associated errors from the GP measurements (Section 6.14). Because there were non-negligible components in the PA-GP relationship that cannot be captured by a simple linear regression (solid lines in Fig. 11A), the second order polynomials (dashed curves in Fig. 11A) were used to explain the variations in GP by the PA regressor, as done by Wyatt (2010). The second order regressor accounted for 48.6 ± 28.9% and 28.7 ± 26.2% of the total variance of the left- and right-eye GPs (large open and filled circles, respectively, in Fig. 11C). In line with the results from the linear correlation analysis, the amount of the variance of GP that could be explained by the second order regressor of PA varied greatly across observers, ranging from 0.02% to 95.4% in the left eye and from 1.0% to 87.1% in the right eye, in a manner dependent on the variance of PA (Pearson's r=0.56, p=0.0001; across observers, for both eyes). Note, however, that despite the large inter-ocular difference in the explained variance of GP by the PA regressor (ANOVA p=0.02, $F_{1,44}$=5.97), the variance in PA itself did not differ significantly between the eyes (ANOVA p=0.77, $F_{1,44}$=0.09). This indicates that the inter-ocular difference in PA-GP relationship could not be attributed to the inter-ocular difference in PA variance. The potential origins of this inter-ocular difference in PA-GP relationship will be discussed in Section 5.4.

The extraction of the GP component associated with PA significantly enhanced both the accuracy and the precision of the remaining GP signal. The signal enhancement was pronounced in the observers with high correlations between GP and PA. In the observer S21, for example, the GPs appeared to drift rightward gradually over time by a large amount (0.16 °/min) before correction (gray curve in the left panel of Fig. 12A), resulting in a large deviation (0.13°) from the FT position during the PR fixation period (gray circle at the center in the Accuracy panel of Fig. 12A; recall Fig. 10E-F) and large standard deviations during both the PR and SL fixation period (gray bars in the Precision panel of Fig. 12A). This gradual drift largely disappeared after correction (black curve in the left panel of Fig. 12A). Consequently, the correction improved the accuracy of GPs during the PR period (dsGP deviation=0.02°) without compromising the GPs during the SL periods (black circles in the Accuracy panel of Fig. 12A) and enhanced the precision of GPs nonspecifically in both the PR and SL periods (black bars in the Precision panel of Fig. 12A). In addition, the correction did not hamper the quality of GP measurements in terms of either accuracy or precision for the individuals who showed weak PA-GP correlations (e.g. S01; Fig. 12B).

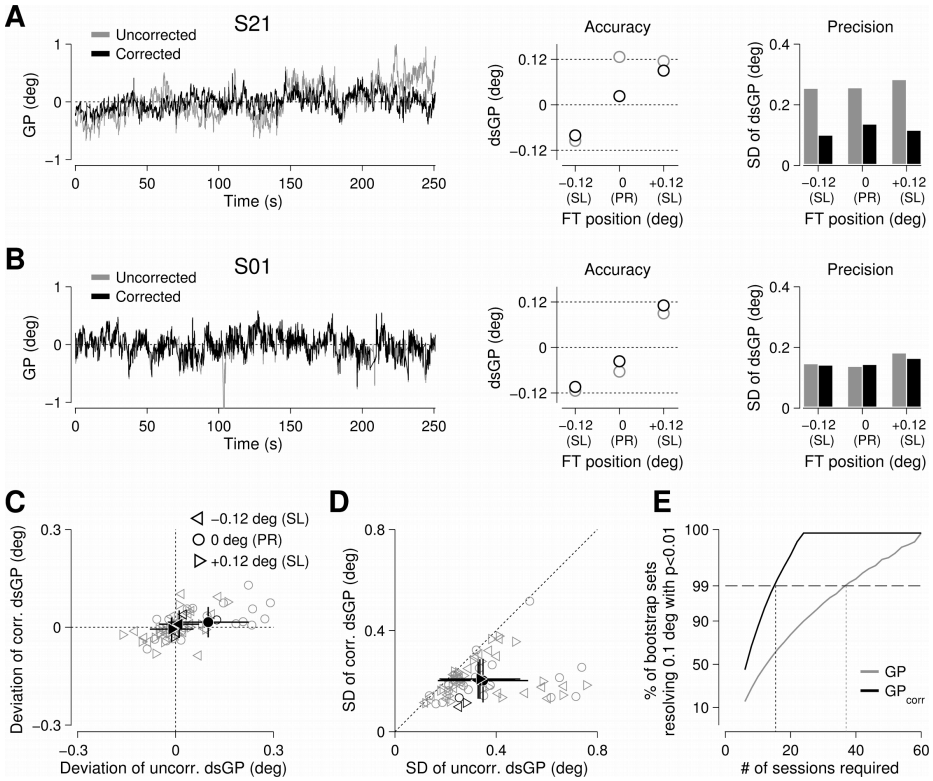

Figure 12. Effects of correction for pupil-confounded errors in gaze position.

A, Effects of correction in observer S21. The GP time-series before (gray) and after (black curves) correction are compared in the left panel. The uncorrected and corrected means of dsGPs are plotted against each other, separately for the one prolonged FT and the two short-lived FTs in the Accuracy (middle) panel. In the Precision (right) panel, the gray and black bars are the SDs of the uncorrected and corrected dsGPs, respectively. ***B,*** Effects of correction in observer S01. The format and legends are identical to those in ***A***. ***C,*** Population summary of accuracy enhancement by correction. The gray open symbols represent individual observers, and their shapes represent the positions of the FT. The black open symbols indicate S21. The large filled circles with error bars are the population average with SDs. ***D,*** Population summary of precision enhancement by correction. The format and legends are identical to those in ***C***. ***E,*** Effects of correction on statistical power of fixation experiments. The number of experimental sessions required for significant differentiation of 0.1° difference in FT position was estimated by running simulations based on the SD of the uncorrected (gray) and corrected (black) GPs (see Section 6.16 for detailed procedure).

The effects of the correction procedure, as exemplified in these two observers, are summarized in Fig. 12C-D. When pooled across observers, the correction improved the accuracy of the GP measurements by decreasing the

deviations in the PR condition (circles in Fig. 12C) from 0.10° ± 0.12° (horizontal axis; mean ± SD across observers) to 0.02° ± 0.05° (vertical axis) and by keeping the deviations in the SL condition (triangles) small, where the uncorrected GPs were already quite accurate before correction (dsGP deviation=0.01° ± 0.06° and -0.01° ± 0.07° for the -0.12° and +0.12° FTs, respectively). On the other hand, the improvement in precision was ubiquitous: it was observed in all of the three FT positions from almost all of the twenty-three observers (Fig. 12D). The SDs of dsGPs at the left (-0.12°), middle (0°) and right (+0.12°) FTs were 0.33° ± 0.15° (mean ± SD across observers), 0.35° ± 0.18°, and 0.34° ± 0.16°, respectively, before correction and decreased to 0.20° ± 0.09°, 0.20° ± 0.07°, and 0.21° ± 0.08°, respectively, after correction. When pooled across the three FT positions, the correction reduced the SDs of GPs by 31.6 ± 24.1% (mean ± SD across observers). To quantify the benefits of this reduction in SDs in practical situations, we computed how many sessions (or observers) would be required to achieve the statistical power of resolving the 0.1° difference in gaze position. The computer simulations based on the pooled SD values of the GPs before and after correction (see Section 6.16 for details) indicated that the corrected GPs can resolve the 0.1° difference in FT position using only the half of the number of sessions required for the uncorrected GPs (Fig. 12E).

It is worth mentioning that the correction procedure not only improved the accuracy and precision of the GP measurements but also diminished the inter-observer and between-condition differences in both accuracy and precision, as evidenced by the reduced standard error values across observers (see the lengths of the horizontal and vertical error bars in Fig. 12C,D) and by the reduced differences in the mean values among the three FT conditions (see the distances among the filled symbols along the horizontal and vertical axes in Fig. 12C,D). We also checked whether the PA-GP relationship differed between during pupillary constriction and during pupillary dilation, for such behavior has been reported in conditions where pupillary responses are evoked by large changes in stimulus luminance (Wyatt, 2010). In our measurements, which did not involve stimulus-evoked changes in pupil size, the differences in PA-GP relationship between during constriction and during dilation were too weak to contribute anything to the correction procedure.

In summary, when corrected by the second order regressor of PA separately for each eye of individual observers, the accuracy and precision of the GP measurements were enhanced and became similar between the fixation conditions and across the observers.

3.4. Comparison between pupil center estimation methods

To learn whether the method used to estimate pupil center affects our conclusions about pupil artifacts, we conducted an auxiliary experiment directly comparing the 'centroid' and 'ellipse' mode of pupil tracking. Using the same experimental setup (Section 6.11), we collected eye-tracking data from 6 naïve observers (4 females, 2 males; aged 19-31 years; one of whom participated in Experiment 2). We administered two runs of the visually guided saccade task (Section 6.12) for each mode, the order of which was counter-balanced across observers; thus in three observers, the centroid mode was used to collect the first two runs of data followed by the ellipse mode to collect the next two runs, and vice versa for the other three observers. Eye tracker calibration was done separately for each mode. This resulted in 24 runs of data in total.

We analyzed the data in the same manner as described in the main text and found that the selection of pupil center estimation method does not affect the PA-GP relationship and thus the resulting correction. The PA-GP correlation values (as in Fig. 11B) with the centroid mode were -0.67 ± 0.21 (mean ± SD across 12 runs) in the left eye and 0.12 ± 0.38 in the right eye, and those with the ellipse mode were -0.64 ± 0.14 in the left eye and -0.05 ± 0.38 in the right eye (Fig. 13A). Also, the second order regressor (as in Fig.

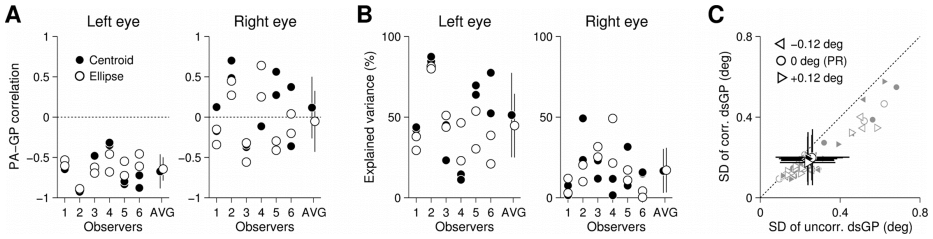

Figure 13. Comparison between pupil center estimation methods.

11C) accounted for 51.2 ± 26.2% (mean ± SD across 12 runs) and 16.6 ± 13.6% of the total variance of the left- and right-eye GPs measured using the centroid mode and accounted for 44.7 ± 19.7% and 17.1 ± 13.8% of those measured using the ellipse mode (Fig. 13B). As a result of correction, the SDs of dsGPs at the left (-0.12°), middle (0°), and right (+0.12°) FTs (as in Fig. 12D) from the centroid mode were 0.24° ± 0.13° (mean ± SD across 12 runs), 0.26° ± 0.18°, and 0.24° ± 0.15°, respectively before correction and decreased to 0.19° ± 0.11°, 0.20° ± 0.14°, and 0.19° ± 0.13°, respectively after correction (Fig. 13C). Those from the ellipse mode were 0.24° ± 0.13° (mean ± SD across 12 runs), 0.26° ± 0.16°, and 0.24° ± 0.14°, respectively before correction and decreased to 0.19° ± 0.10°, 0.20° ± 0.11°, and 0.17° ± 0.08°, respectively after correction. When pooled across the three FT positions, the correction reduced the SDs of dsGPs by 18.8 ± 12.9% (mean ± SD across 12 runs) and 21.0 ± 10.5% for the centroid and ellipse mode, respectively.

There were two notable differences, however, between the previous experimental results (Section 3.3) and these new data. First, the PA-GP correlation values in the right eye changed from -0.27 ± 0.39 (mean ± SD across 23 observers) to 0.03 ± 0.38 (mean ± SD across 24 runs from 6 observers). Second, the reduction of the SDs of dsGP (index of precision) by correction was changed from 31.6 ± 24.1% (mean ± SD across 23 observers) to 19.9 ± 11.6% (mean ± SD across 24 runs from 6 observers). We believe this is mainly due to the large degree of individual differences in the PA-GP relationship and the smaller number of observers used in the auxiliary experiment, further justifying the recommendation that the correction procedure should be applied on an individual-by-individual and an eye-by-eye basis.

4. Experiment 3: Choice signatures in V1 cannot be attributed to fixational eye movements

Although the observers in Experiment 1 were explicitly instructed to maintain their gaze on the fixation mark throughout an entire scan run, their eyes may well have moved unintentionally (Ratliff and Riggs, 1950; Martinez-Conde et al., 2004). And it is known that tiny eye movements such as drifts, tremors and microsaccades can affect V1 neural activity (Gur et al., 1997; Martinez-Conde et al., 2000; Snodderly et al., 2001; Tse et al., 2010). Thus we first considered these involuntary fixational ("miniature") eye movements as a possible origin for the observed *CP*s in V1. Given the advanced temporal locus of the significant *CP*s (0~2.2 s after the onset of the ring stimulus; *CP* and $CP_{\text{stimulus=M}}$ panels in Fig. 6A) and the hemodynamic delay of fMRI signal in the current study (4.4~6.6 s), which can be estimated from the locus of the significant *SP*s (*SP* panel in Fig. 6A), we were particularly interested in whether there were any differential eye movements associated with perceptual choices at the temporal bin spanning 4.4~2.2 s before the ring stimulus onset, when no stimulus was presented other than the small fixation dot. This absence of a ring stimulus at the moment of the *CP*s in our study makes

fixational eye movements unlikely to be the cause of the *CP*s because microsaccades, which occur with the greatest amplitude among the major fixational eye movement types, cause no changes in V1 neural activity in the absence of a stimulus other than a fixation mark (Martinez-Conde et al., 2000, 2002). However, because the impact of various fixational eye movements, including microsaccades, on fMRI measurements in V1 in the absence of stimulation has never been measured directly and because fixational eye movements might alter perception possibly via attention (Hafed et al., 2011) or gain modulation (Hafed, 2013), we explored the possibility that fixational eye movements, which were not monitored during the fMRI experiment, might have generated the choice signature in V1. This possibility can be tested behaviorally because one necessary (but not sufficient) condition for eye movements to generate the *CP*s observed in Experiment 1 is that eye movements occurring around 4.4~2.2 s before stimulus onset should predict observers' choices yet to be made after stimulus presentation. Thus, we searched for eye movements that satisfy this necessary condition by repeating the same experiment on a new batch of observers outside the scanner, but now with their eye movements being tracked throughout the entire experiment (Experiment 3). The stimuli, procedure, and number of observers (23) in this eye-tracking experiment were otherwise identical to those in the original

fMRI experiment (see Section 6.12 for details). As expected, the *SC* and resulting performance of this new batch of observers (0.023 ± 0.006 and 81.6% ± 5.1%, respectively; mean ± SD across observers) were comparable to those of who participated in Experiment 1. Unexpectedly, however, we learned from Experiment 2 that the pupil-confounded errors in gaze position signal were pronounced, particularly when observers were required to maintain prolonged fixation at the same spatial location. So, we investigated the gaze position signal from Experiment 3 for those errors and corrected them before searching for the choice signatures from the signal.

