Attention Modulates Human Visual Responses to Objects by Tuning Sharpening

Narges Doostani
School of Cognitive Sciences, Institute for Research in Fundamental Sciences, Tehran, Iran

Gholam-Ali Hossein-Zadeh
School of Electrical Engineering, University of Tehran, Tehran, Iran
School of Cognitive Sciences, Institute for Research in Fundamental Sciences, Tehran, Iran

Radoslaw Martin Cichy
Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany
Berlin School of Mind and Brain, Faculty of Philosophy, Humboldt-Universität zu Berlin, Berlin, Germany
Bernstein Center for Computational Neuroscience Berlin, Berlin, Germany

Maryam Vaziri-Pashkam
Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, USA

Corresponding Authors: Narges Doostani, Maryam Vaziri-Pashkam
Email: narges.doostani.d@gmail.com, maryam.vaziri-pashkam@nih.gov

The authors declare no competing interests.

Abstract

Visual stimuli compete with each other for cortical processing and attention biases this competition in favor of the attended stimulus. How does the relationship between the stimuli affect the strength of this attentional bias? Here, we used functional MRI to explore the effect of target-distractor similarity in neural representation on attentional modulation in the human visual cortex using univariate and multivariate pattern analyses. Using stimuli from four object categories (human bodies, cats, cars and houses), we investigated attentional effects in the primary visual area V1, the object-selective regions LO and pFs, the body-selective region EBA, and the scene-selective region PPA. We demonstrated that the strength of the attentional bias towards the target is not fixed but decreases with increasing distractor-target similarity. Simulations provided evidence that this result pattern is explained by tuning sharpening rather than an increase in gain. Our findings provide a mechanistic explanation for behavioral effects of target-distractor similarity on attentional biases and suggest tuning sharpening as the underlying mechanism in object-based attention.

Keywords: object-based attention, biased competition, target-distractor similarity, tuning sharpening, response gain, , human visual cortex, fMRI

Introduction

Everyday visual scenes typically contain a large number of stimuli. Since processing all the incoming information is impossible due to the brain's limited neural resources, different stimuli compete for cortical representation and processing[1,2,3,4,5,6]. This competition can be biased by the top-down signal of attention to enhance the parts of input that are most relevant to the task at hand[1,3,7]. Evidence from electrophysiology and fMRI studies have demonstrated the role of attention in biasing the competition by enhancing the response related to the attended stimulus[3,5,6] by approximately 30% compared to its response when unattended, in both electrophysiology studies of the monkey brain[3,8,9,10,11] and fMRI studies of the human brain[6].

Competition and attentional bias likely depend on the nature of the visual scenes rather than being universally uniform. Behavioral studies indicate that the competition between stimuli is content-dependent[12], with higher competition between stimuli that are located closer to each other[13], or between stimuli with more similar cortical representation patterns[12,14]. This suggests that the attentional bias might also be affected by the relationship between the competing stimuli, such as the similarity of their cortical representation. Further, behavioral studies on the effect of target-distractor similarity on performance have proposed that lower performance for more similar target-distractor pairs is due to the fact that the neural resources needed for detailed processing are shared to a greater extent[12]. However, a direct neuroscientific investigation of how target-distractor similarity affects visual representations, and a mechanistic explanation of how shared resources affect attentional biases is missing.

Here, we investigated the impact of similarity in cortical representation on attentional bias and the underlying mechanism with empirical and theoretical tools. First, using functional MRI and uni- as well as multivariate analysis, we investigated how the top-down effect of attention varies as target-distractor similarity changes for multiple presented objects. Specifically, we found that the strength of the attentional bias towards the target decreases with increasing target-distractor similarity in cortical representation.

Second, using simulations of neuronal populations we determined how this effect arises from attentional enhancement of neural responses. We considered two known mechanisms through which attention affects neural firing rate: response gain and tuning sharpening. The response gain model predicts a multiplicative scaling of responses through which neural responses are increased by a gain factor[15,16]. The tuning sharpening model instead proposes that attentional enhancement depends on the neuronal tuning for the attended stimulus, leading to an increase in response for optimal stimuli, and little change in response or at times even response suppression for non-optimal stimuli[17]. We find that the empirically-observed relationship between attentional enhancement and target-distractor similarity are predicted by the tuning sharpening model, but not the response gain model.

Together, our results show that attentional enhancement is dependent on the similarity between the target and the distractor in neural representation, and a more similar distractor causes the target to receive less attentional bias in the competition. Moreover, these results suggest tuning sharpening as the underlying mechanism of attentional enhancement during object-based attention.

Materials and Methods

Main experiment

Participants

17 healthy human participants (9 females, age: mean $\pm$ s.d. $= 29.29 \pm 4.5$ years) with normal or corrected-to-normal vision took part in the study. Participants gave written consent and received payment for their participation in the experiment. Data collection was approved by the Ethics Committee of the Institute for Research in Fundamental Sciences, Tehran.

The behavioral data for two participants was not correctly saved during the scanning due to technical problems. While we used the fMRI data of these two participants, all behavioral reports include the performance of the 15 participants for whom the behavioral data was properly saved.

Stimulus set and experimental design

To determine the effect of target-distractor similarity on attentional modulation, we used object stimuli from four categories (human bodies, cars, houses, and cats). We presented stimuli from each category in semi-transparent form, either in isolation (isolated conditions), or paired with stimuli from another category (paired conditions). Thus, the experiment consisted of 16 conditions: 4 isolated conditions in which isolated stimuli from one of the four categories were presented, and 12 paired conditions (6 category pairs $\times$ 2 attentional targets for each pair) in which a target stimulus from the cued category was superimposed with a distractor stimulus from another category for all category combinations. Figure 1B depicts all stimulus conditions. We used isolated conditions to assess the similarity between different categories, and paired conditions to determine the effect of similarity in a category pair on attentional modulation.

The stimulus set consisted of gray-scaled images from the four object categories of human bodies, cats, cars and houses, similar to stimuli used in previous studies [18,19,20]. Each category consisted of 10 exemplars all varying in identity, 3D-orientation (for houses and cars), and pose (for bodies and cats, see Figure 1A).

Images were presented in a gray background square presented at the center of the screen, subtending 10.2^{o} of visual angle. A red fixation point subtending 0.45^{o} of visual angle was presented at the center of the screen throughout the run (Figure 1C).

