How voluntary and involuntary attention differentially shape spatial resolution

Michael Jigo[1*], David J. Heeger[1,2] & Marisa Carrasco[1,2]

[1] Center for Neural Science and [2] Department of Psychology, New York University

[*] corresponding author

Correspondence: michael.jigo@nyu.edu

Key words:

computational model; texture; visual field; voluntary attention; involuntary attention; normalization; spatial resolution

ABSTRACT

Attention shapes spatial resolution across the visual field, with different perceptual consequences by voluntary and involuntary attention. Extant models of visual attention cannot explain these differences. Here, we develop a model of spatial resolution that explains the differential effects of attention. Our model emulates sensory encoding to segment figures from their background in grayscale images. To explain behavior, each attention type requires separate operating regimes across visual detail (spatial frequency). We focus on texture-based segmentation, which has revealed a clear dissociation between both attention types. The model reproduces segmentation performance across several experiments and simultaneously resolves three unexplained phenomena: (1) the parafoveal advantage in segmentation, (2) the peripheral improvements and foveal impairments by involuntary attention and (3) the uniform improvements across eccentricity by voluntary attention. Our model unveils fundamental differences between voluntary and involuntary attention and provides a generalizable quantitative framework for predicting their effects across the visual field.

INTRODUCTION

Texture segmentation is crucial for detecting objects across the visual field (Bergen and Adelson, 1988; Landy, 2013; Landy and Graham, 2004; Victor et al., 2017). This process isolates regions of local structure from their surroundings. Textures comprise information that span multiple spatial scales (spatial frequency, SF), each conveying different levels of detail (Landy, 2013; Landy and Graham, 2004; Victor et al., 2017). The visual system's spatial resolution (acuity) constrains the visibility of detail and limits segmentation (Kehrer, 1989; Morikawa, 2000). Spatial resolution is high at the fovea and declines with eccentricity, which blurs the neural representation of the periphery (DeValois and DeValois, 1990; Graham, 1989). Covert attention—the prioritization of a spatial location without concurrent eye movements—modifies spatial resolution (Anton-Erxleben and Carrasco, 2013; Carrasco, 2011), shaping the level of detail that can be resolved. As a result, the success of segmentation hinges on the eccentricity of a texture target (the to-be-segmented region) (Kehrer, 1989; Morikawa, 2000) and on the attentional locus (Anton-Erxleben and Carrasco, 2013; Carrasco, 2011).

Conventional models treat texture segmentation as a feedforward process that filters an image into its elementary features (e.g., SF and orientation), transforms them to reflect the local structure (e.g., a region of similarly oriented bars irrespective of their individual positions), then pools across space to emphasize texture-defined contours (Bergen and Adelson, 1988; Landy, 2013; Landy and Graham, 2004). Few of these models account for variations in spatial resolution across eccentricity (Kehrer, 1997; Kehrer and Meinecke, 2003) and all postulate that segmentation is a 'preattentive' (Victor et al., 2017) operation whose underlying neural processing is impervious to attention (Bergen and Adelson, 1988; Kehrer, 1997; Kehrer and Meinecke, 2003; Landy, 2013; Landy and Graham, 2004).

However, attention alters texture segmentation (Anton-Erxleben and Carrasco, 2013; Carrasco, 2011). Attending adjacent textures can improve or impair segmentation, depending on the type of attention deployed and the texture eccentricity. Involuntary spatial attention—the transient, stimulus-driven form of selection (Carrasco, 2011)—improves segmentation for peripheral stimuli but impairs it near the fovea, a consequence of an automatic attentional enhancement of spatial resolution (Carrasco et al., 2006; Jigo and Carrasco, 2018; Talgar and Carrasco, 2002; Yeshurun and Carrasco, 1998, 2000, 2008). Voluntary spatial attention—the sustained, goal-driven form of selection (Carrasco, 2011)—improves segmentation at all attended locations by optimizing spatial resolution to distinguish the textural boundary (Barbot and Carrasco, 2017; Jigo and Carrasco,

2018; Yeshurun et al., 2008b). Owing to these distinct attentional effects, texture segmentation provides a formidable illustration of the stark differences between involuntary and voluntary attention.

Two main hypotheses have been proposed to explain how attention alters spatial resolution. Psychophysical studies ascribe attentional effects to modulations of SF sensitivity (Barbot and Carrasco, 2017; Carrasco et al., 2006). Neurophysiological (Moran and Desimone, 1985; Womelsdorf et al., 2006; Womelsdorf et al., 2008) and neuroimaging (Fischer and Whitney, 2009; Klein et al., 2014) studies bolster the idea that attention modifies spatial profiles of neural receptive fields (Anton-Erxleben and Carrasco, 2013). Both hypotheses qualitatively predict attentional effects but do not specify their underlying neural computations.

Here, we develop a new computational model in which feedforward processing and attentional gain operate across eccentricity and SF and contribute to segmentation performance. (i) Feedforward processing. We augment a conventional model of texture processing (Bergen and Adelson, 1988; Landy, 2013; Landy and Graham, 2004). Our model varies with eccentricity and includes contextual modulation within local regions in the stimulus via normalization (Heeger, 1992). The defining characteristic of normalization is that the response of an individual neuron is (divisively) suppressed by the summed activity of neighboring neurons responsive to different positions and features in the scene. This canonical neural computation is widespread across basic sensory and higher-order perceptual and cognitive processing (Carandini and Heeger, 2011). (ii) Attentional gain. We model attention as multiplicative gains (attentional gain factors (Reynolds and Heeger, 2009)) that vary with eccentricity and SF. Attention shifts sensitivity toward fine or coarse details depending on the range of SFs enhanced.

Our model is image-computable, which allowed us to reproduce behavior directly from grayscale images used in psychophysical experiments (Barbot and Carrasco, 2017; Carrasco et al., 2006; Talgar and Carrasco, 2002; Yeshurun and Carrasco, 1998, 2008; Yeshurun et al., 2008b). The model explains three signatures of texture segmentation hitherto unexplained within a single computational framework (**Figure 1**). (i) The central performance drop (CPD) (Barbot and Carrasco, 2017; Carrasco et al., 2006; Jigo and Carrasco, 2018; Kehrer, 1989; Morikawa, 2000; Talgar and Carrasco, 2002; Yeshurun and Carrasco, 1998, 2000, 2008; Yeshurun et al., 2008b) (**Figure 1A**), i.e., the parafoveal advantage of segmentation over the fovea. (ii) The improvements in the periphery and impairments at foveal locations induced by involuntary attention (Carrasco et al., 2006; Jigo and Carrasco, 2018; Talgar and Carrasco, 2002; Yeshurun and Carrasco, 1998,

2000, 2008; Yeshurun et al., 2008b) (**Figure 1B**). (iii) The equivalent improvements across eccentricity by voluntary attention (Barbot and Carrasco, 2017; Jigo and Carrasco, 2018; Yeshurun et al., 2008b) (**Figure 1C**). Whereas our analyses focused on texture segmentation, our model is general and can be applied to any psychophysical discrimination or detection task with static, grayscale images. For example, the model predicts contrast sensitivity functions across eccentricity (**Supplementary Figure 1**).

To preview our results, model comparisons revealed that normalization is necessary to elicit the CPD and that the effects of involuntary and voluntary attention are generated by separate profiles of gain enhancement across SF (Jigo and Carrasco, 2020). A preferential high-SF enhancement reproduces the impairments by involuntary attention due to a shift in visual sensitivity toward details too fine to distinguish the texture at central locations. The transition from impairments to improvements in the periphery results from involuntary attentional gain gradually shifting to lower SFs that are more amenable to distinguish the texture target. Improvements by voluntary attention result from a uniform enhancement of SFs that encompass the target, optimizing visual sensitivity for the texture across eccentricity.

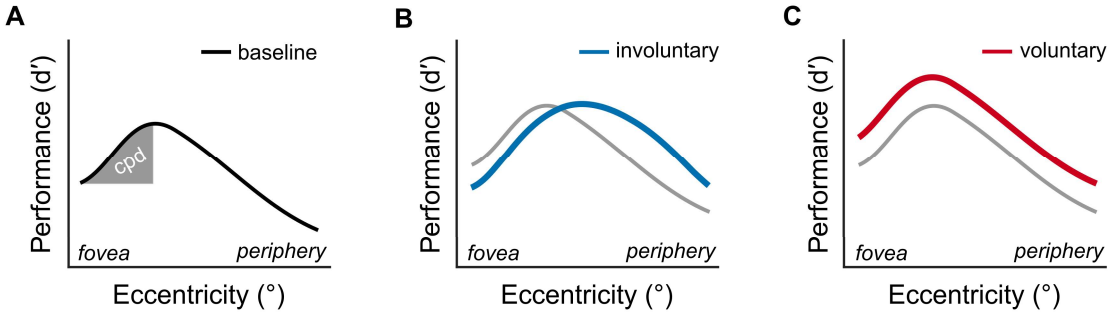


Figure 1. Signatures of texture segmentation.
(**A**) Central performance drop. Shaded region depicts the magnitude of the central performance drop.
(**B**) Involuntary attention modulation. Involuntary attention improves segmentation performance in the periphery and impairs it near the fovea.
(**C**) Voluntary attention modulation. Voluntary attention improves segmentation performance across eccentricity.

RESULTS

Image-computable model of attention and spatial resolution

The model is motivated by established principles of neural computation (Carandini and Heeger, 2011; Heeger, 1992), human pattern (DeValois and DeValois, 1990; Graham, 1989) and texture vision (Bergen and Adelson, 1988; Landy, 2013; Landy and Graham, 2004) and attentional modulation (Reynolds and Heeger, 2009). The model incorporates elements of the Reynolds-Heeger normalization model of attention (NMA) (Reynolds and Heeger, 2009) and illuminates how attention alters contrast and texture sensitivity across SF and eccentricity. Our model incorporates: (i) SF-tuned gain modulation to emulate the decline in contrast sensitivity and peak SF preference with eccentricity. (ii) Spatial summation of normalized inputs to generate texture selectivity. (iii) Separate attentional gain profiles across SF to reproduce effects of involuntary and voluntary attention. The model is composed of four components: stimulus drive, attentional gain, suppressive drive and spatial summation (**Figure 2A**). Following NMA, attention adjusts the gain on the stimulus drive before normalization. For a full description of the model, see **Methods**.

Stimulus drive. We simulate bottom-up responses of a collection of linear receptive fields, each jointly tuned to spatial position, SF and orientation. Images are processed through a filter bank (Simoncelli et al., 1992) covering the visual field at several SFs and orientations using bandwidths compatible with neurophysiological (DeValois and DeValois, 1990) and psychophysical (Graham, 1989) measurements. Filter outputs are combined across quadrature phase (Adelson and Bergen, 1985), yielding contrast energy images corresponding to different SFs and orientations. These outputs simulate the responses of complex cells in primary visual cortex (Adelson and Bergen, 1985; DeValois and DeValois, 1990). The gain on individual receptive fields varies as a function of SF and eccentricity preference (**Figure 2A, green**). Following the behavior of individual neurons (DeValois and DeValois, 1990) and human pattern vision (Graham, 1989), gain modulation is narrowly tuned to high SFs near the fovea and progressively shifts to low SFs in the periphery. Consequently, the stimulus drive reflects local spectral energy within each patch in the image, filtered through feature-selective receptive fields that vary with eccentricity.

Attentional gain. Attention is implemented as a gain control mechanism that multiplicatively scales the gain on the stimulus drive (Reynolds and Heeger, 2009). The magnitude of attentional gain is largest at the cued location (**Figure 2A, orange ring**) and varies with the eccentricity and SF preference of each receptive field. Motivated by findings of psychophysical experiments that manipulated voluntary and involuntary attention (Jigo and Carrasco, 2020), two SF-tuned profiles

are assessed—narrow and broad. The narrow profile selectively enhances a small range of SFs at each eccentricity (**Figure 2A, blue**); the broad profile uniformly enhances SFs (**Figure 2A, red**).

Suppressive drive. Suppression operates via divisive normalization, a canonical computation in sensory processing and cognition (Carandini and Heeger, 2011; Heeger, 1992). Normalized responses are proportional to the attention-scaled stimulus drive divided by a normalization pool plus a constant σ^2 that increases with eccentricity. This constant adjusts the model's overall sensitivity to contrast (i.e., contrast gain; **Figure 2A, black**). The normalization pool consists of the attention-scaled stimulus drive across nearby spatial locations (surround suppression (Cavanaugh et al., 2002)), uniformly across orientation (cross-orientation suppression (Brouwer and Heeger, 2011)) and across preferred and neighboring SFs (cross-frequency suppression (Petrov et al., 2005)) of individual receptive fields. Such broad suppressive pools are supported by physiological (Brouwer and Heeger, 2011; Cavanaugh et al., 2002; Priebe and Ferster, 2006) and psychophysical (Meese and Holmes, 2007; Petrov et al., 2005; Sagi and Hochstein, 1985) findings and models of visual processing (Heeger, 1992).

Spatial summation. Normalized responses are weighted and summed across space within each SF and orientation filter. Motivated by empirical findings (Westrick and Landy, 2013), spatial summation followed normalization, which accentuated texture-defined contours within the image. The size of pooling regions scale with the SF preference of each receptive field (Landy, 2013; Landy and Graham, 2004) (**Figure 2A, purple**); larger for low than for high SFs. This implements an inverse relation between the integration area of individual receptive fields and their SF tuning.

Target discriminability. The model generated measures of discriminability (d′) in a texture segmentation task (**Figure 2B**). The model processed two texture images to generate their respective population responses. One texture contained a target patch of orientation that differed from its surround (target-present) and the other consisted of uniform orientation throughout (target-absent). The vector length (i.e., Euclidean norm) of the difference between their respective population responses indexed stimulus discriminability. This measure is proportional to behavioral performance, assuming the addition of normally distributed noise after normalization (**Methods, Decision mechanism**).

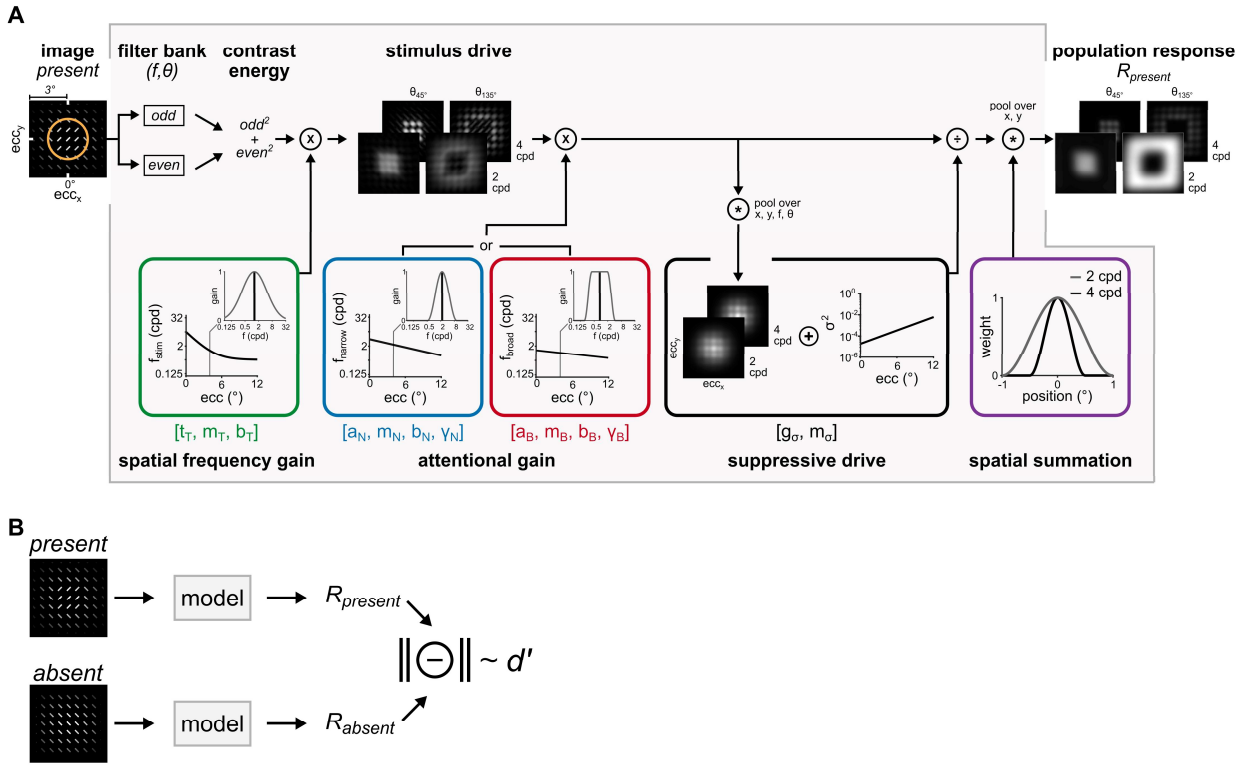


Figure 2. Image computable model of attention and spatial resolution
(**A**) Model structure. A filter bank of linear receptive fields decomposes an image. Receptive fields vary in their preferred position, SF and orientation. Filter responses are squared and summed across quadrature-phase pairs (odd, even), yielding contrast energy outputs. *Spatial frequency gain* scales contrast energy across spatial frequency and eccentricity (green box). The solid black line depicts the center frequency of the tuning function (f_{stim}); the inset displays the full tuning function at a single eccentricity. The *stimulus drive* characterizes the contrast energy at each pixel in the image, filtered through feature-selective and eccentricity-dependent receptive fields. A subset of receptive fields tuned to two SFs and orientations are displayed. *Attentional gain* multiplicatively scales the stimulus drive at a circumscribed region within the image (orange circle in left panel) and varies across SF and eccentricity. The center SF of attentional gain varies with eccentricity (solid black lines in blue and red boxes). Across SF, attentional gain follows distinct profiles. The narrow profile (blue box) selectively enhances a given frequency (f_{narrow}), as depicted in the inset. The broad profile (red box) uniformly enhances a range of SFs (inset), centered on a given frequency (f_{broad}). The *suppressive drive* comprises attention-scaled stimulus drive outputs pooled across a local neighborhood of positions, SFs and uniformly across orientation. Contrast gain, σ^2, adjusts overall suppression magnitude across eccentricity. *Spatial summation* follows normalization (purple box) and generates the *population response*. Pooling area varies inversely with SF tuning. Variables displayed within the square brackets depict model parameters fit to behavior.
(**B**) Target discriminability. Population responses for a texture image containing a patch (*present*) and one without (*absent*) are computed. The vector magnitude of their difference produces a metric proportional to d′, assuming independent and identically distributed Gaussian output noise.