4.1. Correction of gaze position measurements during the fine ring-size classification task for pupil size artifact

In Experiment 3, observers not only tried to maintain fixation at a single position, as they did during the PR period in Experiment 2, but also made difficult visual decisions based on briefly flashed stimuli each requiring an immediate response throughout an entire run of 27 trials interspersed over 356.4 s (recall Section 2.1). On each trial observers viewed a ring stimulus that appeared briefly (300 ms) around the central FT and classified its size into ‘large’ or ‘small’ by making a button press within 1.5 s from the stimulus onset (Fig. 14A). The observers participated in this discrimination

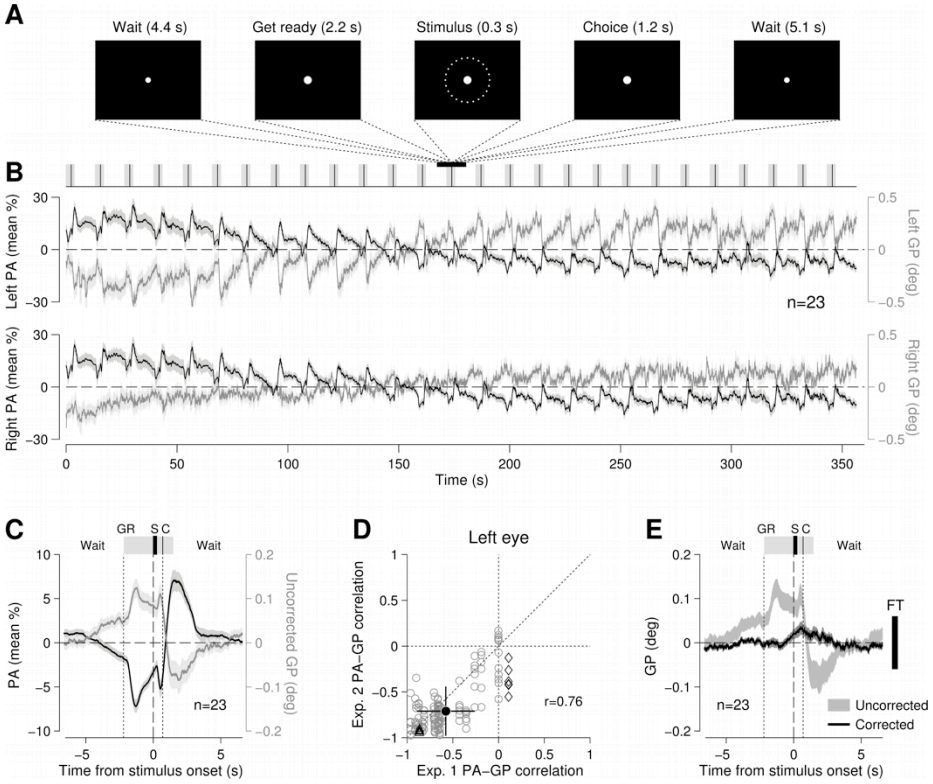

Figure 14. Relationship between pupil size and gaze position measurements during visual discrimination.

A, Fine ring-size discrimination task. The top panel shows the five events comprising a single trial, where the size of the FT is exaggerated for illustration purposes. The bottom panel shows the trial series for an entire run. The horizontal black bar represents a 13.2 s trial. The shaded rectangles demarcate the successive periods of 'Get ready' to 'Choice', during which the FT are enlarged. The vertical black lines indicate the time points of stimulus onset. ***B,*** Across-observer averaged PA (black curve) and GP (gray curve) time-courses over an entire run. The shaded curves represent ±SEM across observers. ***C,*** Across-trial averaged time-course of PAs and GPs with SEM (shaded area). The gray rectangle and vertical bars on the top represent the events described in ***A***. ***D,*** Comparison of PA-GP correlations between Experiment 1 and 2. For each observer, one correlation value in Experiment 1 is plotted against six correlation values in Experiment 2, due to the difference in number of experimental runs. See Fig. **11B** for format. Only the values from the left eye are shown here. ***E,*** Comparison of across-trial averaged time-courses of uncorrrected (SEM in shade only) and corrected GPs. The thick vertical bar on the right indicates the diameter of FT.

immediately following Experiment 2, and their two eyes were monitored in the same manner (Fig. 14B) throughout six experimental runs.

As in Experiment 2, the size of the pupil gradually decreased over the large time scale comprising an entire block of trials (black curves in Fig. 14B). The rate of reduction in PA over time (-4.4 ± 6.6 %/min, mean ± SD across observers and runs) was significantly smaller than zero (t test across observers and runs, $p<0.0001$). In contrast, the trial-locked time-course of pupil size at a small time scale (black curve in Fig. 14C) waxed and waned during the relatively short duration of a trial and contracted slowly between trials.

Although the changes in pupil size were more complex than those in Experiment 2, the relationship between PA and GP was quite similar to that found in Experiment 2 at both the small (13.2 s) and large (~6 min) time scales. At the large time scale, there was a significant inter-observer correlation in the rate of changes during the entire run between PA and GP (Pearson's $r=-0.83$, $p<0.0001$; recall Fig. 10D). Also, at the small time scale, the averaged trial-locked time-course of PA was almost the inverse of that of GP, mirroring all the major peaks and troughs of fluctuation (Fig. 14C). The similarity in PA-GP relationship between the two experiments (Experiments 2 and 3) could also be confirmed by the following direct comparisons. First, the PA-GP correlation values from Experiment 3 (Pearson's $r=-0.67 \pm 0.28$ and -0.30 ± 0.42 for the left and right eye, respectively; mean ± SD across

observers and runs) were comparable to those from Experiment 2 (ANOVA $p=0.15$ and 0.68, $F_{1,159}=2.12$ and 0.18; for the left and right eye, respectively). Second, the PA-GP correlation values were themselves significantly correlated over individuals *between* the two experiments (Fig. 14D; Pearson's $r=0.76$ and 0.57, for the left and right eye, respectively; $p<0.0001$ for both eyes), indicating that the PA-GP correlation values in the two experiments were consistent within given individuals. Lastly, as in Experiment 2, the PA-GP correlation was higher in the left eye than in the right eye (paired t test across observers and runs, $p<0.0001$).

Having confirmed the relationship between PA and GP, we applied the same correction procedure to the GP measurements, separately for different eyes, for individual observers, and for individual runs. The consequences of correction for the PA-confounded errors were replicated. The variances of GP that could be accounted for by the PA regressor (55.8 ± 28.3% and 33.1 ± 26.4%, for the left and right eye, respectively; mean ± SD across observers and runs) were comparable to those derived from the result of Experiment 2 (ANOVA $p=0.26$ and 0.46, $F_{1,159}=1.27$ and 0.54, for the left and right eye, respectively). The correction procedure removed the systematic deviations of the uncorrected GPs (Fig. 14E), the magnitude of which was larger than the size of FT (denoted schematically by the thick

black bar on the right), and enhanced the precision of the GP measurements as illustrated by the reduced variability across trials and runs (as indicated by reduced standard error regions of the black curve).

The observers in Experiment 3 were instructed to maintain their gaze always at the small fixation target, and thus were likely to generate involuntary fixational ("miniature") eye movements, i.e., microsaccades (see Martinez-Conde et al., 2004 for review), which may reflect cognitive processes such as shifts of attention (Hafed and Clark, 2002; Engbert and Kliegl, 2003; Hafed et al., 2011; Yuval-Greenberg et al., 2014) and changes in perception (Laubrock et al., 2008; Hsieh and Tse, 2009; Otero-Millan et al., 2012; Hafed, 2013). To ensure that the correction procedure does not interfere with detecting microsaccades, we compared the number of microsaccades in Experiment 3 between before and after correction. When pooled across all observers and runs, a total of 72,300 and 71,543 microsaccades were detected (Section 6.17) in the uncorrected and corrected GPs (~1% difference), respectively, indicating a negligible influence of the correction procedure on microsaccade detection. This is not surprising, given the large difference in time scale between the events comprising microsaccades (<50 ms) and pupillary responses (>1 s).

4.2. Lack of choice signature in fixational eye movements

After correcting the gaze position measurements for pupil size artifact, we looked for choice signatures in fixational eye movements.

First, we checked whether choices correlated with the frequency of eye blinks, which are known to affect fMRI activity in human V1 in the absence (Bristow et al., 2005) and in the presence (Tse et al., 2010) of retinal stimulation. Although the overall blink frequency (black line trace in the top panel of Fig. 15A) decreased around the time of stimulus onset (dashed vertical lines in Fig. 15A) and increased afterward, its time-course did not differ at any time bins between the two choices either for the M-ring trials only (bottom panel in Fig. 15A; Paired t-test across observers, uncorrected $0.11<p<0.99$) or for the entire set of trials ($0.11<p<0.65$). Next, we checked whether choices correlated with the frequency or direction of microsaccades because a microsaccade may affect the response gain of neurons representing visual regions around its target (Hamker et al., 2008; Hafed, 2013), or the direction of a microsaccade may interact with covert spatial attentional shifts (Hafed and Clark, 2002; Engbert and Kliegl, 2003; Hafed et al., 2011). Consistent with a phenomenon known as 'microsaccadic inhibition' (Rolfs et al., 2008; Hafed and Ignashchenkova, 2013), the overall frequency (dashed line in the top panel of Fig. 15C) decreased after onset of the ready signal

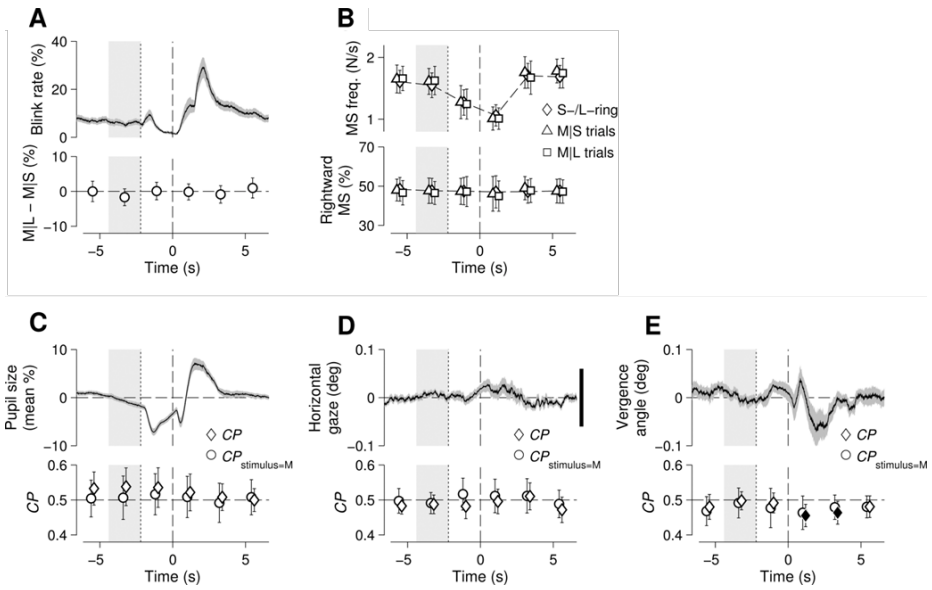

Figure 15. Eye movements during the ring-size discrimination task.

The mean and variability statistics for five different aspects of eye movements are plotted against the time points relative to the stimulus onset, which matched the time axis of the trial-related fMRI matrix after being shifted for hemodynamic delay. The dotted and dashed vertical lines indicate the onset of the 'ready' cue and ring stimulus, respectively. The shaded rectangular areas delineate the time bin at which the significant *CP*s were found in the fMRI experiment. ***A,*** Eye blinks. The top panel shows the across-observer average (black line) time-course of percent blink rate with SEM (shaded area). The bottom panel plots the differences in blink rate between the 'M-ring|S-choice' and 'M-ring|L-choice' trials, with error bars representing SD/2 across observers. ***B,*** Microsaccades. The top and bottom panels show the time-course of microsaccade frequency and the ratio of rightward microssacades, respectively, at each 2.2 s bin. Different symbols represent the trial types. The dashed curve is the average of the S- and L-ring trials and serves as a baseline. Error bars, SD/2 across observers. ***C,*** Changes in pupil size. The top panel shows the time-course of pupil size changes around its mean. The bottom panel shows the *CP*s computed for the entire trials (diamonds) and for M-ring trials (circles). Error bars, SD/2 across observers. ***D,*** Gaze positions. The format is identical to that in ***C***. The vertical black bar represents the width of the fixation mark. ***E,*** Changes in vergence angle. Filled symbols in the bottom panel mark the statistically significant *CP*s (uncorrected $p<0.05$).

(dotted vertical lines in Fig. 15B) and recovered to the baseline level after the choice period. However, neither the frequency (top panel of Fig. 15B) nor the direction (bottom panel of Fig. 15B) differed at any time bins between the two groups of choice-sorted M-ring trials (Paired t-test across observers,

uncorrected $0.20<p<0.97$ and $0.21<p<0.96$, respectively).

