The effect of high-level image structure in early human visual cortex

Sung Jun Joo

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Reading Committee:

Scott O. Murray

Geoffrey M. Boynton

Steven Buck

Program Authorized to Offer Degree:

Psychology

University of Washington

Abstract

A commonly held view about neurons in early visual cortex is that they serve as localized filters. Here, however, we demonstrate that the responses of neurons in early visual cortex are sensitive to perceptual grouping based on global image structures. First we measured neural responses to an oriented Gabor ("target") embedded in various orientation patterns using multiple methodologies—psychophysics, functional magnetic resonance imaging (fMRI), and electroencephalography (EEG). Specifically, we varied whether a central target deviated from its context by changing distant orientations while leaving the immediately neighboring flankers unchanged. The results of psychophysical contrast adaptation and fMRI experiments show that a target that deviates from its context results in more neural activity compared to a target that is grouped into an alternating pattern. For example, the neural response to a vertically oriented target was greater when it deviated from the orientation of flankers (HH**V**HH) compared to when it was grouped into an alternating pattern (VH**V**HV). We then found that this pattern-sensitive response manifests in the earliest sensory component of the event-related potential to the target. In a forced-choice classification task of "noise" stimuli, perceptions are biased to "see" an orientation that deviates from its context. These results show that neurons in early visual cortex are sensitive to large-scale global patterns in images in a way that is more sophisticated than localized feature detection. Finally, we show using event-related potential (ERP) that the well-known orientation-specific surround suppression effect is dependent on the surface structure in the image, demonstrating an important role of high-level, global processes in determining when contextual effects occur in early visual cortex. Overall, our results demonstrate an important role of high-level, global processes in determining when contextual effects occur in early visual cortex.

The effect of high-level image structure in early Human Visual Cortex

Sung Jun Joo

Chair of Supervisory Committee:
Associate Professor Scott O. Murray
Department of Psychology

TABLE OF CONTENTS

LIST OF FIGURES

Figure Number

ACKNOWLEDGEMENTS

I would like to thank my advisors Drs. Scott Murray and Geoffrey Boynton. They have been always supportive and provided invaluable insights and advice for my research. I also thank the members of our lab—vision and cognition lab—for the friendship and being reliable research participants. Finally, I am grateful for my wonderful family: my wife Eun Kyoung and two lovely sons, Daniel and Sean. I could not have done this without them.

DEDICATION

To Eun Kyoung, my beloved wife and a great helper

INTRODUCTION

Neurons in early visual cortex have been assumed to act like *localized* filters that make measurements of a restricted region of an image (Adelson & Bergen, 1985; Carandini, Demb, Mante, Tolhurst, Dan, Olshausen, Gallant, & Rust, 2005; Jones & Palmer, 1987; Lennie & Movshon, 2005; Movshon, Thompson, & Tolhurst, 1978; Ringach, 2002; Rust & Movshon, 2005) without knowing what aspects of the image such as surface, edges, or object are being coded. This view naturally led to a hierarchical feedforward model of vision where local information processed by neurons in early visual cortex is integrated by neurons in higher stages of visual processing with large receptive fields (RFs) to process global information (Riesenhuber & Poggio, 1999; Riesenhuber & Poggio, 2000).

However, this *localized* principle has been challenged by numerous contextual modulation studies (Blakemore & Tobin, 1972; Cavanaugh, Bair, & Movshon, 2002a; Cavanaugh, Bair, & Movshon, 2002b; DeAngelis, Freeman, & Ohzawa, 1994; Levitt & Lund, 1997; Maffei & Fiorentini, 1976; Sillito, Grieve, Jones, Cudeiro, & Davis, 1995) where contextual stimuli placed outside the RF of V1 neurons modulate the neural response to the stimuli placed inside the RF. The modulatory signal is, in general, suppressive (but also see Kapadia, Ito, Gilbert, & Westheimer, 1995; Polat, Mizobe, Pettet, Kasamatsu, & Norcia, 1998) and orientation-tuned and thus contextual modulation is often subsumed by the term “orientation-specific surround suppression.”

Orientation-specific surround suppression has been suggested to be the neural mechanism underlying various high-level, perceptual processes such as figure-ground segregation (Zipser, Lamme, & Schiller, 1996), contour integration (Kapadia et al., 1995; Roelfsema, Lamme, & Spekreijse, 2004), and saliency (Kastner, Nothdurft, & Pigarev, 1997; Knierim & van Essen, 1992). However, many of the apparently complex contextual effects seen in early visual areas can be accounted for through simple, local spatial summation and normalization processes (Cavanaugh et al., 2002a; Shapley, 2004). In addition, many of the studies that have examined global, perceptually-based explanations for surround effects have confounded changes in local stimulation in V1 and perception. For example, most of contextual modulation studies have changed the orientation relationship between a target and immediately "nearby" flankers, and thus it is not possible to determine whether modulations of neural activity are due to changes in the stimulus configuration affecting low-level summation and normalization processes (Cavanaugh et al., 2002a; Shapley, 2004) or are due to changes in global attributes of the image, or both.

The stimuli used in contextual modulation studies can be characterized by a more general description—whether or not the center can be grouped with the surround. For example, the center can be grouped with the surround when the surround matches the center orientation compared to when the surround is orthogonally oriented to the center. Grouping is one of the important tasks in vision because the visual system must decide to which part of the image the local element belong in order to segregate objects from background. How does grouping affect the neural activity in early visual cortex? The answer to this question is crucial

to understand how the visual system is organized to process visual information in real scenes with cluttered objects. However, the results of contextual modulation studies are not converging: on one hand, there is suppression (Bair, Cavanaugh, & Movshon, 2003; Blakemore & Tobin, 1972; Cavanaugh et al., 2002a; Cavanaugh et al., 2002b; Haynes, Roth, Stadler, & Heinze, 2003; Sillito et al., 1995; Zenger-Landolt & Heeger, 2003) when a stimulus can be grouped with the surround, and the other hand, there is enhancement (Kapadia et al., 1995; Kourtzi, Tolias, Altmann, Augath, & Logothetis, 2003; Polat et al., 1998).

We used multiple methodologies—psychophysics, functional magnetic resonance imaging (fMRI), and electroencephalography (EEG)—to study how grouping affects the neural activity in early visual cortex and distinguish the role of local and global processes in surround suppression in early visual cortex.

First, in Chapter I, we measured neural responses to the center ("target") with a flanker above and below the target to establish orientation-specific surround suppression. We used two target orientations—vertical and horizontal—to study whether there is general neural suppression regardless of the target orientation or there is neural enhancement to the vertical target with vertical flankers (collinear facilitation).

Second, in Chapter II, we varied whether a central target was part of a pattern by changing the orientation of distant gratings while maintaining the same local stimulus arrangement (where "local" refers to a central target and its immediate surrounding flankers). For example, a vertically oriented target grating that is flanked locally with horizontal flankers (HVH), can either be part of a repeating pattern of orientations by adding vertical distant

flankers (VHVHV), or can deviate from a pattern by adding horizontal distant flankers (HHVHH). If the surround suppression occurs due to the presence of "nearby" iso-oriented surrounding stimuli, the neural response to the target (V) would be the same between conditions (V**HVH**V vs. H**HVH**H). On the other hand, if the surround suppression is related to high-level, perceptual processes, we expect that there would be neural suppression in one condition (VHVHV) compared to the other (HHVHH) because the target (V) is part of an orientation sequence and grouped with surrounding orientations in VHVHV whereas it deviates from the pattern of surrounding orientations in HHVHH.

Third, in Chapter III, we manipulated perceptual grouping by varying the distance between the target and flankers and placing them in the same or different surface. For example, a target can be isolated when it is surrounded by far-removed flankers (same surface) and when nearby flankers are displayed on different surfaces than the target. However, a target can be grouped into an array when it is surrounded by nearby flankers (same surface) and when far-removed flankers are displayed on the same surface as the target.

The result of these experiments suggest that there is neural suppression when the flankers match the target orientation regardless of the target orientation, and that high-level image structures such as the pattern of orientations and perceived surfaces modulate contextual effects in early visual cortex. These findings are not consistent with the standard model of early visual processing (Adelson & Bergen, 1985; Carandini et al., 2005; Jones & Palmer, 1987; Lennie & Movshon, 2005; Movshon et al., 1978; Ringach, 2002; Rust & Movshon,

2005), and point to a coding strategy in early visual cortex that is sensitive to global attribute of an image.

Chapter I.

Orientation-specific surround suppression in early human visual cortex

The receptive field (RF) of a neuron in primary visual cortex (V1) is defined as the region of retinotopic visual space that elicits an increase in spike rate in response to the presentation of a stimulus. While stimulation outside the RF alone does not increase spike rate, stimuli placed outside the RF ("flankers") can modulate the response to a stimulus placed in the RF ("target"). Though contextual effects are well known, there is little consensus on their specific nature, functional role, or underlying mechanisms. However, most of the reported contextual modulation experiments can be segregated into two broad categories: (1) those reporting enhancement—often referred to as "collinear facilitation" and (2) those reporting suppression—often referred to as "surround suppression."

Collinear facilitation experiments have demonstrated increased behavioral sensitivity and increased spike rate when a target is flanked by elements that have orientations that align with the target orientation. Collinear facilitation is most often observed when the distance between the target and flankers is small, with target-flanker center-to-center separations on the order of 0.15 degrees in visual angle with measurable effects out to approximately 1 degree in visual angle (Kapadia et al., 1995; Kapadia, Westheimer, & Gilbert, 2000; Polat et al., 1998; Polat & Sagi, 1993; Polat & Sagi, 1994). The stimuli are usually small bars or Gabor patches. Though these experiments claim to be placing stimuli outside the RF of the neuron, most used

the minimum response field technique which yields the smallest estimate of RF size (Angelucci & Bressloff, 2006). Thus, it is likely that the flankers occupied portions of the high-threshold spiking RF "fringe" regions that, while not directly eliciting neural responses, could facilitate responses produced by centrally placed stimuli (Cavanaugh et al., 2002a).

In contrast to RFs based on the minimum response field technique, RF estimates based on the summation field technique are much larger (by up to a factor of 2) and include both low and high-threshold spiking regions (Angelucci & Bressloff, 2006). Placing stimuli outside the summation field yield contextual modulation effects that are opposite of those observed for collinear facilitation: suppression that is largest when the flankers match the target orientation. Less suppression and sometimes facilitation is observed for orthogonal flankers (Cavanaugh et al., 2002a; DeAngelis et al., 1994; Levitt & Lund, 1997; Sceniak, Hawken, & Shapley, 2001; Sillito et al., 1995). Experiments reporting surround suppression often use disk and annulus shaped gratings as the targets and flankers, respectively.

Because of the variety of stimulus configurations and inconsistencies across different measurement methodologies (e.g., Kinoshita, Gilbert, & Das, 2009), it is difficult to draw definitive conclusions about the nature of contextual effects. Here we used stimuli that are often used in collinear facilitation experiments (isolated Gabor patches) and relatively large target-flanker distances such that flankers are likely located outside of the summation field of V1 neurons. We observed no evidence of collinear facilitation but show a remarkably consistent pattern of results across four different methodologies—significant suppression when the flanker orientation matches the target orientation.

METHODS

All of the experiments described below examine the neural response to a single parafoveal Gabor stimulus (target) flanked by two Gabors (flankers), one positioned above and the other positioned below the target (dimensions specified below). Across all methodologies we compared three conditions: the target presented alone (*single*), the target with flankers that matched the target orientation (*same*), and the target with flankers that were orthogonal to the target orientation (*orthogonal*). There were two target orientation conditions (vertical and horizontal, Figure 1). The *same* condition with a vertical target specifically examines possible collinear facilitation effects. In the single unit experiments the target orientation was chosen to match the preferred orientation of the cell being recorded.

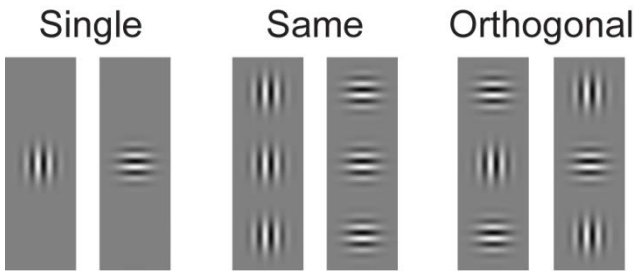

Figure 1. Schematic of the three types of stimulus arrangements and two target orientations. Note that the vertical target orientation tests for collinear facilitation effects.

Functional MRI

Six observers including the authors participated. The Gabor patches had a standard deviation of 0.72°, a spatial frequency of 2 cycles per degree (cpd), and were 50% contrast. The

fixation point was placed 3° below from the center of the display. The target stimuli were displayed in the upper quadrant of both visual fields 3° horizontally and 3° vertically from the fixation point. The center-to-center distance between the stimuli comprising a pattern was 3°.

One experimental session consisted of 2 localizer scans and 8 stimulus scans. A localizer scan (252 s) consisted of two conditions each presented in 12 s blocks: (1) fixation ("F"), (2) stimulus block with checkerboards in the location of the flankers ("S"). The fixation condition and stimulus condition were alternated during the scan (S-F, etc). The checkerboards were counter-phased flickered (6 Hz) and windowed using the same Gaussian envelope used for Gabor stimuli in the main experiment.

A stimulus scan (294 s) consisted of four conditions each presented in 12 s blocks: (1) fixation, (2) *single* (**V** or **H**), (3) *same* (V**V**V and H**H**H), and (4) *orthogonal* (H**V**H and V**H**V). There were ten repetitions per each condition in a stimulus scan. The fixation condition was inserted between the three stimulus conditions (*same*-F-*orthogonal*-F-*alternating*-F, etc). The center orientation alternated between vertical and horizontal in successive stimulus scans. During a stimulus block, flankers were counterphased flickered at 0.5 Hz to preclude fading effects and the target Gabors were flashed on (250 ms) and off (250 ms). The phase of target Gabor was always in-phase with the flanker Gabors. A contrast decrement task was used throughout each stimulus scan to equate attentional state across the conditions. We found no significant differences in both the performance (87.5%, 88.4%, and 82.7% for single, same, and orthogonal condition, respectively; a repeated measures ANOVA, $F(2, 10) = 2.004$, $p = 0.185$) and reaction

times (406 ms, 419 ms, and 418 ms for same, orthogonal, and alternating condition, respectively; $F(2, 10) = 0.535$, $p = 0.601$).

Functional MRI data were acquired using a Philips Achieva 3T scanner using an 8-channel head coil and an echo-planar imaging sequence (repetition time, 2 s; flip angle, 70°; 31 axial slices of 3.5 mm thickness (no gap) and 3.44 × 3.44 mm resolution, field of view, 220 mm). Each scanning session began with a T1-weighted structural scan 1 × 1 × 1 mm used for visualization of retinotopic visual areas. Visual cortical V1 was localized using standard retinotopic mapping and cortical-flattening techniques using BrainVoyager QX. Regions of interest (ROIs) within V1 were determined using the localizer scan described above. Those voxels with a significantly larger response in a comparison between the target and flanker position and confined to the retinotopic location of V1 were included in further analyses.

Time courses for each of the stimulus scans were extracted and averaged across voxels within V1. For each scan, the signal intensity in each condition at each of 8 time points time-locked to the onset of the stimuli was averaged, including 4 s before the onset to 10 s after the onset. The two timepoints (4 s) before trial onset served as baseline. The averaged time courses of signal intensity were converted to percent signal change by subtracting the corresponding baseline measurement and then dividing by that value. The resulting time course for each condition was then averaged across scans. The signal intensity averaged from 4s to 10s post-stimulus presentation was used as the measured response for each condition. This single number was then averaged across observers and a repeated-measures ANOVA and planned-comparison paired t-tests were used for statistical evaluation.

Event-related potentials

Eighteen observers including the author participated. Except for the author, all were naïve observers. They were paid $20 per hour for participation. The stimuli were generated and controlled by Presentation (Neurobehavioral Systems, Inc.), and they were displayed on a 21-in. CRT monitor (60 Hz refresh rate). The viewing distance was approximately 70 cm. The spatial stimulus arrangement was identical to the fMRI experiment. On a given trial, flankers appeared before the target onset and remained throughout the trial. After a random duration chosen from a uniform distribution between 1 and 2 s, targets were briefly flashed for 100 ms. After the target offset, the flankers remained in the display for 500 ms. The inter-trial interval was 2 s. Observers were asked to maintain fixation and to limit eye blinks to the inter-trial interval.

