UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Effects of spatial attention on crowded peripheral processing

Permalink
https://escholarship.org/uc/item/0nf8h1gp

Author
Bowen, Joel Donald

Publication Date
2022

Peer reviewed|Thesis/dissertation

Effects of spatial attention on crowded peripheral processing

by

Joel D Bowen

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Vision Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Michael Silver, Chair
Professor Dennis Levi
Professor Bruno Olshausen

Summer 2022

Effects of spatial attention on crowded peripheral processing

Abstract

Effects of spatial attention on crowded peripheral processing

by

Joel D Bowen

Doctor of Philosophy in Vision Science

University of California, Berkeley

Professor Michael Silver, Chair

The topics of this dissertation center around the facilitation or hindering of performance on cluttered peripheral vision tasks by spatial attention. I focus primarily on two well-characterized visual phenomena: visual crowding and ensemble perception. Visual crowding is defined as the detrimental effect of clutter on object recognition in peripheral vision. Ensemble perception is defined as the ability to quickly extract a summary representation from a set of similar objects. These two phenomena seem to complement each other. For example, peripheral vision is very limited in identifying individual features and objects due to crowding, but it excels at quickly obtaining a summary-statistical representation due to ensemble perception. The roles of spatial attention in visual crowding and ensemble perception, and the commonalities between them on an individual and a group level, are not as well understood.

In the first study, we investigated the effects of involuntary and voluntary attention on crowding. Visual spatial attention can be allocated in two distinct ways: one that is involuntarily captured by salient external stimuli, and one that is voluntarily directed to behaviorally relevant locations in the world. We used an anti-cueing paradigm to separately measure the effects of involuntary and voluntary spatial attention on the critical spacing of crowding, defined as the the minimum target/flanker spacing at which the target is correctly identified at a specified level of performance. We found that involuntary capture of attention led to faster response times (RTs) and smaller critical spacing, while voluntary allocation of attention led to faster RTs but no significant effect on critical spacing.

In the second study, we expanded on the first study by comparing the effects of different sizes of a peripheral involuntary attention cue on performance of an orientation discrimination crowding task and performance of an ensemble mean orientation discrimination task. The stimuli consisted of a central Gabor surrounded by a ring of uniformly-spaced Gabors. We varied the relationship between the central Gabor's orientation and the mean orientation of the entire stimulus array to see how the size of the cues interacted with different patterns of ensemble statistics (Gabor orientations) relative to the orientation of a single cued object within the ensemble. We found that

only the small cue decreased the effect of crowding, while the large cue that encompassed both the target and flankers did not. We also found moderate differential effects of cue size on ensemble perception, but this interaction was primarily observed if the cued Gabor was more salient than the non-cued Gabors within the ensemble.

In the final study, we expanded on the analysis of the second study by fitting response models of different complexities to the psychophysical data. The goal of the modeling approach was to investigate how observers utilized all of the orientations in the stimulus array to make their responses, to determine the role of task relevance in the response process, and to test if common strategies existed between the two tasks. We found that response patterns were better explained by a combination of the Gabor orientations considered independently, as opposed to a spatial-weighted average of these orientations. We also found that spatial-weighting strategies, inferred based on the values of model parameters, were correlated between the two tasks, even though task performance was not correlated for the two tasks.

Contents

List of Figures

Acknowledgments

I am grateful to my friends/family for the support over the years. Liz Lawler for the emotional support and always being the first one to bounce ideas off of. Justin Theiss for supporting me in joining the lab and being a fun collaborator. My advisor, Michael Silver, for making this all possible and making me feel welcome in the lab. I would also specifically like to thank Carissa Alforque, Reem Almagati, and Ahmad Ahmadzada for help with data collection/analysis for the experiments in this dissertation. Also all of the members of the Silver Lab for the helpful discussions at lab meetings and beyond. Finally, I would like to thank the members of my thesis and qualifying exam committees for helping shape this work with their suggestions and comments.

Chapter 1

Introduction

In the 1997 movie, *Good Will Hunting*, Matt Damon's character reads an entire page of a book at the same rate it takes a normal person to read one sentence. The movie portrays him as wicked smart, but there is no way Matt Damon's perfectly-proportioned head is big enough to fit the number of neural cells it would require to achieve such a task. The visual field is not sampled uniformly; only in our most central vision do we have enough densely-packed cells to reliably resolve cluttered information like text.

It is hard to fault the filmmakers for conflating physical impossibility with intelligence. The detriments experienced in peripheral vision are largely different than those that affect central vision, such as defocus blur. For example, it is relatively easy to understand the effects of corrective lenses, but most people in their day-to-day lives do not notice the impoverished nature of their peripheral vision.

As a graduate student, I would often get asked what I study, and I would normally respond with "vision science," until one day someone innocently asked, "you mean like, seeing visions?" I admittedly was a little embarrassed at first, but the more I reflect on it, the more I think the normal visual experience is hallucinatory in a sense. The visual system infers sensory information all the time. It fills in gaps. This is likely the reason the subjective experiences of peripheral and central vision blend well together.

Perhaps the perceptual detriments of peripheral vision are generally not that important for most day-to-day activities. Is it beneficial to see trees over the forest? Grass blades over the meadow? Apples over the orchard? Peripheral vision is very limited in identifying individual features and objects, but it excels at quickly obtaining a summary-statistical representation of regions of the visual scene. So yes, seeing the entire visual field in as high of a resolution as central vision would be nice for activities such as editing the rest of these sixty plus pages, but not everyone is a custodian at Harvard. I will stick to moving my eyes.

In the remainder of this dissertation, I describe three studies that all investigate the effects of spatial attention on the processing of cluttered peripheral stimuli. I compare and contrast two well-characterized visual phenomena: visual crowding and ensemble perception. Visual crowding is defined as the detrimental effect of clutter on object recognition in peripheral vision. Ensemble perception is defined as the ability to quickly extract the gist from a set of similar objects. These

two phenomena have been described as two sides of the same coin; even if it is difficult to identify individual objects in clutter in the periphery due to crowding, the visual system can extract a summary representation of the information due to ensemble perception.

In Chapter 2, we used an anti-cueing paradigm to test the effects of involuntary and voluntary attention on visual crowding. In Chapter 3, we expanded on the first study by investigating the effect of precision of a peripheral cue on visual crowding and ensemble perception. Finally in Chapter 4, we expanded on the analysis of the second study by fitting response models of different complexity to the psychophysical data to investigate how each component of the stimuli informed observers' responses.

Chapter 2

Effects of involuntary and voluntary attention on critical spacing of visual crowding

2.1 Introduction

Processing visual images, reading text, and navigating through the world all require a visual system that continuously parses cluttered scenes. Spatial attention is one mechanism for selecting regions of the visual scene for preferential processing (Carrasco 2011; Anton-Erxleben and Carrasco 2013). Covert spatial attention (i.e., directing attention to a location without accompanying eye movements) can either be voluntarily allocated (endogenous) or involuntarily captured by an external stimulus (exogenous) (Posner, Cohen, and R. D. Rafal 1982; Yantis and Jonides 1990). Effects of involuntary attention occur rapidly after stimulus onset but also dissipate quickly (Posner, Cohen, and R. D. Rafal 1982), while the onset of voluntary attention effects is slower (Posner, Cohen, and R. D. Rafal 1982), but its effects can persist for much longer (Silver, Ress, and Heeger 2007). These two types of attention also have different effects on perceptual factors such as contrast sensitivity (Barbot, Landy, and Carrasco 2012; Jigo and Carrasco 2020), sensory tuning (Fernández, Okun, and Carrasco 2021), and texture discrimination (Yeshurun and Carrasco 1998; Yeshurun, Montagna, and Carrasco 2008; Barbot and Carrasco 2017; Jigo, Heeger, and Carrasco 2021).

The effect of spatial attention on perceptual performance is perhaps most evident in peripheral vision, where perception is limited by a number of factors. Visual crowding, the reduction in the ability to identify target objects in the periphery in the presence of similar flanking objects, is one of the strongest of these limitations (Levi 2008; Whitney and Levi 2011). Previous research on visual crowding has shown that capture of involuntary attention with a peripheral cue improves performance (Felisberti, Solomon, and Morgan 2005; Scolari, Kohnen, et al. 2007) and decreases the critical spacing of crowding (i.e., the minimum target/flanker spacing at which the target is correctly identified at a specified level of performance) (Yeshurun and Rashal 2010; Rashal and

Yeshurun 2014). While these studies have characterized effects of involuntary attention on critical spacing in visual crowding, less is known about the effects of voluntary attention and how the effects of these two types of attention might be related.

In studies of visual crowding, peripheral targets are typically easy to identify in the absence of flanking stimuli, and this differentiates perceptual limitations due to crowding from those based on visual acuity. Crowding has been modeled as arising from inherent limits in the size and density of cortical receptive fields (RFs) in the visual periphery, especially when compared to central vision (Parkes et al. 2001; Balas, Nakano, and Ruth Rosenholtz 2009; Greenwood, Bex, and Dakin 2010; Dakin et al. 2010; Freeman and Simoncelli 2011; Rosenholtz 2016). One mechanism by which visual spatial attention might relieve crowding is by locally increasing the density of RFs that sample the target location (Baruch and Yeshurun 2014; Theiss, Bowen, and Silver 2021). Neurophysiologically, it has been shown that sustained visual spatial attention causes RFs to shift toward the locus of attention and shrink in size (Klein, Harvey, and Dumoulin 2014; Womelsdorf et al. 2006). At the behavioral level, spatial attention can improve texture discrimination performance by increasing the spatial resolution of perception for both involuntary (Yeshurun and Carrasco 1998) and voluntary (Yeshurun, Montagna, and Carrasco 2008) attention. However, compared to voluntary attention, in which changes in perceptual spatial resolution are flexible and based on task demands (Yeshurun, Montagna, and Carrasco 2008; Barbot and Carrasco 2017), involuntary attention always increases the spatial resolution of perception, even when this hinders texture discrimination performance (Yeshurun and Carrasco 1998; Yeshurun and Carrasco 2008). It is currently unknown whether the flexibility in the control of perceptual spatial resolution that has previously been attributed to voluntary attention could also account for attentional effects on critical spacing in visual crowding.

In most studies of spatial attention, involuntary attention is captured by a peripheral cue at the stimulus location, with a very short duration between cue and stimulus onsets ($\sim$ 40-100ms) (Posner 1980). Voluntary attention, on the other hand, is typically directed using a cue that is not at the location to be attended but instead specifies this location through symbolic or abstract information. For example, a central cue that points towards a location at which an upcoming target is likely to appear will enhance target processing at that location. Studies of voluntary attention typically employ a relatively long duration between cue and stimulus onsets ($>$ 300 ms) (Posner, Cohen, and R. D. Rafal 1982) .

In this study, we used an anti-cueing paradigm (Posner, Cohen, and R. D. Rafal 1982; Rokem et al. 2010) and a crowding task to measure the effects of both involuntary and voluntary attention on critical spacing while controlling for a number of experimental factors. Specifically, varying only the duration of the stimulus onset asynchrony (SOA) between presentation of a peripheral cue and a target stimulus allowed us to select whether involuntary or voluntary attention was primarily engaged. We measured the effects of involuntary and voluntary attention over a range of target/flanker spacings to determine the effects of both types of attention on critical spacing in visual crowding. We show that when involuntary attention was directed to the target location, the critical spacing of crowding decreased compared to when involuntary attention was directed elsewhere. However, when voluntary attention was directed to the target location, there was no significant effect on critical spacing. Additionally, we found that the effects of involuntary and

voluntary attention on critical spacing were not strongly correlated across subjects.

2.2   Materials and Methods

Subjects

The UC Berkeley Committee for Protection of Human Subjects approved all experimental pro-
cedures. Twenty-four subjects (sixteen females, eight males; age 20–52 years) with normal or
corrected-to-normal vision participated in this study. Of these subjects, seventeen were naive to
the experimental design, and seven (including all three authors) had at least some knowledge of
the design. Subjects were compensated for their time.

Stimuli and Apparatus

We presented stimuli using the Psychophysics Toolbox (D. H. Brainard 1997; D. G. Pelli 1997;
Kleiner, D. Brainard, and D. Pelli 2007; Cornelissen, Peters, and Palmer 2002) on a 53 cm Dell
UltraSharp LCD monitor with a 1680 x 1050 resolution, 60 Hz refresh rate, and 300 cd/m^2 peak
brightness. We recorded eye position with the Eyelink 1000 (SR Research Ltd., Ottawa, Canada).
Subjects sat at a distance of 50 cm from the screen with their heads on a chinrest in a dark room.
Stimuli consisted of five equally-sized circular Gabor patches (100% contrast, 4 cycles; spatial
frequency varied for each participant (see below); random spatial phase) arranged in a vertical
configuration: one target in the middle and two flankers both above and below the target. The
target was tilted either 45°or 135°away from horizontal, and the flanker orientations were randomly
chosen to be between 0° and 180°. To avoid target pop-out and other forms of saliency based on
differences in orientation between the target and flankers, flanker orientations could not be within
$\pm$ 10°of horizontal (0°), vertical (90°), or either of the two possible target orientations (45°or 135°).
The target was presented at an eccentricity of 14 degrees of visual angle either to the left or right
of fixation.

Procedure

All subjects completed four experimental sessions, with an interval of at least 24 hours between
sessions. During the first session, subjects completed two baseline experiments. The first of these
experiments was used to derive a threshold size for the target in the absence of flankers for each
participant. To do this, we presented a single target (100% contrast, 4 cycles; random spatial
phase; eccentricity of 14 degrees of visual angle) on either the left or right side of the screen
(balanced across subjects) and used a 3-down/1-up staircase procedure to adjust the diameter of
the target in units of degrees of visual angle (133 ms stimulus presentation; 100 trials; 1.8 degrees
initial diameter; 0.1 degree staircase step size). Subjects performed a 2-alternative forced choice
(2AFC) task on the orientation of the target (45°or 135°) using a key press. We then fit a Weibull
cumulative distribution function (equation 2.1; for $s = 1$) to the data using a squared error cost

function weighted by the proportion of trials per target size in the staircase (equation 2.2), and the size of the target and flankers for all subsequent crowding experiments for a given subject was set at $1.5\times$ the diameter at 80% performance on the unflanked orientation discrimination task, as predicted by the fitted psychometric curve. We chose this value for the diameter of the target so that task performance in the subsequent experiments was limited by target/flanker interactions in crowding and not by target visibility. Note that this procedure of picking the size of the Gabor patches resulted in differences across participants in the spatial frequency of the target and flanker Gabor patches. The range of spatial frequencies was 2.27–4.59 cycles/degree.

The second baseline experiment was used to specify a range of target/flanker spacings for each subject. To do this, we presented a target with flankers on the same side of the screen as in the first baseline experiment described above. The size and spatial frequency of the target and flanker stimuli were based on the results of the first baseline experiment for each participant. We used a 3-down/1-up staircase procedure to adjust the center-to-center target/flanker spacing, measured in degrees of visual angle (133 ms stimulus presentation; 150 trials; 5° initial spacing; 0.2° staircase step size). We then fit another Weibull function (as described above) to the spacing data, and the set of target/flanker spacings for all subsequent crowding experiments for each participant was defined as seven evenly-spaced values from a lower limit (the spacing at 55% performance) and an upper limit ($1.5\times$ the spacing at 80% performance), as predicted by the fitted psychometric curve. We selected this range of target/flanker spacings for each subject to avoid floor and ceiling effects that could limit our ability to effectively measure the effects of attention on critical spacing.

For the remainder of the first session and all subsequent sessions, we employed an anti-cueing task (Posner, Cohen, and R. D. Rafal 1982; Rokem et al. 2010) to separately measure the effects of pre-cueing involuntary and voluntary attention on critical spacing of visual crowding. After a 1200 ms fixation period at the start of each trial (Figure 2.1; left panel), one set of vertical bars (presented at 14 degrees of visual angle from fixation on either the left or right side of the screen) became thicker (changing from 0.05° to 0.15° visual angle) and brighter (changing from 25% to 75% maximal luminance) for 40 ms (Figure 2.1; middle panel). Next, the crowded array of Gabor patches was presented for 133 ms, 80% of the time within the vertical bars on the opposite side of the cue, and 20% of the time on the same side as the cue (Figure 2.1; right panel). Subjects performed a 2AFC task on the orientation of the target using a key press. They were instructed to respond as quickly and as accurately as possible, without moving their eyes from the central fixation cross. Subjects were also explicitly told that the stimulus was much more likely to appear on the opposite side than on the cued side. For a given block of trials, the stimulus onset asynchrony (SOA) for the cue and the crowded stimuli was either 40 or 600 ms. At the beginning of each session, subjects completed 32 practice trials (50% long SOAs and 50% short SOAs) with unflanked targets until they achieved 75% correct performance. Each subject then completed eight blocks of 120 trials each (960 trials per session; 3840 total trials for all four sessions). The SOA was fixed for a given block and was randomly ordered across blocks. The eight spacing conditions were randomly interleaved within a block and balanced across each combination of SOA and stimulus location.
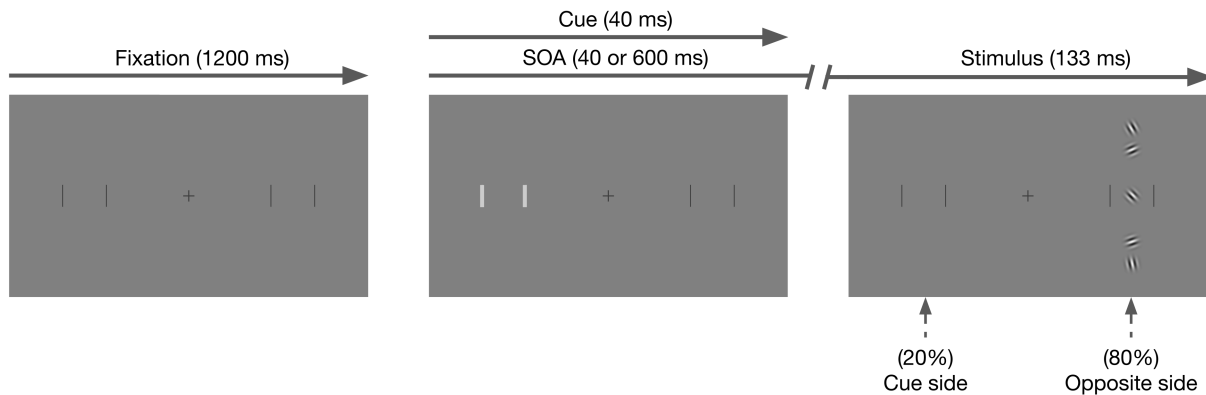
Figure 2.1: Schematic of the anti-cueing task. After a fixation interval, one set of vertical bars became thicker and brighter for 40 ms. After an SOA of 40 ms (short) or 600 ms (long), the crowded stimuli appeared for 133 ms within one of the two sets of vertical bars. On 20% of trials the stimulus appeared on the cued side, and on 80% of trials the stimulus appeared on the opposite side. Stimuli were composed of a central target Gabor patch ($45°$ or $135°$ orientation) and two sets of two flanking Gabor patches with independent random orientations. Target/flanker spacing was varied over a range of center-to-center distances, and the range of spacings and the size of the Gabor patches were customized for each subject (see Materials and Methods). There was also a condition in which the target was presented without flankers. Subjects performed a 2AFC task on the orientation of the target Gabor patch as quickly and as accurately as possible without moving their eyes from the central fixation cross. We recorded response time and accuracy. Gabor patch and cue sizes shown here were increased for visualization purposes and are not representative of actual experimental values.

Analysis

Critical Spacing (CS)   A parameterized Weibull function was fit to the accuracy (percent correct) data across the range of spacings for each combination of SOA (40 ms or 600 ms) and location (Cue or Opposite side). Critical spacing, t, was defined as the spacing at 78% accuracy. We chose this percent correct value because it was the approximate midpoint of the observed range of percent correct values in this study. Therefore, any effects of attention in critical spacing across conditions would be more likely to be reflected at multiple target/flanker spacing values in the psychometric function. The main effects reported in Figure 2.2 are not strictly dependent on this particular percent correct value, and they remain significant over a robust range of values ($75\% - 85\%$ correct).

The parameterized Weibull function was:

$$f_{t,b,s}(x) = s - (s - g)e^{-\left(\frac{kx}{t}\right)^b} \tag{2.1}$$

where $k = -\ln\left(\frac{s-a}{s-g}\right)^{\frac{1}{b}}$, g was chance performance (50%), a was the percent correct value used to define the critical spacing value (78%), and b and s were the slope and asymptote of the psychometric curve, respectively. Parameters b, $s \leq 100\%$ correct, and t were optimized using Matlab's fmincon for each subject's data. We minimized a squared error cost function that was weighted by the proportion of trials at each target/flanker spacing value and regularized by performance on the unflanked trials. Specifically, the optimization was defined as:

$$\min_{t,b,s \leq 1}\left[\left(\sum_i w_i\left(f(i) - y_i\right)^2\right) + w_{\text{unflanked}}\left(s - y_{\text{unflanked}}\right)^2\right] \tag{2.2}$$

where y_i and $y_{\text{unflanked}}$ were the measured percent correct values for each spacing, i, and the unflanked trials, respectively. Similarly, w_i and $w_{\text{unflanked}}$ were the proportion of trials for each spacing, i, and the unflanked trials, respectively.

Response Time (RT)   RT was collected for each trial over the range of target/flanker spacings for each subject. As with critical spacing, which is a summary statistic represented by all the percent correct data in the psychometric fit, we computed an RT summary statistic for making comparisons across conditions. To do this, a single line was fit using least squares regression to median RT values across the range of target/flanker spacings for each combination of SOA and location (Cue or Opposite side). Only RT values from correct trials were used in the fitting process. For each subject, comparisons across conditions were conducted at the predicted RT (from the fitted line) that corresponded to the critical spacing calculated from all SOA/location trials combined. We conducted comparisons across conditions at the target/flanker spacing derived from all combined trials rather than at each condition's critical spacing to avoid any speed/accuracy trade-offs that could be associated with differences in target/flanker spacings. We also conducted RT comparisons for the unflanked trials.

Statistical Analyses   Subjects were removed from analysis if their asymptotic level of performance (s) for at least one of their SOA/location conditions was three or more standard deviations lower than the mean of all conditions (12.5% subjects in total). Trials in which fixation deviated by more than a distance of $3°$ from the fixation cross during target/flanker presentation were also removed from analysis (2.8% of trials). Mean RT and critical spacing were analyzed with a repeated-measures ANOVA with SOA (40 or 600 ms) and stimulus location (Cue or Opposite side) entered as within-subject factors. We additionally conducted a number of planned comparisons to assess the effects of the cue on RT and critical spacing. Specifically, for each SOA we defined the cueing effects as a pairwise difference between values when the stimulus appeared on the cue side and values when the stimulus appeared on the opposite side. We used two-tailed Student's t-tests to assess if the means of the cueing effects were significantly different than zero. For the correlation analyses, Pearson's r values were calculated and tested against the null hypothesis of a correlation coefficient value of zero.

2.3   Results

Participants performed an orientation discrimination task in which presentation of the cue on one side of the screen predicted that the stimulus would appear on the opposite side of the screen 80% of the time and on the same side of the screen 20% of the time (Figure 2.1). For all trials, the cue directed a participant's involuntary attention to its location. With additional time between cue and stimulus presentation, however, the participant could voluntarily direct their attention to the opposite side of the screen (where the stimulus was more likely to appear). We varied the SOA between cue and stimulus presentations to study the effects of involuntary (40 ms SOA) and voluntary attention (600 ms SOA) on crowding. Specifically, we compared critical spacing and response time for each combination of SOA (40 ms or 600 ms) and location (Cue or Opposite side). For each metric, we also computed the magnitude of the cueing effect by calculating within-subject differences between values when the stimulus appeared on the same side as the cue ("Cue" in Figure 2.2) and values when the stimulus appeared on the opposite side of the cue ("Opp." in Figure 2.2).

