

Stimulus similarity modulates competitive interactions in human visual cortex

Diane M. Beck

Department of Psychology and Beckman Institute,
University of Illinois, Urbana-Champaign, IL, USA



Sabine Kastner

Department of Psychology and Center for the Study of Brain,
Mind and Behavior, Princeton University,
Princeton, NJ, USA



When multiple visual stimuli are simultaneously presented in a neuron's receptive field, they often interact with each other by mutually suppressing their visually evoked responses, suggesting that multiple stimuli present at the same time in the visual field compete for neural representation. Previous research has shown that these suppressive interactions can be biased by top-down influences such as spatially directed attention, as well as by the bottom-up factor of visual salience. Using fMRI, we asked whether competitive interactions might also be modulated by other bottom-up factors and tested the effects of stimulus similarity. Specifically, we found that suppressive interactions in area V4, measured by comparing activity evoked by simultaneous (potentially competing) and sequential (noncompeting) presentations, were reduced when four items were identical relative to when the four items differed in color and orientation. Such a result is consistent with the prediction that competition is more likely to occur between groups than within a group.

Keywords: crowding, competition, visual cortex, fMRI, attention

Citation: Beck, D. M., & Kastner, S. (2007). Stimulus similarity modulates competitive interactions in human visual cortex. *Journal of Vision*, 7(2):19, 1–12, <http://journalofvision.org/7/2/19/>, doi:10.1167/7.2.19.

Introduction

Typical visual scenes are crowded and cluttered with many different visual features and objects. However, at any particular point in time only a subset of this information reaches awareness or is stored in memory indicating that there is limited processing capacity within the visual system (Broadbent, 1958; Duncan, 1980; Treisman, 1969). Evidence from single-cell physiology and functional brain imaging studies suggest that multiple objects in cluttered visual scenes compete for neural representation due to limited processing capacity. That is, responses to a visual stimulus are reduced when the stimulus is presented together with a second stimulus in the same classic receptive field (RF; Miller, Gochin, & Gross, 1993; Recanzone, Wurtz, & Schwarz, 1997; Reynolds, Chelazzi, & Desimone, 1999; Snowden, Treue, Erickson, & Andersen, 1991) or together with multiple stimuli in nearby locations (Kastner, De Weerd, Desimone, & Ungerleider, 1998; Kastner et al., 2001) compared to when the stimulus is presented alone. These suppressive interactions were found to be most prominent in extrastriate areas where RFs are large enough to encompass multiple stimuli, suggesting that competitive interactions occur most strongly at the level of the RF (Kastner et al., 1998, 2001).

The biased competition theory of selective attention proposes that these competitive interactions among

multiple stimuli can be influenced by both top-down selective attention and by bottom-up stimulus-driven processes, such as visual salience. Previous work has provided evidence in favor of both these modulatory influences (Beck & Kastner, 2005; Kastner et al., 1998; Luck, Chelazzi, Hillyard, & Desimone, 1997; Moran & Desimone, 1985; Recanzone & Wurtz, 2000; Reynolds & Desimone, 2003; Reynolds et al., 1999). Here we ask whether there are other properties of the stimulus and the context in which it is presented that might influence suppressive interactions. Specifically, we probed the influence of stimulus similarity, which may implicate grouping mechanisms, on suppressive interactions among the stimuli.

Modulation of suppressive interactions by top-down mechanisms

A number of studies have suggested that suppressive interactions among multiple stimuli can be influenced by spatially directing attention to a single item in the display. For example, when a monkey directed attention to one of two competing stimuli within an RF, the responses in extrastriate areas V2, V4, and MT were as large as those

to the attended stimulus presented alone, indicating that directed attention reduced the suppressive influence of the competing stimulus (Recanzone & Wurtz, 2000; Reynolds et al., 1999). Similarly, in human fMRI studies, it was shown that directed attention to one of four competing stimuli enhanced responses evoked by competing (simultaneously presented) stimuli more than those evoked by noncompeting (sequentially presented) stimuli in areas V4 and TEO, suggesting that directed attention operated by counteracting the suppressive influences of nearby distracters (Kastner et al., 1998).

Modulation of suppressive interactions by bottom-up mechanisms

We have recently shown that suppressive interactions are not only modulated by top-down attention, but are also influenced by visual salience (Beck & Kastner, 2005). We found robust suppressive interactions among multiple stimuli in areas V2/VP and V4 when the stimuli were presented in the context of heterogeneous displays, in which four stimuli differed from each other. However, this suppression was eliminated when the same stimuli were presented in the context of pop-out displays, in which a single item differed from the other three.

According to biased competition theory, visual salience is not the only stimulus-driven factor that should influence suppressive interactions among multiple stimuli. Desimone and Duncan (1995) have suggested that suppressive interactions should occur *between* rather than *within* perceptual groups. The support for this prediction comes from behavioral experiments showing that adding distractors to a visual search display has little effect on visual search times if those distractors group with existing distractors, suggesting little competition between target and distractors, whereas the number of perceptual groups in a visual search display has a large effect on the speed with which subjects can find the target (Bundesen & Pedersen, 1983), suggesting competition between groups. Grouping in these experiments was induced by the Gestalt principle of color similarity and proximity (Wertheimer, 1923). In this study, we ask whether the perceptual grouping in a display influences suppressive interactions in visual cortex. In particular, we compared suppressive interactions among four identical items (homogeneous display; grouped) versus four heterogeneous stimuli that differed in both color and orientation (heterogeneous display; not grouped).

There is some evidence to suggest that suppressive interactions among homogeneous stimuli are minimal. As mentioned, although several studies have shown that when

four heterogeneous stimuli are simultaneously presented, they evoked significantly less activity in intermediate visual areas than when the same stimuli are presented sequentially (Beck & Kastner, 2005; Kastner et al., 1998, 2001). Beck and Kastner (2005) found no such difference between simultaneously versus sequentially presented stimuli when all four items were identical. However, in that study, the homogeneous displays were compared to pop-out displays and not to heterogeneous displays, precluding conclusions that suppressive interactions are modulated by stimulus homogeneity. In this study, suppressive interactions evoked by heterogeneous and homogeneous display types were compared directly.