One experimental block consisted of 12 trials [2 orientation (vertical and horizontal) x 3 stimulus conditions (*single*, *same*, and *orthogonal*) x 2 repetitions]. They were randomized within a block. Observers finished 22-51 blocks (264-612 trials). Observers initiated each block after a 5 s break by pressing a designated key in the button box. The first block served as practice.

EEG waveforms were recorded using BioSemi active Ag-AgCl electrodes from 64 sites. The signals were referenced to the left mastoid during online acquisition and re-referenced to the average of right and left mastoids offline. Vertical EOG was measured using an electrode placed below the left eye and horizontal EOG was measured using an electrode placed at the outer canthus of the right eye. The signals were digitized at a sampling rate of 256 Hz.

EEG epochs started 100 ms before the target onset and lasted 400 ms after the target onset. Each waveform was baseline corrected to the average voltage of the interval -100 ms to 0 ms before the target onset and low-pass filtered at 40 Hz to remove high-frequency noise. Trials with waveforms that had a larger than 50 μV peak-to-peak verticalEOG amplitude and that exceeded ±50 μV on other electrodes were excluded as these were trials deemed to be contaminated with eye blinks or other sources of noise. Data from 5 observers were discarded due to excessive artifact rejection (> 40%). The resulting waveforms were averaged across conditions individually for statistical analyses and then averaged across observers for figures.

P1 amplitude on six electrodes (Oz, O1, O2, POz, PO3, and PO4) was measured by averaging the ERP amplitudes during the time window of 130 ms to 170 ms. These electrodes were centered over the maximum of the P1 component as determined through visual inspection of the scalp topography. These individual amplitudes were averaged across six electrodes to represent P1 amplitude.

Psychophysical contrast adaptation

Seven observers participated in each target orientation (vertical and horizontal) version. The first author participated in both target orientation versions. Observers participated in two sessions (left and right visual field). The experiments were conducted in a dark room and head position was stabilized using a chin rest. Stimuli were presented on a 19-inch linearized CRT monitor with vertical refresh rate of 60 Hz. The distance between observers and the monitor was 50 cm. We used a video attenuator device (Video Switcher, Xiangrui Li, Los Angeles, CA) to generate 10-bit gray-scale luminance values (Li, Lu, Xu, Jin, & Zhou, 2003). We used the

MATLAB Psychtoolbox (Brainard, 1997; Pelli, 1997) on a PC to create stimuli, to control stimulus presentation, and to record responses.

Stimuli consisted of Gabor patches (σ = 0.72°, spatial frequency = 2 cpd, sinusoidal counterphase-flickering at 2 Hz). A fixation point (0.48° in diameter) was displayed at the center of the display. The distance between the fixation point and the center of the target location was 6°. To measure the contrast detection threshold for a target, we used two randomly interleaved independent QUEST (Watson & Pelli, 1983) staircases. The detection task was a two-interval forced choice (2IFC) task where observers indicated which interval had the target. Each interval (200 ms) was indicated by a high-pitched tone and there was a gray blank (300 ms) between intervals. Auditory feedback was given for an incorrect response. There were 41 trials (20 trials per each staircase) in a session. The response of the first trial was discarded. The last contrast values of two staircases were averaged to estimate an observer's contrast detection threshold for 82% performance. If the two staircases did not converge observers were asked to re-run the session.

Adapting stimuli consisted of a target and flankers. The center-to-center distance between stimuli was 3°. Observers were initially adapted for 30 s, followed by the first 2-IFC task trial. A 5 s top-up adaptation period was inserted between subsequent trials to maintain stable adaptation. A 500 ms gray blank was inserted before each trial began. During this blank period, a black line was displayed next to the fixation point to indicate the beginning of a trial. The target location was always marked during both the adaptation and task periods to remove effects of location uncertainty on the detection task (Petrov, Verghese, & McKee, 2006). To

equate the attentional state across conditions observers performed a contrast decrement task on the fixation mark during adaptation periods (Bi, Cai, Zhou, & Fang, 2009). The contrast decrement (10%) was displayed for 150 ms and the onset of the contrast decrement was selected randomly from a uniform distribution between 1.5 and 2 s. To quantify the amount of adaptation we defined the threshold ratio between detection threshold before and after adaptation ($threshold_{after}/threshold_{before}$). We measured the amount of adaptation with two different target orientations (vertical, horizontal) embedded in three conditions: *single*, *same*, *and orthogonal*. The target and the flanker contrasts were set to 50% and 25%, respectively.

RESULTS

Functional MRI

The fMRI response in V1 was measured in the retinotopic location of the central target which was localized in a separate scan that alternated between flickering checkerboards in the target location and flanker location (Figure 2A). The responses of the target-ROI were well localized—the checkerboards in the flanker location did not evoke an fMRI response significantly above the response to a blank screen fixation period (Figure 2B).

A repeated measures ANOVA revealed that there was no significant main effect of target orientation ($F(1, 5) = 0.19$, $p = 0.681$) or an interaction ($F(2, 10) = 1.158$, $p = 0.353$). However, there was a significant effect of flanker condition ($F(2, 10) = 9.432$, $p = 0.005$), indicating contextual effects on the target response.

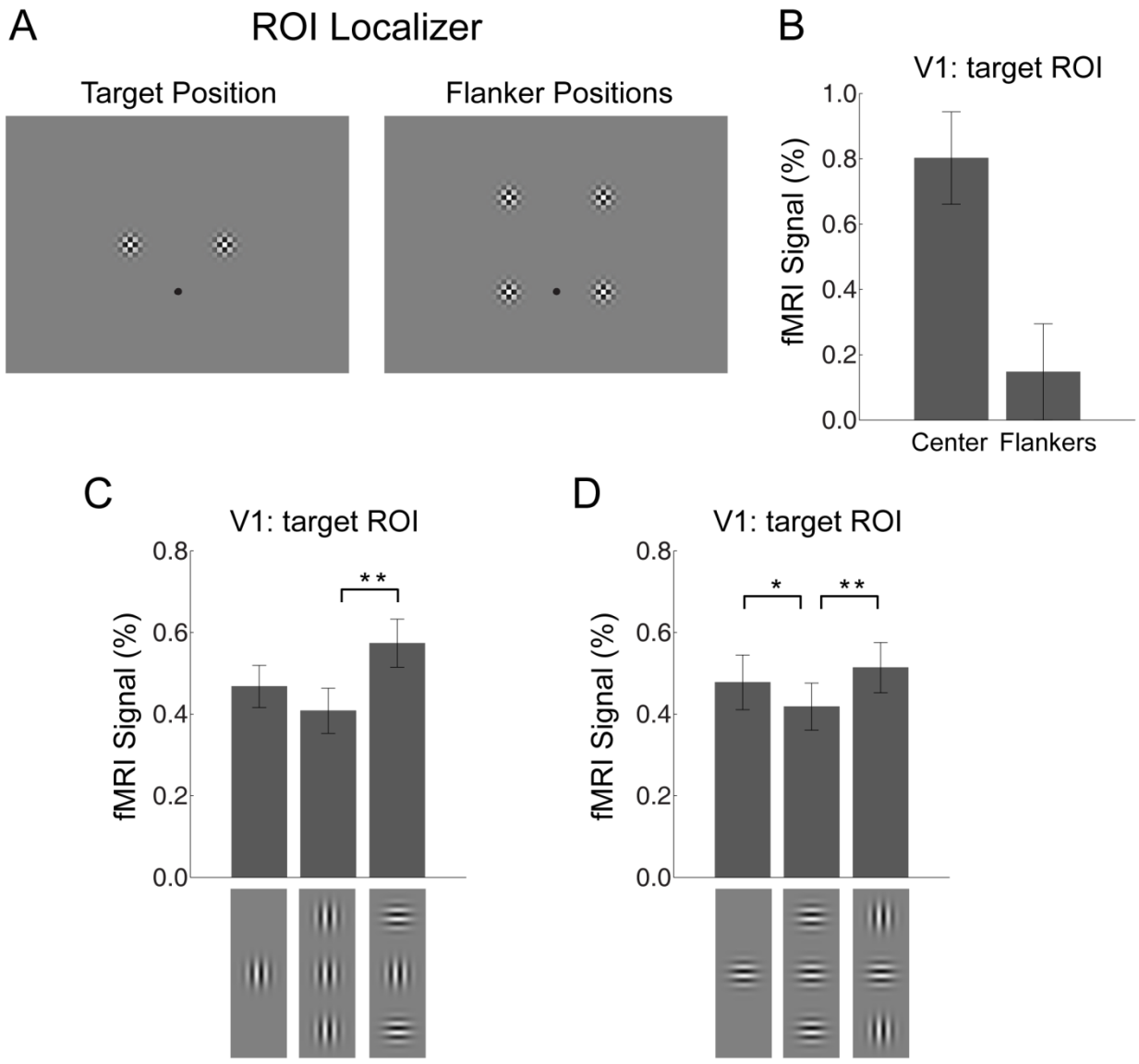

Figure 2. fMRI experiment. (A) ROI localizer. The localizer scan alternated flickering checker boards in target and flanker positions. (B) fMRI signal in V1 target ROI. (C) fMRI signal in V1 target ROI for the vertical target. (D) fMRI signal in V1 target ROI for the horizontal target. Error bars represent the standard error of the mean (SEM) across observers. * $p < 0.05$; ** $p < 0.01$.

We separately analyzed each target orientation. Figures 2C and 2D show the fMRI response to the vertical and horizontal target in the different flanker configurations, respectively. With both horizontal and vertical targets there was a significantly smaller

response in the *same* compared to the *orthogonal* condition ($F(1, 5) = 17.813$, $p = 0.008$ for vertical target; $F(1, 5) = 14.234$, $p = 0.013$ for horizontal target). Furthermore, there was significant suppression in the *same* compared to *single* condition with a horizontal target ($F(1, 5) = 6.723$, $p = 0.049$). These results are consistent with previous electrophysiology findings demonstrating orientation-tuned surround suppression. The fact that we observed suppression in the vertical target condition in the *same* versus *orthogonal* condition is not consistent with findings of collinear facilitation.

Event-related potentials

We also performed an ERP experiment using similar stimulus configurations to examine the timing of the effect. Our prediction was that if the flankers preceded the onset of the targets by a sufficient length of time, the suppression mechanism (e.g., via lateral connections and/or feedback) would be in place and stabilized before the onset of the target. Consequently, any differential response we observed between the *same* and *orthogonal* conditions should appear early in the ERP response to the target.

The ERP response was measured to the onset of the target in the three different flanker arrangements. Figure 3A shows the timeline for each trial. Importantly, the flankers were presented first for a random duration between 1 and 2 seconds, followed by the target for 100 ms. Again, the flankers preceding the target was done to "configure" the surround suppression mechanism before target onset. The flankers remained for an additional 500 ms so that their offset would not contaminate the ERP response to the target.

The first prominent ERP component that was observed in response to the target was a positive deflection that peaked at approximately 150 ms (P1). Previous research has suggested that the P1 originates from early extrastriate visual areas (Clark, Fan, & Hillyard, 1994; Di Russo, Martinez, Sereno, Pitzalis, & Hillyard, 2002). The C1 component—which is believed to originate in V1 (Clark et al., 1994; Di Russo et al., 2002)—was not observed in our data. The scalp distribution of the P1 was confined to the occipital pole (Figure 3B). To characterize the magnitude of the P1, signals from the six electrodes around the occipital pole (Oz, O1, O2, POz, PO3, and PO4) were averaged. We calculated the magnitude of the P1 as the average amplitude between 130 ms and 170 ms. A repeated measures ANOVA showed no significant effect of orientation ($F(1, 12) = 0.995$, $p = 0.338$) or interaction ($F(2, 24) = 0.354$, $p = 0.705$). Figures 2C and 2D show the average waveforms and P1 amplitudes for each target orientation in each flanker conditions. Planned contrasts showed that the P1 amplitude was smaller for the *same* condition compared to *orthogonal* ($F(1,12) = 5.607$, $p = 0.036$ for vertical target; $F(1,12) = 12.633$, $p = 0.004$ for horizontal target) and *single* conditions ($F(1,12) = 11.911$, $p = 0.005$ for vertical target; $F(1,12) = 23.160$, $p < 0.001$ for horizontal target). These results indicate that when the flankers precede the presentation of the target, the initial processing of the target is affected by whether the target orientation matches the flanker orientation.

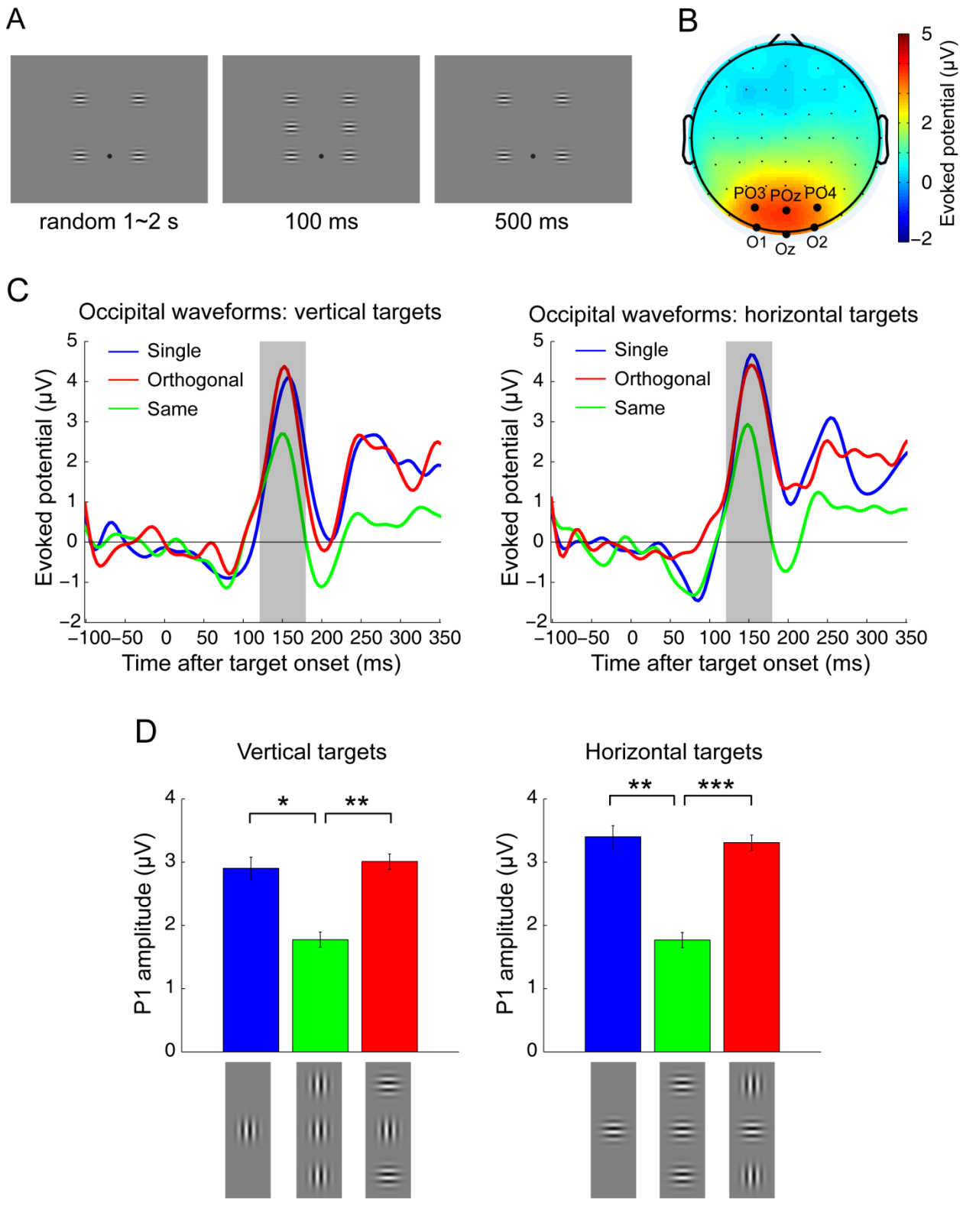
A
random 1~2 s
100 ms
500 ms
B
PO3
POz
PO4
O1
Oz
O2
5
2
0
−2
Evoked potential (μV)
C
Occipital waveforms: vertical targets
Occipital waveforms: horizontal targets
Single
Orthogonal
Same
Evoked potential (μV)
Time after target onset (ms)
−100 −50 0 50 100 150 200 250 300 350
D
Vertical targets
Horizontal targets
P1 amplitude (μV)
*
**
***

Figure 3. ERP experiment. (A) The experimental procedure. On a given trial, flankers appeared before target onset. After a random duration 1~2 s, the target was flashed for 100 ms. After target offset, flankers remained in the display for 500 ms. (B) Scalp topographical map at 150 ms after target onset. The color bar represents the amplitude of evoked potential. The ERP from six occipital electrodes (Oz, O1, O2, POz, PO3, and PO4) were averaged to define the signal. (C) ERP waveforms for each target orientation: vertical target in left panel and horizontal target in right panel. Blue, green, and red lines represent ERPs for *single*, *same*, and *orthogonal* condition, respectively. (D) P1 amplitudes for each target orientation: vertical target in left panel and horizontal target in right panel. Blue, green, and red bars represent P1 amplitude for *single*, *same*, and *orthogonal* condition, respectively. Error bars represent the SEM across observers. * $p < 0.05$; ** $p < 0.01$.