Next, we checked whether choices correlated with pupil diameter because changes in pupil size are known to accompany changes in arousal (Hess and Polt, 1960; Bradshaw, 1967; Henson and Emuh, 2010), changes in perceptual interpretation (Einhauser et al., 2008; Einhauser et al., 2010), and task-related factors (Hess and Polt, 1964; Kahneman and Beatty, 1966; Nassar et al., 2012; de Gee et al., 2014). These kinds of uncontrolled changes in pupil size in turn would produce variations in the retinal image (Campbell and Gubisch, 1966) that could induce changes in V1 activity. As suggested by its kin relationship with task structure, the overall pupil size (black line trace in the top panel of Fig. 15C) fluctuated substantially around the stimulus onset in a manner similar to the blink rate (top panel of Fig. 15A) and the microsaccade frequency (top panel of Fig. 15B). However, the pupil size failed to predict the choices made by observers, as indicated by the absence of significant *CP* at any time bins (bottom panel of Fig. 15C), including the one spanning 4.4~2.2 s before stimulus onset, either for the M-ring trials (Student's t-test across observers, uncorrected $0.52<p<0.85$) or for the entire trials ($0.17<p<0.99$).

Next, we checked whether choices correlate with gaze position, which may affect V1 activity via gain modulation (Trotter and Celebrini,

1999; Rosenbluth and Allman, 2002; Sharma et al., 2003; Merriam et al., 2013). Unlike the previous measurements (blink rate, microsaccade frequency, and pupil size), the overall gaze position (black line trace in the top panel of Fig. 15D) remained virtually stationary, showing only negligible amounts of fluctuation inside the fixation mark (thick black bar on the right side of Fig. 15D). In agreement with the previous measurements, however, the gaze position did not carry choice information at all (bottom panel of Fig. 15D; Student's t-test across observers, uncorrected $0.13<p<0.80$ and $0.39<p<0.85$; for *CP* and $CP_{stimulus=M}$, respectively).

Lastly, we checked whether choices correlated with changes in vergence angle because vergence eye movements are known to affect the perceived size of an object (Mon-Williams et al., 1997; Sperandio et al., 2013), which in turn could affect V1 activity (Murray et al., 2006). The vergence angle (black line trace in the top panel of Fig. 15E) was not biased toward the either near or far side before and during the stimulus presentation. However, the significant correlations were found at the two consecutive time bins after stimulus onset (black diamonds in the bottom panel of Fig. 15E; *CP*= 0.46±0.06 and 0.46±0.07, respectively; mean ± SD across observers; Student's t-test, uncorrected p=0.01 and 0.04, respectively), but not at the time bin (4.4~2.2 s before stimulus onset) associated with the significant *CP*s

found in the V1 fMRI activity (p=0.63). This temporal mismatch disqualifies vergence angle as a potential origin for the choice signature in V1. Instead, the temporally-delayed changes in vergence angle are likely to reflect the vergence control system's automatic reaction to the perceived size of a ring stimulus. This interpretation is supported by the fact that divergence was greater after the ring stimulus was judged to be large than when it was judged to be small (vergence angle difference at 0~2.2 s= -0.05° ± 0.12°; mean ± SD across observers).

In summary, our eye movement measurements were accurate and reliable enough to reveal the previously-known subtle changes associated with task structure. However, none of the five aspects of eye movements were related to choice behavior. Hence there is no reason to attribute the choice-related V1 activity found in Experiment 1 to eye movements. In the following section (Section 5.1) we consider other possible reasons for that activity.

5. Discussions

5.1. The role of V1 in perceptual decision-making

In Experiment 1 human observers made speeded perceptual judgments about the size of ring stimuli while, at the same time, neural responses were being measured from their primary visual cortex by fMRI. Ring stimuli highly similar in size were briefly presented, creating a difficult discrimination task that induced variability in choice behavior across trials, thereby allowing us to identify when and where stimulus-correlated and choice-correlated responses arose relative to stimulus onset. Raw fMRI measurements failed to exhibit reliable neural signatures selective either for the particular stimulus presented or for the choice made by the observer. When the omnibus, untuned component was filtered from the raw responses, however, both 'stimulus' and 'choice' signatures were revealed within two separate constellations of activation, located far apart from one another in both space and time. The spatial and temporal loci of the 'stimulus' signature precisely matched those derived from the eccentricity-tuning curves and those predicted by hemodynamic delay of sensory responses, respectively. Compared to the 'stimulus' signature, the 'choice' signature was quite

advanced in time, appearing before onset of the stimulus, and located farther away from the stimulation site in space.

Besides our study, there is another human fMRI study, by Ress and Heeger (2003), that also found significant correlations between V1 fMRI responses and observers' perceptual choices. Unlike us, however, Ress and Heeger found those correlations within the same voxels as those associated with fMRI responses triggered by stimulus presentation. We can envision several possible considerations that reconcile their findings with ours. For one thing, we employed a difficult stimulus discrimination task that probably relies on information carried by neurons that are not maximally responsive to the presented stimulus. Ress and Heeger, on the other hand, employed a contrast detection task that relies on the overall magnitude of responses associated with the presence or absence of a weak stimulus, responses likely to arise in neurons maximally responsive to the stimulus. For another thing, we purposefully tailored our task (difficult size discrimination), stimuli (thin, highly localized rings) and fMRI protocol (sparse event-related) to achieve very high spatial resolution. Ress and Heeger used a detection task in concert with an fMRI protocol (dense event-related) that was not optimized to uncover possible dissociations in the spatial and temporal origins of SPs and CPs.

What might our results say about neural coding of ring size? The most reliable stimulus signals associated with ring size were not those appearing within voxels maximally responsive to a given-sized ring but, instead, were signals arising at locations to either side of the sites registering the precise peak response to the given-sized ring (*SP* panel in Fig. 6A). This finding comports with the idea that the information capacity of cortical neurons in primary sensory cortex is not always governed by their maximal responses to sensory stimuli. By definition, the preferred stimulus of a visual neuron is the one producing the strongest responses in that neuron (i.e., the peak of its tuning curve), but when it comes to discriminating subtle, near-threshold differences between stimuli (e.g., ring size) maximally-responsive neurons may not provide the optimal information for that discrimination (Regan and Beverley, 1985; Jazayeri and Movshon, 2006). This idea has been corroborated by results from single-cell recordings in V1 (Graf et al., 2011; Berens et al., 2012) and extra-striate cortex (Purushothaman and Bradley, 2005), and now we have evidence supporting this principle within human V1.

Turning to the choice-related activity revealed by our study, our results are noteworthy in two respects. First, significant *CP*s were not located at neural sites where significant *SP*s were arising, and second, those *CP*s

were appearing much earlier in time than were the *SP*s. The absence of significant *CP*s within the spatiotemporal window in which significant *SP*s were found is consistent with previous single-cell studies reporting the absence of significant *CP*s in the responses of neurons within primary sensory cortex that participate in encoding sensory inputs relevant to given tasks (Grunewald et al., 2002; de Lafuente and Romo, 2005; Nienborg and Cumming, 2006; Hernandez et al., 2010). And in the one neurophysiological study that did find significant choice-related responses in V1 (Palmer et al., 2007) those responses were not arising within the neurons maximally sensitive to the evoking stimulus, again consistent with what we observed in our fMRI results.

What do our results say about V1's role in choice behavior? We reasoned if sensory neurons in V1 played a causal role in determining perceptual judgments, trial-to-trial variability in choice behavior should be linked to the same neural activity that signifies differences in stimuli relevant to the task. A recent computational study provided a formal proof that this relationship should hold, if neuronal populations are read-out in a (sub-) optimal manner (see Eq. 6 and Fig. 6 of Haefner et al., 2013). Our observation of a clear-cut dissociation between the signatures of stimulus and choice undermines the 'causal' view as it pertains to V1. Of course, one can

never definitively rule out the possibility that fMRI measurements simply failed to detect choice-related responses from a small subset of neurons within the stimulus-related voxels. We are satisfied, however, that our measurement techniques are sufficiently sensitive to distinguish both the stimulus- and choice-related responses with good spatial and temporal resolution.

At this stage we are cautious about generalizing our finding of a dissociation between stimulus and choice signatures in V1 activity to other perceptual tasks such as simple detection (Ress and Heeger, 2003; Palmer et al., 2007), coarse discrimination (Britten et al., 1996). It will be interesting to look for dissociations between stimulus- and choice-associated population responses in other contexts and, for that matter, in extra-striate visual areas including V2 and V3, to learn whether the dissociation is unique to primary sensory cortex. The stimuli and task employed in our study were optimized to exploit the high spatial resolution of the V1 retinotopic map, so they may not work so well when targeting extra-striate visual areas, due to the relatively larger receptive field sizes (and, hence, poorer spatial resolution) of neurons in those higher tier areas. In all events, however, our findings make a unique contribution to the elucidation of the roles of V1 in perceptual decision-making by directly comparing choice-related responses and

stimulus-encoding responses at a level of population activity.

5.2. Potential origin of the choice signature in V1

The choice signature observed in our study originates 3~4 seconds *before* the brief appearance of the task-related stimulus. This is remarkable because the only stimulus present at that time is the small fixation dot seen against an otherwise dark background. So, what is responsible for producing this pre-stimulus choice signature? A few previous studies found a tendency for eye movements in the absence of visual stimulation to produce spiking (Kagan et al., 2008) or fMRI activity (Sylvester et al., 2005) in V1. Thus, we considered the possibility that eye movements made 3~4 seconds before the stimulus onset might be correlated with the choices made by the observers. This possibility was tested exhaustively by searching for any choice-related changes in the five different aspects of fixational eye movements: blink rate, frequency and direction of microsaccades, pupil size, gaze position, and vergence angle (see Section 4). However, none of those measurements were significantly correlated with the choices made by the observers (bottom panels in Figs. 15A-E). This lack of correlation between pre-stimulus eye movements and choices cannot be attributed to the inaccuracy nor unreliability of our measurements because the eye-tracking signals in the

current study were accurate and reliable enough to resolve as small a difference as 0.1° (see Section 3) and to exhibit previously known task-locked changes in overall microsaccades and pupil diameter (top panels in Figs. 15B-C).

Given this absence of the pre-stimulus, choice-related eye movements, we speculate about top-down expectation as an alternative candidate for the origin of the *CP*s found in V1. The choice signature's specific location in space and in time – remote from the retinotopic site activated by the stimuli and advanced in time relative to appearance of those stimuli – suggests that this signature could arise from pre-trial expectations ('short-term' priors, in other words) originating from high-tier areas involved in perceptual inference. Given the high degree of uncertainty imposed by the subtle ring-size differences and by time pressure to make judgments in our study, perhaps the brain builds up 'trial-to-trial priors' in advance of trial onset, with those priors based on non-sensory sources of information (e.g., frequencies of choices made in recent trials). The trial-to-trial priors may be implemented by feedback projections carrying neural signals anticipating the forthcoming visual stimulus. Those feedback signals (pale-red and pale-blue curves in Fig. 16A) are unlikely to be as precise as stimulus-evoked feedforward signals (dark-red and dark-blue curves in Fig. 16A), thus

shifting the cortical sites with the maximum differences away from the boundary (gray triangles in Fig. 16B). At the same time, the dissociated signatures of stimulus and choice suggest that the prior signals preceding in V1 does not interfere with V1 neural activity that encodes stimulus features, as would be expected if sensory-evoked activity quenches pre-existing neural variability (Churchland et al., 2010).

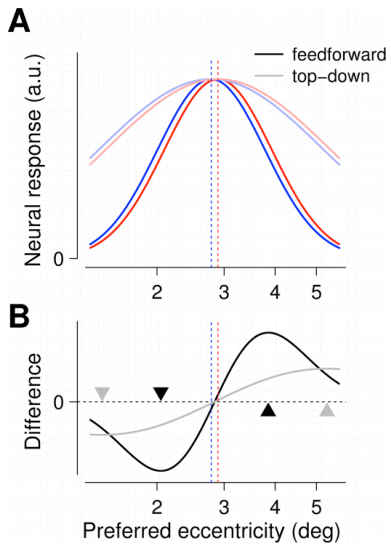

Figure 16. Conceptual implementation of trial-to-trial prior.

A, Hypothetical feedforward and top-down signals in V1. The horizontal axis specifies the preferred eccentricity. The dotted blue and red vertical lines indicate the two cortical sites that respond maximally to the S- and L-rings, respectively. The solid blue and red curves are population responses evoked by the S- and L-ring stimuli, respectively, representing feedforward signals. The pale-blue and pale-red curves are hypothetical population responses reflecting spatially blurred top-down signals. ***B,*** Cortical sites with the maximum response difference. The solid black curve specifies the difference of two solid curves in ***A***, and the gray curve the difference of two pale-colored curves in ***A***. The black and gray triangles indicate the cortical sites with the maximum response differences between the solid curves and between the pale-colored curves, respectively.

5.3. Pupil size dynamics during fixation

Pupil size varies in response to fluctuations in stimulus luminance (Alpern et al., 1963; Ellis, 1981), but other factors can impact pupil size including arousal level (Hess and Polt, 1960; Bradshaw, 1967; Henson and Emuh, 2010) and task-related cognitive demands (Hess and Polt, 1964; Kahneman and Beatty, 1966; Nassar et al., 2012; de Gee et al., 2014). Our study has corroborated and extended these findings by demonstrating that changes in pupil size occur at different time scales and may reflect multiple functional components while observers perform either a fixation task or a visual discrimination task over an extended period of time. At a long time scale (4~6 mins), the pupils contracted gradually over time in the both tasks (mean -6.9%/min for Experiment 2; mean -4.4 %/min for Experiment 3). Perhaps, this large-scale pupil contraction reflects a gradual decrease in overall arousal level. Regardless of its exact origin, the magnitude of this task-invariant, gradual contraction in pupil size can produce larger errors in estimation of gaze position. What makes this gradual change in pupil size even more problematic is the fact that it varies among individuals (range, -25.3 ~ 4.0 %/min, SD, 9.1 %/min for Experiment 2; range -17.7 ~ 3.7 %/min, SD across observers 5.2 %/min for Experiment 3), thereby compromising the statistical power and possibly promoting false positive effects when one tries

to probe differences in gaze position in experiments using a between-subjects design.