Texture stimuli, behavioral protocol and optimization strategy

Stimuli. Model parameters were constrained by data from ten published psychophysical experiments. Involuntary attention was manipulated in six experiments (Carrasco et al., 2006; Talgar and Carrasco, 2002; Yeshurun and Carrasco, 1998, 2008; Yeshurun et al., 2008b) (**Figure 3A**) and voluntary attention in four experiments (Barbot and Carrasco, 2017; Yeshurun et al., 2008b) (**Figure 3B**). In each experiment, observers distinguished a patch of one orientation embedded within a background of differing orientation at several possible eccentricities. Grayscale images of the experiments' stimuli served as input to the model.

Behavioral protocol. Performance was typically measured with a two-interval forced choice protocol (**Figure 3C**). Observers maintained fixation at the display's center while viewing two intervals of texture stimuli, one of which randomly contained a target texture. Different pre-cues at their optimal timing manipulated involuntary or voluntary attention. Brief peripheral pre-cues manipulated involuntary attention and appeared before both intervals, but near the upcoming target location in the interval containing the target (Carrasco et al., 2006; Jigo and Carrasco, 2018; Talgar and Carrasco, 2002; Yeshurun and Carrasco, 1998, 2000, 2008; Yeshurun et al., 2008b). Symbolic pre-cues manipulated voluntary attention, appeared near fixation and indicated the target location in the target-present interval (Barbot and Carrasco, 2017; Yeshurun et al., 2008b). Attention effects were determined relative to a neutral condition, in which observers distributed attention across all possible target locations. Behavioral performance displayed the three signatures of texture segmentation: (i) The CPD emerged in the neutral condition (**Figure 1A**). (ii) Peripheral pre-cues improved performance in the periphery and impaired it at central locations (**Figure 1B**). (iii) Central, symbolic pre-cues improved performance at all eccentricities (**Figure 1C**).

Optimization. To identify the computations that underlie each signature of segmentation, we separately fit the model to three subsets of behavioral data. First, the CPD was isolated from attentional effects by fitting to the neutral condition from all ten experiments. Second, involuntary attentional effects were assessed by fitting to neutral and peripheral cueing conditions from the six involuntary attention experiments. Third, voluntary attentional effects were assessed by fitting to neutral and central cueing conditions from the four voluntary attention experiments (**Methods, Optimization strategy**).

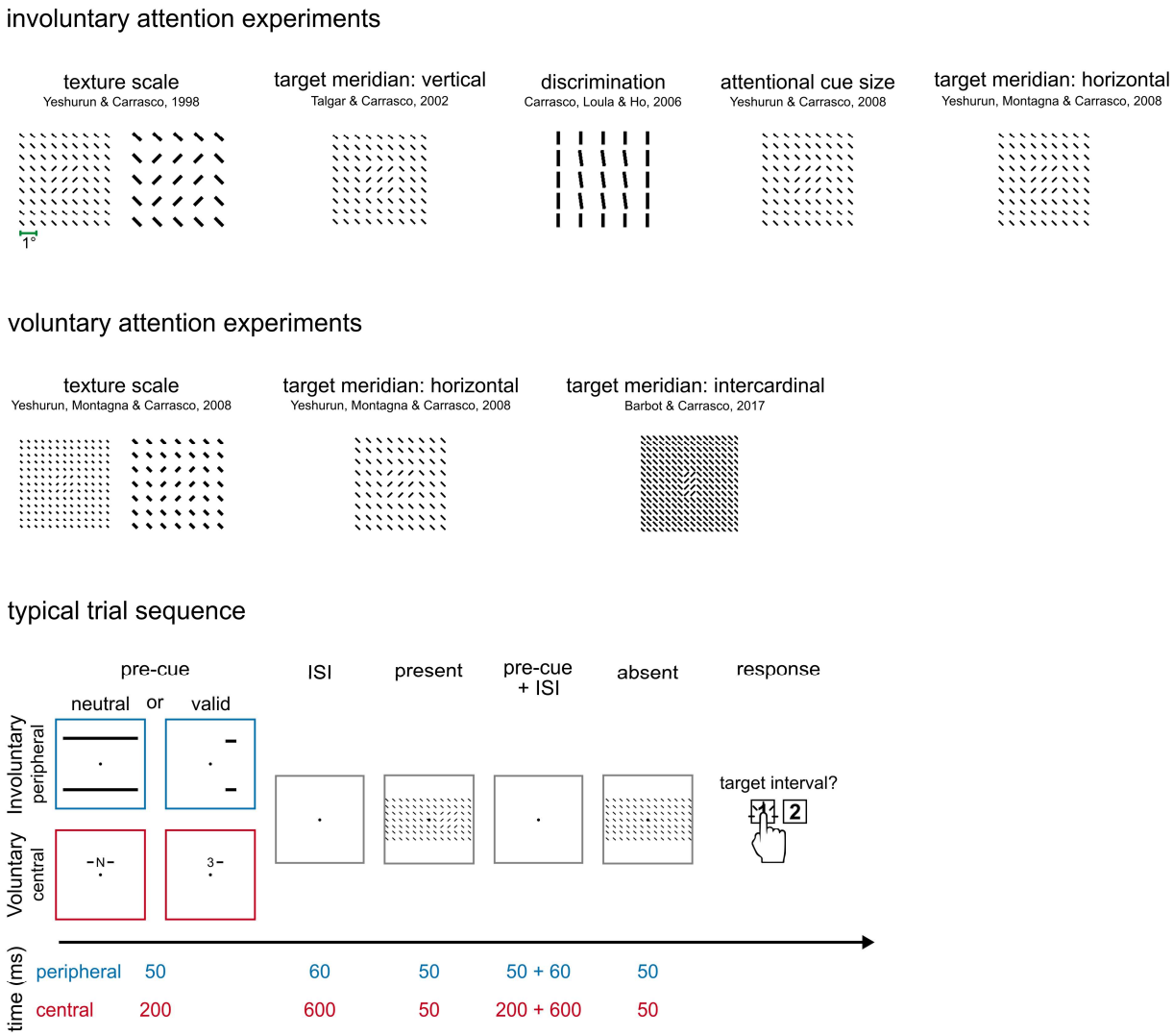


Figure 3. Texture stimuli and a typical behavioral protocol.
Target-present texture stimuli, displayed at their respective spatial scales, used in experiments in which (**A**) involuntary attention and (**B**) voluntary attention were manipulated. Scale bar (green) applies to all texture stimuli. Target-absent stimuli were homogenous textures. Labels above each texture indicate the experimental manipulation conducted in the corresponding experiment.
(**C**) Two-interval forced choice protocol typically used to assess texture segmentation performance across eccentricity. Numbers correspond to the representative timing information for each pre-cue—peripheral (blue) and central (red)—and their corresponding inter-stimulus intervals (ISI). Peripheral and central pre-cues manipulated involuntary and voluntary attention, respectively. Neutral pre-cues equally distributed attention to all possible target locations. Valid peripheral pre-cues appeared near the upcoming target location whereas valid central pre-cues symbolically indicated the upcoming target location. In the displayed example, the number “3” and the adjacent line indicate that the target would appear at a peripheral eccentricity in the right visual hemifield.

Contextual modulation and spatial summation mediate the CPD

To identify the computations mediating the CPD, we simultaneously fit the model to group-average performance in each experiment's neutral condition (103 data points). 15 model parameters constrained performance (**Methods, Table 2**). To account for differences in contrast sensitivity due to variable display properties among experiments (e.g., mean luminance), foveal contrast gain (g_σ; **Figure 2A**) was independently determined for each of ten experiments (10 parameters). Two separate parameters determined foveal SF preference (t_T)–one shared among involuntary attention studies and another among voluntary attention studies. The remaining three parameters–SF bandwidth (b_T), the gradual increase in contrast gain (m_σ) and the progressive shift to lower SFs with eccentricity (m_T)–were shared among all experiments. Attentional gain was not included for these fits.

The model reproduced the CPD and its dependence on texture scale (**Figure 4**). With no change in parameters, the model predicted contrast sensitivity functions (Jigo and Carrasco, 2020; Pointer and Hess, 1989) reasonably well (**Supplementary Figure 1**). For a fine-scale texture—characterized by narrow, densely spaced lines—performance peaked within the parafovea (4 deg) and declined toward the fovea and periphery (**Figure 4A**). Model fits revealed that differences between target-present and target-absent stimuli were largest within the 2 cpd filter (**Figure 4A, middle**). This filter best differentiated the target patch from a homogenous texture; we denote its center SF as f_{fine}. A coarser texture was best distinguished by lower SFs (1 cpd, f_{coarse}), which exaggerated the CPD, moving peak performance to a farther eccentricity (~6 deg; **Figure 4B**). The CPD was well-fit by the model for all experiments (**Figure 4C**) and explained 77% of the variance (95% bootstrapped CI = [70 80]), with the best-fitting regression line falling close to the unity line.

Previous models qualitatively matched the CPD through spatial summation (Kehrer, 1997; Kehrer and Meinecke, 2003), but ignored the contributions of contextual modulation via normalization. To assess the contribution of each operation to behavior, we compared the full model to variants that either lacked components of the suppressive drive (cross-orientation, cross-frequency, and/or surround suppression) or spatial summation (**Figure 4D**). We separately restricted the pool of orientations (-θ), SFs (-f), spatial positions (-x,y) or all simultaneously (-all) such that suppressive modulations due to featural attributes and/or spatial positions outside each receptive field's tuning were removed. In the –context model, each receptive field was suppressed by its response alone, unaffected by its surrounding population. The final variant lacked spatial summation (-sum), which resulted in a population response that consisted of only normalized outputs. Removing spatial summation results in an attenuated response to regions of similar orientation (e.g., target patch).

Each model was fit to behavioral performance in the neutral condition across all experiments and compared using Akaike Information Criterion (Akaike, 1974) (AIC).

Removing contextual modulation or spatial summation attenuated the CPD (**Supplementary Figure 2**). We measured model performance relative to the full model, which yielded a ΔAIC score; positive values represent a decrease in model performance. We use “M” and “CI” to denote the median and 95% confidence interval of the bootstrapped distribution of ΔAIC scores. Model performance fell without cross-orientation suppression (M=3.8, CI=[-0.6 8.6]), cross-frequency suppression (M=7.6, CI=[2.3 13.1]), surround suppression (M=9.1, CI=[3.7 14.3]) and without all forms of contextual modulation (M=18.0, CI=[13.0 23.1]). Without spatial summation, model performance decreased as well (M=35.9, CI=[31.7 40.3]). Thus, reliable reproduction of the CPD requires both contextual modulation and spatial summation.

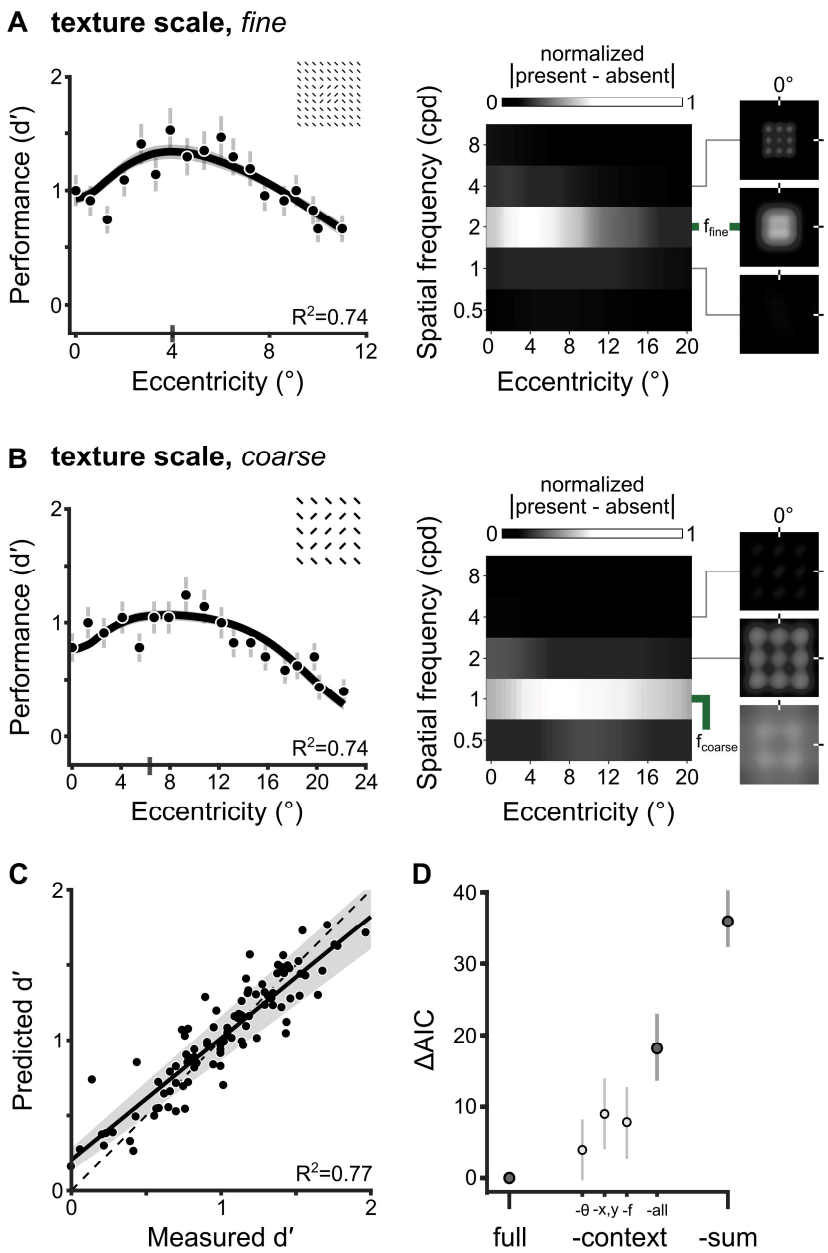


Figure 4. Contextual modulation and spatial summation mediate the CPD
(**A**) *Left*. Fit to Experiment 1 in (Yeshurun and Carrasco, 1998) which used a fine-scale texture. Black dots and error bars depict group-average performance and ±1 SEM. The black line and shaded region indicate the median and 68% bootstrapped confidence interval. The model predicted performance simultaneously for all ten experiments. The gray bar on the x-axis indicates the eccentricity at which performance peaked. The inset depicts the experiment's texture stimulus.
Middle. The matrix depicts the absolute value of differences between target-present and target-absent population responses, normalized by the maximum across eccentricity and averaged across orientation and space. f_{fine} denotes the SF filter containing the largest difference between the target-present and target-absent fine-scale textures. Note that the vector length of differences indexed target discriminability (**Figure 2B**); we use absolute differences to visualize the SFs that drove discriminability.
Right. Spatial distributions of the absolute value of differences between target-present and target-absent population responses. Each panel depicts a subset of receptive fields centered on the fovea (indicated by

tick marks). Receptive fields were tuned to one of three spatial frequencies (4, 2, 1 cpd) and an orientation of 30°.
(**B**) Fit to Experiment 2 in (Yeshurun and Carrasco, 1998), which used a coarse-scale texture. The model predicted performance simultaneously with the fit shown in *a*. Visualization follows the conventions in *a*. Note that eccentricity (x-axis) is twice that of *a*. f_{coarse} denotes the spatial frequency filter that best distinguished the coarse-scale target.
(**C**) Goodness-of-fit for the neutral condition across ten experiments. Each dot depicts the measured (x-axis) and predicted (y-axis) behavioral performance at a given eccentricity. The solid line and shaded area depict the best-fitting regression line and its 95% confidence interval, respectively. The dashed line indicates the unity line y=x.
(**D**) Model comparisons using AIC. ΔAIC is the AIC of each model minus the AIC of the full model. '-θ' denotes the model variant without cross-orientation suppression, '-f' without cross-frequency suppression, '-x,y' without surround suppression, '-context' devoid of all contextual modulation and '-sum', without spatial summation. The dots represent the median ΔAIC and error bars denote 95% confidence interval of the bootstrap distribution.