Procedure

We used a blocked design for the main experiment. At the beginning of each block, participants were cued by a word to attend to either bodies, cars, houses, or cats. During the block, participants maintained attention on the images from the cued category, and performed a one-back repetition detection task on them by pressing the response button when the same stimulus from the attended category appeared in two consecutive trials. Repetition occurred 2-3 times at random times in each block. The experiment consisted of 16 block types, corresponding to the 16 task conditions (Figure 1B).

Each block lasted for 10 s, starting with the cue word presented for 1 s, followed by 1 s of fixation. Then, ten images from the cued category were presented in isolation or paired with ten images from another category. Each image was presented for 400 ms, followed by 400 ms of fixation (Figure 1C). There were 8 s of fixation in between the blocks, and a final 8-s fixation after the last block.

We organized blocks in runs, each lasting 4 min 56 s. Each run started with 8 s of fixation followed by block presentations. The presentation order of the 16 task conditions was counterbalanced across each experimental run. 10 participants completed 16 runs and 7 participants completed 12 runs of the main experiment.

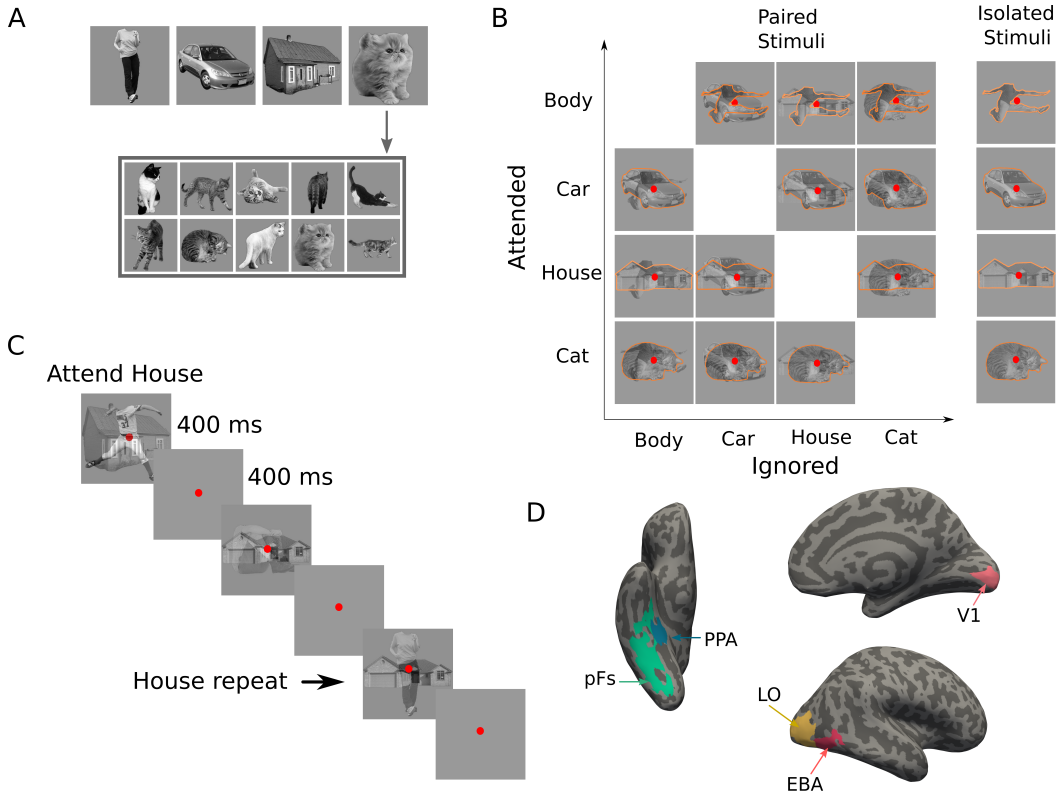

Figure 1: Stimuli, paradigm and regions of interest.
(A) Top images represent the four categories used in the main experiment: body, car, house and cat. The stimulus set consisted of 10 exemplars from each category (here: cats), with exemplars differing in pose and 3D-orientation. (B) The experimental design comprised 16 task conditions (12 paired, 4 isolated). The 4×4 matrix on the left illustrates the 12 paired conditions, with the to-be-attended category (outlined in orange for illustration purposes, not present in the experiment) on the y-axis and the to-be-ignored category on the x-axis. The right column illustrates the four isolated conditions. (C) Experimental paradigm. A paired block is depicted with superimposed body and house stimuli. In this example block, house stimuli were cued as target, and the participant responded on the repetition of the exact same house in two consecutive trials, as marked here by the arrow. (D) Regions of interest for an example participant; the primary visual cortex V1, the object-selective regions LO and pFs, the body-selective region EBA, and the scene-selective region PPA.

Localizer experiments

We investigated five different regions of interest: the primary visual cortex (V1), the object-selective areas lateral occipital cortex (LO) and posterior fusiform (pFs), the body-selective extrastriate body area (EBA), and the scene-selective parahippocampal place area (PPA). To define these regions of interest (ROIs), each participant completed four localizer runs described in detail below.

Early visual area localizer

We used meridian mapping to localize the primary visual cortex V1. Participants viewed a black-and-white checkerboard pattern through a 60 degree polar angle wedge aperture. The wedge was presented either horizontally or vertically. Participants were asked to detect luminance changes in the wedge in a blocked-design paradigm. Each run consisted of four horizontal and four vertical blocks, each lasting 16 s, with 16 s of fixation in between. A final 16 s fixation followed the last block. Each run lasted 272 s. The order of the blocks was counterbalanced within each run. Participants completed two runs of this localizer.

Category localizer

We used a category localizer to localize the cortical regions selective to scenes (PPA), bodies (EBA), and objects (LO, pFs). In a blocked-design paradigm, participants viewed stimuli from the five categories of faces, scenes, objects, bodies, and scrambled images. Each localizer run contained two 16-s blocks of each category, with the presentation order counterbalanced within each run. An 8-s fixation period was presented at the beginning, in the middle, and at the end of the run. In each block, 20 stimuli from the same category were presented. Stimuli were presented for 750 ms followed by 50 ms of fixation on a gray background screen. Participants were asked to maintain their fixation on a red circle at the center of the screen throughout and press a key when they detected a slight jitter in the stimuli that happened 2-3 times per block. Each run lasted 344 s. Participants completed two runs of this localizer.

Stimulus presentation inside the scanner

We back-projected the stimuli onto a screen positioned at the rear of the magnet using an LCD projector with a refresh rate of 60 Hz and a spatial resolution of 768×1024. Participants observed the screen through a mirror attached to the head coil.