We also found that pupil size varied over a relatively small time scale (≤16 s). Unlike the tonic changes found at a large time scale, the dynamics of the phasic changes plausibly reflect task-specific responses of the pupillary system. In the visually guided saccade task we have employed, pupil size remained constant when the position of the fixation target was frequently (mean 0.5 Hz) updated in an unpredictable manner, but pupil size started to contract steadily as soon as the target came to a standstill (Fig. 10E). This difference between the two fixation conditions implies that the pupillary system reacts instantaneously to subtle changes in uncertainty of fixation targets (Yu and Dayan, 2005; Sara, 2009). Similar to the large-scale tonic changes, the fixation-duration-dependent changes in pupil size varied greatly across individuals (Fig. 10F) and thus invite the same problems posed by the idiosyncratic nature of the tonic changes when it comes to its impacts on gaze position measurements. In the visual discrimination task we used, pupil size underwent trial-locked multi-phasic fluctuations around its mean (black curve in Fig. 14C). Several different factors, which were associated with a series of events constituting a trial, are likely to contribute to these trial-locked fluctuations. First, the biphasic response locked to the 'get ready'

event probably reflect pupil contraction due to the increase in total luminance associated with the size increase of the fixation dot (Fig. 14A; from 0.12° to 0.18° in diameter), followed by pupil dilation caused by an increasing level of alertness associated with observers' expectation of the upcoming target stimulus. Likewise, the next biphasic response locked to the 'stimulus onset' and 'choice' events could reflect a mixture of an initial contraction, caused by the luminance increase associated with the target ring shown in the periphery, followed by rapid dilation associated with observers' deliberative effort to make a decision when faced with a high degree of uncertainty (Preuschoff et al., 2011). The monotonic decrease in pupil size during the inter-trial fixation period plausibly reflects a quick return to its baseline size initially and followed by a moderate, gradual decrease probably due to prolonged fixation, similar to what we observed in the visually guided saccade task. Regardless of what caused these event-locked pupil size fluctuations, our findings show that the pupillary system is sensitive to subtle changes in external and internal task-related demands transpiring over a short period of time.

5.4. Pupil size artifact in video-based gaze estimation

Despite its multi-source origins and associated individual differences, pupil size was found to be robustly linked to gaze position measurements, as evidenced by the invariance of the PA-GP relationship to the differences in task and time scale. Although the factors that contributed to the changes in pupil size differed between the two tasks, the PA-GP correlations were quite comparable between the two tasks within individual observers (Fig. 14D; Pearson's r=0.75 and 0.57 for the left and right eyes, respectively). In addition, the inter-observer correlation between the decrease in PA and the shift in GP that were measured at a large time scale (Fig. 10D; Pearson's r = -0.76) was almost identical to the correlation calculated at a small time scale (Fig. 10F; Pearson's r = -0.73). This, in turn, means that changes in pupil size could be mistaken for shifts in gaze position if eye movement measurements were not corrected for dynamic pupil size changes, as illustrated in Fig. 14C and as discussed in a recent paper (Chen and Hafed, 2013), which compared the gaze positions that were simultaneously measured by the video-based eye tracker and the scleral search coil system. This mistaken identification would confound attempts to test the effects of different experimental conditions on gaze positions since the pupillary system reacts differently to subtle differences in exogenous and endogenous

factors associated with different conditions.

However, it should be noted that the PA-GP relationship, although quite robust for a single eye of a given observer, was different between the eyes and across the observers (Fig. 11B). We reckon that the inter-observer differences were probably due to the idiosyncrasy in pupil shape and to the magnitude of the center shift during constriction ('decentration') relative to the corneal center (Wilson et al., 1992; Charlier et al., 1994; Rynders et al., 1995; Wyatt, 1995; Wildenmann and Schaeffel, 2013) -- pupil shape and centration are two of the most crucial constants in estimating gaze position with the P-CR method. We suspect that the inter-ocular differences arose from our specific positioning of the camera and LED illuminator in relative to the center of gaze: the camera was off from the axis created between the center of the display (large white circle in Fig. 9A), where the fixation targets appeared, and the center of an observer's head. This differential geometry between the eyes probably produced slightly different 2D images of the pupils that, in turn, could have contributed to the different PA-GP relationships due to viewing-direction-dependent nonlinear distortions that occur when recovering 3D information from 2D images (Fedtke et al., 2010; Gagl et al., 2011; Brisson et al., 2013; Mathur et al., 2013). There might be an alternative way of setting up an eye tracker such that those distortions are

mitigated (e.g., aligning the camera, instead of the LED illuminator, to the center of gaze). However, those distortions may be intrinsic to the estimation of 3D from 2D information, and that identifying an error-minimizing setup separately for individual observers would require a substantial amount of time and, therefore, be impractical in many experimental situations

The robust PA-GP relationship allowed us to correct gaze position measurements for pupil size based on a regression-based method, without resorting to any special hardware or separate calibration runs. As Drewes and colleagues (2012) suggested, one might consider calibrating the eye tracker under different pupil-size conditions to identify and correct for the errors associated with changes in pupil size. Although this pupil-size-specific calibration should work in principle, it would not be easily applied to typical experiments, particularly those that involve many human observers, because it requires access to raw camera outputs and needs to be carried out frequently between experimental runs to compensate for subtle changes in head position or angle. In contrast, the regression-based correction method can be applied to "out-of-the-box" data without additional calibrations, thus affording an easily implemented, effective means for correcting for pupil-size-dependent errors that will improve both the accuracy and precision of gaze estimation. Indeed, the corrected GP measurements could readily

resolve spatial displacements as small as 0.1°. It should also be noted that the standard deviation value (0.21° ± 0.07°) of the corrected GPs obtained during the prolonged fixation was comparable to the radius (0.21°) of the fixation span (483 arcmin^2, 68th percentile) reported by a recent study using a Dual Purkinje Image eye tracker (Cherici et al., 2012), an expensive device that is free from pupil artifacts.

There are at least three caveats associated with the regression-based correction method that should be kept in mind. First, the current correction method might not be sufficient in experiments involving large eye movements such as those associated with reading (Gagl et al., 2011) or with unconstrained viewing of natural scenes. With those kinds of eye movements, rotation of eyes may cause notable nonlinear distortions to the 2D images of the pupil (Fedtke et al., 2010; Gagl et al., 2011; Brisson et al., 2013; Mathur et al., 2013), calling for a more sophisticated correction method. Second, we did not attempt to determine whether the current correction method is robust in the presence of small head movements, which goes beyond the purpose of the current study. Finally, the regression-based correction is post-hoc by nature, and thus cannot be applied to the data in an online fashion, such as initiating task events based on gaze position measurements in an interactive way.

6. Experimental procedures

6.1. Observers

Nineteen individuals (nine females; 20~30 year-old) participated in the fMRI experiment (Experiment 1), and twenty-three (eleven females; 18~36 year-old; 2 of whom also participated in the main experiment) participated in the eye-tracking experiment (Experiment 2 and 3), with informed consent in accordance with the guidelines and approval of the Institutional Review Board at Seoul National University. None of the participants reported any history of reading problems or symptoms of abnormal vision. All except one observer were naïve to the purpose of the study.

6.2. fMRI experimental setup

MR data were collected using a 3 Tesla Siemens Tim Trio scanner equipped with a 12-channel Head Matrix coil at the SNU Brain Imaging Center. Stimuli were generated using Matlab (The MathWorks Inc., Natick, MA) in conjunction with MGL (URL: http://justingardner.net/mgl) on a Macintosh computer. Observers looked through an angled mirror attached to the head coil to view stimuli displayed via an LCD projector (Canon XEED SX60)

onto a back-projection screen at the end of the magnet bore at a viewing distance of 87 cm, yielding a field of view of 22x17°.

6.3. fMRI behavioral protocol (Experiment 1)

Observers participated in one fMRI session of retinotopy-mapping runs, wherein V1 boundaries, a population eccentricity-tuning map, and a hemodynamic impulse response function (HIRF) were defined, and one session of main experimental runs, wherein observers performed a fine ring-size discrimination task (Fig. 1). On each trial of this task, the observer initially viewed a small fixation dot (diameter 0.12°, luminance 321 cd/m^2) appearing at the center of a dark screen (luminance 38 cd/m^2). A small but foveally visible increase in the size of the fixation dot (from 0.12° to 0.18° in diameter) forewarned the observer of an upcoming presentation of the test stimulus. That test stimulus consisted of the brief (300 ms) presentation of a thin (full-width at half-maximum of a Gaussian envelope, 0.17°), white (321 cd/m^2), dashed (radial frequency, 32 cycles/360°) ring that counter-phase-flickered at 10 Hz. Following each brief ring presentation, observers reported the ring's size – "small" or "large" -- using a left-hand or right-hand key, guessing if necessary. Observers were instructed to maintain strict fixation on the central dot, for otherwise they would be unable to detect the change in

the fixation dot signaling a forthcoming brief target stimulus and would invariably hamper their performance on the ring-size discrimination task.

Inside the scanner but without being scanned, observers performed 54 practice trials then 180 threshold-estimation trials before the main experiments. On each of the threshold-estimation trials, dubbed 'short trials', one of 20 different-sized rings was presented according to a multiple random staircase procedure (four randomly-interleaved 1-up-2-down staircases, two starting from the easiest stimulus and the other two starting from the hardest one) with trial-to-trial feedback and with inter-stimulus interval of 2.7 s. The psychometric curves obtained from the short trials were fit by a Weibull function using a maximum-likelihood procedure. From the fitted Weibull function, the size contrast (*SC*) associated with 70.7% correct was estimated to determine the radii (r_M, r_S, and r_L) of the three ring stimuli (S: small, L: large, M: medium; Fig. 1B) used in the main experiments: $r_M = 2.84°$; $r_S = (1\text{-}SC)*r_M$; $r_L = (1+SC)*r_M$. In the main experiments, observers performed 156 'long' trials (see Fig. 1A for trial structure; note that the trials mentioned below are all 'long' unless otherwise specified) in total, while being scanned over six, 343.2 s functional scan runs, on these three different-sized rings, which were presented in the order defined by an m-sequence (base=3, power=3; nine S- and L-rings and eight M-rings were presented; all scan

runs started with two M-rings; Buracas and Boynton, 2002) to null the autocorrelation between stimuli. Before participating in the fMRI experiments, each observer practiced on the task intensively (about 6000 short trials over 6 sessions) outside the scanner.

6.4. Acquisition and preprocessing of MRI data

For each observer's brain, two 3D, T1-weighted, high-resolution (1x1x1 mm) anatomical scans were acquired with an optimized protocol (MPRAGE; field of view (FOV), 256 mm; repetition time (TR), 1.9 s; time for inversion, 700 ms; time to each (TE), 2.36 ms; flip angle (FA), 9°), averaged to improved image fidelity and segmented/flattened to be aligned with the data from the retinotopy-mapping and main experimental scan sessions using FreeSurfer (URL: http://surfer.nmr.mgh.harvard.edu; Dale et al., 1999).

T2*-weighted functional images were obtained with a gradient echo planar imaging (EPI) pulse sequence for the retinotopy-mapping and main experimental scans. The parameters for these two scan types, which differed slightly, were as follows (retinotopy followed by experimental): TR, 2.7 s, 2.2 s; TE, 40 ms, 40 ms; FA, 77°, 73°; FOV, 208 mm, 207 mm; image matrix, 104x104, 90x90; slice thickness, 1.8 mm with 11% gap, 2 mm with 15% slice gap; slices, 30, 22 oblique transverse slices; bandwidth, 858 Hz/px,

750 Hz/px; effective voxel size 2.0x2.0x1.998 mm, 2.3x2.3x2.3 mm). At the beginning of each functional session, a high-resolution (1.078x1.078x2.0 mm, 1.083x1.083x2.3 mm) T1-weighted inplane image was acquired with the same slice prescription as the functional images (MPRAGE; TR, 1.5 s; TI, 700 ms; TE 2.79 ms; FA, 9°) for the image-based registration.

All functional EPI images were motion-corrected using SPM8 (URL: http://www.fil.ion.ucl.ac.uk/spm; Friston et al., 1996; Jenkinson et al., 2002), and then co-registered to the reference high-resolution anatomical volume of the same observer's brain via the high-resolution inplane image (Nestares and Heeger, 2000). After co-registration, the images of the retinotopy-mapping scan were resliced, but not spatially smoothed, in alignment with the spatial dimensions of the main experimental scans. The area V1 was manually defined on the flattened gray-matter cortical surface mainly based on the meridian representations, resulting in 825.4 ± 140.7 (mean ± standard deviation (SD) across observers) voxels. The individual voxels' time-series were divided by their means to convert them from arbitrary intensity units to percentage modulations and were linearly de-trended and high-pass filtered (Smith et al., 1999) using custom scripts in Matlab. The cut-off frequency was 0.0185 Hz for the retinotopy-mapping session and 0.0076 Hz for the main session. The first 10 (out of 90; a length

of a cycle) and 6 (out of 156; a length of a trial) frames of each retinotopy-mapping and main scan, respectively, were discarded to minimize the effect of transient magnetic saturation and allow the hemodynamic response to reach steady state. The 'blood-vessel-clamping' voxels, which show unusually high variances of fMRI responses, were discarded (Olman et al., 2007; Shmuel et al., 2007); a voxel was classified as 'blood-vessel-clamping' if its variance exceeds 10 times of the median variance value of the entire voxels.