Narrow high-SF enhancement mediates involuntary attention effects

The model simultaneously predicted group-average data from neutral and peripheral cueing conditions across size experiments (146 data points). Involuntary attention was modeled as a narrow SF gain profile (**Figure 2, blue**), motivated by psychophysical measurements (Jigo and Carrasco, 2020). 14 free parameters constrained model behavior (**Methods, Table 3**). Model parameters that determined neutral cueing performance—foveal contrast gain (g_σ), SF tuning (t_T), SF bandwidth (b_T), the increase in contrast gain (m_σ) and the decline in SF preference with eccentricity (m_T)—were configured identically as described above. Four parameters, shared among experiments, determined attentional gain–foveal SF preference (a_N), the gradual shift to lower SFs with eccentricity (m_N), SF bandwidth (b_N) and amplitude (γ_N). Consequently, attention operated identically on each texture stimulus, with the spatial spread of attention fixed across experiments (**Methods, Table 1**).

The model reproduced the central impairments, peripheral improvements and their variation with texture scale. For a fine-scale texture, the narrow gain profile across SF yielded improvements within the parafovea (4-12°), impairments across a small range of central eccentricities (0-2°) and shifted peak performance toward the periphery (~6°; **Figure 5A**). For the coarser texture, the same attentional gain profile generated improvements in the periphery (8-22°), impairments within the parafovea (0-8°) and shifted peak performance farther toward the periphery (~15°; **Figure 5B**).

A gradual shift of attentional gain to lower SFs (Jigo and Carrasco, 2020) reproduced the transition from impairments to improvements across eccentricity (**Figure 5C**). At the fovea, attentional gain was centered on a SF (4 cpd) higher than those distinguishing the fine- (2 cpd, f_{fine}) or coarse-scale

(1 cpd, f_{coarse}) textures. As a result, the population response shifted away from the target and impaired performance. With increasing eccentricity, attentional gain progressively overlapped the SFs of fine- and coarse-scale targets, improving performance. Attention enhanced the fine-scale target SF within the parafovea (4-12°) then enhanced the coarse-scale target at farther eccentricities (8-22°). Overall, across both neutral and peripheral cueing conditions in six experiments, the model explained 77% of the variance (95% bootstrapped CI = [49 82]; **Figure 5D**).

Attentional gain on SFs higher than the target yielded impairments at central locations. This pattern was consistent across all six experiments (**Figure 5E**). Consequently, the overlap between fine- (f_{fine}) or coarse-scale (f_{coarse}) target SFs and the SF tuning of attentional gain was minimal at the fovea and peaked in the periphery (**Figure 5F**). This mismatch between the SF tuning of attention (f_{narrow}) and the target is suggested to be driven by involuntary attention operating above intrinsic SF preferences at each eccentricity (Jigo and Carrasco, 2020). We corroborated this relation. We compared f_{narrow} to the model's baseline SF tuning, indexed by the stimulus drive (f_{stim}, **Figure 2A**). Consistent with empirical measurements, we found that the narrow attentional gain profile preferred SFs higher than baseline (**Supplementary Figure 3**).

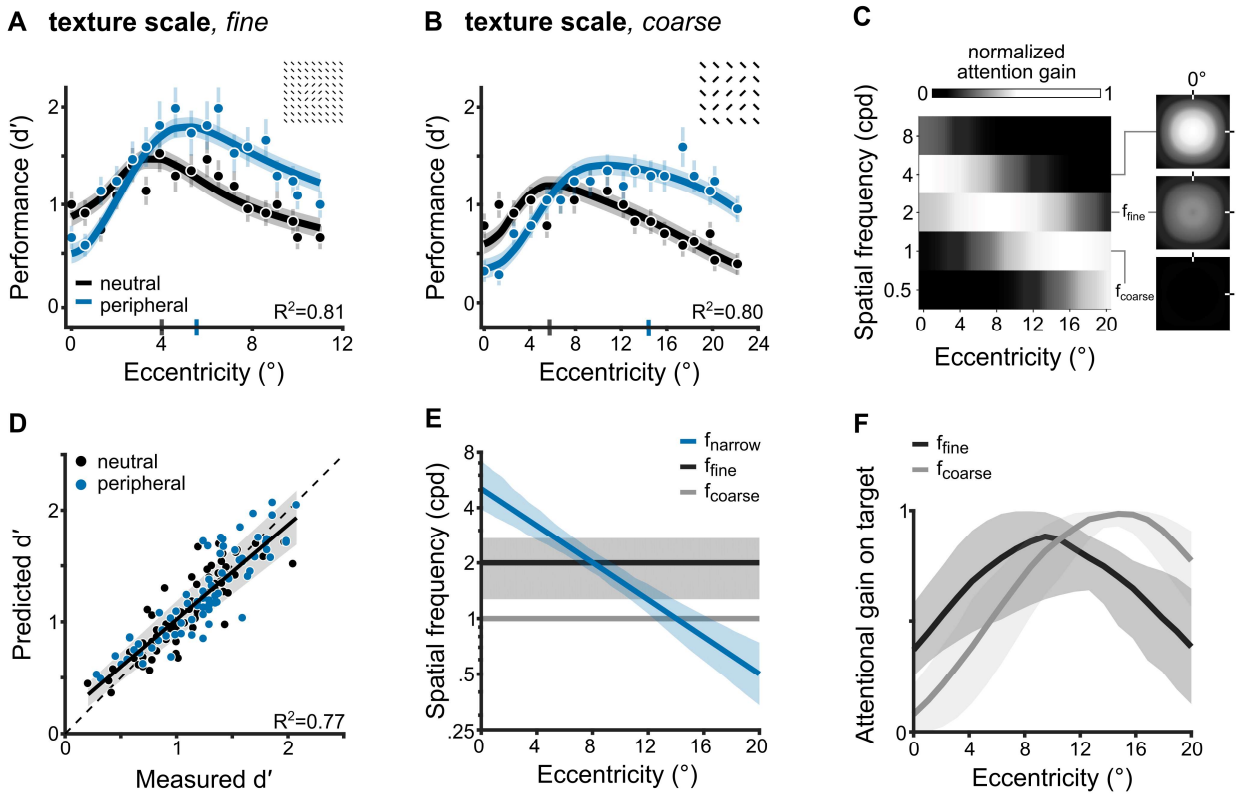


Figure 5. Narrow high-SF enhancement mediates involuntary attention effects
(**A**) Fit to Experiment 1 in (Yeshurun and Carrasco, 1998), which used a fine-scale texture. The dots depict the group-average and error bars denote ±1 SEM. Solid lines indicate the median and shaded regions indicate the 68% confidence interval of the bootstrapped distribution of model fits. The vertical blue bar on the x-axis indicates the eccentricity at which performance peaked with valid peripheral cues. The model predicted attention effects simultaneously for all six involuntary attention experiments.
(**B**) Fit to Experiment 2 in (Yeshurun and Carrasco, 1998), which used a coarse texture scale. This fit was determined simultaneously with that shown in *a*. Visualization follows the conventions in *a*.
(**C**) Best-fitting narrow gain profile.
Left. The matrix depicts attentional gain across eccentricity, normalized by the maximum and averaged across space and orientation. The SFs that best distinguished fine- (f_{fine}) and coarse-scale targets (f_{coarse}) in *a* and *b* are displayed.
Right. The spatial distribution of attentional gain on a subset of receptive fields centered on the fovea (indicated by tick marks). Receptive fields were tuned to one of three spatial frequencies (1, 2, 4 cpd) and an orientation of 30°.
(**D**) Goodness-of-fit for simultaneous model fit to neutral and peripheral-cued performance. Plotted as in **Figure 4C.**
(**E**) SF preference of the narrow attentional gain profile (f_{narrow}) and the SF that best distinguished target-present from target-absent stimuli for fine- (f_{fine}) and coarse-scale textures (f_{coarse}). The solid lines indicate the median and the shaded areas denote the 68% confidence interval of bootstrapped distribution. Shaded area for f_{coase} overlaps the solid line.
(**F**) Normalized magnitude of attentional gain on the fine- and coarse-scale target SF across eccentricity (median and 68% confidence interval of bootstrapped distribution).

Broad SF enhancements yield voluntary attention effects

The model simultaneously predicted group-average data from neutral and central cueing conditions across four experiments (60 data points). Voluntary attention was modeled as a broad gain profile (**Figure 2A, red**), motivated by psychophysical measurements (Jigo and Carrasco, 2020). 12 free parameters constrained model behavior (**Methods, Table 4**). Four parameters, shared among experiments, determined attentional gain–foveal SF preference (a_B), the decline in SF preference with eccentricity (m_B), SF bandwidth (b_B) and amplitude (γ_B).

The model reproduced improvements across eccentricity for both fine- (**Figure 6A**) and coarse-scale textures (**Figure 6B**). To generate these improvements, attentional gain encompassed the target SF for each texture scale (**Figure 6C**). Across all four experiments, the model explained 89% of the variance (95% bootstrapped CI [67 92]; **Figure 6D**).

Broad SF attentional gain, centered near the target SF across eccentricity (f_{broad} in **Figure 6E**), reproduced voluntary attention effects. This contrasts with the narrow SF gain profile that modulated higher SFs at central locations to reproduce involuntary attention effects (**Figure 5E**). Although the center SF of attention declined with eccentricity, the modulation profile's plateau ensured that it overlapped both fine- and coarse-scale target SFs across eccentricity (**Figure 6F**). Psychophysical measurements of attentional effects on contrast sensitivity (Jigo and Carrasco,

2020) suggest that the SF range enhanced by voluntary attention is centered near those intrinsically preferred by an observer at each eccentricity. However, our model fits to texture segmentation experiments revealed that attentional gain enhanced lower SFs than baseline tuning (f_{stim}) at central locations (**Supplementary Figure 4**).

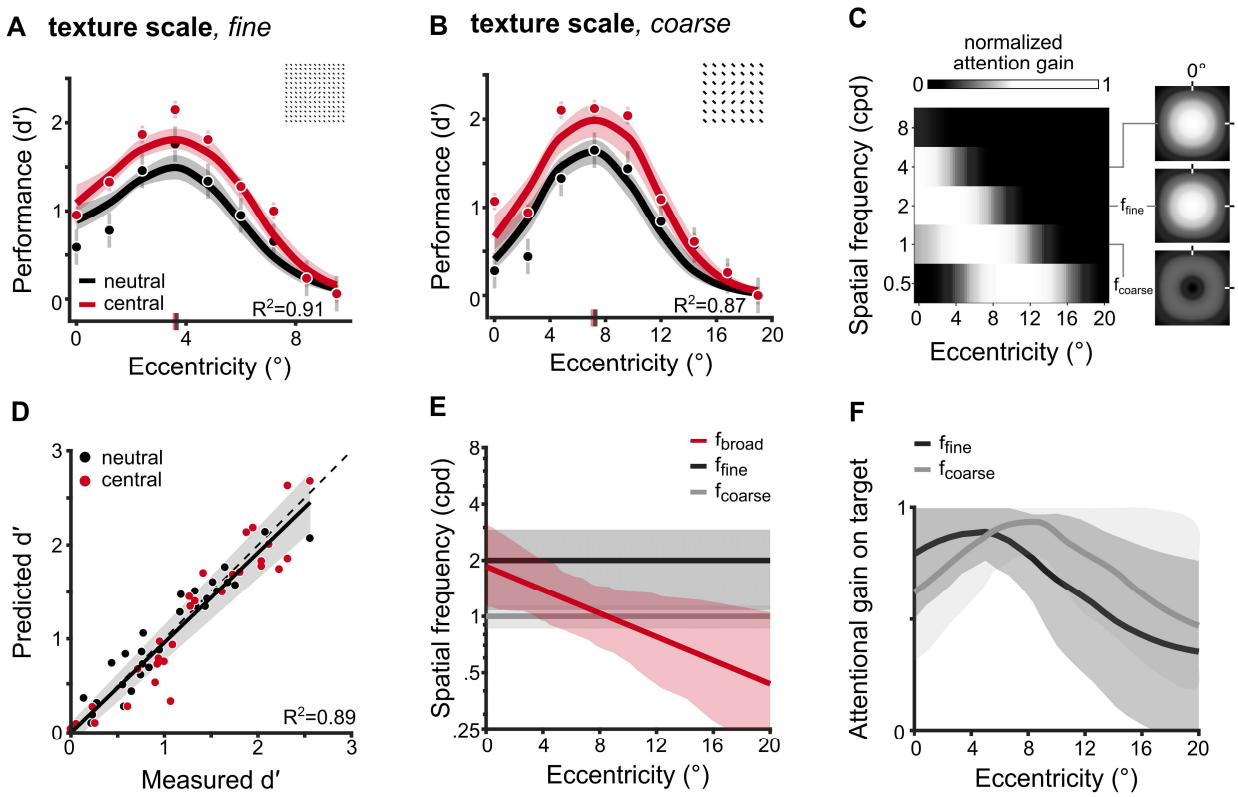


Figure 6. Broad SF enhancements yield voluntary attention effects
(**A**) Fit to Experiment 3 in (Yeshurun et al., 2008b), which used a fine-scale texture. The dots depict group-average performance and error bars denote ±1 SEM. Solid lines indicate the median and shaded regions indicate 68% confidence intervals of the bootstrapped distribution of model fits. The model simultaneously predicted performance for all four voluntary attention experiments.
(**B**) Fit to Experiment 4 in (Yeshurun et al., 2008b), which used a coarse-scale texture. This fit was determined simultaneously with that shown in *a*. Visualization follows the conventions in *a*.
(**C**) Best-fitting broad gain profile. Plotted as in **Figure 5C**.
(**D**) Goodness-of-fit for simultaneous fits to performance on all neutral and central cueing conditions. Each dot corresponds to the measured (x-axis) and predicted (y-axis) behavioral performance at a given eccentricity. Plotted as in **Figure 4C**.
(**E**) Average SF preference of the broad attentional gain profile (f_{broad}) and the SF that best distinguished target-present from target-absent stimuli for fine- (f_{fine}) and coarse-scale textures (f_{coarse}). Solid lines indicate the median and the shaded areas denote 68% confidence intervals of bootstrapped distributions.
(**F**) Normalized magnitude of attentional gain on the fine- and coarse-scale target spatial frequency across eccentricity (median and 68% confidence interval of bootstrapped distribution).

Different SF gain profiles govern involuntary and voluntary attention effects

We directly assessed the necessity of different SF gain profiles—narrow or broad—for generating the effects of involuntary and voluntary attention. In addition, we compared the efficacy of SF-tuned gain against a model wherein the spatial extent of attention varied across experiments while the gain across SF was uniform. The spatial spread of attention is a key factor of the NMA (Reynolds and Heeger, 2009), which posits that its extent relative to the stimulus size helps reconcile apparent discrepancies between involuntary and voluntary attention effects on contrast sensitivity. These predictions have been empirically tested and confirmed (Herrmann et al., 2010). We compared each of these models and measured their performance using ΔAIC (**Figure 7**).

Tuned SF gain modulation reproduced the effects of attention. The spatial extent alone was insufficient to capture the effects of either involuntary (M=21.2, CI=[14.9 27.5]; **Figure 7A**) or voluntary attention (M=12.6, CI=[5.9 19.3]; **Figure 7B**). SF-tuned gain was required. For involuntary attention, the narrow profile outperformed the broad profile (M=27.1, CI=[21.7 32.7]; **Figure 7A**). For voluntary attention, the broad profile outperformed the narrow profile (M=24.2, CI=[18.1 30.2]; **Figure 7B**). Decrements in model performance manifested as an inability to capture impairments or improvements at eccentricities demarcating the CPD (**Supplementary Figure 5**). Thus, these model comparisons corroborated psychophysical measurements (Jigo and Carrasco, 2020): involuntary and voluntary attention effects are best explained by different attentional gain profiles across SF.

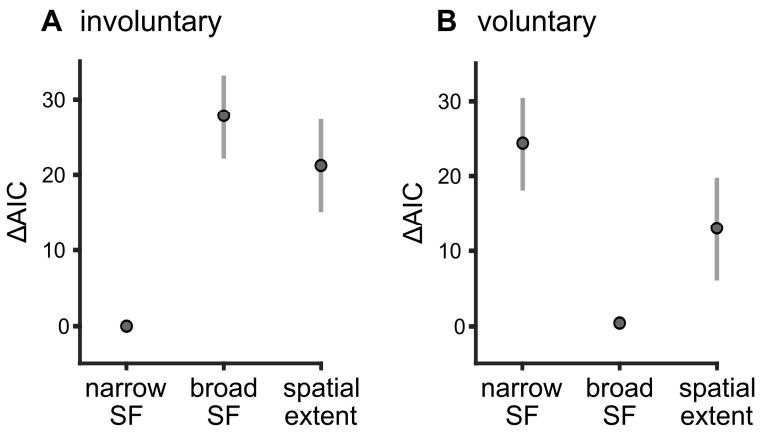


Figure 7. Different SF gain profiles govern involuntary and voluntary attention effects
AIC model comparison of different regimes of attentional modulation for (**A**) involuntary attention and (**B**) voluntary attention. Dots represent median and error bars denote 95% confidence interval of the bootstrap distribution. For involuntary attention, models were compared to the narrow SF profile. For voluntary attention, models were compared to the broad SF profile.