In the heterogeneous display condition, four stimuli that differed in color and orientation were presented in nearby locations. In the homogeneous display condition, four stimuli that were identical in color and orientation were presented in the same locations (see Figure 1). Integrated over time, however, the physical stimulus conditions in each of the four locations were identical with the two display conditions; the only variable that was manipulated was the context of the visual stimuli. As in previous studies (Beck & Kastner, 2005; Kastner et al., 1998, 2001), suppressive interactions were assessed by comparing activity evoked by each display type under two presentation conditions: sequential and simultaneous. In the sequential condition, each stimulus was presented alone in one of the four locations. In the simultaneous condition, the stimuli were shown together in the four locations. Again, integrated over time, the amount of visual stimulation in each of the four locations was identical under the two conditions. However, suppressive interactions among stimuli within RFs could take place only in the simultaneous condition and not in the sequential condition. Based on predictions from biased competition theory, we expected to find a significant interaction of presentation condition and display type. Specifically, we predicted significant differences between sequentially and simultaneously presented heterogeneous stimuli, in accordance with previous studies (Beck & Kastner, 2005; Kastner et al., 1998, 2001) and weaker effects of homogeneous displays on suppressive interactions than heterogeneous displays in intermediate processing levels of extrastriate cortex.

Methods

Subjects

Seven subjects (four females; age: 21–38 years) participated in the study, which was approved by the Institutional Review Panel of Princeton University. All subjects were in good health with no past history of

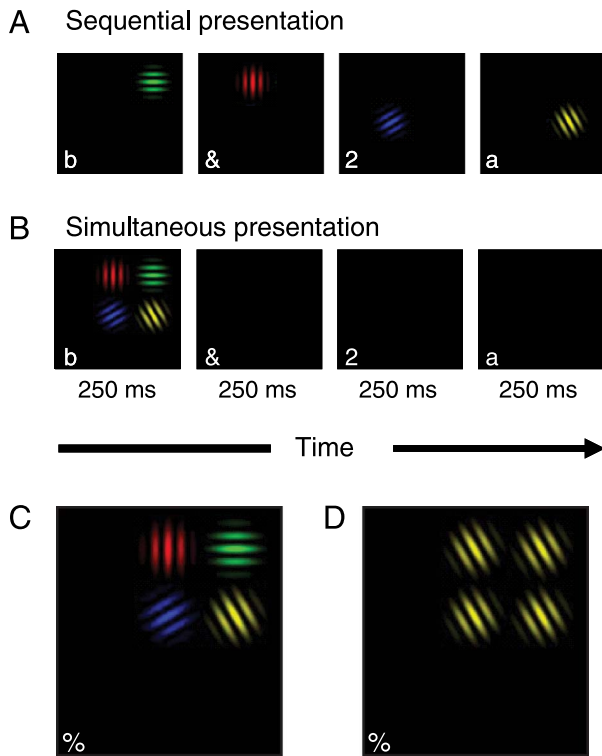


Figure 1. Four Gabor stimuli were presented in four nearby locations in the periphery of the upper right quadrant either (A) sequentially or (B) simultaneously. The displays were either (C) heterogeneous, in which all four stimuli differed in color and orientation, or (D) homogeneous, in which all four stimuli were identical. A stimulation period of 1 s, which was repeated in blocks of 16 s, is shown for a heterogeneous display (A and B). During all conditions, subjects detected target letters at fixation (illustrated in lower left corner of each display).

psychiatric or neurological diseases and gave their written informed consent. Subjects had normal or corrected-to-normal vision.

Visual stimuli and experimental design

Visual stimuli were four Gabor patches (wavelength, 0.47° ; standard deviation of Gaussian envelope, 0.73°) presented in four nearby locations in the upper right quadrant of the visual field at $6\text{--}10^\circ$ eccentricity from a fixation point. Each Gabor stimulus subtended approximately 2° of visual angle, and the entire four-stimulus display subtended approximately 4° . The stimuli were either red, blue, green, or yellow, and they had an orientation of 0° (vertical), 60° , 90° (horizontal), and 150° , respectively (Figure 1). Stimuli were shown under two presentation conditions: sequential (SEQ) and simultaneous (SIM). In the sequential condition, each of the four Gabor stimuli was presented alone in one of the four

locations for 250 ms (Figure 1A). In the simultaneous condition, the same four stimuli were presented together in the same four locations for 250 ms (Figure 1B). The order of the stimuli and locations was randomized but were integrated over time; the stimulation parameters at each location were identical for sequential and simultaneous presentations.

In addition to the two presentation conditions, two display types were used: heterogeneous (HET) and homogeneous (HOM). In the heterogeneous display condition, all four stimuli differed in both orientation and color (Figure 1C). In the homogeneous display condition, all four stimuli were identical (Figure 1D). However, in both cases the display types were blocked such that the same four colors and orientations occurred with equal probability in both blocks. That is, in both the homogeneous and the heterogeneous blocks, each color was presented four times in each of the four locations. Thus, integrated across an entire block the stimulation parameters at each location were identical for homogeneous and heterogeneous display conditions, and it was only the context in which they were presented that differed. All stimuli were presented on a dark background. Stimuli were generated on a Power Mac G4 using Matlab software (Mathworks, Natick, MA) and the Psychophysical Toolbox (Brainard, 1997; Pelli, 1997).

During a given scan, presentation conditions (SEQ vs. SIM) and display type conditions (HOM vs. HET) were combined to produce four block types of 16 s each (SEQ HOM, SEQ HET, SIM HOM, SIM HET) that were interleaved with blank periods of the same duration. Each scan began with a 24-s blank period and had a total duration of 152 s. Presentation conditions were presented in the sequence SEQ–SIM–SIM–SEQ, and display type conditions were counterbalanced across scans with the first two blocks being homogeneous on half the scans and heterogeneous on the other half.

Letters, digits, and keyboard symbols (0.5° in size) were presented for 250 ms in random order at 4 Hz at a central fixation point. The subjects' task was to count target letters (a, b, or c) at fixation throughout the scan. This letter counting task was designed to engage the subjects' attention at fixation and not at the peripherally presented stimuli, which were always ignored. Before being scanned, subjects participated in a training session outside the scanner to learn to fixate well over several minutes while performing the fixation task.

Data acquisition and analysis

Images were acquired with a 3-T head scanner (Allegra, Siemens, Erlangen, Germany) using a standard head coil. Subjects were comfortably placed on their backs with their heads surrounded by soft foam to reduce head movements. Data were acquired in a single scan session, lasting about 1.5 hr. In addition, retinotopic mapping was

performed for all subjects in a separate scan session. Functional images were taken with a gradient echo, echoplanar sequence (TR, 2 s; TE, 30 ms; flip angle, 90°; matrix, 64 × 64 voxels; FOV, 160 × 160 mm).

