An image-computable model of how endogenous and exogenous attention differentially alter visual perception

Michael Jigo[a,1] , David J. Heeger[a,b] , and Marisa Carrasco[a,b]

[a]Center for Neural Science, New York University, New York, NY 10003; and [b]Department of Psychology, New York University, New York, NY 10003

Attention alters perception across the visual field. Typically, endogenous (voluntary) and exogenous (involuntary) attention similarly improve performance in many visual tasks, but they have differential effects in some tasks. Extant models of visual attention assume that the effects of these two types of attention are identical and consequently do not explain differences between them. Here, we develop a model of spatial resolution and attention that distinguishes between endogenous and exogenous attention. We focus on texture-based segmentation as a model system because it has revealed a clear dissociation between both attention types. For a texture for which performance peaks at parafoveal locations, endogenous attention improves performance across eccentricity, whereas exogenous attention improves performance where the resolution is low (peripheral locations) but impairs it where the resolution is high (foveal locations) for the scale of the texture. Our model emulates sensory encoding to segment figures from their background and predict behavioral performance. To explain attentional effects, endogenous and exogenous attention require separate operating regimes across visual detail (spatial frequency). Our model reproduces behavioral performance across several experiments and simultaneously resolves three unexplained phenomena: 1) the parafoveal advantage in segmentation, 2) the uniform improvements across eccentricity by endogenous attention, and 3) the peripheral improvements and foveal impairments by exogenous attention. Overall, we unveil a computational dissociation between each attention type and provide a generalizable framework for predicting their effects on perception across the visual field.

computational model | endogenous attention | exogenous attention | spatial resolution | texture

Endogenous and exogenous spatial attention prioritize subsets of visual information and facilitate their processing without concurrent eye movements (1–3). Selection by endogenous attention is goal-driven and adapts to task demands, whereas exogenous attention transiently and automatically orients to salient stimuli (1–3). In most visual tasks, both types of attention typically improve visual perception similarly [e.g., acuity (4–6), visual search (7, 8), perceived contrast (9–11)]. Consequently, models of visual attention do not distinguish between endogenous and exogenous attention (e.g., refs. 12–19). However, stark differences also exist. Each attention type differentially modulates neural responses (20, 21) and fundamental properties of visual processing, including temporal resolution (22, 23), texture sensitivity (24), sensory tuning (25), contrast sensitivity (26), and spatial resolution (27–34).

The effects of endogenous and exogenous attention are dissociable during texture segmentation, a visual task constrained by spatial resolution [reviews (1–3)]. Whereas endogenous attention optimizes spatial resolution to improve the detection of an attended texture (32–34), exogenous attention reflexively enhances resolution even when detrimental to perception (27–31, 34). Extant models of attention do not explain these well-established effects.

Two main hypotheses have been proposed to explain how attention alters spatial resolution. Psychophysical studies ascribe attentional effects to modulations of spatial frequency (SF) sensitivity (30, 33). Neurophysiological (13, 35, 36) and neuroimaging (37, 38) studies bolster the idea that attention modifies spatial profiles of neural receptive fields (RFs) (2). Both hypotheses provide qualitative predictions of attentional effects but do not specify their underlying neural computations.

Differences between endogenous and exogenous attention are well established in segmentation tasks and thus provide an ideal model system to uncover their separate roles in altering perception. Texture-based segmentation is a fundamental process of midlevel vision that isolates regions of local structure to extract figures from their background (39–41). Successful segmentation hinges on the overlap between the visual system's spatial resolution and the levels of detail (i.e., SF) encompassed by the texture (39, 41, 42). Consequently, the ability to distinguish between adjacent textures varies as resolution declines toward the periphery (43–46). Each attention type differentially alters texture segmentation, demonstrating that their effects shape spatial resolution [reviews (1–3)].

Current models of texture segmentation do not explain performance across eccentricity and the distinct modulations by attention. Conventional models treat segmentation as a feedforward

process that encodes the elementary features of an image (e.g., SF and orientation), transforms them to reflect the local structure (e.g., regions of similarly oriented bars), and then pools across space to emphasize texture-defined contours (39, 41, 47). Few of these models account for variations in resolution across eccentricity (46, 48, 49) or endogenous (but not exogenous) attentional modulations (18, 50). All others postulate that segmentation is a "preattentive" (42) operation whose underlying neural processing is impervious to attention (39, 41, 46–49).

Here, we develop a computational model in which feedforward processing and attentional gain contribute to segmentation performance. We augment a conventional model of texture processing (39, 41, 47). Our model varies with eccentricity and includes contextual modulation within local regions in the stimulus via normalization (51), a canonical neural computation (52). The defining characteristic of normalization is that an individual neuron is (divisively) suppressed by the summed activity of neighboring neurons responsive to different aspects of a stimulus. We model attention as multiplicative gains [attentional gain factors (15)] that vary with eccentricity and SF. Attention shifts sensitivity toward fine or coarse spatial scales 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 (6, 26, 27, 29–33). The model explains three signatures of texture segmentation hitherto unexplained within a single computational framework (Fig. 1): 1) the central performance drop (CPD) (27–34, 43–46) (Fig. 1A), that is, the parafoveal advantage of segmentation over the fovea; 2) the improvements in the periphery and impairments at foveal locations induced by exogenous attention (27–32, 34) (Fig. 1B); and 3) the equivalent improvements across eccentricity by endogenous attention (32–34) (Fig. 1C).

Whereas our analyses focused on texture segmentation, our model is general and can be applied to other visual phenomena. We show that the model predicts the effects of attention on contrast sensitivity and acuity, i.e., in tasks in which both endogenous and exogenous attention have similar or differential effects on performance. To preview our results, model comparisons revealed that normalization is necessary to elicit the CPD and that separate profiles of gain enhancement across SF (26) generate the effects of exogenous and endogenous attention on texture segmentation. A preferential high-SF enhancement reproduces the impairments by exogenous attention due to a shift in visual sensitivity toward details too fine to distinguish the target at foveal locations. The transition from impairments in the periphery results from exogenous attentional gain gradually shifting to lower SFs that are more amenable for target detection. Improvements by endogenous attention result from a uniform enhancement of SFs that encompass the
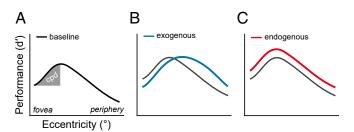
target, optimizing visual sensitivity for the attended stimulus across eccentricity.

Results

Image-Computable Model of Attention and Spatial Resolution.
We developed an observer model based on established principles of neural computation (51, 52), pattern (53, 54), and texture vision (39, 41, 47), and attentional modulation (15). The model incorporates elements of the Reynolds-Heeger normalization model of attention (NMA) (15) and illuminates how attention alters contrast and texture sensitivity across SF and eccentricity. We implement 1) SF-tuned gain modulation to emulate the decline in contrast sensitivity and peak SF preference with eccentricity; 2) spatial summation of normalized inputs to generate texture selectivity; and 3) separate attentional gain profiles across SF to reproduce effects of exogenous and endogenous attention. The model is composed of four components: stimulus drive, attentional gain, suppressive drive, and spatial summation (Fig. 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 RFs, each jointly tuned to spatial position, SF, and orientation. Images are processed through a filter bank (55) covering the visual field at several SFs and orientations using bandwidths compatible with neurophysiological (54) and psychophysical (53) measurements. Filter outputs are combined across quadrature phase (56), yielding contrast energy images corresponding to different SFs and orientations. These outputs simulate the responses of complex cells in primary visual cortex (54, 56). The gain on individual RFs varies as a function of SF and eccentricity preference (Fig. 2A, green). Following the behavior of individual neurons (54) and pattern vision (53), gain modulation is narrowly tuned to high SFs near the fovea and progressively shifts to low SFs with eccentricity. Consequently, the stimulus drive reflects local spectral energy within each patch in an image, filtered through feature-selective RFs that vary with eccentricity.

Attentional gain. Attention is implemented as a gain control mechanism that scales the gain on the stimulus drive (15). The magnitude of attentional gain is largest at the cued location (Fig. 2A, orange) and varies with the eccentricity and SF preference of each RF. Motivated by findings of psychophysical experiments that manipulated endogenous and exogenous attention (26), two SF-tuned profiles are assessed—narrow and broad. The narrow profile selectively enhances a small range of SFs at each eccentricity (Fig. 2A, blue); the broad profile uniformly enhances SFs (Fig. 2A, red).

Suppressive drive. Suppression operates via divisive normalization (51, 52). 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; Fig. 2A, black). The normalization pool consists of the attention-scaled stimulus drive across nearby spatial locations [surround suppression (57)], uniformly across orientation [cross-orientation suppression (58)], and across preferred and neighboring SFs [cross-frequency suppression (59)] of individual RFs. Such broad suppressive pools are supported by physiological (57, 58, 60) and psychophysical (59, 61, 62) findings and models of visual processing (51).