A parsimonious explanation for several experimental manipulations

Figure 8 depicts behavioral data for a variety of experimental manipulations. Whereas we focus on the impact of texture scale in **Figure 5 and 6**, the model is general. It simultaneously accounted for multiple target locations (vertical, **Figure 8A**; horizontal, **Figure 8C-E**; and intercardinal meridians, **Figure 8F**), behavioral tasks (orientation discrimination, **Figure 8B**; **Methods, Stimulus generation**) and attentional manipulations (cue size, **Figure 8C**). Although the model was fit using texture images with fixed positions and orientations (**Figure 3**), it behaved similarly for textures with randomly jittered elements (**Supplementary Figure 6**). Therefore, the proposed model provides a parsimonious explanation for and a quantitative match to human segmentation performance (**Figure 8**) and contrast sensitivity (**Supplementary Figure 1**).

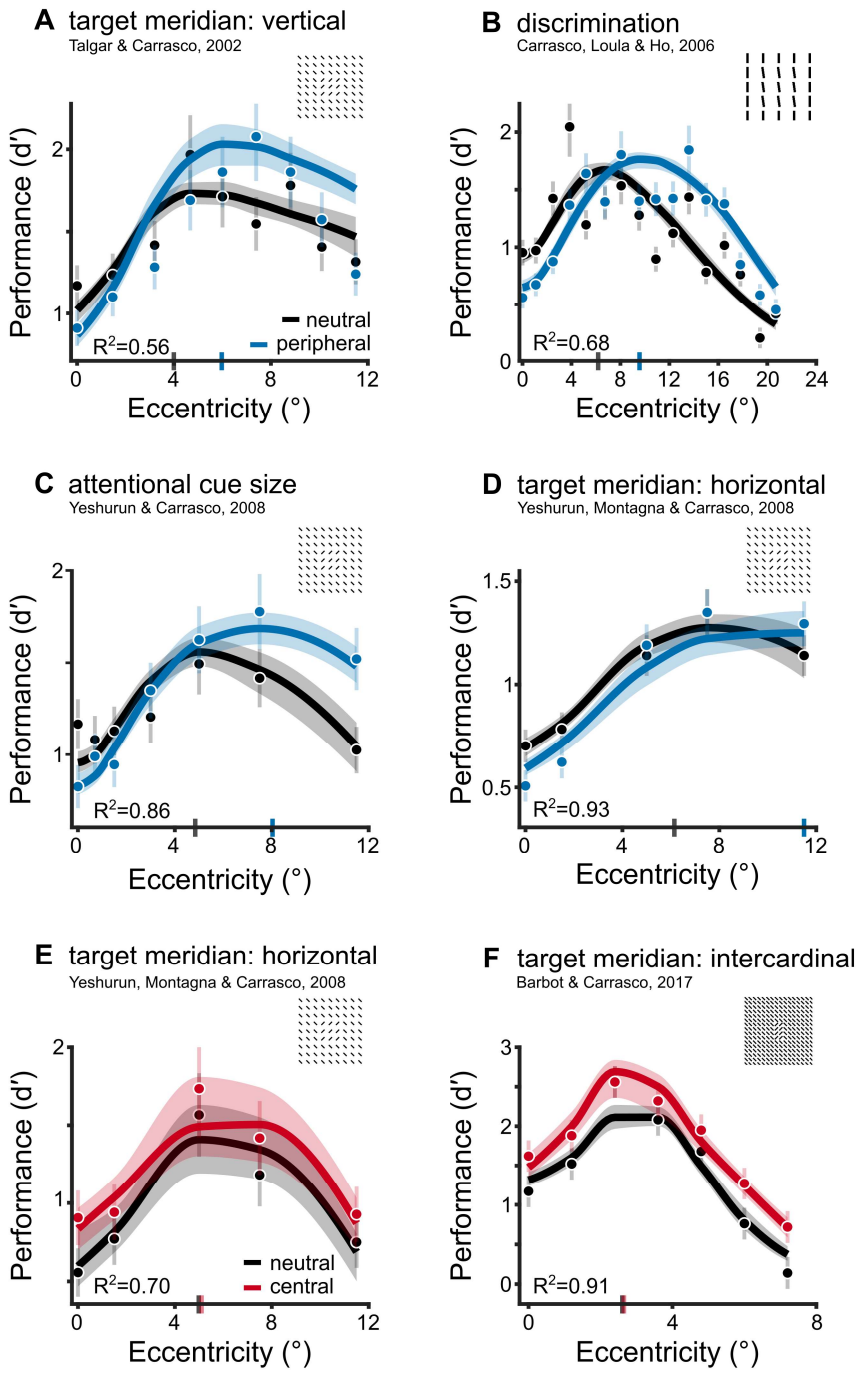


Figure 8. A parsimonious explanation for several experimental manipulations (**A-D**) Fits to involuntary attention experiments that used a narrow gain profile, determined simultaneously with data displayed in **Figure 5**. (**E-F**) Fits to voluntary attention experiments that used a broad gain profile, determined simultaneously with data displayed in **Figure 6**. Dots depict group-average and error bars indicate ±1 SEM. Solid lines indicate the median and shaded regions indicate 68% confidence intervals of the bootstrapped distribution of model fits. Labels above each panel match the labels in **Figure 3**, which describe the experimental manipulation used in the corresponding study.

DISCUSSION

We developed an image-computable model that reproduces human texture segmentation performance and reveals how involuntary and voluntary attention differentially modulate spatial resolution. This model links neural computations to three visual phenomena. (i) Divisive normalization and spatial summation mediate the CPD (Barbot and Carrasco, 2017; Carrasco et al., 2006; Jigo and Carrasco, 2018; Kehrer, 1989; Morikawa, 2000; Talgar and Carrasco, 2002; Yeshurun and Carrasco, 1998, 2000, 2008; Yeshurun et al., 2008b). (ii) Narrow high-SF enhancement drives involuntary attention effects (Carrasco et al., 2006; Jigo and Carrasco, 2018; Talgar and Carrasco, 2002; Yeshurun and Carrasco, 1998, 2000, 2008; Yeshurun et al., 2008b). (iii) Broad SF gain drives voluntary attention effects (Barbot and Carrasco, 2017; Jigo and Carrasco, 2018; Yeshurun et al., 2008b).

Normalization models of attention have described how spatial attention affects neural responses and behavior (Boynton, 2009; Ni et al., 2012; Reynolds and Heeger, 2009). Our model adopts the same algorithm specified by the Reynolds-Heeger normalization model of attention (Reynolds and Heeger, 2009) (NMA)—attentional gain modulates the stimulus drive before divisive normalization. Predictions by NMA have been empirically confirmed with psychophysical experiments (Herrmann et al., 2010). These experiments equated seemingly distinct effects of voluntary and involuntary attention on contrast sensitivity by manipulating and accounting for the spatial extent of attention.

Here, we demonstrate a critical limitation of extant models of attention. Their predictions do not extend to the differential effects on spatial resolution and cannot explain the dissociation between voluntary and involuntary attention. We find that the spatial extent of attention alone cannot reproduce the disparate effects of each attention type. This result is consistent with the empirical finding that manipulating the spread of attention during texture segmentation does not yield the typical effects of involuntary and voluntary attention (Yeshurun and Carrasco, 2008). We find that distinct SF gain profiles are required. These results reveal a computational dissociation between involuntary and voluntary attention that substantiates their differential impact on sensory processing.

To capture these differential effects by involuntary and voluntary attention across eccentricity we implemented: (i) Eccentricity-dependent and SF-tuned multiplicative gains that emulate neural (DeValois and DeValois, 1990) and psychophysical (Graham, 1989) SF selectivity. (ii) Spatial summation, which emphasizes textural contours (Bergen and Adelson, 1988; Landy, 2013; Landy

535 and Graham, 2004). (iii) Distinct SF gain profiles (Jigo and Carrasco, 2020) that scale responses
536 prior to normalization (Reynolds and Heeger, 2009), thereby adjusting the balance between fine
537 and coarse-scale visual sensitivity.
538
539 The model links SF-tuned attentional gain to changes in spatial resolution that are well-established
540 (Barbot and Carrasco, 2017; Carrasco et al., 2006; Jigo and Carrasco, 2018; Talgar and Carrasco,
541 2002; Yeshurun and Carrasco, 1998, 2000, 2008; Yeshurun et al., 2008b) (reviews (Anton-
542 Erxleben and Carrasco, 2013; Carrasco, 2011)) but hitherto unexplained in a computational model.
543 Previous explanations attribute the impact of involuntary attention to an automatic increase in
544 spatial resolution (Carrasco et al., 2006; Jigo and Carrasco, 2018; Talgar and Carrasco, 2002;
545 Yeshurun and Carrasco, 1998, 2000, 2008; Yeshurun et al., 2008b) (reviews (Anton-Erxleben and
546 Carrasco, 2013; Carrasco, 2011)) along with concomitant costs in temporal resolution (Yeshurun
547 and Levy, 2003). These spatial and temporal tradeoffs have been related to an engagement of the
548 spatially-fine but temporally-coarse sampling of parvocellular neurons (Yeshurun, 2004; Yeshurun
549 and Levy, 2003). Psychophysical experiments further demonstrate that sensitivity to high SFs
550 mediates the attentional enhancement of spatial resolution (Carrasco et al., 2006). Here, we
551 specify the computations that underlie these changes in resolution and corroborate empirical
552 evidence that narrow high-SF enhancement improves spatial resolution.
553
554 The involuntary attentional gain in our model closely matched empirical measurements made with
555 grating stimuli (Jigo and Carrasco, 2020). These measurements demonstrated that involuntary
556 attention operates above intrinsic (i.e., baseline) SF preferences. The fact that our model was
557 constrained by texture stimuli yet produced a similar preference for higher SFs, highlights the
558 hallmark inflexibility of involuntary attention (Carrasco et al., 2006; Giordano et al., 2009; Jigo and
559 Carrasco, 2018; Talgar and Carrasco, 2002; Yeshurun and Carrasco, 1998, 2000, 2008; Yeshurun
560 et al., 2008b) (review (Carrasco, 2011)). In addition to involuntary attention effects being invariant
561 to cue predictability (Giordano et al., 2009) and sometimes detrimental to perception (Carrasco et
562 al., 2006; Jigo and Carrasco, 2018; Talgar and Carrasco, 2002; Yeshurun and Carrasco, 1998,
563 2000, 2008; Yeshurun et al., 2008b), our modeling suggests that involuntary attention is also
564 invariant to the type of stimulus being attended.
565
566 The impact of voluntary attention has been attributed to an optimization of resolution for the
567 attended texture (Barbot and Carrasco, 2017; Jigo and Carrasco, 2018; Yeshurun et al., 2008b)
568 (reviews (Anton-Erxleben and Carrasco, 2013; Carrasco, 2011)). We propose that voluntary
569 attention achieves this via a broad SF gain modulation. At central eccentricities, this gain profile

enhances SFs lower than baseline preferences. Our proposal seemingly conflicts with an earlier explanation that voluntary attention suppresses sensitivity to high SFs to improve texture segmentation (Barbot and Carrasco, 2017). However, suppressed high-SF sensitivity at foveal locations would decrease cross-frequency suppression (Petrov et al., 2005; Sagi and Hochstein, 1985) and result in an effective dominance of lower SFs, which is compatible with our findings. Importantly, the model's SF preference during texture segmentation differed from those measured with gratings (Jigo and Carrasco, 2020). This discrepancy suggests that the impact of voluntary attention may depend on the properties of the attended stimulus and the nature of the task, consistent with the notion of a flexible voluntary attentional mechanism (Barbot and Carrasco, 2017; Giordano et al., 2009; Jigo and Carrasco, 2018; Yeshurun et al., 2008b).

In addition to distinguishing between voluntary and involuntary attention, the model also resolves the role of divisive normalization in texture segmentation. Previous studies demonstrate that when the pool of SFs contributing to normalization is restricted, the CPD is attenuated (Barbot and Carrasco, 2017; Carrasco et al., 2006; Morikawa, 2000). However, existing models of the CPD (Kehrer, 1997; Kehrer and Meinecke, 2003) relate the phenomenon solely to an increase in receptive field size with eccentricity. Our model directly links the summation area of receptive fields to their SF tuning. Consequently, the dominant summation area increases with eccentricity as SF preferences decrease. Without accounting for the surrounding context via normalization, we could not reliably capture the CPD. Thus, we show that normalization is critical for segmentation.

Although we are agnostic as to the biophysical implementation of our model components, each is motivated by and emulate known properties of the human and non-human primate visual system. The stimulus drive simulates bottom-up responses of phase-invariant complex cells in V1 (Adelson and Bergen, 1985) that vary with SF and eccentricity (DeValois and DeValois, 1990; Graham, 1989). The model's response to texture is generated through pooling bottom-up inputs, consistent with the gradual emergence of texture selectivity along the visual hierarchy (El-Shamayleh and Movshon, 2011; Hallum et al., 2011; Larsson et al., 2006).

In the model, attention alters sensitivity to contrast and texture differently across SF. Neural measurements of involuntary attention are sparse. It modulates activity in visuo-occipital areas and beyond (Busse et al., 2008; Dugué et al., 2018, 2020; Liu et al., 2005; Wang et al., 2015). Transcranial magnetic stimulation (TMS) of early visual cortex reveals that its activity plays a key role in the generation of involuntary attention effects (Fernández and Carrasco, 2020). However,

604 future studies are required to determine how the SF gain modulation we report manifests in neural
605 populations.
606
607 In contrast to involuntary attention, it is well-established that voluntary attention modulates cortical
608 responses (Anton-Erxleben and Carrasco, 2013; Carrasco, 2011; Dugué et al., 2018, 2020;
609 Fischer and Whitney, 2009; Klein et al., 2014; Pestilli et al., 2011; Poort et al., 2012; Reynolds and
610 Chelazzi, 2004; Womelsdorf et al., 2006; Womelsdorf et al., 2008). During texture segmentation
611 tasks, voluntary attention selectively enhances V1 and V4 responses to the embedded figure with
612 little modulation of its surrounding boundaries, suggesting that attention spreads across the target
613 object to facilitate its segmentation (Poort et al., 2012). Our model provides complementary
614 evidence –voluntary attention optimizes SF sensitivity to improve segmentation across texture
615 scale. Yet, it is unclear how neural activity generates these SF modulations. Neuroimaging (Fischer
616 and Whitney, 2009; Klein et al., 2014) and electrophysiological (Womelsdorf et al., 2006;
617 Womelsdorf et al., 2008) recordings demonstrate that spatial tuning profiles are altered by
618 voluntary attention. Such changes are consistent with, but not necessary for, the modulations of SF
619 sensitivity we report.
620
621 Our model explains how attention shapes spatial resolution across eccentricity. Few computational
622 models have implemented possible ways in which attention alters spatial resolution. Some have
623 proposed that attention modifies how finely a spatial region is analyzed. Such changes are either
624 driven by an attention field that adjusts the spatial profile of receptive fields (Womelsdorf et al.,
625 2008) or by attracting receptive fields toward and contracting them around the attended location
626 (Baruch and Yeshurun, 2014). Other models suggest an attentional prioritization that selectively
627 tunes responses for a given spatial location and attenuates responses to surrounding regions
628 (Tsotsos, 2011; Tsotsos et al., 1995). However, these models neither account for differences
629 across eccentricity nor explain attentional shifts toward fine or coarse-scale details. Critically, these
630 models do not distinguish between involuntary and voluntary attention effects. In contrast to these
631 previous models, we do not propose any modifications to the structure of receptive fields. Instead,
632 we attribute changes in spatial resolution to modulations of SF, a fundamental dimension of early
633 visual processing.
634
635 Our model advances the understanding of attention as a mechanism that not only improves low-
636 level visual processes that encode elementary features (e.g., contrast, orientation, motion) but also
637 alters mid-level processes that extract structure from visual elements. Texture segmentation
638 embodies this intermediate level of visual processing and its underlying computations bridge the

gap between sensory encoding and object recognition (Landy, 2013; Landy and Graham, 2004; Roelfsema, 2006). Therefore, the distinct impact by each type of attention and their computational differences at this processing stage have broad implications for natural visual behavior.

The fact that our model operates on arbitrary images facilitates its generalization to other complex visual phenomena. For example, the model can be extended to explain disparate voluntary and involuntary attentional effects on second-order texture perception (Jigo and Carrasco, 2018; Yeshurun and Carrasco, 2000) and on temporal resolution (Hein et al., 2006; Yeshurun, 2004; Yeshurun and Levy, 2003), as well as the interaction of attention and adaptation (Barbot and Carrasco, 2017; Carrasco et al., 2006; Ling and Carrasco, 2006; Pestilli et al., 2007)

In conclusion, we quantitatively reproduce key signatures of texture segmentation (Barbot and Carrasco, 2017; Carrasco et al., 2006; Jigo and Carrasco, 2018; Kehrer, 1989; Morikawa, 2000; Talgar and Carrasco, 2002; Yeshurun and Carrasco, 1998, 2000, 2008; Yeshurun et al., 2008b) and characterize the contributions of attention to a process commonly considered ‘preattentive’ (Bergen and Adelson, 1988; Kehrer, 1997; Kehrer and Meinecke, 2003; Landy, 2013; Landy and Graham, 2004; Morikawa, 2000; Victor et al., 2017). Moreover, we reveal the neural computations that underlie how attention modifies spatial resolution (Anton-Erxleben and Carrasco, 2013; Carrasco, 2011). Attention scales sensitivity to high and/or low SFs, adjusting the balance between fine and coarse-scale spatial resolution. Involuntary attention preferentially enhances fine details whereas voluntary attention uniformly enhances fine and coarse details to optimize task performance. Because the model distinguishes between voluntary and involuntary attention, varies with stimulus eccentricity, flexibly implements psychophysical tasks and operates on arbitrary grayscale images, it provides a general-purpose tool for assessing theories of vision and attention across the visual field.