In addition, another high-resolution anatomical scan of the whole brain (MPRAGE; TR, 2.5 s; TE, 4.38 ms, flip angle, 8°; matrix, 256 × 256 voxels; FOV, 256 × 256 mm) was acquired for each subject to perform spatial normalization in BrainVoyager (Brain Innovation, Maastricht, Netherlands) and detailed retinotopic analyses.

Visual stimuli were projected from a PowerLite 7250 LCD projector (Epson, Long Beach, CA) onto a translucent screen located at the back of the scanner. Stimuli were viewed from inside the bore of the magnet via a mirror system attached to the head coil, providing a maximal visual angle of 28° × 36° and a viewing distance of 60 cm. Synchronization of the video presentation with the MR data acquisition was accomplished by synchronizing the display computer with the scanner using a trigger pulse.

Between-scan head movements were corrected by aligning each image to a reference image obtained in the middle of the session using Automatic Image Registration (AIR) software (Woods, Mazziotta, & Cherry, 1993). Statistical analyses were restricted to brain voxels with adequate signal intensity (average intensity of >20% of the maximum value across voxels). The first four images of each scan were excluded from further analyses. Statistical analyses were performed using multiple regression in the framework of the general linear model (Friston et al., 1994, 1995) with the National Institutes of Health functional imaging data analysis program (FIDAP) software. Square-wave functions matching the time course of the experimental design were defined as effects of interest in the multiple regression model. The square-wave functions contrasted (1) visual stimulation versus blank periods and (2) sequential versus simultaneous presentations. For each effect of interest, square-wave functions were convolved with a Gaussian model of the hemodynamic response (lag, 4.8 s; dispersion, 1.8 s) to generate idealized response functions, which were used as regressors in the multiple regression model. Additional regressors were included in the model to factor out between-run changes in mean intensity and within-run linear drifts.

Statistical maps were thresholded at a Z score of 2.33 ($p < .01$, corrected for multiple comparisons) for all subjects except one, whose statistical map was thresholded at a Z score of 3.07 ($p < .001$). Statistical maps were then overlaid on structural T1-weighted scans taken in the same session and in the same plane, and activity in visual cortex was assigned to retinotopically organized areas (see below).

For each subject, time series of fMRI intensities were averaged over all voxels activated in a given visual area during visual stimulation versus blank presentations. Mean signals were computed by averaging across peak

intensity values obtained in a given condition and are given as percent signal change, which was computed relative to the mean signal obtained during the last half of the blank presentation blocks. The first half of the blank presentation blocks were excluded from the baseline to minimize the contribution from the poststimulus undershoot. Time series data and mean signal changes were averaged across subjects and are presented as group data. To quantify the effect of sequential and simultaneous presentations, we separately computed a sensory suppression index [$SSI = (R_{SEQ} - R_{SIM}) / (R_{SEQ} + R_{SIM})$; R_{SEQ} is the mean signal change obtained during sequential presentation conditions and R_{SIM} is the mean signal change obtained during simultaneous presentation conditions] for the homogeneous and heterogeneous display types. Statistical significance of the SSI and mean signal changes were assessed using repeated measures ANOVAs and paired t tests with subject as the random variable.

Mapping visual areas

Retinotopic mapping was performed for each subject in a separate scanning session using procedures similar to those established by Sereno et al. (1995) and described in detail elsewhere (Kastner et al., 2001). Areas V1, V2, ventral V3 (referred to as VP), and V3A were identified by the alternating representations of the vertical and the horizontal meridians, which form the borders of these areas (e.g., DeYoe et al., 1996; Engel, Glover, & Wandell, 1997; Sereno et al., 1995). This was accomplished by presenting flickering checkerboards of high color and luminance contrast along the meridians. Area V4 was identified by its characteristic upper (UVF) and lower (LVF) visual field retinotopy. The UVF and LVF are separated in V4 and located medially and laterally, respectively, on the posterior part of the fusiform gyrus. Area V4 in this study likely corresponds to area V4 of McKeefry and Zeki (1997) and Wade, Brewer, Rieger, and Wandell (2002) and appears to overlap with V4 and V8 described by Hadjikhani, Liu, Dale, Cavanagh, and Tootell (1998). UVF and LVF retinotopy was accomplished by presenting the colorful, flickering checkerboards either to the upper or to the lower quadrants.

Behavioral control tasks

The efficiency of the letter counting task in engaging subjects' attentional resources at fixation was tested in behavioral control experiments outside the scanner, in which five of the seven subjects participated. The subject was placed in front of a CRT monitor, with his or her head in a chin rest at a distance of 60 cm from the screen (the same distance as in the scanner). As in the scanner, stimuli were displayed on a Power Mac G4 using Matlab software

and the Psychophysical Toolbox (Brainard, 1997; Pelli, 1997); all stimulus parameters were kept as similar as possible to those in the scanner. Subjects were tested in both a counting version of the task that was similar to the one they performed in the scanner and a reaction time (RT) version of the task, which was included because changes in the distribution of attention are known to affect RTs more than accuracy. In the counting version of the task, subjects reported the number of detected targets at the end of every 16-s block. Each subject completed 12 repetitions of each of the five different block types (blank, sequential homogeneous, simultaneous homogeneous, simultaneous heterogeneous, and sequential heterogeneous). The order of block types was the same as in the scanning session, except that a blank block only occurred in between each set of four visual presentation blocks. In the RT version of the task, instead of counting the target letters, subjects were asked to press a button as quickly as possible when the target appeared. In this experiment, subjects completed 10 repetitions of the same block sequences run in the scanner (see [Visual stimuli and experimental design](#) section), including the interleaved blank periods.

Error rate on the counting task was determined by computing the difference between the number of targets counted and the total number of targets presented. In the RT task, RTs were computed relative to the onset of the target stimulus. Mean error rate and RT were then computed as a function of block type for each subject and were submitted to a repeated measures ANOVA.

Results

The Gabor stimuli, as compared with blank intervals, evoked significant activity in areas V1, V2, VP, and V4, as determined based on retinotopic mapping. The locations of the activations were in the ventral parts of these areas in the left hemisphere, consistent with the locations of stimuli in the upper right visual field. In two subjects, area VP was insufficiently activated, so that the analysis for this area was based on data from the remaining five subjects. An analysis of mean signal changes across subjects confirmed a main effect of visual stimulation, $F(1, 4) = 277.52$, $p < .001$, and no interaction of visual stimulation and area, $F < 1$.