MRI data acquisition

We recorded the data of 10 participants using the Siemens 3T Tim Trio MRI system with a 32-channel head coil at the Institute for Research in Fundamental Sciences (IPM). We collected the data of 7 additional participants on a Siemens Prisma MRI system using a 64-channel head coil at the National Brain-mapping Laboratory (NBML). For each participant, we performed a whole-brain anatomical scan using a T1-weighted MPRAGE sequence. For the functional scans, including the main experiment and the localizer experiments, we acquired 33 slices parallel to the AC-PC line using $T2^*$-weighted gradient-echo echo-planar imaging (EPI) sequences covering the whole brain (TR=2 s, TE=30 ms, flip angle = 90°, voxel size=$3 \times 3 \times 3$ mm^3, matrix size = 64×64).

fMRI data preprocessing

We performed fMRI data analysis using FreeSurfer (https://surfer.nmr.mgh.harvard.edu), Freesurfer Functional Analysis Stream (FsFast [21]) and in-house MATLAB codes. fMRI data preprocessing steps included 3D motion correction, slice timing correction, and linear and quadratic trend removal. We performed no spatial smoothing on the data. We used a double gamma function to model the hemodynamic response function. We eliminated the first four volumes (8 s) of each run to avoid the initial magnetization transient.

fMRI data analysis

For the main experiment, we performed a general linear model (GLM) analysis for each participant to estimate voxel-wise regression coefficients in each of the 16 task conditions. The onset and duration of each block were convolved with a hemodynamic response function and were then entered to the GLM as regressors. We also included movement parameters and linear and quadratic nuisance regressors in the GLM. We used these voxel-wise coefficients from the regions of interest (ROIs) as the basis for all further analyses.

For the early visual area localizer experiment, we estimated voxel regression coefficients in each of the two conditions (i.e., vertical and horizontal wedge) using a separate GLM. After convolving with a hemodynamic response function, the onset and duration of each block were entered to the GLM as regressors of interest. We also included movement parameters and linear and quadratic nuisance regressors in the GLM. We used the obtained coefficients to define the V1 ROI.

For the category localizer, we used another GLM to estimate voxel-wise regression coefficients in the five task conditions (i.e. faces, scenes, objects, bodies, and scrambled images). The GLM procedure was similar to the other two experiments. We then used these estimated coefficients to define the LO, pFs, EBA, and PPA ROIs.

Definition of ROIs

We determined the V1 ROI using a contrast of horizontal versus vertical polar angle wedges that reveals the topographic maps in the occipital cortex[22,23]. To define the object-selective areas LO in the lateral occipital cortex and pFs in the posterior fusiform gyrus[24,25], we used a contrast of objects versus scrambled images. Active voxels in the lateral occipital and ventral occipitotemporal cortex were selected as LO and pFS, respectively, following the procedure described by Kourtzi and Kanwisher[26]. We used a contrast of scenes versus objects for defining the scene-selective area PPA in the parahippocampal gyrus[27], and a contrast of bodies versus objects for defining the body-selective area EBA in the lateral occipitotemporal cortex[28]. We thresholded the activation maps for both the early visual localizer and the category localizer at $p < 0.001$, uncorrected.

Univariate fMRI analysis

We first used a univariate analysis to determine the effect of attention for different category pairs. Using the voxel-wise coefficients of the isolated conditions associated with each category, we examined the relative response of each voxel to the two categories for each category pair. This relative response determined which of the two categories was more preferred by the voxel. Therefore, for each category pair and each voxel, the category that elicited a higher response in the isolated condition was assigned the relatively more preferred category (M) label and the other the relatively less preferred category (L) label.

Univariate distance based on the isolated conditions

We had 6 pairs of categories: Body-Car, Body-House, Body-Cat, Car-House, Car-Cat and House-Cat. As a measure of the difference between the response evoked by each of the two categories in a pair, we defined a univariate distance. We calculated the univariate distance for each pair of categories simply as the difference in voxel responses of the two isolated conditions (Equation 1):

$$Univariate\ distance = R_{M^{at}} - R_{L^{at}} \tag{1}$$

Here, R denotes the average voxel response across runs, and the subscripts M and L denote the presence of the more preferred and the less preferred stimuli, respectively. The superscript at denotes the attended stimulus. Note that in the isolated conditions, the presented stimulus was always attended. Thus, $R_{M^{at}}$ is the average response related to the isolated preferred stimulus, and $R_{L^{at}}$ is the average response to the isolated less preferred stimulus. For example, the Body-Car univariate distance was assessed by $R_{Body^{at}} - R_{Car^{at}}$ for voxels more responsive to bodies than cars, and by $R_{Car^{at}} - R_{Body^{at}}$ for voxels more responsive to cars than bodies. Thus, according to this measure, two categories that elicited closer responses had less univariate distance, indicating more similarity in univariate response between the two categories.

Univariate attentional modulation based on the paired conditions

For each of the 6 category pairs, we had two paired conditions, in which stimuli from both categories were presented, but with attention directed to either one or the other category (for example, $Body^{at}Car$ and $BodyCar^{at}$ conditions for the Body-Car pair, with the superscript at denoting the to-be-attended stimulus). Since these paired conditions differed only in the attentional target and not in the stimuli, any difference observed in cortical response can be uniquely ascribed to the shift in

attention[18,29,30,31]. We thus defined attentional modulation for each pair of categories as the change in response when attention shifted from the more preferred stimulus to the other:

$$Attentional\ modulation = R_{M^{at}L} - R_{ML^{at}} \tag{2}$$

Here, $R_{M^{at}L}$ denotes the response related to the paired condition with attention directed to the more preferred category, while $R_{ML^{at}}$ is the elicited response when attending the less preferred category in the pair. For example, considering the Body-Car pair, attentional modulation was assessed by $R_{Body^{at}Car} - R_{BodyCar^{at}}$ for a voxel preferring bodies to cars, and by $R_{BodyCar^{at}} - R_{Body^{at}Car}$ for a voxel preferring cars to bodies.

Multivariate pattern analysis

To determine the effect of attention at the multivariate level, and to examine the attentional bias that the representation of each stimulus receives, we used a multivariate pattern analysis. Here, rather than comparing the mean values of voxel-wise coefficients in each ROI, we instead considered the ROI response pattern in each condition as a response vector, with the voxel-wise coefficients as its elements. Therefore, we had 16 response vectors, one for each task condition, in each ROI. Similar to the univariate analysis, we used the responses in the isolated conditions to assess category distance, and the responses in the paired conditions to evaluate the effect of attention.