6.5. Eccentricity-tuning mapping in V1 voxels

Retinotopy-mapping scans. Standard traveling wave methods (Engel et al., 1994; Sereno et al., 1995) were used (1) to define V1 (Fig. 2A), (2) to estimate each observer's hemodynamic impulse response function (HIRF) of V1 (Fig. 2D), and (3) to estimate V1 voxels' receptive field center and width (Figs. 2B and 2C). High-contrast and flickering (1.33 Hz) dartboard patterns were presented either as 0.89°-thick expanding or contracting rings in two scan runs, as 40°-width clockwise or counterclockwise rotating wedges in four runs, or in one run as four stationary, 15°-wide wedges forming two bow-ties centered on the vertical and horizontal meridians. Each scan run consisted of 9 repetitions of 27 s period of stimulation. The fixation behavior

during the scans was assured by monitoring observers' performance on a fixation task, in which they had to detect any reversal in direction of a small dot rotating around the fixation.

HIRF estimation. For each observer, the data from the bowtie scan provided the estimation of HIRF. The procedure of HIRF estimation was as follows. First, a group of voxels that were driven by the bowtie stimuli was defined by identifying the ones whose SNR (the ratio of Fourier power at the stimulus frequency (0.037 Hz) to at frequencies higher than the 3rd harmonics (> 0.111 Hz)) was greater than 3. Second, the time-series from those voxels (204.6 ± 50.8 and 136.9 ± 30.9 voxels locked to the vertical and horizontal meridians, respectively; mean ± SD across observers) were aligned each to the stimulus onset and then all pooled and averaged across voxels to enhance SNR, resulting in a single representative time-series. Third, the HIRF was parameterized using a difference of two gamma functions (Friston et al., 1998; Glover, 1999) by fitting the predicted fMRI time-series to the representative time-series using a least-square procedure, which was implemented by the *ga* function (for initial estimation) in conjunction with *fminsearch* function in the Global Optimization Toolbox in Matlab. The model explained a large fraction of the total variance in the representative

time-series (91.8% ± 3.7%; mean ± SD across observers).

Estimation of population eccentricity-tuning curves. The map of population eccentricity-tuning curves (Fig. 2C) was defined by fitting a one-dimensional Gaussian function simultaneously to the time-series of fMRI responses to the expanding and contracting ring stimuli, which were also used for definition of V1. Details of this procedure are as follows.

First, as in the HIRF estimation, the time-series of fMRI were extracted only from a relevant group of voxels with SNR higher than 3 in both of the ring sessions. Second, an eccentricity-tuning curve (gain over eccentricity, in other words) of a single voxel, $g(\varepsilon)$, was modeled by a Gaussian as a function of the eccentricity in a visuotopic space, ε, and it was parameterized by a peak eccentricity, e, and a tuning width, σ:

$$g_e(\varepsilon) = exp^{-\left(\frac{(\varepsilon-e)^2}{2\sigma^2}\right)}. \qquad \text{(Eq. 1)}$$

Third, the collective responses of neurons within that voxel with a particular $g(\varepsilon)$, at a given time frame t, $n(t)$, were predicted by multiplying $g(\varepsilon)$ to spatial layout of stimulus input at that time frame, $s(\varepsilon,t)$:

$$n(t) = \sum_{\varepsilon} s(\varepsilon, t)g(\varepsilon). \qquad \text{(Eq. 2)}$$

Fourth, the predicted time-series of fMRI responses of that voxel, $fMRI_p(t)$, were generated by convoluting $n(t)$ with a scaled (by β) copy of the HIRF acquired from the meridian scans (as described above), $h(t)\beta$ plus a baseline

response, b:

$$fMRI_p(t) = n(t) * h(t)\beta + b. \qquad \text{(Eq. 3)}$$

Fifth, the model parameters (e, σ, β, b) were found by fitting $fMRI_p(t)$ to the predicted time-series of fMRI responses to actual stimulation, $fMRI_o(t)$, by minimizing the residual sum of squared errors between $fMRI_p(t)$ and $fMRI_o(t)$ over all time frames, RSS:

$$RSS = \sum_t (fMRI_o - fMRI_p)^2. \qquad \text{(Eq. 4)}$$

Sixth, a valid group of voxels was further refined by discarding voxels with goodness of fit, estimated by R^2, the squared correlation between $fMRI_p(t)$ and $fMRI_o(t)$, below a criterion, which itself was established by a bootstrap procedure. The bootstrap distribution of R^2 was created by computing R^2s based on the fits of $fMRI_p(t)$ to bootstrap sample time-series of $fMRI_o$, which was in turn obtained by superposing the 8 repetitions of 10 normally-distributed random values (mean 0; SD $\sqrt{3}$) onto 80 normally-distributed random values (mean 0; SD 1). The bootstrapped 99% confidence criterion for R^2 was 0.4. Seventh, the estimated parameters, e and σ, were roughly consistent with previous fMRI studies (Dumoulin and Wandell, 2008; Kay et al., 2008; Harvey and Dumoulin, 2011), but we noticed two notable exceptions with respect to the value of σ: its sudden increase and decrease in the foveal and peripheral regions and a large gap around σ=2°. Therefore, the

voxels whose values of e were either smaller than 1.26° or greater than 5.65° or whose values of σ were greater than 2 were discarded from further analyses. Overall, the number of the valid voxels was 191.5 ± 54.5 (mean ± SD across observers), resulting in the selection rate of 23.5% ± 6.8%. Our conservative rule of voxel selection was supported by the well-known linear regression of σ by e with a power of 1.1 (Duncan and Boynton, 2003):

$$\sigma(e) = ce^{1.1} + d, \qquad \text{(Eq. 5)}$$

where the estimated c and d are 0.0952 and 0.5953, respectively. Based on this relationship, we constructed the matrix of population eccentricity-tuning curves by assigning σ values to e values following Equation 5. The pattern of results we found, incidentally, remained unchanged when we reanalyzed our data applying a less conservative voxel selection criterion (R^2>0.2) that increased the selection rate (56.1% ± 7.8%).

6.6. Definition of eccentricity bins

The twenty-one eccentricity bins and their fMRI responses were defined by the following steps. First, the estimated eccentricity values (e), which were defined in visuotopic scale, were converted into values in units of "relative cortical distance" (e_{rcd}), which scales positions relative to the cortical region representing r_M, based on the canonical cortical magnification factor for

human visual cortex (Horton and Hoyt, 1991):

$$e_{rcd} = 17.3\, log(e + 0.75) - 17.3\, log(r_M + 0.75). \quad \text{(Eq. 6)}$$

Second, the e_{rcd} values were split into twenty-one equal-sized bins, such that the central (11^{th}) bin represents r_M (2.84°) whereas the foveal and peripheral ends represent 1.26° and 5.65°, respectively. Due to differences in cortical coverage, the number of voxels in single bins monotonically increased as a function of the preferred eccentricity. Third, the fMRI responses from V1 voxels (Fig. 3A) were transformed into those from the twenty-one eccentricity bins by applying the Gaussian kernels, the centers of which were at each bin center and the full-width half-maximum of which were the two units of bin size (Fig. 3B). Fourth, the binned fMRI responses of each scan run from each observer were scaled to match the size of the matrix (21 eccentricity bins x 150 time frames; Fig. 3C), as we did to the simulated time-series.

6.7. Computation of stimulus and choice probabilities

To quantify the ability of an ideal observer to predict from the matrix of fMRI responses whether the stimulus presented was an S-ring or an L-ring, we computed stimulus probabilities (*SP*s) in the following way. For each observer, trials were sorted into 6 classes jointly defined by a stimulus and

by a choice: 'stimulus|choice' = 'S|S', 'S|L', 'M|S', 'M|L', 'L|S' and 'L|L' (Fig. 4A). By contrasting the stimulus factor conditional on the choice factor ('**S**|S' vs. '**L**|S' or '**S**|L' vs. '**L**|L'; top panel in Fig. 4B), 'receiver operating characteristic (ROC)' curves (Fig. 4C) were constructed by defining α and β, two integrals of the conditional response probability density p[r|'**S**'] and p[r|'**L**'] as a function of c, a classification threshold:

$$\alpha(c) = \int_c^{\infty} dr\ p[r|'S'];$$

$$\beta(c) = \int_c^{\infty} dr\ p[r|'L'], \qquad \text{(Eq. 7)}$$

where r is a response at a cell within the trial-related matrix of fMRI responses (Fig. 3D). Then, because the ROC curve is β plotted as a function of α, the probability of the correct classification for a given '**S** vs **L**' contrast pair, *P[correct]* is equivalent to the area under the ROC curve, which can be computed by integrating β over all values of α:

$$P[correct] = \int_0^1 d\alpha\ \beta. \qquad \text{(Eq. 8)}$$

In a similar manner, we quantified the ability of an ideal observer to predict an observer's choice by computing choice probabilities (*CP*s) from the fMRI responses at the same local cortical sites used to compute *SPs*. This entailed contrasting the choice factor conditional on the stimulus factor ('S|**S**' vs. 'S|**L**', 'M|**S**' vs. 'M|**L**', and 'L|**S**' vs. 'L|**L**'; bottom panel in Fig. 4B).

When the procedure above is applied to the two 'stimulus-contrast'

pairs and the three 'choice-contrast' pairs, the resulting *P[correct]*s are the two individual *SP*s ($SP_{choice=S}$ and $SP_{choice=L}$; horizontal brackets in Fig. 4A) and the three *CP*s ($CP_{stimulus=S}$, $CP_{stimulus=M}$, and $CP_{stimulus=L}$; vertical brackets in Fig. 4A), respectively. The grand *SP* and *CP* (Fig. 4D; Fig. 6A) were the averages of those individual *SP*s and *CP*s. Note that our definitions of *SP*s and *CP*s are different from the definition used in single-cell neurophysiological studies. Specifically, our definition means that values of *SP* >.5 denote larger fMRI responses on the L-ring trials than on the S-ring trials, regardless of eccentricity bin preference (and conversely that *SP* values <.5 denote larger fMRI responses on S-ring trials relative to L-ring trials). Likewise, by our definitions, *CP* values >.5 denote larger fMRI responses on the L-choice trials than on the S-choice trials (and conversely *CP*<.5 indicates larger fMRI responses on the S-choice trials than on the L-choice trials). The proportions of S- or L-choice trials within the S- or L-ring trials could be unbalanced, simply due to observers' above-chance-level performance. However, the numbers in the most unbalanced trial group across all observers were 48 to 5. In earlier studies, data set containing at least five trials of each alternative are considered valid (Nienborg and Cumming, 2006).

The spatiotemporal cells with significant (corrected $p<0.05$) *SP* or

CP values were identified with the threshold-free cluster-enhancement method (TFCE; Smith and Nichols, 2009): at each of 2000 permutations, the maximum TFCE (computed with dh=0.1, H=2, and E=2/3) value out of all 126 spatiotemporal cells was taken to build up the null distribution, against which the observed TFCE values of each cells were compared.

6.8. Prediction of V1 population responses to ring stimuli

The model prediction of fMRI population responses to the different-sized rings based on the map of population eccentricity-tuning curves (Fig. 2E) was generated in the following steps. First, a vector of stimulus events was defined by an m-sequence of the S-, M-, and L-rings (80 trials with base of 3 and power of 4) with inter-stimulus interval of 13.2 s (6 time bins with 2.2 s repetition time, representing -1.1 s, 1.1 s, 3.3 s, 5.5 s, 7.7 s and 9.9 s after stimulus onset, respectively), which replicated the temporal structure of stimulus events in the main experiment except for the sequence length. Second, for each simulation trial, a profile of responses of the twenty-one eccentricity bins (see Section 6.6 *Definition of eccentricity bins* for how these bins were defined) to a given ring (the across-observer averaged *SC* was used to determine the size of S- and L-ring) was defined by a set of gains of population eccentricity-tuning curves at the eccentricity of the

stimulus across of the bins (blue and red bell-shape curves with dotted lines at center, plotted on the right-hand vertical axis in Fig. 2C). The generation of the response profiles across all the simulation trials resulted in a spatiotemporal matrix of neural responses to the sequence of ring stimuli (21 eccentricity bins x 480 time frames). For ease of comparison with the experimental data, sum of squares of the matrix was scaled to match the size of the matrix (21x480). Third, the convolution of this matrix with the observer-averaged V1 HIRF (Fig. 2D) predicted a matrix of noise-free fMRI responses to the sequence of rings. Finally, the 2-D matrix of differential responses to the L- and S-rings was obtained by subtracting the trial-locked average of predicted fMRI responses to the S-ring trials from that to the L-ring trials (Fig. 2E).

6.9. Decomposition of fMRI population responses into tuned and untuned components

At each time frame, t, and at each eccentricity bin, i, the average of raw responses (RRs) across the n_e (=21) eccentricity bins was subtracted from the RR_i to derive the tuned response (TR_i) (Fig. 5A):

$$TR_i(t) = RR_i(t) - \sum_{i=1}^{n_e} RR_i(t)/n_e. \qquad \text{(Eq. 9)}$$

6.10. Population decoding of stimulus and choice information

The procedure for computing the population *SP*s and *CP*s was identical to that for computing the *SP*s and *CP*s at the individual local cells, except that '*r*' in Eq. 7 was replaced with '*pr*', weighted (*w*) sum of population raw responses (RRs) over the eccentricity bins (i=[1, n_e]):

$$pr = \sum_{i=1}^{n_e} w_i RR_i. \qquad \text{(Eq. 10)}$$

The three different weighting profiles, each representing the contributions of the individual eccentricity bins assessed by the three different schemes (the 'uniform', the 'discriminability' and the 'log-likelihood-ratio' schemes), were defined as follows. The 'uniform' scheme (Fig. 7A) assigned three discrete values to the eccentricity bins depending on which flanking side of the M-size ring their preferred eccentricities (e) belonged to:

$$\mathrm{w}\,(e) = \begin{cases} -1, & \text{for } e < \mathrm{r}_M \\ 0, & \text{for } e = \mathrm{r}_M \\ 1, & \text{for } e > \mathrm{r}_{M.} \end{cases} \qquad \text{(Eq. 11)}$$

The 'discriminability' scheme (Fig. 7B) defined weights in proportion to the differential responses of given eccentricity bins to the L-size and the S-size rings, which were derived from the eccentricity-tuning curves defined from the retinotopy-mapping session:

$$w(e) = g_e(r_L) - g_e(r_S) - \delta, \qquad \text{(Eq.12)}$$

where g_e is the eccentricity-tuning curve of the eccentricity bin with preferred eccentricity, *e*, as defined by Eq. 1, and the baseline offset, δ, is $\sum_e [g_e(r_L) - g_e(r_S)]/n_e$.