ACKNOWLEDGEMENTS

This work was supported by NIH RO1-EY019693 to M.C. We thank Jonathan Winawer, Michael Landy, Antoine Barbot, Hsin-Hung Li, as well as Antonio Fernández, Nina Hanning, Marc Himmelburg, Luke Huszar & Yong-Jun Lin and other members of the Carrasco lab for their helpful comments and discussion.

AUTHOR CONTRIBUTIONS

M.J.: Conceptualization, Methodology, Software, Formal analysis, Investigation, Writing – Original Draft, Writing – Review & Editing, Visualization. **D.J.H.:** Conceptualization, Methodology, Resources, Writing – Review & Editing. **M.C.:** Conceptualization, Methodology, Investigation, Resources, Writing – Review & Editing, Supervision, Project administration, Funding acquisition.

METHODS

Model

Our model simulates the response of a collection of receptive fields (RFs) each narrowly tuned to spatial position (*x,y*), orientation (θ) and SF (*f*). Responses varied with eccentricity (α). The population response (*R*) output results from four components: the stimulus drive (*E*), attentional gain (*A*), suppressive drive (S and σ), and spatial summation (*F*), where * represents convolution:

$$R(f,\theta,x,y) = \frac{E(f,\theta,x,y)A(f,\alpha)}{\sigma^2(\alpha) + S(f,\theta,x,y)} * F(f) \quad (1)$$

All model parameters are given in **Table 1**.

Table 1. Model parameters. Bolded entries indicate a model component.

Parameter	**Description**
Stimulus drive	
t_T	peak SF ($\log_2$-cpd)
t_{min}	minimum preferred SF (cpd); *fixed at 0.5*
m_T	SF tuning change across eccentricity (octaves/°)
b_T	SF bandwidth (octaves)
Contrast gain	
g_σ	gain at fovea
m_σ	slope along eccentricity
Normalization pool	
δ_f	SF pool bandwidth (octaves); *fixed at 1*
δ_θ	Orientation pool bandwidth (°); *fixed at 180*
δ_{pos}	Spatial pool width (°); *fixed at* $\frac{2}{f}$
Spatial summation	
δ_{pos}	Summation pool width (°); *fixed at* $\frac{2}{f}$

Attentional gain profile	
a_N or a_B	attentional SF tuning at the fovea ($\log_2$-cpd)
m_N or m_B	slope of SF tuning across eccentricity (octaves/°)
γ_N or γ_B	amplitude
b_N or b_B	SF bandwidth (octaves)
b_{pos}	spatial spread (°); *fixed at FWHM of 4*

687

688 Stimulus drive

689 The stimulus drive characterizes responses of linear RFs in the absence of suppression, attention
690 and spatial summation. Stimulus drive responses were determined using a steerable pyramid
691 (Simoncelli et al., 1992). The pyramid decomposed stimulus images into several SF and orientation
692 subbands. SF and orientation selectivity were defined by weighted sums of the image (i.e., linear
693 filters). The weights were parameterized by raised-cosine functions that evenly tiled SFs,
694 orientations and positions (eccentricities).

695

696 The number of SF and orientation subbands are parameters that can be flexibly chosen. We used
697 a set of 30 subbands that comprised five SF bands and six orientation bands. The number of SF
698 subbands is determined by the size of the stimulus image and the subband bandwidth. In our
699 simulations, we used images that were 160 x 160 pixels (see **Stimulus generation**), and a full-
700 width at half-maximum (FWHM) bandwidth of 1 octave, which allowed for five different SF
701 subbands. The chosen bandwidth is comparable to empirical tuning curves measured in primate
702 electrophysiological recordings (De Valois et al., 1982) and human psychophysical (Graham, 1989)
703 measurements. The FWHM orientation bandwidth (60°) of each subband is comparable to
704 empirical tuning curves measured in primate electrophysiological recordings (Ringach et al., 2002).
705 Using narrower (30°) or wider (90°) bandwidths yielded similar results supporting the same
706 conclusions.

707

708 For each SF and orientation subband, the pyramid includes RFs in quadrature phase. We
709 computed a 'contrast energy' response (Adelson and Bergen, 1985), which is the sum of squared
710 responses across phase. This response depends on the local spectral energy at each position in
711 the image for a given SF and orientation subband and is fundamental to models of texture

perception (Bergen and Adelson, 1988; Landy, 2013; Landy and Graham, 2004). The contrast energy response to SF (*f*) and orientation (θ) across space (*x*,y) is denoted as *C(f,*θ*,x,y).*

SF gain. Human (Graham, 1989; Jigo and Carrasco, 2020; Pointer and Hess, 1989) and non-human primate (DeValois and DeValois, 1990) contrast sensitivity is narrowly tuned to SF across eccentricity. SF tuning shifts from high to low SFs with eccentricity. To model this behavior, contrast energy was multiplied point-by-point by a SF gain function, *T*, defined by a log-parabola (Lesmes, 2010; Watson and Ahumada, 2005):

$$T(f,\alpha) = \exp(-[\frac{\log_2(\frac{f}{\lambda_T(\alpha)})}{b_T}]^2) \tag{2}$$

where α denotes the eccentricity of a RF and b_T determines the function's SF FWHM bandwidth. The preferred SF (λ_T) at a given eccentricity is given by:

$$\lambda_T(\alpha) = 2^{t_T - m_T\alpha} + t_{min} \tag{3}$$

Motivated by psychophysical measurements (Pointer and Hess, 1989), SF preferences converge onto a single value in the far periphery, t_{min}. As a result, the preferred SF at the fovea is given by $2^{t_T} + t_{min}$ and progressively shifts towards t_{min} at the rate m_T (octaves/degree). Whereas t_T was free to vary during simulations (see **Table 1-4**), t_{min} was fixed at 0.5 cpd because texture stimuli produced minimal contrast energy below that SF subband. Allowing t_{min} to vary yielded similar results supporting the same conclusions.

In sum, the stimulus drive (*E*) characterizes the contrast energy response of a collection of RFs whose SF preferences vary with eccentricity, computed as:

$$E(f,\theta,x,y) = T(f,\alpha)C(f,\theta,x,y) \tag{4}$$

<u>Attentional gain</u>

Attention is implemented as an attentional gain field, *A*, that multiplies the stimulus drive point-by-point as in the Reynolds-Heeger normalization model of attention (Reynolds and Heeger, 2009). Attentional gain was uniform across orientation. Across SF and spatial position, cosine window functions, *w*, characterized the distribution of gain:

$$w(z;\mu,b) = \begin{cases} 0.5 + 0.5\cos(\frac{\pi[z-\mu]}{b}) & \\ 0 & \mu - b < z > \mu + b \end{cases} \quad (5)$$

where μ defined its center and *b* defined its FWHM. The units of μ and *z* depended on the dimension: for SF each variable was in units of $\log_2$-transformed cycles per degree, and for spatial position they were in units of degrees of visual angle. Thus, the window was defined on a logarithmic axis for SF but on a linear axis for position. SF and spatial position functions were multiplied, point-by-point, to characterize the full distribution of attentional gain.

Spatial spread. Attentional gain was centered on the target location. In our simulations, the target was always placed along the horizontal meridian at eccentricity α_{targ} (see **Stimulus generation**). The spread of attention was determined by multiplying two separate cosine functions (*w*, equation 5): one that varied as a function of *x* and another as a function of *y*, each with an identical width b_{pos}. The width did not vary across eccentricity. This produced a two-dimensional spatial window A_{pos}:

$$A_{pos}(x,y) = w(x;\alpha_{targ},b_{pos})w(y;0,b_{pos}) \quad (6)$$

The precise spatial spread of attention is controversial (Yeshurun, 2019) and can change based on task demands (Herrmann et al., 2010; Stigchel et al., 2009). Critically, the spread of attention has not been explicitly manipulated during texture segmentation tasks by varying the target's spatial uncertainty. Such a protocol has been used to test predictions of the NMA and, through functional magnetic resonance imaging, has been demonstrated to adjust the size of the attention field (Herrmann et al., 2010). Instead, a previous study (Yeshurun and Carrasco, 2008) measured involuntary attention effects while manipulating the size of a peripheral pre-cue. The authors found that involuntary attention altered performance as long as the cue was the same or smaller than the target size. In our simulations, the spread of attention was fixed at a FWHM of 4° (**Table 1**). This size was chosen because it encompassed the largest target size used to constrain model parameters (**Table 5**). As a result, the spatial extent of attention was identical across eccentricity and experiments. Similar results were observed when the spread was fixed at 2° and 3°. However, in the model variant wherein the spatial extent could change (see **Model alternatives**), the FWHM of attentional spread (b_{pos}) was free to vary between experiments.

SF gain profile. Motivated by empirical findings (Jigo and Carrasco, 2020), we implemented two attentional gain profiles: narrow and broad.

Narrow profile. In the narrow model (A_N), attentional gain was bandpass across SF. Attentional gain peaked at a given SF, λ_N, and fell gradually toward neighboring frequencies within its bandwidth, b_N. The window was characterized by a cosine function as defined in equation (5):

$$A_N(f,\alpha) = w(f;\lambda_N(\alpha), b_N) \tag{7}$$

The center SF of attentional gain profiles (λ_N for narrow, λ_B for broad) varied with eccentricity:

$$\lambda_N(\alpha) = 2^{a_N - m_N\alpha} \tag{8}$$

where a_N (or a_B for the broad profile) defined the center frequency at the fovea, which gradually changed with eccentricity at the rate m_N (m_B for broad).

Broad profile. In the broad profile (A_B), attentional gain was broadband across SF, characterized by the sum of three overlapping cosine functions (w, equation 5):

$$A_B(f,\alpha) = w_1 + w_2 + w_3 \tag{9}$$

where $w_1 = w(f;\lambda_B(\alpha), b_B)$, $w_2 = w(f;\lambda_B(\alpha) - b_B, b_B)$, and $w_3 = w(f;\lambda_B(\alpha) + b_B, b_B)$. The bandwidth of each function was given by b_B. Relative to the center SF, λ_B, the adjacent functions were centered $\pm b_B$ apart, which ensured that their sum yielded a plateau that spanned $\frac{2b_B}{3}$ octaves.

In sum, attentional gain multiplicatively scaled the stimulus drive uniformly across orientation, but differently across spatial frequency and eccentricity. Overall, attentional gain was given by:

$$A(f,\alpha) = \gamma_B A_{pos} A_B \tag{10}$$

where A_{pos} and A_B (or A_N) were four-dimensional matrices that characterized the distribution of attentional gain on spatial position, spatial frequency and orientation. The strength of attention was determined by γ_B (or γ_N). To simulate the neutral cueing condition, attentional gain was set to 1 across spatial frequency, orientation and spatial position, which removed modulations by attention. In addition, to assess the explanatory power of the spatial spread of attention (see **Model alternatives**), A_B (or A_N) were set to 1 and only γ and A_{pos} varied.

Suppressive drive

The suppressive drive comprised contextual modulation, computed through pooling the attention-scaled stimulus drive (Reynolds and Heeger, 2009) across nearby spatial positions, all orientations and across neighboring SFs. This pooling procedure implemented lateral interactions between RFs that differed in their tuning profiles. Suppression magnitude was computed via convolution (Reynolds and Heeger, 2009). Convolution kernels were cosine window functions (*w*, equation 5), defined across SF, orientation and spatial position.

The bandwidth of the spatial frequency kernel, δ_f, equaled 1 octave:

$$K_f = \begin{cases} 1 & f_i - \delta_f \leq f_i \leq f_i + \delta_f \\ 0 & otherwise \end{cases} \tag{11}$$

where f_i denotes the center SF of a steerable pyramid subband. This kernel summed the squared outputs of the steerable pyramid's subbands within and ±1 octave around each SF band. Given the parameters for the steerable pyramid's subband bandwidth (1 octave, see **Stimulus drive**), this sum yielded the response of a single large filter with a FWHM of 3 octaves. Consequently, each RF was suppressed by the response to spatial frequencies within and around its tuning bandwidth.

The bandwidth of the orientation kernel, δ_θ, equaled 180°, which encompassed all orientation subbands:

$$K_\theta = \begin{cases} 1 & \theta_i - \delta_\theta \leq \theta_i \leq \theta_i + \delta_\theta \\ 0 & otherwise \end{cases} \tag{12}$$

where θ_i denotes the center orientation of a steerable pyramid subband. This kernel summed the squared outputs of the steerable pyramid's subbands across orientation. This sum yielded the response of a single large filter that spanned all orientations. As a result, each RF was suppressed by the contrast energy response to all orientations in the spatial and SF neighborhood of the RF.

Spatial position kernels were determined by multiplying two cosine windows defined in equation (5):

$$K_{pos}(x, y; f) = w(x; 0, \delta_{pos})w(y; 0, \delta_{pos}) \tag{13}$$

One window varied across x, another across y and their centers traversed across the image during convolution. The two-dimensional kernel summed to unity and therefore computed the average contrast energy within the pooled spatial area. The width of each kernel, δ_{pos}, equaled $\frac{2}{f}$ and was thus inversely proportional to subband SF *f*, which yielded two-dimensional spatial kernels, K_{pos}, which were broader for lower SFs than for higher SF subbands. Kernel widths did not vary with eccentricity.

Overall, contextual modulation was characterized via separable convolution across the attention-scaled stimulus drive:

$$S(f,\theta,x,y) = K_f * (K_\theta * (K_{pos} * [E(f,\theta,x,y)A(f,\alpha)])) \quad (14)$$

where * denotes convolution of the suppression kernels, *K*. The magnitude of suppression was adjusted across eccentricity by adding σ^2, which controlled the level of contrast at which neural responses reached half-maximum and is referred to as contrast gain. Motivated by psychophysical measurements (Jigo and Carrasco, 2020; Pointer and Hess, 1989), contrast gain was implemented as an exponential function across eccentricity:

$$\sigma^2(\alpha) = 10^{2(g_\sigma - m_\sigma \alpha)} \quad (15)$$

where g_σ and m_σ are free parameters that determine contrast gain at the fovea and the rate at which it varies with eccentricity, respectively.

<u>Spatial summation</u>

Following divisive normalization, responses were weighted and summed across space and within each SF and orientation subband. Spatial summation was accomplished via convolution by two-dimensional cosine windows, *F*, computed using equation (13). Importantly, the width of each filter scaled with subband spatial frequency such that narrow (wide) regions of space were pooled for high (low) SFs (Landy and Graham, 2004). Filter widths did not vary with eccentricity. A spatial summation stage is common in models of texture perception (Landy, 2013; Landy and Graham, 2004).

865 **Decision mechanism**

866 We used signal detection theory to relate population responses to behavioral performance (*d′*). The
867 available signal *s* for an observer was computed as the Euclidean norm of the difference between
868 target-present ($\mathbf{r_t}$) and target-absent ($\mathbf{r_n}$) neural population responses: $s = ||\boldsymbol{r_t} - \boldsymbol{r_n}||$. Performance
869 on a discrimination task is proportional to the neural responses given the assumption of additive,
870 independent and identically distributed (IID) noise. An alternative model with Poisson noise and a
871 maximum-likelihood decision rule would yield the same linkage between neural response and
872 behavioral performance (Jazayeri and Movshon, 2006; Pestilli et al., 2009). Based on signal
873 detection theory, the signal and the magnitude of noise (σ_n) were used to compute behavioral
874 performance $d' = \frac{s}{\sigma_n}$. In our simulations, $\sigma_n = \overline{r_{neutral}} \big/ \overline{s_{neutral}}$ where $\overline{r_{neutral}}$ denotes the observed
875 neutral performance averaged across eccentricity and $\overline{s_{neutral}}$ denotes the average signal across
876 eccentricity for the neutral condition. This ratio scaled the model's predicted behavioral
877 performance to match the observed data.

878

879 **Model fitting**

880 Models were optimized by minimizing the residual sum of squared error (SSE) between model and
881 behavioral *d′* using Bayesian adaptive direct search (BADS (Acerbi and Ma, 2017)). This
882 optimization procedure is well-suited for the number of parameters and cost function evaluation
883 time of the model. Behavioral data were extracted from the original study publications. When
884 applicable, performance data for a psychophysical experiment were converted from proportion
885 correct, *p*, to *d′* with the assumption of no interval bias (Macmillan and Creelman, 2005): $d' =$
886 $\sqrt{2}z(p)$ where *z* denotes the inverse normal distribution. Although performance on 2IFC tasks can
887 exhibit biases between intervals (Yeshurun et al., 2008a), our conversion algorithm operated
888 uniformly across eccentricity, which preserved the performance variation (i.e., the CPD) that is
889 critical for the goals of this study.

890

891 **Optimization strategy**

892 <u>Central performance drop</u>

893 To address whether normalization is critical for eliciting the CPD, the model was first fit to
894 performance on all neutral conditions across the 10 texture segmentation experiments (103 data
895 points). Peripheral and central cueing conditions (for involuntary and voluntary attentional
896 conditions, respectively) were excluded to isolate the CPD. 15 free parameters fit all 103 data
897 points (**Table 2**).