Spatial summation. Normalized responses are weighted and summed across space within each SF and orientation filter. Spatial summation followed normalization (63), which accentuated texture-defined contours within the image. The size of pooling regions scale with the SF preference of each RF (39, 41) (Fig. 2A, purple), larger for low than for high SFs. This implements an inverse relation between the integration area of individual RFs and their SF tuning.

Fig. 1. Signatures of texture segmentation. (A) CPD. Shaded region depicts the magnitude of the CPD. Identical axis labels are omitted in B and C. (B) Exogenous attention modulation. Exogenous attention improves segmentation performance in the periphery and impairs it near the fovea. (C) Endogenous attention modulation. Endogenous attention improves segmentation performance across eccentricity.
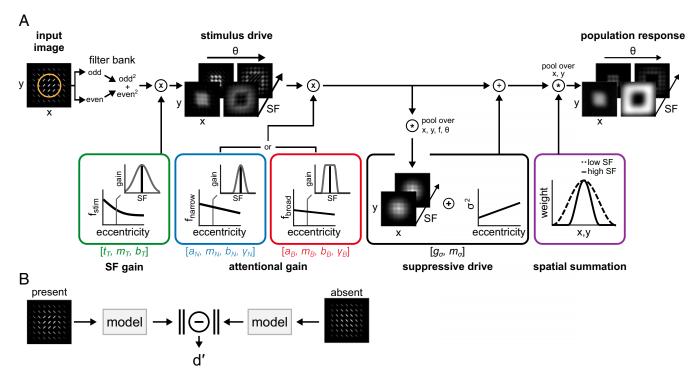
Fig. 2. Image-computable model of attention and spatial resolution. (*A*) Model structure. A filter bank of linear RFs decomposes an image. Filter responses are squared and summed across quadrature-phase pairs (odd, even), yielding contrast energy outputs. SF gain scales contrast energy across SF and eccentricity (green box). The solid black line depicts the center frequency of the tuning function (f_{stim}); insets display the full SF tuning function at a single eccentricity. The stimulus drive characterizes contrast energy at each pixel in the image, filtered through feature-selective and eccentricity-dependent RFs. 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 either a narrow profile (blue box) or a broad profile (red box), each centered on a given frequency (f_{narrow} or f_{broad}). The suppressive drive comprises the attention-scaled stimulus drive pooled across a local neighborhood of positions, SFs, and uniformly across orientation. Contrast gain, σ^2, adjusts 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 texture images with (present) or without (absent) a target patch are computed. The vector magnitude of their difference produces a metric proportional to d′, assuming independent and identically distributed Gaussian output noise.

Target discriminability. The model generated measures of discriminability (d′) in a texture segmentation task (Fig. 2*B*). The model generated population responses to two texture images. One contained a target patch whose orientation 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 population responses indexed d′. This measure is proportional to behavioral performance, assuming the addition of normally distributed noise after normalization.

Texture Stimuli, Behavioral Protocol, and Optimization Strategy.
Stimuli. Model parameters were constrained by data from 10 published psychophysical experiments. Exogenous attention was manipulated in six experiments (27, 29–32) (Fig. 3 *A–F*) and endogenous attention in four experiments (32, 33) (Fig. 3 *G–J*). In each experiment, observers distinguished a patch of one orientation embedded within a background of differing orientation at several possible eccentricities.
Behavioral protocol. Performance was typically measured with a two-interval forced choice protocol (Fig. 3*K*). Observers maintained fixation at the display's center while viewing two intervals of texture stimuli, one of which randomly contained a target texture. Different precues at their optimal timing manipulated exogenous or endogenous attention. Brief peripheral precues manipulated exogenous attention and appeared before both intervals but near the upcoming target location in the interval containing the target (27–32, 34). Symbolic precues manipulated

endogenous attention. Precues appeared near fixation and indicated the target location in the target-present interval (32, 33). 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: 1) the CPD emerged in the neutral condition (Fig. 1*A*); 2) peripheral precues improved performance in the periphery and impaired it at foveal locations (Fig. 1*B*); and 3) central symbolic precues improved performance at all eccentricities (Fig. 1*C*).
Optimization. To identify the computations that underlie each signature, 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 10 experiments. Second, exogenous attentional effects were assessed by fitting to neutral and peripheral cueing conditions from the six exogenous attention experiments. Third, endogenous attentional effects were assessed by fitting to neutral and central cueing conditions from the four endogenous attention experiments. The model was jointly fit to each subset of data, with model parameters shared among experiments within a subset (*SI Appendix*, Tables S2–S4).

Contextual Modulation and Spatial Summation Mediate the CPD. To identify the computations mediating the CPD, we fit the model to group-average performance across all experiments' neutral condition (103 data points). Fifteen model parameters constrained performance (*SI Appendix*, Table S2). To account for differences
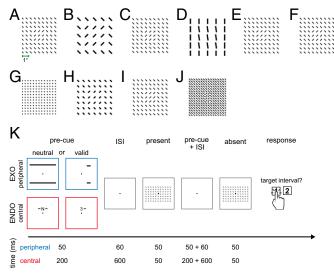
Jigo et al.
An image-computable model of how endogenous and exogenous attention differentially alter visual perception

PNAS | 3 of 12
https://doi.org/10.1073/pnas.2106436118

Fig. 3. Texture stimuli and a typical texture segmentation behavioral protocol. Target-present texture stimuli used in (*A–F*) exogenous attention and (*G–J*) endogenous attention experiments, displayed at their respective spatial scales. Textures displayed include (*A*) fine and (*B*) coarse-scale textures used in Yeshurun and Carrasco (27); (*C*) Talgar and Carrasco (29), with targets placed on the vertical meridian; (*D*) Carrasco, Loula, and Ho (30), wherein observers discriminated the target's orientation; (*E*) Yeshurun and Carrasco (31), where the cue's size was manipulated; (*F*) Experiment 2 of Yeshurun, Montagna, and Carrasco (32) with targets placed on the horizontal meridian; (*G*) Experiment 3 and (*H*) Experiment 4 of Yeshurun, Montagna, and Carrasco (32), wherein fine and coarse scale textures were displayed, respectively; (*I*) Experiment 1 of Yeshurun, Montagna, and Carrasco (32) with targets placed on the horizontal meridian; and (*J*) Barbot and Carrasco (33) with targets placed on the intercardinal meridians. (*K*) Two-interval forced choice protocol typically used to assess texture segmentation performance. EXO corresponds to exogenous attention and ENDO to endogenous attention. Numbers denote the representative timing information for each precue—peripheral (blue) and central (red)—and their corresponding interstimulus intervals (ISI). Neutral precues equally distributed attention to all possible target locations. Valid peripheral precues appeared near the upcoming target location, whereas valid central precues 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.

in contrast sensitivity due to variable display properties among experiments (e.g., mean luminance), foveal contrast gain (g_σ; Fig. 2*A*) was independently determined for each of 10 experiments (10 parameters). Two separate parameters determined foveal SF preference (t_T)—one shared among exogenous attention studies and another among endogenous 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 (Fig. 4). For a fine-scale texture—characterized by narrow, densely spaced lines—performance peaked within the parafovea (4°) and declined toward the fovea and periphery (Fig. 4*A*). Differences between target-present and target-absent stimuli were largest within the 2 cpd filter (Fig. 4 *A*, *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°; Fig. 4*B*). The CPD was well-fit in all experiments (Fig. 4*C*); 77% of the variance was explained (95% bootstrapped CI = [70 to 80]), with the best-fitting regression line falling close to the unity line.

Previous models qualitatively matched the CPD through spatial summation (46, 48, 49), but ignored the contributions of contextual modulation via normalization. To assess the contribution
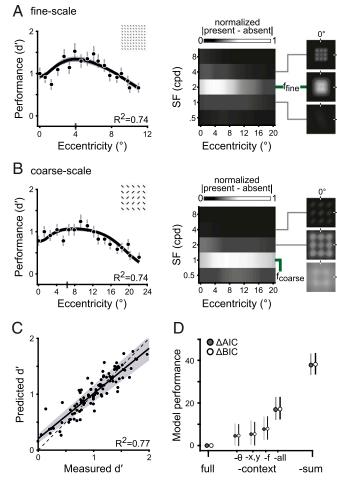

Fig. 4. Contextual modulation and spatial summation mediate the CPD. (*A*, *Left*) Fit to Experiment 1 in Yeshurun and Carrasco (27). Dots (*n* = 18) and error bars depict group-average performance and ±1 SEM. The black line and shaded regions depict the median and 68% bootstrapped confidence interval of model fits. The gray vertical bar on the *x*-axis indicates the eccentricity of peak performance. The inset shows the textures 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 with the largest difference between population responses. We use absolute differences only to visualize the SFs that drove discriminability. (*Right*) Spatial distribution of the absolute value of differences between target-present and target-absent population responses. Each panel depicts a subset of RFs centered on the fovea and tuned to one of three SFs (4, 2, 1 cpd) and an orientation of 30°. (*B*) Fit to Experiment 2 (*n* = 18) in Yeshurun and Carrasco (27). The model jointly fits neutral performance with parameters shared among all 10 experiments, including the data shown in *A*. Visualization follows the conventions in *A*. Note that eccentricity (*x*-axis) is twice that of *A*. f_coarse denotes the SF filter that best distinguished the coarse-scale target. (*C*) Goodness-of-fit for the neutral condition across 10 experiments (*n* = 103). Each dot depicts the measured (*x*-axis) and predicted (*y*-axis) performance at a given eccentricity. The solid line and shaded area depict the best-fitting regression line and its 95% CI. The dashed line indicates the unity line *y = x*. (*D*) Model comparisons using AIC and BIC. Positive values indicate models underperforming, relative to the full model. "-context" describes restrictions of contextual modulation: "-θ" denotes the variant without cross-orientation suppression, "-f" without cross-frequency suppression, "-x,y" without surround suppression, and "-all" devoid of all contextual modulation. "-sum" denotes the model variant without spatial summation. The dots and error bars denote the median and 95% CI of the bootstrap distribution.