Figure 2 illustrates the response vectors for two stimulus categories (here termed x and y) in both the isolated and the paired conditions. The four vectors $V_{x^{att}}$, $V_{y^{att}}$, $V_{x^{att}y}$, and $V_{xy^{att}}$ represent the response patterns of the four conditions related to the x-y category pair, with V representing the response vector in an ROI, subscripts x and y denoting the presence of the x and y stimuli, respectively, and the superscript att denoting the attended stimulus. Therefore, $V_{x^{att}}$ represents the response vector related to the isolated x condition (in which x was automatically attended), and $V_{xy^{att}}$ represents the response vector related to the paired xy condition with attention directed to the y stimulus.

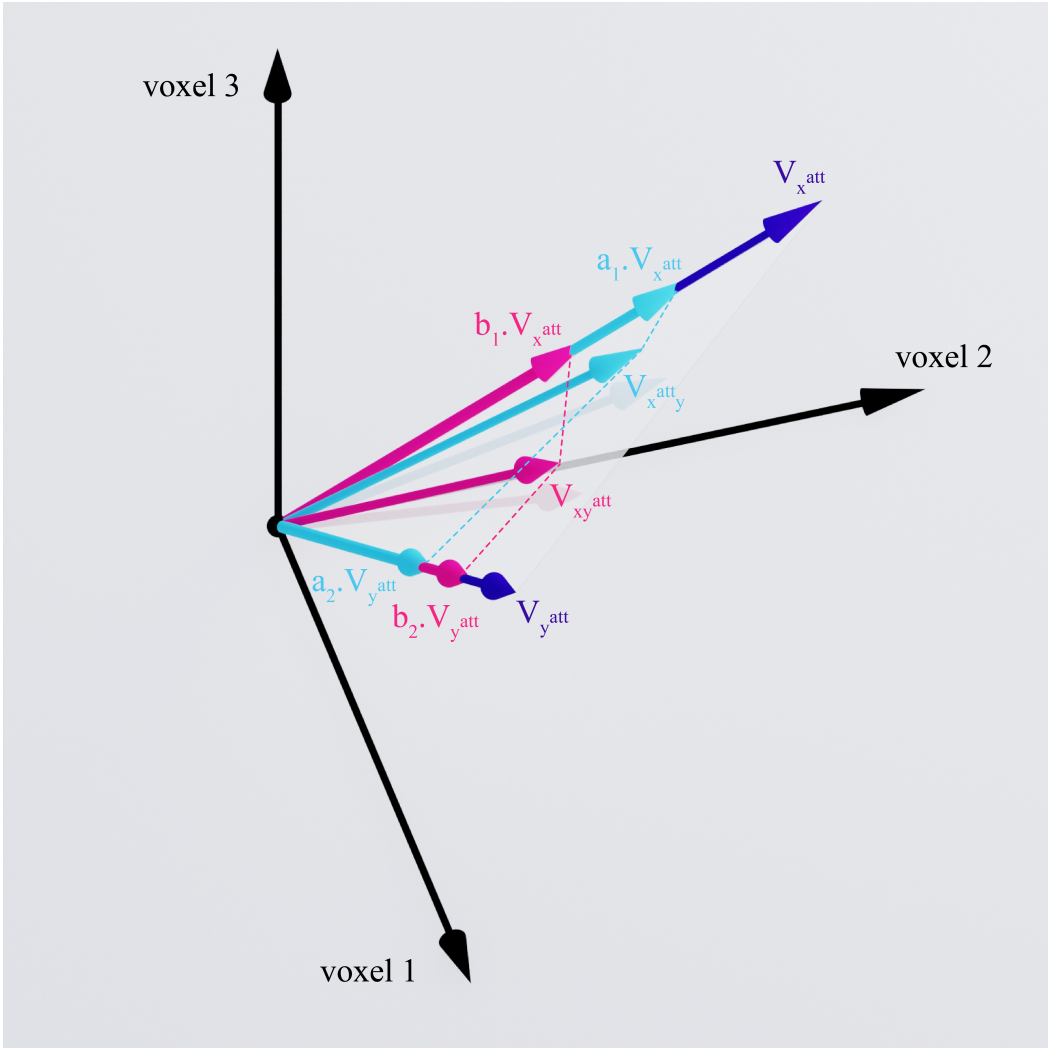

Figure 2: Response vectors related to x and y stimuli in isolated and paired conditions. $V_{x^{att}}$ and $V_{y^{att}}$ denote the response vectors related to isolated x and isolated y conditions. $V_{x^{att}y}$ and $V_{xy^{att}}$ illustrate the response in paired conditions, when attention is directed to stimulus x and y, respectively. Each paired response was projected on the two isolated responses $V_{x^{att}}$ and $V_{y^{att}}$. a_1 and b_1 represent the weight of isolated x response in the pair response, respectively for the $V_{x^{att}y}$ and $V_{xy^{att}}$ responses. a_2 and b_2 represent the weight of the isolated y response in the $V_{x^{att}y}$ and $V_{xy^{att}}$ paired response, respectively.

Multivariate distance based on the isolated conditions

As illustrated in Figure 2, the two isolated response vectors $V_{x^{att}}$ and $V_{y^{att}}$ have a certain distance because the response across the voxels varies for the two stimuli. For two stimuli that elicit more similar response patterns in an ROI, the isolated response vectors are closer to each other. Thus, we defined the multivariate distance between the two isolated response vectors $V_{x^{att}}$ and $V_{y^{att}}$ in each ROI using Pearson's correlation , as shown in Equation 3:

$$Multivariate\ distance = 1 - \rho(V_{x^{att}}, V_{y^{att}}) \tag{3}$$

where $V_{x^{att}}$ and $V_{y^{att}}$ represent the response vectors related to isolated x and y conditions, and ρ denotes Pearson's correlation between the two response vectors. For stimuli with more similar response

patterns, the correlation between their response vectors will be higher, leading to lower multivariate distance.

Multivariate effect of attention based on the paired conditions (attentional weight shift)

Similar to the isolated conditions, we considered the response pattern in the paired conditions as vectors, $V_{x^{att}y}$ and $V_{xy^{att}}$ (Figure 2). The response vectors in the paired conditions can be written as a linear combination of the response vectors in the isolated conditions, with an error term denoting the deviation of the paired-condition responses from the plane defined by the isolated-condition responses[6], as shown in Equation 4:

$$V_{x^{att}y} = a_1 \cdot V_{x^{att}} + a_2 \cdot V_{y^{att}} + \epsilon_1 \tag{4a}$$

$$V_{xy^{att}} = b_1 \cdot V_{x^{att}} + b_2 \cdot V_{y^{att}} + \epsilon_2 \tag{4b}$$

Here, parameters a_1 and a_2 are the weights of the isolated x and y responses, respectively, when x is attended, and parameters b_1 and b_2 are the respective weights of isolated x and y responses when y is attended. The weights are determined by projecting the paired vectors on each of the isolated vectors (Figure 2). ϵ_1 and ϵ_2 denote the error terms related to the deviation of the $V_{x^{att}y}$ and $V_{xy^{att}}$ from the $V_{x^{att}} - V_{y^{att}}$ plane, respectively. A higher a_1 compared to a_2 indicates that the paired response pattern is more similar to $V_{x^{att}}$ compared to $V_{y^{att}}$, and vice versa.