The 'log-likelihood-ratio' scheme (Fig. 7C) defined weights by taking the differences between the log-likelihoods of obtaining a given response if the stimulus were the L-ring stimulus, $logL_L$, and if the stimulus were the S-ring stimulus, $logL_S$. Because the eccentricity-tuning curves were assumed to be described by a Gaussian function, the log-likelihood-ratio weights at preferred eccentricity, *e*, can be simplified to the following formula:

$$w(e) = logL_L - logL_S = -\frac{1}{2\sigma_L^2}(e - r_L)^2 + \frac{1}{2\sigma_S^2}(e - r_S)^2 - \delta, \qquad \text{(Eq. 13)}$$

where σ_L and σ_S are the tuning width derived from Eq. 5 with r_L and r_S, and the baseline offset, δ, is $\sum_e [-\frac{1}{2\sigma_L^2}(e - r_L)^2 + \frac{1}{2\sigma_S^2}(e - r_S)^2]/n_e$.

6.11. Eye-tracking experimental setup

Stimuli were presented in a dimly lit room on a gamma-linearized 22-inch CRT monitor (Totoku CV921X CRT monitor) operating at vertical refresh

rate of 180 Hz and a spatial resolution of 800 x 600 pixels. Stimuli were generated using MATLAB (MathWorks) in conjunction with MGL (http://justingardner.net/mgl) on a Macintosh computer. Observers viewed the monitor at a distance of 90 cm while their binocular eye positions were sampled at 500 Hz by an infrared eye tracker (EyeLink 1000 Desktop Mount, SR Research; instrument noise, 0.01° RMS; Fig. 9A). The LED illuminator and camera (broken-line boxes in Fig. 9A) were positioned side by side, at a distance of 65 cm from the observer (broken line with arrow ends in Fig. 9A), and angled toward the observer's face to insure that infrared light illuminated both eyes and was being reflected from both eyes and imaged on the camera sensor.

An observer sat on a height-adjustable chair with his/her head supported by a forehead and chin rest (HeadSpot, UHCOTech), which were, together with the monitor, mounted on a height-adjustable table (Fig. 9B). To minimize body and head movements that compromise the quality of eye tracking measurements, the following procedure was applied. First, an observer was given enough time to find a comfortable arrangement of the chair, table, forehead, and chin rest by adjusting the heights of those devices. Second, the lower part of the head was harnessed by wrapping a memory-foam cushion around the neck such that the cushion's ends were tightly

secured to the head post and the sides of the chin. Third, the upper and middle part of the head was constrained by fastening a wide buckled cotton strap over the forehead, the head post, and the lower back of the head. To mitigate discomfort associated with tight head fixation, baby-proofing cushion tapes were attached on the contact surfaces of the chin-rest.

The eye tracker was calibrated using the built-in five point calibration routine (HV5), not only at the beginning of each daily session but also whenever the observer was disengaged from a previously calibrated head positioning setup. During a session, the observer was allowed to take as many breaks as desired, disengaging from the eye tracking setup and moisturizing the eyes using disposable artificial tears as needed. Eye tracking signals were acquired in a 'pupil-corneal reflection (P-CR)' mode, and the pupil center was estimated using the ellipsoid fitting method, which is known to be robust to pupil occlusion by the eyelids. To check the possibility that the relationship between pupil size and gaze position is dependent on pupil tracking methods, we also collected data using the centroid method, an alternative method of pupil center estimation. But the results using the ellipsoid and centroid methods did not differ.

6.12. Eye-tracking behavioral protocol (Experiment 2 and 3)

Each observer participated in a total of three daily sessions, one for practice (315 short trials), one for threshold *SC* estimation (315 short trials plus four runs of the main task, 108 trials) and the other for six runs of the ring-size discrimination trials with eye position being monitored (162 trials). In addition, a visually-guided saccade task (Tse et al., 2010) taking about 4-minutes was conducted at the beginning of each session to measure the sensitivity of eye-tracker.

Visually guided saccade task (Experiment 2). Observers performed a visually guided saccade task (Tse et al., 2010) by shifting or holding their gaze on a fixation target that appeared at three different positions on the monitor. An experimental run consisted of two alternating blocks of eye tracking measurements. In 'prolonged-fixation (PR)' blocks, a central gray (30 cd/m^2) dot (0.12° in diameter) was presented as a fixation target (FT) for 16 s against dark (3 cd/m^2) background. In 'short-lived fixation (SL)' blocks, which lasted for 31 s, the position of the FT was updated at 1 Hz, appearing either in the left (−0.12°) or in the right (+0.12°) side of the center of the monitor. The position of the FT was determined by an m-sequence (31 trials with base of 2 and power of 5), making location order unpredictable over

time with zero autocorrelation. Every run started and ended with a PR block, and contained 5 SL blocks, resulting in a total of 251 s (16 s x 6 PR blocks plus 31 s x 5 SL blocks) for one single run. Each observer performed a single run.

Fine ring-size discrimination task (Experiment 3). The stimuli and procedure matched those of Experiment 1 (see Section 6.3) except for the following. First, although the luminance contrast between the stimuli and the background remained comparable to that in Experiment 1, their absolute luminance values were changed to 30 cd/m^2 and 3 cd/m^2, respectively. Second, the threshold *SC* value was determined using an even larger number of trials (315 instead of 180). Third, one more M-ring trial was added to a given run, resulting in slightly more M-ring trials (48 trials in total per observer) available for data analysis. Lastly, the number of practice trials was smaller compared to that in Experiment 1.

6.13. Eye-tracking data preprocessing

The EyeLink system estimates gaze position and pupil area using built-in proprietary software and provides those estimates to end users in a digitized format called 'EDF.' In this file format gaze position estimates are in units of

pixels of the stimulus monitor, and pupil area estimates are in arbitrary units. Although it was possible to use raw camera output, these EDF-format data were used because the heuristic filter adopted by the proprietary software is known to be superior to any other known algorithms in suppressing noise in video-based eye tracking data (Kimmel et al., 2012) and because we wanted end users to readily apply the data preprocessing and analysis procedures proposed in the current study to their native, "out-of-the-box" data. The EDF files were imported to MATLAB using an open source script (https://github.com/iandol/opticka/blob/master/communication/edfmex.m) and analyzed using the custom MATLAB scripts. Here, we only focused on the horizontal gaze estimation data for further analyses because the vertical gaze measurements were much noisier; the SDs of raw gaze measurement from Experiment 2 were 0.53° ± 0.12° (mean ± SD across observers) and 0.43 ± 0.82° for the horizontal measurements from the left and right eyes, respectively, and 0.85° ± 0.64° and 0.75° ± 0.48° for the vertical measurements from the left and right eyes, respectively. Vertical gaze measurements become unreliable in video-based eye trackers when the upper eyelid occludes a significant portion of pupil margin, making the elliptical description of the margin unreliable. This depends on eyelid anatomy, lid posture, and pupil diameter of individual observers.

Because pupil information is unavailable during eye blinks, video-based gaze position measurements can be contaminated by eye blinks. Thus, eye blinks and associated time-series of gaze measurements were identified and excluded from subsequent analyses. As in previous studies (Troncoso et al., 2008; Otero-Millan et al., 2012), eye blinks were judged to occur if any of the following three conditions was met: (i) Pupil data were missing for either eye; (ii) pupil area measurements fluctuated abruptly with unrealistically large amplitudes (>50 units per sample); (iii) gaze position measurements deviated substantially (>10°) from the screen center. Then, the data acquired immediately before and after (±200 ms) were likely to be contaminated by eye blinks, hence were removed and replaced with Not-a-Number (NaN) values. The blink-confounded samples comprised 8.5 ± 7.2% (mean ± SD across observers) of the total data for Experiment 2 and 10.8 ± 9.1% for Experiment 3.

The blink-free samples of gaze position and pupil area were then scaled and normalized in a way that allowed them to be merged or compared across runs or observers. Gaze position values were converted to units of visual angle and were re-centered to their median value in a given measurement run (subtractive normalization) based on the assumption that observers' fixation positions are centered around the FT in a given run.

These blink-free, scaled-in-visual-angle, normalized-to-the-median gaze position samples will be referred to as GP. Pupil area values, which were originally provided in arbitrary units (mean ± SD across observers; 2629 ± 910 (left eye) and 2408 ± 871 (right eye) for Experiment 2; 2231 ± 579 (left eye) and 2044 ± 606 (right eye) for Experiment 3), were converted to units of 'percent change from the mean' to control for individual differences in absolute pupil size. These blink-free, normalized pupil area values will be referred to as PA. Note that the eye-averaged GP and PA values are used throughout the Sections 3 and 4, unless otherwise specified.

6.14. Correction of gaze position measurements for pupil size artifact

In video-based methods of gaze position estimation, changes in pupil size are known to engender systematic estimation errors (Wyatt, 2010; Ivanov and Blanche, 2011; Drewes et al., 2012; Kimmel et al., 2012). These pupil size-associated estimation errors were confirmed in our own data and regressed out from the original GP time-series using a simple second order polynomial:

$$GP_{corr} = GP - (\alpha + \beta*PA + \gamma*PA^2), \qquad \text{(Eq. 14)}$$

where a gaze position corrected for pupil size-associated estimation errors (GP_{corr}) is modeled as a subtraction of a second order polynomial of PA from

an original uncorrected GP value, and α, β and γ are the best fitting parameters that minimize the squared errors between the two sides of the equation. Note that a similar method has been presented in an abstract form (Ivanov and Blanche, 2011). The correction for pupil size was done separately for the two eyes and for each observer because the exact relationship between gaze estimation errors and PA is likely to differ between the eyes and individuals due to differences in multiple factors including eye geometry, eyelid position, and the magnitude of pupil decentration (Wyatt, 1995, 2010).

6.15. Accuracy and precision analysis of gaze position signal

The uncorrected and corrected GP estimates during the visually guided saccade task were compared, respectively, by evaluating how similar their means were to true FT positions (‘accuracy’) and how similar individual estimates were to each other under the same fixation condition (‘precision’). The terms of ‘accuracy’ and ‘precision’ were adopted to characterize two orthogonal qualities of estimation as in recent eye tracking studies (Kimmel et al., 2012; Otero-Millan et al., 2014; Reingold, 2014). Note that, in this context, ‘precision’ means the same as ‘repeatability’ or ‘reliability’. To

exclude the samples during saccades in computing accuracy and precision, and to make the data in the PR and SL blocks comparable, specific portions, not the entire parts, of the original GP time-series were extracted and down-sampled in the following procedure. We first divided the original time-series of GP at 500-Hz sampling frequency into discrete 1-s bins (thus 500 samples for each bin), between which the FT could have shifted to a new location thus triggered a saccade (Section 6.12), and then took only the central one third (blink-free samples between 0.333 s ~ 0.667 s) portion of each of those 1-s bins. Next, for each bin, the samples within the central portion were averaged, resulting in three down-sampled (1 Hz) sets of GPs (dsGPs), one for each of the three FT positions (-0.12°, 0°, and +0.12°). When more than 30 samples within its central portion of any given bin were confounded with eye blinks, the dsGPs in that bin was judged unreliable and discarded from the accuracy and precision analysis. The fractions of the discarded, eye blink-contaminated samples were 4.6% ± 5.7% (mean ± SD across observers), 4.5% ± 6.5% and 5.3% ± 5.5% of the dsGP samples toward -0.12°, 0°, and +0.12° FTs, respectively. In addition, the dsGPs in the initial PR block were discarded to match the number of trials between the three fixation conditions. For each observer, a mean and a standard deviation were computed separately for the three sets of valid dsGPs, each corresponding to

one of the three fixation conditions, respectively. A deviation of a mean dsGP from its true FT position was used as an error metric for estimation accuracy, whereas a standard deviation itself directly served as an error metric for estimation precision.

6.16. Statistical power analysis of gaze position signal

To quantify the benefit from correcting gaze estimates for pupil size in an experimental setting, the statistical powers of the uncorrected and corrected GPs were inferred, respectively, and compared by counting how many observations were required to reliably resolve a small difference in fixation position via Monte Carlo simulation. In the simulation, it was assumed that twenty measurements are collected respectively for two fixation positions separated only by 0.1° in a single experimental session. Those two sets of twenty measurements were acquired independently by randomly sampling twenty values from two Gaussian distributions whose means differed by 0.1° but whose standard deviations were identical. Two different standard deviation values, one reflecting the level of precision for the uncorrected GPs and the other for the corrected GPs, were determined by taking the quadratic means (root mean of the squares) of all observers' SDs of uncorrected and corrected dsGPs, respectively, which were pooled across three sets of dsGPs

by weighted averaging. Simulations were carried out by varying the number of experimental sessions from 6 to 60. For a given number of sessions, 50,000 bootstrap sample observations were generated and examined using a paired t test with p value set to 0.01. The statistical power was assessed by how quickly the percentage of the bootstrap sample experiments with significant p values increased as a function of session numbers.

6.17. Microsaccades analysis

We defined microsaccades conservatively by designating them as events when the position measurements from the both eyes met the following set of criteria which have been routinely used in previous studies (Engbert and Kliegl, 2003; Engbert and Mergenthaler, 2006): median velocity threshold, $\lambda=6$; minimum duration of 6 ms; minimum inter-saccadic interval of 20 ms; maximum amplitude of 2°. However, only the horizontal positions were used for detecting those binocular microsaccades because the vertical eye positions were often noisy due to pupil occlusion in some observers and because most microsaccades are known to occur along the horizontal meridian (Tse et al., 2004). The distribution of microsaccade amplitudes (n=72,300, merged across observers; median, 0.19°; gamma parameters, $\alpha=1.66$ and $\beta=0.14$) was comparable to those reported in previous studies

using video-based eye-trackers (Tse et al., 2010; Kimmel et al., 2012; Otero-Millan et al., 2012; Hafed, 2013).