Heterogeneous displays

In accordance with previous findings (Beck & Kastner, 2005; Kastner et al., 1998, 2001), an analysis of the fMRI time series and the mean signal changes averaged across all subjects revealed that simultaneous presentations of

heterogeneous displays evoked less response than sequential presentations in areas V2, VP, and V4, $t(6) = 5.36$, $p < .01$, $t(4) = 5.02$, $p < .01$, and $t(6) = 6.41$, $p < .01$, respectively ([Figures 2 and 3A](#)), indicating mutually suppressive interactions among the stimuli in extrastriate visual areas. The difference in activations between sequential and simultaneous presentations gradually increased from area V1, which showed no difference, $t < 1$, to area V4. This gradual increase of the suppression effects across ventral visual areas is also reflected in the SSI, $F(3, 12) = 4.2$, $p < .05$. The SSI quantifies the differences in responses to sequential and simultaneous presentations. Positive values indicate stronger responses to sequential than to simultaneous presentations, negative values indicate the opposite, and values around 0 indicate the absence of response differences. The gradual increase of the SSI from V1 to V4 can be seen most clearly in [Figure 3B](#), with significantly larger suppressive effects in the latter areas (V1/V2 vs. VP/V4), $t(6) = 3.77$, $p < .01$. These data are in accordance with previous results (Kastner et al., 1998, 2001) and suggest that suppression is scaled to the increasing RF sizes of neurons in these areas.

Homogeneous displays

Results from the heterogeneous display condition confirmed response differences between sequentially and simultaneously presented stimuli suggesting the existence of suppressive interactions among nearby stimuli in extrastriate cortex, but the critical question in our study was whether homogeneous displays produced similar suppressive interactions. We submitted the mean signal changes for each subject to a repeated measures ANOVA on presentation condition (sequential vs. simultaneous) and display type condition (homogeneous vs. heterogeneous). This analysis revealed a significant main effect of presentation condition for all areas except area V1; V2: $F(1, 6) = 7.48$, $p < .05$; VP: $F(1, 4) = 33.80$, $p < .01$; V4: $F(1, 6) = 11.84$, $p < .05$ ([Figures 2 and 3A](#)). However, importantly, there was a significant interaction of the presentation condition and display type condition in area V4, $F(1, 6) = 7.57$, $p < .05$, such that the response differences between sequential and simultaneous presentation conditions were smaller for homogeneous relative to heterogeneous displays. In fact, in area V4, simultaneous presentations of the homogeneous displays evoked similar responses as sequential presentations of the homogeneous displays (1.73% vs. 1.77% signal change, respectively, $t < 1$). The interaction of presentation condition and display type condition was most clearly seen in comparing the SSI for heterogeneous and homogeneous displays ([Figure 3B](#)). Although the suppressive effects also appeared to be reduced for homogeneous displays in V2 and VP, the interaction of presentation condition and display type condition did not reach

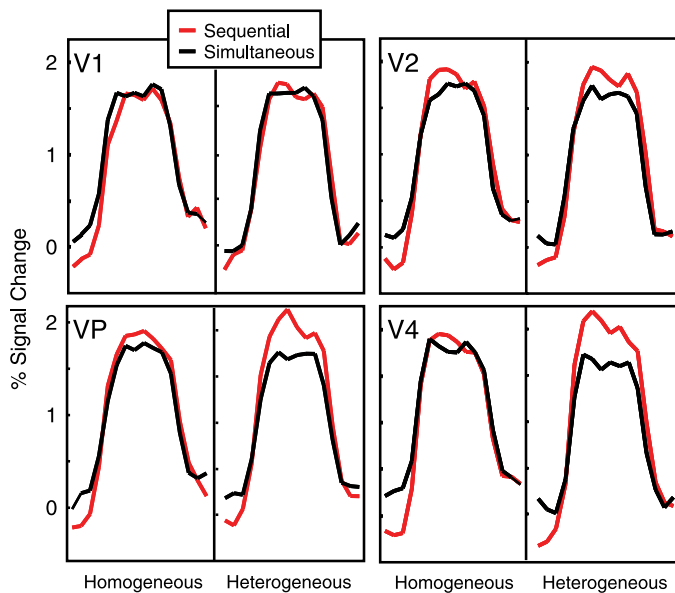


Figure 2. Time series of fMRI signals in visual cortex ($n = 7$). Red curves indicate activity evoked by sequential presentations and black curves that evoked by simultaneous presentations for homogeneous and heterogeneous displays.

significance in either area, V2: $F < 1$ and VP: $F(1, 6) = 2.4$. In other words, the reduction in suppressive effects was only significant in V4, suggesting that the mechanisms responsible for the reduced suppression effect for homogeneous displays may be specific to V4, depending perhaps on the larger RF sizes of that area. In area V1, there were no significant effects of presentation condition for either heterogeneous or homogeneous displays.

Behavioral results

Despite the difficulty of the letter counting task, subjects' performance in the scanner was reasonably good. The average accuracy rate across subjects was 90% (range: 78–94%). During scanning sessions, subjects reported a single number at the end of a scan. Thus, we were not able to assess performance as a function of the different peripheral display conditions. It is possible that the obtained fMRI effects were due to differences in the degree to which the peripheral displays captured attention. Such an effect should be reflected in subjects' performance on the letter counting task, with poorer performance in those conditions that more strongly captured attention. Five of the seven subjects' participated in both a counting and RT version of the task performed outside the scanner, during which performance was assessed as a function of the five stimulus conditions. The counting task was included because it most closely matched the task in the scanner, the only difference being that subjects reported

the number of targets at the end of every 16-s block rather than at the end of a 152-s four-block run. Not surprisingly, due to the fact that subjects were able to concentrate their efforts in short blocks in this behavioral version of the task, subjects made very few errors, and indeed there were no differences in error rate across block types, $F(4, 16) = 1.08$, ns (Table 1).