898

Table 2. Free parameters for the fits to the neutral condition of all ten psychophysical experiments. The mapping between the experiment labels (a-j) and the respective references is given below the table; italicized text describe the manipulation conducted in a given experiment. Bold values indicate the median and values within square brackets denote the 95% CI of the bootstrapped distribution of fitted parameters. min. = minimum; SF = spatial frequency; bw = bandwidth.

		Contrast gain		Stimulus drive		
parameter		g_σ	m_σ	t_T	m_T	b_T
description		*min.*	*slope*	*SF peak*	*SF slope*	*SF bw*
Experiment	a	**2.3** [2.2 2.5]	**-0.09** [-0.09 -0.08]	**1.9** [1.5 2.4]	**-0.8** [-0.9 -0.6]	**1.7** [1.6 2.0]
	b	**2.3** [2.1 2.5]				
	c	**2.7** [2.5 2.8]				
	d	**2.7** [2.6 2.8]				
	e	**2.2** [2.0 2.5]				
	f	**2.7** [2.4 2.8]				
	g	**1.9** [1.5 2.3]		**2.8** [2.4 3.1]		
	h	**1.8** [1.7 2.0]				
	i	**2.6** [2.5 2.8]				
	j	**2.7** [2.6 2.8]				

[a] (Yeshurun and Carrasco, 1998). *Fine-scale texture; experiment 1.*
[b] (Yeshurun and Carrasco, 1998). *Coarse-scale texture; experiment 2.*
[c] (Talgar and Carrasco, 2002). *Target meridian: lower vertical*.
[d] (Carrasco et al., 2006). *Orientation discrimination: baseline adaptation*.
[e] (Yeshurun and Carrasco, 2008). *Attentional cue size: cue size 1.*
[f] (Yeshurun et al., 2008b). *Target meridian: horizontal; experiment 2*.
[g] (Yeshurun et al., 2008b). *Target meridian: horizontal; experiment 1*
[h] (Yeshurun et al., 2008b). *Fine-scale texture; experiment 3.*
[i] (Yeshurun et al., 2008b). *Coarse-scale texture; experiment 4.*
[j] (Barbot and Carrasco, 2017). *Target meridian: intercardinal; baseline adaptation.*

Ten separate free parameters independently controlled the minimum contrast gain at the fovea (g_σ; equation 15) for each of the 10 experiments. These parameters accounted for possible differences in overall contrast sensitivity that can arise from variable display properties (e.g., mean luminance) among experimental protocols (reviews (DeValois and DeValois, 1990; Graham, 1989)).

Sensitivity to contrast and SF varies for stimuli placed at isoeccentric locations around the visual field; it is higher at the horizontal meridian and decreases gradually towards the vertical meridian (Barbot et al., 2021; Cameron et al., 2002; Carrasco et al., 2001; Himmelberg et al., 2020). Whereas 5 out of 6 involuntary attention experiments used texture targets placed on the horizontal meridian, 3 out of 4 voluntary attention experiments used targets presented along the diagonal meridians (**Table 5**). Because SF selectivity depends on stimulus polar angle, two parameters separately determined the highest preferred spatial frequency (t_T; equation 3)—one was shared among involuntary attention experiments and the other was shared among voluntary attention experiments. Alternatively, we could have fit separate parameters for horizontal, vertical and diagonal meridians. However, this approach would have added a third free parameter, reducing the parsimony of the model. The configuration we used (i.e., separate parameters for involuntary and voluntary attention experiments; two parameters total) yielded reasonably good fits.

The remaining three parameters were shared among the neutral conditions for all experiments. Each controlled the bandwidth (b_T; equation 2) of the tuning function T, the gradual shift toward lower SFs with eccentricity (m_T; equation 3) and the increase in contrast gain across eccentricity (m_σ; equation 15).

<u>Attentional modulation</u>

To assess whether narrow or broad attentional gain profiles generated the effects of attention, the model was fit separately to involuntary and voluntary attention experiments. Importantly, a pair of cueing conditions—neutral and valid—were tested in each experiment, as the neutral condition provided a baseline from which the effects of attention were measured. We preserved the link between the pair by jointly fitting the model to neutral and valid conditions of each experiment. We refer to the valid condition for involuntary attention as “peripheral” and as “central” for voluntary attention to match the different cues used in each set of experiments.

Involuntary attention. All six involuntary attention experiments were fit simultaneously (146 data points) with fourteen free parameters (**Table 3**). Minimum contrast gain at the fovea (g_σ; equation 15) was determined independently for each of six experiments, yielding six free parameters. The

remaining eight parameters were shared among all experiments. Four determined the stimulus drive: its bandwidth (b_T; equation 2), the highest preferred SF at the fovea (t_T; equation 3), the shift to lower SFs with eccentricity (m_T; equation 3) and the slope of contrast gain across eccentricity (m_σ; equation 15). The remaining four controlled the narrow SF attentional gain profile, specifically its bandwidth (b_N; equation 7), center SF (λ_N; equation 8), the shift to lower SFs with eccentricity (m_N; equation 8), and its amplitude (γ_N; equation 10).

Table 3. Free parameters for the fits to the neutral and peripheral cueing conditions of the six involuntary attention experiments. Bold values indicate the median and values in square brackets depict the 95% CI of the bootstrapped distribution of fitted parameters. min. = minimum; SF = spatial frequency; bw = bandwidth; amp. = amplitude.

		Contrast gain		**Stimulus drive**			**Narrow gain profile**			
	parameter	g_σ	m_σ	t_T	m_T	b_T	a_N	m_N	b_N	γ_N
	description	*min.*	*slope*	*SF peak*	*SF slope*	*SF bw*	*SF peak*	*SF slope*	*SF bw*	*amp.*
Experiment	a	**1.9** [1.7 2.2]	**-0.05** [-0.07 -0.04]	**2.1** [1.8 2.8]	**-0.7** [-0.9 -0.5]	**1.6** [1.4 2.1]	**2.3** [1.8 3.0]	**-0.2** [-0.3 -0.1]	**2.9** [1.5 4.8]	**4.3** [3.0 7.8]
	b	**1.7** [1.5 2.1]								
	c	**2.6** [2.2 2.8]								
	d	**2.1** [2.8 2.5]								
	e	**1.5** [1.5 2.0]								
	f	**2.0** [1.7 2.7]								

Voluntary attention. All four voluntary attention experiments were fit simultaneously (60 data points) with 12 free parameters (**Table 4**). Minimum contrast gain at the fovea (g_σ; equation 15) was determined independently for each experiment, which yielded four free parameters. The remaining eight parameters were shared among experiments, as described above for involuntary attention.

Table 4. Free parameters for the fits to the neutral and central cueing conditions of the four voluntary attention experiments. Bold values indicate the median and values within square brackets depict the 95% CI of the bootstrapped distribution of fitted parameters. min. = minimum; SF = spatial frequency; bw = bandwidth; amp. = amplitude.

		Contrast gain		**Stimulus drive**			**Broad gain profile**			
parameter		g_σ	m_σ	t_T	m_T	b_T	a_B	m_B	b_B	γ_B
description		*min.*	*slope*	*SF peak*	*SF slope*	*SF bw*	*SF peak*	*SF slope*	*SF bw*	*amp.*
Experiment	g	**1.6** [1.5 2.1]	**-0.09** [-0.1 -0.06]	**3.4** [2.8 4.0]	**-0.8** [-1.0 -0.5]	**1.3** [1.1 1.6]	**1.66** [0.5 2.3]	**-0.2** [-0.3 -0.01]	**3.0** [1.4 4.4]	**3.1** [2.2 6.2]
	h	**1.7** [1.5 1.9]								
	i	**2.5** [2.2 2.8]								
	j	**1.7** [1.5 1.9]								

Model alternatives

To assess whether contextual modulation, via normalization, and spatial summation are critical for generating the CPD, we implemented five model variants. The impact of contextual modulation was assessed by iteratively removing individual components of the suppressive drive: cross-orientation suppression ('-θ'), cross-frequency suppression ('-f'), surround suppression ('-x,y') and all components simultaneously ('-context'). In a separate variant, spatial summation was removed ('-sum'). We fit each variant separately to neutral performance data from all ten psychophysical experiments using the configurations described in **Optimization strategy, Central performance drop**.

In the '-context' model, each RF was suppressed by its own response, simulating an extremely narrow suppressive pool. Specifically, the extent of suppressive pools (δ_f, δ_θ, δ_{pos}; equations 11-13) were set to 0. As a result, the contributions of surround, cross-orientation and cross-frequency suppression were absent. The other contextual modulation variants only had a single parameter set to 0 (e.g., δ_f for cross-frequency suppression).

The '-sum' variant removed spatial summation (i.e., *F* in equation 1) from the model. As a result, normalized responses were not pooled across space and the response to regions of homogenous orientation were attenuated.

We additionally compared the efficacy of each attentional gain profile across SF—narrow or broad—in generating the effects of involuntary and voluntary attention. To this end, each profile was fit to involuntary and voluntary attention experiments. To assess the explanatory power of the spatial extent of attentional gain, a third attention model was compared. In this model, the spatial extent of attention (b_{pos}, equation 6) could vary between experiments and the gain across SF was uniform. Each model fit followed the configurations described in **Parameter configuration, Attentional modulation**.

Model comparisons

We compared models using the Akaike Information Criterion (Akaike, 1974) (AIC), computed with the assumption of a normal error distribution. The difference in AIC values between model variants indexed model performance. '-θ', '-f', '-x,y', '-context' and '-sum' models were compared to the full model; positive values reflect the full model outperforming the variants. Additionally, narrow and broad SF gain profiles as well as the spatial extent model were compared; positive values reflect a decrease in model performance.

Stimulus generation

Target-present and target-absent textures were re-created to match the stimulus parameters used in each psychophysical study (**Table 5**). For all stimuli, each pixel subtended 0.03125° (i.e., 32 pixels/°), roughly matching the spatial resolution of a 1280 × 960 monitor display placed 57 cm away from the observer. Although these monitor display parameters were not identical to those used in each psychophysical study, they are typical of visual psychophysics experiments.

The full texture stimulus used in each experiment typically spanned the entire display. We generated 5°-wide square cutouts of the texture stimulus, centered on the target location. The target was displayed along the vertical meridian in (Talgar and Carrasco, 2002) to assess the impact of asymmetries in visual function between the cardinal meridians (Barbot et al., 2021; Cameron et al., 2002; Carrasco et al., 2001; Himmelberg et al., 2020). Additionally, targets were displayed at the diagonal (intercardinal) meridian in (Barbot and Carrasco, 2017) and in experiments 3 and 4 in (Yeshurun et al., 2008b) to avoid the cardinal asymmetries. All other experiments presented targets along the horizontal meridian (**Table 5**). Because the model implemented visual sensitivity that varied with eccentricity, but was uniform at isoeccentric locations, all targets were assumed to be presented along the horizontal eccentricity for simplicity (as in equation 6).

Each texture array was composed of lines oriented 135°. For target-present textures, the target comprised a patch of lines that were oriented 45°. The exception was for the (Carrasco et al., 2006) study, in which the texture array comprised vertical lines (0°) and the target patch contained lines tilted ±8° (**Figure 3**). In this study, observers' performed an orientation discrimination task wherein they reported the orientation of the target patch presented on each trial. To simulate orientation discrimination performance, the target-present and target-absent stimuli that were input to the model always contained a patch but their orientation differed.

To avoid edge artifacts, texture stimuli were windowed by the sum of three cosine window functions (as in equation 8) centered on the target that produced a uniform plateau covering the central 3.75 deg and fell off with cosine edges. Pixel intensities in each stimulus were constrained between 0 and 1.

To maximize model discriminability and minimize random fluctuations in model output, textures used to fit the model were generated without spatial or orientation jittering. In additional simulations, the stimuli of two representative experiments were jittered. The stimuli for experiment 1 of (Yeshurun and Carrasco, 1998) were spatially jittered (0.3 deg jitter), and the stimuli in experiment 4 of (Yeshurun et al., 2008b), were jittered spatially (0.34 deg jitter) and in orientation (55° bandwidth). Parameters of the jitter were compatible with those specified in each study.

Table 5. Stimulus parameters for each texture segmentation experiment. Experiment numbers match those described in **Table 2**.

Experiment	**Line spacing**: horz × vert	**Line size**: width × height	**Target size**: width × height	**Target meridian:**
a	0.68° × 0.71°	0.1° × 0.4°	1.97° × 2.03°	horizontal
b	1.36° × 1.42°	0.2° × 0.8°	3.78° × 3.84°	horizontal
c	0.71° × 0.68°	0.1° × 0.4°	2.03° × 1.97°	vertical
d	1.43° × 1.37°	0.2° × 1°	4.09° × 3.97°	horizontal
e	0.68° × 0.71°	0.1° ×0.7°	2.28° × 2.34°	horizontal
f	0.68° × 0.71°	0.1° × 0.7°	2.28° × 2.34°	horizontal
g	0.68° × 0.71°	0.1° × 0.7°	2.28° × 2.34°	horizontal

h	0.46° × 0.46°	0.1° × 0.2°	1.34° × 1.34°	diagonal
i	0.91° × 0.91°	0.2° × 0.4°	2.47° × 2.47°	diagonal
j	0.4° × 0.4°	0.1° × 0.24°	1.03° × 1.84°	diagonal

Resampling procedures

We obtained confidence intervals on the parameter estimates, model predictions and AIC values by bootstrapping the data and refitting the model 100 times per configuration (**Optimization strategy**) and for each model variant (**Model alternatives**). Bootstrap samples were generated by drawing random samples from Gaussian distributions. Gaussians were centered on group-average performance at a given eccentricity and the standard error of the mean for each study defined the width of the distribution. The model was fit to each resampled dataset, which yielded 100 fits of resampled data. Confidence intervals on parameter estimates (**Tables 2-4**), model fits at each eccentricity (**Figures 4-6, 8, Supplementary Figures 2, 5 & 6**) and AIC values (**Figures 4 & 7**) were calculated from the bootstrapped estimates.

To generate bootstrap samples for simulations with jittered texture stimuli, the model was first fit to the data for each experiment using a non-jittered texture. Then, the model parameters were fixed and jittered stimuli were input to the model. One hundred unique jittered stimuli were presented to the model, which yielded 100 estimates of texture segmentation performance at each eccentricity.

Predicted contrast sensitivity functions

Contrast sensitivity functions were generated from the model to assess their compatibility with two datasets of psychophysical measurements (Jigo and Carrasco, 2020; Pointer and Hess, 1989). These datasets were chosen as they measured contrast sensitivity across a wide range of spatial frequencies and eccentricities.

Two important differences exist between these two datasets. First, (Jigo and Carrasco, 2020) used a fixed grating size (4° wide, cosine window) for all SFs, whereas (Pointer and Hess, 1989) adjusted grating size so that 3.2 periods of each SF were contained in a Gaussian window. As a result, gratings of low SF were larger than gratings of high SF. Second, (Jigo and Carrasco, 2020) used an orientation discrimination task to measure contrast sensitivity, whereas (Pointer and Hess, 1989) used a detection task.

The flexibility of our model allowed us to simulate contrast sensitivity for these different experiments. Gratings that matched the stimulus properties of each study were input to the model. To simulate the signal available to an observer in a detection task, the population response to a grating was compared to the population response to a uniform gray image. For the discrimination task, the difference in population response between orthogonal gratings determined the available signal.

For each grating SF, model population responses were evaluated for 0-20 degrees of eccentricity and at 9 levels of contrast that were interpolated to characterize the full contrast response function at a given eccentricity. Interpolation was used to reduce computational load; similar contrast response functions were generated when the model was evaluated at finer contrast steps. The available signal was scaled by internal noise magnitude to yield model discriminability (**Decision mechanism**). Because internal noise varies with spatial frequency (Silvestre et al., 2018), the available signal across eccentricity was scaled so that the maximum *d'* for each SFat the fovea equaled 2. Contrast thresholds were then determined as the level of contrast needed to reach *d'*=1. The inverse of contrast threshold served as our index of contrast sensitivity.

To generate contrast sensitivity estimates, we randomly sampled (with replacement) from the full distribution of best-fitting parameters for the fits to the neutral condition across all ten texture segmentation experiments (**Table 2**). Parameters were resampled 100 times, resulting in 100 different contrast sensitivity estimates for each SF and eccentricity.

Cross-validation procedure

To characterize how the operating range of involuntary and voluntary attention varied with eccentricity, relative to baseline tuning preferences (**Supplementary Figures 3 and 4**), we fit polynomials to empirical measurements made by (Jigo and Carrasco, 2020). Leave-one-subject-out cross-validation determined the best-fitting polynomial order. Specifically, the ratio, in octaves, between the peak SF of the neutral contrast sensitivity function and the preferred SF of attentional modulation were computed for individual observers. Eccentricities were aggregated between each of the two experiments conducted. The ratio for one observer was set aside, and the remaining were averaged across observers. Zero to second-order polynomials were fit to the group-average ratio across eccentricities. Cross-validation error was computed as the sum of squared residuals to the left-out data point. This process was iterated until each observation was left-out once, resulting in 19 total iterations. The best-fitting polynomial order was defined as one that produced the lowest median cross-validation error across all iterations.

REFERENCES

Acerbi, L., and Ma, W.J. (2017). Practical Bayesian optimization for model fitting with Bayesian adaptive direct search. Advances in neural information processing systems 30, 1836-1846.

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

Akaike, H. (1974). A new look at the statistical model identification. IEEE Trans Automat Contr 19, 716-723.