For half of the blocks, the SOA was 40 ms to maximize involuntary spatial attention directed to the cue location while not allowing enough time for allocation of voluntary attention (Posner, Cohen, and R. D. Rafal 1982; Rokem et al. 2010). For these trials (Figure 2.2; top left), mean RT was faster when the stimulus appeared on the cue side (494 ms) compared to the opposite side (506 ms), indicating that involuntary attention was allocated to the cue side. The magnitude of this cueing effect (cue RT - opposite RT) was -12 ms (Figure 2.2; bottom left) and was significantly less than zero ($t_{20} = -2.66$, $p = 0.015$). The mean r^2 values for the lines fit to RT over the range of target/flanker spacings (see Materials and Methods) were fairly low (0.75 and 0.47 for the opposite and cue side, respectively). To confirm that this cueing effect for involuntary attention was not just a consequence of the quality of the linear fit, we computed the effect of cueing on RT for correct trials in the unflanked condition. This mean RT effect was also significantly less than zero (-17 ms; $t_{20} = -3.77$, $p = 0.001$), providing further evidence that the cue was effective in capturing involuntary attention.

For the 40 ms SOA, mean critical spacing was smaller when the stimulus appeared on the cue side (3.03 degrees) compared to the opposite side (3.18) (Figure 2.2; top right). The magnitude of this cueing effect was -0.15 degrees (Figure 2.2; bottom right) and was significantly less than zero ($t_{20} = -2.93$, $p = 0.008$). Overall, the Weibull function fit the accuracy (percent correct) data well for the short SOA trials: mean r^2 values were 0.93 and 0.87 for the opposite and cue sides, respectively. Taken together, these results demonstrate that involuntary attention leads to both faster RT and smaller critical spacing in visual crowding.

For the other half of the blocks, the SOA was 600 ms, leaving sufficient time for subjects to overcome the initial involuntary capture of attention by the cue and to then maximize allocation of voluntary attention to the opposite side, where the stimulus most often appeared (Posner, Cohen, and R. D. Rafal 1982; Rokem et al. 2010). For these trials (Figure 2.2; top left), mean RT was faster when the stimulus appeared on the opposite side (489 ms) compared to the cue side (520 ms), indicating that voluntary attention was successfully allocated to the higher-probability opposite side. The magnitude of this cueing effect was 31 ms (Figure 2.2; bottom left) and was significantly
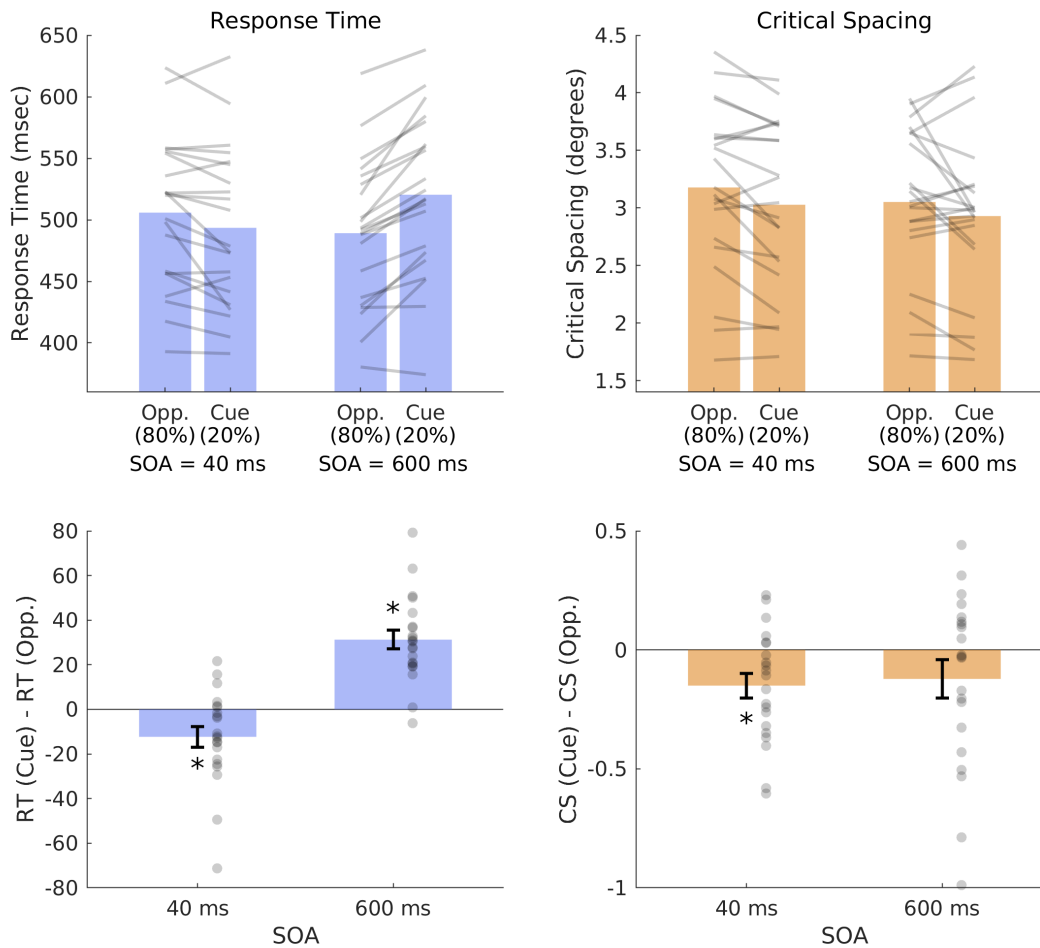
Figure 2.2: (*Top*) The effect of stimulus location relative to cue location (opposite or cue side) and SOA (40 or 600 ms) on response time (left; blue) and critical spacing (right; orange) for the crowding task. Gray lines represent matched individual subject data across location conditions. (*Bottom*) Mean within-subject cueing effects for each metric, defined as the difference between values when the stimulus appeared on the same side as the cue (Cue) and values when the stimulus appeared on the opposite side of the cue (Opp.). Gray dots represent individual subjects, and asterisks indicate significance level $\alpha < 0.05$ from a planned comparison paired t-test. Error bars are standard errors of the mean.

greater than zero ($t_{20} = 7.44$, $p = 3.53\mathrm{e} - 7$). As for the 40 ms SOA, the r^2 values for the lines fit to the RTs for the 600 ms SOA over the range of target/flanker spacings were low: 0.82 and 0.52 for the opposite and cue sides, respectively. However, once again, correct trials from the

unflanked condition confirmed a large cueing effect that was significantly greater than zero (43 ms; $t_{20} = 7.87$, $p = 1.49\text{e} - 7$).

Interestingly, and in contrast to RT, mean critical spacing was greater when the stimulus appeared on the opposite side (3.05 degrees) compared to the cue side (2.93 degrees) for the long SOA (Figure 2.2; top right). However, the magnitude of the cueing effect on critical spacing (−0.12 degrees) (Figure 2.2; bottom right), was not significantly different from zero ($t_{20} = -1.53$, $p = 0.14$). Overall, the Weibull function fit the accuracy (percent correct) data well for the long SOA trials: mean r^2 values were 0.96 and 0.87 for the opposite and cue sides, respectively. Taken together, these results indicate that allocation of voluntary attention to the higher-probability opposite side leads to much faster RT but no significant change in critical spacing. A repeated-measures ANOVA revealed that the interaction of SOA and location was significant for RT ($F_{1,20} = 37.8$, $p = 5.34\text{e} - 6$) but was not significant for critical spacing ($F_{1,20} = 0.13$, $p = 0.72$). These results indicate that long and short SOAs produced significantly different patterns for RT but not for critical spacing.
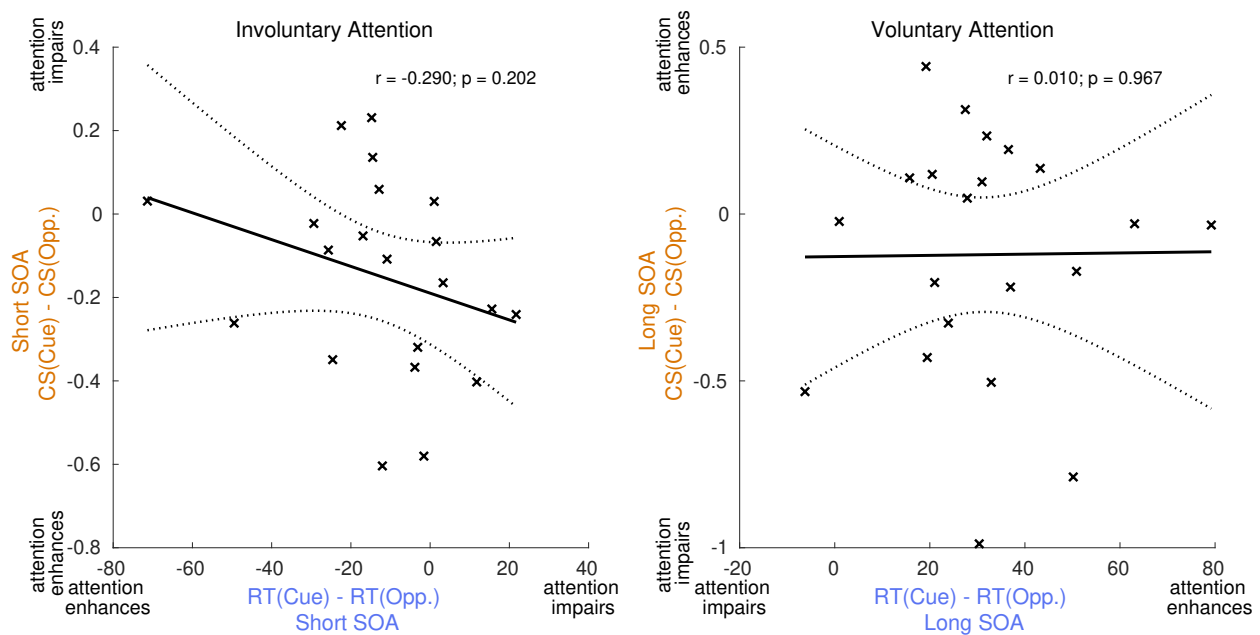
Figure 2.3: Cueing effects on RT were not strongly correlated with cueing effects on critical spacing for either short SOA (*left*) or long SOA (*right*) trials. The 'x's represent individual subjects. The solid and dashed lines represent the linear regression fits and 95% confidence intervals, respectively. Pearson's r and p values for the correlations are displayed in the upper right corner of each plot. Attention "enhances" or "impairs" labels correspond to the direction of the cueing effect for each of these metrics. Specifically, a negative cueing effect (Cue < Opp.) for involuntary attention and a positive cueing effect (Cue > Opp.) for voluntary attention are associated with enhanced processing due to attention (i.e., faster RT/smaller critical spacing).

One possible explanation for the non-significant negative cueing effect for critical spacing but a significant positive cueing effect for RT (Figure 2.2) for the long SOA trials is a trade-off between attentional effects on RT and on critical spacing. For example, it is possible that a subject who responded much faster when voluntary attention was directed to the target compared to when it was directed elsewhere (attentional enhancement of RT) consequently had no change (or even an impairment) in critical spacing with attention. To test this relationship, we correlated the magnitude of individual subjects' RT cueing effect with the magnitude of their critical spacing cueing effect for each SOA (Figure 2.3). We found no significant correlation between these two cueing effects for either short SOA (Figure 2.3; left; $r = -0.290$; $p = 0.202$) or long SOA (Figure 2.3; right; $r = 0.010$; $p = 0.967$) trials. The lack of strong correlations between cueing effects for RT and for critical spacing for both short and long SOAs indicates that subjects in this study likely did not have significant trade-offs between attentional effects on RT and on critical spacing.
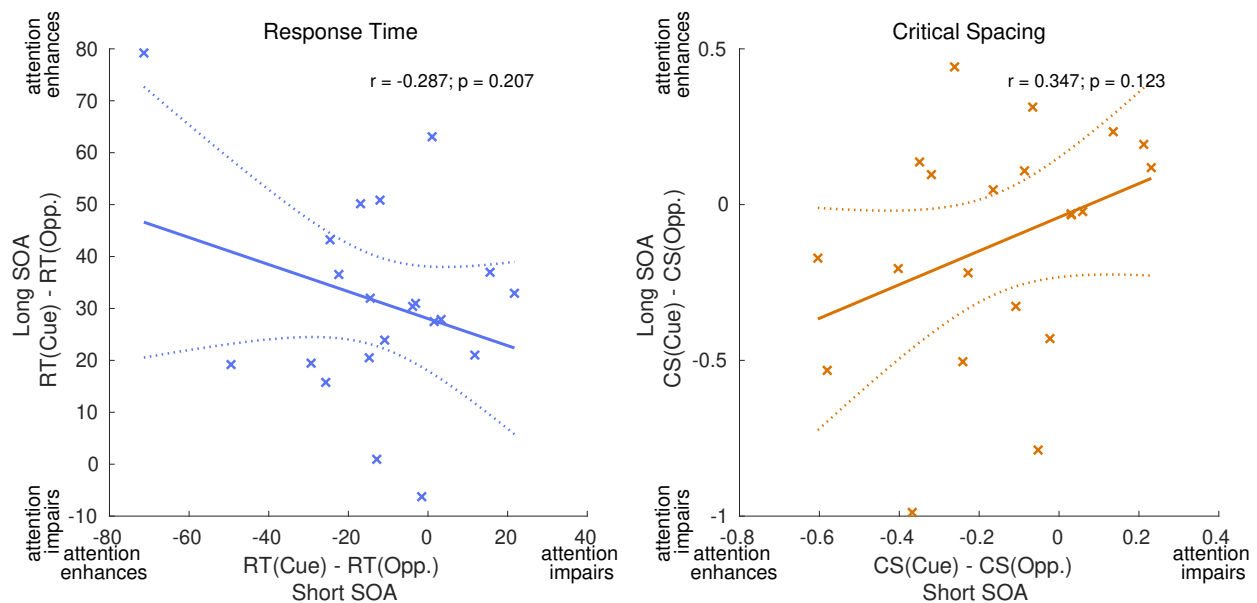
Figure 2.4: Cueing effects for involuntary (short SOA trials) and voluntary attention (long SOA trials) were not significantly correlated across individual subjects for both RT (*left*; blue) and critical spacing (*right*; orange). The 'x's represent individual subjects. The solid and dashed lines represent the linear regression fits and 95% confidence intervals, respectively. Pearson's r and p values for the correlations are displayed in the upper right corner of each plot. Attention "enhances" or "impairs" labels correspond to the direction of the cueing effect for each of these metrics. Specifically, a negative cueing effect (Cue $<$ Opp.) for involuntary attention and a positive cueing effect (Cue $>$ Opp.) for voluntary attention are associated with enhanced processing due to attention (i.e., faster RT/smaller critical spacing).

Finally, we tested whether the magnitudes of the effects of involuntary and voluntary attention were correlated across subjects. Some studies have described how competition between involun-

tary and voluntary attention can affect perceptual processing (Berger, Henik, and R. Rafal 2005; Fukuda and Vogel 2009). Perhaps a subject who more successfully overcame the initial involuntary capture of attention by the cue (short SOA trials) was also able to more effectively direct their attention to the opposite side (long SOA trials), where the stimulus most often appeared. To test this relationship, we correlated the magnitude of cueing effects for short and long SOA trials across subjects for both RT and critical spacing. We found no significant correlation between these two cueing effects for either RT (Figure 2.4; left; $r = -0.287$, $p = 0.207$) or critical spacing (Figure 2.4; right; $r = 0.347$, $p = 0.123$). These results indicate that the magnitude of an individual's cueing effect for one spatial attention mechanism (voluntary or involuntary) does not reliably predict the other.

2.4 Discussion

We used an anti-cueing paradigm to measure the effects of involuntary and voluntary spatial attention on a visual crowding task. In our study, all stimulus and task factors were identical for the involuntary and voluntary conditions except for the SOA between the cue and stimulus presentations. Additionally, by customizing the range of target/flanker spacings for each subject, we more effectively avoided floor and ceiling effects on performance compared to other similar studies of attention and crowding. Our data set includes a large number of participants (21 that were included in the analyses presented here), within-subject statistical comparisons to assess the effects of involuntary and voluntary attention and the correlations between them, and a substantial amount of data per subject (3840 trials of the anti-cueing task). Our study therefore has high sensitivity for detecting possible cueing effects on RT and critical spacing.

For both involuntary and voluntary attention, we found reductions in RT for target orientation discrimination when spatial attention was directed to the target location compared to when attention was directed elsewhere. We also showed that directing involuntary attention to the target location with a peripheral cue decreased critical spacing compared to when attention was directed elsewhere. Interestingly, we did not find any significant difference in critical spacing when voluntary attention was directed to the target location compared to when it was directed elsewhere, and we showed that this lack of a voluntary attention effect could not be explained by trade-offs between attentional effects on RT and on critical spacing. Overall, short and long SOAs did not produce significantly different patterns of critical spacing, even though our RT results provide evidence that the cues effectively engaged involuntary and voluntary attention. Finally, we showed that, for both RT and critical spacing, involuntary attention cueing effects were not strongly correlated with voluntary attention cueing effects across participants.

Effects of spatial attention on crowding and critical spacing. There is much debate about how attention affects visual crowding in general and critical spacing in particular. Most studies of the effects of involuntary attention have found enhanced perceptual performance on crowding tasks, and some of these also reported significant effects of involuntary attention on critical spacing (Yeshurun and Rashal 2010; Rashal and Yeshurun 2014), while others did not (Felisberti, Solomon, and

Morgan 2005; Scolari, Kohnen, et al. 2007). Yeshurun and Rashal 2010 hypothesized that they observed a significant decrease in critical spacing with involuntary attention while other investigators did not because their peripheral cue did not act as a forward mask on processing of the subsequently-presented target. For the short SOA trials in our study, we also found a significant decrease in critical spacing when involuntary attention was directed to the target location compared to when it was directed elsewhere, and our cue (vertical lines on either side of the possible target locations) also avoided forward masking by cueing locations that were adjacent to the target instead of the target location itself.

Reports in the literature on possible effects of voluntary attention on critical spacing in crowding are more limited. Albonico et al. 2018 used a long SOA (400 ms) to test the effects of three different cue types (dot, small box, and large box) on critical spacing. They found that only the dot cue significantly decreased critical spacing. However, the goal of their study was to distinguish between orienting and focusing of attention, so they always used a peripheral cue to direct attention to the crowded stimulus. The lack of a symbolic cue makes it difficult to separate the effects of voluntary from involuntary attention in their study. Perhaps the large box cue employed by Albonico et al. 2018, which encompasses both target and flanker locations, is most conceptually similar to our anti-cueing approach to isolating involuntary and voluntary attention effects. While we used differences in cue/target SOA to accomplish this, the large box cue in Albonico et al. 2018 spatially separates the region cued by involuntary attention from the region cued by voluntary attention. The lack of a significant cueing effect on critical spacing for the large cue in their study is consistent with what we found for our long SOA trials.

Although studies of the effects of voluntary attention on critical spacing have been limited and inconclusive, this type of attention has been shown to modulate other aspects of crowding. Instead of using a cueing design, Mareschal, Morgan, and Solomon 2010 instructed participants to attend to different aspects of a crowded stimulus. Specifically, they measured how attending to a radial target/flanker configuration differed from attending to a tangential target/flanker configuration when both were presented at the same time in a cross-shaped stimulus array. In general, flankers along a radial axis crowd more strongly than tangential flankers (Toet and Levi 1992), and Mareschal, Morgan, and Solomon 2010 found a similar radial/tangential asymmetry that was based only on allocation of voluntary attention.

Some studies have examined the effects of attention on brain responses to crowded stimuli (Fang and S. He 2008; Chen et al. 2014). Using fMRI, Chen et al. 2014 found that the magnitude of the behavioral crowding effect was closely linked with a suppressive cortical interaction in V1. Specifically, the peak amplitudes of the fMRI signal were greater for large target/flanker spacings compared to small spacings, indicating that stronger crowding was associated with greater physiological suppression of visual responses. Furthermore, this suppression effect was more prominent when the stimuli were attended versus when they were passively viewed.

Distinct effects of involuntary and voluntary attention on the spatial resolution of perception. Because performance on crowding tasks is dependent on both target/flanker spacing and on the eccentricity of the target (Bouma 1970; Whitney and Levi 2011), it has been thought that

it is limited by the spatial resolution of the visual system (Parkes et al. 2001; Balas, Nakano, and Ruth Rosenholtz 2009; Greenwood, Bex, and Dakin 2010; Dakin et al. 2010; Freeman and Simoncelli 2011; Rosenholtz 2016). Attention has been shown to influence the spatial resolution of texture discrimination, with involuntary and voluntary attention showing distinct effects (Yeshurun and Carrasco 1998; Yeshurun and Carrasco 2008; Yeshurun, Montagna, and Carrasco 2008; Barbot and Carrasco 2017; Jigo, Heeger, and Carrasco 2021). Without spatial attention cues, texture discrimination performance varies as an inverted U-shape function of eccentricity, with maximal performance at mid-peripheral locations. Involuntary attention causes peak texture discrimination performance to shift towards more peripheral eccentricities for all cue locations, thereby decreasing performance at more central locations compared to a neutral attention cue (Yeshurun and Carrasco 1998; Yeshurun and Carrasco 2008). However, voluntary attention enhances performance across all eccentricities, including both peripheral and central locations (Yeshurun, Montagna, and Carrasco 2008). From these results, it has been hypothesized that involuntary attention always increases perceptual spatial resolution even if it hinders performance, while voluntary attention is more capable of flexibly adapting spatial resolution to match the demands of the task (Barbot and Carrasco 2017).

Physiologically, the effects of spatial attention on the resolution of stimulus representations are related to neuronal RF sizes. Directing spatial attention to one of multiple objects within a single RF biases responses in favor of the attended object (Desimone and Duncan 1995). At the single-cell level, these attentional effects have been observed as both a scaling of neuronal responses to an attended stimulus by a gain factor (McAdams and Maunsell 1999) and a shrinking of neuronal RFs around an attended stimulus (Anton-Erxleben, Stephan, and Treue 2009). Furthermore, RFs in humans and other animals have been observed to both shrink in size and shift toward the locus of attention with sustained spatial attention (Womelsdorf et al. 2006; Klein, Harvey, and Dumoulin 2014). Using computational modeling approaches, Baruch and Yeshurun 2014 showed that this reconfiguration of RFs with attention could explain a number of attentional effects on neural responses, and Theiss, Bowen, and Silver 2021 showed that a similar mechanism could be implemented in a convolutional neural network, resulting in attentional enhancement of performance on a visual crowding task. Additionally, D. He, Wang, and Fang 2019 showed that following perceptual learning of a crowded orientation discrimination task, decreases in RF size of individual fMRI voxels in cortical area V2 correlated with improved performance resulting from perceptual learning. For a separate group of subjects, training on the orientation discrimination task in the absence of flankers also resulted in improvement in performance, but this perceptual improvement with learning was not correlated with changes in RF size as measured with fMRI.

Given the substantial behavioral, physiological, and computational evidence supporting the beneficial effects of increased spatial resolution by sustained attention, it is surprising we saw no significant change in critical spacing when voluntary attention was directed to the target location compared to when it was directed elsewhere. However, we note that we did not include a baseline or neutral cue condition in our study, so fully differentiating possible beneficial and detrimental effects of attention on performance was not possible. Moreover, one difference between our anti-cueing study and the texture discrimination tasks described above is that the voluntary attention cue in our crowding task is less informative than the cue in the texture discrimination tasks, given

that we employed a interleaved range of target/flanker spacings. More specifically, both types of tasks contain a cue that provides information about target location, but in our crowding task, there is also task-irrelevant trial-to-trial uncertainty about the spacing between the target and its flankers. One direction for future work is to investigate how cues that provide information about target/flanker spacing or similarity (Scolari, Byers, and Serences 2012) impact the ability to find optimal solutions for crowding tasks.