Psychophysical contrast adaptation

Across 3 separate methodologies (human fMRI, human ERP, monkey electrophysiology) we observed smaller neural responses to the target when it matched the orientation of the flankers. An important question is whether this neural modulation has direct perceptual consequences. We measured psychophysical contrast detection after adaptation to the different target configurations. Figure 4A shows a schematic of the adaptation procedure. Observers were initially adapted to a stimulus for 30 s before performing a two-interval forced choice (2IFC) detection task. A 5 s top-up adaptation period was inserted between trials to maintain stable adaptation (Figure 4A). The target location was always marked during both the adaptation and task periods to remove any effects of location uncertainty on the detection task (Petrov et al., 2002). To equate the attentional state across conditions observers performed a contrast decrement detection task on the fixation mark during adaptation periods (Bi et al., 2009).

To quantify adaptation strength, we calculated the ratio of each observer's contrast detection threshold for a target before and after adaptation. Our assumption is that more

adaptation—as indexed by an increase in post-adaptation detection thresholds—reflects stronger neural activity in response to the adapting stimulus (Blake, Tadin, Sobel, Raissian, & Chong, 2006; Blakemore & Campbell, 1969; Carandini, Movshon, & Ferster, 1998; Dragoi, Sharma, & Sur, 2000; Engel, 2005; Larsson, Landy, & Heeger, 2006; Priebe, Churchland, & Lisberger, 2002).

We found less adaptation in the *same* compared to the *orthogonal* condition, (paired two-tailed t test, $t(6) = 6.145$, $p < 0.001$ for vertical target; $t(6) = 2.801$, $p = 0.031$ for horizontal target; Figure 4B]. Furthermore, there was less adaptation with the vertical target in the *same* condition compared to the *single* condition ($t(6) = 2.998$, $p = 0.024$). In an additional experiment we found that the horizontal and vertical flankers alone (no stimulus in the target position) did not produce any adaptation, indicating that differences in adaptation observed between the same and orthogonal condition were due to differences in the orientation relationship between the target and flankers and not due to differences in the flankers themselves (Figure 4C).

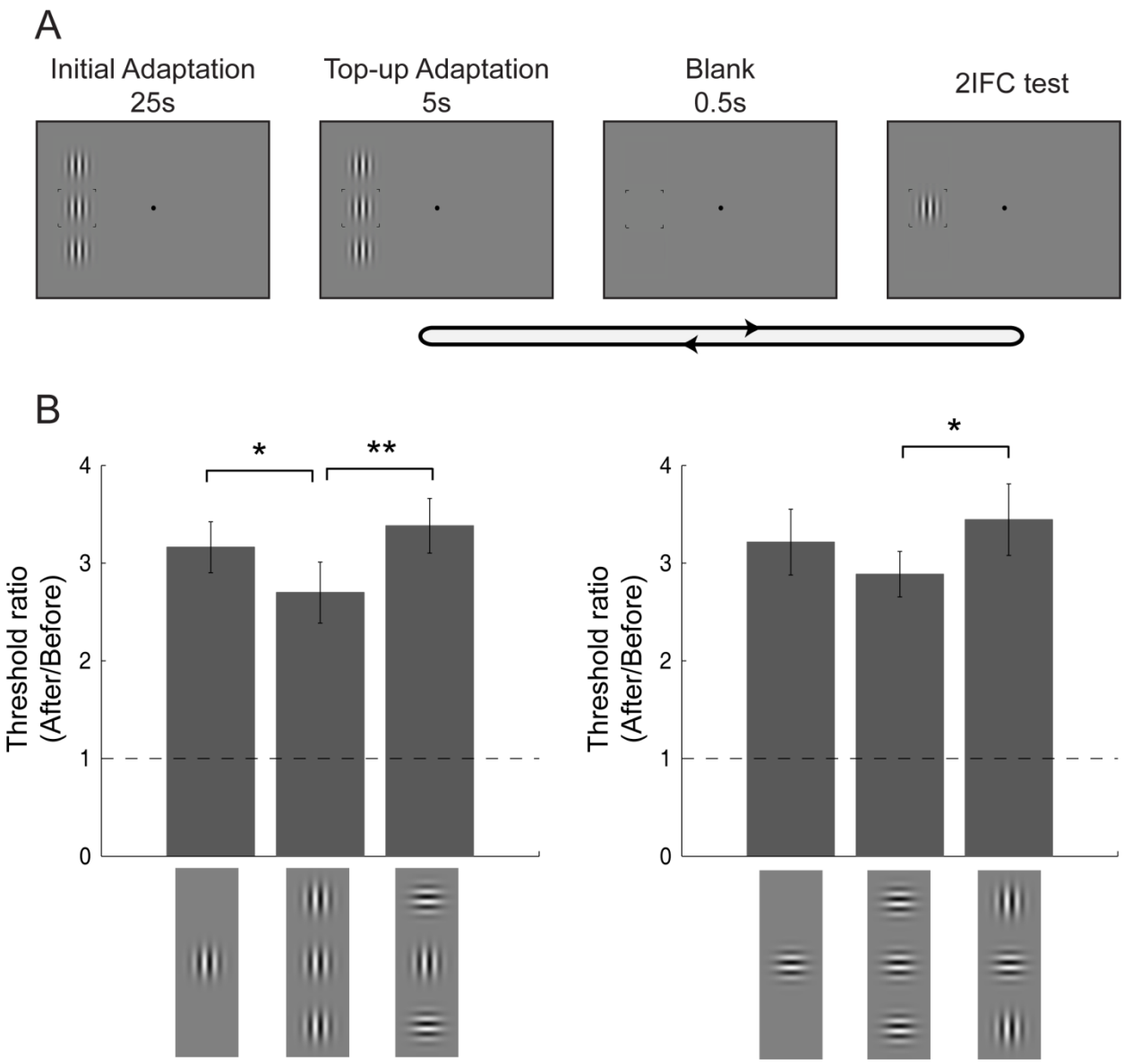

Figure 4. Contrast adaptation experiment. (A) The experimental procedure. Observers initially adapted to the adapting stimuli for 30 s. The detection threshold was measured using 2-IFC task trials. A 5 s top-up adaptation period was inserted between trials to maintain adaptation level. (B) Threshold ratio for each target orientation: vertical target in left panel and horizontal target in right panel. Dotted line indicates no adaptation. Error bars represent the SEM across observers. * $p < 0.05$; ** $p < 0.01$.

DISCUSSION

We observed across multiple methodologies a reduced neural response when the orientation of a pair of flankers matched the target orientation. In particular, we observed no evidence for response enhancement induced by the flankers—for example, similar to that described as "collinear facilitation."

Most of psychophysical experiments that study collinear facilitation measured detection threshold with and without flankers. What is typically observed is that the detection threshold of the target with collinear flankers decreases compared to the detection threshold of the target alone (Polat & Sagi, 1993; Polat & Sagi, 1994), and it is used as evidence for collinear facilitation. However, this interpretation might be problematic because the enhancement in sensitivity of a stimulus at near threshold might be related to position uncertainty of the target (Petrov et al., 2006). In our adaptation experiment we tried to preclude the effect of position uncertainty by presenting a position cue during adaptation and test periods.

The electrophysiology studies that observed collinear facilitation effect (Kapadia et al., 1995; Polat et al., 1998) used so-called minimum response field (MRF; Barlow, Blakemore, & Pettigrew, 1967) as a RF. In those experiments the fact that the flankers presented outside the MRF did not drive the cells to fire was used as evidence for stimulating outside the RF. However, the region outside the MRF can be a subthreshold region that modulates the neural response (Angelucci & Bressloff, 2006; Angelucci, Levitt, Walton, Hupe, Bullier, & Lund, 2002; Cavanaugh et al., 2002a). Furthermore, the RF size measured by the summation field technique that yields

the largest RF size (2.3-fold larger than the MRF size) depends on the stimulus contrast—the RF is 2.2-fold larger for low contrast stimulus compared to the RF for high contrast stimulus (Sceniak, Ringach, Hawken, & Shapley, 1999). Thus, in collinear facilitation experiments it is likely that the flankers encroaches the subthreshold regions of a RF, and indeed neural enhancement was observed for a low contrast target whereas neural suppression was found for a high contrast target in collinear stimulus configurations (Polat et al., 1998).

Our results are most consistent with contextual modulation experiments that observed orientation-specific surround suppression by placing contextual stimuli outside the summation field (Cavanaugh et al., 2002a; DeAngelis et al., 1994; Levitt & Lund, 1997; Sceniak et al., 2001; Sillito et al., 1995). In addtion, the results across different methodologies are remarkably similar. So, we conclude that we have established experimental procedures using multiple methodologies to measure contextual modulation that resembles the well-established surround suppression measured at individual neuron level.

As discussed in Introduction, contextual modulation measured using “nearby” flankers can be explained by a number of potential accounts—ranging from local normalization models (Cavanaugh et al., 2002a; Shapley, 2004) to our high-level, grouping hypothesis. In the next two Chapters we tested our hypothesis using the same experimental procedures in multiple methodologies.

Chapter II.

Long-range, pattern-dependent contextual effects in early human visual cortex

In separate experiments we measured neural responses using psychophysical contrast adaptation, human fMRI, and event-related potentials (ERPs) to an oriented Gabor stimulus ("target") embedded in various visual patterns as defined by the relative orientation of flanking stimuli. Specifically, we varied whether a central target deviated from its context by changing the orientation of distant gratings while leaving the immediately neighboring flankers unchanged. For example, we hypothesized that the neural response to a vertically oriented target grating would be greater when it deviated from the orientation of its flankers (horizontal flankers, HH**V**HH) compared to when it was grouped into an alternating sequence of orientations (VH**V**HV). Keeping the local orientation configuration around the target the same across conditions (e.g., H**V**H) eliminates effects of simple spatial summation (Cavanaugh et al., 2002a; Cavanaugh et al., 2002b; Shapley, 2004).

METHODS

All observers had normal or corrected-to-normal vision and all gave informed written consent in accordance with the University of Washington Institutional Review Board.

Observers and general setup for behavioral experiments

The experiments were conducted in a dark room and head position was stabilized using a chin rest. Stimuli were presented on a 19-inch linearized CRT monitor with vertical refresh rate of 60 Hz. The distance between observers and the monitor was 50 cm. We used a video attenuator device (Video Switcher, Xiangrui Li, Los Angeles, CA) to generate 10-bit gray-scale luminance values (Li et al., 2003). We used the MATLAB Psychtoolbox (Brainard, 1997; Pelli, 1997) on a PC to create stimuli, to control stimulus presentation, and to record responses.

Adaptation experiments

Stimuli consisted of Gabor patches (σ = 0.72°, spatial frequency = 2 cycles per degree (cpd), sinusoidal counterphase-flickering at 2 Hz). A fixation point (0.48° in diameter) was displayed at the center of the display. The distance between the fixation point and the center of the target location was 6°. To measure the contrast detection threshold for a target, we used two randomly interleaved independent QUEST (Watson & Pelli, 1983) staircases. The detection task was a two-interval forced choice (2IFC) task where observers indicated which interval had the target. Each interval (200 ms) was indicated by a high-pitched tone and there was a gray blank (300 ms) between intervals. Auditory feedback was given for an incorrect response. There were 41 trials (20 trials per each staircase) in a session. The response of the first trial was discarded. The last contrast values of two staircases were averaged to estimate an observer's contrast detection threshold for 82% performance. If the two staircases did not converge observers were asked to re-run the session.

Adapting stimuli consisted of a target and flankers. The center-to-center distance between stimuli was 3°. Observers were initially adapted for 30 s, followed by the first 2-IFC

task trial. A 5 s top-up adaptation period was inserted between subsequent trials to maintain stable adaptation. A 500 ms gray blank was inserted before each trial began. During this blank period, a black line was displayed next to the fixation point to indicate the beginning of a trial. The target location was always marked during both the adaptation and task periods to remove effects of location uncertainty on the detection task (Petrov et al., 2006). To equate the attentional state across conditions observers performed a contrast decrement task on the fixation mark during adaptation periods (Bi et al., 2009). The contrast decrement (10%) was displayed for 150 ms and the onset of the contrast decrement was selected randomly from a uniform distribution of [1.5, 2] s. To quantify the amount of adaptation we defined the threshold ratio between detection threshold before and after adaptation ($threshold_{after}/threshold_{before}$).

We measured the amount of adaptation with three different target orientations (vertical, horizontal, and oblique) embedded in five pattern conditions: *single*, *same* (3 and 5-element), *orthogonal*, and *alternating*. For a vertical target example, these patterns were **V**, V**V**V, VV**V**VV, H**V**H, and VH**V**HV. The target and the flanker contrasts were set to 50% and 25%, respectively. Seven observers participated in each target orientation version. The first author participated in all target orientation versions. Observers participated in two sessions (left and right visual field). Each session consisted of five blocks (five pattern conditions). The order of blocks was randomized for each observer. A 3 minute break was given between blocks.

fMRI experiments

Six observers including the author participated. The fixation point was placed 3° below from the center of the display. The target stimuli were displayed in the upper quadrant of both visual fields 3° horizontally and 3° vertically from the fixation point. The center-to-center distance between the stimuli comprising a pattern was 3°.

One experimental session consisted of 2 localizer scans and 8 stimulus scans. A localizer scan (372 s) consisted of three conditions each presented in 12 s blocks: (1) fixation ("F"), (2) stimulus block with checkerboards in the location of the first flankers ("S1"), and (2) stimulus block with the checkerboards in the location of the second flankers ("S2"). The fixation condition was inserted between the two stimulus conditions (S1-F-S2-F, etc). The checkerboards were counter-phased flickered and windowed using the same Gaussian envelope used for Gabor stimuli in the main experiment.

A stimulus scan (372 s) consisted of four conditions each presented in 12 s blocks: (1) fixation, (2) *same* (VV**V**VV or HH**H**HH), (3) *orthogonal* (HH**V**HH and VV**H**VV), and (4) *alternating* (VH**V**HV and HV**H**VH). There were five repetitions per each condition in a stimulus scan. The fixation condition was inserted between the three stimulus conditions (*same*-F-*orthogonal*-F-*alternating*-F, etc). The center orientation alternated between vertical and horizontal in successive stimulus scans. During a stimulus block, flankers were counterphased flickered at 0.5 Hz to preclude fading effects and the target Gabors were flashed on (250 ms) and off (250 ms). The phase of target Gabor was always in-phase with the flanker Gabors. The same contrast decrement task used in the adaptation experiment above was used throughout each stimulus scan to equate attentional state across the conditions. We found no significant differences in

both the performance (87.5%, 88.4%, and 82.7% for same, orthogonal, and alternating condition, respectively; a repeated measures ANOVA, $P = 0.601$) and reaction times (406 ms, 419 ms, and 418 ms for same, orthogonal, and alternating condition, respectively; $P = 0.185$).

Functional MRI data were acquired using a Philips Achieva 3T scanner using an 8-channel head coil and an echo-planar imaging sequence (repetition time, 2 s; flip angle, 70°; 31 axial slices of 3.5 mm thickness (no gap) and 3.44 × 3.44 mm resolution, field of view, 220 mm). Each scanning session began with a T1-weighted structural scan 1 × 1 × 1 mm used for visualization of retinotopic visual areas. Visual cortical areas, V1, V2, and V3 were localized using standard retinotopic mapping and cortical-flattening techniques using BrainVoyager QX. Regions of interest (ROIs) within these visual areas were determined using the localizer scan described above. Those voxels with a significantly larger response in a comparison between the target and flanker-1 position and confined to the retinotopic locations of each visual area were included in further analyses.