The RT task was included in an effort to obtain a more sensitive measure of attentional capture by the peripheral stimulus displays. Moreover, the same continuous block structure was used as in the scanner; the only difference was that subjects pressed a button in response to the targets rather than count them. An analysis of subjects RTs to correctly detect a target also showed no differences across block types, $F(4, 16) = 1.22$, ns (Table 1). The comparison between the simultaneous heterogeneous and simultaneous homogeneous conditions is of particular interest because these conditions led to different evoked responses in the fMRI experiment. If these differences could be attributed to difference in the extent to which the peripheral stimuli automatically attracted attention to themselves, then we might expect subjects' letter detection

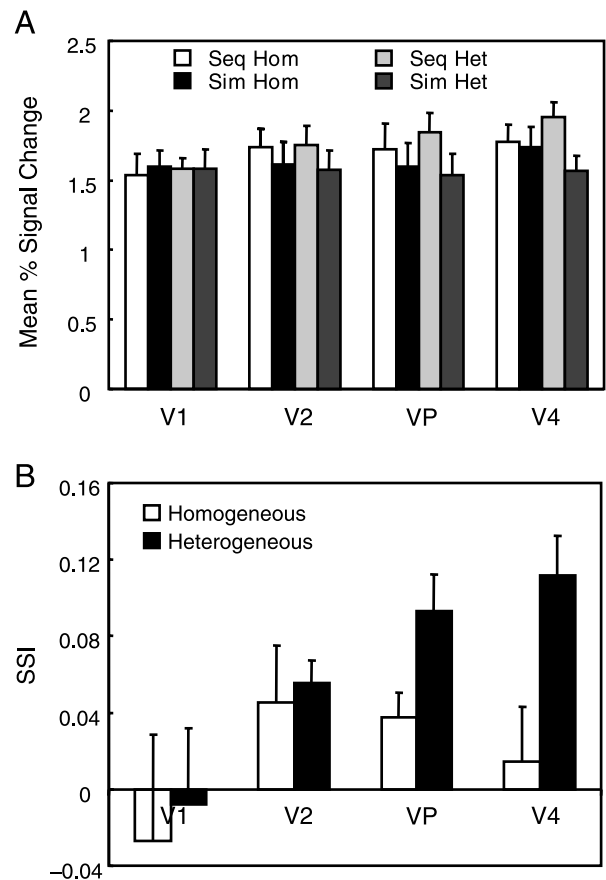


Figure 3. (A) Mean signal changes for each area, and each of the four conditions was averaged across subjects ($n = 7$). (B) Sensory suppression indexes (SSIs) were derived from the data shown in panel A. Vertical bars indicate SEM.

Stimulus condition	Percent error ($\pm SE$)	RT ($\pm SE$)
Blank	6 (3)	487 (20)
Sequential homogeneous	3 (1)	481 (17)
Simultaneous homogeneous	3 (1)	491 (20)
Sequential heterogeneous	4 (2)	491 (19)
Simultaneous heterogeneous	4 (1)	494 (20)

Table 1. Error rates for counting task and reaction times for RT task performed outside the scanner.

performance in these two conditions to differ. However, paired t tests for both the counting data and the RT data do not support such an explanation, $t(4) = 1.02$, ns , and $t < 1$, respectively.

Discussion

In accordance with previous data, simultaneous presentation of four heterogeneous visual stimuli significantly evoked less activity in areas V2, VP, and V4 than the same stimuli presented sequentially. Such a result is consistent with the idea that stimuli compete for neural representation. Furthermore, the reduction in suppressive effects with smaller RF sizes, with the suppressive effect disappearing in V1, is consistent with the hypothesis that competition is occurring at the level of the RF. When the four stimuli were identical, however, the suppression was considerably reduced relative to the heterogeneous conditions. The interaction of presentation condition and display type was most evident in V4, but a similar pattern was found in V2 and VP. Such a result suggests that competition is sensitive to the context in which the stimuli are presented: heterogeneous stimuli evoke more competition than homogeneous stimuli.

Grouping

As mentioned in the introduction, biased competition theory predicts that suppressive interactions should occur *between* rather than *within* perceptual groups (Desimone & Duncan, 1995). A homogenous group of stimuli is likely to form a strong perceptual group based on similarity (Wertheimer, 1923), and thus we might not expect much competition or suppressive interactions. On the other hand, stimuli that differ in both orientation and color are less likely to be perceived as a single group and thus more susceptible to suppressive interactions. Thus, the pattern of results is consistent with the predictions of Desimone and Duncan (1995).

There are two ways in which grouping may play a role in effects we found with homogeneous stimuli. Grouping

may occur at the same level as the competition, for instance in area V4. In fact, grouping may even be a consequence of the competition: If less competition is evoked by similar items, there is no need to select or filter any one of the items and instead the items are processed as a group. Alternatively, grouping mechanisms from elsewhere in the cortex may boost the activity related to the set of stimuli as it enters V4, effectively counteracting any competition that may have occurred between stimuli. Such a perspective is consistent with effects of grouping and figure-ground segmentation found as early as in V1 and V2 (Kapadia, Ito, Gilbert, & Westheimer, 1995; Kastner, Nothdurft, & Pigarev, 1999; Lamme, 1995; Nothdurft, Gallant, & Van Essen, 1999; Zhou, Friedman, & von der Heydt, 2000).

It is also possible that these effects may be modulated by later areas involved in attention. However, this does not seem likely in our paradigm. Throughout the experiment, subjects ignored the heterogeneous and the homogeneous stimuli and instead were occupied with a demanding letter counting task at fixation. Moreover, behavioral experiments using the same stimuli and letter detection task showed no effect of the peripheral stimuli on subjects' ability to count or to rapidly detect target letters. Importantly, there was no difference in subjects' performance on the letter detection task when homogeneous stimuli versus heterogeneous stimuli appeared in the periphery, suggesting that the two sets of stimuli did not differentially capture attention. These results are consistent with the idea that some aspects of perceptual organization proceed in an automatic preattentive fashion (Driver, Baylis, & Rafal, 1992; Duncan, 1984; Kahneman & Henik, 1981; Moore & Egeth, 1997; Neisser, 1967; Treisman 1982).