Another possibility is that the critical spacing of visual crowding is not sensitive to small changes in the spatial resolution of perception or specifically neuronal RF size. The neurotransmitter acetylcholine (ACh) reduces excitatory RF size in marmoset primary visual cortex (Roberts et al. 2005), and enhancement of cholinergic signaling with the cholinesterase inhibitor donepezil decreases the spatial spread of the fMRI response to visual stimuli in human visual cortex (Silver, Shenhav, and D'Esposito 2008), a result that is consistent with smaller excitatory RFs. Perceptually, donepezil reduces visual surround suppression in humans (Kosovicheva et al. 2012) and sharpens visual spatial perception in a contrast decrement detection task in the presence of flankers (Gratton et al. 2017). In contrast to these reports of improved spatial resolution of visual perception in humans following cholinergic enhancement, donepezil was found to have no significant effect on critical spacing of visual crowding in a letter identification task (Kosovicheva et al. 2012). These pharmacological results suggest a possible distinction between critical spacing in crowding and other perceptual and neurophysiological measures of the spatial resolution of stimulus representations.

Finally, it is somewhat surprising we observed no significant correlation between voluntary and involuntary cueing effects across participants. Other cognitive factors such as memory capacity have been shown to be positively correlated with the ability to resist attentional capture from salient involuntary cues (Fukuda and Vogel 2009). This result and those from other similar studies (Berger, Henik, and R. Rafal 2005) suggest that subjects with stronger voluntary attention may be better able to overcome/ignore capture by involuntary attention. However, we did not observe a significant relationship between a subject's ability to use the peripheral cue to direct their attention to the more likely opposite side (long SOA trials) and their ability to overcome capture by involuntary attention (short SOA trials). It could be informative to conduct a similar correlation study for other tasks that show strong effects of attention on perception, such as contrast sensitivity or texture discrimination. Furthermore, it would be interesting to not only correlate performance or RT for the two attention types but also response amplitudes (Dugué et al. 2020) and RF sizes, as measured physiologically.

Chapter 3

Effects of precision of a peripheral cue on visual crowding and ensemble perception

3.1 Introduction

In peripheral vision, it is difficult to identify individual objects within clutter, a phenomenon known as visual crowding (Bouma 1970; Levi 2008; Whitney and Levi 2011). However, the actual impact that these perceptual limitations have on day-to-day life can be starkly different from what is measured in the lab (Ruth Rosenholtz 2020). For example, the ability to see an individual's face in detail at a glimpse may not be that useful for quickly navigating through a crowd. While peripheral vision is very limited in identifying individual features and objects, it excels at quickly obtaining a summary-statistical representation of regions of the visual scene. This phenomenon of extracting the gist of a scene is known as ensemble perception (Alvarez 2011; Haberman and Whitney 2012; Whitney and Yamanashi Leib 2018). The visual system additionally uses spatial attention to select certain regions for preferentially processing (Carrasco 2011). Normally, spatial attention that is directed to a location in the visual field increases the ability to quickly and accurately react to all types of stimuli at that location (Posner 1980; Posner, Cohen, and R. D. Rafal 1982). However, this enhanced perception at attended locations is accompanied by impaired perception at unattended locations (Smith, Singh, and Greenlee 2000).

Attention that is involuntarily captured by salient visual cues (exogenous attention) has been shown to increase the spatial resolution of perception in peripheral texture discrimination tasks (Yeshurun and Carrasco 1998; Barbot and Carrasco 2017) and to improve performance in visual crowding tasks (Yeshurun and Rashal 2010; Scolari, Kohnen, et al. 2007). However, there still remains debate over the effect of cue size/type on crowding. Here, we focus on the precision of an attention cue, which we define as the size of the physical region the cue encompasses. In texture segmentation tasks, Yeshurun and Carrasco 2008 showed that a cue had to be smaller than or equal to the size of the region that encompassed the target texture patch in order to increase the resolution of perception. In visual crowding of letter stimuli, Albonico et al. 2018 showed that the crowding effect was modulated most by a small dot cue and not by cues that encompassed the space beyond

the target location. In contrast, Strasburger and Malania 2013 found that the benefits of attention on crowding were independent of cue size. As we will outline below, crowding is largely affected by the relationship between target and flanker features, so we explore further if modulations via an exogenous cue depend on the features of the surrounding flankers and/or their relationship to the target features.

In crowding tasks, there are a number of factors that influence participants' perceptual reports, especially those that are erroneous. For example, in a target identification task, participants are much more likely to incorrectly report the identity of one of the present flankers than that of a flanker that was not present (Ester, Klee, and Awh 2014; Hanus and Vul 2013; Coates, Bernard, and Chung 2019). There are also categorical target/flanker effects (Reuther and Chakravarthi 2014), global/contextual effects (Manassi, Sayim, and Herzog 2012; Herzog et al. 2015), and holistic effects (Farzin, Rivera, and Whitney 2009) that can all influence perception of crowded stimuli. In visual crowding, there is evidence of averaging (Parkes et al. 2001), spatial weighting (Greenwood, Bex, and Dakin 2010), and pooling of feature representations into summary-statistical representations (Balas, Nakano, and Ruth Rosenholtz 2009; Freeman and Simoncelli 2011; Rosenholtz 2016). Pooling of information is more likely to occur in the periphery than in central vision due to the larger integration regions of peripheral receptive fields (RFs) (Gattass, Gross, and Sandell 1981; Gattass, Sousa, and Gross 1988). Sustained spatial attention has been shown to affect the size and position of visual cortical neuronal RFs (Womelsdorf et al. 2006; Klein, Harvey, and Dumoulin 2014). Therefore, RF modulation may be one mechanism by which attention affects crowding performance (Baruch and Yeshurun 2014; Theiss, Bowen, and Silver 2021; D. He, Wang, and Fang 2019).

While it is difficult and time consuming to segment and identify individual objects in clutter, it is relatively easy to calculate summary statistics from a group of similar objects (Alvarez 2011; Haberman and Whitney 2012; Whitney and Yamanashi Leib 2018). Specifically, observers can extract the mean and variance of a number of features, such as size, orientation, color, motion, facial expression, and emotion of an ensemble (Alvarez 2011), and size and modality of a distribution of features can be extracted in both central (Chetverikov, Campana, and Kristjánsson 2016; Chetverikov, Campana, and Kristjansson 2017) and peripheral (Tanrıkulu, Chetverikov, and Kristjánsson 2020) vision. However, the role of attention in ensemble perception is not as well understood. Recent studies have shown that the reported mean orientation of an ensemble can be biased toward more salient objects (Iakovlev and Utochkin 2021), and attention directed to one object in a ensemble can bias the reported mean shape/size (De Fockert and Marchant 2008) and mean facial expression (Ying 2022) toward that of the attended object. However, there are also studies that have shown minimal effects of attention on ensemble perception (Ji, Rossi, and Pourtois 2018; Talipski, Goodhew, and Edwards 2021).

In a crowded target identification task, it may be more beneficial to focus attention on the target location. Alternatively, in an ensemble perception task, it may be more beneficial to distribute attention over the entire stimulus array. However, if target and flanker features are inherently combined into a summary-statistic representation in peripheral vision, will observers utilize similar strategies when asked to report the identity of one of the elements versus a summary statistic of the entire ensemble. How does performance change when attention is directed to one of the elements

versus the entire stimulus array? In the current study, we aim to address these open questions to gain a better understanding of the interactions between exogenous spatial attention and the processing of cluttered ensemble information.

In Experiment 1, we employed a spatial attention cueing paradigm (Posner 1980; Posner, Cohen, and R. D. Rafal 1982) to compare the effects of different sizes of a peripheral cue on performance of an orientation discrimination crowding task (Target task; session 1) versus performance of an ensemble mean orientation discrimination task (Ensemble task; session 2). The stimuli consisted of a central Gabor surrounded by a ring of uniformly-spaced Gabors. The 50% valid peripheral cue could either encompass the spatial extent of just the central Gabor (Small) or the extent of the entire stimulus array (Large). For both tasks, we varied the relationship between the central Gabor's orientation and the mean orientation of the entire stimulus array to see how the size of the cues interacted with different patterns of ensemble statistics relative to a single cued object within the ensemble.

In Experiment 2, a new set of subjects performed an ensemble mean orientation discrimination task on a 4x4 grid of uniformly-spaced Gabors to better control for the saliency of the cued Gabor relative to the other Gabors. The spatial cueing procedure was the same as Experiment 1, but now the cue and corresponding cued Gabor could appear at any one of the four locations within the center of the grid.

Overall, we found that the small cue, and not the large cue, reduced crowding in the Target task. For the same observers and stimuli, cue size also moderately modulated ensemble perception. However, by controlling for the saliency of individual Gabors with the uniform grid stimuli, we found that this interaction of cue size on ensemble perception was most likely dependent on the saliency of the cued Gabor relative to the other Gabors.

3.2 General Methods

Stimuli and Apparatus

The UC Berkeley Committee for Protection of Human Subjects approved all experimental procedures. We presented stimuli using the Psychophysics Toolbox (D. H. Brainard 1997; D. G. Pelli 1997; Kleiner, D. Brainard, and D. Pelli 2007; Cornelissen, Peters, and Palmer 2002) on a 53 cm Dell UltraSharp LCD monitor with 1680 x 1050 resolution, 60 Hz refresh rate, and 300 cd/m^2 peak brightness. We recorded eye position with the Eyelink 1000 (SR Research Ltd., Ottawa, Canada). Subjects sat at a distance of 50 cm from the screen with their heads on a chinrest in a dark room.

We conducted two experiments on two separate sets of subjects. Stimuli for both experiments consisted of an ensemble of equally-sized circular Gabor patches (100% contrast, 4 cycles; spatial frequency varied for each participant (see Experiment-specific methods); random spatial phase). A single "cued" Gabor within the ensemble was tilted 45°, 90°, or 135°away from horizontal, depending on the condition/task (see Experiment-specific methods).

The orientations of the remaining Gabors were pseudo-randomly chosen such that the arithmetic mean orientation of all Gabors in the ensemble (including the target) was 45°, 90°, or

135°away from horizontal, depending on the condition/task (see Experiment-specific methods). Gabor orientations in the range [0°, 180°] were chosen in pairs/groups using an iterative algorithm that maintained a consistent overall arithmetic mean across the iterations. For each iteration of the algorithm, one orientation was sampled from a logistic distribution (centered on the desired mean; 45°standard deviation; truncated to [0°, 180°]). Next, the symmetric opposite orientation (with respect to the mean) was chosen in order to maintain the desired arithmetic mean. If the symmetric opposite orientation exceeded the limits [0°, 180°], then the minimum number of equal non-symmetric opposite orientations (with respect to the mean) were drawn from the logistic distribution to maintain the desired arithmetic mean. For example, if 45° was the ensemble mean, and 110°was the sampled orientation, then two Gabors with 12.5° orientations were used to maintain the mean. This procedure was repeated until all orientations were chosen for a given ensemble.

The cued Gabor's orientation was always used for the first iteration of the algorithm. If a given iteration's sampled orientation led to a number of orientations exceeding the desired number of total orientations, then the iteration was repeated with a different orientation sample. If all but one orientation in the ensemble was chosen, then the final orientation was set to the desired arithmetic mean. The non-target orientations were then randomly assigned to Gabors that were spatially arranged in the Experiment-specific stimulus arrays. This procedure generated a large amount of trial-to-trial variability in the individual orientations while maintaining a consistent arithmetic mean of orientations for a given condition/task.

Spatial Attention Cueing

For both experiments, we employed a spatial attention cueing paradigm (Posner, Cohen, and R. D. Rafal 1982) to direct exogenous spatial attention to one side of the screen. After a 1200 ms fixation period at the start of each trial, a pink (RGB color code: [1.00 0.714 0.757]) circular peripheral cue (centered at the cued Gabor's location on either the left or right side of the screen) briefly flashed for 40 ms on a gray background. For half of the trials, the size of the cue was set to the size of a single Gabor (Small). For the other half of the trials, the size of the cue was set to encompass both the cued and surrounding Gabors (Large). The cues contained a cosine taper at the edge that started at one-half the radius from the center of the cue so that the edges would blend with the background. Immediately following the 40 ms cue presentation, the stimulus was presented for 133 ms either on the same (valid, 50% of trials) or opposite (invalid, 50% of trials) side of the screen as the cue. Subjects performed a 2AFC task (45°or 135°) using a key press either reporting the orientation of just the cued Gabor target (Experiment 1, session 1) or the mean orientation of the ensemble of Gabors (Experiment 1, session 2 and Experiment 2).

Statistical Analyses

Percent correct values on the target discrimination and Ensemble tasks were analyzed with a repeated-measures ANOVA with cue size (small or large), cue validity (valid or invalid), and cued Gabor/mean orientation relationship (congruent, neutral, or incongruent) entered as within-subject factors. If the overall three-way interaction of the ANOVA was significant, we conducted ad-

ditional two-way (cue size x validity) ANOVAs on each of the three isolated cued Gabor/mean orientation relationship conditions. Finally, we conducted planned comparisons to assess the effects of the different cue size on performance. Specifically, for each cue size, we defined the cueing effect as a pairwise difference between percent correct values when the stimulus appeared on the same side as the cue (valid) and values when the stimulus appeared on the opposite side (invalid). We used two-tailed Student's t-tests to assess if the means of the cueing effects were significantly different than zero. For correlation analyses, Pearson's r values were calculated and tested against the null hypothesis of a correlation coefficient value of zero.

3.3 Experiment 1

Materials and Methods

Subjects

Nineteen subjects (12 females, 7 males; age 18–30 years) with normal or corrected-to-normal vision participated in this experiment. All subjects were naive to the experimental design, and they were compensated for their time.

Stimulus Array

The stimulus array consisted of an ensemble of ten equally-sized circular Gabor patches (100% contrast, 4 cycles; spatial frequency varied for each participant (see below); random spatial phase) arranged such that one central Gabor (the target) was surrounded by a circle of nine uniformly-spaced Gabors. The target was tilted 45°, 90°, or 135°away from horizontal, depending on the condition/task (see below). The arithmetic mean of the ensemble (described in General Methods) could be 45°, 90°, or 135°, depending on the condition/task (see below)

Procedure

All subjects completed two experimental sessions with an interval of at least 24 hours between sessions. During the first session, subjects completed two baseline experiments. The first of these experiments was used to derive a threshold size for a single Gabor without surrounding Gabors for each participant. To do this, we presented a single Gabor (100% contrast, 4 cycles; random spatial phase; eccentricity of 10 degrees of visual angle) on either the left or right side of the screen (balanced across subjects) and used a 3-down/1-up staircase procedure to adjust the diameter of the Gabor in units of degrees of visual angle (133 ms stimulus presentation; 100 trials; 1.5 degrees initial diameter; 0.1 degree staircase step size). Subjects performed a 2-alternative forced choice (2AFC) task on the orientation of the Gabor (45°or 135°) using a key press. The size of all Gabors for all subsequent experimental sessions for a given subject was set to 1.5 times the average of the last six size reversals in the staircase from the first baseline experiment. We chose this value for the diameter of the Gabors so that orientation discrimination in the subsequent experiments was
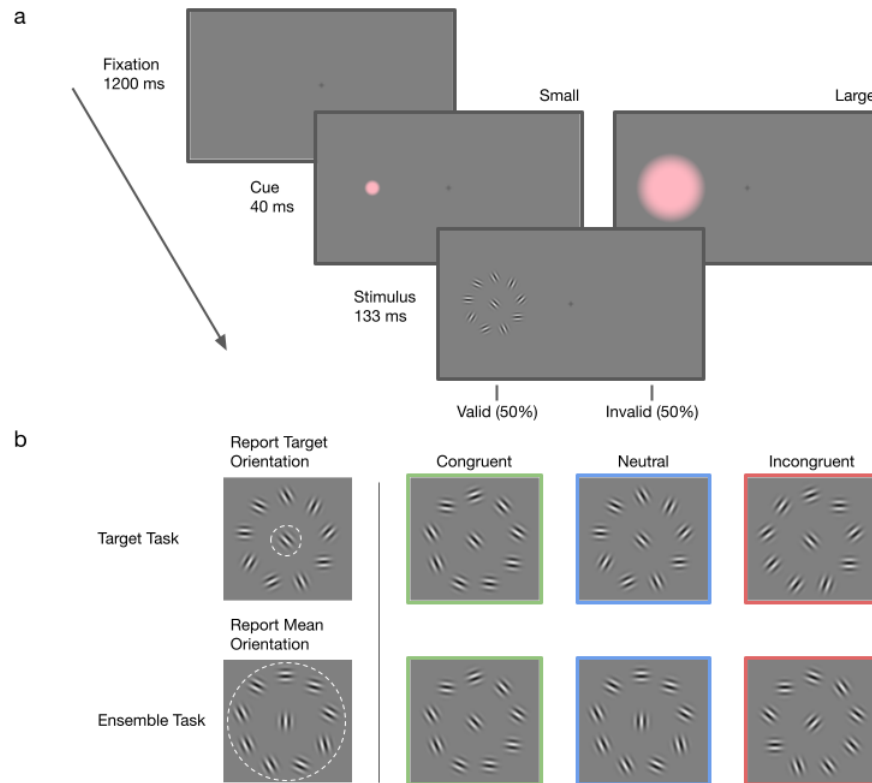
Figure 3.1: *(a) Schematic of the spatial attention cueing task (Experiment 1).* After a fixation interval, a pink circular peripheral cue briefly appeared on one side of the screen for 40 ms. Immediately following the cue, the stimulus array, consisting of a central Gabor surrounded by nine additional Gabors, either appeared on the same side as the cue (valid, 50% of trials) or on the opposite side of the cue (invalid, 50% of trials) for 133 ms. On half of the trials (Small cue), the cue was the size of the the central Gabor. On the other half of the trials (Large cue), the cue was the size of the entire stimulus array. *(b) Task type and cued Gabor/mean orientation relationship.* During the first session, participants were instructed to report the orientation (45°or 135°) of the central Gabor (Target task; upper row). For a given trial of the Target task, the mean orientation of the ensemble (central plus surrounding Gabors) was equally likely to be congruent (green), neutral (blue), or incongruent (red) with respect to the central Gabor's orientation. During the second session, participants were instructed to report the mean orientation (45°or 135°) of the ensemble (Ensemble task; lower row). For a given trial of the Ensemble task, the orientation of the central Gabor was equally likely to be congruent (green), neutral (blue), or incongruent (red) with respect to the ensemble mean orientation.

limited by cued Gabor/mean orientation interactions and not by the visibility of individual Gabors. Note that this procedure of customizing the size of the Gabor patches resulted in differences across participants in the spatial frequency of the Gabor patches. The range of spatial frequencies was

3.48–6.40 cycles/degree.

The second baseline experiment was used to derive a threshold spacing between the central Gabor and surrounding circle of Gabors for each subject. To do this, we presented the central and surrounding Gabors on the same side of the screen as in the first baseline experiment. The size and spatial frequency of all Gabors were based on the results of the first baseline experiment for each participant. The arithmetic mean of the orientations in the ensemble was set to $90°$ (i.e., vertical) for all trials in this baseline experiment. The orientations of the individual surrounding Gabors were selected from a logistic distribution (see above).

We used a 2-down/1-up staircase procedure to adjust the center-to-surround Gabor spacing, measured in degrees of visual angle from the center of the central Gabor to the center of any surrounding Gabor (133 ms stimulus presentation; 150 trials; $7°$ initial spacing; $0.2°$ staircase step size). Subjects performed a 2AFC task on the orientation of the central Gabor ($45°$ or $135°$) using a key press. The center-to-surround spacing for all subsequent experimental sessions for a given subject was set to the average of the last six spacing reversals in the staircase ($\sim 75\%$ threshold). We selected a threshold spacing for each subject to avoid floor and ceiling effects that could limit our ability to effectively measure the effects of attention on performance.

For the remaining sessions, we employed a spatial attention cueing paradigm (see General Methods) to compare the effects of the precision of a peripheral cue on an orientation discrimination crowding task (Target task; session 1) versus an ensemble mean orientation discrimination task (Ensemble task; session 2) that had the same stimulus parameters.

Spatial Cues   Figure 3.1.a shows the display sequence. The cue was centered at 10 degrees of visual angle from fixation on either the left or right side of the screen. The small cue was set to the size of the central Gabor. The large cue was set to a size that encompassed both the central and surrounding Gabors. Subjects performed a 2AFC task ($45°$ or $135°$) using a key press, either reporting the orientation of just the central Gabor (Target task; session 1) or the mean orientation of the ensemble of Gabors (Ensemble task; session 2).

Target task   During the Target task (first session), participants were instructed to report the orientation of the cued target Gabor (central Gabor; $45°$ or $135°$) using a key press. On each trial, the mean orientation of the ensemble (central plus surrounding Gabors) could either be congruent (ensemble mean orientation equal to cued Gabor orientation), neutral (ensemble mean orientation equal to vertical ($90°$)), or incongruent (ensemble mean orientation was orthogonal to the cued Gabor orientation) (Figure 3.1.b, upper row). Participants were not explicitly made aware of the three types of cued Gabor/mean orientation relationships. Each subject completed one practice block, followed by eight blocks of 120 trials each (960 trials total). All cue size (small or large), cue validity (valid or invalid), and cued Gabor/mean orientation (congruent, neutral, incongruent) conditions were randomly interleaved and counterbalanced (80 trials per unique combination of conditions).

Ensemble task   During the Ensemble task (second session), the same participants that previously completed the Target task in the first session were instructed to report the mean orientation of the ensemble (central and surrounding Gabors; 45°or 135°) with a key press. Additionally, the cued Gabor's orientation (central Gabor) could either be congruent (cued Gabor orientation equal to ensemble mean orientation), neutral (cued Gabor orientation equal to vertical (90°)), or incongruent (cued Gabor orientation was orthogonal to the ensemble mean orientation) (Figure 3.1.b, lower row). Participants were not explicitly made aware of the three types of cued Gabor/mean orientation relationships, and they were instructed to weight all orientations equally in determining the mean. Each subject first completed one practice block, followed by eight blocks of 120 trials each (960 trials total). Like the Target task, all cue size (small or large), cue validity (valid or invalid), and cued Gabor/mean orientation (congruent, neutral, incongruent) conditions were randomly interleaved and counterbalanced (80 trials per unique combination of conditions).

Results

Small cue benefits Target task performance.   During the first session of the experiment, participants were instructed to report the orientation of a central Gabor that was surrounded by a ring of distractor Gabors (Target task). Immediately following a brief peripheral cue, the Gabor array was equally likely to appear on the same side of the screen as the cue (valid trials) or in a matched peripheral location on the opposite side of the screen (invalid trials). For all trials, the cue directed a participant's involuntary attention to its location. We varied the size of the cue to study the effects of spatial precision of stimulus-driven attention on crowding performance. Specifically, we compared target identification performance when the cue precisely covered the spatial extent of just the central Gabor (small cue) versus when the cue covered the spatial extent of the entire stimulus array (large cue). Additionally, we varied the relationship between the cued target Gabor orientation and the mean orientation of the entire stimulus array to determine how the precision of the cues interacted with different patterns of ensemble statistics. Specifically, the mean orientation of the array could be congruent, neutral, or incongruent with the cued target Gabor orientation.

The top of Figure 3.2 shows overall Target task performance for each of the cueing and cued Gabor/mean orientation conditions. A repeated-measures ANOVA revealed a significant main effect of cued Gabor/mean orientation relationship on performance ($F_{2,36} = 27.4$, $p = 5.86e-08$), with highest performance for the congruent condition and lowest performance for the incongruent condition. Additionally, the two-way interaction between cue size and validity was significant ($F_{1,18} = 6.05$, $p = 0.024$). However, the three-way interaction of cue size, validity, and cued Gabor/mean orientation relationship was not significant ($F_{2,36} = 0.71$, $p = 0.499$). Taken together, these results indicate that cue size and validity produce significantly different patterns of Target task performance but that these interactions were not significantly affected by cued Gabor/mean orientation relationship.