Anton-Erxleben, K., and Carrasco, M. (2013). Attentional enhancement of spatial resolution: linking behavioural and neurophysiological evidence. Nature Reviews Neuroscience 14, 188-200.

Barbot, A., and Carrasco, M. (2017). Attention modifies spatial resolution according to task demands. Psychological science 28, 285-296.

Barbot, A., Xue, S., and Carrasco, M. (2021). Asymmetries in visual acuity around the visual field. Journal of Vision 21, 2-2.

Baruch, O., and Yeshurun, Y. (2014). Attentional attraction of receptive fields can explain spatial and temporal effects of attention. Visual Cognition 22, 704-736.

Bergen, J.R., and Adelson, E.H. (1988). Early vision and texture perception. Nature 333, 363-364.

Boynton, G.M. (2009). A framework for describing the effects of attention on visual responses. Vision research 49, 1129-1143.

Brouwer, G.J., and Heeger, D.J. (2011). Cross-orientation suppression in human visual cortex. Journal of Neurophysiology 106, 2108-2119.

Busse, L., Katzner, S., and Treue, S. (2008). Temporal dynamics of neuronal modulation during exogenous and endogenous shifts of visual attention in macaque area MT. Proceedings of the National Academy of Sciences 105, 16380-16385.

Cameron, E.L., Tai, J.C., and Carrasco, M. (2002). Covert attention affects the psychometric function of contrast sensitivity. Vision Research 42, 949-967.

Carandini, M., and Heeger, D.J. (2011). Normalization as a canonical neural computation. Nature Reviews Neuroscience.

Carrasco, M. (2011). Visual attention: The past 25 years. Vision Research 51, 1484-1525.

Carrasco, M., Loula, F., and Ho, Y.-X. (2006). How attention enhances spatial resolution: Evidence from selective adaptation to spatial frequency. Attention, Perception, & Psychophysics 68, 1004-1012.

Carrasco, M., Talgar, C.P., and Cameron, E.L. (2001). Characterizing visual performance fields: Effects of transient covert attention, spatial frequency, eccentricity, task and set size. Spatial vision 15, 61-75.

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

De Valois, R.L., Albrecht, D.G., and Thorell, L.G. (1982). Spatial frequency selectivity of cells in macaque visual cortex. Vision Research 22, 545-559.

DeValois, R.L., and DeValois, K.K. (1990). Spatial vision, 2nd ed edn (New York: Oxford Univ. Press).

Dugué, L., Merriam, E.P., Heeger, D.J., and Carrasco, M. (2018). Specific Visual Subregions of TPJ Mediate Reorienting of Spatial Attention. Cerebral Cortex 28, 2375-2390.

Dugué, L., Merriam, E.P., Heeger, D.J., and Carrasco, M. (2020). Differential impact of endogenous and exogenous attention on activity in human visual cortex. Scientific Reports 10, 21274.

El-Shamayleh, Y., and Movshon, J.A. (2011). Neuronal Responses to Texture-Defined Form in Macaque Visual Area V2. Journal of Neuroscience 31, 8543-8555.

Fernández, A., and Carrasco, M. (2020). Extinguishing Exogenous Attention via Transcranial Magnetic Stimulation. Current Biology.

Fischer, J., and Whitney, D. (2009). Attention narrows position tuning of population responses in V1. Current Biology 19, 1356-1361.

Giordano, A.M., McElree, B., and Carrasco, M. (2009). On the automaticity and flexibility of covert attention: A speed-accuracy trade-off analysis. Journal of Vision 9, 30-30.

Graham, N.V.S. (1989). Visual pattern analyzers (New York: Oxford University Press).

Hallum, L.E., Landy, M.S., and Heeger, D.J. (2011). Human primary visual cortex (V1) is selective for second-order spatial frequency. Journal of Neurophysiology 105, 2121-2131.

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

Hein, E., Rolke, B., and Ulrich, R. (2006). Visual attention and temporal discrimination: Differential effects of automatic and voluntary cueing. Visual Cognition 13, 29-50.

Herrmann, K., Montaser-Kouhsari, L., Carrasco, M., and Heeger, D.J. (2010). When size matters: attention affects performance by contrast or response gain. Nature Neuroscience 13, 1554-1559.

Himmelberg, M.M., Winawer, J., and Carrasco, M. (2020). Stimulus-dependent contrast sensitivity asymmetries around the visual field. Journal of vision 20, 18-18.

Jazayeri, M., and Movshon, J.A. (2006). Optimal representation of sensory information by neural populations. Nature Neuroscience 9, 690-696.

Jigo, M., and Carrasco, M. (2018). Attention alters spatial resolution by modulating second-order processing. Journal of Vision 18, 2.

Jigo, M., and Carrasco, M. (2020). Differential impact of exogenous and endogenous attention on the contrast sensitivity function across eccentricity. Journal of Vision 20, 11.

Kehrer, L. (1989). Central performance drop on perceptual segregation tasks. Spatial vision 4, 45-62.

Kehrer, L. (1997). The central performance drop in texture segmentation: A simulation based on a spatial filter model. Biological Cybernetics 77, 297-305.

Kehrer, L., and Meinecke, C. (2003). A space-variant filter model of texture segregation: Parameter adjustment guided by psychophysical data. Biological Cybernetics 88, 183-200.

Klein, Barrie P., Harvey, Ben M., and Dumoulin, Serge O. (2014). Attraction of position preference by spatial attention throughout human visual cortex. Neuron 84, 227-237.

Landy, M.S. (2013). Texture analysis and perception. In The New Visual Neurosciences, J.S. Werner, and L.M. Chalupa, eds. (Cambrige, MA: MIT Press), pp. 639-652.

Landy, M.S., and Graham, N. (2004). Visual Perception of Texture. In The Visual Neurosciences (Cambrige, MA: MIT Press), pp. 1106-1118.

Larsson, J., Landy, M.S., and Heeger, D.J. (2006). Orientation-Selective Adaptation to First- and Second-Order Patterns in Human Visual Cortex. Journal of Neurophysiology 95, 862-881.

Lesmes, L.A. (2010). Bayesian adaptive estimation of the contrast sensitivity function: The quick CSF method. Journal of Vision 10, 1-21.

Ling, S., and Carrasco, M. (2006). When sustained attention impairs perception. Nature neuroscience 9, 1243.

Liu, T., Pestilli, F., and Carrasco, M. (2005). Transient attention enhances perceptual performance and fMRI response in human visual cortex. Neuron 45, 469-477.

Macmillan, N.A., and Creelman, C.D. (2005). Detection theory: A user's guide (Mahwah, NJ: Lawrence Erlbaum Associates).

Meese, T.S., and Holmes, D.J. (2007). Spatial and temporal dependencies of cross-orientation suppression in human vision. Proc R Soc B 274, 127-136.

Moran, J., and Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. Science 229, 782-784.

Morikawa, K. (2000). Central performance drop in texture segmentation: the role of spatial and temporal factors. Vision Research 40, 3517-3526.

Ni, A.M., Ray, S., and Maunsell, J.H. (2012). Tuned normalization explains the size of attention modulations. Neuron 73, 803-813.

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

Pestilli, F., Ling, S., and Carrasco, M. (2009). A population-coding model of attention's influence on contrast response: Estimating neural effects from psychophysical data. Vision research 49, 1144-1153.

Pestilli, F., Viera, G., and Carrasco, M. (2007). How do attention and adaptation affect contrast sensitivity? Journal of Vision 7, 9.

Petrov, Y., Carandini, M., and McKee, S. (2005). Two distinct mechanisms of suppression in human vision. Journal of Neuroscience 25, 8704-8707.

Pointer, J.S., and Hess, R.F. (1989). The contrast sensitivity gradient across the human visual field: With emphasis on the low spatial frequency range. Vision research 29, 1133-1151.

Poort, J., Raudies, F., Wannig, A., Lamme, Victor A.F., Neumann, H., and Roelfsema, Pieter R. (2012). The Role of Attention in Figure-Ground Segregation in Areas V1 and V4 of the Visual Cortex. Neuron 75, 143-156.

Priebe, N.J., and Ferster, D. (2006). Mechanisms underlying cross-orientation suppression in cat visual cortex. Nature Neuroscience 9, 552-561.

Reynolds, J.H., and Chelazzi, L. (2004). Attentional modulation of visual processing. Annu Rev Neurosci 27, 611-647.

Reynolds, J.H., and Heeger, D.J. (2009). The Normalization Model of Attention. Neuron 61, 168-185.

Ringach, D.L., Shapley, R.M., and Hawken, M.J. (2002). Orientation Selectivity in Macaque V1: Diversity and Laminar Dependence. The Journal of Neuroscience 22, 5639-5651.

Roelfsema, P.R. (2006). Cortical algorithms for perceptual grouping. Annu Rev Neurosci 29, 203-227.

Sagi, D., and Hochstein, S. (1985). Lateral inhibition between spatially adjacent spatial-frequency channels? Perception & Psychophysics 37, 315-322.

Silvestre, D., Arleo, A., and Allard, R. (2018). Internal noise sources limiting contrast sensitivity. Scientific Reports 8, 2596.

Simoncelli, E.P., Freeman, W.T., Adelson, E.H., and Heeger, D.J. (1992). Shiftable multiscale transforms. IEEE Transactions on Information Theory 38, 587-607.

Stigchel, S.V.d., Belopolsky, A.V., Peters, J.C., Wijnen, J.G., Meeter, M., and Theeuwes, J. (2009). The limits of top-down control of visual attention. Acta Psychologica 132, 201-212.

Talgar, C.P., and Carrasco, M. (2002). Vertical meridian asymmetry in spatial resolution: Visual and attentional factors. Psychonomic Bulletin & Review 9, 714-722.

Tsotsos, J.K. (2011). A computational perspective on visual attention (MIT Press).

Tsotsos, J.K., Culhane, S.M., Wai, W.Y.K., Lai, Y., Davis, N., and Nuflo, F. (1995). Modeling visual attention via selective tuning. Artificial intelligence 78, 507-545.

Victor, J.D., Conte, M.M., and Chubb, C.F. (2017). Textures as probes of visual processing. Annual review of vision science 3, 275-296.

Wang, F., Chen, M., Yan, Y., Zhaoping, L., and Li, W. (2015). Modulation of Neuronal Responses by Exogenous Attention in Macaque Primary Visual Cortex. Journal of Neuroscience 35, 13419-13429.

Watson, A.B., and Ahumada, A.J. (2005). A standard model for foveal detection of spatial contrast. Journal of Vision 5, 6.

Westrick, Z.M., and Landy, M.S. (2013). Pooling of first-order inputs in second-order vision. Vision Research 91, 108-117.

Womelsdorf, T., Anton-Erxleben, K., Pieper, F., and Treue, S. (2006). Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. Nature Neuroscience 9, 1156-1160.

Womelsdorf, T., Anton-Erxleben, K., and Treue, S. (2008). Receptive Field Shift and Shrinkage in Macaque Middle Temporal Area through Attentional Gain Modulation. Journal of Neuroscience 28, 8934-8944.

Yeshurun, Y. (2004). Isoluminant stimuli and red background attenuate the effects of transient spatial attention on temporal resolution. Vision Research 44, 1375-1387.

Yeshurun, Y. (2019). The spatial distribution of attention. Current Opinion in Psychology 29, 76-81.

Yeshurun, Y., and Carrasco, M. (1998). Attention improves or impairs visual performance by enhancing spatial resolution. Nature 396, 72.

Yeshurun, Y., and Carrasco, M. (2000). The locus of attentional effects in texture segmentation. Nature Neuroscience 3, 622-627.

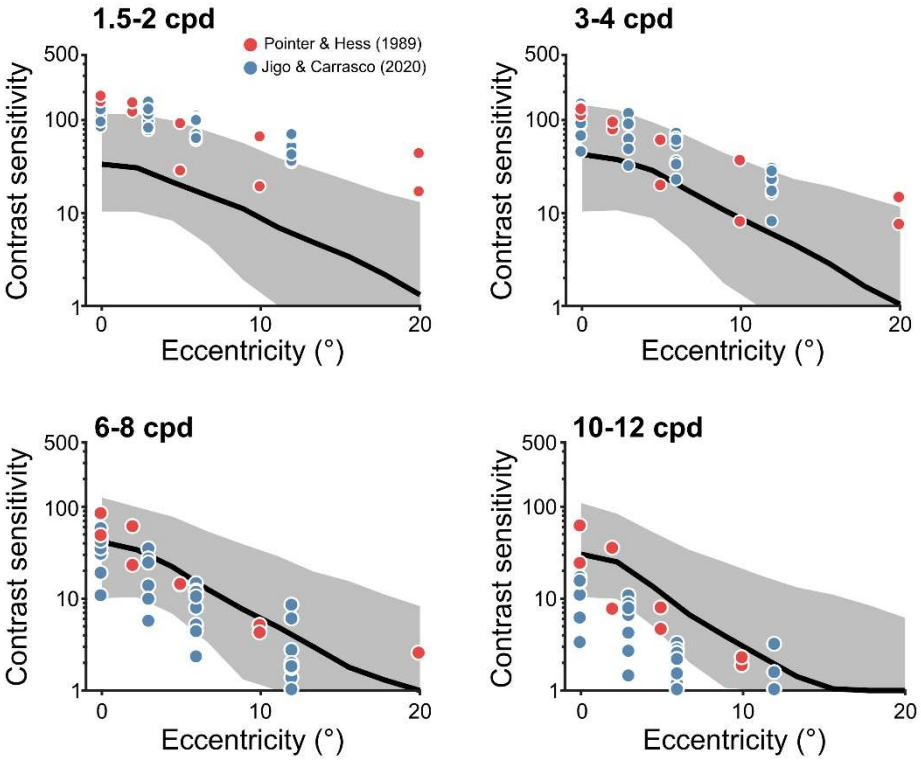
Yeshurun, Y., and Carrasco, M. (2008). The effects of transient attention on spatial resolution and the size of the attentional cue. Perception & Psychophysics 70, 104-113.

Yeshurun, Y., Carrasco, M., and Maloney, L.T. (2008a). Bias and sensitivity in two-interval forced choice procedures: Tests of the difference model. Vision Research 48, 1837-1851.

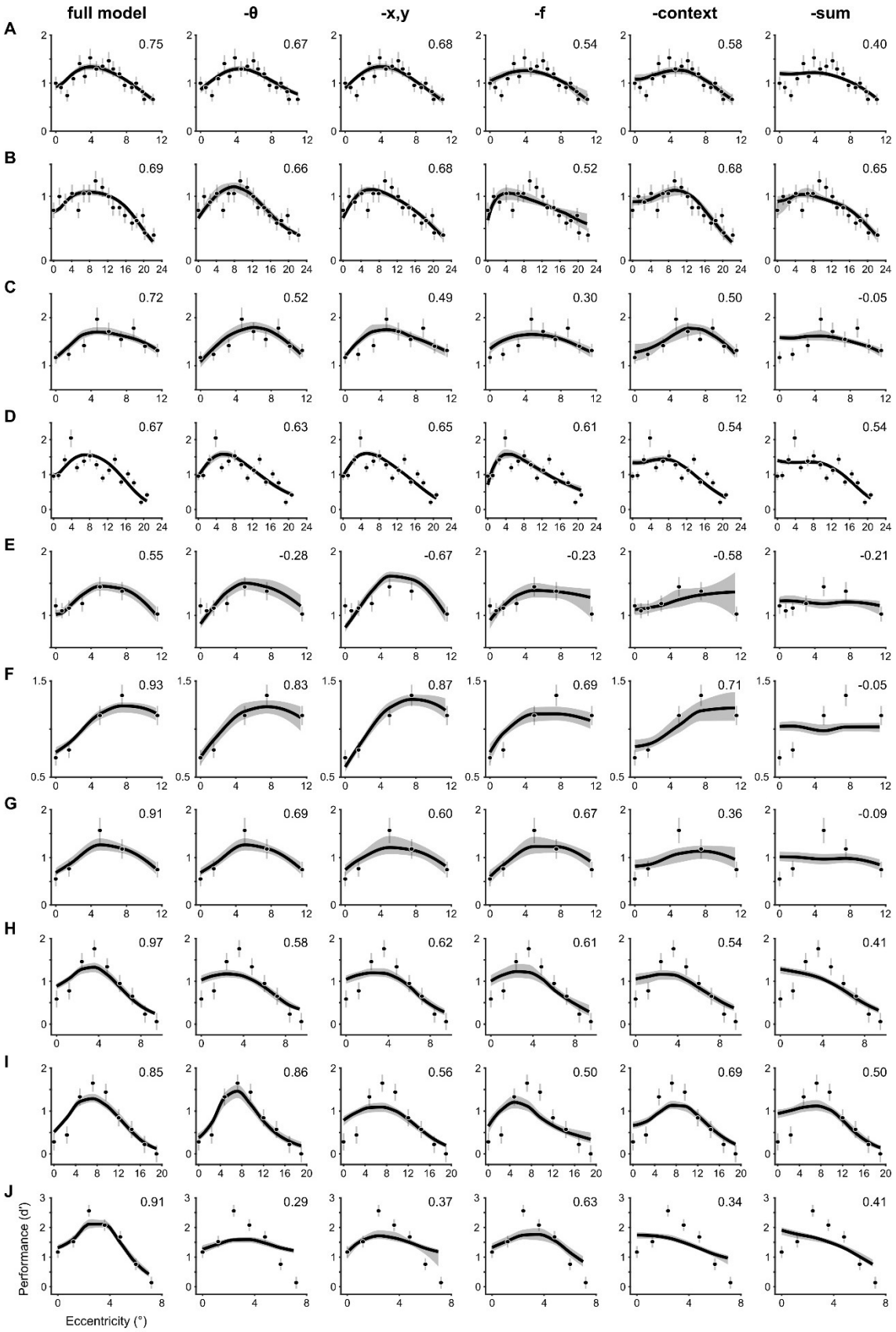
Yeshurun, Y., and Levy, L. (2003). Transient spatial attention degrades temporal resolution. Psychological Science 14, 225-231.