Time courses for each of the stimulus scans were extracted and averaged across voxels within an ROI. For each scan, the signal intensity in each condition at each of 8 time points time-locked to the onset of the stimuli was averaged, including 4 s before the onset to 10 s after the onset. The two timepoints (4 s) before trial onset served as baseline. The averaged time courses of signal intensity were converted to percent signal change by subtracting the corresponding baseline measurement and then dividing by that value. The resulting time course for each condition was then averaged across scans. The signal intensity averaged from 4s to 10s post-stimulus presentation was used as the measured response for each condition. This single

number was then averaged across observers and a repeated-measures ANOVA and planned-comparison paired t-tests were used for statistical evaluation.

ERP experiment

Seventeen observers including the author participated. Except for the author, all were naïve observers. They were paid $20 per hour for participation. The stimuli were generated and controlled by Presentation (Neurobehavioral Systems, Inc.), and they were displayed on a linearized 21-in. Samsung SyncMaster CRT monitor (85 Hz refresh rate). The viewing distance was approximately 70 cm. The spatial stimulus arrangement was identical to the fMRI experiment. On a given trial, flankers appeared before the target onset and remained throughout the trial. After a random duration chosen from a uniform distribution [1, 2] s, targets were briefly flashed for 100 ms. After the target offset, the flankers remained in the display for 500 ms. The inter-trial interval was 2 s. Observers were asked to maintain fixation and to limit eye blinks to the inter-trial interval.

One experimental block consisted of 12 trials [2 orientation (vertical and horizontal) x 3 stimulus conditions (*same*, *orthogonal*, and *alternating*) x 2 repetitions]. They were randomized within a block. Observers finished 28-50 blocks (336-600 trials). Observers initiated each block after a 5 s break by pressing a designated key in the button box. The first block served as practice.

EEG waveforms were recorded using BioSemi active Ag-AgCl electrodes from 64 sites. The signals were referenced to the right mastoid during online acquisition and re-referenced to the average of right and left mastoids offline. Vertical EOG was measured using an electrode

placed below the left eye and horizontal EOG was measured using an electrode placed at the outer canthus of the right eye. The signals were digitized at a sampling rate of 256 Hz.

EEG epochs started 100 ms before the target onset and lasted 500 ms after the target onset. Each waveform was baseline corrected to the average voltage of the interval -100 ms to 0 ms before the target onset and low-pass filtered at 30 Hz to remove high-frequency noise. Trials with waveforms that had a larger than 50 μV peak-to-peak amplitude were excluded as these were trials deemed to be contaminated with eye blinks or other sources of noise. On average 18.3% of trials were removed from each observer (minimum = 1.8% of trials, maximum = 47.2% of trials). Data from three observers were discarded due to excessive artifact rejection (> 40%). The resulting waveforms were averaged across conditions individually for statistical analyses and then averaged across observers for figures.

P1 amplitude on six electrodes (Oz, O1, O2, POz, PO3, and PO4) was measured by averaging the ERP amplitudes during the time window of 130 ms to 170 ms. These electrodes were centered over the maximum of the P1 component as determined through visual inspection of the scalp topography. These individual amplitudes were averaged across six electrodes to represent P1 amplitude.

Noise classification experiment

Twelve observers participated. The stimulus parameters were identical to those of the experiments above except as specified below. Noise stimuli were constructed by applying a log Gaussian filter to white noise in the spatial frequency domain producing stimuli that were isotropic for orientation, with a peak spatial frequency at 3 cpd, and a bandwidth of 1 octave.

The luminance contrast of the noise stimuli was 30%. Gabor stimuli had spatial frequency of 3 cpd and a luminance contrast of either 8% or 24%. Target stimuli were constructed by the adding the noise and Gabor stimuli. There were three target conditions: noise-only, noise + 8% Gabor, and noise + 24% Gabor. A target could be displayed within five flanker conditions: *single* (no flankers), *vertical* (VVNVV), *horizontal* (HHNHH), and *alternating* (VHNHV and HVNVH) where N represents the target. The experimental session (450 trials) consisted of 30 blocks of 15 trials (3 target types x 5 flanker conditions). The trial order was counterbalanced such that all 15 trials within each block were randomized and presented before the next block started. Thirty noise stimuli were generated and each noise stimulus was used for each block. So, if there were any response biases due to the noise image *per se*, it would have affected the response to all flanker conditions within the block in the same way. The target orientation was randomly chosen for each trial. Observers initiated each trial by pressing the space key on the computer keyboard. The flankers appeared first in both visual fields. A random inter-stimulus interval (ISI) selected from a random distribution of [1, 2] s was inserted before the target was displayed. The target was displayed for 150 ms in either left or right visual field and the entire display disappeared together. Observers reported the target orientation. Fifteen trials were given as practice trials prior to the main experiment.

RESULTS

Contrast adaptation: deviations from the context result in more adaptation

Our first experiment used psychophysical contrast adaptation to infer the magnitude of the neural response in early visual cortex to the target stimulus (Blakemore & Campbell, 1969; Bradley, Switkes, & De Valois, 1988; Movshon & Lennie, 1979). To quantify adaptation strength, we calculated the ratio of each observer's contrast detection threshold for a target before and after adaptation. Our assumption is that more adaptation—as indexed by an increase in post-adaptation detection thresholds—reflects stronger neural activity in response to the adapting stimulus (Blake et al., 2006; Blakemore & Campbell, 1969; Carandini et al., 1998; Dragoi et al., 2000; Engel, 2005; Fang, Murray, Kersten, & He, 2005; Kohn & Movshon, 2003; Larsson et al., 2006; Priebe et al., 2002; also see Figure 5A). Observers were initially adapted to a stimulus for 30 s before performing a two-interval forced choice (2IFC) detection task. A 5 s top-up adaptation period was inserted between trials to maintain stable adaptation (Figure 5B). The target location was always marked during both the adaptation and task periods to remove any effects of location uncertainty on the detection task (Petrov et al., 2006). To equate the attentional state across conditions observers performed a contrast decrement detection task on the fixation mark during adaptation periods (Bi et al., 2009).

Using a vertically oriented target (**V**), we found more adaptation with orthogonal, horizontally oriented local flankers (H**V**H, *orthogonal* condition) compared to vertically oriented local flankers (V**V**V, 3-element *same* condition), [paired two-tailed t test, $t(6) = 6.15$, $p < 0.001$; Figure 6A]. In an additional experiment we found that the horizontal and vertical flankers alone (no stimulus in the target position) did not produce any adaptation, indicating that differences in adaptation observed between the same and orthogonal condition were due to differences in

the orientation relationship between the target and flankers and not due to differences in the flankers themselves (Figure 5A). Larger adaptation to a target with orthogonally oriented flankers is consistent with electrophysiology studies showing larger neural responses in V1 to a stimulus when it is surrounded by orthogonal than iso-oriented stimuli (Cavanaugh et al., 2002b; Knierim & van Essen, 1992; Sillito et al., 1995). One possible interpretation of this result is that neurons respond more strongly to stimuli that deviate from the context. However, because this condition also manipulates the local orientation arrangement of the target and flankers, this result in isolation can also be explained through local orientation-selective spatial summation or normalization (Cavanaugh et al., 2002a; Cavanaugh et al., 2002b; Shapley, 2004).

We then added distant flankers to the original 3-element stimuli to create 5-element stimuli. This allowed us to manipulate whether the target deviated from the context without changing the *local* (the target and its immediate neighbors) feature arrangement. Adding distant vertical flankers to the 3-element *same* condition, generating a 5-element *same* condition (VV**V**VV) did not change the amount of adaptation [$t(6) = 1.23$, $p = 0.26$], demonstrating that additional vertical flankers do not necessarily alter responses to the target. Importantly, we observed a significant difference in our critical condition: adaptation to the 3-element orthogonal condition (H**V**H) is significantly reduced when distant vertical flankers are added, producing a 5-element *alternating* condition (VH**V**HV) [$t(6) = 4.25$, $p = 0.005$], presumably because the target does not deviate from—and is grouped into—the alternating pattern induced by the flankers. As a further test, we directly compared the amount of adaptation between the *alternating* VHVHV condition and the *orthogonal* HHVHH condition.

Consistent with our hypothesis, the amount of adaptation was significantly less in the alternating condition (see Figure 5C).

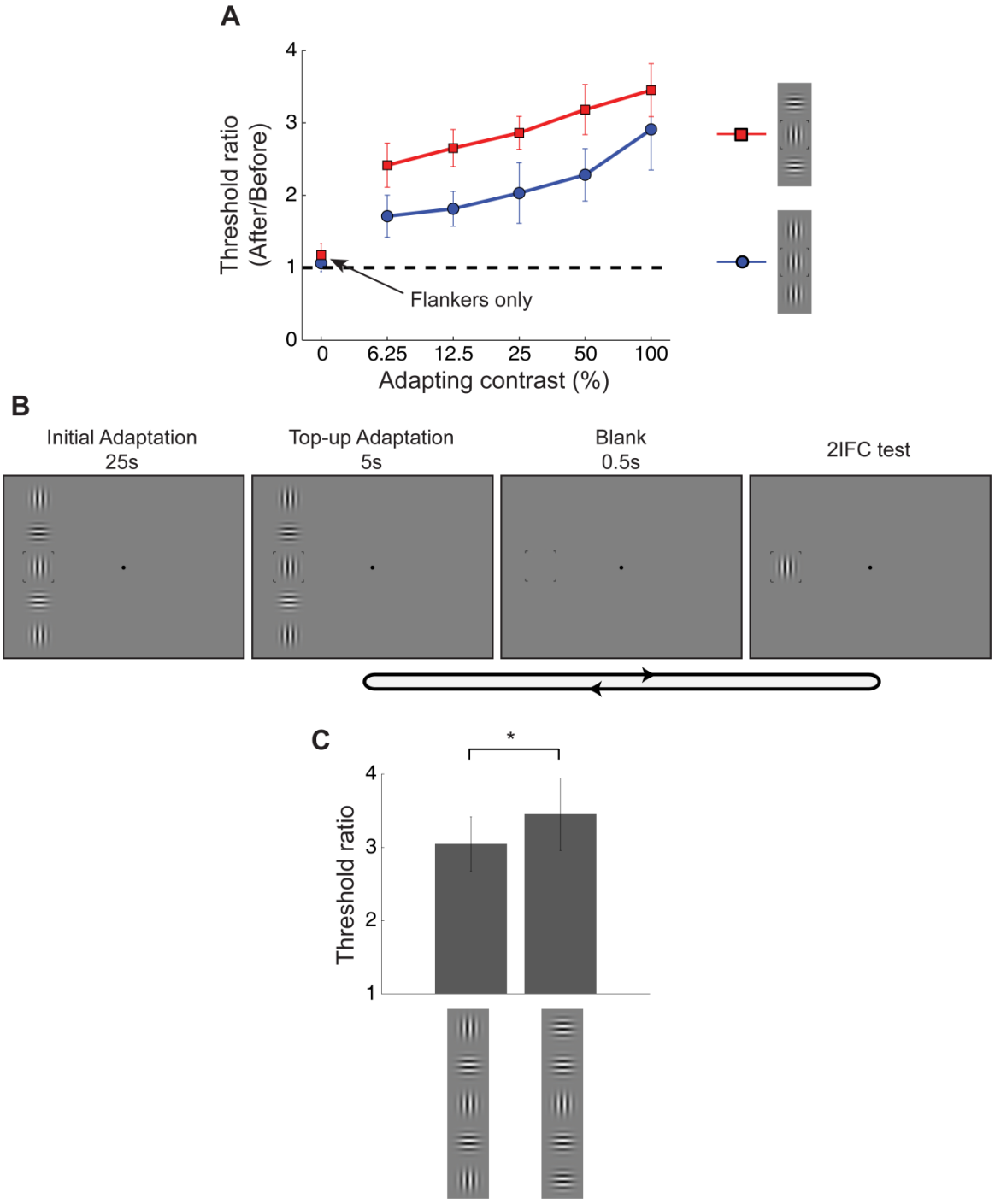

Figure 5. Contrast adaptation control experiment. (A) We measured the threshold ratio on a vertical target as a function of the adapting contrast (0%, 6.25%, 12.5%, 25%, 50%, and 100%) in two pattern conditions: same (VVV) and orthogonal (HVH). The contrast of the flankers was fixed at 25%. For both the same and orthogonal conditions, as the adapting contrast of the target increased, the threshold ratio also increased presumably due to higher neural activity for stimuli with higher contrast (a repeated measures ANOVA, $p = 0.005$). Importantly, the target in the orthogonal condition produced more adaptation than in the same condition across all the adapting contrasts, implying that the neural response to the target was higher in the orthogonal condition ($p = 0.002$; no interaction between adapting contrast and pattern condition, $p = 0.634$). Note that the flankers alone – with no adapting stimulus presented at the target location – did not affect the amount of adaptation (paired two-tailed t tests, $p = 0.617$ and $p = 0.323$ in the same and the orthogonal condition respectively, $n = 5$; isolated data points at 0% contrast). Dashed line indicates no changes in detection threshold after adaptation. Error bars are the SEM across observers. (B) Contrast adaptation experimental procedure. Observers initially adapted for 30 s. A gray blank screen (500 ms) was inserted before the two-interval forced choice (2IFC) detection task. Each interval (200 ms) was indicated by short high-pitched tones. A blank screen (300 ms) was inserted between the two intervals. Observers reported which interval contained the target. Top-up adaptation (5 s) was inserted between 2IFC detection tasks to maintain stable adaptation. The target location was always marked by thin lines throughout the procedure. (C) Additional adaptation control experiment. Five observers participated in an additional comparison using identical stimulus procedures as the main adaptation experiment. Detection thresholds following adaptation to the alternating condition were significantly lower than following adaptation to the orthogonal condition. * $p < 0.05$. Error bars are the SEM.

We observed the same results for horizontally (Figure 6B) and obliquely (Figure 6C) oriented targets. In all cases, the 3-element *orthogonal* condition resulted in more adaptation than the 3-element *same* condition [$t(6) = 2.80$, $p = 0.03$ for horizontal target; $t(6) = 3.95$, $p < 0.01$ for oblique target]. Also in all cases, the 3-element *same* condition did not differ from the 5-element *same* condition [$t(6) = 0.70$, $p = 0.51$ for horizontal target; $t(6) = 0.46$, $p = 0.66$ for oblique target]. Critically, in all cases, adding distant flankers to the 3-element *orthogonal* condition to create a 5-element *alternating* sequence significantly reduced the amount of adaptation [$t(6) = 4.09$, $p < 0.01$ for horizontal target; $t(6) = 3.76$, $p < 0.01$ for oblique target]. Overall, these results indicate that distant flankers that make the target part of an alternating sequence reduce neural responses in early human visual cortex.