The interaction between cue size and validity was further explored by calculating cueing effects for the the two cue sizes. The bottom of Figure 3.2 shows the magnitude of these cueing effects (percent correct values for valid trials minus invalid trials) for all cued Gabor/mean orientation relationships. For the trials in which the cue precisely covered just the central Gabor (small cue),

Figure 3.2: Target task results. *(Top)* The effect of cue size (Small or Large) and cue validity (Valid or Invalid) on Target task performance for congruent (left; green bars), neutral (middle; blue bars), and incongruent (right; red bars) cued Gabor/mean orientation relationships. Gray lines represent matched individual subject data across validity conditions. *(Bottom)* Mean within-subject cueing effects, defined as the difference between percent correct values when the stimulus appeared on the same side as the cue (Valid) and when the stimulus appeared on the opposite side of the cue (Invalid). Gray dots represent individual subjects, and asterisks indicate significant cueing effects (level $\alpha < 0.05$, based on a planned comparison paired t-test). Error bars are standard errors of the mean.

performance was better for valid compared to invalid trials. Specifically, for the trials in which the mean ensemble orientation had the same orientation as the central Gabor (congruent; green bars), the magnitude of the cueing effect for the small cue was 3.75% correct (88.6% valid - 84.8% invalid) (Figure 3.2; bottom left), and this was significantly greater than zero ($t_{18} = 3.21, p = 0.005$). For the trials in which the mean ensemble orientation was always vertical 90° (neutral; blue bars), the magnitude of the cueing effect for the small cue was 6.18% correct (77.3% valid - 71.1% invalid) (Figure 3.2; bottom center), and this was also significantly greater than zero ($t_{18} = 2.52, p = 0.021$). Finally, for the trials in which the mean ensemble orientation was orthogonal to the orientation of the central Gabor (incongruent; red bars), the magnitude of the cueing effect for

the small cue was 6.51% correct (76.7% valid - 70.2% invalid) (Figure 3.2; bottom right), and this was once again significantly greater than zero ($t_{18} = 3.23, p = 0.005$).

For the trials in which the cue covered an area containing both the central Gabor and the surrounding Gabors (large cue), performance was similar for valid and invalid trials (Figure 3.2; bottom). Specifically, across all cued Gabor/mean orientation relationships, the magnitude of the cueing effect for the large cue was $\leq 1.71\%$, and this was not significantly different from zero for any cued Gabor/mean orientation relationship ($t_{18} \leq 1.31, p \geq 0.205$). These results suggest that stimulus-driven attention does not substantially benefit performance on a crowding task if the cue encompasses the area containing both the target and the surrounding distractors.

Figure 3.3: Target task performance binned by standard deviation of ensemble orientations. (*left*) Mean percent correct values (combining valid and invalid trials and small-cue and large-cue trials) binned by the standard deviation of the ensemble's orientations (equal number of trials in each of five bins). The lines were fit by least-squares linear regression. (*right*) The effect of cue size and validity on the slope of the regression lines fit to each subject's binned data (top row), and the corresponding cueing effects (bottom row) for each cued Gabor/mean orientation condition. Individual subject data, error bars, and significance levels are displayed as in Figure 3.2.

Greater standard deviation of ensemble orientations affects Target task performance differently for each cued Gabor/mean orientation condition. For all trials of the experiment, the arithmetic mean orientation of the ensemble was fixed to be one of three values ($45°$, $90°$, or $135°$). However, the deviation of orientations from this mean varied from trial to trial. To examine the effect of standard deviation of ensemble orientations on performance, we binned trials for each participant based on standard deviation.

The left part of Figure 3.3 shows mean percent correct values (combining valid and invalid trials and small-cue and large-cue trials) as a function of standard deviation (in degrees) of the ensemble's orientations for each cued Gabor/mean orientation condition. We fit a linear regression

line (slope with units of percent correct/standard deviation in degrees) to the mean percent correct values of the five bins (equal number of trials per bin). Overall, as standard deviation increased, mean performance significantly decreased for congruent trials (green line; slope = -0.17, $t_{18} = -5.94$, $p = 1.27e - 05$), stayed relatively the same for neutral trials (blue line; slope = -0.06, $t_{18} = -1.44$, $p = 0.166$), and significantly increased for incongruent trials (red line; slope = 0.40, $t_{18} = 4.55$, $p = 2.46e - 04$). These results suggest that target orientation discrimination was largely affected by the magnitude of deviation of the individual orientations and not only by the mean orientation of the ensemble.

The right part of Figure 3.3 shows mean slope values for each cueing and cued Gabor/mean orientation condition separately (top row) and the corresponding cueing effects (bottom row). A repeated-measures ANOVA revealed a main effect of cued Gabor/mean orientation relationship on slope ($F_{2,36} = 22.5$, $p = 4.89e - 07$). Furthermore, the three-way interaction of cue size, validity, and cued Gabor/mean orientation condition on slope was also significant ($F_{2,36} = 5.07$, $p = 0.011$). Results from planned-comparison paired t-tests of the individual cueing effects only showed a significant effect for the small cue in the congruent trials (mean slope difference = -0.18; $t_{18} = -2.11$, $p = 0.049$). All other cueing effects on slope were not significantly different from zero ($-1.76 \leq t_{18} \leq 1.87$, $p \geq 0.078$). Since the three-way interaction was significant, we conducted additional two-way (cue size x validity) ANOVAs for each of the cued Gabor/mean orientation conditions. Only the incongruent trials showed a significant two-way interaction on slope ($F_{1,18} = 5.44$, $p = 0.032$). All other cued Gabor/mean orientation conditions did not have a significant interaction ($F_{1,18} \leq 2.95$, $p \geq 0.102$). These results provide evidence that, for the incongruent trials, the magnitude of the cueing effect for the small cue was more sensitive (steeper slope) to changes in the deviation of the ensemble's orientation, compared to the large cue.

The small cue biases reports of the mean orientation in the direction of the central Gabor's orientation more than the large cue.   During the second session of the experiment, participants were instructed to report the arithmetic mean orientation of the entire stimulus array (Ensemble task). The stimuli used in the Ensemble task had almost identical parameters to those used in the Target task, and the same cueing procedure was employed in the two tasks. For the Ensemble task, we examined how the spatial precision of stimulus-driven attention affected performance on a task that required a more distributed attentional focus. As in the Target task, we varied the relationship between the mean orientation and the cued central Gabor's orientation, this time to examine if the different cue sizes directed ensemble perception away from or towards the central Gabor's orientation. The central cued Gabor could be congruent, neutral or incongruent with respect to the mean orientation.

The top of Figure 3.4 shows mean task performance for each of the cueing and cued Gabor/mean orientation conditions. As for the Target task, a repeated-measures ANOVA revealed a significant main effect of cued Gabor/mean orientation relationship on performance ($F_{2,36} = 114.22$, $p = 2.58e - 16$), with highest performance for the congruent condition and lowest performance for the incongruent condition. Unlike the Target task, the two-way interaction between cue size and validity on performance was not significant for the Ensemble task ($F_{1,18} = 1.74$,
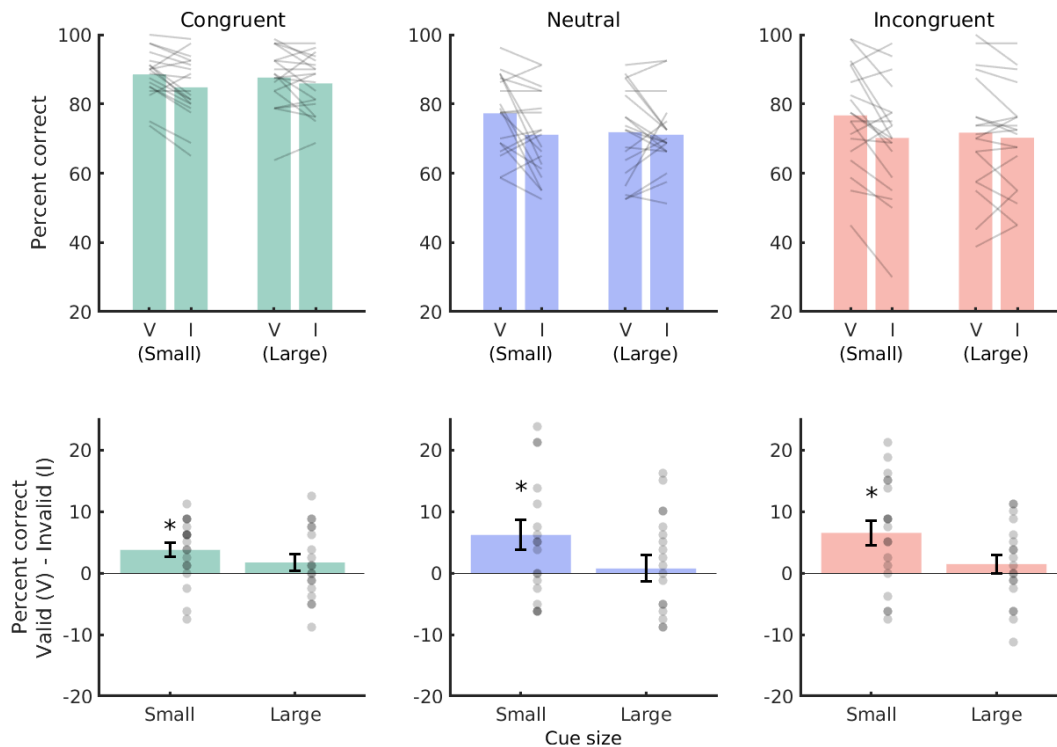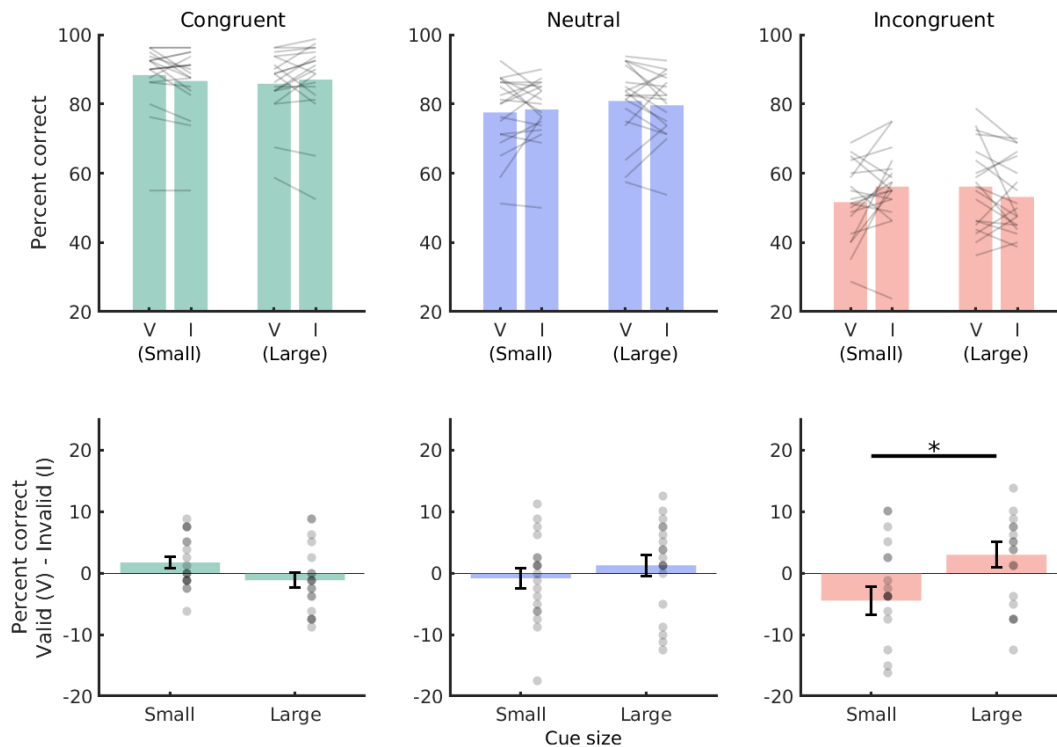
Figure 3.4: Ensemble task results. *(Top)* The effect of cue size (Small or Large) and cue validity (Valid or Invalid) on Ensemble task performance for congruent (left; green bars), neutral (middle; blue bars), and incongruent (right; red bars) cued Gabor/mean orientation relationships. Gray lines represent matched individual subject data across validity conditions. *(Bottom)* Mean within-subject cueing effects, defined as the difference between performance when the stimulus appeared on the same side as the cue (Valid) and when the stimulus appeared on the opposite side of the cue (Invalid). Gray dots represent individual subjects, and the asterisk with a bar underneath indicates significance level $\alpha < 0.05$ from a condition-specific two-way repeated-measures ANOVA (cue size x validity). Error bars are standard errors of the mean.

$p = 0.204$). However, there was a significant three-way interaction of cue size, validity, and cued Gabor/mean orientation relationship ($F_{2,36} = 4.54$, $p = 0.017$).

The bottom of Figure 3.4 shows cueing effect magnitudes. For both cue sizes, performance was not significantly different when the cue was valid compared to when it was invalid for any of the cued Gabor/mean orientation relationships ($-1.96 \leq t_{18} \leq 1.77, p \geq 0.066$). However, since the three-way effect of cue size, validity, and target/ensemble relationship from the ANOVA suggested that there were significant differences in how the cue size affected attentional modulation across the different cued Gabor/mean orientation groups, we conducted two-way ANOVAs for each of the cued Gabor/mean orientation groups to test for cue size/validity interactions. Results from

these ANOVAs showed a significant interaction between cue size and validity for the incongruent condition ($F_{1,18} = 4.74$, $p = 0.043$) but not for the congruent or neutral conditions ($F_{1,18} \leq 3.25, p \geq 0.088$). These results and the directions of the effects in Figure 3.4 (bottom row) provide some evidence that, for incongruent trials, a valid small cue biased perception toward the central Gabor's orientation more than the large cue did. This is further examined in the next section, where we characterize how standard deviation of the ensemble affects performance.

Greater deviation from the mean orientation negatively impacts Ensemble task performance.
As for the Target task, we also examined Ensemble task performance for trials binned by the standard deviation of the ensemble's orientations. The left part of Figure 3.5 shows the mean percent correct values (combining valid and invalid trials and small-cue and large-cue trials) as a function of standard deviation of the orientations. Overall, larger standard deviations were associated with lower Ensemble task performance for all cued Gabor/mean orientation relationships (slope $\leq -0.37, t_{18} \leq -8.53, p \leq 9.68e - 08$). Interestingly, for the incongruent trials at the largest deviation, performance actually fell below chance (50 %) performance. This is further evidence that reports of the mean orientation were biased in the direction of the central Gabor's orientation.

The right part of Figure 3.5 shows mean slope values for each cueing and cued Gabor/mean orientation condition separately (top row) and the corresponding cueing effects (bottom row). Results from a repeated-measures ANOVA revealed a main effect of cued Gabor/mean orientation relationship on slope ($F_{2,36} = 14.5$, $p = 2.38e - 05$), but neither the two-way interaction of cue size and validity nor the overall three-way interaction were significant ($F_{2,36} \leq 2.00$, $p \geq 0.149$). Finally, none of the cueing effects on slope were significantly different than zero, based on planned comparison paired t-tests ($-1.88 \leq t_{18} \leq 1.03, p \geq 0.076$). These results suggest that deviation from the mean orientation negatively affected Ensemble task performance and that these effects depended on the cued Gabor/mean orientation relationship. However, the cues did not have much of an effect on the rate at which the orientation deviation affected performance.

No detectable correlation in magnitude of cueing effects in the target and Ensemble tasks.
On average, when the small cue was valid, it consistently led to increased performance on the Target task, compared to when it was invalid. In contrast, the small cue did not always improve Ensemble task performance, and in the case of the incongruent trials, it actually decreased performance in the majority of participants. However, it is unclear if subjects who most effectively utilized the small cue in the Target task were also most affected by it in the Ensemble task. To answer this question, for each cue size, we correlated individual subjects' cueing effect magnitudes (valid trials - invalid trials) for the Target task with their cueing effects for the Ensemble task. Neither the cueing effects for the small cue (Figure 3.6; top row) nor for the large cue (Figure 3.6; bottom row) were significantly correlated across the two tasks for any of the cued Gabor/mean orientation conditions (See Figure 3.6 for Pearson's r and p-values).
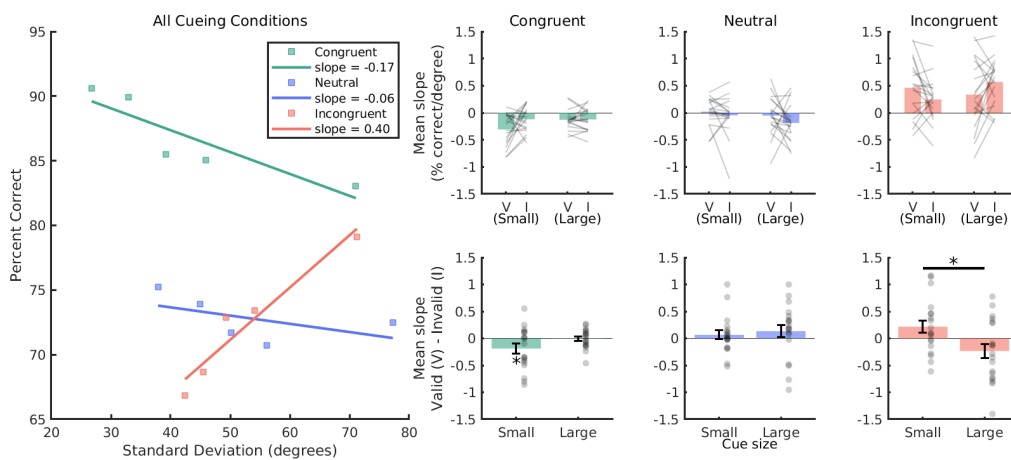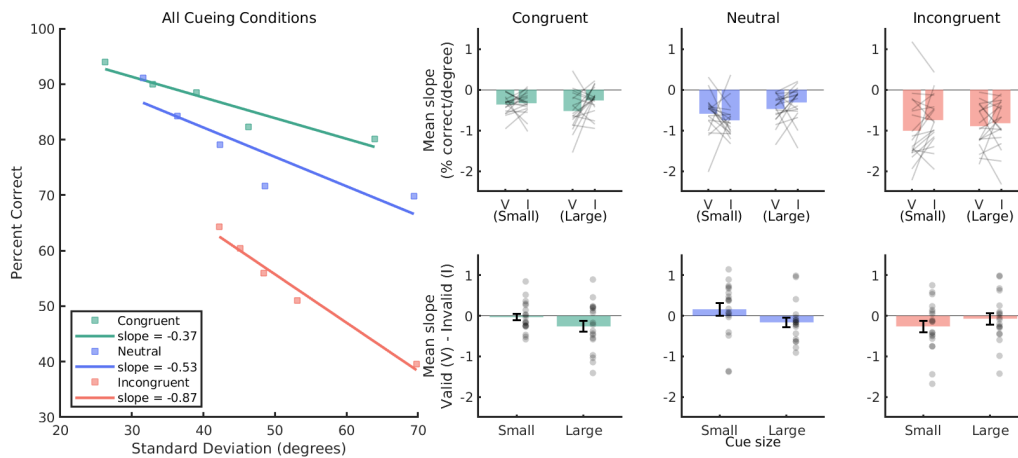
Figure 3.5: Ensemble task performance, binned by standard deviation of ensemble orientations. (*left*) Mean percent correct values (combining valid and invalid trials and small-cue and large-cue trials) were binned by standard deviation of the ensemble's orientations (equal number of trials in each of five bins). The lines were fit by least-squares linear regression. (*right*) The effects of cue size and validity on the slopes of the regression lines fit to each subject's binned data (top row), and their corresponding cueing effects (bottom row) for each cued Gabor/mean orientation relationship condition. Individual subject data and error bars are displayed as in Figure 3.2.

3.4 Experiment 2

Material and Methods

Subjects

A new group of twenty subjects (13 females, 7 males; age 22–32 years) with normal or corrected-to-normal vision participated in this experiment. All subjects were naive to the experimental design, and they were compensated for their time.

Stimulus Array

The stimulus array consisted of a grid of sixteen (4x4) uniformly-spaced circular Gabor patches (100% contrast, 4 cycles; random spatial phase). Each Gabor had a diameter of $0.85°$, which was the mean Gabor size across all subjects in Experiment 1. The grid spacing was $2°$, which was also the mean center-to-center spacing across subjects for the surrounding Gabors in Experiment 1. The grid was centered at $10°$ of visual angle to the left or right of fixation. Unlike the fixed central cued Gabor in Experiment 1, the cued Gabor now was equally likely to be any one of the four most central Gabors within the grid. Centering the cue on only one of these Gabors ensured that the cued Gabor always had flanking stimuli on all four sides. The cued Gabor could be tilted
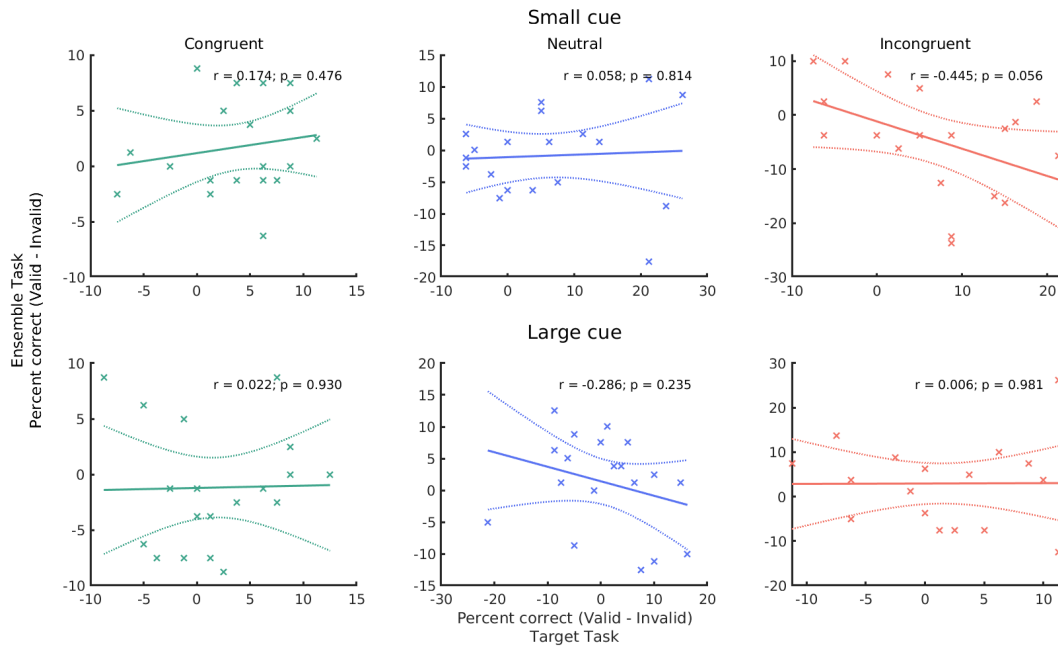
Figure 3.6: Cueing effects (percent correct for valid trials - invalid trials) for the Target task and the Ensemble task were not significantly correlated across individual subjects for either cue size (small and large) or for any of the cued Gabor/mean orientation conditions (congruent: green; neutral: blue; incongruent: red). The 'x's represent individual subjects. The solid and dashed lines represent linear regression fits and 95% confidence intervals, respectively. Pearson's r and p values for the correlations are displayed in the upper right corner of each plot.

$45°$, $90°$, or $135°$ away from horizontal, depending on the condition (see below), and the arithmetic mean orientation of the ensemble (described in General Methods) was either $45°$ or $135°$.

Procedure

The spatial attention cueing procedure was the same as in Experiment 1 (Figure 3.1.a), except here the cue was equally likely to be centered at the location of any one of the four Gabors within the center of the grid on either side of the screen. The black dashed line in Figure 3.7 shows an example of one possible cue location. The size of the small cue was equal to the size of a single Gabor in the grid ($0.85°$), and the size of the large cue was equal to $8.85°$ to encompass the target and the adjacent Gabors. The stimulus array was either on the same side as the cue (valid trials) or on the opposite side as the cue (invalid trials).