In the presence of two stimuli, if attention could completely remove the effect of the unattended stimulus, the paired response would be the same as the response to the isolated attended stimulus. However, the information related to the unattended stimulus is not fully removed and attention has been shown to only increase the weight of the response related to the attended stimulus in the paired response[6]. To examine whether this increase in the weight of the attended stimulus is constant or if it depends on the similarity of the two stimuli in cortical representation, we defined the weight shift as the multivariate effect of attention:

$$\Delta w = \frac{a_1}{a_1 + a_2} - \frac{b_1}{b_1 + b_2} \tag{5}$$

Here, a_1, a_2, b_1,and b_2 are the weights of the isolated responses, estimated using Equation 4. The weight shift, Δw, is the change in the relative weight of the isolated x response in the paired response when attention shifts from x to y. A higher Δw for a category pair indicates that attention is more efficient in removing the effect of the unattended stimulus in the pair.

Simulations

We investigated the mechanisms underlying the observed effect of stimulus similarity on attentional modulation using simulations. We considered two models for attentional enhancement: a response gain model[15,16] and a tuning sharpening model[17,32].

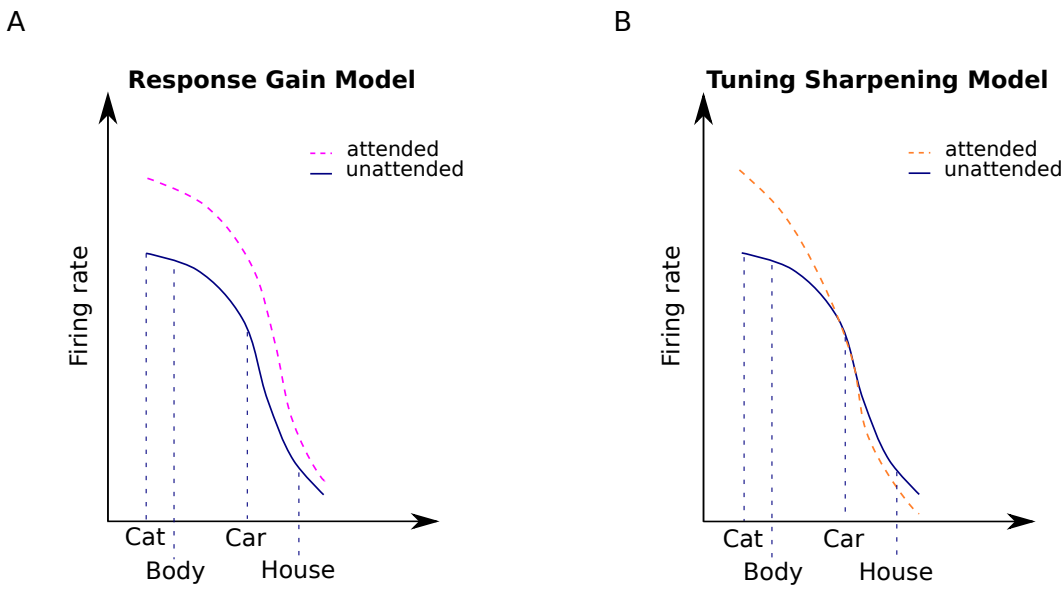

Figure 3: Attentional modulation by the response gain model and tuning sharpening model.
We illustrate the models here for the example of a neuron with high selectivity for cat stimuli. Solid curves denote the response to unattended stimuli and dashed curves denote the response to attended stimuli. (A) According to the response gain model, the response of the neuron to attended stimuli is scaled by a constant attention factor. Therefore, the response of the cat-selective neuron to an attended stimulus is enhanced to the same degree for all stimuli. (B) According to the tuning sharpening model, the response modulation by attention depends on the neuron's tuning for the attended stimulus. Therefore, for optimal and near-optimal stimuli such as cat and body stimuli the response is highly increased, while for non-optimal stimuli such as houses, the response is suppressed.

According to the response gain model, attention to an object multiplicatively increases neural responses to that object (Figure 3A). For instance, for a body-selective neuron, this mechanism can be implemented using Equation 6:

$$R_{Body^{at}} = \beta \cdot R_{Body} \tag{6a}$$

$$R_{Car^{at}} = \beta \cdot R_{Car} \tag{6b}$$

Here, R_{Body} is the neuron's response to an ignored body stimulus, and $R_{Body^{at}}$ is the response of the neuron to the attended body stimulus, which is enhanced by the attention factor, β. R_{Car} and $R_{Car^{at}}$ in Equation 6b denote the response of the same body-selective neuron to an ignored and an attended car stimulus, respectively. The response gain model posits that attention to either stimulus enhances the response of the neuron by the same attention factor. This multiplicative scaling preserves the shapes of the neurons' tuning curves[33,15].

In contrast, according to the tuning sharpening model, attention to an object increases neural responses relative to their responsiveness to that object (Figure 3B). Therefore, while the response of a neuron is substantially enhanced when an optimal stimulus is attended, its response to an attended non-optimal stimulus is increased to a lesser degree, or even decreased. The tuning sharpening model thus predicts a sharpening of the neurons' tuning curve with attention[32].

We implemented this mechanism using Equation 7:

$$R_{Body^{at}} = \beta \cdot s_1 \cdot R_{Body} \tag{7a}$$

$$R_{Car^{at}} = \beta \cdot s_2 \cdot R_{Car} \tag{7b}$$

$$s_1 = \frac{R_{Body}}{R_{max}} \tag{7c}$$

$$s_2 = \frac{R_{Car}}{R_{max}} \tag{7d}$$

In the above equations, R_{Body}, R_{Car}, $R_{Body^{at}}$ and $R_{Car^{at}}$ denote the neuron's response to ignored body, ignored car, attended body, and attended car stimuli, respectively. Parameters s_1 and s_2 likewise denote the degree of the neuron's selectivity to body and car stimuli. Parameter β is the attention factor. R_{max} is the response of the neuron to its optimal stimulus.