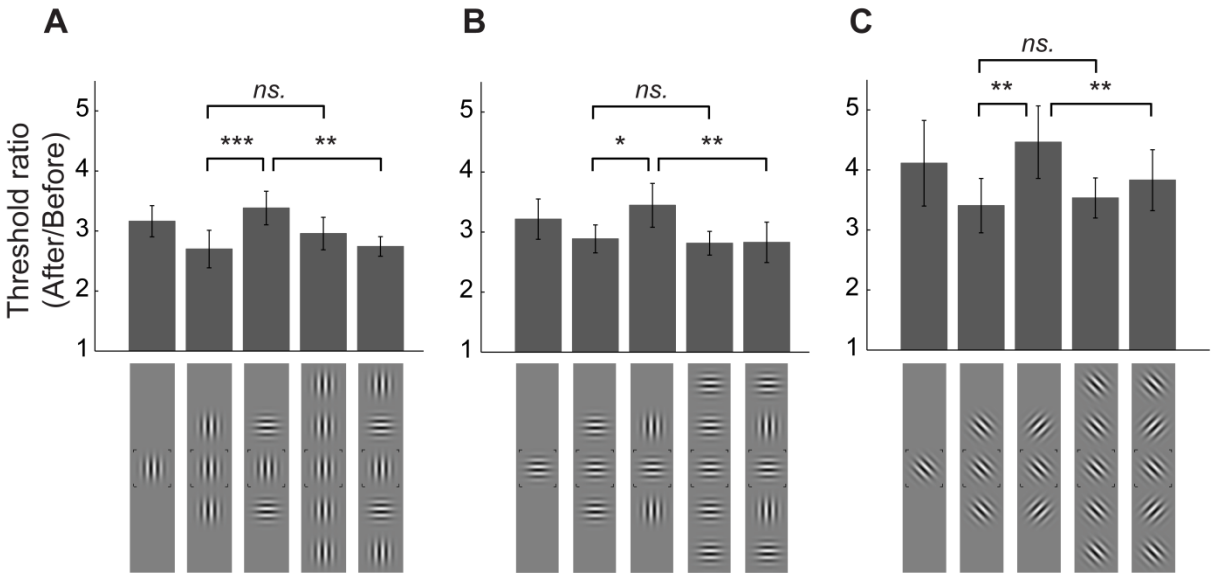

Figure 6. Adaptation experiment. (A-C) Threshold ratio – calculated as the contrast threshold after adaptation divided by the contrast threshold before adaptation – of each target orientation in five conditions. The threshold ratio of the vertical target, the horizontal target, and the oblique target is shown in A, B, and C, respectively. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; *ns*. not significant. Error bars are the SEM across observers.

fMRI: deviations from the context result in larger responses

Given these initial behavioral results using adaptation, our expectation was that the fMRI response to the target within early visual cortex would show a similar dependence on whether the target was grouped into or deviated from the pattern of orientations induced by the contextual elements. We measured the fMRI signal in early visual cortex (areas V1, V2, and V3) in response to the target under similar orientation configurations. To maximize data collection five-element patterns were simultaneously presented in both the left and right visual field. Stimuli were positioned so that the target was centered in the upper visual quadrants so as to optimize the localization of their cortical representation.

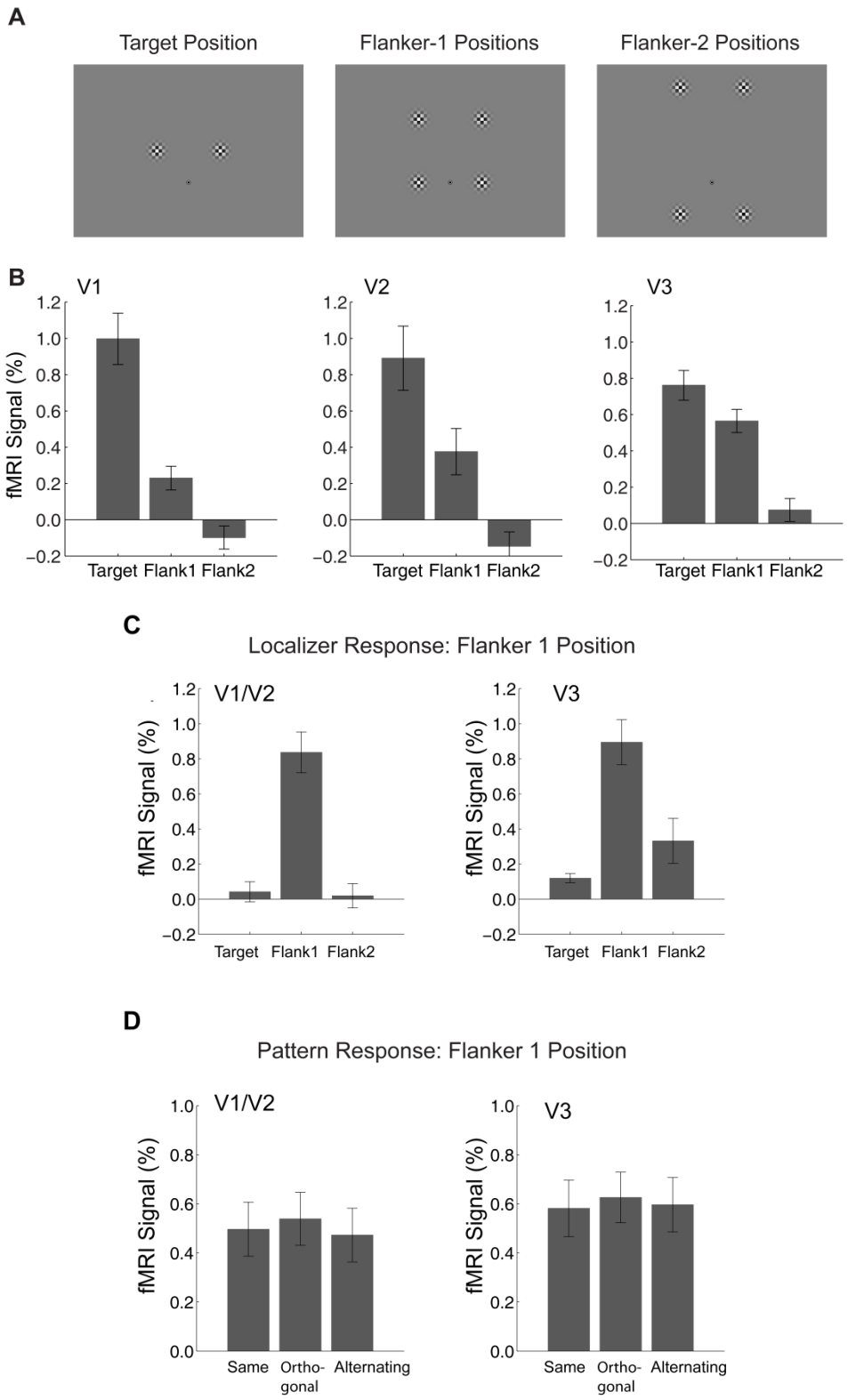

Figure 7. fMRI localizer scan and control analysis. (A) Localizer scans that presented flickering checkerboards in the target position, in the positions of local flankers (flanker-1), and in the positions of

the distant flankers (flanker-2). (B) Percent signal change from blank fixation in the target ROI for each of the three checkerboard locations measured in V1, V2, V3. We assessed the spatial specificity of the target-ROI by calculating the percent signal change (from blank fixation) in response to the checkerboards in the two flanker locations. The response of the target-ROI to the checkerboards in the flanker-1 position was significantly greater than zero in V1 ($p = 0.02$), V2 ($p = 0.03$), and V3 ($p < 0.0001$) indicating that the target-ROI response is at least partially affected by stimuli in the first flanker location. However there was no significant response in the target ROI to stimuli in the flanker-2 position in V1, V2, or V3. Error bars are the SEM. (C-D) Functional MRI experiment analyzing the response in the cortical representation of the Flanker 1 position. (C) Percent signal change from blank fixation in the Flanker 1 ROI for each of the three checkerboard locations measured in V1/V2 and V3. (D) Percent signal change from blank fixation in the Flanker 1 ROI for each of the three stimulus conditions in V1/V2 and V3. No significant differences or trends were observed. Error bars are the SEM.

Localizer scans of Gaussian windowed counterphase-flickering checkerboards were used to isolate the cortical region in each visual area that represents the target position. These localizer scans contained three conditions: target-position, flanker-1 positions, and flanker-2 positions (Figure 7A) intermixed with blank screen fixation in a block design. Target regions-of-interest (ROIs) were defined as those regions with a larger fMRI response to the target position than the flanker-1 position. We then assessed the spatial specificity of the target-ROI by calculating the percent signal change (from blank fixation) in response to the checkerboards in the two flanker locations. The response of the target-ROI to the checkerboards in the flanker-1 position was significantly greater than zero in V1 ($p = 0.02$), V2 ($p = 0.03$), and V3 ($p < 0.0001$) indicating that the target-ROI response is at least partially affected by stimuli in the first flanker location. However there was no significant response in the target ROI to stimuli in the flanker-2 position in V1, V2, or V3 (Figure 7B).

We then measured the fMRI response in the target ROI to three different 5-element stimulus conditions: target matching all flankers [*same* (VVVVV or HHHHH)], target orthogonal to all flankers [*orthogonal* (VVHVV or HHVHH)], and the target part of an alternating sequence

[*alternating* (VHVHV or HVHVH)]. Importantly, the local configuration between the target and the flankers is the same in the *orthogonal* and *alternating* conditions (VHV or HVH) – only the orientation of the distant flankers varies. Conditions were presented in a block design that switched between the three stimulus conditions (*same*, *orthogonal*, and *alternating,* Figure 8A) intermixed with blank-screen fixation periods. Vertical and horizontal targets were presented in separate scans. The flankers were continuously present and flickered in square-wave counterphase at 0.5 Hz to remove unwanted perceptual fading effects. Targets were flashed on (250 ms) and off (250 ms) repeatedly during the block to drive neural responses. To maintain an equivalent attentional state across all conditions, participants performed a demanding central fixation task that required detecting brief luminance changes in the fixation mark (see Methods for performance data). Because the central target deviates from its context in the *orthogonal* compared to the *alternating* condition we expected a larger fMRI response in the *orthogonal* condition.

There was no main effect or interaction between horizontal and vertical target conditions so their responses are averaged. In all three visual areas, there was a significantly larger response for the *orthogonal* (HHVHH, or VVHVV) than the *same* condition [HHHHH or VVVVV, paired two-tailed t tests: V1, $t(5) = 2.75$, $p = 0.04$; V2, $t(5) = 3.41$, $p = 0.02$; V3, $t(5) = 5.37$, $p = 0.003$; Figure 8B]. In the critical comparison we saw, as expected, a greater response to the target in the *orthogonal* condition (VVHVV or HHVHH) than for the *alternating* condition (VHVHV or HVHVH) in both V2 [$t(5) = 2.87$, $p = 0.035$] and V3 [$t(5) = 3.80$, $p = 0.01$]. In V1, the reduction in response was similar but did not reach statistical significance [$t(5) = 2.29$, $p = 0.07$].

We also conducted a similar analysis in ROIs corresponding to the cortical representations of the immediately adjacent flankers ("Flanker 1" position). This analysis was done to ensure that the differences between conditions in the target-related ROI did not reflect differences encroaching from neighboring cortical regions. No significant differences were observed (see Figures 7C and 7D) suggesting that the effects are confined to the target-ROI. Overall, our results show that the fMRI response across early visual cortex is affected, not only by the specific local orientation relationship between the target and immediately adjacent flankers, but also by whether or not the target is part of an alternating sequence.

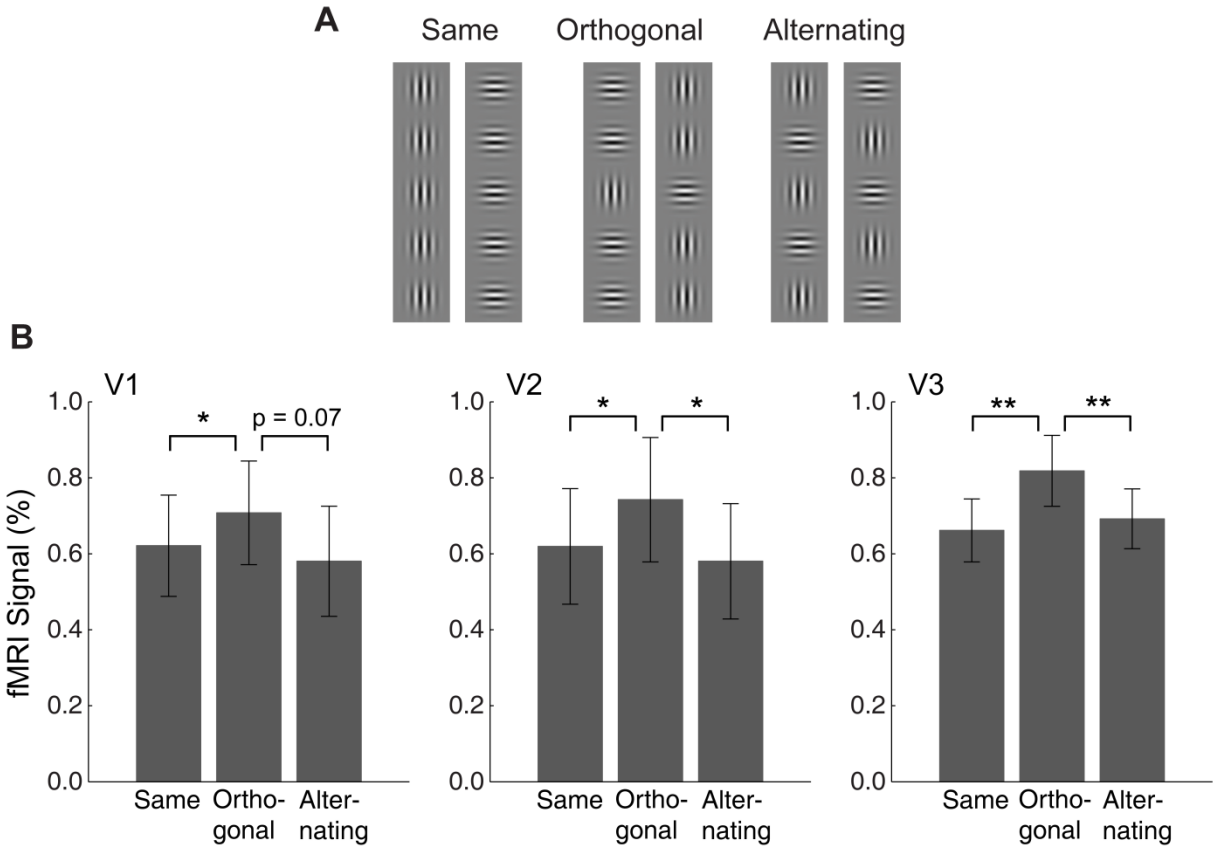

Figure 8. Functional MRI experiment. (A) The stimulus conditions used in the main fMRI experiment. (B) Percent signal change from blank fixation in the target ROI for each of the three stimulus conditions in V1, V2, and V3. * $p < 0.05$; ** $p < 0.01$. Error bars are the SEM across observers.

Event-related potentials (ERP): deviations from the context affect early components

Our assumption is that the pattern-related effects we have observed are the result of a dynamic feedback process between early visual areas with small receptive fields representing the individual components of the stimuli and higher visual areas with large receptive fields encompassing the entire pattern. If our pattern-related effects are indeed due to such a feedback process then we can make predictions about their expected time course and, specifically, how these effects will depend upon the relative timing of flanker and target onset. In particular, our prediction is that if the flankers precede the onset of the target by a sufficient length of time the putative feedback process is likely to be in place and stabilized *before* the onset of the target. Consequently, the differential response between *alternating* and *same* conditions as compared to the *orthogonal* conditions should appear early in the neural response to the target.

To test this prediction we performed an ERP experiment using similar stimulus configurations as our previous experiments. Two arrays of Gabors were displayed in the left and right visual fields. The target Gabors were displaced 3° horizontally and vertically upward from a central fixation cross. This closely matches their position in our previous experiments and also minimizes the cancelling of the ERP signal from V1 as stimuli presented along the horizontal meridian stimulate both sides of calcarine sulcus and may generate dipoles of opposite orientation on each side (Clark et al., 1994).

On a given trial, the flankers appeared first and after a duration randomly chosen from a uniform distribution [1, 2] s, targets were presented for 100 ms. The flankers remained in the

display for 500ms after the target offset to ensure that the flanker offset did not contaminate the evoked potential to the target. Six conditions [*same* (VV**V**VV and HH**H**HH)], target orthogonal to all flankers [*orthogonal* (VV**H**VV and HH**V**HH)], and the target part of an alternating sequence [*alternating* (VH**V**HV and HV**H**VH)] were randomly intermixed within an experimental block.

The first prominent ERP component that was observed in response to the target was a positive deflection that peaked at approximately 150 ms (P1). Previous research has suggested that the P1 originates from early extrastriate visual areas (Clark et al., 1994; Di Russo et al., 2002). The C1 component – which is believed to originate in V1 (Clark et al., 1994; Di Russo et al., 2002) – was not observed in our data. The scalp distribution of the P1 was confined to the occipital pole (see Figure 3B). To characterize the magnitude of the P1, signals from the six electrodes around the occipital pole (Oz, O1, O2, POz, PO3, and PO4) were averaged. We calculated the magnitude of the P1 as the average amplitude between 130 ms and 170 ms. A repeated measures ANOVA showed no significant effect of orientation [$F(1, 13) = 1.23$, $p = 0.29$] or interaction [$F(2, 26) = 0.49$, $p = 0.62$], so the data from the two target orientations were collapsed for further analysis. Figure 9 shows the average waveforms and P1 amplitudes for each flanker conditions. There was a significant effect of flanker conditions [$F(2, 26) = 9.46$, $p < 0.001$] and planned contrasts showed that the P1 amplitude was greater for the *orthogonal* condition compared to *same* [$F(1,13) = 15.08$, $p = 0.002$] and *alternating* conditions [$F(1,13) = 9.11$, $p = 0.01$]. These results indicate that when the flankers precede the presentation of the

target, the initial feedforward processing of the target is affected by whether the target is grouped into or deviates from the contextual elements.