Participants performed a similar task as the Ensemble task in session 2 of Experiment 1. They were instructed to report the mean orientation of the ensemble (all sixteen Gabors; $45°$ or $135°$) using a key press. The cued Gabor's orientation (equally likely to be one of the four Gabors within the center of the grid) could be congruent, neutral, or incongruent (Figure 3.7) with respect to the
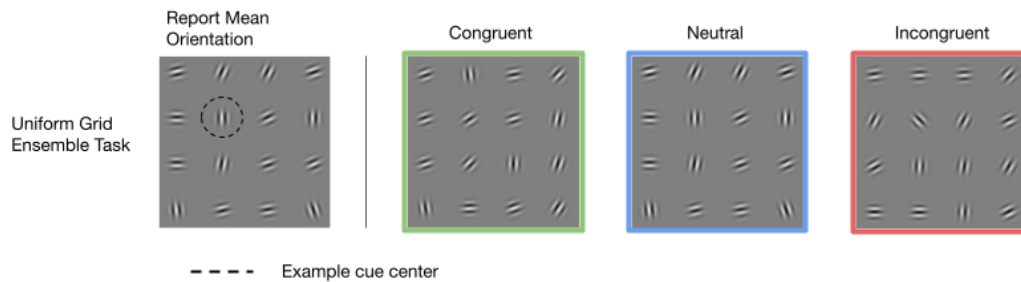
Figure 3.7: Uniform grid task and cued Gabor/mean orientation relationships. Similar to session 2 of Experiment 1 (Ensemble task), participants were instructed to report the mean orientation (45°or 135°) of the ensemble. However, here the ensemble was a 4x4 grid, and the cued Gabor could be any one of the four Gabors within the center of the grid. The black dashed line in the first panel shows an example of a possible cued Gabor location. The orientation of the cued Gabor was equally likely to be congruent (green), neutral (blue), or incongruent (red) with respect to the ensemble mean orientation.

mean orientation of the ensemble. Participants were not explicitly made aware of the three types of cued Gabor/mean orientation relationships, and they were instructed to weight all orientations equally in determining the mean. Each subject first completed practice blocks of 60 trials until they were able to achieve at least 70% accuracy. The practice blocks were followed by eight blocks of 120 trials each (960 trials total). All cue size (small or large), cue validity (valid or invalid), and cued Gabor/mean orientation (congruent, neutral, incongruent) conditions were randomly inter-leaved and counterbalanced (80 trials per unique combination of conditions).

Results

In Experiment 1, the same participants completed both the crowding task and the Ensemble task, and the two tasks had almost identical stimulus parameters. However, this decision had a number of trade-offs. 1) To avoid awareness that there were systematic differences between the central orientation and the mean of the ensemble in the Target task, participants always completed the Target task before the Ensemble task. 2) Ensemble task performance fell below chance in the incongruent trials with the largest standard deviation (left part of Figure 3.5), suggesting that the saliency of the cued central Gabor (independent of the cue) might have influenced performance for stimulus arrays containing a central Gabor surrounded by a circle of flanking Gabors.

To determine whether the three-way interaction of cue size, validity, and cued Gabor/mean orientation that was observed in the Ensemble task in Experiment 1 was still evident after controlling for these factors, a new set of subjects reported the mean orientation of a 4x4 grid of uniformly-spaced Gabors. The cueing procedure was the same as in Experiment 1, but here the cue could be centered at any one of the four locations of the Gabors within the center of the grid. We tested the

same three cued Gabor/mean orientation relationships as in Experiment 1: congruent, neutral, and incongruent.

The top of Figure 3.8 shows mean task performance for each of the cueing and cued Gabor/mean orientation conditions. A repeated-measures ANOVA revealed a significant main effect of cued Gabor/mean orientation relationship on performance ($F_{2,38} = 112$, $p = 1.259e - 16$), with highest performance for the congruent condition and lowest performance for the incongruent condition. However, neither the two-way interaction between cue size and validity ($F_{1,19} = 0.121$, $p = 0.732$) nor the three-way interaction of cue size, validity and cued Gabor/mean orientation relationship was significant ($F_{2,38} = 1.95$, $p = 0.156$).

The bottom of Figure 3.8 shows cueing effect magnitudes. Interestingly, only the cueing-effect magnitude for the large cue in the congruent trials was significantly different from zero (-4.56 percent correct; $t_{19} = -2.94$, $p = 0.008$; all other cueing effects: $-0.812 \leq t_{19} \leq 0.944$, $p \geq 0.357$).

As for Experiment 1, we also examined task performance for trials binned by the standard deviation of the ensemble's orientations. The left part of Figure 3.9 shows mean percent correct values (combining valid and invalid trials and small-cue and large-cue trials) as a function of standard deviation of the orientations. Similar to the Ensemble task in Experiment 1, larger standard deviations were associated with lower task performance for all cued Gabor/mean orientation relationships (slopes ≤ -0.57, $t_{19} \leq -10.0$, $p \leq 4.95e - 09$).

The right part of Figure 3.9 shows mean slope values for each cueing and cued Gabor/mean orientation condition separately (top row) and the corresponding cueing effects (bottom row). Results from a repeated-measures ANOVA revealed a main effect of cued Gabor/mean orientation relationship on slope ($F_{2,38} = 3.92$, $p = 0.028$), but neither the two-way interaction of cue size and validity ($F_{1,19} = 2.61e - 04$, $p = 0.987$) nor the three-way interaction of cue size, validity, and cued Gabor/mean orientation ($F_{2,38} = 2.38$, $p = 0.106$) was significant. Finally, none of the cueing effects on slope were significantly different than zero, based on results from planned comparison paired t-tests ($-2.06 \leq t_{19} \leq 1.10$, $p \geq 0.053$). These findings are consistent with those observed for the Ensemble task of Experiment 1.

3.5 Discussion

We varied the size of a peripheral cue to compare the effects of the spatial precision of stimulus-driven attention on visual crowding (Experiment 1, session 1) and ensemble perception (Experiment 1, session 2 and Experiment 2). We also tested whether attention interacted with the relationship between a single cued Gabor and the mean orientation of the ensemble. For all experiments, the cued Gabor's orientation and the mean ensemble orientation were congruent (cued Gabor and mean orientations were parallel), neutral (either the cued Gabor or mean orientation was fixed at vertical 90°), or incongruent (cued Gabor and mean orientations were orthogonal). Subjects were not explicitly made aware of these cued Gabor/mean orientation relationships.

In Experiment 1, the same set of subjects completed two sessions: 1) an orientation discrimination crowding task (Target task), and 2) an ensemble mean orientation discrimination task (En-
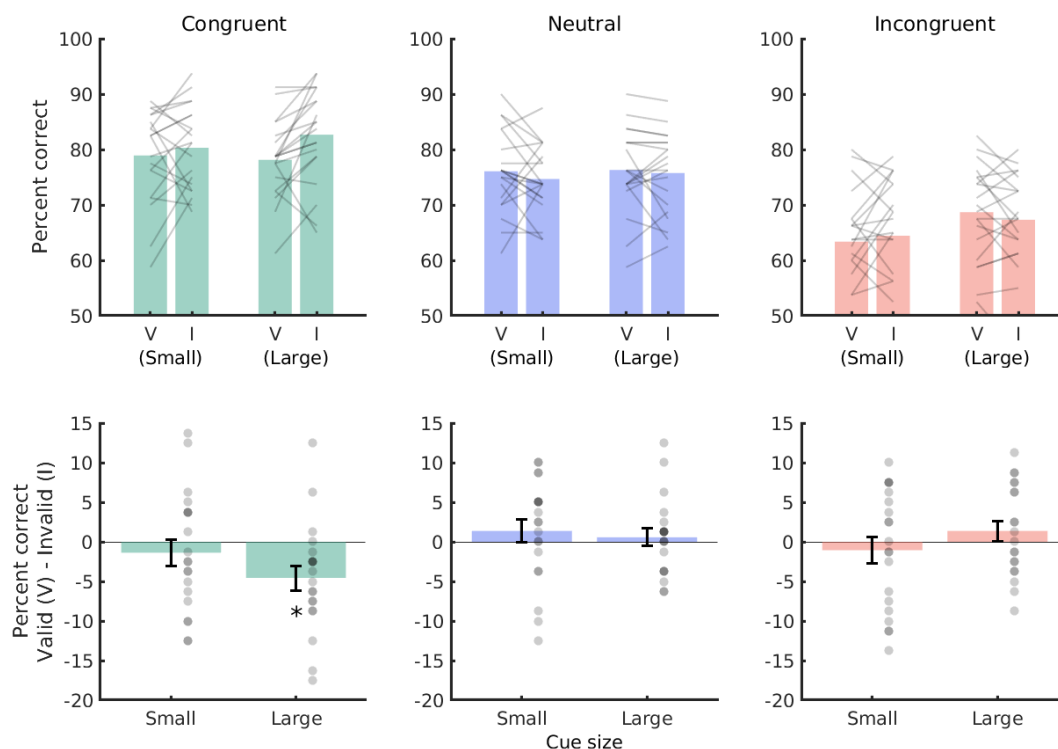
Figure 3.8: Uniform grid Ensemble task results. *(Top)* The effect of cue size (Small or Large) and cue validity (Valid or Invalid) on performance for congruent (left; green bars), neutral (middle; blue bars), and incongruent (right; red bars) cued Gabor/mean orientation relationships. Gray lines represent matched individual subject data across validity conditions. *(Bottom)* Mean within-subject cueing effects, defined as the difference between performance when the stimulus appeared on the same side as the cue (Valid) and when the stimulus appeared on the opposite side of the cue (Invalid). Gray dots represent individual subjects, and asterisks indicate significant cueing effects (level $\alpha < 0.05$ from a planned comparison paired t-test). Error bars are standard errors of the mean.

semble task). For the Target task, we found that the small cue significantly increased performance, while the large cue had no significant effect on performance. These cueing effects were the same regardless of the relationship between the cued Gabor and the mean of the ensemble. Binning the trials based on standard deviation of orientations in the ensemble further revealed differences in how this relationship affected performance.

For the Ensemble task, neither the small nor the large cue produced significant cueing effects on performance, but we did observe a three-way interaction among cue size, validity, and cued Gabor/mean orientation relationship. Condition-specific analyses provided additional evidence that the small cue more strongly biased reports of the mean orientation towards that of the central cued
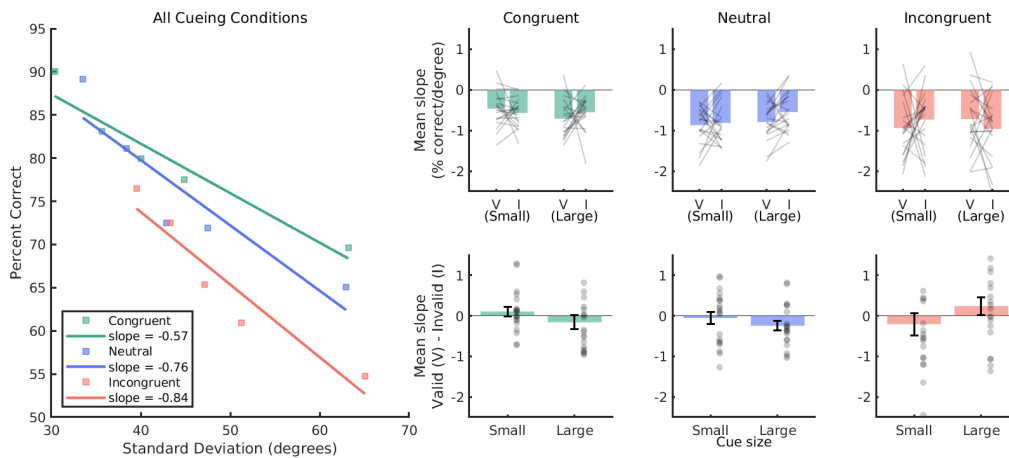
Figure 3.9: Uniform grid Ensemble task performance binned by standard deviation of orientations. (*left*) Mean percent correct values (combining valid and invalid trials and small-cue and large-cue trials) binned by standard deviation of the ensemble's orientations (equal number of trials in each of five bins). The lines were fit by least-squares linear regression. (*right*) The effects of cue size and validity on the slope of the regression lines fit to each subject's binned data (top row), and the corresponding cueing effects (bottom row) for each cued Gabor/mean orientation relationship condition. Individual subject data and error bars are displayed as in Figure 3.2.

Gabor's orientation than the large cue, but this interaction was only significant for the incongruent trials. Finally, we found no detectable correlation in the magnitudes of individual subjects' cueing effect between the target and Ensemble tasks.

In Experiment 2, a separate set of subjects performed a task that was similar to the Ensemble task in Experiment 1, except that the ensemble was a 4x4 grid of uniformly-spaced Gabors. This experiment was designed to better control for the saliency of the central cued Gabor relative to the other Gabors when they were arranged in a circular pattern. Interestingly, we did not observe the same significant three-way interaction among cue size, validity, and cued Gabor/mean orientation relationship in Experiment 2 as we did in Experiment 1. Taken together, these results suggest that the effects of spatial precision of a peripheral cue on ensemble perception that were observed in Experiment 1 were most likely dependent on the saliency of the individual cued Gabor relative to the non-cued Gabors.

The spatial precision of exogenous attention   It has been shown that stimulus-driven attention decreases the effects of crowding on identification and discrimination (Scolari, Kohnen, et al. 2007; Yeshurun and Rashal 2010; Albonico et al. 2018) and increases the spatial resolution of perception (Yeshurun and Carrasco 1998; Yeshurun and Carrasco 2008; Barbot and Carrasco 2017) at the cued location. However, there is some debate regarding the effects of cue size on perception.

For a peripheral texture segmentation task, Yeshurun and Carrasco 2008 found that a cue had to

be smaller than or equal to the size of the region that encompassed the target texture patch in order to increase the resolution of perception. Also, Yeshurun and Carrasco 2008 found that this effect of exogenous transient attention was similar to an all-or-nothing effect, in that they did not find a gradual change in performance by gradually increasing the cue size beyond the size of the target texture patch. Similarly, Albonico et al. 2018 tested the effects of three different cue types/sizes (dot, small box, and large box) on the critical spacing of crowding in a letter identification task. They found that only the dot cue significantly decreased critical spacing, while the larger cues did not.

In our study, the results from Experiment 1 support the idea that crowding is only relieved when the cue is small (i.e., the size of the target). However, in contrast to Yeshurun and Carrasco 2008, Albonico et al. 2018, and our study, Strasburger and Malania 2013 showed that an attention cue enhanced target contrast sensitivity in a crowding task in a manner that was independent of cue size. They further claimed that the cue had no effect on positional errors (i.e. reporting one of the flankers instead of the target) and only affected contrast thresholds of the target. Perhaps cue sizes modulate the effects of target/flanker spacing on target identification (i.e. the critical spacing in Albonico et al. 2018) but not contrast thresholds. However, our results provide some evidence that the increase in performance due to the small cue was independent of flanker confusions, as it was independent of the relationship between the cued Gabor and the mean orientation.

There are also open questions about how the size of a spatial cue interacts with the goals or properties of a visual task. Using a dual-task design, Burnett, d'Avossa, and Sapir 2013 found that identification of apparent motion was modulated differently by cue size, compared to localization of a single red-dot probe. The apparent motion task may have required more distributed attentional focus (like our Ensemble task) than the localization task (like our Target task). Therefore, the finding of Burnett, d'Avossa, and Sapir 2013 that cue size only had affected localization is consistent with our finding that cue size only had affected performance of the Target task and not the Ensemble task.

The role of attention in ensemble perception   Observers can extract summary representations of a number of features, such as size, orientation, color, motion, facial expression, and emotional content of an ensemble (Alvarez 2011). However, the role of attention in ensemble perception is not well understood. Spatial attention directed to one location in the visual field increases the apparent size of the adjacent spatial area (Fortenbaugh, Prinzmetal, and Robertson 2011; Kirsch, Heitling, and Kunde 2018), suggesting that directing attention to one element in an ensemble should influence the summary percept of the ensemble. Indeed, recent studies have shown that the reported mean orientation of an ensemble can be biased toward more salient objects (Iakovlev and Utochkin 2021), and attention directed to one object in a ensemble can bias the reported mean shape/size (De Fockert and Marchant 2008) and mean facial expression (Ying 2022) toward that of the attended object. However, there are also studies that have shown minimal effects of attention on ensemble perception (Talipski, Goodhew, and Edwards 2021; Ji, Rossi, and Pourtois 2018). In our study, we showed that observers performed above chance on a mean orientation discrimination task even if the individual Gabor orientations were difficult to identify. However, we only observed

attention effects on ensemble perception when the cued Gabor was more salient relative to the non-cued Gabors (i.e., the center/surround incongruent stimuli in Experiment 1).

Chapter 4

Response models of visual crowding and ensemble perception

4.1 Introduction

Identification of individual objects in clutter is severely limited in peripheral vision (Levi 2008; Whitney and Levi 2011; Rosenholtz 2016), but categorical perception remains intact. For example, it is possible to perceive a cluttered object in the periphery as a face without being able to determine its identity. Additionally, an ensemble of similar objects in the periphery can reliably result in summary statistical representations, such as the mean and variance of a set, even if the individual features are inaccessible (Whitney and Yamanashi Leib 2018; Haberman and Whitney 2012; Alvarez 2011; Manassi and Whitney 2018).

Limitations on target identification in visual crowding depend on visual eccentricity and target/flanker spacing, which are likely the result of the visual field not being sampled uniformly: neuronal receptive fields (RFs) scale as a function of eccentricity (Gattass, Gross, and Sandell 1981; Gattass, Sousa, and Gross 1988). Indeed, behavioral studies have shown that flanking stimuli that are presented more peripherally, relative to a target stimulus location, crowd more than those that are presented more foveally (Banks, Bachrach, and Larson 1977; Petrov and Meleshkevich 2011). In the last two decades, visual crowding has often been modeled as a pooling mechanism in which relative spatial information of features is discarded while retaining summary statistical representations of the features (Parkes et al. 2001; Balas, Nakano, and Ruth Rosenholtz 2009; Freeman and Simoncelli 2011; Van den Berg, Roerdink, and Cornelissen 2010; Keshvari and Ruth Rosenholtz 2016). However, there are additional aspects of crowding that cannot be wholly explained by pooling models, such as substitution errors (in which subjects report the identity of one of the flankers instead of the target (Hanus and Vul 2013; Ester, Klee, and Awh 2014; Coates, Bernard, and Chung 2019)), categorical target/flanker effects (Reuther and Chakravarthi 2014), global/contextual effects (Manassi, Sayim, and Herzog 2012; Herzog et al. 2015), holistic effects (Farzin, Rivera, and Whitney 2009), and attention effects (Yeshurun and Rashal 2010).

The combination of features into summary statistical representations that occurs for crowded

stimuli is similar to ensemble perception, defined as the ability to extract the gist from a group of visual features. Even in the periphery, where it is difficult or impossible to identify individual features, observers can still extract the mean and variance of a number of features, such as size, orientation, color, motion, facial expression, and emotion (Alvarez 2011; Manassi and Whitney 2018), as well as the range and modality of a distribution of features (Tanrıkulu, Chetverikov, and Kristjánsson 2020). Although observer's erroneous reports of a single object in a visual crowding task resemble ensemble perception, it is less clear which mechanisms are used to arrive at these similar outcomes. Furthermore, it is not currently known if observers share common strategies between the two tasks.

In this study, we modeled response patterns for visual crowding and ensemble perception tasks. One limitation of perceptual metrics, such as percent correct, is that it is difficult to know which aspects of the stimuli informed the observers' responses. By quantitatively modeling how each element of a crowded stimulus array (task-relevant or irrelevant) informs responses in visual crowding and ensemble perception, we can learn more about how these two similar tasks compare on both individual and group levels.

Previous studies have also modeled response patterns (or error distributions) of behavioral data from visual crowding tasks (Hanus and Vul 2013; Ester, Klee, and Awh 2014; Jimenez, Kimchi, and Yashar 2022). For example, Ester, Klee, and Awh 2014 showed that errors in a target identification crowding task were better explained by probabilistic substitutions of target/flanker features than by a simple averaging of the features. More recently, Jimenez, Kimchi, and Yashar 2022 used a mixture-modeling approach to show that observers use both local and global orientation information to make crowded target identification responses. However, to our knowledge, quantitative comparisons of individual response patterns from both visual crowding and ensemble perception tasks is a novel concept.

We apply our response models to the psychophysics data described in Chapter 3, in which individual participants completed both a target identification crowding task and a mean orientation Ensemble task. Both tasks used similar stimuli that consisted of a central Gabor surrounded by a ring of nine evenly-spaced Gabors. We fit a number of models of varying complexity to data from the the two types of tasks. Each model included a combination of two factors: spatial weighting and orientation similarity. Spatial weighting describes how much each of the oriented Gabors contributes to the response patterns, based on its location within the stimulus array. Orientation similarity describes how much each Gabor contributes, based on how similar its orientation is to that of one of the response categories.

For both tasks, we found that a spatially-weighted linear or multiplicative combination of the individual orientations considered independently better accounted for observers' response patterns than a spatial-weighted mean of the orientations. Furthermore, by analyzing the fitted model parameter values, we found that spatial-weighting strategies were correlated across observers between the two tasks, even though task performance was not correlated for the two tasks.

4.2   General Methods

Psychophysics Experiments

We modeled pre-existing psychophysics data described in Chapter 3. In that study, a set of nineteen
subjects first performed an orientation discrimination crowding task (Target task) followed by a
mean orientation discrimination ensemble perception task (Ensemble task) with the same stimulus
parameters. Stimuli consisted of a central Gabor surrounded by a circle of nine uniformly-spaced
Gabors. Threshold Gabor size and center-to-surround spacing were customized for each subject.
Additionally, a separate set of twenty subjects completed a similar peripheral ensemble perception
task on a 4x4 grid of uniformly-spaced Gabors (Uniform task). The stimuli for both experiments
were centered at 10°of visual angle away from fixation, either on the left or right side of the screen,
and appeared for 133 ms. Each experiment had a total of 960 trials per subject.

Target task   Participants were instructed to report the orientation of a single target Gabor (central
Gabor; 45°or 135°) within the ensemble using a key press (Figure 4.1, top left panel). The mean of
the orientations in the ensemble was balanced across the trials to either be congruent (mean orien-
tation was the same as the target orientation), neutral (mean orientation was always 90°vertical), or
incongruent (mean orienation was orthogonal to the target orientation) with respect to the target's
orientation. Subjects were not explicitly made aware of these target/mean orientation relationships.

Ensemble task   The same participants came back for another experimental session on a different
day and were instructed to report the mean orientation of the ensemble (central and surrounding
Gabors; 45°or 135°) using a key press (Figure 4.1, top middle panel). The orientation of the central
Gabor was balanced across the trials to either be congruent (central Gabor had the same orientation
as the mean orientation), neutral (the central Gabor was always 90°), or incongruent (the central
Gabor had an orthogonal orientation to the mean orientation) with respect to the mean orientation.
Participants were instructed to weight all Gabors equally when making their responses.

Uniform Spacing Task   Finally, a new set of participants performed a similar Ensemble task,
but this time on a 4x4 grid of uniformly-spaced Gabors (Figure 4.1, top right panel). They were
instructed to report the mean orientation of the ensemble (all sixteen; 45°or 135°) using a key
press. Additionally, one of the central Gabors' orientations (equally likely to be one of the four
Gabors within the center of the grid) was balanced across the trials to either be congruent, neutral,
or incongruent with respect to the mean orientation of the ensemble.