We simulated the action of the response gain model and the tuning sharpening model using numerical simulations. We composed a neural population of 10^6 neurons in equal proportions body-, car-, cat- or house-selective. Each neuron also responded to object categories other than its preferred category, but to a lesser degree and with variation. The neural responses and the attention factor were randomly chosen from a range comparable with neural studies of attention and object recognition in the ventral visual cortex[30,34].

Attention was implemented according to the above equations. Using Equations 6 and 7, we calculated the response of each neuron to the same 16 conditions as our main fMRI experiment. Then, we randomly chose 1000 neurons with similar selectivity from the population, and averaged their responses to make up a voxel.

We modeled two neural populations: an object-selective population with mixed preference across voxels, and a second population with similar preference for all voxels. The former population models the object-selective cortex that does not have a clear preference for any one category. The latter population models category-specific regions that respond more strongly to one category of objects than to others. Finally, we performed the same univariate and multivariate analyses as those used for the fMRI data to compare the predictions of each model with the observed data.

Results

Behavioral results

Participants performed a one-back task to maintain attention towards the cued stimuli. Accuracy in each experimental run was checked during the scan to ensure that participants followed the instructions. Participants had an average performance of 90.49% across all runs, confirming effective attention towards the cued stimuli (chance level without attention 50%). As expected, average performance in the isolated conditions (94.82%$\pm$0.046) was significantly higher than in the paired conditions (89% $\pm$ 0.07, with $t(14) = 7.2$ and $p < 0.0001$), since detecting a repetition in the superimposed case was more difficult.

Attentional modulation varies dependent on target-distractor difference in response

We considered the effect of attention in five ROIs: the primary visual cortex V1, the object-selective regions LO and pFs, the body-selective region EBA, and the scene-selective region PPA. We obtained the voxel-wise responses through a general linear model in those ROIs for all task conditions, consisting of four isolated conditions (blocks with isolated stimuli from one category) and 12 paired conditions (blocks with superimposed stimuli from two categories, see Figure 1B). There were 6 combinations of category pairs: Body-Car, Body-House, Body-Cat, Car-House, Car-Cat and House-Cat. For each

voxel we determined its relative preference for the two categories of each category pair, based on its response to the two categories in isolation. Thus, for each pair, one category was labeled as the more preferred category (M), and the other as the less preferred category (L). Considering the isolated and paired conditions related to each category pair, we hereafter refer to the conditions related to each category pair as M^{at}, $M^{at}L$, ML^{at}, and L^{at}, with M and L denoting the more preferred and the less preferred categories for each voxel, and the superscript at denoting the attended stimulus.

For instance, for the Body-Car pair, for a voxel that showed a higher response to body stimuli than to car stimuli, the four associated conditions related to the pair were referred to as M^{at} (attended body stimuli), $M^{at}L$ (attended body stimuli paired with ignored car stimuli), ML^{at} (attended car stimuli paired with ignored body stimuli), and L^{at} (attended car stimuli). If the same voxel was more responsive to cats than bodies, then the four conditions related to the Body-Cat pair would be referred to as: M^{at} (attended cat stimuli), $M^{at}L$ (attended cat stimuli paired with ignored body stimuli), ML^{at} (attended body stimuli paired with ignored cat stimuli), and L^{at} (attended body stimuli).

We next determined the amount of attentional modulation for each category pair using the voxel-wise coefficients related to the two paired conditions, $M^{at}L$ and ML^{at}. We defined attentional modulation for each category pair as the change in response when attention shifted from the M category to the L category in the presence of both stimuli[18,29,30,31] (see Figure 4, illustrated for all pairs in EBA).

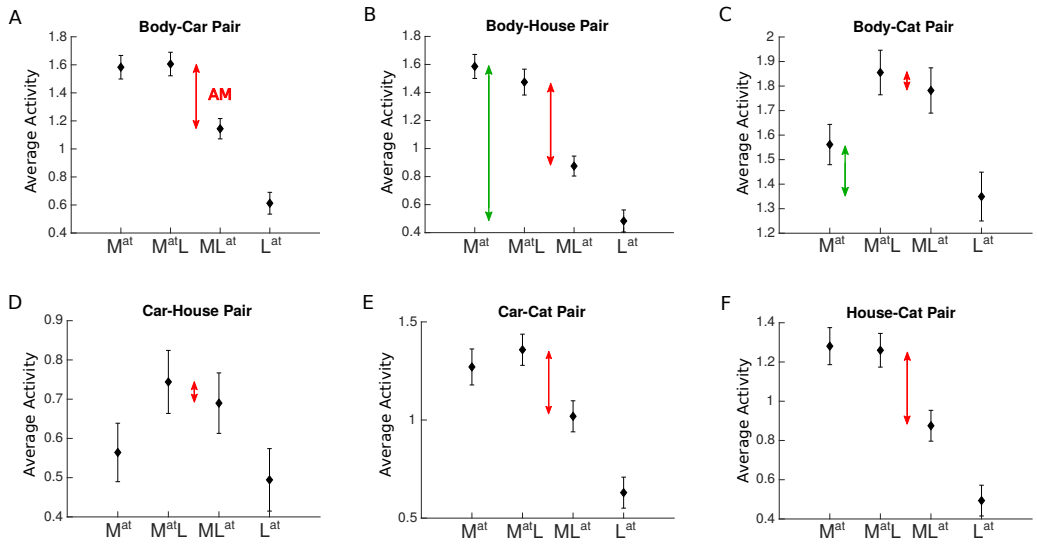

Figure 4: Average voxel response in EBA for each pair of stimulus categories.
The x-axis labels represent the 4 conditions related to each category pair, M^{at}, $M^{at}L$, ML^{at}, L^{at}, with M and L denoting the presence of the more preferred and the less preferred category and the superscript at denoting the attended category. For instance, M^{at} refers to the condition in which the more preferred stimulus was presented in isolation (and automatically attended), and ML^{at} refers to the paired condition in which the less preferred stimulus was attended to. Red arrows in each panel illustrate the observed attentional modulation (AM) caused by the shift of attention from the more preferred to the less preferred stimulus. Green arrows in panels B and C illustrate the difference in the response to isolated stimuli. Error bars represent standard errors of the mean. N = 17 human participants.