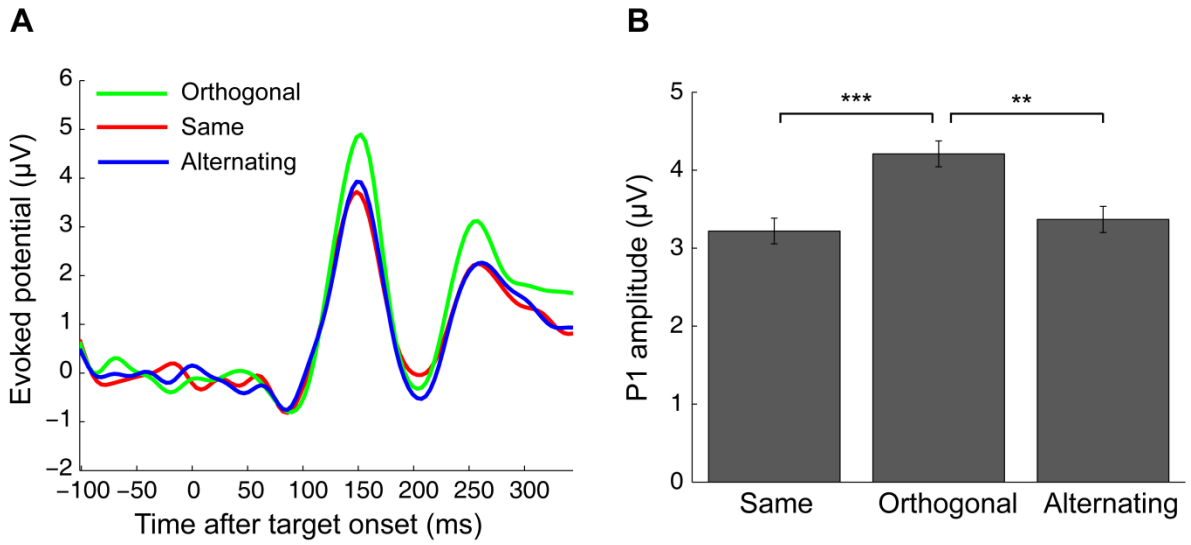

Figure 9. ERP Experiment. (A) The averaged ERP waveforms across observers for the three stimulus conditions: *same* (red), *orthogonal* (green), and *alternating* (blue). The ERPs were the averaged evoked potentials of the six electrodes (Oz, O1, O2, POz, PO3, and PO4). (B) The P1 amplitudes for three stimulus conditions. The P1 amplitudes of each electrode were measured by averaging the amplitude within the temporal window of 130-170 ms after the target onset. Then P1 amplitudes of each electrode were averaged and used for the statistical analysis. ** $p = 0.01$; *** $p < 0.01$. Error bars are the SEM across observers.

Classifying noise images: biased responses toward deviations from the context

Across 3 separate methodologies (psychophysics, fMRI, and ERP) we observed larger neural responses to the target when it deviated from the context. An important question is whether this neural modulation has direct perceptual consequences. We hypothesized that when asked to report whether a noise target containing no specific orientation information was horizontal or vertical, observers would be more likely to "see" the orientation that deviates

from the context. In other words, we expected that the influence of the flankers would result in orientation-specific modulation of the noise stimulus to bias perception *away* from orientations contained in the contextual elements. Such results would be consistent with a model that subtracts or inhibits context-matching orientations in the noise target.

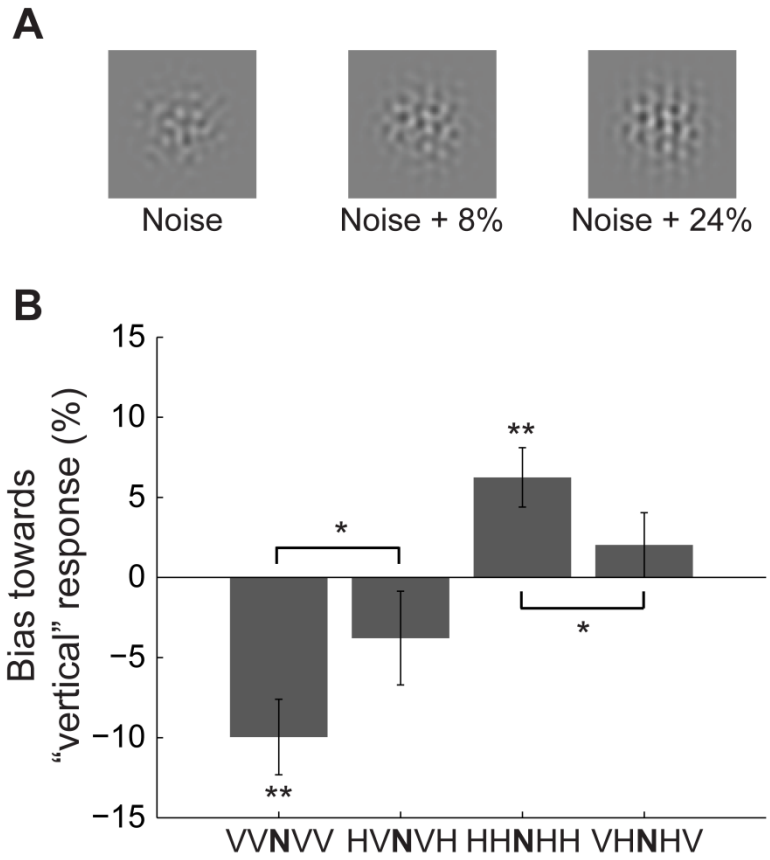

Figure 10. Noise classification experiment. (A) The first column shows an example of the noise-only image. The second and the third columns show examples of the combination of noise + orientation images (8% and 24%, respectively). Only examples of the addition of vertical orientation are shown. (B) The y-axis represents the response bias towards "vertical" orientation to noise-only images with different flanker configurations as compared to the *single* condition (noise-only without flankers). The x-axis represents context conditions, V: vertical orientation; H: horizontal orientation; N: the noise image. * $p < 0.05$; ** $p < 0.01$. Error bars are the SEM across observers.

Observers were asked to classify briefly presented targets as either horizontal or vertical. There were three target conditions: noise-only (30% luminance contrast), noise+8% contrast orientated Gabor (either horizontally or vertically oriented), and noise+24% contrast oriented Gabor (Figure 10A). Responses to the noise-only conditions were our primary interest – the noise+8% and noise+24% trials were used to maintain observer motivation and to assess individual performance in the task (see Figure 11 for performance data). A noise target (N) could be displayed within one of five flanker conditions: *single* (no flankers)*, vertical (*VVNVV), *horizontal* (HHNHH), and *alternating* (VHNHV and HVNVH). The *single*, noise-only condition was used to measure any baseline bias toward vertical or horizontal responses. On trials containing flankers, the flankers appeared first, followed by a random inter-stimulus interval selected from a random distribution of [1, 2] s, and then the target was displayed for 150 ms along with the flankers. The flankers were presented first to ensure that the hypothesized feedback processes induced by the flankers could be in place at the time of target processing.

Our goal was to see whether orientation judgments in the noise-only condition were influenced by the pattern of flanking elements. We quantified each observer's response bias toward "vertical" responses by subtracting the proportion of "vertical" responses for the *single* condition from "vertical" responses in each flanker condition: a positive value indicates that the presence of flanking stimuli biased observers towards more "vertical" responses.

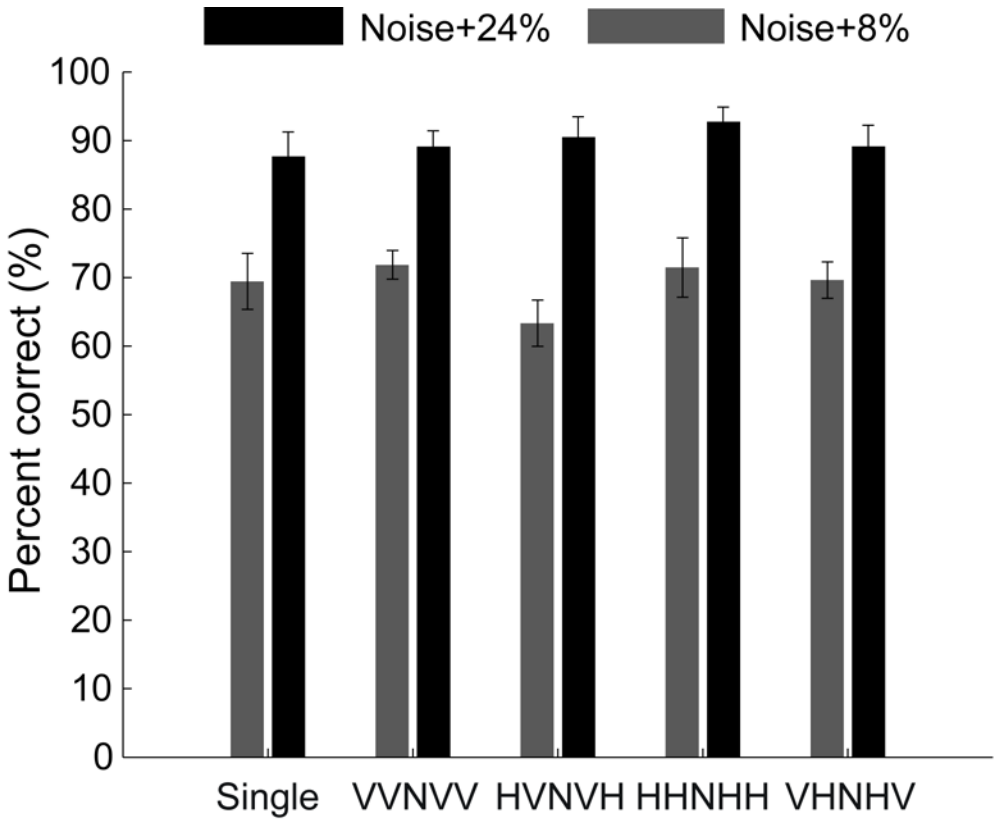

Figure 11. The performance data. The performance in Noise+8% (gray bars) and Noise+24% (black bars) target conditions. Single represents no flanker condition. V and H represent vertical and horizontal orientation, respectively. N indicates the target. As expected, performance was significantly higher in the noise+24% condition (M = 90%, SD = 8.73, n = 12) than in the noise+8% condition (M = 69%, SD = 8.71; a repeated measures of ANOVA, $p < 0.001$). There was no significant main effect of pattern condition ($p = 0.09$) or interaction ($p = 0.25$). Error bars are the SEM.

We found that the orientation of flanking stimuli significantly biased the perceived orientation of noise stimuli. We saw a response bias towards "horizontal" in the *vertical* (VV**N**VV) condition [one sample two-tailed t tests, mean = -9.96%, $t(11) = 4.24$, $p = 0.0014$; Figure 10B] and a bias towards "vertical" in the *horizontal* (HH**N**HH) condition [mean = 6.25%, $t(11) = 3.37$, $p = 0.0062$]. This indicates that a noise stimulus is 'seen' to be biased toward the orientation orthogonal to the flanking stimulus. In addition, even though the local

configurations were equivalent, changing the distant flankers eliminated the response bias [$t(11) = 1.29$, $p = 0.22$ in HVNVH condition and $t(11) = 0.10$, $p = 0.34$ in VHNHV condition]. Pairwise comparisons also confirmed the effect of distant flankers [$t(11) = 2.37$, $p = 0.04$, VVNVV vs. HVNVH; $t(11) = 2.50$, $p = 0.03$, HHNHH vs. VHNHV] on the perceptual decision of the central noise image within global patterns. Overall, our results show that perceptions of the target are biased towards orientations that deviate from the surrounding context.

DISCUSSION

The standard model of information processing in early visual cortex is that neurons behave like localized, linear, band-pass filters that are optimized to detect specific features in restricted regions of an image (Adelson & Bergen, 1985; Carandini et al., 2005; Jones & Palmer, 1987; Lennie & Movshon, 2005; Movshon et al., 1978; Ringach, 2002; Rust & Movshon, 2005). However, inconsistent with the standard model, it is well known that responses of V1 neurons to a stimulus inside the classical receptive field (CRF) can be modulated by stimuli presented outside the CRF (Albright & Stoner, 2002; Allman, Miezin, & McGuinness, 1985; Angelucci & Bressloff, 2006; Fitzpatrick, 2000). These findings have led to revisions of the standard model that include various forms of divisive normalization from neighboring neurons (Carandini, Heeger, & Movshon, 1997; Heeger, 1992; Schwartz & Simoncelli, 2001) and, to a first approximation, these revised models can account for a wide variety of contextual, surround effects (Cavanaugh et al., 2002a; Cavanaugh et al., 2002b; Shapley, 2004). Importantly,

normalization models still act like localized feature detectors but offer a mechanism for nonlinear gain control.

But the view that neurons in early visual cortex are localized feature detectors has been challenged in recent years (Lee & Mumford, 2003; Olshausen & Field, 2005) and recent studies have pointed to a potential role of early visual areas in processing high-level, "global" attributes of an image. For example, it has been suggested that early visual areas may be modulated by processes such as perceptual pop-out (Kastner et al., 1997; Knierim & van Essen, 1992; Nothdurft, Gallant, & Van Essen, 1999; Sillito et al., 1995), figure-ground segmentation (Lamme, 1995; Zipser et al., 1996) and contour integration Field, Hayes, & Hess, 1993; Kapadia et al., 1995; Kapadia et al., 2000; Polat et al., 1998). However, many of the apparently complex contextual effects seen in early visual areas can be accounted for through simple, local spatial summation and normalization processes (Cavanaugh et al., 2002a). In addition, many of the studies that have examined global, perceptually-based explanations for surround effects have confounded changes in local stimulation in V1 and perception. For example, studies that have examined perceptual pop-out (Kastner et al., 1997; Knierim & van Essen, 1992; Nothdurft et al., 1999; Sillito et al., 1995) have changed the orientation relationship between a target and immediately adjacent flankers. Similarly, contour integration has been investigated by varying the distance between a target and collinear flankers (Kapadia et al., 1995; Kapadia et al., 2000). In both instances, it is not possible to determine whether modulations of neural activity are due to changes in the stimulus configuration affecting low-level summation and normalization

processes (Cavanaugh et al., 2002a; Cavanaugh et al., 2002b) or are due to changes in global attributes of the image, or both.

Our stimuli eliminated the confound between the local stimulus arrangement and global attributes of an image. The addition of the second flankers allowed us to maintain the same local stimulus arrangement (target plus immediately adjacent flankers) while changing whether the target was grouped with or deviated from the orientation pattern of the flankers. Our results from different methodologies—psychophysics, fMRI, and ERP—clearly show that the neural response to the target is affected by global attributes of the image: when the target could be grouped with the flankers neural responses were smaller compared to when it deviated from the flankers. We consider this to be a global process because to recognize whether the central target belongs to or deviates from the flankers necessarily means that its orientation relationship to all other features in the image is analyzed. This is a novel finding that appears to contradict standard models of visual processing that emphasize the role of early visual areas in localized feature detection (Adelson & Bergen, 1985; Carandini et al., 2005; Jones & Palmer, 1987; Lennie & Movshon, 2005; Movshon et al., 1978; Ringach, 2002; Rust & Movshon, 2005).

Revisions of the standard model designed to account for nonlinear neural activity in V1 can explain some types of contextual effects such as surround suppression. While many of these models are not orientation-specific, there do exist versions of the standard model that have been proposed to account for orientation-specific surround effects between a target and immediately adjacent flankers (Cavanaugh et al., 2002a; Cavanaugh et al., 2002b). Our results

are inconsistent with these models (and any natural variant of them). Suppose a model is constructed – based purely on local mechanisms between the receptive field and immediately adjacent surround – that predicts larger responses when flankers are orthogonal (i.e., larger response to the target in a V**H**V configuration than a H**H**H configuration). This model would necessarily predict a further *enhancement* of the central target when distant flankers are added that are orthogonal to the first flankers (HV**H**VH) because the second distant flankers would enhance the response to the first flankers (since they are orthogonal) which would in turn further enhance the response to the target. Instead we show that the response to the target is *reduced* in a HV**H**VH configuration. Our results show that, in addition to the known local, normalization processes in early visual cortex, there is also an additional global, pattern-based process that is sensitive to orientation patterns across a large spatial scale.