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 (Fig. 4D). We restricted contextual modulation (-context) by separately limiting 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 RF's tuning were removed. The final variant lacked spatial summation (-sum), which resulted in a population response that consisted of only normalized inputs. Removing spatial summation attenuates the 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 (AIC) (64) and Bayesian Information Criterion (BIC) (65).

Removing contextual modulation or spatial summation attenuated the CPD (*SI Appendix*, Fig. S1). We measured model performance relative to the full model, which yielded ΔAIC and ΔBIC scores; positive values represent a decrease in model performance. We use "M" and "CI" to denote the median and 95% CI of the bootstrapped distribution. Model performance fell without cross-orientation suppression (ΔAIC: M = 4.8, CI = [−0.1 to 9.7]; ΔBIC: M = 4.6, CI = [−0.2 to 9.6]), cross-frequency suppression (ΔAIC: M = 7.9, CI = [2.7 to 13.2]; ΔBIC: M = 7.7, CI = [2.4 to 13.8]), surround suppression (ΔAIC: M = 5.4, CI = [0.03 to 11.0]; ΔBIC: M = 5.4, CI = [−0.1 to 11.5]), and without all forms of contextual modulation (ΔAIC: M = 17.0, CI = [11.5 to 22.1]; ΔBIC: M = 16.9, CI = [11.6 to 22.4]). Without spatial summation, model performance decreased as well (ΔAIC: M = 37.8, CI = [33.3 to 42.6]; ΔBIC: M = 37.8, CI = [33.1 to 42.8]). Thus, reliable reproduction of the CPD requires both contextual modulation and spatial summation.
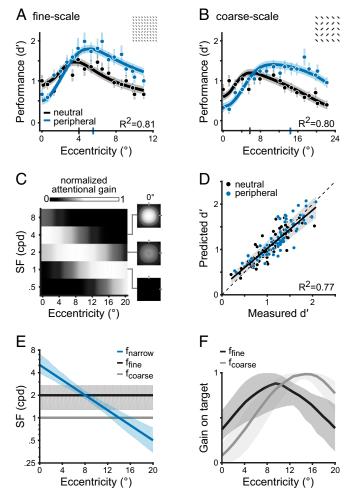
Narrow High-SF Enhancement Generates Exogenous Attention Effects. The model predicted behavior in neutral and peripheral cueing conditions across six experiments (146 data points). Exogenous attention was modeled as a narrow SF gain profile (Fig. 2, blue), motivated by psychophysical measurements (26). Fourteen free parameters constrained model behavior (*SI Appendix*, Table S3). 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 in *Contextual Modulation and Spatial Summation Mediate the CPD*. 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.

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

A gradual shift of attentional gain toward lower SFs (26) reproduced the transition from impairments to improvements across eccentricity (Fig. 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 SF of each target, improving performance. Attention enhanced the fine-scale target SF within

the parafovea (4° to 12°) and then enhanced the coarse-scale target at farther eccentricities (8° to 22°). Across the six experiments, the model explained 77% of the variance (95% bootstrapped CI = [49 to 82]; Fig. 5D).

Attentional gain on SFs higher than the target yielded impairments at foveal locations. This pattern was consistent across all six experiments (Fig. 5E). Consequently, the overlap between fine- (f_{fine}) or coarse-scale (f_{coarse}) targets and the SF tuning of attentional gain was minimal at the fovea and peaked in the periphery (Fig. 5F). This mismatch between the SF tuning of attention (f_{narrow}) and the target is suggested to be driven by

Fig. 5. Narrow high-SF enhancement generates exogenous attention effects. (A) Fit to Experiment 1 in Yeshurun and Carrasco (27). The dots ($n = 36$) depict group-average performance and error bars denote ±1 SEM. The solid lines and shaded regions indicate the median and 68% CIs of the bootstrapped distribution of model fits. The vertical blue bar on the x-axis indicates the eccentricity of peak performance with peripheral cues. (B) Fit to Experiment 2 ($n = 36$) in Yeshurun and Carrasco (27). The model jointly fits performance on neutral and peripheral cue conditions with parameters shared among all six experiments, including the data shown in A. Visualization follows the conventions in A. (C) Best-fitting narrow gain profile. The matrix depicts attentional gain across eccentricity, normalized by the maximum and averaged across space and orientation. Matrix visualization and the panels on the right follow the conventions of Fig. 4A. (D) Goodness-of-fit for neutral and peripheral-cued performance ($n = 146$). Plotted as in Fig. 4C. (E) SF preference of the narrow attentional gain profile (f_{narrow}) and the SF that best distinguished fine- (f_{fine}) and coarse-scale targets (f_{coarse}). The solid lines and shaded areas indicate the median and 68% bootstrapped CI. The 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% CI of bootstrapped distribution).

Jigo et al.
An image-computable model of how endogenous and exogenous attention differentially alter visual perception

PNAS | 5 of 12
https://doi.org/10.1073/pnas.2106436118

Fig. 6. Broad SF enhancements yield endogenous attention effects. (*A*) Fit to Experiment 3 in Yeshurun, Montagna, and Carrasco (32). The dots (*n* = 18) depict group-average performance and error bars denote ±1 SEM. The solid lines and shaded regions indicate the median and 68% CIs of the bootstrapped distribution of model fits. The vertical red bar on the x-axis indicates the eccentricity of peak performance with central cues. (*B*) Fit to Experiment 4 (*n* = 18) in Yeshurun, Montagna, and Carrasco (32). The model jointly fits performance on neutral and central cue conditions with parameters shared among all four experiments, including the data shown in *A*. Visualization follows the conventions in *A*. (*C*) Best-fitting broad gain profile. Plotted as in Fig. 5C. (*D*) Goodness-of-fit for neutral and central-cued performance (*n* = 60). Plotted as in Fig. 4C. (*E*) SF preference of the broad attentional gain profile (f_{broad}) and the SF that best distinguished fine- (f_{fine}) and coarse-scale targets (f_{coarse}). The solid lines and shaded areas indicate the median and 68% bootstrapped CI. (*F*) Normalized magnitude of attentional gain on the fine- and coarse-scale target SF across eccentricity (median and 68% CI of bootstrapped distribution).
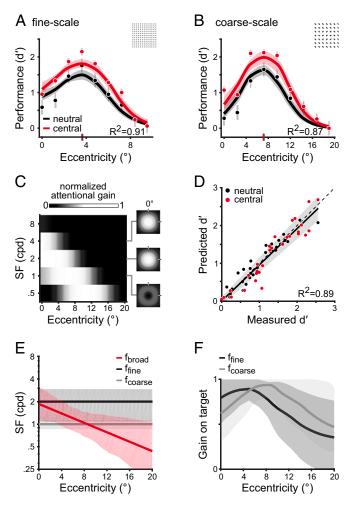
exogenous attention operating above intrinsic SF preferences at each eccentricity (26). We corroborated this relation. We compared f_{narrow} to the model's baseline SF tuning, indexed by the peak SF of the stimulus drive (f_{stim}, Fig. 2A). Consistent with empirical measurements, we found that the narrow SF profile preferred SFs higher than baseline tuning (*SI Appendix*, Fig. S2).

Broad SF Enhancements Yield Endogenous Attention Effects. The model predicted group-average data from neutral and central cueing conditions across four experiments (60 data points). Endogenous attention was modeled as a broad SF gain profile (Fig. 2A, red) (26). Twelve free parameters constrained model behavior (*SI Appendix*, Table S4). 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- (Fig. 6A) and coarse-scale textures (Fig. 6B). To generate these improvements, attentional gain encompassed the target SF for each texture scale (Fig. 6C). Across all four experiments, the model explained 89% of the variance (95% bootstrapped CI [67 to 92]; Fig. 6D).