6.18. Pupil size analysis

The raw pupil size data varied substantially across individuals (2135 ± 574 arbitrary size units; deblinked mean ± SD across observers). To make data comparable across different runs and observers, we normalized the raw pupil size by dividing them by their deblinked means within each run and converting into percent mean values. The pupil size *CP*s were calculated using the same procedure described above, except that 'r' in Eq. 7 was replaced with the mean pupil size in each 2.2 s interval, matched to the duration associated with acquisition of an fMRI volume in the main experiment. Our definition means that values of $CP > .5$ denote larger pupil size on the L-choice trials than on the S-choice trials (and conversely that *CP* values $< .5$ denote larger pupil size on S-choice trials relative to L-choice trials). In calculating group statistics for *CP*s (diamonds in the bottom panel of Fig. 15C), the values from six observers whose data set contained less than five incorrect trials in either S- or L-ring conditions were excluded; none were excluded in calculating $CP_{\text{stimulus=M}}$ (circles in the bottom panel of Fig. 15C).

6.19. Analysis of gaze position and vergence angle

Eye positions were sampled binocularly, so the final estimates of eye position were defined by taking the average of the position measurements from the both eyes. In addition, we estimated relative vergence angles, deviations from the default vergence angle determined by the distance between the two eyes and the fixation point on the display: the positive angle values indicate convergent eye movement whereas the negative values indicate divergent eye movement. The eye position and vergence *CP*s (bottom panels of Figs. 15D and 15E, respectively) were calculated using the procedure described above, except that '*r*' in Eq. 7 was replaced with the mean horizontal eye position and the mean vergence angle, respectively, in each 2.2 s interval of each trial. Our definition means that values of *CP* >.5 denote more rightward eye position or convergent eye movement, respectively, on the L-choice trials than on the S-choice trials (and conversely that *CP* values <.5 denote more rightward eye position or convergent eye movement, respectively, on S-choice trials relative to L-choice trials).

Bibliography

Alpern M, McCready DW, Jr., Barr L (1963) The Dependence of the Photopupil Response on Flash Duration and Intensity. J Gen Physiol 47:265-278.

Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. Nat Rev Neurosci 7:358-366.

Benucci A, Ringach DL, Carandini M (2009) Coding of stimulus sequences by population responses in visual cortex. Nat Neurosci 12:1317-1324.

Berens P, Ecker AS, Cotton RJ, Ma WJ, Bethge M, Tolias AS (2012) A fast and simple population code for orientation in primate V1. J Neurosci 32:10618-10626.

Bianciardi M, Fukunaga M, van Gelderen P, Horovitz SG, de Zwart JA, Duyn JH (2009) Modulation of spontaneous fMRI activity in human visual cortex by behavioral state. Neuroimage 45:160-168.

Bradshaw J (1967) Pupil size as a measure of arousal during information processing. Nature 216:515-516.

Brefczynski-Lewis JA, Datta R, Lewis JW, DeYoe EA (2009) The topography of visuospatial attention as revealed by a novel visual field mapping technique. J Cogn Neurosci 21:1447-1460.

Brisson J, Mainville M, Mailloux D, Beaulieu C, Serres J, Sirois S (2013) Pupil diameter measurement errors as a function of gaze direction in corneal reflection eyetrackers. Behav Res Methods 45:1322-1331.

Bristow D, Haynes JD, Sylvester R, Frith CD, Rees G (2005) Blinking suppresses the neural response to unchanging retinal stimulation. Curr Biol 15:1296-1300.

Britten KH, Newsome WT, Shadlen MN, Celebrini S, Movshon JA (1996) A relationship between behavioral choice and the visual responses of neurons in macaque MT. Vis Neurosci 13:87-100.

Buracas GT, Boynton GM (2002) Efficient design of event-related fMRI experiments using M-sequences. Neuroimage 16:801-813.

Campbell FW, Gubisch RW (1966) Optical quality of the human eye. J Physiol 186:558-578.

Carandini M, Heeger DJ (2012) Normalization as a canonical neural computation. Nat Rev Neurosci 13:51-62.

Cardoso MM, Sirotin YB, Lima B, Glushenkova E, Das A (2012) The neuroimaging signal is a linear sum of neurally distinct stimulus- and task-related components. Nat Neurosci 15:1298-1306.

Celebrini S, Newsome WT (1994) Neuronal and psychophysical sensitivity to motion signals in extrastriate area MST of the macaque monkey. J Neurosci 14:4109-4124.

Charlier JR, Behague M, Buquet C (1994) Shift of the pupil center with pupil constriction. Investigative Ophthalmology and Visual Science 35:1278.

Chen CY, Hafed ZM (2013) Postmicrosaccadic enhancement of slow eye movements. J Neurosci 33:5375-5386.

Chen Y, Geisler WS, Seidemann E (2006) Optimal decoding of correlated neural population responses in the primate visual cortex. Nat Neurosci 9:1412-1420.

Cherici C, Kuang X, Poletti M, Rucci M (2012) Precision of sustained fixation in trained and untrained observers. J Vis 12.

Churchland MM et al. (2010) Stimulus onset quenches neural variability: a widespread cortical phenomenon. Nat Neurosci 13:369-378.

Cohen MR, Newsome WT (2009) Estimates of the contribution of single neurons to perception depend on timescale and noise correlation. J Neurosci 29:6635-6648.

Collewijn H, van der Mark F, Jansen TC (1975) Precise recording of human eye movements. Vision Res 15:447-450.

Dale AM, Fischl B, Sereno MI (1999) Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage 9:179-194.

de Gee JW, Knapen T, Donner TH (2014) Decision-related pupil dilation reflects upcoming choice and individual bias. Proc Natl Acad Sci U S A 111:E618-625.

de Lafuente V, Romo R (2005) Neuronal correlates of subjective sensory experience. Nat Neurosci 8:1698-1703.

Donner TH, Sagi D, Bonneh YS, Heeger DJ (2008) Opposite neural signatures of motion-induced blindness in human dorsal and ventral visual cortex. J Neurosci 28:10298-10310.

Donner TH, Sagi D, Bonneh YS, Heeger DJ (2013) Retinotopic patterns of correlated

fluctuations in visual cortex reflect the dynamics of spontaneous perceptual suppression. J Neurosci 33:2188-2198.

Drewes J, Masson GS, Montagnini A (2012) Shifts in reported gaze position due to changes in pupil size: ground truth and compensation. In: Proceedings of the Symposium on Eye Tracking Research and Applications, pp 209-212. Santa Barbara, California: ACM.

Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex. Neuroimage 39:647-660.

Duncan RO, Boynton GM (2003) Cortical magnification within human primary visual cortex correlates with acuity thresholds. Neuron 38:659-671.

Einhauser W, Koch C, Carter OL (2010) Pupil dilation betrays the timing of decisions. Front Hum Neurosci 4:18.

Einhauser W, Stout J, Koch C, Carter O (2008) Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. Proc Natl Acad Sci U S A 105:1704-1709.

Ellis CJ (1981) The pupillary light reflex in normal subjects. Br J Ophthalmol 65:754-759.

Engbert R, Kliegl R (2003) Microsaccades uncover the orientation of covert attention. Vision Res 43:1035-1045.

Engbert R, Mergenthaler K (2006) Microsaccades are triggered by low retinal image slip. Proc Natl Acad Sci U S A 103:7192-7197.

Engel SA, Rumelhart DE, Wandell BA, Lee AT, Glover GH, Chichilnisky EJ, Shadlen MN (1994) fMRI of human visual cortex. Nature 369:525.

Fedtke C, Manns F, Ho A (2010) The entrance pupil of the human eye: a three-dimensional model as a function of viewing angle. Opt Express 18:22364-22376.

Fiser J, Chiu C, Weliky M (2004) Small modulation of ongoing cortical dynamics by sensory input during natural vision. Nature 431:573-578.

Fox MD, Snyder AZ, Zacks JM, Raichle ME (2006) Coherent spontaneous activity accounts for trial-to-trial variability in human evoked brain responses. Nat Neurosci 9:23-25.

Friston KJ, Williams S, Howard R, Frackowiak RS, Turner R (1996) Movement-related effects in fMRI time-series. Magn Reson Med 35:346-355.

Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R (1998) Event-related fMRI:

characterizing differential responses. Neuroimage 7:30-40.

Gagl B, Hawelka S, Hutzler F (2011) Systematic influence of gaze position on pupil size measurement: analysis and correction. Behav Res Methods 43:1171-1181.

Ghose GM, Harrison IT (2009) Temporal precision of neuronal information in a rapid perceptual judgment. J Neurophysiol 101:1480-1493.

Glover GH (1999) Deconvolution of impulse response in event-related BOLD fMRI. Neuroimage 9:416-429.

Gold JI, Shadlen MN (2001) Neural computations that underlie decisions about sensory stimuli. Trends Cogn Sci 5:10-16.

Graf ABA, Kohn A, Jazayeri M, Movshon JA (2011) Decoding the activity of neuronal populations in macaque primary visual cortex. Nat Neurosci 14:239-245.

Grunewald A, Bradley DC, Andersen RA (2002) Neural correlates of structure-from-motion perception in macaque V1 and MT. J Neurosci 22:6195-6207.

Gu Y, DeAngelis GC, Angelaki DE (2007) A functional link between area MSTd and heading perception based on vestibular signals. Nat Neurosci 10:1038-1047.

Gu Y, Angelaki DE, Deangelis GC (2008) Neural correlates of multisensory cue integration in macaque MSTd. Nat Neurosci 11:1201-1210.

Gur M, Beylin A, Snodderly DM (1997) Response variability of neurons in primary visual cortex (V1) of alert monkeys. J Neurosci 17:2914-2920.

Haefner RM, Gerwinn S, Macke JH, Bethge M (2013) Inferring decoding strategies from choice probabilities in the presence of correlated variability. Nat Neurosci 16:235-242.

Hafed ZM (2013) Alteration of visual perception prior to microsaccades. Neuron 77:775-786.

Hafed ZM, Clark JJ (2002) Microsaccades as an overt measure of covert attention shifts. Vision Res 42:2533-2545.

Hafed ZM, Ignashchenkova A (2013) On the Dissociation between Microsaccade Rate and Direction after Peripheral Cues: Microsaccadic Inhibition Revisited. J Neurosci 33:16220-16235.

Hafed ZM, Lovejoy LP, Krauzlis RJ (2011) Modulation of microsaccades in monkey during a covert visual attention task. J Neurosci 31:15219-15230.

Hamker FH, Zirnsak M, Calow D, Lappe M (2008) The peri-saccadic perception of objects and space. PLoS Comput Biol 4:e31.

Harvey BM, Dumoulin SO (2011) The Relationship between Cortical Magnification Factor and Population Receptive Field Size in Human Visual Cortex: Constancies in Cortical Architecture. J Neurosci 31:13604-13612.

Henson DB, Emuh T (2010) Monitoring vigilance during perimetry by using pupillography. Invest Ophthalmol Vis Sci 51:3540-3543.

Hernandez A, Nacher V, Luna R, Zainos A, Lemus L, Alvarez M, Vazquez Y, Camarillo L, Romo R (2010) Decoding a perceptual decision process across cortex. Neuron 66:300-314.

Hess EH, Polt JM (1960) Pupil size as related to interest value of visual stimuli. Science 132:349-350.

Hess EH, Polt JM (1964) Pupil Size in Relation to Mental Activity during Simple Problem-Solving. Science 143:1190-1192.

Hol K, Treue S (2001) Different populations of neurons contribute to the detection and discrimination of visual motion. Vision Res 41:685-689.

Horton JC, Hoyt WF (1991) The representation of the visual field in human striate cortex. A revision of the classic Holmes map. Arch Ophthalmol 109:816-824.

Hsieh PJ, Tse PU (2009) Microsaccade rate varies with subjective visibility during motion-induced blindness. PLoS One 4:e5163.

Hubel DH, Wiesel TN (1962) Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J Physiol 160:106-154.

Ibbotson M, Krekelberg B (2011) Visual perception and saccadic eye movements. Curr Opin Neurobiol 21:553-558.

Ivanov P, Blanche T (2011) Improving gaze accuracy and predicting fixation in real time with video based eye trackers. J Vis 11:505.

Jack AI, Shulman GL, Snyder AZ, McAvoy M, Corbetta M (2006) Separate Modulations of Human V1 Associated with Spatial Attention and Task Structure. Neuron 51:135-147.

Jazayeri M, Movshon JA (2006) Optimal representation of sensory information by neural populations. Nat Neurosci 9:690-696.

Jazayeri M, Movshon JA (2007) A new perceptual illusion reveals mechanisms of sensory decoding. Nature 446:912-915.

Jenkinson M, Bannister P, Brady M, Smith S (2002) Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17:825-841.

Kagan I, Gur M, Snodderly DM (2008) Saccades and drifts differentially modulate neuronal activity in V1: effects of retinal image motion, position, and extraretinal influences. J Vis 8:19 11-25.

Kahneman D, Beatty J (1966) Pupil diameter and load on memory. Science 154:1583-1585.

Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. Nature 452:352-355.

Kimmel DL, Mammo D, Newsome WT (2012) Tracking the eye non-invasively: simultaneous comparison of the scleral search coil and optical tracking techniques in the macaque monkey. Front Behav Neurosci 6:49.

Kolster H, Peeters R, Orban GA (2010) The retinotopic organization of the human middle temporal area MT/V5 and its cortical neighbors. J Neurosci 30:9801-9820.

Larsson J, Landy MS, Heeger DJ (2006) Orientation-selective adaptation to first- and second-order patterns in human visual cortex. J Neurophysiol 95:862-881.

Laubrock J, Engbert R, Kliegl R (2008) Fixational eye movements predict the perceived direction of ambiguous apparent motion. J Vis 8:13 11-17.

Law CT, Gold JI (2008) Neural correlates of perceptual learning in a sensory-motor, but not a sensory, cortical area. Nat Neurosci 11:505-513.

Law CT, Gold JI (2009) Reinforcement learning can account for associative and perceptual learning on a visual-decision task. Nat Neurosci 12:655-663.

Lee SH, Blake R, Heeger DJ (2005) Traveling waves of activity in primary visual cortex during binocular rivalry. Nat Neurosci 8:22-23.