Of course, our results should not be taken as evidence that grouping is necessarily preattentive. There is behavioral evidence to suggest that our perception of grouping and figure-ground relies on a sophisticated interplay of multiple visual processes located throughout the processing stream (Beck & Palmer, 2002; Ben-Av, Sagi, & Braun, 1992; Mack, Tang, Tuma, Kahn, & Rock, 1992; Palmer, Neff, & Beck, 1996; Palmer & Nelson, 2000; Peterson, Harvey, & Weidenbacher, 1991; Rock & Brosgole, 1964; Rock, Nijhawan, Palmer, & Tudor, 1992), and thus perceptual organization is best thought of as a result of both feedforward and feedback mechanisms. However, we should point out that, regardless of whether grouping involves feedback or recurrent connections, these processes have their origin in the parameters of the visual stimulus, and thus grouping can still be considered primarily stimulus driven. Moreover, the effects found here should be contrasted sharply with the top-down effects of directed attention (Kastner et al., 1998; Luck et al., 1997; Moran & Desimone, 1985; Motter, 1993; Recanzone & Wurtz, 2000; Reynolds et al., 1999). In the directed attention studies, the subjects were required to "select" one of several competing targets. The

result was a reduction in suppressive effects. In the case of the homogeneous stimuli, selection is not necessary. Indeed, it is more difficult to select one of a group of similar items (Driver & Baylis, 1989; Duncan & Humphreys, 1989; Kramer & Jacobson, 1991) than one of a set of heterogeneous items. Instead, the reduced competition found in homogeneous displays is more likely due to aggregative mechanisms rather than segregative ones.

Context effects

There have been numerous physiology studies to show that the activity evoked by a stimulus is often dependent on the visual context in which it occurs (Hegd  & Felleman, 2003; Kapadia et al., 1995; Kastner et al., 1999; Knierim & Van Essen, 1992). However, these experiments have all been confined to primary visual cortex. Here we show effects of context in later visual areas. Indeed, our context effect is greatest in V4, reduced in V2 and VP, and nonexistent in V1. This latter result is consistent with the single-cell results of Knierim and Van Essen (1992). They also failed to find a difference between heterogeneous (randomly oriented) and homogeneous textures in V1 at the single-cell level. Thus, it may be that contextual effects related to stimulus similarity only occur in later areas where RFs are large enough to contain multiple stimuli.

We also wish to point out another significant difference between the current study and previous studies on context effects. Unlike the effects presented here, previous studies on context effects have been concerned with influences of contextual stimuli placed outside the classic RF. The competitive interactions that we are referring to occur within the RF. Indeed, our finding of decreased suppression for homogeneous stimuli relative to heterogeneous stimuli is in marked contrast to the increased suppression induced by similar stimuli, as compared to contrasting stimuli, presented outside the RF (Kastner et al., 1999; Knierim & Van Essen, 1992; Schein & Desimone, 1990). If such *surround suppression* was at work here, it would work against, rather than cause, our observed effects.

RF effects

The results obtained with the heterogeneous displays are consistent with both single-cell data as well as previous fMRI experiments that use the same sequential/simultaneous paradigm to define competitive interactions among multiple stimuli (Kastner et al., 1998, 2001; Reynolds et al., 1999), suggesting that competitive interactions scale to the size of the RF. Specifically, the SSI increased with increasing RF size. The idea that suppressive interactions are scaled to RF size was directly tested in a study in which the spatial separation among the

stimuli was increased (Kastner et al., 2001). According to the RF hypothesis, the magnitude of the suppressive interactions should be inversely related to the degree of spatial separation among the stimuli. In agreement with this idea, separating the stimuli by 4° abolished suppressive interactions in V2, reduced them in V4, but did not affect them in TEO. Separating the stimuli by 6° led to a further reduction of suppression effects in V4 but again had no effect in TEO.

The effects of stimulus similarity shown here may also depend on RF size, in that the interaction of display type (homogeneous vs. heterogeneous) and presentation condition was only significant in V4 where RF sizes are the largest. The interaction approached significance in VP, $F(1, 4) = 5.97$, $p = .07$, but was clearly not significant in V2 and V1. Such a pattern of results is consistent with the idea that the mechanisms responsible for reduced suppressive interactions may also be operating at the level of the RF, such that they require multiple stimuli to fall within the same RF.

Finally, as mentioned above, the fact that the effects of homogeneous stimuli found here (i.e., decreased suppression) are in opposition to those found from outside the RF (i.e., increased surround suppression) is consistent with the idea that our effects stem from interactions within the RF rather than outside it.

Other differences between sequential and simultaneous presentations

There are a number of differences, other than their ability to induce suppressive interaction, between sequential and simultaneous presentations. For example, the visual presentation period of each trial extends over an entire 1 s during the sequential presentation condition and over 250 ms in the sequential condition. The sequential condition also contains four onsets to every one onset in simultaneous condition. However, there are a number of results that speak against accounts that seek to explain the differential activity evoked by sequential and simultaneous in these terms. If the stimulation duration or the number of onsets are responsible for the response difference between sequential and simultaneous presentations, then this should be true for all stimulus configurations. However, as noted above, this response difference decreases with increasing spatial separation (Kastner et al., 2001). Similarly, in both the current experiment and the previous experiments on visual salience (Beck & Kastner, 2005), the differential stimulation durations and onset factor do not change, and yet there is a clear reduction in the differential activity evoked by the homogeneous and pop-out displays relative to the heterogeneous display, just as biased competition predicts. Finally, allocating attention to the stimulus location closest to fixation resulted in greater response enhancement in the simultaneous condition than

in the sequential condition (Kastner et al., 1998). This is in agreement with a prediction from single-cell physiology that the allocation of spatial attention counteracts competitive interactions induced by nearby distracters. It is not clear why a single 250-ms stimulus should be more modulated by attention than four sequential presentations of the same duration, but such a result is expected according to single-cell physiology that has shown that the allocation of spatial attention counteracts competitive interactions induced by nearby distracter.

Finally, the sequential condition affords the possibility of apparent motion, whereas the simultaneous condition does not. However, it is not clear how such an effect would evoke differential activity within the ventral stream. Moreover, again, an account based on apparent motion could not explain the diminished difference for homogeneous and pop-out displays relative to heterogeneous displays, as all three conditions would be subject to apparent motion considerations in the sequential condition.

Relationship to monkey physiology studies

Single-cell physiology studies in monkeys (Moran & Desimone, 1985; Reynolds et al., 1999) have found that multiple stimuli simultaneously presented in a neuron's RF are not independently processed but interact with each other in a suppressive fashion. More specifically, it was shown that the response of V4 neurons to a pair of stimuli is best described as a weighted average of the responses of the two stimuli when presented alone.

It is difficult to directly extrapolate from what is known at the single-cell level to a response at the population level (such as the BOLD). The BOLD signal may be sensitive to interactions among neighboring neurons that will not necessarily be apparent at the single-cell level. In fact, even the spiking activity of multiple neurons may not predict the BOLD response, as it has been found that local field potentials are a better predictor of the BOLD response than multiple-unit spiking activity (Logothetis et al., 2001). Moreover, all neurons are presumably not weighted equally at the population level. For instance, the population response should be dominated by stimuli presented closest to the fovea due to cortical magnification of more foveal stimuli.