Chapter III.

Surface structure modulates contextual effects in human visual cortex

It is well established that there can be neural suppression in early visual cortex measured at the individual neuron level (Allman et al., 1985; Blakemore & Tobin, 1972; Cavanaugh et al., 2002a; Maffei & Fiorentini, 1976; Sillito et al., 1995), by human fMRI (Zenger-Landolt & Heeger, 2003), and by human ERP (Haynes et al., 2003; Joo, Boynton, & Murray, 2012) when surrounding stimuli ("flankers") match the center ("target") orientation. Despite the importance of surround suppression in potentially mediating a number of important perceptual effects including saliency (Knierim & van Essen, 1992; Zipser et al., 1996), contour integration (Dobbins, Zucker, & Cynader, 1987; Kapadia et al., 1995), and orientation discrimination (Mareschal, Sceniak, & Shapley, 2001), the underlying neural mechanisms remain unknown. For example, surround suppression appears as early as the retina (Solomon, Lee, & Sun, 2006) and thalamus (Alitto & Usrey, 2008) and recent evidence has shown an important role of local cortical circuits in V1 (Adesnik, Bruns, Taniguchi, Huang, & Scanziani, 2012). However, a number of recent psychophysical findings have shown that high-level perceptual grouping can influence basic visual detection and discrimination performance in stimulus configurations that resemble those used in surround suppression experiments (Joo et al., 2012; Manassi, Sayim, & Herzog, 2012; Mareschal et al., 2001; Sayim, Westheimer, & Herzog, 2008). Thus, we hypothesized that perceptual grouping would influence surround suppression such that it would occur only when

the target and flankers were grouped into a single array. For example, we predicted that stimulus manipulations that isolated the target from the flankers—such as increasing spacing or making them appear to be on different surfaces—would eliminate surround suppression. On the other hand, manipulations that grouped the target and flankers into a single array—such as decreasing spacing or making them appear to be on the same surface—would promote surround suppression.

METHODS

Observers

All observers had normal or corrected-to-normal vision and all gave informed written consent in accordance with the University of Washington Institutional Review Board. Eighteen observers including the author participated in Experiment 1. Sixteen observers participated in Experiments 2-4. Except for the author, all were naïve observers who volunteered for either course credit or monetary compensation ($20 per hour).

Stimuli and procedure

The Gabor patches had a standard deviation of 0.72°, a spatial frequency of 2 cycles per degree (cpd), and were 50% contrast. The surface patches were generated using the "drop shadow" function in Adobe Illustrator CS3 (Adobe Systems, Inc.). In Experiment 4, we adjusted the shadow position and size to prevent overlap between the shadow edge and the target. The fixation point was placed 3° below from the center of the display. The target stimuli were

displayed in the upper quadrant of both visual fields 3° horizontally and 3° vertically from the fixation point. The center-to-center distance between the stimuli comprising a pattern was 3° in Experiment 1 and Experiment 4 and 6° in Experiment 2 and Experiment 3. The stimuli were generated and controlled by Presentation (Neurobehavioral Systems, Inc.), and they were displayed on a 21-in. CRT monitor (60 Hz refresh rate). The viewing distance was approximately 70 cm. On a given trial, flankers (or flankers and drop-shadow in Experiments 3 & 4) appeared before the onset of the target. After a random duration chosen from a uniform distribution between 1 and 2 s, targets were briefly flashed for 100 ms. After the target offset, the flankers remained in the display for 500 ms. The inter-trial interval was 3 s. Observers were asked to maintain fixation and to limit eye blinks to the inter-trial interval.

Target orientation could be either vertical or horizontal orientation. There were three flanker conditions (*single*, *same*, and *orthogonal*). The flanker orientation varied according to the target orientation. The flanker orientation matched the target orientation in the *same* condition but was orthogonal to the target orientation in the *orthogonal* condition. A thin circle (0.2°) that matched the size and contrast of Gabor stimuli was displayed in the flanker positions in the single condition to equate the stimulus timing.

One experimental block consisted of 12 trials [2 target orientation conditions x 3 flanker conditions x 2 repetitions]. They were randomized within a block. Observers finished 22-51 blocks (264-612 trials). Observers initiated each block after a 5 s break by pressing a designated key in the button box. The first block served as practice.

EEG recording and data analysis

EEG waveforms were recorded using BioSemi active Ag-AgCl electrodes from 64 sites. The signals were referenced to the left mastoid during online acquisition and re-referenced to the average of right and left mastoids offline. Vertical EOG was measured using an electrode placed below the left eye and horizontal EOG was measured using an electrode placed at the outer canthus of the right eye. The signals were digitized at a sampling rate of 256 Hz.

EEG epochs started 100 ms before the target onset and lasted 400 ms after the target onset. Each waveform was baseline corrected to the average voltage of the interval -100 ms to 0 ms before the target onset and low-pass filtered at 40 Hz to remove high-frequency noise. Trials with waveforms that had a larger than 50 μV peak-to-peak vertical and that exceeded ±50 μV on other electrodes were excluded as these were trials deemed to be contaminated with eye blinks or other sources of noise. Data from 5 observers (Experiment 1) and 1 observer (Experiment 2, 3, and 4) were discarded due to excessive artifact rejection (> 50%). The resulting waveforms were averaged across conditions individually for statistical analyses and then averaged across observers for figures.

P1 amplitude on six electrodes (Oz, O1, O2, POz, PO3, and PO4) was measured by averaging the ERP amplitudes during the time window of 130 ms to 170 ms. These electrodes were centered over the maximum of the P1 component (150 ms after target onset) as determined through visual inspection of the scalp topography (see Figure 3B). These individual amplitudes were averaged across six electrodes to represent P1 amplitude. We conducted a repeated measures ANOVA for the statistical analysis. The data from the *single* condition was not included in the analysis because the response to the *single* condition (interaction between

an oriented target and circles) was categorically different from the other conditions (interaction between an oriented target and oriented flankers). The single condition was only used to assess any difference between vertical and horizontal targets in the absence of oriented flankers (e.g., see Figure 13).

RESULTS

In Experiment 1 we established the basic surround suppression effect using stimulus configurations where the target was surrounded by "nearby" flankers (Figure 12A, target-flanker distance = 3°). We used the amplitude of the P1 component (150 ms after target onset) to index neural activity in early visual cortex (Clark et al., 1994; Joo et al., 2012). The amplitude of the P1 was defined by averaging ERP amplitudes during the time window between 130 ms and 170 ms after target onset on six occipital electrodes (Oz, O1, O2, POz, PO3, and PO4). These electrodes were centered over the maximum of the ERP amplitudes at 150 ms after target onset as determined through visual inspection of the scalp topography (see Figure 3B).

A repeated measures ANOVA revealed that there was no significant effect of target orientation ($F_{1,12} = 0.016$, $P = 0.901$) or interaction between target orientation and flanker condition ($F_{1,12} = 0.002$, $P = 0.964$). However, P1 amplitude was suppressed in the *same* condition compared to the *orthogonal* condition ($F_{1,12} = 16.787$, $P = 0.001$), consistent with orientation-specific surround suppression. Separate analyses for each target orientation

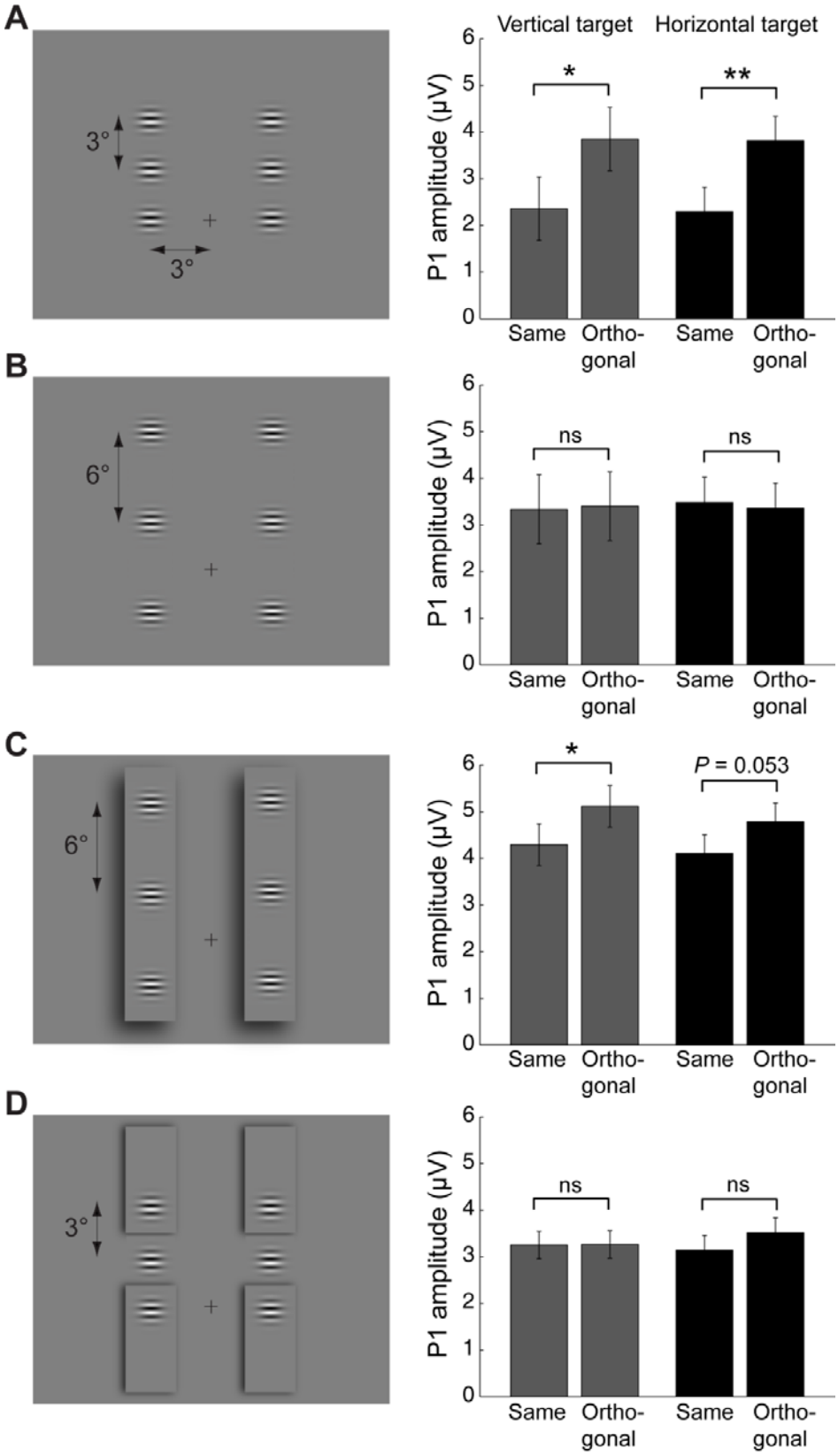
A
3°
3°
Vertical target
Horizontal target
P1 amplitude (μV)
*
**
Same
Ortho-gonal
Same
Ortho-gonal
B
6°
ns
ns
C
6°
*
P = 0.053
D
3°
ns
ns

Figure 12. The stimuli and results of Experiments 1-4. The left panel shows the stimulus configurations in each experiment and the right panel shows P1 amplitude measured in each condition for each target orientation. Grey and black bars represent P1 amplitudes for the vertical target and horizontal target, respectively. (A) Experiment 1: near flankers. The center-to-center distance between the target and flankers was 3°. (B) Experiment 2: far flankers. The center-to-center distance between the target and flankers was 6°. (C) Experiment 3: far flankers on the same surface as the target. The center-to-center distance between the target and flankers was 6°. (D) Experiment 4: near flankers on a different surface from the target. The center-to-center distance between the target and flankers was 3°. * $p < 0.05$; ** $p < 0.01$; ns, not significant. Error bars represent within-subject 95% confidence interval (Loftus & Masson, 1994).

confirmed that the trend was similar for both vertical ($F_{1,12} = 7.654$, $P = 0.017$) and horizontal ($F_{1,12} = 13.455$, $P = 0.003$) targets (Figure 12A).

After establishing orientation-specific surround suppression in the P1 amplitude of our ERP data, we tested our grouping hypothesis using the same procedures. First in Experiment 2, we simply increased the distance between the target and flankers by doubling the center-to-center distance (6°, Figure 12B) used in Experiment 1. Perceptually, with the increased distance, the stimulus now appeared to be 3 isolated Gabors rather than a single array of 3 Gabors. Thus, because the target and flankers were no longer grouped into a single array, we predicted that surround suppression would be eliminated. As predicted, we found no difference in P1 amplitude between the *same* and *orthogonal* conditions for both target orientations (Figure 12B; vertical, $F_{1,14} = 0.013$, $P = 0.911$; horizontal, $F_{1,14} = 0.081$, $P = 0.781$). However, this finding is consistent with a number of potential explanations—ranging from local normalization models (Cavanaugh et al., 2002a; Shapley, 2004) to our high-level, grouping hypothesis.

To distinguish between these alternatives, in Experiment 3 we used the distant-flanker configuration of Experiment 2 but made the target and flankers appear to be grouped on a common surface that was distinct from the background. This was done by adding a small “drop

shadow" in the region around the stimuli to create a surface that appeared to be at a closer depth-plane than the background (Figure 12C). Both the flankers and the drop-shadow surface appeared before the onset of the target—using the same timing structure as Experiments 1 and 2. Any model of surround suppression that emphasized local orientation interactions between the target and flankers would again predict no surround suppression—as in Experiment 2. However, based on our grouping hypothesis, we expected to observe surround suppression because the target and flankers were now grouped on a common surface separate from the background. Consistent with the grouping hypothesis, P1 amplitude was suppressed in the *same* condition compared to the *orthogonal* condition (Figure 12C; vertical target, $F_{1,14} = 5.223$, $P = 0.038$; horizontal target, $F_{1,14} = 4.461$, $P = 0.053$).

If surface representations are indeed important for determining when surround suppression effects occur, we predicted that moving the target and flankers to different surfaces would eliminate surround suppression—even with spatial parameters that would otherwise result in strong surround suppression. In Experiment 4, we used the same spatial parameters of Experiment 1—where we observed strong surround suppression—but moved the flankers to different surfaces than the target (Figure 12D). Although the flankers were displayed in the near proximity of the target, we found no evidence of orientation-specific surround suppression in the P1 amplitude (Figure 12D; vertical target, $F_{1,14} = 0.002$, $P = 0.965$; horizontal target, $F_{1,14} = 2.143$, $P = 0.165$).

DISCUSSION

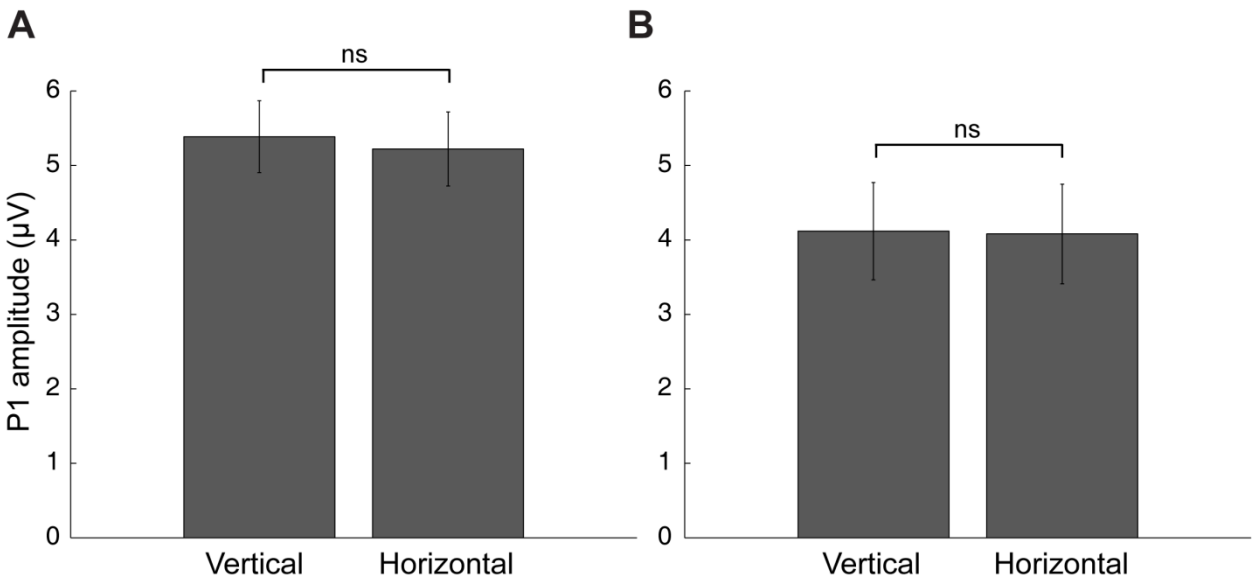

Figure 13. P1 amplitue in single condition. To assess whether there was any significant low-level contribution of the cast-shadow surfaces to the response of the target, we examined the P1 amplitude measured in the *single* condition (no oriented flankers) in Experiments 3 and 4. If there was a low-level contribution we would expect to observe a difference based on target orientation. (A) Experiment 3. There was no significant difference in P1 amplitude between target orientations (t_{14} = 0.592, p = 0.563). (B) Experiment 4. There was no significant difference in P1 amplitude between target orientations (t_{14} = 0.116, p = 0.909). ns, not significant.