Endogenous attention effects were reproduced by a broad SF attentional gain that was centered near the target SF across eccentricity (f_{broad} in Fig. 6E). This contrasts with the narrow SF gain profile that modulated higher SFs at central locations to reproduce exogenous attention effects (Fig. 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 (Fig. 6F). Psychophysical measurements of attentional effects on contrast sensitivity (26) suggest that the SF range enhanced by endogenous 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 (*SI Appendix*, Fig. S3).

Different SF Gain Profiles Govern Exogenous and Endogenous Attention Effects. We directly assessed whether different SF gain profiles—narrow or broad—generate the effects of exogenous and endogenous 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 remained uniform. The spatial spread of attention is a key factor of the NMA (15), which posits that its extent relative to the stimulus size helps reconcile apparent discrepancies between each attention type's effects on contrast sensitivity. These predictions have been empirically tested and confirmed (66). By comparing the narrow and broad SF models to the spatial extent model, we directly assessed the separate contributions of SF gain and the spatial spread of attention to segmentation performance (Fig. 7).

Tuned SF gain modulation reproduced the effects of attention. The spatial extent alone was insufficient to capture the effects of either exogenous (ΔAIC: M = 21.2, CI = [18.8 to 26.0]; ΔBIC: M = 31.7, CI = [27.9 to 34.9]; Fig. 7A) or endogenous attention (ΔAIC: M = 11.4, CI = [3.9 to 18.9]; ΔBIC: M = 13.5, CI = [5.7 to 20.8]; Fig. 7B). For exogenous attention, the narrow profile outperformed the broad profile (ΔAIC: M = 39.1, CI = [35.5 to 42.5]; ΔBIC: M = 39.1, CI = [35.9 to 42.5]; Fig. 7A). For
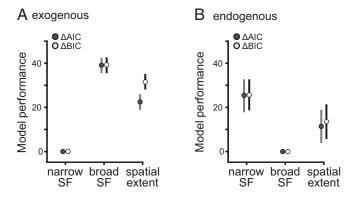

Fig. 7. Different SF gain profiles govern exogenous and endogenous attention effects. AIC and BIC model comparisons for different regimes of attentional modulation for (*A*) exogenous attention and (*B*) endogenous attention. The dots and error bars represent the median and 95% CIs of the bootstrap distributions.

endogenous attention, the broad profile outperformed the narrow profile (ΔAIC: M = 25.4, CI = [17.8 to 32.7]; ΔBIC: M = 25.5, CI = [18.0 to 32.7]; Fig. 7*B*). Decrements in model performance manifested as an inability to capture impairments or improvements at eccentricities demarcating the CPD (*SI Appendix*, Fig. S4). Thus, these model comparisons substantiate psychophysical measurements (25, 26): exogenous and endogenous attention effects are best explained by different attentional gain profiles across SF.

A Parsimonious Explanation for Several Experimental Manipulations in Texture Segmentation. Fig. 8 depicts behavioral data for a variety of texture segmentation experiments. Although we focus on the impact of texture scale in Figs. 5 and 6, the model is general. It jointly accounted for multiple target locations (vertical, Fig. 8*A*; horizontal, Fig. 8 *C–E*; and intercardinal meridians, Fig. 8*F*), behavioral tasks (orientation discrimination, Fig. 8*B*) and attentional manipulations (cue size, Fig. 8*C*). Although the model was fit using texture images with fixed positions and orientations

(Fig. 3), it behaved similarly for textures with randomly jittered elements (*SI Appendix*, Fig. S5). Overall, the proposed model provides a parsimonious explanation for and a quantitative match to segmentation performance (Fig. 8).

Model Predictions Generalize to Basic Visual Tasks. To test whether this model generalizes to other basic visual tasks, we applied it to tasks mediated by acuity (6) and contrast sensitivity (26), with no additional model parameters (*SI Appendix*, section S6). These studies separately manipulated exogenous and endogenous attention and highlight how attention effects depend on the stimulus and task. In the acuity task, observers discriminated the location of a small gap (<1°) in a Landolt square (*SI Appendix*, Fig. S6) whereas contrast sensitivity was measured with gratings in an orientation discrimination task (*SI Appendix*, Fig. S7).

The model reproduced the improvements to acuity and contrast sensitivity for each attention type. On the one hand, both exogenous and endogenous attention improve acuity similarly (6). Model simulations yielded consistent visual acuity improvements for both exogenous (Fig. 9*A*) and endogenous (Fig. 9*B*) attention, despite different SF gain profiles underlying each attention type. On the other hand, each type of attention alters contrast sensitivity across SF differently (26). Model simulations captured the differences between exogenous (Fig. 9*C*) and endogenous attention (Fig. 9*D*). The model reproduced the narrow SF bandwidth of exogenous attention that is centered on SFs higher than baseline tuning preferences (*SI Appendix*, Fig. S7D). It also captured the broad SF modulation by endogenous attention that spanned SFs above and below baseline tuning (*SI Appendix*, Fig. S7E). Attention effects derived from our observer model closely matched descriptive fits to the data from ref. 26 (Fig. 9 *C* and *D*).

The attention parameters were consistent across tasks (*SI Appendix*, Table S6). The SF bandwidth of endogenous attentional gain consistently spanned a larger range than exogenous attention (*SI Appendix*, Table S6, SF bw). Moreover, the rate at which SF selectivity declined with eccentricity also differed. The peak SF decreased with eccentricity (*SI Appendix*, Table S6, SF slope), but less so for exogenous than endogenous attention, indicating that exogenous attention consistently enhanced SFs higher than the peak SF of the stimulus drive (*SI Appendix*, Fig. S2). Lastly, we observed tradeoffs between the amplitude and spatial spread of attention (*SI Appendix*, Table S6). In the acuity task, the amplitude was large (>8) and the spatial spread was narrower (0.6°) than the stimulus (1°), whereas in contrast sensitivity, the amplitude was lower (<1.5) and the spatial spread was broader (>5°) than the stimulus (4°). Texture segmentation yielded intermediate values wherein the amplitude was ~4 for a fixed spread of 4°. Independent of attentional effects, differences in the experimental protocol and stimuli used across experiments resulted in subtle differences in the best-fitting model parameters for contrast gain and the stimulus drive. Importantly, similar attention parameters reproduce endogenous and exogenous attention effects in a variety of visual tasks.

Discussion

We used texture segmentation as a model system to dissociate endogenous and exogenous attention. To this end, we developed an image-computable model that reproduces human segmentation performance and the modulations by each attention type. This model links neural computations to three visual phenomena: 1) divisive normalization and spatial summation mediate the CPD (27–34, 43–46), 2) narrow high-SF enhancement drives exogenous attentional effects (27–32, 34), and 3) broad SF gain drives endogenous attentional modulations (32–34).

Normalization models of attention have described how spatial attention affects neural responses and behavior (e.g., refs. 14, 15, 17). Our model adopts the same algorithm specified by the Reynolds-Heeger NMA (15)—attentional gain modulates the stimulus
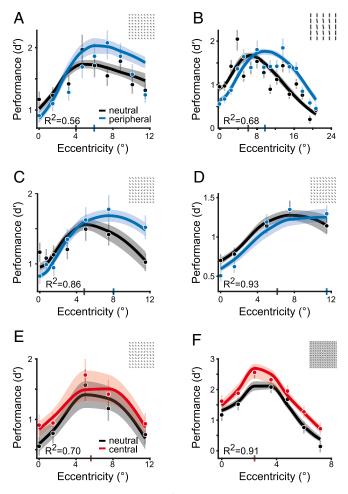

Fig. 8. A parsimonious explanation for several experimental manipulations in texture segmentation. (*A–D*) Narrow SF gain profile fit to exogenous attention experiments. The model jointly fits these data and those displayed in Fig. 5, with parameters shared among all six experiments. Insets in each panel depict the same textures displayed in Fig. 3 *C–F*, respectively. (*E* and *F*) Broad SF gain profile fit to endogenous attention experiments. The model jointly fits these data and those displayed in Fig. 6, with parameters shared among all four experiments. Insets in each panel depict the same textures displayed in Fig. 3 *I* and *J*, respectively. The dots and error bars depict group-average and ±1 SEM. The solid lines and shaded regions depict the median and 68% CIs of the bootstrapped distribution of model fits.