We observed a significant reduction in response when attention shifted from the M stimulus to the L stimulus for all pairs in the higher-level ROIs ($ts > 3$, $ps < 0.04$, *corrected*) except for the Body-Car, Body-Car, and Car-Cat pairs in PPA ($ts < 2$, $ps > 0.3$, *corrected*) and the Car-House pair in EBA ($t(16) = 1$, $p = 0.9$, *corrected*). In V1, we observed no significant attentional modulation for any pairs ($ts < 2.5$, $ps > 0.1$, *corrected*) except for the Body-Car pair ($t(16) = 3.8$, $p < 0.01$, *corrected*). Thus, the observed effect was limited to higher-level visual areas. Since presented stimuli were the

same in both conditions, this effect is unequivocally due to attentional modulation caused by the shift in attention.

Closer comparison of the results suggests that for pairs with significant attentional modulation, the modulation is not uniform. Instead, attentional modulation was greater for pairs in which the M and L stimuli elicited more different responses compared to pairs with M and L stimuli eliciting closer responses. For example, we observed a larger attentional modulation for the Body-House pair (Figure 4B) compared to the Body-Cat pair (Figure 4C) in all ROIs ($ts > 4,\ ps < 0.001$, Figures 4B-C, compare the size of the red arrows) except for V1 ($t(16) = 0.65,\ p = 0.5$). Comparing the isolated responses for these two pairs, we observed that the difference between the response of the isolated Body and isolated House conditions was generally higher than the difference between the isolated Body and isolated Cat conditions in all ROIs ($ts > 4,\ ps < 0.001$, Figure 4B-C, compare the size of the green arrows).

To examine this relationship quantitatively for all category pairs, we used two approaches. First, in a univariate analysis using average voxel responses, we determined the relationship between the observed attentional modulation and the difference in isolated responses. Next, in a multivariate pattern analysis, we considered the response patterns in each ROI and looked for the underlying basis of this variation in attentional modulation at the multivariate level. This analysis enabled us to determine whether the bias of attention on the representation of the attended stimulus differed for different category pairs.

Attentional modulation decreases for target-distractor pairs that elicit closer responses

We first used a univariate analysis to determine the relationship between attentional modulation and category distance across pairings and in different ROIs. After determining the voxel-wise M and L categories for each category pair, we calculated the difference in the isolated response elicited by the two categories (univariate category distance) using the two isolated conditions M^{at} and L^{at} (Equation 1).

We then assessed the attentional modulation related to the pair as the amount of the reduction in response when attention shifted from the M stimulus to the L stimulus in the paired presentation of both stimuli (Equation 2). For instance, for the Body-Car pair and a voxel more responsive to bodies than cars, univariate category distance was calculated by $R_{Body^{at}} - R_{Car^{at}}$, and univariate attentional modulation was calculated by $R_{Body^{at}Car} - R_{BodyCar^{at}}$.

We observed a significantly positive correlation between univariate attentional modulation and category distance in all ROIs ($ts > 4,\ ps < 0.001$) except V1 ($t(16) = 0.67,\ p = 0.51$, see Figure 5). These results demonstrate that for stimuli that elicit more different responses, attention causes a greater response modulation, while the shift of attention between stimuli with more similar responses causes little response change.This indicates that the amount of attentional modulation is related to the response difference between the two presented stimuli.

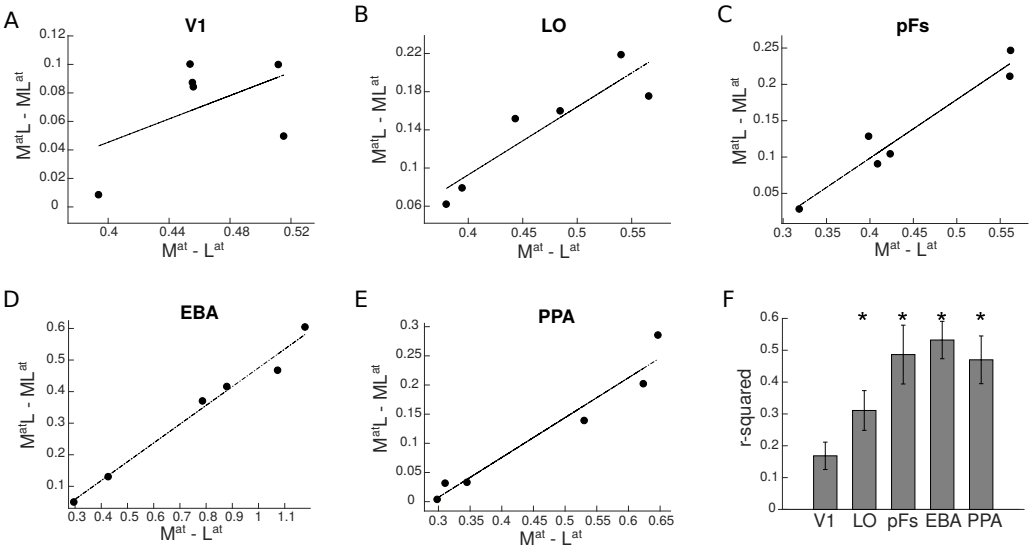

Figure 5: Attentional modulation versus category distance in each ROI.
(A-E) The value of attentional modulation versus category distance. $M^{at}L$ and ML^{at} denote the two paired conditions with attention directed to the more preferred (M) or less preferred (L) stimulus, respectively. M^{at} and L^{at} represent the isolated conditions, respectively, with the more preferred or the less preferred stimulus presented in isolation. Each circle represents the values related to one category pair. Note that the data illustrated here are averaged across subjects only for illustration purposes. R^2 was calculated for individual participants and statistical significance using t-tests across participants as illustrated in panel F. (F) R^2 for the correlation between attentional modulation and category distance in each ROI. Asterisks indicate that the correlation coefficients are significantly positive. Error bars represent standard errors of the mean. N = 17 human participants.

The multivariate effect of attention decreases for more similar target-distractor pairs

The univariate analysis above considers average response only and thus cannot capture other aspects of response variance. For example, in an object-selective region with diverse selectivity for different objects, the average response to body and house stimuli is close, but the response pattern may be very different since voxels highly responsive to bodies do not show high responses to houses, and vice versa. Thus, we had to consider voxel preferences in the univariate analysis to observe the difference in response between the two categories.

We complement the univariate approach with a multivariate pattern analysis to assess the relationship between the effect of attention and category distance at the multivariate level. By considering the whole response pattern in an ROI to each stimulus, we can compare the responses to each stimulus without considering voxel preferences. Moreover, using this method we can determine the weight of the response to each isolated stimulus in the total response, and determine the attentional bias related to each category pair.