Our results demonstrate that grouping—specifically mediated by the surface placement of the target and flankers—modulates surround suppression. We assume that the surface structure of the images in our experiments is represented in higher stages of the visual system that have neurons with sufficiently large receptive fields and complex tuning properties sensitive to relative depth. This may include regions such as the lateral occipital complex (Kourtzi & Kanwisher, 2001; Murray, Kersten, Olshausen, Schrater, & Woods, 2002; Murray, Olshausen, & Woods, 2003). We suggest that feedback from these regions modulates surround suppression in early visual areas (e.g., V1-V3). However, the signal that we measured in response to the target—the P1 component—is believed to represent early, feedforward neural

activity (Luck, Woodman, & Vogel, 2000). How do we reconcile the potential role of feedback with the modulation of an early feedforward neural signal? It is important to emphasize the relative timing of our stimulus presentation. The flankers and the drop-shadow surfaces (in Experiments 3 and 4) were presented first, 1-2 seconds in advance of the briefly presented target stimulus. Thus, there was sufficient time for the putative feedback process to be in place and stabilized *before* the onset of the target. How our results generalize to other timing configurations—such as the simultaneous presentation of the target and flankers—remains an open question.

The target stimuli were behaviorally irrelevant (i.e., under no specific task instruction), briefly flashed, of unpredictable orientation and peripherally located. Thus, it is unlikely that there are any simple confounds related to attention or motivation that could potentially explain our results. Further, the lack of any significant interaction between the horizontal and vertical target orientations makes it unlikely that, for example, the vertical orientation energy of the shadow trivially affected the results. If the shadow was somehow interacting with the target stimulus in a low-level manner, a signal difference between the vertical and horizontal targets would be expected. We also found no difference between vertical and horizontal target orientations when presented alone without oriented flankers but in the presence of the drop-shadow surfaces (see Supplementary Fig. 5). Finally, any potential low-level explanation of the shadow to our results would somehow have to account for surround suppression effects appearing in Experiment 3 but not in Experiment 4. Instead, the high-level surface structure of the image—specifically, whether the target and flankers shared a common surface and thus

were grouped into a single array—offers the most consistent explanation for our results. Indeed, our results are consistent with behavioral evidence that demonstrate a fundamental role of surface structure in perceptual grouping (Nakayama & Mackeben, 1989; Nakayama & Shimojo, 1992; Nakayama, Shimojo, & Silverman, 1989). Overall, the results of the present study, together with our recent findings (Joo et al., 2012), suggest a coding scheme in early visual cortex that is sensitive to high-level image structure.

BIBLIOGRAPHY

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A, 2*, 284-299.

Adesnik, H., Bruns, W., Taniguchi, H., Huang, Z. J., & Scanziani, M. (2012). A neural circuit for spatial summation in visual cortex. *Nature, 490*, 226-231.

Albright, T. D., & Stoner, G. R. (2002). Contextual influences on visual processing. *Annual Review of Neuroscience, 25*, 339-379.

Alitto, H. J., & Usrey, W. M. (2008). Origin and dynamics of extraclassical suppression in the lateral geniculate nucleus of the macaque monkey. *Neuron, 57*, 135-146.

Allman, J., Miezin, F., & McGuinness, E. (1985). Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local-global comparisons in visual neurons. *Annual Review of Neuroscience, 8*, 407-430.

Angelucci, A., & Bressloff, P. C. (2006). Contribution of feedforward, lateral and feedback connections to the classical receptive field center and extra-classical receptive field surround of primate V1 neurons. *Progress in Brain Research, 154*, 93-120.

Angelucci, A., Levitt, J. B., Walton, E. J., Hupe, J. M., Bullier, J., & Lund, J. S. (2002). Circuits for local and global signal integration in primary visual cortex. *Journal of Neuroscience, 22*, 8633-8646.

Bair, W., Cavanaugh, J. R., & Movshon, J. A. (2003). Time course and time-distance relationships for surround suppression in macaque V1 neurons. *Journal of Neuroscience, 23*, 7690-7701.

Barlow, H. B., Blakemore, C., & Pettigrew, J. D. (1967). The neural mechanism of binocular depth discrimination. *Journal of Physiology, 193*, 327-342.

Bi, T., Cai, P., Zhou, T., & Fang, F. (2009). The effect of crowding on orientation-selective adaptation in human early visual cortex. *Journal of Vision, 9*, 1-10.

Blake, R., Tadin, D., Sobel, K. V., Raissian, T. A., & Chong, S. C. (2006). Strength of early visual adaptation depends on visual awareness. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 4783-4788.

Blakemore, C., & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology, 203*, 237-260.

Blakemore, C., & Tobin, E. A. (1972). Lateral inhibition between orientation detectors in the cat's visual cortex. *Experimental Brain Research, 15*, 439-440.

Bradley, A., Switkes, E., & De Valois, K. (1988). Orientation and spatial frequency selectivity of adaptation to color and luminance gratings. *Vision Research, 28*, 841-856.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433-436.

Carandini, M., Demb, J. B., Mante, V., Tolhurst, D. J., Dan, Y., Olshausen, B. A., Gallant, J. L., & Rust, N. C. (2005). Do we know what the early visual system does? *Journal of Neuroscience, 25*, 10577-10597.

Carandini, M., Heeger, D. J., & Movshon, J. A. (1997). Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience, 17*, 8621-8644.

Carandini, M., Movshon, J. A., & Ferster, D. (1998). Pattern adaptation and cross-orientation interactions in the primary visual cortex. *Neuropharmacology, 37*, 501-511.

Cavanaugh, J. R., Bair, W., & Movshon, J. A. (2002a). Nature and interaction of signals from the receptive field center and surround in macaque V1 neurons. *Journal of Neurophysiology, 88*, 2530-2546.

Cavanaugh, J. R., Bair, W., & Movshon, J. A. (2002b). Selectivity and spatial distribution of signals from the receptive field surround in macaque V1 neurons. *Journal of Neurophysiology, 88*, 2547-2556.

Clark, V. P., Fan, S., & Hillyard, S. A. (1994). Identification of early visual evoked potential generators by retinotopic and topographic analyses. *Human Brain Mapping, 2*, 170-187.

DeAngelis, G. C., Freeman, R. D., & Ohzawa, I. (1994). Length and width tuning of neurons in the cat's primary visual cortex. *Journal of Neurophysiology, 71*, 347-374.

Di Russo, F., Martinez, A., Sereno, M. I., Pitzalis, S., & Hillyard, S. A. (2002). Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping, 15*, 95-111.

Dobbins, A., Zucker, S. W., & Cynader, M. S. (1987). Endstopped neurons in the visual cortex as a substrate for calculating curvature. *Nature, 329*, 438-441.

Dragoi, V., Sharma, J., & Sur, M. (2000). Adaptation-induced plasticity of orientation tuning in adult visual cortex. *Neuron, 28*, 287-298.

Engel, S. A. (2005). Adaptation of oriented and unoriented color-selective neurons in human visual areas. *Neuron, 45*, 613-623.

Fang, F., Murray, S. O., Kersten, D., & He, S. (2005). Orientation-tuned FMRI adaptation in human visual cortex. *Journal of Neurophysiology, 94*, 4188-4195.

Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: evidence for a local "association field". *Vision Research, 33*, 173-193.

Fitzpatrick, D. (2000). Seeing beyond the receptive field in primary visual cortex. *Current Opinion in Neurobiology, 10*, 438-443.

Haynes, J. D., Roth, G., Stadler, M., & Heinze, H. J. (2003). Neuromagnetic correlates of perceived contrast in primary visual cortex. *Journal of Neurophysiology, 89*, 2655-2666.

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9*, 181-197.

Jones, J. P., & Palmer, L. A. (1987). An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *Journal of Neurophysiology, 58*, 1233-1258.

Joo, S. J., Boynton, G. M., & Murray, S. O. (2012). Long-range, pattern-dependent contextual effects in early human visual cortex. *Current Biology, 22*, 781-786.

Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. *Neuron, 15*, 843-856.

Kapadia, M. K., Westheimer, G., & Gilbert, C. D. (2000). Spatial distribution of contextual interactions in primary visual cortex and in visual perception. *Journal of Neurophysiology, 84*, 2048-2062.

Kastner, S., Nothdurft, H. C., & Pigarev, I. N. (1997). Neuronal correlates of pop-out in cat striate cortex. *Vision Research, 37*, 371-376.

Kinoshita, M., Gilbert, C. D., & Das, A. (2009). Optical imaging of contextual interactions in V1 of the behaving monkey. *Journal of Neurophysiology, 102*, 1930-1944.
Knierim, J. J., & van Essen, D. C. (1992). Neuronal responses to static texture patterns in area V1 of the alert macaque monkey. *Journal of Neurophysiology, 67*, 961-980.
Kohn, A., & Movshon, J. A. (2003). Neuronal adaptation to visual motion in area MT of the macaque. *Neuron, 39*, 681-691.
Kourtzi, Z., & Kanwisher, N. (2001). Representation of perceived object shape by the human lateral occipital complex. *Science, 293*, 1506-1509.
Kourtzi, Z., Tolias, A. S., Altmann, C. F., Augath, M., & Logothetis, N. K. (2003). Integration of local features into global shapes: monkey and human FMRI studies. *Neuron, 37*, 333-346.
Lamme, V. A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience, 15*, 1605-1615.
Larsson, J., Landy, M. S., & Heeger, D. J. (2006). Orientation-selective adaptation to first- and second-order patterns in human visual cortex. *Journal of Neurophysiology, 95*, 862-881.
Lee, T. S., & Mumford, D. (2003). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A, 20*, 1434-1448.
Lennie, P., & Movshon, J. A. (2005). Coding of color and form in the geniculostriate visual pathway (invited review). *Journal of the Optical Society of America A, 22*, 2013-2033.
Levitt, J. B., & Lund, J. S. (1997). Contrast dependence of contextual effects in primate visual cortex. *Nature, 387*, 73-76.
Li, X., Lu, Z. L., Xu, P., Jin, J., & Zhou, Y. (2003). Generating high gray-level resolution monochrome displays with conventional computer graphics cards and color monitors. *Journal of Neuroscience Methods, 130*, 9-18.
Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends Cogn Sci, 4*, 432-440.
Maffei, L., & Fiorentini, A. (1976). The unresponsive regions of visual cortical receptive fields. *Vision Research, 16*, 1131-1139.
Manassi, M., Sayim, B., & Herzog, M. H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *Journal of Vision, 12*, 1-14.
Mareschal, I., Sceniak, M. P., & Shapley, R. M. (2001). Contextual influences on orientation discrimination: binding local and global cues. *Vision Research, 41*, 1915-1930.
Movshon, J. A., & Lennie, P. (1979). Pattern-selective adaptation in visual cortical neurones. *Nature, 278*, 850-852.
Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978). Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. *Journal of Physiology, 283*, 101-120.
Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences of the United States of America, 99*, 15164-15169.
Murray, S. O., Olshausen, B. A., & Woods, D. L. (2003). Processing shape, motion and three-dimensional shape-from-motion in the human cortex. *Cerebral Cortex, 13*, 508-516.

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research, 29*, 1631-1647.

Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science, 257*, 1357-1363.

Nakayama, K., Shimojo, S., & Silverman, G. H. (1989). Stereoscopic depth: its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception, 18*, 55-68.

Nothdurft, H. C., Gallant, J. L., & Van Essen, D. C. (1999). Response modulation by texture surround in primate area V1: correlates of "popout" under anesthesia. *Visual Neuroscience, 16*, 15-34.

Olshausen, B. A., & Field, D. J. (2005). How close are we to understanding v1? *Neural Computation, 17*, 1665-1699.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision, 10*, 437-442.

Petrov, Y., Verghese, P., & McKee, S. P. (2006). Collinear facilitation is largely uncertainty reduction. *Journal of Vision, 6*, 170-178.

Polat, U., Mizobe, K., Pettet, M. W., Kasamatsu, T., & Norcia, A. M. (1998). Collinear stimuli regulate visual responses depending on cell's contrast threshold. *Nature, 391*, 580-584.

Polat, U., & Sagi, D. (1993). Lateral interactions between spatial channels: suppression and facilitation revealed by lateral masking experiments. *Vision Research, 33*, 993-999.

Polat, U., & Sagi, D. (1994). The architecture of perceptual spatial interactions. *Vision Research, 34*, 73-78.

Priebe, N. J., Churchland, M. M., & Lisberger, S. G. (2002). Constraints on the source of short-term motion adaptation in macaque area MT. I. the role of input and intrinsic mechanisms. *Journal of Neurophysiology, 88*, 354-369.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*, 1019-1025.

Riesenhuber, M., & Poggio, T. (2000). Models of object recognition. *Nature Neuroscience, 3 Suppl*, 1199-1204.

Ringach, D. L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology, 88*, 455-463.

Roelfsema, P. R., Lamme, V. A., & Spekreijse, H. (2004). Synchrony and covariation of firing rates in the primary visual cortex during contour grouping. *Nature Neuroscience, 7*, 982-991.

Rust, N. C., & Movshon, J. A. (2005). In praise of artifice. *Nature Neuroscience, 8*, 1647-1650.

Sayim, B., Westheimer, G., & Herzog, M. H. (2008). Contrast polarity, chromaticity, and stereoscopic depth modulate contextual interactions in vernier acuity. *Journal of Vision, 8*, 12 11-19.

Sceniak, M. P., Hawken, M. J., & Shapley, R. (2001). Visual spatial characterization of macaque V1 neurons. *Journal of Neurophysiology, 85*, 1873-1887.

Sceniak, M. P., Ringach, D. L., Hawken, M. J., & Shapley, R. (1999). Contrast's effect on spatial summation by macaque V1 neurons. *Nature Neuroscience, 2*, 733-739.

Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature Neuroscience, 4*, 819-825.

Shapley, R. (2004). A new view of the primary visual cortex. *Neural Networks, 17*, 615-623.

Sillito, A. M., Grieve, K. L., Jones, H. E., Cudeiro, J., & Davis, J. (1995). Visual cortical mechanisms detecting focal orientation discontinuities. *Nature, 378*, 492-496.

Solomon, S. G., Lee, B. B., & Sun, H. (2006). Suppressive surrounds and contrast gain in magnocellular-pathway retinal ganglion cells of macaque. Journal of Neuroscience, 26, 8715-8726.

Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. Perception & Psychophysics, 33, 113-120.

Zenger-Landolt, B., & Heeger, D. J. (2003). Response suppression in v1 agrees with psychophysics of surround masking. Journal of Neuroscience, 23, 6884-6893.

Zipser, K., Lamme, V. A., & Schiller, P. H. (1996). Contextual modulation in primary visual cortex. Journal of Neuroscience, 16, 7376-7389.

Vita

Sung Jun Joo was born in Seoul, Korea. He moved to Seattle the graduate school with his beloved wife, Eun Kyoung Lee, and now he has two sons, Daniel and Sean. He received a Bachelor of Engineering degree from Korea University, Seoul, and a Master of Science degree from Yonsei University, Seoul. In 2012 he earned a Doctor of Philosophy at the University of Washington in Psychology.