However, we argue that our results are qualitatively in agreement with the single-cell data. Although the population response will be dominated by the stimulus closest to fixation, in the simultaneous heterogeneous case the three more peripheral stimuli would serve to reduce the response of the more central stimulus, via competitive interactions. Whereas, in the sequential case, there are no competing stimuli with which to reduce the dominant response elicited by the stimulus closest to fixation. Moreover, there is an increasing number of experiments using similar paradigms that support the interpretation that

the suppression associated with simultaneous presentations is due to competitive interactions at the level of the RF. As mentioned, suppressive effects decrease with increasing spatial separation between stimuli and increase with increasing RF size (Beck & Kastner, 2005; Kastner et al., 1998, 2001). Allocating attention to one of the competing stimuli reduces suppressive effects (Kastner et al., 1998) and the effects of attention on simultaneously presented stimuli decrease with increasing spatial separation (Bles, Schwarzbach, De Weerd, Goebel, & Jansma, 2006), in agreement with predictions from single-cell physiology that the allocation of spatial attention counteracts competitive interactions at the level of the RF (Luck et al., 1997; Moran & Desimone, 1985; Recanzone & Wurtz, 2000; Reynolds et al., 1999). Finally, stimulus salience (Beck & Kastner, 2005) and stimulus similarity, as shown in the experiments presented here, modulate the suppressive effects associated with simultaneously presented stimuli, in agreement with predictions from biased competition theory that bottom-up stimulus-driven factors can influence competitive interactions.

It should also be noted that it is possible to draw the same qualitative relationship between the single-cell data and our own in the case of the homogeneous displays. Specifically, the lack of a difference between simultaneous and sequential presentations for the homogeneous displays may reflect the fact that response of V4 neurons to a pair of stimuli is best described as a weighted average of the responses to the two stimuli when presented alone (Luck et al., 1997; Reynolds et al., 1999). If the two stimuli that comprise the pair, each produces identical responses when presented alone, then the response to the pair is indistinguishable from the response to each of the individual stimuli (Reynolds et al., 1999), just as the weighted average model would predict. Thus, we may not need to appeal to additional grouping mechanisms to explain our data. Instead, the reduced competition present in the homogeneous displays, relative to the heterogeneous displays, may simply be the result of the averaging procedure performed by the cell. However, such an explanation is still consistent with the hypothesis that grouping may be a *consequence* of competition at the level of the RF. Specifically, it may be the lack of competition that is responsible for the grouping. Without strong competitive interactions, there is no need to filter one item at the expense of the others and instead the items may be processed as a group.

Modulation of suppressive interactions by bottom-up factors

Suppressive interactions are not only modulated by top-down attention, but also an expanding list of bottom-up factors. Both stimulus salience (Beck & Kastner, 2005) and spatial proximity (Kastner et al., 2001) modulate

suppressive interactions in intermediate level visual areas. To this list, this study adds stimulus similarity: identical stimuli result in less suppressive interactions among stimuli, than heterogeneous stimuli do.

Conclusions

We have shown that competition for neural representation is reduced in the context of homogeneous stimuli relative to heterogeneous stimuli. This effect may reflect early grouping operations that provide initial structure to complex scenes and suggests that competition is not only biased by top-down factor of directed attention but is also modulated by the bottom-up stimulus-driven factor of stimulus similarity.

Acknowledgments

This research was supported by grants from the National Institutes of Mental Health (RO1 MH-64043, P50 MH-62196) and the Whitehall Foundation.

Commercial relationships: none.

Corresponding author: Diane M. Beck.

Email: dmbeck@uiuc.edu.

Address: Beckman Institute, 405 North Mathews Avenue, Urbana, IL 61801.

References

- Beck, D. M., & Kastner, S. (2005). Stimulus context modulates competition in human extrastriate cortex. *Nature Neuroscience*, 8, 1110–1116. [[PubMed](#)] [[Article](#)]
- Beck, D. M., & Palmer, S. E. (2002). Top-down influences on perceptual grouping. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 1071–1084. [[PubMed](#)]
- Ben-Av, M. B., Sagi, D., & Braun, J. (1992). Visual attention and perceptual grouping. *Perception & Psychophysics*, 52, 277–294. [[PubMed](#)]
- Bles, M., Schwarzbach, J., De Weerd, P., Goebel, R., & Jansma, B. M. (2006). Receptive field size-dependent attention effects in simultaneously presented stimulus displays. *Neuroimage*, 30, 506–511. [[PubMed](#)]
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10, 433–436. [[PubMed](#)]
- Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon.
- Bundesen, C., & Pedersen, L. F. (1983). Color segregation and visual search. *Perception & Psychophysics*, 33, 487–493. [[PubMed](#)]
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222. [[PubMed](#)]
- DeYoe, E. A., Carman, G. J., Bandettini, P., Glickman, S., Wieser, J., Cox, R., et al. (1996). Mapping striate and extrastriate visual areas in human cerebral cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 2382–2386. [[PubMed](#)] [[Article](#)]
- Driver, J., & Baylis, G. C. (1989). Movement and visual attention: The spotlight metaphor breaks down. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 448–456. [[PubMed](#)]
- Driver, J., Baylis, G. C., & Rafal, R. D. (1992). Preserved figure-ground segregation and symmetry perception in visual neglect. *Nature*, 360, 73–75. [[PubMed](#)]
- Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, 87, 272–300. [[PubMed](#)]
- Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113, 501–517. [[PubMed](#)]
- Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96, 433–458. [[PubMed](#)]
- Engel, S. A., Glover, G. H., & Wandell, B. A. (1997). Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cerebral Cortex*, 7, 181–192. [[PubMed](#)] [[Article](#)]
- Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C., Frackowiak, R. S., et al. (1995). Analysis of fMRI time-series revisited. *Neuroimage*, 2, 45–53. [[PubMed](#)]
- Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J. P., Frith, C. D., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2, 189–210.
- Hadjikhani, N., Liu, A. K., Dale, A. M., Cavanagh, P., & Tootell, R. B. (1998). Retinotopy and color sensitivity in human visual cortical area V8. *Nature Neuroscience*, 1, 235–241. [[PubMed](#)]
- Hegd e, J., & Felleman, D. J. (2003). How selective are V1 cells for pop-out stimuli? *Journal of Neuroscience*, 23, 9968–9980. [[PubMed](#)] [[Article](#)]
- Kahneman, D., & Henik, A. (1981). Perceptual organization and attention. In M. Kubovy & J. R. Pomerantz