Jigo et al.
An image-computable model of how endogenous and exogenous attention differentially alter visual perception

PNAS | 7 of 12
https://doi.org/10.1073/pnas.2106436118
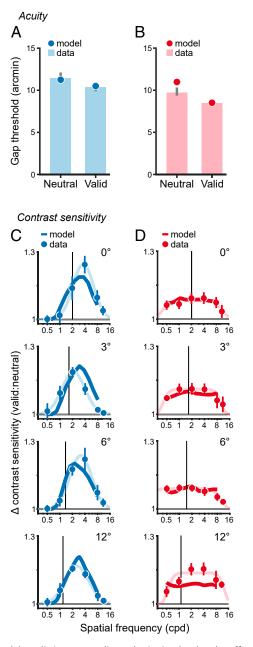
PSYCHOLOGICAL AND COGNITIVE SCIENCES

Fig. 9. Model predictions generalize to basic visual tasks. The effects of (*A*) exogenous and (*B*) endogenous attention on gap discrimination thresholds in an acuity task. Data from ref. 6. Lower thresholds indicate higher acuity. Bars depict group-average thresholds in neutral and valid cueing conditions. Error bars are ±1 SEM. Dots depict model-derived gap thresholds for the acuity task. (*C*) Exogenous and (*D*) endogenous attention effects on contrast sensitivity across SF and eccentricity, quantified as the ratio between valid and neutral contrast sensitivity. Data from ref. 26. Values above 1 indicate an attentional enhancement of contrast sensitivity. The dots and error bars depict the group-average and ±1 SEM. The vertical black lines show baseline SF preferences measured in the neural condition (*SI Appendix*, Fig. S7). The solid colored lines show model fits to the data, whereas lightly shaded lines are descriptive fits to the data from ref. 26. In all panels, the narrow SF profile was fit to exogenous attention effects, whereas the broad SF profile was fit to endogenous attention effects.

drive before divisive normalization. Predictions by NMA have been empirically confirmed with psychophysical experiments (66). These experiments equated seemingly distinct effects of endogenous and exogenous 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 do not explain the dissociation between endogenous and exogenous attention. Although the spatial extent of attention is critical for explaining effects on contrast sensitivity, our model comparisons demonstrate that it is not vital for reproducing attention effects on texture segmentation ("spatial extent" model in Fig. 7 and *SI Appendix*, Fig. S4). These results corroborate empirical evidence that manipulating the spread of attention during texture segmentation does not yield shifts between the typical effects of endogenous and exogenous attention (31).

To capture the effects of attention on texture segmentation, we implemented 1) eccentricity-dependent and SF-tuned multiplicative gains that emulate neural (54) and psychophysical (53) SF selectivity; 2) spatial summation, which emphasizes textural contours (39, 41, 47); and 3) distinct SF gain profiles for endogenous and exogenous attention (25, 26) that scale responses prior to normalization (15), thereby adjusting the balance between fine and coarse-scale visual sensitivity. The model's distinct SF profiles instantiate a computational dissociation between each attention type that substantiates their differential impact on sensory processing.

The necessity for different SF profiles is supported by empirical evidence (25, 26) and provides insights toward the distinct roles of endogenous and exogenous attention in guiding visual behavior. Previous models (e.g., refs. 14, 15, 17) demonstrate that both forms of attention improve low-level visual processes that encode elementary features (e.g., contrast, orientation, motion). Here, we show that attention differentially interacts with normalization to shape the competition inherent in midlevel processes such as texture segmentation. Exogenous attention preferentially enhances a narrow range of high SFs. Consequently, its effects prioritize fine-grained visual details at the expense of competing coarse-scale features within a stimulus. In contrast, endogenous attention consistently improves midlevel processing by broadly enhancing sensory encoding across fine and coarse spatial scales. The computations underlying midlevel processing bridge the gap between sensory encoding and object recognition (39–42). 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 model provides a computational framework for understanding the mechanisms underlying established effects of exogenous attention on spatial resolution (27–34) (reviews in refs. 1–3). Previous studies offered qualitative descriptions that exogenous attention automatically increases spatial resolution (27–32, 34) (reviews in refs. 1–3) with concomitant costs in temporal resolution (22) attributed to an engagement of parvocellular neurons (22, 67). Here, we develop an observer model that anchors these qualitative descriptions onto established neural computations. In doing so, we corroborate previous psychophysical experiments that found a similar high-SF preference of exogenous attention (25, 26, 30, 68), specify how attentional gain changes across the visual field and demonstrate its computational validity for explaining effects on perception.

We also provide converging evidence that exogenous attention alters perception inflexibly. By comparing the model's exogenous attentional gain on textures to empirical measurements made with gratings (26), we found that it consistently operates above intrinsic (i.e., baseline) SF preferences despite large differences in stimuli (*SI Appendix*, Fig. S2). These findings suggest that in addition to exogenous attentional effects being invariant to cue validity (8) and sometimes detrimental to perception (27–32, 34), its operating range across SF is also invariant to the type of stimulus being attended.

The model provides insights on the mechanisms underlying endogenous attention effects on spatial resolution. Previous research has established that endogenous attention modulates texture segmentation (18, 32–34, 69) and its impact has been described as an optimization of spatial resolution (reviews in refs. 1–3). We

propose that a broad SF gain control mechanism yields these perceptual improvements. Our proposal complements previous reports that endogenous attention uniformly excludes noise across SF (70), but seemingly conflicts with an earlier explanation that endogenous attention suppresses sensitivity to high SFs to improve texture segmentation (33). However, suppressed high-SF sensitivity at foveal locations would decrease cross-frequency suppression (59, 61) and result in an effective dominance of lower SFs, which is compatible with our findings (*SI Appendix*, Fig. S3).

Moreover, we provide converging evidence of the flexibility of endogenous attention. We found that the model's SF preference during texture segmentation differed from those measured with gratings (26). This discrepancy suggests that the impact of endogenous attention depends on the properties of the attended stimulus and the nature of the task, consistent with the notion of a flexible endogenous attentional mechanism (8, 32–34).

The effects of attention depend on divisive normalization. Without normalization, we could not reliably capture the CPD, which served as the foundation of our analyses. Previous studies demonstrate that when the pool of SFs contributing to normalization is restricted, the CPD is attenuated (30, 33, 44). However, existing models of the CPD (46, 48, 49) relate the phenomenon solely to an increase in RF size with eccentricity. Our model directly links the summation area of RFs to their SF tuning. Consequently, the dominant summation area increases with eccentricity as SF preferences decrease. Despite implementing an increase in RF size, we could not capture the CPD without accounting for the surrounding context via normalization.

Additionally, we demonstrate that spatial constraints mediate the CPD independently from limitations in temporal processing across eccentricity. The proposal that the CPD may result from slow information accrual at the fovea, which yields poor performance particularly when a backward mask limits processing time (43), has been criticized (45, 46, 71). We note that our model accounts equally well for the findings of texture segmentation studies regardless of whether they contained or omitted a mask, which minimized temporal contributions to task performance (*SI Appendix*, Table S5). Importantly, both endogenous and exogenous attention speed information accrual (72) across the visual field (73, 74) and across different levels of cue validity (8). Thus, effects of attention on temporal processing would predict similar improvements by each attention type on the CPD, a prediction clearly contradicted by the modeled studies here (27, 29–33).

The computations implemented in the model are based on the known properties of the human and nonhuman primate visual system. The stimulus drive simulates bottom-up responses of phase-invariant complex cells in V1 (56) that vary with SF and eccentricity (53, 54). 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 (75–77).

Exogenous attentional gain in the model result in changes to texture sensitivity; however, little is known about the neural underpinnings of these effects. There are sparse demonstrations of exogenous attentional modulations in visuo-occipital areas and beyond (20, 21, 78–80). Transcranial magnetic stimulation of early visual cortex reveals that its activity plays a key role in the generation of exogenous attention effects (81). However, future studies are required to determine how the SF gain modulation we report manifests in neural populations.

In contrast, it is established that endogenous attention modulates cortical responses (1, 2, 13, 18, 20, 21, 36–38, 82, 83). During texture segmentation tasks, endogenous attention selectively enhances V1 and V4 responses to the embedded figure, suggesting that attention spreads across the target object to facilitate its segmentation (18). Our model provides complementary evidence that endogenous attention optimizes SF sensitivity to improve segmentation across texture scale. Yet, it is unclear how neural activity generates these SF modulations. Neuroimaging

(37, 38) and electrophysiological (13, 36) recordings demonstrate that spatial tuning profiles are altered by endogenous attention. Such changes are consistent with, but not necessary for, the modulations of spatial resolution we report.

Few computational models have implemented possible ways in which attention alters spatial resolution. Some have proposed that attention modifies how finely a spatial region is analyzed. Such changes are either driven by an attention field that adjusts the spatial profile of RFs (13) or by attracting RFs toward and contracting them around the attended location (19). Other models suggest an attentional prioritization that selectively tunes responses for a given spatial location and attenuates responses to surrounding regions (12, 16). However, these models neither account for differences across eccentricity nor explain attentional shifts toward fine or coarse spatial scales. Critically, these models do not distinguish between endogenous and exogenous attention. In contrast to these previous models, we do not propose any modifications to the structure of RFs. Instead, we attribute changes in spatial resolution to modulations of SF, a fundamental dimension of early visual processing.