Dataset

We recorded each subject's two-alternative forced-choice responses for each trial of the three tasks
as well as the Gabor orientations that made up the stimulus arrays. Response models (described
below) were fit on a combination of trials from the three congruency conditions (congruent, neu-
tral, incogruent). Before fitting models to the response/orientation data, we also mirrored Gabor
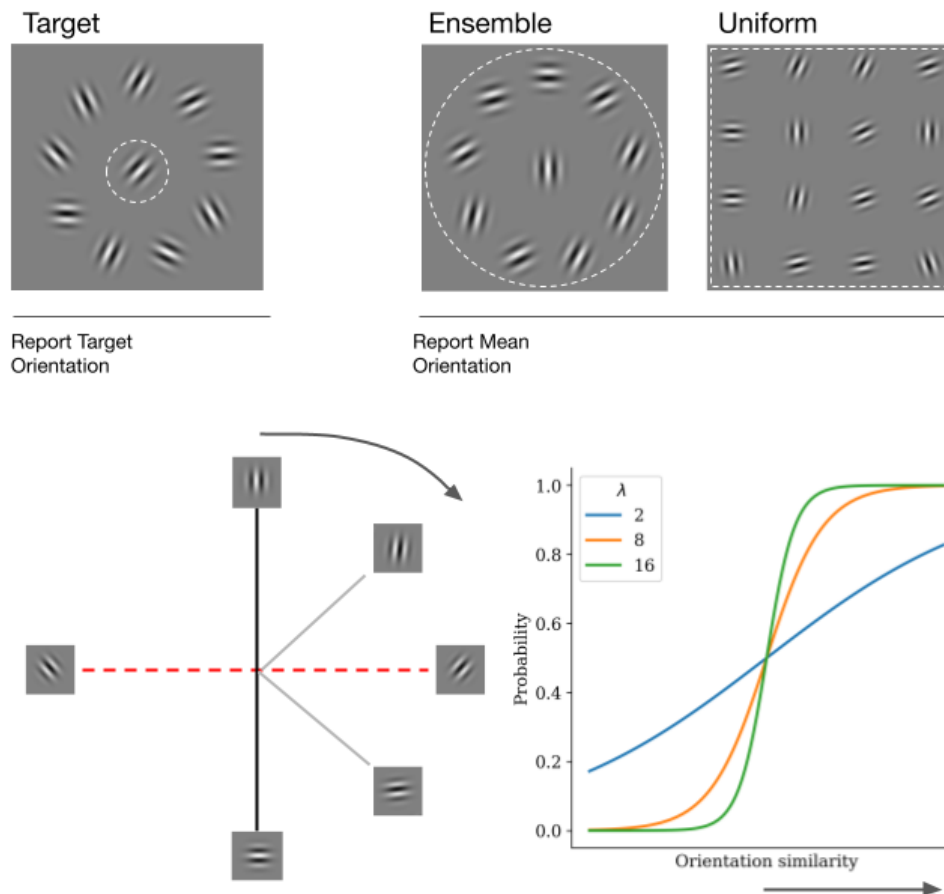
Figure 4.1: (*Top*) Stimuli for the three tasks. A set of nineteen subjects first performed an ori-
entation discrimination crowding task (Target task) followed by a mean orientation discrimina-
tion ensemble perception task (Ensemble task) with the same stimulus parameters. A separate
set of twenty subjects completed a similar peripheral ensemble perception task on a 4x4 grid of
uniformly-spaced Gabors (Uniform task). The stimuli for all three tasks were centered at $10°$ of
visual angle away from fixation, either on the left or right side of the screen, and appeared for 133
ms. (*Bottom*) Illustration of orientation similarity (equation 4.1). The diagram on the left shows
orientations over the circular distribution $(0, \pi)$ with the two response categories represented by
the red-dashed line. The function on the right shows the probability of observing a response as a
function of orientation similarity (see equation 4.1) for a few different values of scale factor (λ).

locations within the stimulus about a vertical plane (centered at fixation) such that trials in which
the stimuli appeared on the left side of the screen would have their visual field locations matched
with the right. This allowed us to fit a single set of spatial weights (see below) for each subject,
combining trials with stimuli to the left or right of the central fixation point.

4.3 Response Models

To compare the response patterns for visual crowding and ensemble perception tasks, we considered a number of response models. The goal of these models was to gain more information about how the orientations of individual Gabors within the stimulus arrays contributed to observers' orientation discrimination choices. We compared these models across the tasks to better understand if changing the task-relevance of certain orientations affected response patterns. We considered two main factors for each response model: spatial weighting and orientation similarity. Spatial weighting describes how much each of the oriented Gabors contributes to the response patterns, based on their locations within the stimulus array. Orientation similarity describes how much each Gabor contributes to the response patterns, based on how similar its orientation is to one of the two response categories.

Spatial Weights

One influence on response patterns is the weighting of certain locations more than others. We define a vector of weights w such that for each spatial position $i = 0, 1, \ldots, n$ of the Gabors in the stimulus array, $w_i \in [0, 1]$ and $\sum_{i=0}^{n} w_i = 1$. For an example, if $i = 0$ indexes the central Gabor in the center/surround stimulus array, and if a participant achieved 100% correct accuracy on the Target task, then the optimal weighting is $w_0 = 1$ (i.e., the target Gabor fully determined the responses). Similarly, if an observer achieved 100% correct accuracy on the ensemble or Uniform task, then the optimal weighting is $w_i = \frac{1}{n}$ for all $i = 0, 1, \ldots, n$ (i.e., all Gabors weighted equally).

Orientation Similarity

It is possible that one of the two response categories is more likely to be reported if the orientations in the stimulus array are more similar to it. We modeled the probability of observing a response r, given an orientation x in the stimulus array, based on the similarity between orientations and the two response categories. To achieve this, we define the following logistic sigmoid function:

$$S(x, r | \lambda) = \frac{1}{1 - e^{-\lambda\left(\frac{\pi}{4} - |x - r|\right)}} \tag{4.1}$$

where λ is a scale parameter that controls the slope of the function, and $|x - r|$ is the absolute distance between x and the response orientation r over the circular distribution $[0, \pi]$ (i.e., $0 \leq |x - r| \leq \frac{\pi}{2}$). Note that $S \to 0$ when $|x - r| \to \frac{\pi}{2}$, $S = 0.5$ when $|x - r| = \frac{\pi}{4}$ (corresponding to a vertical orientation for x), and $S \to 1$ when $|x - r| \to 0$. The probability of reporting the opposite orientation to r is $1 - S(x, r | \lambda)$.

The scale factor λ controls the sensitivity over which similar orientations contribute to the probability of observing one of the responses. If λ is small, then a given response is more likely to be observed when the orientation is very similar to the response ($|x - r|$ is small). If λ is large, S approaches a step function, which indicates that stimulus orientations only have to be slightly similar to one of the response categories in order to substantially influence the response.

The bottom of Figure 4.1 shows an illustration of orientation similarity and the logistic sigmoid function for few λ values.

Models

We used a number of response models to examine the patterns of subject responses for the three tasks. For each model, we define a probability of observing a response $\left(r = \frac{\pi}{4} \text{ or } \frac{3\pi}{4}\right)$, given the orientations X $(0 < X_i < \pi)$ in the stimulus array for each trial. The models predicted responses are based on a mixture of random guessing and the contributions of the orientations within stimulus arrays. We define each model below.

Random Guessing   For a given trial, it is possible a response was based on a random guess. Random guessing reflects that the subject has low (or no) confidence about the true orientation to be reported. This model predicts a uniform probability over both response types:

$$P(r|X) = \frac{1}{2}. \tag{4.2}$$

The random guessing model does not consider any of the orientations in its prediction. We consider this to be the baseline model that all other models in this paper are compared to. Furthermore, we included a random guessing component in all of the models in this paper, so all other models reduce to this model (i.e., if $p = 1$).

Correct Response: Target Orientation   Depending on the task, participants were given different instructions for making their responses. The goal of the Target task was to identify a single orientation within the set of orientations in the stimulus array. This model predicts that a single target orientation (X_0) within the stimulus array was correctly reported with probability $(1 - p)$ or based on a random uniform guess with probability p:

$$P(r|X) = p\left(\frac{1}{2}\right) + (1 - p)\delta\left[X_0, r\right], \quad p \in [0, 1]. \tag{4.3}$$

For the Target task, X_0 is the central Gabor's orientation. Note that participants in the Target task were instructed to only report on a single orientation, as opposed to the mean orientation of the ensemble. In the other two tasks, however, a single Gabor within the array was always congruent, neutral, or incongruent with respect to the mean orientation. Therefore, we still apply this model to the other tasks, and we choose X_0 to be the central Gabor for the Ensemble task and one of the four central grid Gabors for the Uniform task. The δ function captures the three possible orientations of X_0:

$$\delta\left[x, r\right] = \begin{cases} 1 & \text{if } |x - r| < \frac{\pi}{4} \\ \frac{1}{2} & \text{if } |x - r| = \frac{\pi}{4} \\ 0 & \text{if } |x - r| > \frac{\pi}{4} \end{cases} \tag{4.4}$$

Specifically, the function is equal to 1 if the target was the same as the response, $\frac{1}{2}$ if the target was neutral with respect to the response, or 0 if the target was the opposite of the response. The orientations of non-target Gabors within the stimulus array do not contribute to this model's predictions.

Correct Response: Arithmetic Mean   The orientations of Gabors on the stimulus arrays were pseudo-randomly chosen such that they would have a specific arithmetic mean. For the ensemble and Uniform tasks, participants reported the mean of these orientations. This model predicts that the arithmetic mean of orientations was correctly reported with probability $(1-p)$ or was a random uniform guess with probability p:

$$P(r|X) = p\left(\frac{1}{2}\right) + (1-p)\delta\left[\left(\frac{1}{n+1}\sum_{i=0}^{n}X_i\right), r\right]. \tag{4.5}$$

This model assumes that each orientation is weighted equally in contributing to the response. Note that participants were instructed to report the arithmetic mean of the orientations only in the ensemble and Uniform tasks and not in the the Target task. In the Target task, the mean of the orientations was balanced across trials to be congruent, neutral, or incongruent with respect to the target response categories. Therefore, we still apply this model to the Target task data, and as for the Correct Response: Target Orientation model, the δ function captures the three possible arithmetic means of the stimulus arrays.

Spatial Weighted Mean   Many studies have shown that observers extract summary statistics from an ensemble of similar looking objects, such as the mean and/or variance, even if the individual objects are difficult or impossible to identify (Haberman and Whitney 2012). Furthermore, there is evidence that the summary statistics are biased in the direction of the more salient objects of the ensemble (Iakovlev and Utochkin 2021). Therefore, we created a spatial weighted mean model which predicts that a response was based on a spatially-weighted mean of the orientations in the stimulus array with probability $(1 - p)$ or a random uniform guess with probability p:

$$P(r|X) = p\left(\frac{1}{2}\right) + (1-p)S\left(\sum_{i=0}^{n}w_iX_i, r|\lambda\right). \tag{4.6}$$

Like the Correct Response: Arithmetic Mean model, this model also assumes that the mean orientation drives response patterns. However, here we include non-uniform weighting which may lead to favoring certain Gabor orientations over others depending on the demands/goals of the task. Note that this model reduces to the Correct Response: Target Orientation model if λ is large and $w_0 = 1$ or to the Correct Response: Arithmetic Mean model if λ is large and $w_i = \frac{1}{n}$ for all $i = 0, 1, \ldots, n$.

Spatial Weighted Substitution   In visual crowding, substitution errors occur when subjects report the identity of one of the flanking stimuli instead of the target identity (Ester, Klee, and Awh

2014). Therefore, we created the Spatial Weighted Substitution model which predicts a response was based on the presented orientations considered independently with probability $(1 - p)$ or a random uniform guess with probability p: :

$$P(r|X) = p\left(\frac{1}{2}\right) + (1 - p) \sum_{i=0}^{n} w_i S(X_i, r|\lambda) \tag{4.7}$$

This model is similar to the Spatial Weighted Mean model, except here the weighted sum is computed after calculation of orientation similarity. Therefore, it describes the weighted mean of the orientation similarities, not the orientation similarity of the weighted mean. This model can determine the extent to which one or more of the Gabor orientations within the stimulus array drove response patterns. This model also reduces to both Correct Response models.

Spatial Weighted Multiplicative Combination   Finally, our last model examines if orientations are combined multiplicatively instead of linearly to produce observed response patterns. Specifically, this model predicts a response was based on a multiplicative combination of the presented orientations with probability $(1 - p)$ or a random uniform guess with probability p:

$$P(r|X) = p\left(\frac{1}{2}\right) + (1 - p)\frac{\prod_{i=0}^{n} S(X_i, r|\lambda)^{w_i}}{\prod_{i=0}^{n} S(X_i, r|\lambda)^{w_i} + \prod_{i=0}^{n} (1 - S(X_i, r|\lambda))^{w_i}} \tag{4.8}$$

This model assesses the extent to which multiple orientations together drove response patterns.

Model Fitting

Each of the models described above predicts a likelihood of the response patterns, given the data. For the three tasks, we find the parameter(s) that maximize the likelihood estimates for each subject separately. Specifically, for all trials t of a given subject/task, we find $[\hat{p}, \hat{s}, \hat{\lambda}]$ that maximize the following log likelihood estimate:

$$\arg\max_{\hat{p}, \hat{w}, \hat{\lambda}} \left[\sum_{t \in T} \log P\left(r^{(t)} | X^{(t)}\right) \right]. \tag{4.9}$$

We used PyTorch's Stochastic Gradient Descent for 5000 iterations and a learning rate of 0.001 to fit all models for each task and subject combination.

Statistical Analyses

We conducted planned comparisons to assess how well each model explained the observed response patterns. Specifically, we used paired two-tailed Student's t-tests to assess if the mean log-likelihood values across the trials/subjects were significantly different between the models. Additionally, we conducted two-tailed Student's t-tests to assess if specified mean model parameters were above baseline values. For correlation analyses, Pearson's r values were calculated and tested against the null hypothesis of a correlation coefficient value of zero.

4.4 Results
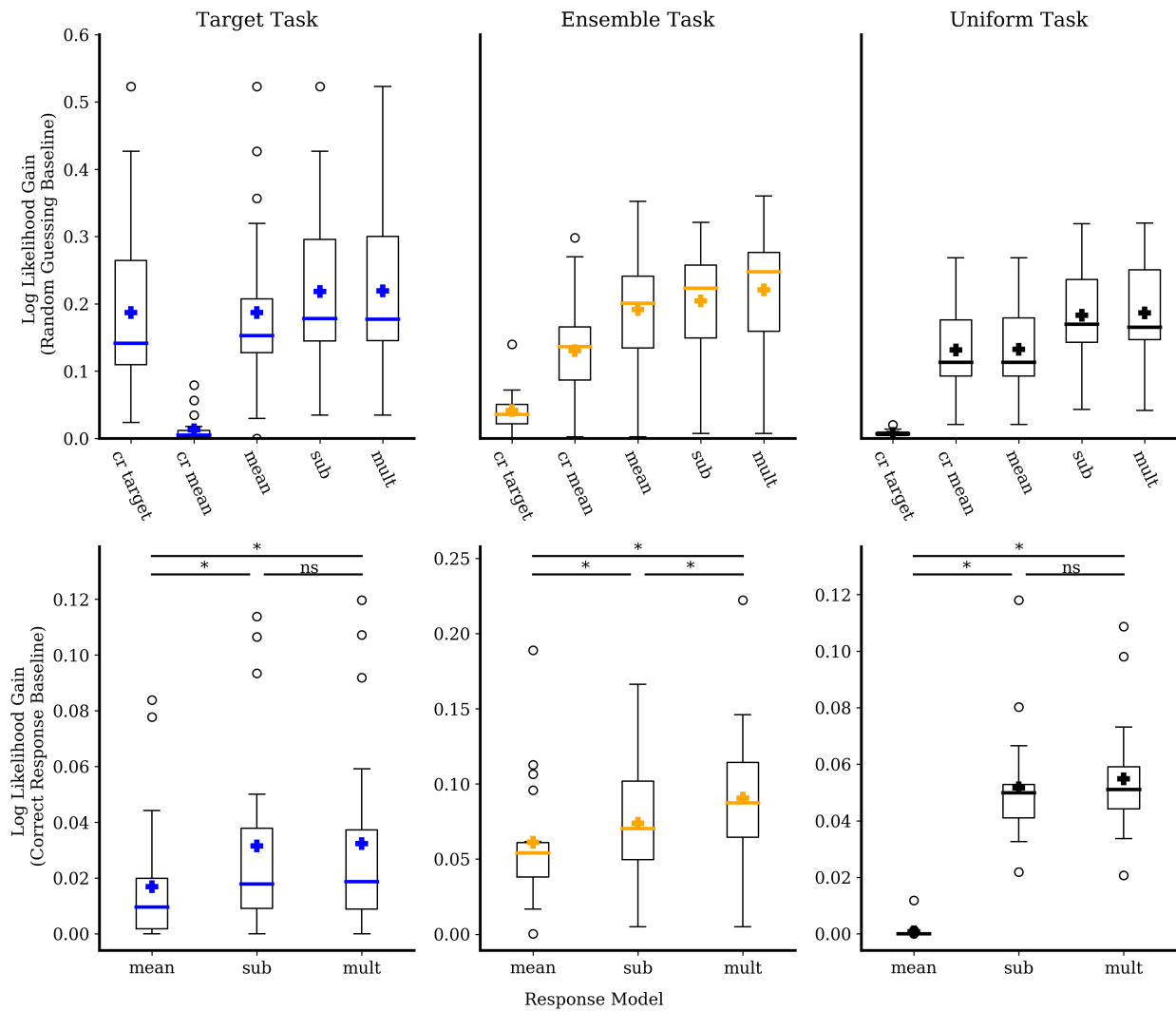
Figure 4.2: Log likelihood gain for response models, relative to the Random Guessing baseline (*Top*) and the Correct Response model baseline (*Bottom*). The Target task (left; blue) has the Correct Response: Target Orientation (CR Target) model as the baseline, while the ensemble (middle; orange) and uniform (right; black) tasks have the Correct Response: Arithmetic Mean (CR Mean) model as their baseline. Box and whisker plots are shown for the three tasks. Colored lines and crosses represent median and mean log likelihood gain, respectively. Box limits represent 1st and 3rd quartiles across observers. Whiskers extend beyond the box by 1.5 times the inter-quartile range. Small circles represent individual subject outliers beyond the whisker range. Asterisks represent significance level $\alpha < 0.05$ on paired t-tests.

Model comparison

We used pre-existing psychophysics data from three visual perceptual tasks. In the first task, subjects were instructed to report the orientation of a central Gabor that was crowded by nine additional Gabors (Target task). In the second task, the same subjects were instructed to report the mean orientation of the ensemble of Gabors with the same spatial arrangement of center/surround stimuli as in the Target task (Ensemble task). Finally, a separate set of subjects completed a similar mean orientation discrimination task but on a 4x4 grid of uniformly-spaced Gabors (Uniform task).

We fit response models to the data from the three tasks for each subject to compare the response patterns for target identification in visual crowding (Target task) and for ensemble perception (Ensemble and Uniform tasks). The fitted parameters maximize the likelihood estimates for each task. We averaged log likelihood estimates across the trials for each subject/task combination. The mean log likelihood across the trials provides a metric for how well each model predicts response patterns.

The top of Figure 4.2 shows box and whisker plots for the trial-averaged log likelihood gain for each model. The gain was computed as a difference in log likelihood estimates between each model and a baseline derived from the Random Guessing model (equation 4.2). The Random Guessing model has a constant log likelihood of -0.693. The colored lines and crosses within the boxes of figure 4.2 represent median and mean, respectively.

The bottom of Figure 4.2 shows box and whisker plots for the trial-averaged log likelihood gain relative to baselines derived from Correct Response models. Specifically, the Target task has the Correct Response: Target Orientation model estimates as its baseline, and the Ensemble and Uniform tasks have the Correct Response: Arithmetic Mean model estimates as their baselines.

For the Target task (Figure 4.2; bottom left; blue), the Substitution and Multiplicative models fit response patterns significantly better than the Mean model ($t_{18} \geq 4.99, p \leq 9.43e-05$). However, the Multiplicative model did not fit the response patterns significantly better than the Substitution model ($t_{18} = 1.22, p = 0.237$). For the Ensemble task (Figure 4.2; bottom middle; orange), The Multiplicative model fit response patterns significantly better than either the Substitution ($t_{18} = 4.40, p = 3.42e-04$) or Mean ($t_{18} = 6.20, p = 7.48e-06$) models. The Substitution model also fit response patterns significantly better than the Mean model ($t_{18} = 2.51, p = 0.022$). For the Uniform task (Figure 4.2; bottom right; black), the Substitution and Multiplicative models once again fit response patterns significantly better than the Mean model ($t_{18} \geq 12.6, p \leq 1.11e-10$). Unlike the results from the Ensemble task, the Multiplicative model did not fit response patterns significantly better than the Substitution model ($t_{18} = 1.83, p = 0.083$). Together, these results suggest that a Weighted Substitution model that considers the similarity of individual orientations of the Gabors accounts for subject responses better than a Weighted Mean model for visual crowding and ensemble perception tasks. Interestingly, the Multiplicative model did not provide a much better fit for responses than the Substitution model except for the Ensemble task.

Effect of task type on model parameters

Another way to compare visual crowding and ensemble perception is by examining the values of the fitted parameters across the tasks. For example, the distribution of spatial weights may differ depending on the focus/goal of the task. We first evaluated $1 - p$, which is the correct response rate and inversely related to task difficulty. Next we assessed the distributions of spatial weights, w, across the tasks/models. Lastly, we examined the orientation similarity scaling factor, λ.

Correct response rate The instructions for the Target task were to identify the orientation of the central Gabor (45°or 135°), while the instructions for the ensemble and Uniform tasks were to identify the arithmetic mean of all the orientations in the array of Gabors. In this section, we examined the rate of correct responses (based on these instructions) that could not be explained by random guessing. Specifically, we compared $(1 - p)$ for each task.

The left part of Figure 4.3 shows box and whisker plots for correct-response rates $(1-p)$ for the two Correct Response models (CR Target and CR Mean). The colored lines and crosses represent median and mean values, respectively. Generally, the mean correct-response rate that could not be accounted for by random guessing was around 0.5 for all three tasks. This result is consistent with the measured percent correct values across the three tasks (roughly 75%, Chapter 3), because the remaining 25% of correct responses could be explained by random guessing. This analysis indicates that overall task difficulty was similar across the three tasks.

Since the same subjects performed both the target and Ensemble tasks, the right part of Figure 4.3 shows individual subjects' correct response rates for the Target task against their results for the Ensemble task. $(1 - p)$ values were not correlated across the two tasks ($r = -0.122, p = 0.619$).

Central spatial weight and orientation similarity scale factor One limitation of a task performance metric like $(1 - p)$ is that it does not reveal possible differences in strategies that are utilized by subjects. In this section, we compared fitted parameter values for the three spatial-weighted models (Mean, Substitution, and Multiplicative combination) to examine the extent to which individual subjects employ similar strategies for the target and Ensemble tasks.

First, we considered the orientation similarity scaling factor, λ. λ controls the steepness of the slope of the orientation similarity logistic sigmoid function (equation 4.1). A large value of λ indicates that stimulus orientations only have to be slightly similar to one of the response categories in order to substantially influence the response, while a small value of λ indicates that the orientations have to be very similar to a response category in order to influence a response.

The top row of Figure 4.4 shows the similarity scaling factor box-and-whisker plots for each combination of task type and model. Overall, fitted λ values were larger for the Ensemble task than they were for the Target task. The bottom row of Figure 4.4 shows individual subjects' λ values for the Target task against their values for the Ensemble task. λ values between the two tasks were positively correlated across subjects for all three models, and they were significantly correlated for the Mean ($r = 0.605, p = 0.006$) and Multiplicative Combination ($r = 0.585, p = 0.009$) models.