Yeshurun, Y., Montagna, B., and Carrasco, M. (2008b). On the flexibility of sustained attention and its effects on a texture segmentation task. Vision Research 48, 80-95.

SUPPLEMENTARY FIGURES

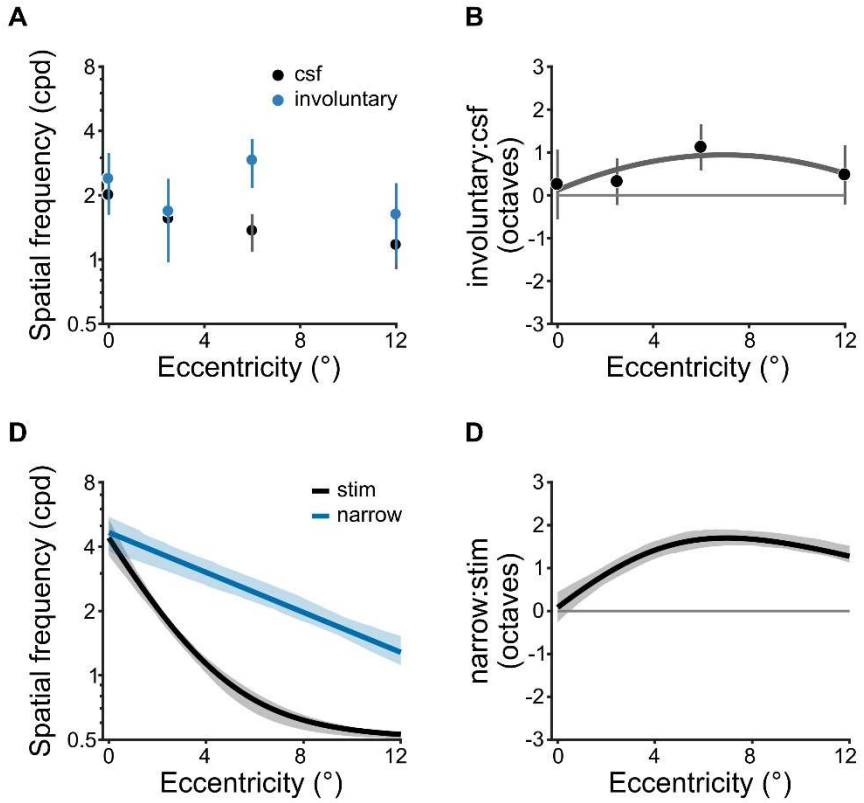


Supplementary Figure 1. Contrast sensitivity across eccentricity.
The black line indicates the median and shaded regions denote 68% confidence intervals of the bootstrapped distribution of the model's contrast sensitivity. Using parameters fit to texture segmentation performance and grating images that matched each of the two experiments displayed above, the model predicted contrast sensitivity functions (see **Methods, Predicted contrast sensitivity functions**). Model parameters were randomly selected, with replacement, from the bootstrapped distribution of parameters that simultaneously fit the central performance drop of all ten texture segmentation experiments. Dots indicate contrast sensitivity, measured psychophysically, for individual observers.

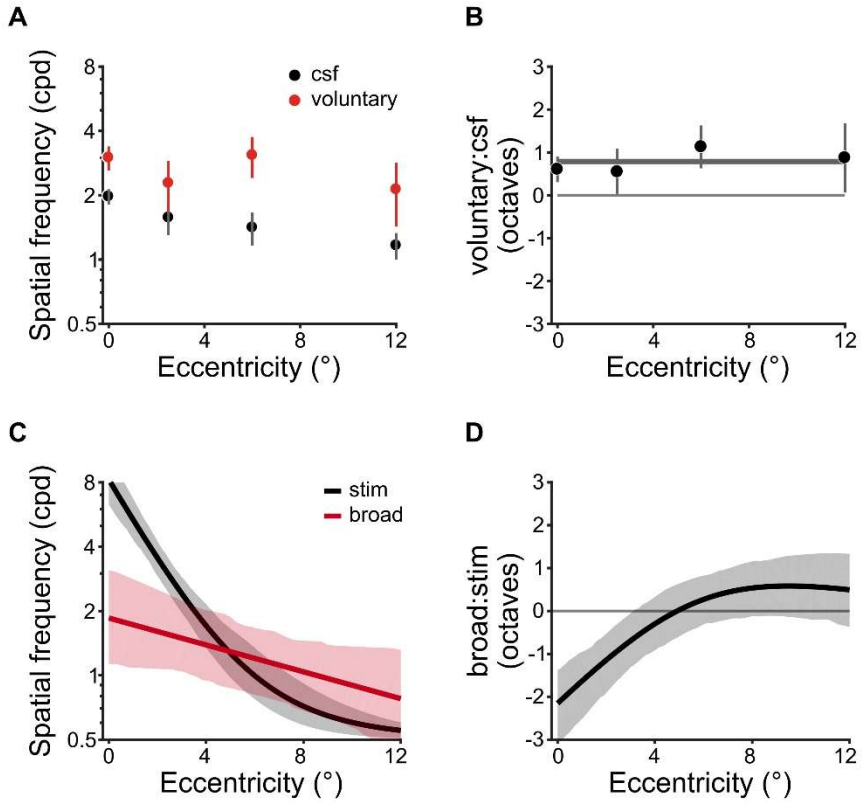

full model
-θ
-x,y
-f
-context
-sum
A
0.75
0.67
0.68
0.54
0.58
0.40
B
0.69
0.66
0.68
0.52
0.68
0.65
C
0.72
0.52
0.49
0.30
0.50
-0.05
D
0.67
0.63
0.65
0.61
0.54
0.54
E
0.55
-0.28
-0.67
-0.23
-0.58
-0.21
F
0.93
0.83
0.87
0.69
0.71
-0.05
G
0.91
0.69
0.60
0.67
0.36
-0.09
H
0.97
0.58
0.62
0.61
0.54
0.41
I
0.85
0.86
0.56
0.50
0.69
0.50
J
0.91
0.29
0.37
0.63
0.34
0.41
Performance (d')
Eccentricity (°)

1390

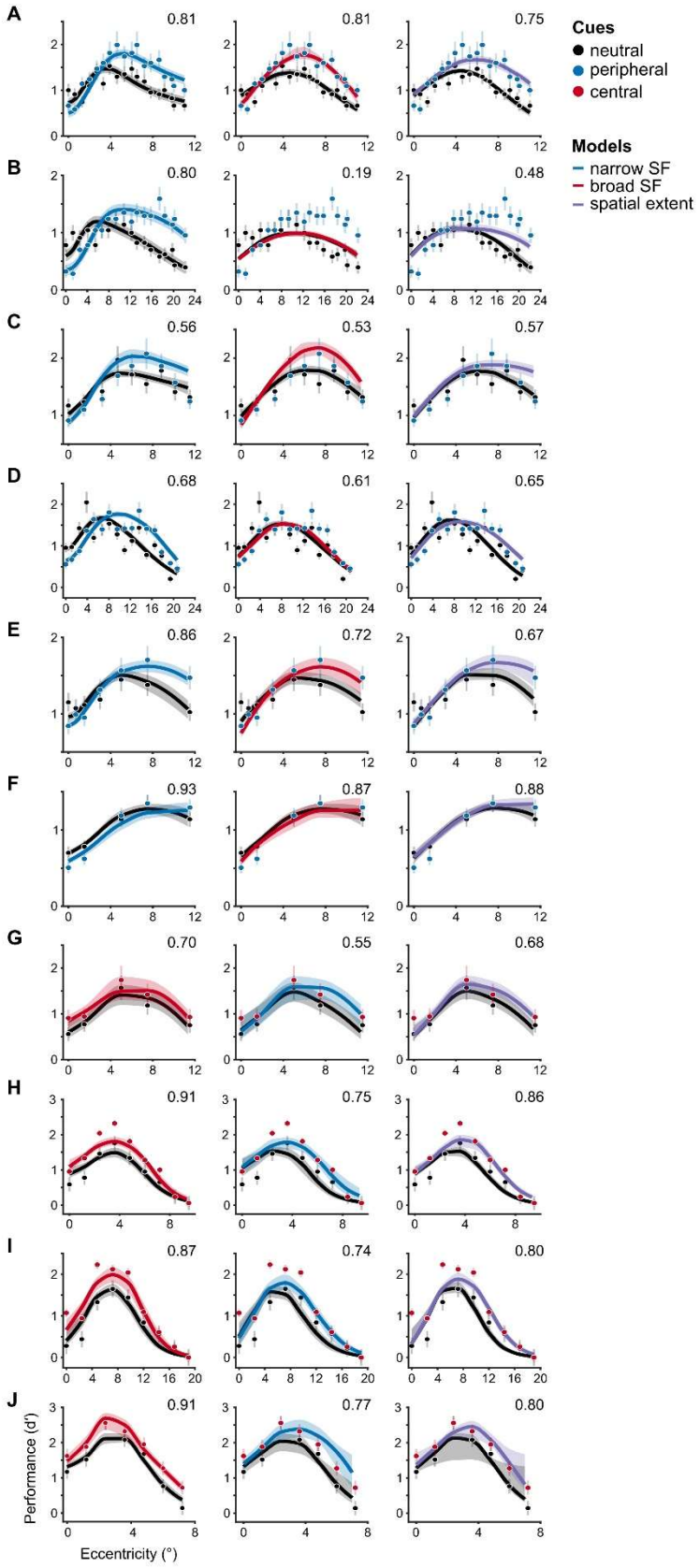
Supplementary Figure 2. Model variants fit to the neutral condition of all ten psychophysical experiments. Each row depicts the behavioral data from (**A**) (Yeshurun and Carrasco, 1998), Experiment 1; (**B**) (Yeshurun and Carrasco, 1998), Experiment 2; (**C**) (Talgar and Carrasco, 2002); (**D**) (Carrasco et al., 2006); (**E**) (Yeshurun and Carrasco, 2008); (**F**) (Yeshurun et al., 2008b), Experiment 2; (**G**) (Yeshurun et al., 2008b), Experiment 1; (**H**) (Yeshurun et al., 2008b), Experiment 3; (**I**) (Yeshurun et al., 2008b), Experiment 4; (**J**) (Barbot and Carrasco, 2017). Each column shows the fit of different model variants arranged in order of best-to-worst according to the model comparisons displayed in **Figure 4D**: 'full' denotes the full model, '-θ' lacks cross-orientation suppression, '-x,y', lacks surround suppression, '-f' lacks cross-frequency suppression, '-context' lacks all contextual modulation and '-sum' lacks spatial summation. Dots and error bars denote group-average performance and ±1 SEM. The solid lines depict the median and shaded regions depict 68% confidence intervals of the bootstrapped distribution of model fits. Values in top-right of each panel denote the median R^2 of the bootstrapped distribution of model fits. Negative R^2 values indicate that the model fit captures less variance in the data than a horizontal line passing through the average d′ across eccentricity.



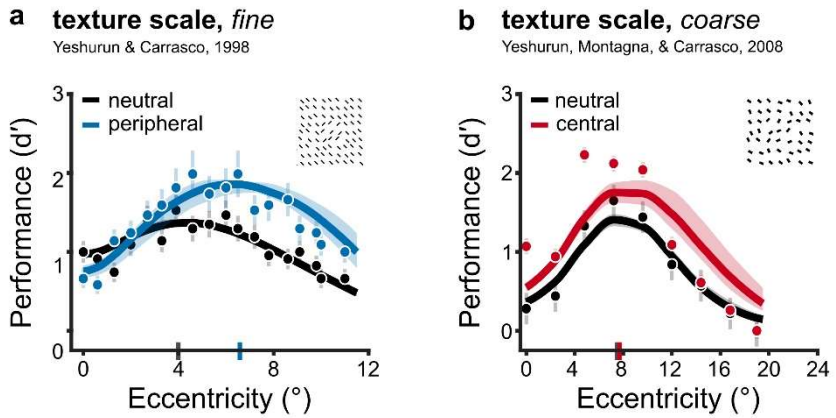
Supplementary Figure 3. Spatial frequency operating range of involuntary attention. (**A**) Peak spatial frequency of baseline contrast sensitivity (CSF) and involuntary attentional modulation from(Jigo and Carrasco, 2020). Estimates were based on human contrast sensitivity, measured psychophysically with narrowband gratings. (**B**) Ratio (in octaves) of attentional and baseline peak spatial frequency tuning across eccentricity. Positive values denote an attentional preference for spatial frequencies higher than baseline. The solid line depicts the best-fitting second-order polynomial (i.e., parabola). Polynomial order was determined using leave-one-subject-out cross-validation (**Methods, Cross-validation procedure**). Dots in *a* and *b* depict group-average and error bars depict ±1 SEM. (**C**) Peak spatial frequency of the stimulus drive (stim) and the narrow attention gain profile (narrow). Estimates were derived from model fits to texture segmentation performance across all six involuntary attention experiments. (**D**) Ratio of the preferred spatial frequency for the stimulus drive and attentional gain. Solid lines indicate the median and shaded area denotes 68% confidence interval of bootstrapped distribution in *c* and *d*.



Supplementary Figure 4. Spatial frequency operating range of voluntary attention. (**A**) The peak spatial frequency of baseline contrast sensitivity (CSF) and the center frequency of broad voluntary attentional modulation from (Jigo and Carrasco, 2020). Estimates were based on human contrast sensitivity, measured psychophysically with narrowband gratings. (**B**) Ratio (in octaves) of attentional and baseline and spatial frequency preferences across eccentricity. Negative values denote an attentional preference for spatial frequencies lower than baseline. The solid line depicts the best-fitting zero-order polynomial (i.e., constant). Polynomial order was determined using leave-one-subject-out cross-validation (**Methods, Cross-validation procedure**). Dots in *a* and *b* depict group-average and error bars depict ±1 SEM (**C**) The center spatial frequency of the stimulus drive (stim) and the broad attentional gain profile (broad). Estimates were derived from model fits to texture segmentation performance across all six voluntary attention experiments. (**D**) Ratio of the preferred spatial frequency for the stimulus drive and attentional gain. Solid lines indicate the median and shaded areas denote 68% confidence intervals of the bootstrapped distribution in *c* and *d*.



Supplementary Figure 5. Attention model variants fit to behavioral data of all ten experiments.
Each row depicts a different experiment: (**A**) (Yeshurun and Carrasco, 1998), Experiment 1; (**B**) (Yeshurun and Carrasco, 1998), Experiment 2; (**C**) (Talgar and Carrasco, 2002); (**D**) (Carrasco et al., 2006); (**E**) (Yeshurun and Carrasco, 2008); (**F**) (Yeshurun et al., 2008b), Experiment 2; (**G**) (Yeshurun et al., 2008b), Experiment 1; (**H**) (Yeshurun et al., 2008b), Experiment 3; (**I**) (Yeshurun et al., 2008b), Experiment 4; (**J**) (Barbot and Carrasco, 2017). Each column depicts a different set of attentional gain models and are ordered from best-to-worst from left-to-right. The numbers in the top-right of each panel denote the median R^2 of the bootstrapped distribution of model fits.
Left. (**A-F**) Narrow SF gain profile fit to involuntary attention experiments and (**G-J**) broad SF gain profile fit to voluntary attention experiments.
Middle. (**A-F**) Broad SF gain profile fit to involuntary attention experiments and (**G-J**) narrow SF gain profile fit to voluntary attention experiments.
Right. Spatial extent model (i.e., uniform attentional gain across SF) fit to all experiments.



Supplementary Figure 6. Model fits to jittered texture stimuli.
(**A**) Predicted performance for Experiment 1 in (Yeshurun and Carrasco, 1998) using texture stimuli generated with line elements spatially jittered within the stimulus parameters of the experiment (**Methods, Stimulus generation**). An example jittered stimulus is shown in the top-right. Solid lines indicate the median and shaded regions depict 68% confidence intervals of the bootstrap distributions of model predictions. Gray and blue ticks on x-axis indicate peak of performance in the neutral and peripheral cueing condition, respectively. To generate the bootstrap distributions, model parameters were fixed to those that best-fit all involuntary experiments simultaneously then new jittered texture stimuli were input to the model on each iteration.
(**B**) Predicted performance for experiment 4 in (Yeshurun et al., 2008b) using texture stimuli generated with line elements whose orientation and spatial location were randomly jittered within the parameters of the experiment (**Methods, Stimulus generation**). Solid lines indicate the median and shaded regions depict 68% confidence intervals of bootstrap distributions of model predictions. Gray and red ticks on x-axis indicate peak of performance in the neutral and central cueing condition, respectively. To generate the bootstrap distribution, model parameters were fixed to those that best-fit all voluntary experiments simultaneously and new jittered texture stimuli were input to the model on each iteration.