- (Eds.), *Perceptual organization* (pp. 181–211). Hillsdale, NJ: Erlbaum.
- 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. [[PubMed](#)] [[Article](#)]
- Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science*, *282*, 108–111. [[PubMed](#)]
- Kastner, S., De Weerd, P., Pinsk, M. A., Elizondo, M. I., Desimone, R., & Ungerleider, L. G. (2001). Modulation of sensory suppression: Implications for receptive field sizes in the human visual cortex. *Journal of Neurophysiology*, *86*, 1398–1411. [[PubMed](#)] [[Article](#)]
- Kastner, S., Nothdurft, H. C., & Pigarev, I. N. (1999). Neuronal responses to orientation and motion contrast in cat striate cortex. *Visual Neuroscience*, *16*, 587–600. [[PubMed](#)]
- 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. [[PubMed](#)]
- Kramer, A. F., & Jacobson, A. (1991). Perceptual organization and focused attention: The role of objects and proximity in visual processing. *Perception & Psychophysics*, *50*, 267–284. [[PubMed](#)]
- Lamme, V. A. (1995). The neurophysiology of figure-ground segregation in primary visual cortex. *Journal of Neuroscience*, *15*, 1605–1615. [[PubMed](#)] [[Article](#)]
- Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., & Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature*, *412*, 150–157. [[PubMed](#)]
- Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology*, *77*, 24–42. [[PubMed](#)] [[Article](#)]
- Mack, A., Tang, B., Tuma, R., Kahn, S., & Rock, I. (1992). Perceptual organization and attention. *Cognitive Psychology*, *24*, 475–501. [[PubMed](#)]
- McKeefry, D. J., & Zeki, S. (1997). The position and topography of the human colour centre as revealed by functional magnetic resonance imaging. *Brain*, *120*, 2229–2242. [[PubMed](#)] [[Article](#)]
- Miller, E. K., Gochin, P. M., & Gross, C. G. (1993). Suppression of visual responses of neurons in inferior temporal cortex of the awake macaque by addition of a second stimulus. *Brain Research*, *616*, 25–29. [[PubMed](#)]
- Moore, C. M., & Egeth, H. (1997). Perception without attention: Evidence of grouping under conditions of inattention. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 339–352. [[PubMed](#)]
- Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*, 782–784. [[PubMed](#)]
- Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of Neurophysiology*, *70*, 909–919. [[PubMed](#)]
- Neisser, U. (1967). *Cognitive psychology*. East Norwalk, CT, US: Appleton-Century-Crofts.
- 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. [[PubMed](#)]
- Palmer, S. E., Neff, J., & Beck, D. (1996). Late influences on perceptual grouping: Amodal completion. *Psychonomic Bulletin & Review*, *3*, 75–80.
- Palmer, S. E., & Nelson, R. (2000). Late influences on perceptual grouping: Illusory figures. *Perception & Psychophysics*, *62*, 1321–1331. [[PubMed](#)]
- Pelli, D. G. (1997). The VideoToolBox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442. [[PubMed](#)]
- Peterson, M. A., Harvey, E. M., & Weidenbacher, H. J. (1991). Shape recognition contributions to figure-ground reversal: Which route counts? *Journal of Experimental Psychology: Human Perception & Performance*, *17*, 1075–1089. [[PubMed](#)]
- Recanzone, G. H., & Wurtz, R. H. (2000). Effects of attention on MT and MST neuronal activity during pursuit initiation. *Journal of Neurophysiology*, *83*, 777–790. [[PubMed](#)] [[Article](#)]
- Recanzone, G. H., Wurtz, R. H., & Schwarz, U. (1997). Responses of MT and MST neurons to one and two moving objects in the receptive field. *Journal of Neurophysiology*, *78*, 2904–2915. [[PubMed](#)] [[Article](#)]
- Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience*, *19*, 1736–1753. [[PubMed](#)] [[Article](#)]
- Reynolds, J. H., & Desimone, R. (2003). Interacting roles of attention and visual salience in V4. *Neuron*, *37*, 853–863. [[PubMed](#)] [[Article](#)]
- Rock, I., & Brosgole, L. (1964). Grouping based on phenomenal proximity. *Journal of Experimental Psychology*, *67*, 531–538. [[PubMed](#)]

- Rock, I., Nijhawan, R., Palmer, S., & Tudor, L. (1992). Grouping based on phenomenal similarity of achromatic color. *Perception*, *21*, 779–789. [[PubMed](#)]
- Schein, S. J., & Desimone, R. (1990). Spectral properties of V4 neurons in the macaque. *Journal of Neuroscience*, *10*, 3369–3389. [[PubMed](#)] [[Article](#)]
- Sereno, M. I., Dale, A. M., Reppas, J. B., Kwong, K. K., Belliveau, J. W., Brady, T. J., et al. (1995). Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, *268*, 889–893. [[PubMed](#)]
- Snowden, R. J., Treue, S., Erickson, R. G., & Andersen, R. A. (1991). The response of area MT and V1 neurons to transparent motion. *Journal of Neuroscience*, *11*, 2768–2785. [[PubMed](#)] [[Article](#)]
- Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance*, *8*, 194–214. [[PubMed](#)]
- Treisman, A. M. (1969) Strategies and models of selective attention. *Psychological Review*, *76*, 282–299. [[PubMed](#)]
- Wade, A. R., Brewer, A. A., Rieger, J. W., & Wandell, B. A. (2002). Functional measurements of human ventral occipital cortex: Retinotopy and colour. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, *357*, 963–973. [[PubMed](#)] [[Article](#)]
- Wertheimer, M. (1923). Untersuchungen zur lehre von der Gestalt: II. *Psychologische Forschung*, *4*, 301–350. [Partial translation in W. D. Ellis (Ed.), *A sourcebook of Gestalt psychology* (pp. 71–81). New York: Humanities Press].
- Woods, R. P., Mazziotta, J. C., & Cherry, S. R. (1993). MRI-PET registration with automated algorithm. *Journal of Computer Assisted Tomography*, *17*, 536–546. [[PubMed](#)]
- Zhou, H., Friedman, H. S., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *Journal of Neuroscience*, *20*, 6594–6611. [[PubMed](#)] [[Article](#)]