Second, we considered the fitted central spatial weight, w_0, for the center/surround stimulus arrays. The top row of Figure 4.5 shows the central spatial weight box-and-whisker plots. Fitted
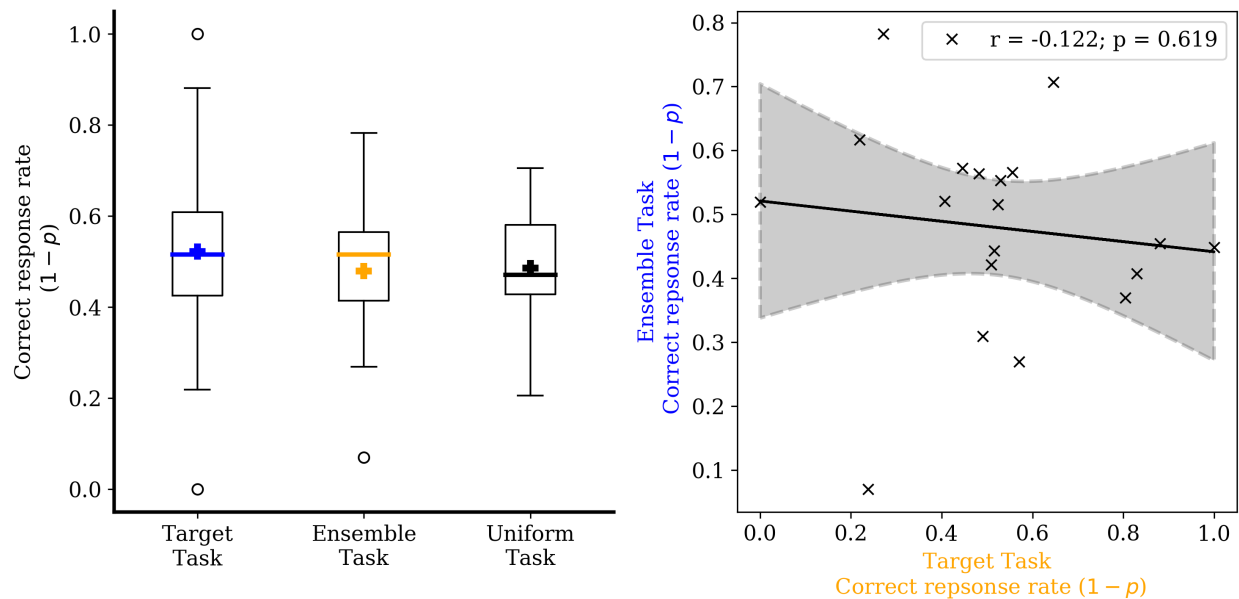
Figure 4.3: Task difficulty estimated by model parameter values. (*Left*) Model-fitted correct response rates (1-p) are displayed in box-and-whisker plots for the Correct Response models for each task type (CR Target for Target task; CR Mean for Ensemble and Uniform tasks). (*Right*) Individual subjects' correct response rate values for the Target task against values for the Ensemble task. The 'x's represent data from individual subjects. The solid and dashed lines represent regression line and 95% confidence intervals, respectively. Pearson's r and p-value for the correlation are presented.

w_0 values were larger for the Target task than they were for the Ensemble task. This result is expected, given that subjects reported the orientation of the central Gabor in the Target task. The bottom row of Figure 4.5 shows individual subjects' w_0 values for the two tasks plotted against each other. w_0 values between the two tasks were positively correlated across subjects for all three models, and they were significantly correlated for the Substitution ($r = 0.634, p = 0.004$) and Multiplicative Combination ($r = 0.538, p = 0.017$) models.

The consistent positive correlations for both fitted parameters (λ, w_0) across three spatial-weighted models suggests that subjects employed a similar strategy for both tasks. Specifically, subjects who weighted the central Gabor's orientation heavily in the Target task were likely to also weight it heavily in the Ensemble task. Furthermore, subjects with a high orientation similarity sensitivity in the Target task were likely to also have a higher sensitivity in the Ensemble task.

Other spatial weights In the previous section, we examined the central spatial weight and found that participants assigned more weight to the central Gabor for the Target task compared to the Ensemble task. In this section, we characterize the spatial weights associated with the surrounding Gabor orientations. For the Target task, the surrounding Gabors impede identification of the central

Figure 4.4: Effect of task type on similarity scaling factor (λ) for the center/surround stimuli. (*Top*) Box-and-whisker plots (properties described in Figure 4.2) for each task type and response model. (*Bottom*) Correlation of similarity scaling factor values between the two tasks across individual subjects. The 'x's represent individual subjects. The solid and dashed lines represent regression line and 95% confidence intervals, respectively. Pearson's r and p-value for the correlation are displayed for each plot.

Gabor's orientation, but for the Ensemble task, the surrounding Gabors are all equally relevant for estimating the mean orientation of the stimulus array.

The top row of Figure 4.6 shows radar plots of spatial weights for each of the surrounding Gabors for each model. The left and right sides of the plots represent the foveal and peripheral visual field locations, respectively. Overall, the spatial weights for the surrounding Gabors were larger in the Ensemble task than they were in the Target task, suggesting that subjects emphasized the surrounding Gabors more when computing the mean orientation (Ensemble task), compared to discriminating the central Gabor (Target task).

Interestingly, the distributions of the spatial weights in Figure 4.6 (top row) are anisotropic. In
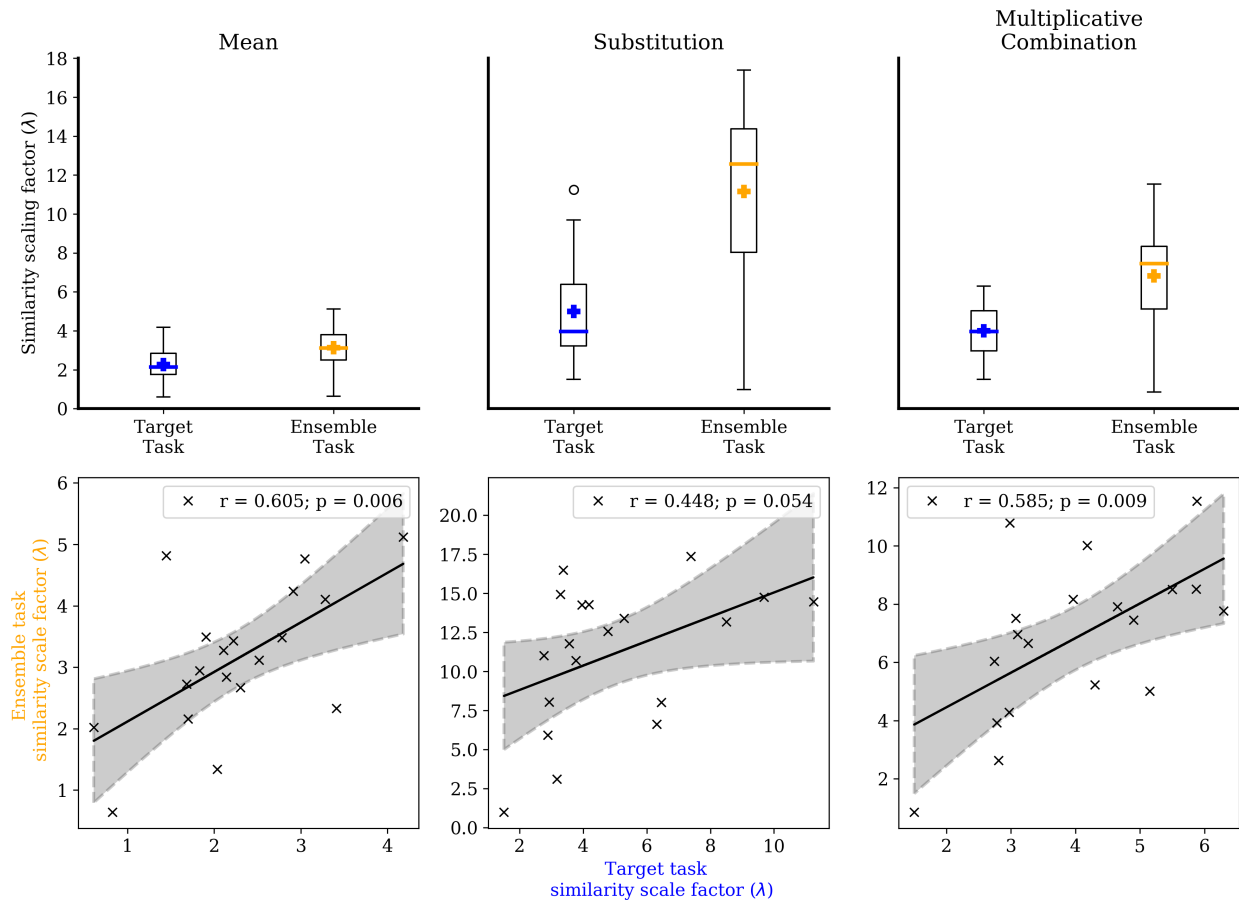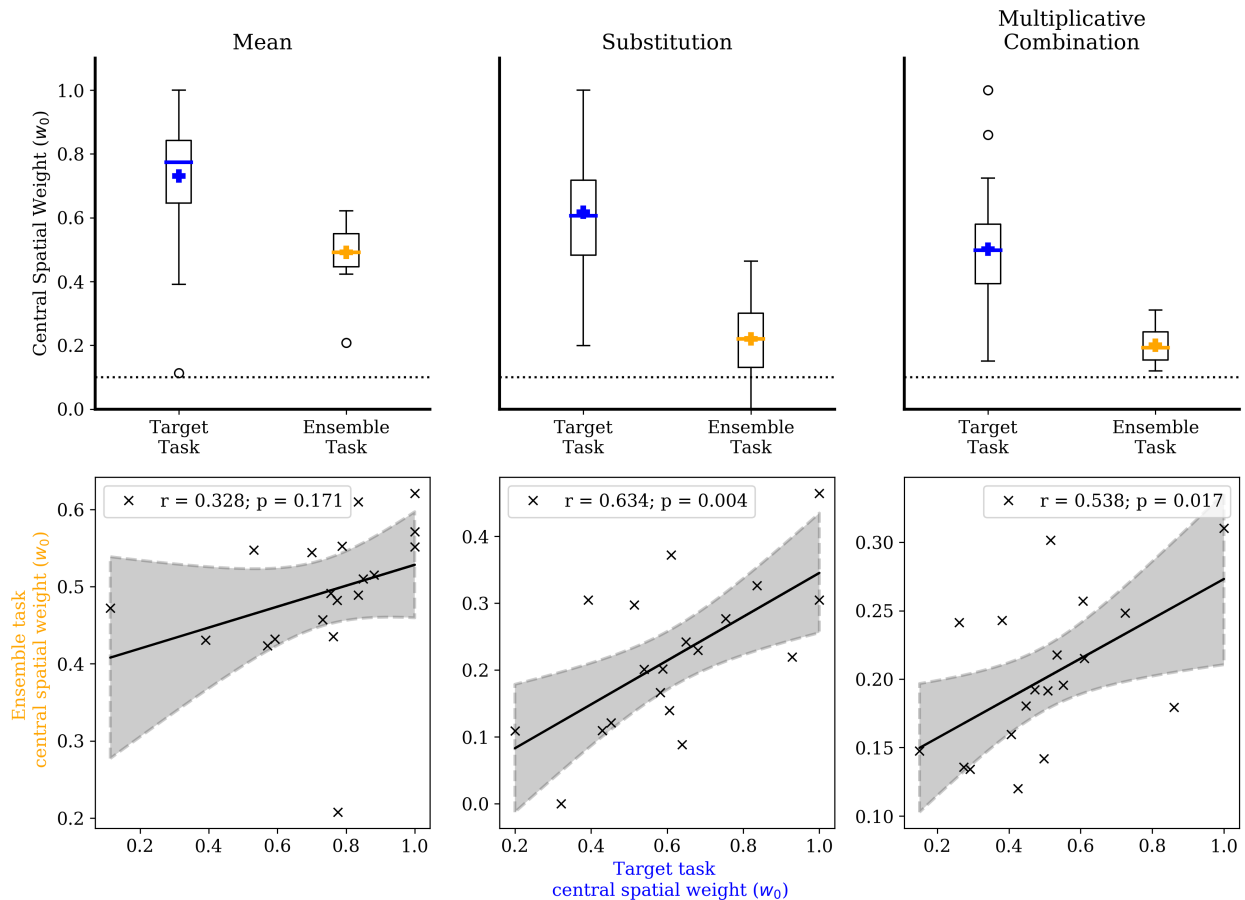
Figure 4.5: Effect of task type on central spatial weight (w_0) for the center/surround stimuli. (*Top*) Box-and-whisker plots (properties described in Figure 4.2) for each task type and response model. The dashed black line at 0.10 represents a hypothetical central spatial weight when all ten Gabors in the stimulus array are equally weighted. (*Bottom*) Correlation of central spatial weight between the two tasks across individual subjects. The 'x's represent individual subjects. The solid and dashed lines represent regression line and 95% confidence intervals, respectively. Pearson's r and p-value for the correlation are displayed for each plot.

humans, visual crowding is influenced by target/flanker configuration: a single inner flanker on the foveal side of the target causes less crowding than the same outer flanker on the peripheral side of the target (Banks, Bachrach, and Larson 1977). Additionally, flankers on either side of the target along a radial axis emanating from the fovea cause more crowding than flankers along a tangential axis that is perpendicular to the radial axis (Toet and Levi 1992; Chen et al. 2014).

To examine if our observed distributions of spatial weights replicate these known crowding phenomena, we fit ellipses to the distribution of spatial weights for each subject using linear regression. We defined the distance of the center of the ellipse (x-dimension) from the origin as the

metric for the inner/outer effect and the ratio between the length of a horizontal line and vertical line that pass through the center of the ellipse as the metric for the radial/tangential effect.

The middle row of Figure 4.6 shows the inner/outer crowding effect with box-and-whisker plots for each task and model combination. Surprisingly, none of the models for the Target task had an spatial weight ellipse center value that was significantly different than the origin ($t_{18} \leq 1.19, p \geq 0.254$). The Mean and the Multiplicative Combination models for the Ensemble task, however, did show a significant outer bias (positive x-center values) for the surrounding spatial weights ($t_{18} \geq 3.33, p \leq .004$), but the Substitution model did not ($t_{18} = 2.05, p = 0.055$).

The bottom row of Figure 4.6 shows the radial/tangential crowding effect with box-and-whisker plots for each task/model combination. For the Target task, all three models showed a significant radial bias (ratio greater than 1) ($t_{18} \geq 2.51, p \leq 0.026$). For the Ensemble task, only the substitution model showed a significant radial bias ($t_{18} = 2.38, p = 0.028$), and the other two models did not show a significant effect ($t_{18} \leq 0.734, p \geq 0.472$).

To summarize, most of the Target task spatial weights exhibited significant radial/tangential biases but not inner/outer biases. For the Ensemble task, two of three models showed a significant inner/outer bias, and one showed a significant radial/tangential bias. We note that the psychophysics experiments that are typically used to measure configuration effects on crowding present different spatial arrangement of flankers on different trials. In our experiment, we presented all of the flankers at once, and this may have influenced possible configuration effects.

Spatial attention cueing   The psychophysics experiments from Chapter 3 also included a 40 ms pre-cue immediately before stimulus presentation. The spatial extent of the cue corresponded to either the size of a single Gabor (small cue) or the entire stimulus array (large cue). The cue appeared 50% of the time in the same location as the center of the stimulus (valid) and 50% of the time in the matched location on the opposite side of the screen (invalid). In this section, we examined the effect of stimulus-driven attention on the fitted model parameter values for each task. We present results only from the Substitution model, as it showed the greatest correlation in central spatial weight between the two tasks (Figure 4.5; bottom).

The top of Figure 4.7 shows the magnitudes of the cueing effect (values for valid trial - invalid trials) with box-and-whisker plots for correct response rate (left) and central spatial weight (right) for each task/cue size combination. Overall, the magnitude of the cueing effect on correct response rate ($1 - p$) was significantly greater than zero only for the small cue in the Target task ($t_{18} = 3.71, p = 0.002$). Similarly, the magnitude of the cueing effect on central spatial weight (w_0) values was significantly greater than zero only for the small cue for the Target task ($t_{18} = 2.35, p = 0.03$). Additionally, cueing-effect magnitudes for the small cue were not significantly correlated between the two tasks across subjects ($r = 0.436, p = 0.062$). These results are consistent with and further support the analysis from Chapter 3.
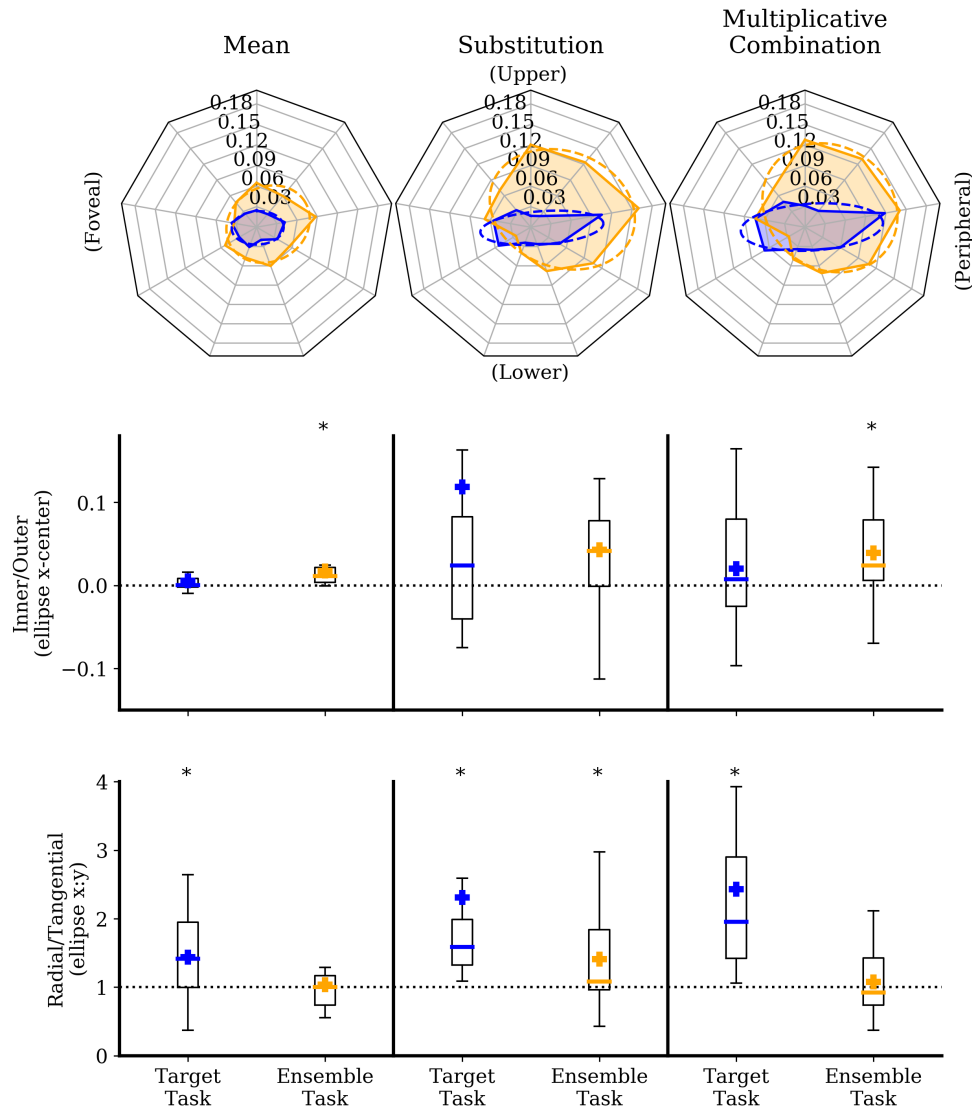
Figure 4.6: Other spatial weights of the center/surround stimuli. (*Top*) Radar plots of the spatial weights for each of the surrounding Gabors. The left and right sides of the plots represent foveal and peripheral Gabors, respectively. We used linear regression to fit each subject's spatial weights with an ellipse. (*Middle*) Inner/Outer crowding effect box-and-whisker plots for each task/model, measured as the distance between the center of the fitted ellipse and the origin in the x-dimension. (*Bottom*) Radial/Tangential crowding effect box-and-whisker plots for each task/model, measured as the ratio between the lengths of the horizontal and vertical lines that pass through the center of the ellipse. Asterisks represent significance level $\alpha < 0.05$ from two-tailed t-tests. Box and whisker properties are the same as described in Figure 4.2.

4.5 Discussion

We modeled the response patterns of pre-existing data from individual subjects who completed both a visual crowding and an ensemble perception task with almost identical stimulus parameters. The goal of the modeling was to better understand how the orientations within the arrays of Gabors contributed to how observers made their orientation discrimination choices and how this was modulated by task relevance. The set of models of different complexities combined spatial weighting and orientation similarity in different ways.

First we found that, for both the Target and Ensemble task, a spatial-weighted linear (equation 4.7) or multiplicative combination (equation 4.8) of the similarities of individual orientations to the two response categories better accounted for observer response patterns than the similarity of a spatial-weighted mean of the orientations to these response categories (equation 4.6) (Figure 4.2). Second, we found that model parameters, such as the sensitivity of the similarity of responses to the orientations (λ in equation 4.1), or the central spatial weight of the center/surround stimuli, were significantly positively correlated for the two tasks across observers for most models (bottom of Figures 4.4, 4.5), even though performance on the two tasks was not correlated (Figure 4.3). Finally, we showed that the relationships of spatial weights for the surrounding Gabor orientations supported radial/tangential configuration effects of visual crowding. There was also some evidence for inner/outer effects in the Ensemble task (Figure 4.6).

Response models of visual crowding   Other studies have also modeled the response patterns of perceptual judgements of visual crowding stimuli. Our modeling approach is similar to that of (Hanus and Vul 2013), who quantified the error distributions of crowded letter stimuli. They investigated different combinations of spatial weighting and letter-to-letter similarity/confusability. They observed that a spatial-weighted multiplicative combination of letter confusability did not explain response patterns better than a spatial-weighted linear combination. Our study replicates these findings for crowded oriented Gabors (Figure 4.2; left).

Similar approaches have also been used to compare pooling and substitution theories of visual crowding. For example, Ester, Klee, and Awh 2014 found that errors in a target orientation identification crowding task were better explained by probabilistic substitutions of target/flanker features than by a simple averaging of the features. In a follow-up study, Ester, Zilber, and Serences 2015 found that this result was obtained even when target/distractor features were very dissimilar. These studies are in contrast to the findings of Parkes et al. 2001, who reported that the response patterns were consistent with a compulsory averaging of the orientations of Gabors. We expand on previous work by including a spatial-weighting function in which certain orientations could be weighted more heavily than the others. We found that response patterns were better explained by a combination of the orientations considered independently (Substitution model) than by a spatial-weighted average of the orientations (Mean model). However, we acknowledge that our response data was limited to only two response categories (as opposed to a continuous adjustment method).

Although a spatial-weighted mean of orientations was not the best explanation of our crowding responses, we do not rule out more complicated averaging models (Balas, Nakano, and Ruth Rosenholtz 2009; Rosenholtz 2016) and/or image-computable models that consider non-uniform

spatially-weighted integration regions (Freeman and Simoncelli 2011; Theiss, Bowen, and Silver 2021). Furthermore, crowding has been shown to affect multiple levels of visual processing (Manassi and Whitney 2018). Perhaps models that consider multiple levels of processing and/or grouping/holistic effects could provide a better account of the response patterns of visual crowding (Jimenez, Kimchi, and Yashar 2022).

Visual crowding versus ensemble perception   While many studies have explored visual crowding and ensemble perception separately, only a handful have used the same observers for both tasks. Lin, Gong, and Li 2022 showed that attention modulated crowding more than ensemble perception in the same observers. We also found that a small exogenous attention cue had a significant effect on the correct response rate and the central spatial weight of visual crowding, but not on ensemble perception (Figure 4.7; top) (see Chapter 3 for more detail).