Liu S, Gu Y, DeAngelis GC, Angelaki DE (2013) Choice-related activity and correlated noise in subcortical vestibular neurons. Nat Neurosci 16:89-97.

Martinez-Conde S, Macknik SL, Hubel DH (2000) Microsaccadic eye movements and firing of single cells in the striate cortex of macaque monkeys. Nat Neurosci 3:251-258.

Martinez-Conde S, Macknik SL, Hubel DH (2002) The function of bursts of spikes during

visual fixation in the awake primate lateral geniculate nucleus and primary visual cortex. Proc Natl Acad Sci U S A 99:13920-13925.

Martinez-Conde S, Macknik SL, Hubel DH (2004) The role of fixational eye movements in visual perception. Nat Rev Neurosci 5:229-240.

Mathur A, Gehrmann J, Atchison DA (2013) Pupil shape as viewed along the horizontal visual field. J Vis 13.

Merchant J, Morrissette R, Porterfield JL (1974) Remote measurement of eye direction allowing subject motion over one cubic foot of space. IEEE Trans Biomed Eng 21:309-317.

Merriam EP, Gardner JL, Movshon JA, Heeger DJ (2013) Modulation of visual responses by gaze direction in human visual cortex. J Neurosci 33:9879-9889.

Mon-Williams M, Tresilian JR, Plooy A, Wann JP, Broerse J (1997) Looking at the task in hand: vergence eye movements and perceived size. Exp Brain Res 117:501-506.

Murray SO, Boyaci H, Kersten D (2006) The representation of perceived angular size in human primary visual cortex. Nat Neurosci 9:429-434.

Nassar MR, Rumsey KM, Wilson RC, Parikh K, Heasly B, Gold JI (2012) Rational regulation of learning dynamics by pupil-linked arousal systems. Nat Neurosci 15:1040-1046.

Nestares O, Heeger DJ (2000) Robust multiresolution alignment of MRI brain volumes. Magn Reson Med 43:705-715.

Newsome WT, Britten KH, Movshon JA (1989) Neuronal correlates of a perceptual decision. Nature 341:52-54.

Nienborg H, Cumming BG (2006) Macaque V2 neurons, but not V1 neurons, show choice-related activity. J Neurosci 26:9567-9578.

Nienborg H, Cumming BG (2009) Decision-related activity in sensory neurons reflects more than a neuron's causal effect. Nature 459:89-92.

Nienborg H, Cumming B (2010) Correlations between the activity of sensory neurons and behavior: how much do they tell us about a neuron's causality? Curr Opin Neurobiol 20:376-381.

Nienborg H, Cohen MR, Cumming BG (2012) Decision-related activity in sensory neurons: correlations among neurons and with behavior. Annu Rev Neurosci 35:463-483.

Nishimoto S, Vu AT, Naselaris T, Benjamini Y, Yu B, Gallant JL (2011) Reconstructing

visual experiences from brain activity evoked by natural movies. Curr Biol 21:1641-1646.

Olman CA, Inati S, Heeger DJ (2007) The effect of large veins on spatial localization with GE BOLD at 3 T: Displacement, not blurring. Neuroimage 34:1126-1135.

Otero-Millan J, Macknik SL, Martinez-Conde S (2012) Microsaccades and blinks trigger illusory rotation in the "rotating snakes" illusion. J Neurosci 32:6043-6051.

Otero-Millan J, Castro JL, Macknik SL, Martinez-Conde S (2014) Unsupervised clustering method to detect microsaccades. J Vis 14.

Palmer C, Cheng SY, Seidemann E (2007) Linking Neuronal and Behavioral Performance in a Reaction-Time Visual Detection Task. J Neurosci 27:8122-8137.

Paradiso MA (1988) A theory for the use of visual orientation information which exploits the columnar structure of striate cortex. Biol Cybern 58:35-49.

Park SH, Cha K, Lee SH (2013) Coaxial anisotropy of cortical point spread in human visual areas. J Neurosci 33:1143-1156.

Parker AJ, Krug K, Cumming BG (2002) Neuronal activity and its links with the perception of multi-stable figures. Philos Trans R Soc Lond B Biol Sci 357:1053-1062.

Pestilli F, Carrasco M, Heeger DJ, Gardner JL (2011) Attentional enhancement via selection and pooling of early sensory responses in human visual cortex. Neuron 72:832-846.

Pouget A, Dayan P, Zemel R (2000) Information processing with population codes. Nat Rev Neurosci 1:125-132.

Preuschoff K, 't Hart BM, Einhauser W (2011) Pupil Dilation Signals Surprise: Evidence for Noradrenaline's Role in Decision Making. Front Neurosci 5:115.

Price NS, Born RT (2010) Timescales of sensory- and decision-related activity in the middle temporal and medial superior temporal areas. J Neurosci 30:14036-14045.

Purushothaman G, Bradley DC (2005) Neural population code for fine perceptual decisions in area MT. Nat Neurosci 8:99-106.

Ratliff F, Riggs LA (1950) Involuntary motions of the eye during monocular fixation. J Exp Psychol 40:687-701.

Regan D, Beverley KI (1985) Postadaptation orientation discrimination. J Opt Soc Am A 2:147-155.

Reingold EM (2014) Eye Tracking Research and Technology: Towards Objective Measurement of Data Quality. Vis cogn 22:635-652.

Ress D, Heeger DJ (2003) Neuronal correlates of perception in early visual cortex. Nat Neurosci 6:414-420.

Robinson DA (1963) A Method of Measuring Eye Movement Using a Scleral Search Coil in a Magnetic Field. IEEE Trans Biomed Eng 10:137-145.

Rolfs M, Kliegl R, Engbert R (2008) Toward a model of microsaccade generation: the case of microsaccadic inhibition. J Vis 8:5 1-23.

Romo R, Hernandez A, Zainos A, Lemus L, Brody CD (2002) Neuronal correlates of decision-making in secondary somatosensory cortex. Nat Neurosci 5:1217-1225.

Rosenbluth D, Allman JM (2002) The effect of gaze angle and fixation distance on the responses of neurons in V1, V2, and V4. Neuron 33:143-149.

Rynders M, Lidkea B, Chisholm W, Thibos LN (1995) Statistical distribution of foveal transverse chromatic aberration, pupil centration, and angle psi in a population of young adult eyes. J Opt Soc Am A Opt Image Sci Vis 12:2348-2357.

Salzman CD, Britten KH, Newsome WT (1990) Cortical microstimulation influences perceptual judgements of motion direction. Nature 346:174-177.

Sara SJ (2009) The locus coeruleus and noradrenergic modulation of cognition. Nat Rev Neurosci 10:211-223.

Schölvinck ML, Friston KJ, Rees G (2012) The influence of spontaneous activity on stimulus processing in primary visual cortex. Neuroimage 59:2700-2708.

Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. Science 268:889-893.

Shadlen MN, Britten KH, Newsome WT, Movshon JA (1996) A computational analysis of the relationship between neuronal and behavioral responses to visual motion. J Neurosci 16:1486-1510.

Sharma J, Dragoi V, Tenenbaum JB, Miller EK, Sur M (2003) V1 neurons signal acquisition of an internal representation of stimulus location. Science 300:1758-1763.

Sharon D, Grinvald A (2002) Dynamics and constancy in cortical spatiotemporal patterns of orientation processing. Science 295:512-515.

Shmuel A, Yacoub E, Chaimow D, Logothetis NK, Ugurbil K (2007) Spatio-temporal point-spread function of fMRI signal in human gray matter at 7 Tesla. Neuroimage 35:539-552.

Shtoyerman E, Arieli A, Slovin H, Vanzetta I, Grinvald A (2000) Long-term optical imaging and spectroscopy reveal mechanisms underlying the intrinsic signal and stability of cortical maps in V1 of behaving monkeys. J Neurosci 20:8111-8121.

Sirotin YB, Das A (2009) Anticipatory haemodynamic signals in sensory cortex not predicted by local neuronal activity. Nature 457:475-479.

Sirotin YB, Cardoso M, Lima B, Das A (2012) Spatial homogeneity and task-synchrony of the trial-related hemodynamic signal. Neuroimage 59:2783-2797.

Smith AM, Lewis BK, Ruttimann UE, Ye FQ, Sinnwell TM, Yang Y, Duyn JH, Frank JA (1999) Investigation of low frequency drift in fMRI signal. Neuroimage 9:526-533.

Smith JE, Zhan CA, Cook EP (2011) The functional link between area MT neural fluctuations and detection of a brief motion stimulus. J Neurosci 31:13458-13468.

Smith SM, Nichols TE (2009) Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44:83-98.

Snodderly DM, Kagan I, Gur M (2001) Selective activation of visual cortex neurons by fixational eye movements: implications for neural coding. Vis Neurosci 18:259-277.

Sperandio I, Kaderali S, Chouinard PA, Frey J, Goodale MA (2013) Perceived size change induced by nonvisual signals in darkness: the relative contribution of vergence and proprioception. J Neurosci 33:16915-16923.

Sylvester R, Haynes JD, Rees G (2005) Saccades differentially modulate human LGN and V1 responses in the presence and absence of visual stimulation. Curr Biol 15:37-41.

Tolhurst DJ, Movshon JA, Dean AF (1983) The statistical reliability of signals in single neurons in cat and monkey visual cortex. Vision Res 23:775-785.

Troncoso XG, Macknik SL, Martinez-Conde S (2008) Microsaccades counteract perceptual filling-in. J Vis 8:15 11-19.

Trotter Y, Celebrini S (1999) Gaze direction controls response gain in primary visual-cortex neurons. Nature 398:239-242.

Tse PU, Sheinberg DS, Logothetis NK (2004) The distribution of microsaccade directions need not reveal the location of attention. Psychol Sci 15:708-710.

Tse PU, Baumgartner FJ, Greenlee MW (2010) Event-related functional MRI of cortical activity evoked by microsaccades, small visually-guided saccades, and eyeblinks in human visual cortex. Neuroimage 49:805-816.

Uka T, DeAngelis GC (2004) Contribution of area MT to stereoscopic depth perception: choice-related response modulations reflect task strategy. Neuron 42:297-310.

Walsh G (1988) The effect of mydriasis on the pupillary centration of the human eye. Ophthalmic Physiol Opt 8:178-182.

Wildenmann U, Schaeffel F (2013) Variations of pupil centration and their effects on video eye tracking. Ophthalmic Physiol Opt 33:634-641.

Wilson MA, Campbell MC, Simonet P (1992) Change of pupil centration with change of illumination and pupil size. Optom Vis Sci 69:129-136.

Wyatt HJ (1995) The form of the human pupil. Vision Res 35:2021-2036.

Wyatt HJ (2010) The human pupil and the use of video-based eyetrackers. Vision Res 50:1982-1988.

Yang Y, Thompson K, Burns SA (2002) Pupil location under mesopic, photopic, and pharmacologically dilated conditions. Invest Ophthalmol Vis Sci 43:2508-2512.

Young LR, Sheena D (1975) Survey of eye movement recording methods. Behavior Research Methods & Instrumentation 7:397-429.

Yu AJ, Dayan P (2005) Uncertainty, neuromodulation, and attention. Neuron 46:681-692.

Yuval-Greenberg S, Merriam EP, Heeger DJ (2014) Spontaneous microsaccades reflect shifts in covert attention. J Neurosci 34:13693-13700.

국 문 초 록

지각판단 중 인간 1차시각피질의 역할 규명을 위한 뇌영상연구

서울대학교 대학원
자연과학대학 뇌인지과학과
최 경 환

1차시각피질(V1)은 대뇌피질 중 최초로 외부 환경의 사물과 사건에 대한 신경표상을 형성하는 곳이며, 이러한 정보를 시각정보처리 위계상 상위에 위치하는 여타 뇌영역으로 전달한다. 지난 수십여년의 연구를 통해 망막에 맺힌 영상자극이 V1에 어떠한 신경활동을 촉발하는지에 대해서는 잘 확립되었지만 그러한 V1반응이 지각판단과 어떤 상관이 있는지에 대해서는 아직 잘 알려져 있지 않다. 특히, 관찰자가 동일한 자극에 대해 매번 다른 지각판단을 내리는데 V1이 얼마나 기여하는지, 혹시 매 순간 다른 V1의 신경활동이 인과적으로 지각판단을 결정하는지에 대해서 활발하게 연구가 진행되고 있다. 이러한 맥락에서 본 연구는 V1의 신경활동이 과연 지각판단을 인과적으로 결정하는지 알아보기 위해 인과론이 필연적으로 예측하는 바- V1에 선택상관반응이 존재한다면 이는 자극상관반응과 동일한 시공간 상에

존재해야 함-를 실험을 통해 조사해봤다. 기능성 자기공명영상(fMRI)를 이용해서 어려운 시지각 판단과제를 수행하는 인간피험자의 V1 집단반응을 조사한 결과, V1에서 피험자가 내린 선택과 관련된 신경활동이 발견됐으나 이는 자극에 의한 신경활동에 비해 시간적으로 훨씬 앞서있고 공간적으로 다른 피질상에 위치해 있었다. 즉, 본 실험결과는 V1이 지각판단을 인과적으로 결정한다는 이론을 지지하지 않았다. V1의 선택관련 신경활동의 기원이 과연 안구운동에 의한 피드포워드 정보인지 혹은 상위피질의 하향피드백인지 알아보기 위해 시선추적기(eye-tracker)를 이용해서 안구운동을 측정하며 동일한 실험을 반복했다. 실험 결과 5가지 다른 미세안구운동은 모두 피험자의 선택과 무관했다. 본 실험결과를 종합해 볼 때, V1의 선택관련 신경활동은 지각판단 중 상위피질에서 매 시행별로 미리 계산되는 예상(prior information)이 V1에 하향피드백 됨으로써 발생하지만, 이러한 신경활동은 자극에 의해 촉발된 신경활동에 영향을 주지 않으며 따라서 지각판단 과정에는 기여하지 않는다.

주요어: 시지각, 지각판단, 1차시각피질, 기능성 자기공명영상, 시선·동공추적, 선택확률

학 번: 2010-31300