The fact that our model operates on arbitrary images facilitates its generalization to other visual stimuli and tasks. We show that the model reproduces the differential endogenous and exogenous attention effects on contrast sensitivity (Fig. 9 *C* and *D*). Notably, the model recreates behavior in visual acuity tasks where the improvements by each attention type are similar (Fig. 9 *A* and *B*). Unlike texture segmentation, acuity tasks always benefit from heightened spatial resolution, which obscures differences between these two attention types. Recent studies that compared both attention types head-to-head with the same observers, stimuli and task found that they produced similar behavioral effects but modulated neural activity differently in the temporo-parietal junction (20) and occipital cortex (21). Our model is consistent with these findings and highlights that differences in the underlying computations can yield similar perceptual effects between endogenous and exogenous attention depending on the stimulus and task.

Future work may extend the model to other visual phenomena. For instance, it could capture the differential effects by each attention type on second-order texture perception (28, 34), second-order texture contrast sensitivity (24) and temporal resolution (22, 23, 67). Lastly, it is unknown how interactions between both forms of attention may affect midlevel processes like texture segmentation. Endogenous attention attenuates the transient effects of exogenous attention on stimulus discriminability when both are deployed concurrently (84). Therefore, it is possible that endogenous attentional benefits will outweigh the costs induced by exogenous attention when both are deployed simultaneously during texture segmentation. Although the experimental designs of the studies we have modeled cannot address this open question, our model framework may facilitate predictions of the perceptual consequences when both forms of attention are deployed.

In conclusion, we reproduce signatures of texture segmentation (27–34, 43–46) and characterize the contributions of attention to a process commonly considered "preattentive" (39, 41, 42, 44–49). Moreover, we reveal the neural computations that underlie how attention modifies spatial resolution (1–3). Attention scales sensitivity to high and/or low SFs, adjusting the balance between fine and coarse-scale spatial resolution. Exogenous attention preferentially enhances fine details whereas endogenous attention uniformly enhances fine and coarse features to optimize task performance. Because the model distinguishes between endogenous and exogenous 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.

Methods

Extended methods are available in *SI Appendix*.

Jigo et al.
An image-computable model of how endogenous and exogenous attention differentially alter visual perception

PNAS | 9 of 12
https://doi.org/10.1073/pnas.2106436118

Model. We developed an observer model that simulates the response of a collection of RFs each narrowly tuned to spatial position (x,y), orientation (θ), and SF (f). Responses varied with eccentricity (α). The population response (R) is generated by four components: the stimulus drive (E), attentional gain (A), suppressive drive (S and σ), and spatial summation (F), where $*$ represents convolution:

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

All model parameters are given in *SI Appendix*, Table S1.

Stimulus Drive. The stimulus drive characterizes responses of linear RFs in the absence of suppression, attention, and spatial summation. A steerable pyramid (55) decomposed stimulus images into several SF and orientation subbands, defined by weighted sums of the image (i.e., linear filters). Weights were parameterized by raised-cosine functions that evenly tiled SFs, orientations and positions.

The number of SF and orientation subbands are parameters that can be flexibly chosen. We used a set of 30 subbands comprising five SF bands and six orientation bands. The size of the stimulus image and the subband bandwidth determine the total number of SF subbands. In our simulations, images were 160 × 160 pixels (*SI Appendix*, section S3) and SF bandwidth (i.e., full-width at half-maximum, FWHM) was 1 octave, which allowed for five different SF subbands. The chosen bandwidth is comparable to empirical tuning curves measured in primate electrophysiological recordings (85) and human psychophysical (53) measurements. The FWHM orientation bandwidth (60°) is comparable to physiological tuning curves measured in primates (86). Using narrower (30°) or wider (90°) bandwidths yielded similar results supporting the same conclusions.

The pyramid includes RFs in quadrature phase. We computed a "contrast energy" response (56) (i.e., the sum of squared responses across phase) which depends on the local spectral energy at each SF, orientation and position in the image. Contrast energy is fundamental to texture perception models (39, 41, 47), and we denote it as $C(f, \theta, x, y)$.

SF gain. Human (26, 53, 87) and nonhuman primate (54) contrast sensitivity is narrowly tuned to SF. SF tuning shifts from high to low SFs with eccentricity. To model this behavior, contrast energy is multiplied point-by-point by a SF gain function, T, defined by a log-parabola (88, 89):

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

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

$$\lambda_T(\alpha) = 2^{t_T - m_T\alpha} + t_{min}.$$ **[3]**

SF preferences converge onto a single value in the far periphery, t_{min} (87). The preferred SF at the fovea is given by $2^{t_T} + t_{min}$ and progressively shifts toward t_{min} at the rate m_T. Whereas t_T varied during simulations (*SI Appendix*, Tables S1–S4), 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 contrast energy responses that vary with SF and eccentricity, computed as

$$E(f, \theta, x, y) = T(f, \alpha)C(f, \theta, x, y).$$ **[4]**

Attentional Gain. Attention is implemented as an attentional gain field, A, that multiplies the stimulus drive point-by-point as in the Reynolds-Heeger NMA (15). Attentional gain was uniform across orientation. Across SF and position, gain was distributed according to a cosine function, w:

$$w(z; \mu, b) = \begin{cases} 0.5 + 0.5\cos\left(\frac{\pi[z - \mu]}{b}\right) & \\ 0 & \mu - b < z > \mu + b \end{cases},$$ **[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 position, they were in units of degrees of visual angle. 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 fell along the horizontal meridian at eccentricity α_{targ} (*SI Appendix*, section S3). The product of two cosine functions (w, Eq. 5) defined the spread of attention: one varied as a function of x and another as a function of y, each with an identical width b_{pos}. Widths did not vary across eccentricity. A_{pos} defined the spatial spread of attention:

$$A_{pos}(x, y) = w(x; \alpha_{targ}, b_{pos})w(y; 0, b_{pos}).$$ **[6]**

The precise spatial spread of attention is controversial (90) and can change based on task demands (66, 91). Critically, it 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 has been demonstrated to adjust the size of the attention field (66). Instead, a previous study (31) measured exogenous attention effects while manipulating the size of a peripheral precue. The authors found that exogenous 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° (*SI Appendix*, Table S1) because it encompassed the largest target size used to constrain model parameters (*SI Appendix*, Table S5). 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.

Narrow SF profile. In the narrow model (A_N), attentional gain was bandpass across SF (26). Attentional gain peaked at a given SF, λ_N, and fell gradually toward neighboring frequencies within its bandwidth, b_N, characterized by a cosine function:

$$A_N(f, \alpha) = w(f; \lambda_N(\alpha), b_N).$$ **[7]**

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

$$\lambda_N(\alpha) = 2^{a_N - m_N\alpha},$$ **[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 SF profile. The broad profile (A_B) implemented broadband attentional gain (26), characterized by the sum of three overlapping cosine functions:

$$A_B(f, \alpha) = w_1 + w_2 + w_3,$$ **[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, ensuring that their sum yielded a plateau spanning b_B octaves and a FWHM of $1.5b_B$.

In sum, attentional gain multiplicatively scaled the stimulus drive uniformly across orientation, but differently across SF and eccentricity given by

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

where A_{pos} and A_B (or A_N) were four-dimensional matrices characterizing attentional gain across position, SF and orientation. γ_B (or γ_N) defined attentional amplitude. To simulate the neutral cueing condition, amplitude was set to 1. 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 (15) across nearby positions, all orientations and neighboring SFs. This pooling procedure implemented lateral interactions between RFs and was computed via convolution (15). Convolution kernels were cosine functions (w, Eq. 5).

The bandwidth of the SF kernel, δ_f, equaled 1 octave:

$$K_f = \begin{cases} 1 & f_i - \delta_f \leq f_i \leq f_i + \delta_f \\ 0 & \text{otherwise} \end{cases},$$ **[11]**

where f_i denotes the center SF of a subband. This kernel summed contrast energy within and ± 1 octave around each SF subband.

Jigo et al.
An image-computable model of how endogenous and exogenous attention
differentially alter visual perception

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

$$K_\theta = \begin{cases} 1 & \theta_i - \delta_\theta \le \theta_i \le \theta_i + \delta_\theta \\ 0 & \text{otherwise} \end{cases}, \qquad [12]$$

where θ_i denotes the center orientation of a steerable pyramid subband. This kernel summed contrast energy across all orientations.

Spatial position kernels were determined by multiplying two cosine windows:

$$K_{pos}(x, y; f) = w(x; 0, \delta_{pos})w(y; 0, \delta_{pos}). \qquad [13]$$

One window varied across x, another across y, and their centers traversed across the image during convolution. The two-dimensional kernel summed to unity, which computed the average energy within the pooled area. Kernel width, δ_{pos}, equaled $\frac{2}{f}$ and was inversely proportional to subband SF f and yielded two-dimensional spatial kernels, K_{pos}. Kernel widths were identical across eccentricity.