Some studies have applied other dual tasks to ensemble perception (i.e. instead of crowding). For example, Haberman, Brady, and Alvarez 2015 used a dual task where observers either reported the identity of one element or the mean (orientation or face) of an ensemble. They found that individuals' performance on a mean Ensemble task was related to their performance on identification of a single element within the ensemble, for both orientation and face stimuli. In their study, individual elements were not crowded. In our study, we did not find a significant correlation between overall correct response rate between the two tasks. However, for most models, we did find a significant correlation between two tasks for both the weight of the central Gabor's orientation and the orientation similarity factor.

Taken together, these results suggest that observers utilized similar spatial-weighting strategies for orientation discrimination of a target embedded in a crowded array and extraction of mean orientation from the array, even though this similarity was not evident as correlations in performance for the two tasks. We argue that by quantitatively considering how each element of the stimulus, as well as task relevance of the individual elements, contributed to the response, we obtained insights not available from analysis of only percent correct data. We hope to inspire future studies that apply similar approaches to visual crowding and/or ensemble perception.

Possible future directions include the study of different feature dimensions, such as shapes, faces, emotions, etc. A similar modeling approach could illuminate how ensemble perception and visual crowding compare at multiple levels of visual processing. Furthermore, including image-computable feature representations of the stimuli, such as those extracted from wavelet decomposition (Balas, Nakano, and Ruth Rosenholtz 2009; Freeman and Simoncelli 2011) or a neural network (Chaney, Fischer, and Whitney 2014; Theiss, Bowen, and Silver 2021), instead of just the raw orientation values, may provide an even better account for the response patterns for crowded peripheral visual tasks.
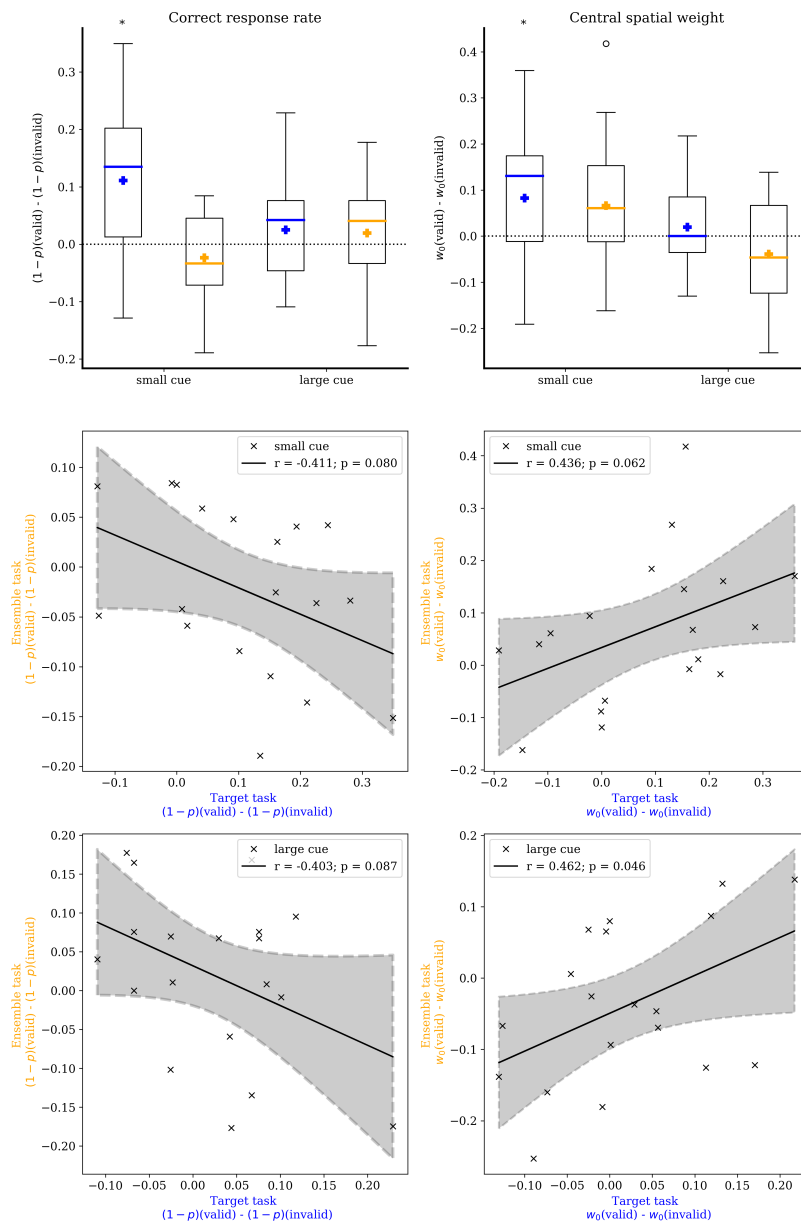
Figure 4.7: Effects of exogenous spatial attention on model parameter values. (*Top*) Attention cueing-effect magnitudes (values for valid trial - invalid trials) displayed in box-and-whisker plots for correct response rate (left) and central spatial weight (right) for each task (target - blue; ensemble - orange) and cue size (small or large) combination. Asterisks represent significance level $\alpha < 0.05$ from paired t-tests. Box and whisker properties are the same as described in Figure 4.2. (*Bottom*) Correlation of cueing-effect magnitude values between the two tasks. The 'x's represent individual subjects. The solid and dashed lines represent regression line and 95% confidence intervals, respectively. Pearson's r and p-values for the correlation are presented.

Chapter 5

Conclusion

In this dissertation, I described three studies that investigated the effects of spatial attention on crowded peripheral processing of oriented Gabors. In the first study, we used an anti-cueing paradigm to test the effects of involuntary and voluntary attention on visual crowding. We found that involuntary attention to the target location decreased response time and lowered the critical spacing of crowding. Interestingly, in the same observers, voluntary attention directed to the target location decreased response time but did not affect critical spacing.

In the second study, we expanded on the first study by investigating the effect of precision of a peripheral cue on visual crowding. We found that only the small cue that encompassed the target location decreased the effect of crowding while the large cue that encompassed both the target and flankers did not. In the same study, and with the same group of observers, we also investigated the effect of cue size on a peripheral ensemble perception task with stimulus parameters that matched the visual crowding task. We found that the precision of stimulus-driven attention moderately modulated ensemble perception but that this modulation occurred only if the cued Gabor was more salient than the non-cued Gabors within the ensemble.

In the third and final study, we expanded on the analysis of the second study by fitting response models of different complexity to the psychophysical data. The goal of the modeling approach was to investigate how observers utilized all of the orientations of stimulus array to make their responses, to characterize the influences of task relevance, and to test if common strategies were shared between the two tasks. We found that response patterns were better explained by a combination of the orientations considered independently (our Substitution model) than by a spatial-weighted average of the orientations (our Mean model). Furthermore, by analyzing the fitted model parameter values, we found that spatial-weighting strategies were correlated between the two tasks, even though task performance was not correlated for the two tasks.

The main impacts of this body of work are three-fold: 1) A better understanding of the interactions between spatial attention and crowded peripheral processing can inform treatments and therapies for certain visual impairments such as central vision loss in macular degeneration or stereo vision loss in amblyopia. 2) This research could inform safety or usability concerns in human-technology interactions, such as dashboards/displays for vehicles or page layouts for websites. 3) Debunking obscure scenes from popular 1990's movies.

Future directions include incorporation of physiological data during performance of visual tasks. We did not observe strong correlations between attention effects for either involuntary and voluntary attention or visual crowding and ensemble perception. It could be informative to not only correlate performance or RT for these two phenomena but also these effects on physiological measurements, such as response amplitudes and RF sizes.

Bibliography

Albonico, Andrea et al. (2018). "Focusing and orienting spatial attention differently modulate crowding in central and peripheral vision". In: *Journal of vision* 18.3, pp. 4.1–4.17.

Alvarez, George A (2011). "Representing multiple objects as an ensemble enhances visual cognition". In: *Trends in cognitive sciences* 15.3, pp. 122–131.

Anton-Erxleben, Katharina and Marisa Carrasco (2013). "Attentional enhancement of spatial resolution: Linking behavioural and neurophysiological evidence". In: *Nature Reviews Neuroscience* 14.3, p. 188.

Anton-Erxleben, Katharina, Valeska M Stephan, and Stefan Treue (2009). "Attention reshapes center-surround receptive field structure in macaque cortical area MT". In: *Cerebral Cortex* 19.10, pp. 2466–2478.

Balas, Benjamin, Lisa Nakano, and Ruth Rosenholtz (2009). "A summary-statistic representation in peripheral vision explains visual crowding". In: *Journal of vision* 9.12, pp. 13.1–13.18.

Banks, William P, Kenneth M Bachrach, and Douglas W Larson (1977). "The asymmetry of lateral interference in visual letter identification". In: *Perception & Psychophysics* 22.3, pp. 232–240.

Barbot, Antoine and Marisa Carrasco (2017). "Attention modifies spatial resolution according to task demands". In: *Psychological science* 28.3, pp. 285–296.

Barbot, Antoine, Michael S Landy, and Marisa Carrasco (2012). "Differential effects of exogenous and endogenous attention on second-order texture contrast sensitivity". In: *Journal of Vision* 12.8, pp. 6.1–6.25.

Baruch, Orit and Yaffa Yeshurun (2014). "Attentional attraction of receptive fields can explain spatial and temporal effects of attention". In: *Visual Cognition* 22.5, pp. 704–736.

Berger, Andrea, Avishai Henik, and Robert Rafal (2005). "Competition between endogenous and exogenous orienting of visual attention." In: *Journal of Experimental Psychology: General* 134.2, pp. 207–221.

Bouma, Herman (1970). "Interaction effects in parafoveal letter recognition". In: *Nature* 226.5241, pp. 177–178.

Brainard, David H (1997). "The psychophysics toolbox". In: *Spatial vision* 10.4, pp. 433–436.

Burnett, Katherine E, Giovanni d'Avossa, and Ayelet Sapir (2013). "Matching cue size and task properties in exogenous attention". In: *Quarterly Journal of Experimental Psychology* 66.12, pp. 2363–2375.

Carrasco, Marisa (2011). "Visual attention: The past 25 years". In: *Vision research* 51.13, pp. 1484–1525.

Chaney, Wesley, Jason Fischer, and David Whitney (2014). "The hierarchical sparse selection model of visual crowding". In: *Frontiers in integrative neuroscience* 8, pp. 73.1–73.11.

Chen, Juan et al. (2014). "Attention-dependent early cortical suppression contributes to crowding". In: *Journal of Neuroscience* 34.32, pp. 10465–10474.

Chetverikov, Andrey, Gianluca Campana, and Arni Kristjansson (2017). "Rapid learning of visual ensembles". In: *Journal of Vision* 17.2, pp. 21.1–21.15.

Chetverikov, Andrey, Gianluca Campana, and Árni Kristjánsson (2016). "Building ensemble representations: How the shape of preceding distractor distributions affects visual search". In: *Cognition* 153, pp. 196–210.

Coates, Daniel R, Jean-Baptiste Bernard, and Susana TL Chung (2019). "Feature contingencies when reading letter strings". In: *Vision research* 156, pp. 84–95.

Cornelissen, Frans W, Enno M Peters, and John Palmer (2002). "The Eyelink Toolbox: eye tracking with MATLAB and the Psychophysics Toolbox". In: *Behavior Research Methods, Instruments, & Computers* 34.4, pp. 613–617.

Dakin, Steven C et al. (2010). "Probabilistic, positional averaging predicts object-level crowding effects with letter-like stimuli". In: *Journal of Vision* 10.10, pp. 14.1–14.16.

De Fockert, Jan W and Alexander P Marchant (2008). "Attention modulates set representation by statistical properties". In: *Perception & Psychophysics* 70.5, pp. 789–794.

Desimone, Robert and John Duncan (1995). "Neural mechanisms of selective visual attention". In: *Annual review of neuroscience* 18.1, pp. 193–222.

Dugué, Laura et al. (2020). "Differential impact of endogenous and exogenous attention on activity in human visual cortex". In: *Scientific reports* 10.1, pp. 1–16.

Ester, Edward F, Daniel Klee, and Edward Awh (2014). "Visual crowding cannot be wholly explained by feature pooling." In: *Journal of Experimental Psychology: Human Perception and Performance* 40.3, pp. 1022–1033.

Ester, Edward F, Emma Zilber, and John T Serences (2015). "Substitution and pooling in visual crowding induced by similar and dissimilar distractors". In: *Journal of vision* 15.1, pp. 4.1–4.12.

Fang, Fang and Sheng He (2008). "Crowding alters the spatial distribution of attention modulation in human primary visual cortex". In: *Journal of Vision* 8.9, pp. 6.1–6.9.

Farzin, Faraz, Susan M Rivera, and David Whitney (2009). "Holistic crowding of Mooney faces". In: *Journal of vision* 9.6, pp. 18.1–18.15.

Felisberti, Fatima M, Joshua A Solomon, and Michael J Morgan (2005). "The role of target salience in crowding". In: *Perception* 34.7, pp. 823–833.

Fernández, Antonio, Sara Okun, and Marisa Carrasco (2021). "Differential effects of endogenous and exogenous attention on sensory tuning". In: *BioRxiv*.

Fortenbaugh, Francesca C, William Prinzmetal, and Lynn C Robertson (2011). "Rapid changes in visual-spatial attention distort object shape". In: *Psychonomic Bulletin & Review* 18.2, pp. 287–294.

Freeman, Jeremy and Eero P Simoncelli (2011). "Metamers of the ventral stream". In: *Nature neuroscience* 14.9, pp. 1195–1201.

Fukuda, Keisuke and Edward K Vogel (2009). "Human variation in overriding attentional capture". In: *Journal of Neuroscience* 29.27, pp. 8726–8733.

Gattass, Ricardo, CG Gross, and JH Sandell (1981). "Visual topography of V2 in the macaque". In: *Journal of Comparative Neurology* 201.4, pp. 519–539.

Gattass, Ricardo, AP Sousa, and CG Gross (1988). "Visuotopic organization and extent of V3 and V4 of the macaque". In: *Journal of neuroscience* 8.6, pp. 1831–1845.

Gratton, Caterina et al. (2017). "Cholinergic, but not dopaminergic or noradrenergic, enhancement sharpens visual spatial perception in humans". In: *Journal of Neuroscience* 37.16, pp. 4405–4415.

Greenwood, John A, Peter J Bex, and Steven C Dakin (2010). "Crowding changes appearance". In: *Current Biology* 20.6, pp. 496–501.

Haberman, Jason, Timothy F Brady, and George A Alvarez (2015). "Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation." In: *Journal of Experimental Psychology: General* 144.2, p. 432.

Haberman, Jason and David Whitney (2012). "Ensemble perception: Summarizing the scene and broadening the limits of visual processing". In: *From perception to consciousness: Searching with Anne Treisman*, pp. 339–349.

Hanus, Deborah and Edward Vul (2013). "Quantifying error distributions in crowding". In: *Journal of Vision* 13.4, pp. 17.1–17.27.

He, Dongjun, Yingying Wang, and Fang Fang (2019). "The critical role of V2 population receptive fields in visual orientation crowding". In: *Current Biology* 29.13, pp. 2229–2236.

Herzog, Michael H et al. (2015). "Crowding, grouping, and object recognition: A matter of appearance". In: *Journal of vision* 15.6, pp. 5.1–5.18.

Iakovlev, Aleksei U and Igor S Utochkin (2021). "Roles of saliency and set size in ensemble averaging". In: *Attention, Perception, & Psychophysics* 83.3, pp. 1251–1262.

Ji, Luyan, Valentina Rossi, and Gilles Pourtois (2018). "Mean emotion from multiple facial expressions can be extracted with limited attention: Evidence from visual ERPs". In: *Neuropsychologia* 111, pp. 92–102.

Jigo, Michael and Marisa Carrasco (2020). "Differential impact of exogenous and endogenous attention on the contrast sensitivity function across eccentricity". In: *Journal of Vision* 20.6, pp. 11.1–11.25.

Jigo, Michael, David J Heeger, and Marisa Carrasco (2021). "An image-computable model of how endogenous and exogenous attention differentially alter visual perception". In: *Proceedings of the National Academy of Sciences* 118.33.

Jimenez, Mikel, Ruth Kimchi, and Amit Yashar (2022). "Mixture-modeling approach reveals global and local processes in visual crowding". In: *Scientific Reports* 12.1, pp. 1–9.

Keshvari, Shaiyan and Ruth Rosenholtz (2016). "Pooling of continuous features provides a unifying account of crowding". In: *Journal of Vision* 16.3, pp. 39.1–39.15.

Kirsch, Wladimir, Bastian Heitling, and Wilfried Kunde (2018). "Changes in the size of attentional focus modulate the apparent object's size". In: *Vision Research* 153, pp. 82–90.

Klein, Barrie P, Ben M Harvey, and Serge O Dumoulin (2014). "Attraction of position preference by spatial attention throughout human visual cortex". In: *Neuron* 84.1, pp. 227–237.

Kleiner, Mario, David Brainard, and Denis Pelli (2007). "What's new in Psychtoolbox-3?" In.

Kosovicheva, Anna A et al. (2012). "Cholinergic enhancement reduces orientation-specific surround suppression but not visual crowding". In: *Frontiers in behavioral neuroscience* 6, p. 61.

Levi, Dennis M (2008). "Crowding—An essential bottleneck for object recognition: A mini-review". In: *Vision research* 48.5, pp. 635–654.

Lin, Zhen, Mingliang Gong, and Xiang Li (2022). "On the relation between crowding and ensemble perception: Examining the role of attention". In: *PsyCh Journal*.

Manassi, Mauro, Bilge Sayim, and Michael H Herzog (2012). "Grouping, pooling, and when bigger is better in visual crowding". In: *Journal of Vision* 12.10, pp. 13.1–13.14.

Manassi, Mauro and David Whitney (2018). "Multi-level crowding and the paradox of object recognition in clutter". In: *Current Biology* 28.3, R127–R133.

Mareschal, Isabelle, Michael J Morgan, and Joshua A Solomon (2010). "Attentional modulation of crowding". In: *Vision Research* 50.8, pp. 805–809.

McAdams, Carrie J and John HR Maunsell (1999). "Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4". In: *Journal of Neuroscience* 19.1, pp. 431–441.

Parkes, Laura et al. (2001). "Compulsory averaging of crowded orientation signals in human vision". In: *Nature neuroscience* 4.7, pp. 739–744.

Pelli, Denis G (1997). "The VideoToolbox software for visual psychophysics: Transforming numbers into movies". In: *Spatial vision* 10, pp. 437–442.

Petrov, Yury and Olga Meleshkevich (2011). "Asymmetries and idiosyncratic hot spots in crowding". In: *Vision research* 51.10, pp. 1117–1123.

Posner, Michael I (1980). "Orienting of attention". In: *Quarterly journal of experimental psychology* 32.1, pp. 3–25.

Posner, Michael I, Yoram Cohen, and Robert D Rafal (1982). "Neural systems control of spatial orienting". In: *Philosophical Transactions of the Royal Society of London. B, Biological Sciences* 298.1089, pp. 187–198.

Rashal, Einat and Yaffa Yeshurun (2014). "Contrast dissimilarity effects on crowding are not simply another case of target saliency". In: *Journal of vision* 14.6, pp. 9.1–9.12.

Reuther, Josephine and Ramakrishna Chakravarthi (2014). "Categorical membership modulates crowding: Evidence from characters". In: *Journal of Vision* 14.6, pp. 5.1–5.13.

Roberts, MJ et al. (2005). "Acetylcholine dynamically controls spatial integration in marmoset primary visual cortex". In: *Journal of neurophysiology* 93.4, pp. 2062–2072.

Rokem, Ariel et al. (2010). "Cholinergic enhancement increases the effects of voluntary attention but does not affect involuntary attention". In: *Neuropsychopharmacology* 35.13, pp. 2538–2544.

Rosenholtz, R (2016). "Capabilities and Limitations of Peripheral Vision." In: *Annual review of vision science* 2, pp. 437–457.

Rosenholtz, Ruth (2020). "Demystifying visual awareness: Peripheral encoding plus limited decision complexity resolve the paradox of rich visual experience and curious perceptual failures". In: *Attention, Perception, & Psychophysics* 82.3, pp. 901–925.

Scolari, Miranda, Anna Byers, and John T Serences (2012). "Optimal deployment of attentional gain during fine discriminations". In: *Journal of Neuroscience* 32.22, pp. 7723–7733.

Scolari, Miranda, Andrew Kohnen, et al. (2007). "Spatial attention, preview, and popout: Which factors influence critical spacing in crowded displays?" In: *Journal of Vision* 7.2, pp. 7.1–7.23.

Silver, Michael A, David Ress, and David J Heeger (2007). "Neural correlates of sustained spatial attention in human early visual cortex". In: *Journal of neurophysiology* 97.1, pp. 229–237.

Silver, Michael A, Amitai Shenhav, and Mark D'Esposito (2008). "Cholinergic enhancement reduces spatial spread of visual responses in human early visual cortex". In: *Neuron* 60.5, pp. 904–914.

Smith, A T, K D Singh, and M W Greenlee (2000). "Attentional suppression of activity in the human visual cortex". In: *Neuroreport* 11.2, pp. 271–278.

Strasburger, Hans and Maka Malania (2013). "Source confusion is a major cause of crowding". In: *Journal of vision* 13.1, pp. 24.1–24.20.

Talipski, Louisa A, Stephanie C Goodhew, and Mark Edwards (2021). "No effect of spatial attention on the processing of a motion ensemble: Evidence from Posner cueing". In: *Attention, Perception, & Psychophysics*, pp. 1–13.

Tanrıkulu, Ömer Dağlar, Andrey Chetverikov, and Árni Kristjánsson (2020). "Encoding perceptual ensembles during visual search in peripheral vision". In: *Journal of Vision* 20.8, pp. 20.1–20.18.

Theiss, Justin D, Joel D Bowen, and Michael A Silver (2021). "Spatial Attention Enhances Crowded Stimulus Encoding Across Modeled Receptive Fields by Increasing Redundancy of Feature Representations". In: *Neural Computation* 34.1, pp. 190–218.

Toet, Alexander and Dennis M Levi (1992). "The two-dimensional shape of spatial interaction zones in the parafovea". In: *Vision research* 32.7, pp. 1349–1357.

Van den Berg, Ronald, Jos BTM Roerdink, and Frans W Cornelissen (2010). "A neurophysiologically plausible population code model for feature integration explains visual crowding". In: *PLoS Comput Biol* 6.1, e1000646.

Whitney, David and Dennis M Levi (2011). "Visual crowding: A fundamental limit on conscious perception and object recognition". In: *Trends in cognitive sciences* 15.4, pp. 160–168.

Whitney, David and Allison Yamanashi Leib (2018). "Ensemble perception." In: *Annual review of psychology* 69.1, pp. 105–129.

Womelsdorf, Thilo et al. (2006). "Dynamic shifts of visual receptive fields in cortical area MT by spatial attention". In: *Nature neuroscience* 9.9, pp. 1156–1160.

Yantis, Steven and John Jonides (1990). "Abrupt visual onsets and selective attention: voluntary versus automatic allocation." In: *Journal of Experimental Psychology: Human perception and performance* 16.1, pp. 121–134.

Yeshurun, Yaffa and Marisa Carrasco (1998). "Attention improves or impairs visual performance by enhancing spatial resolution". In: *Nature* 396.6706, pp. 72–75.

— (2008). "The effects of transient attention on spatial resolution and the size of the attentional cue". In: *Perception & Psychophysics* 70.1, pp. 104–113.

Yeshurun, Yaffa, Barbara Montagna, and Marisa Carrasco (2008). "On the flexibility of sustained attention and its effects on a texture segmentation task". In: *Vision research* 48.1, pp. 80–95.

Yeshurun, Yaffa and Einat Rashal (2010). "Precueing attention to the target location diminishes crowding and reduces the critical distance". In: *Journal of Vision* 10.10, pp. 16.1–16.12.

Ying, Haojiang (2022). "Attention modulates the ensemble coding of facial expressions". In: *Perception* 51.4, pp. 276–285.