Contextual modulation was characterized via separable convolution:

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

where * denotes convolution of the suppression kernels, K. Suppression magnitude was adjusted across eccentricity by σ^2, which controlled the level of contrast at which neural responses reached half-maximum and is referred to as contrast gain. Contrast gain was implemented as an exponential function across eccentricity (26, 87):

$$\sigma^2(\alpha) = 10^{2(g_\sigma - m_\sigma \alpha)}, \qquad [15]$$

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.

Spatial Summation. Following divisive normalization, responses were weighted and summed across space, within each SF and orientation subband. Summation was accomplished via convolution by cosine functions, F, computed using Eq. 13. The width of each filter scaled with SF: narrow (wide) regions of space were pooled for high (low) SFs (39) and did not vary with eccentricity.

Decision Mechanism. We used signal detection theory to relate population responses to behavioral performance (d'). The available signal s was computed as the Euclidean norm of the difference between target-present (r_t) and target-absent (r_n) population responses: $s = ||r_t - r_n||$. Performance on a discrimination task is proportional to the neural responses given the assumption of additive, independent and identically distributed noise. An alternative model with Poisson noise and a maximum-likelihood decision rule yields the same linkage between neural response and behavioral performance (92, 93). The signal and noise magnitude (σ_n) defined behavioral performance $d' = \frac{s}{\sigma_n}$. $\sigma_n = \overline{r_{neutral}} / \overline{s_{neutral}}$ where $\overline{r_{neutral}}$ denotes the observed neutral performance averaged across eccentricity and $\overline{s_{neutral}}$ denotes the eccentricity-average of the signal. This ratio scaled the model's predicted behavioral performance to match the observed data.

Model Fitting. Models were optimized by minimizing the residual sum of squared error between model and behavioral d' using Bayesian adaptive direct search [BADS (94)]. When applicable, performance data for a psychophysical experiment were converted from proportion correct, p, to d' with the assumption of no interval bias (95): $d' = \sqrt{2}z(p)$, where z denotes the inverse normal distribution. Although performance on 2IFC tasks can exhibit biases between intervals (96), our conversion algorithm operated uniformly across eccentricity, which preserved the performance variation (i.e., the CPD) critical for the goals of this study.

Model Alternatives. To assess whether contextual modulation and spatial summation are critical for the CPD, we implemented five model variants. Individual components of the suppressive drive were iteratively removed: cross-orientation suppression ("-θ"), cross-frequency suppression ("-f"), surround suppression ("-x,y") and all components simultaneously ("-all"). In a separate variant, spatial summation was removed ("-sum"). We fit each variant separately to neutral performance data from all 10 psychophysical experiments using the configuration described in *SI Appendix*, section S1.1).

In the "-all" 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}; Eqs. 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 Eq. 1) from the model.

We additionally compared the efficacy of each attentional gain profile across SF—narrow or broad—in generating the effects of exogenous and endogenous attention by fitting each profile to exogenous and endogenous attention experiments. To assess the explanatory power of the spatial extent of attention, a third model was compared in which the spatial spread of attention (b_{pos}, Eq. 6) varied between experiments and the gain across SF was uniform. Each model fit followed the configurations described in *SI Appendix*, section S1.2.

Data Availability. Data and code for model fitting and plotting are available at https://github.com/CarrascoLab/modelCPD.

1. M. Carrasco, Visual attention: The past 25 years. *Vision Res.* **51**, 1484–1525 (2011).
2. K. Anton-Erxleben, M. Carrasco, Attentional enhancement of spatial resolution: Linking behavioural and neurophysiological evidence. *Nat. Rev. Neurosci.* **14**, 188–200 (2013).
3. M. Carrasco, A. Barbot, How attention affects spatial resolution. *Cold Spring Harb. Symp. Quant. Biol.* **79**, 149–160 (2014).
4. Y. Yeshurun, M. Carrasco, Spatial attention improves performance in spatial resolution tasks. *Vision Res.* **39**, 293–306 (1999).
5. M. Carrasco, P. E. Williams, Y. Yeshurun, Covert attention increases spatial resolution with or without masks: Support for signal enhancement. *J. Vis.* **2**, 467–479 (2002).
6. B. Montagna, F. Pestilli, M. Carrasco, Attention trades off spatial acuity. *Vision Res.* **49**, 735–745 (2009).
7. M. Carrasco, Y. Yeshurun, The contribution of covert attention to the set-size and eccentricity effects in visual search. *J. Exp. Psychol. Hum. Percept. Perform.* **24**, 673–692 (1998).
8. A. M. Giordano, B. McElree, M. Carrasco, On the automaticity and flexibility of covert attention: A speed-accuracy trade-off analysis. *J. Vis.* **9**, 30 (2009).
9. M. Carrasco, S. Ling, S. Read, Attention alters appearance. *Nat. Neurosci.* **7**, 308–313 (2004).
10. S. Ling, M. Carrasco, Transient covert attention does alter appearance: A reply to Schneider (2006). *Percept. Psychophys.* **69**, 1051–1058 (2007).
11. T. Liu, J. Abrams, M. Carrasco, Voluntary attention enhances contrast appearance. *Psychol. Sci.* **20**, 354–362 (2009).
12. J. K. Tsotsos et al., Modeling visual attention via selective tuning. *Artif. Intell.* **78**, 507–545 (1995).
13. T. Womelsdorf, K. Anton-Erxleben, S. Treue, Receptive field shift and shrinkage in macaque middle temporal area through attentional gain modulation. *J. Neurosci.* **28**, 8934–8944 (2008).
14. G. M. Boynton, A framework for describing the effects of attention on visual responses. *Vision Res.* **49**, 1129–1143 (2009).
15. J. H. Reynolds, D. J. Heeger, The normalization model of attention. *Neuron* **61**, 168–185 (2009).
16. J. K. Tsotsos, *A Computational Perspective on Visual Attention* (MIT Press, Cambridge, MA, 2011).
17. A. M. Ni, S. Ray, J. H. Maunsell, Tuned normalization explains the size of attention modulations. *Neuron* **73**, 803–813 (2012).
18. J. Poort et al., The role of attention in figure-ground segregation in areas V1 and V4 of the visual cortex. *Neuron* **75**, 143–156 (2012).
19. O. Baruch, Y. Yeshurun, Attentional attraction of receptive fields can explain spatial and temporal effects of attention. *Vis. Cogn.* **22**, 704–736 (2014).
20. L. Dugué, E. P. Merriam, D. J. Heeger, M. Carrasco, Specific visual subregions of TPJ mediate reorienting of spatial attention. *Cereb. Cortex* **28**, 2375–2390 (2018).
21. L. Dugué, E. P. Merriam, D. J. Heeger, M. Carrasco, Differential impact of endogenous and exogenous attention on activity in human visual cortex. *Sci. Rep.* **10**, 21274 (2020).
22. Y. Yeshurun, L. Levy, Transient spatial attention degrades temporal resolution. *Psychol. Sci.* **14**, 225–231 (2003).
23. E. Hein, B. Rolke, R. Ulrich, Visual attention and temporal discrimination: Differential effects of automatic and voluntary cueing. *Vis. Cogn.* **13**, 29–50 (2006).
24. A. Barbot, M. S. Landy, M. Carrasco, Differential effects of exogenous and endogenous attention on second-order texture contrast sensitivity. *J. Vis.* **12**, 6 (2012).

Jigo et al.
An image-computable model of how endogenous and exogenous attention differentially alter visual perception

25. A. Fernandez, S. Okun, M. Carrasco, Differential effects of endogenous and exogenous attention on sensory tuning. *bioRxiv* [Preprint] https://doi.org/10.1101/2021.04.03.438325 (Accessed 4 April 2021).

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

27. Y. Yeshurun, M. Carrasco, Attention improves or impairs visual performance by enhancing spatial resolution. *Nature* **396**, 72–75 (1998).

28. Y. Yeshurun, M. Carrasco, The locus of attentional effects in texture segmentation. *Nat. Neurosci.* **3**, 622–627 (2000).

29. C. P. Talgar, M. Carrasco, Vertical meridian asymmetry in spatial resolution: Visual and attentional factors. *Psychon. Bull. Rev.* **9**, 714–722 (2002).

30. M. Carrasco, F. Loula, Y.-X. Ho, How attention enhances spatial resolution: Evidence from selective adaptation to spatial frequency. *Percept. Psychophys.* **68**, 1004–1012 (2006).

31. Y. Yeshurun, M. Carrasco, The effects of transient attention on spatial resolution and the size of the attentional cue. *Percept. Psychophys.* **70**, 104–113 (2008).

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

33. A. Barbot, M. Carrasco, Attention modifies spatial resolution according to task demands. *Psychol. Sci.* **28**, 285–296 (2017).

34. M. Jigo, M. Carrasco, Attention alters spatial resolution by modulating second-order processing. *J. Vis.* **18**, 2 (2018).

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

36. T. Womelsdorf, K. Anton-Erxleben, F. Pieper, S. Treue, Dynamic shifts of visual receptive fields in cortical area MT by spatial attention. *Nat. Neurosci.* **9**, 1156–1160 (2006).

37. J. Fischer, D. Whitney, Attention narrows position tuning of population responses in V1. *Curr. Biol.* **19**, 1356–1361 (2009).

38. B. P. Klein, B. M. Harvey, S. O. Dumoulin, Attraction of position preference by spatial attention throughout human visual cortex. *Neuron* **84**, 227–237 (2014).

39. M. S. Landy, N. Graham, "Visual perception of texture" in *The Visual Neurosciences*, L. M. Chalupa, J. S. Werner, Eds. (MIT Press, Cambridge, MA, 2004), pp. 1106–1118.

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

41. M. S. Landy, "Texture analysis and perception" in *The New Visual Neurosciences*, J. S. Werner, L. M. Chalupa, Eds. (MIT Press, Cambrige, MA, 2013), pp. 639–652.

42. J. D. Victor, M. M. Conte, C. F. Chubb, Textures as probes of visual processing. *Annu. Rev. Vis. Sci.* **3**, 275–296 (2017).

43. L. Kehrer, Central performance drop on perceptual segregation tasks. *Spat. Vis.* **4**, 45–62 (1989).

44. K. Morikawa, Central performance drop in texture segmentation: The role of spatial and temporal factors. *Vision Res.* **40**, 3517–3526 (2000).

45. C. Potechin, R. Gurnsey, Backward masking is not required to elicit the central performance drop. *Spat. Vis.* **16**, 393–406 (2003).

46. R. Gurnsey, D. Di Lenardo, C. Potechin, Backward masking and the central performance drop. *Vision Res.* **44**, 2587–2596 (2004).

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

48. L. Kehrer, The central performance drop in texture segmentation: A simulation based on a spatial filter model. *Biol. Cybern.* **77**, 297–305 (1997).

49. L. Kehrer, C. Meinecke, A space-variant filter model of texture segregation: Parameter adjustment guided by psychophysical data. *Biol. Cybern.* **88**, 183–200 (2003).

50. A. Thielscher, H. Neumann, A computational model to link psychophysics and cortical cell activation patterns in human texture processing. *J. Comput. Neurosci.* **22**, 255–282 (2007).

51. D. J. Heeger, Normalization of cell responses in cat striate cortex. *Vis. Neurosci.* **9**, 181–197 (1992).

52. M. Carandini, D. J. Heeger, Normalization as a canonical neural computation. *Nat. Rev. Neurosci.* **13**, 51–62 (2011).

53. N. V. S. Graham, *Visual Pattern Analyzers* (Oxford University Press, New York, 1989).

54. R. L. DeValois, K. K. DeValois, *Spatial Vision* (Oxford University Press, New York, ed. 2, 1990).

55. E. P. Simoncelli, W. T. Freeman, E. H. Adelson, D. J. Heeger, Shiftable multiscale transforms. *IEEE Trans. Inf. Theory* **38**, 587–607 (1992).

56. E. H. Adelson, J. R. Bergen, Spatiotemporal energy models for the perception of motion. *J. Opt. Soc. Am. A* **2**, 284–299 (1985).

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

58. G. J. Brouwer, D. J. Heeger, Cross-orientation suppression in human visual cortex. *J. Neurophysiol.* **106**, 2108–2119 (2011).

59. Y. Petrov, M. Carandini, S. McKee, Two distinct mechanisms of suppression in human vision. *J. Neurosci.* **25**, 8704–8707 (2005).

60. N. J. Priebe, D. Ferster, Mechanisms underlying cross-orientation suppression in cat visual cortex. *Nat. Neurosci.* **9**, 552–561 (2006).

61. D. Sagi, S. Hochstein, Lateral inhibition between spatially adjacent spatial-frequency channels? *Percept. Psychophys.* **37**, 315–322 (1985).

62. T. S. Meese, D. J. Holmes, Spatial and temporal dependencies of cross-orientation suppression in human vision. *Proc. Biol. Sci.* **274**, 127–136 (2007).

63. Z. M. Westrick, M. S. Landy, Pooling of first-order inputs in second-order vision. *Vision Res.* **91**, 108–117 (2013).

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

65. G. Schwarz, Estimating the dimension of a model. *Ann. Stat.* **6**, 461–464 (1978).

66. K. Herrmann, L. Montaser-Kouhsari, M. Carrasco, D. J. Heeger, When size matters: Attention affects performance by contrast or response gain. *Nat. Neurosci.* **13**, 1554–1559 (2010).

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

68. N. Megna, F. Rocchi, S. Baldassi, Spatio-temporal templates of transient attention revealed by classification images. *Vision Res.* **54**, 39–48 (2012).

69. C. Casco, A. Grieco, G. Campana, M. P. Corvino, G. Caputo, Attention modulates psychophysical and electrophysiological response to visual texture segmentation in humans. *Vision Res.* **45**, 2384–2396 (2005).

70. Z.-L. Lu, B. A. Dosher, Spatial attention excludes external noise without changing the spatial frequency tuning of the perceptual template. *J. Vis.* **4**, 955–966 (2004).

71. R. Gurnsey, P. Pearson, D. Day, Texture segmentation along the horizontal meridian: Nonmonotonic changes in performance with eccentricity. *J. Exp. Psychol. Hum. Percept. Perform.* **22**, 738–757 (1996).

72. M. Carrasco, B. McElree, Covert attention accelerates the rate of visual information processing. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 5363–5367 (2001).

73. M. Carrasco, A. M. Giordano, B. McElree, Temporal performance fields: Visual and attentional factors. *Vision Res.* **44**, 1351–1365 (2004).

74. M. Carrasco, A. M. Giordano, B. McElree, Attention speeds processing across eccentricity: Feature and conjunction searches. *Vision Res.* **46**, 2028–2040 (2006).

75. J. Larsson, M. S. Landy, D. J. Heeger, Orientation-selective adaptation to first- and second-order patterns in human visual cortex. *J. Neurophysiol.* **95**, 862–881 (2006).

76. Y. El-Shamayleh, J. A. Movshon, Neuronal responses to texture-defined form in macaque visual area V2. *J. Neurosci.* **31**, 8543–8555 (2011).

77. L. E. Hallum, M. S. Landy, D. J. Heeger, Human primary visual cortex (V1) is selective for second-order spatial frequency. *J. Neurophysiol.* **105**, 2121–2131 (2011).

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

79. L. Busse, S. Katzner, S. Treue, Temporal dynamics of neuronal modulation during exogenous and endogenous shifts of visual attention in macaque area MT. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 16380–16385 (2008).

80. F. Wang, M. Chen, Y. Yan, L. Zhaoping, W. Li, Modulation of neuronal responses by exogenous attention in macaque primary visual cortex. *J. Neurosci.* **35**, 13419–13429 (2015).

81. A. Fernández, M. Carrasco, Extinguishing exogenous attention via transcranial magnetic stimulation. *Curr. Biol.* **30**, 4078–4084.e3 (2020).

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

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

84. M. A. Grubb, A. L. White, D. J. Heeger, M. Carrasco, Interactions between voluntary and involuntary attention modulate the quality and temporal dynamics of visual processing. *Psychon. Bull. Rev.* **22**, 437–444 (2015).

85. R. L. De Valois, D. G. Albrecht, L. G. Thorell, Spatial frequency selectivity of cells in macaque visual cortex. *Vision Res.* **22**, 545–559 (1982).

86. D. L. Ringach, R. M. Shapley, M. J. Hawken, Orientation selectivity in macaque V1: Diversity and laminar dependence. *J. Neurosci.* **22**, 5639–5651 (2002).

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

88. A. B. Watson, A. J. Ahumada Jr, A standard model for foveal detection of spatial contrast. *J. Vis.* **5**, 717–740 (2005).

89. L. A. Lesmes, Z. L. Lu, J. Baek, T. D. Albright, Bayesian adaptive estimation of the contrast sensitivity function: The quick CSF method. *J. Vis.* **10**, 17.1–21 (2010).

90. Y. Yeshurun, The spatial distribution of attention. *Curr. Opin. Psychol.* **29**, 76–81 (2019).

91. S. Van der Stigchel et al., The limits of top-down control of visual attention. *Acta Psychol. (Amst.)* **132**, 201–212 (2009).

92. M. Jazayeri, J. A. Movshon, Optimal representation of sensory information by neural populations. *Nat. Neurosci.* **9**, 690–696 (2006).

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

94. L. Acerbi, W. J. Ma, Practical Bayesian optimization for model fitting with Bayesian adaptive direct search. *Adv. Neural Inf. Process. Syst.* **30**, 1836–1846 (2017).

95. N. A. Macmillan, C. D. Creelman, *Detection Theory: A User's Guide* (Lawrence Erlbaum Associates, Mahwah, NJ, 2005).

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