The multivariate representation of two simultaneously-presented stimuli can be written as the weighted average of the representations of the two stimuli presented in isolation[6]: When one stimulus is attended, the weight of the response to that stimulus increases in the multivariate representation.

Taking this approach, for each category pair (e.g. Body-Car), we considered the multivariate representation of the two paired conditions ($V_{Body^{at}Car}$ and $V_{BodyCar^{at}}$, with V denoting the multivariate response pattern of each condition), and determined the weight of each of the isolated-stimulus responses ($V_{Body^{at}}$ and $V_{Car^{at}}$) in the paired response (Figure 2). We then calculated the difference between the weight of each stimulus when it was the target and when it was the distractor (e.g. for the Body-Car pair, the difference between the weight of $V_{Body^{at}}$ in $V_{Body^{at}Car}$ and $V_{BodyCar^{at}}$).

If attention could perfectly remove the effect of the distractor, the weight of the attended stimulus

would equal one and the representation of the pair would be identical to the representation of the isolated target. In this case, the difference between the weight of the stimulus representation when attended and ignored would be a maximal value of one. However, if the distractor is not completely removed, this leads to a weight shift value smaller than one. Thus, the magnitude of the weight shift is an indicator of the efficiency of attention, with greater values indicating a higher efficiency of attention in removing the distractor.

To compare the efficiency of attention across category pairs, we calculated the weight shift for each category pair (Equation 5). Then, to determine whether this multivariate effect of attention was dependent on the similarity between the target and the distractor in their cortical representation, we calculated the multivariate category distance for each category pair (Equation 3).

We observed that the attentional weight shift was not constant for different category pairs, and that weight shift and category distance correlated positively in LO, pFs and EBA ($ts > 4.4,\ ps < 5 \times 10^{-3}$, Figure 6), marginally significantly in PPA ($t(16) = 1.8,\ p = 0.09$), and not in V1 ($t(16) = 0.42,\ p = 0.68$). These results indicate that the attentional bias towards a stimulus in a pair decreases as the similarity between the two stimuli in neural representation increases.

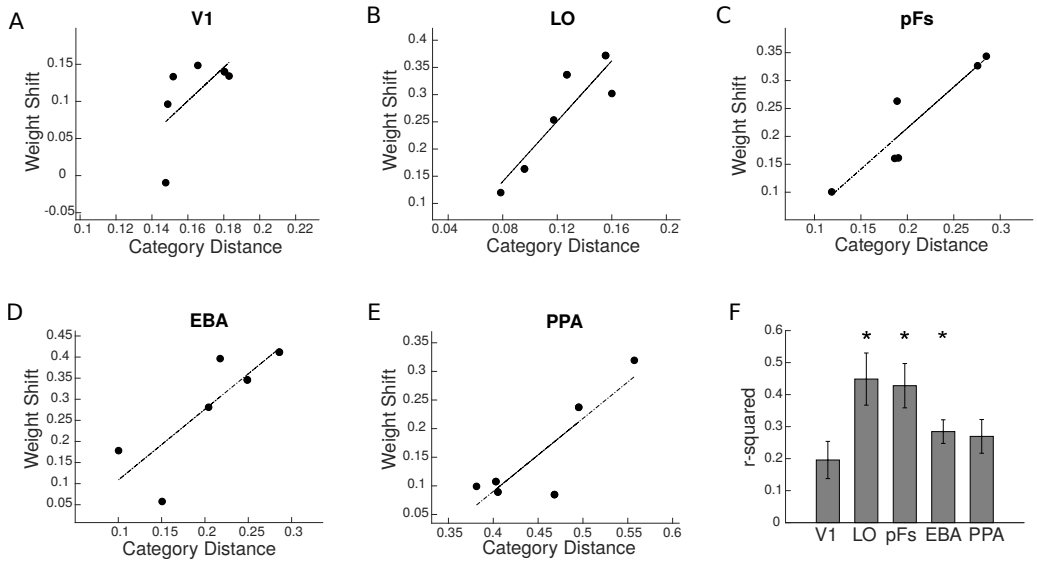

Figure 6: Weight shift versus category distance in each ROI.
(A-E) Attentional weight shift versus category distance. Each circle represents the values related to one category pair. Note that the data illustrated here are averaged across subjects only for illustration purposes. R^2 was calculated for individual participants and statistical significance using t-tests across participants as illustrated in panel F. (F) R^2 for the correlation between attentional weight shift and category distance in each ROI. Asterisks indicate that the correlations are significantly positive. Error bars represent standard errors of the mean. N = 17 human participants.

Tuning sharpening predicts the dependence of attentional modulation on target-distractor similarity

We observed empirically that attentional enhancement is not constant and content-independent, but rather depends on the response similarity between the target and the distractor. We next asked whether gain increase or tuning changes predict the observed effect of target-distractor similarity on attentional modulation.

Based on the response gain model, attention increases neural responses by scaling the responses by a constant attention factor[15,16]. Therefore, the response gain model predicts that attention scales the neurons' tuning function without affecting its shape (Figure 3A).

In contrast, the tuning sharpening model proposes that attention enhances the response of each

neuron based on its preference to the attended stimulus[17,32]. Therefore, this model predicts that attention causes a sharpening of the neurons' tuning function, with a sharp increase in the response to optimal stimuli, and no increase in the response to the non-optimal stimuli (Figure 3B).

To examine which of these mechanisms could account for the observed results, we simulated the responses of a neural population to isolated or paired stimuli from the four categories of bodies, cars, houses and cats. Equivalent to the fMRI experiment, we determined neuronal responses to stimuli presented either in isolation or paired with stimuli from another category (Figure 1B). We implemented attentional modulation of the neural responses either using the response gain model (Equation 6), or the tuning sharpening model (Equation 7). We then used the univariate and multivariate analyses equivalent to those used for the fMRI data to determine which model predicts the empirical data.

We created two neural populations: i) a population with similar selectivity across all neurons to represent a region with strong preference for a specific object category, in which neurons generally show high response to stimuli from that category, and ii) a population with varying selectivity across neurons, representing object-selective regions, in which neurons show different selectivities. Then we assessed attentional modulation using the reduction in response when attention shifted from the stronger to the weaker stimulus in a pair (Equation 2), and examined its relationship with univariate category distance (Equation 1).

We found that the response gain model predicted no relationship between attentional modulation and category distance in either population (Figure 7A-B). In contrast, the tuning sharpening model predicted a positive correlation between attentional modulation and category distance in both neural populations (Figure 7C-D). Thus, the tuning sharpening model provides a better prediction of the empirical data compared to the response gain model.

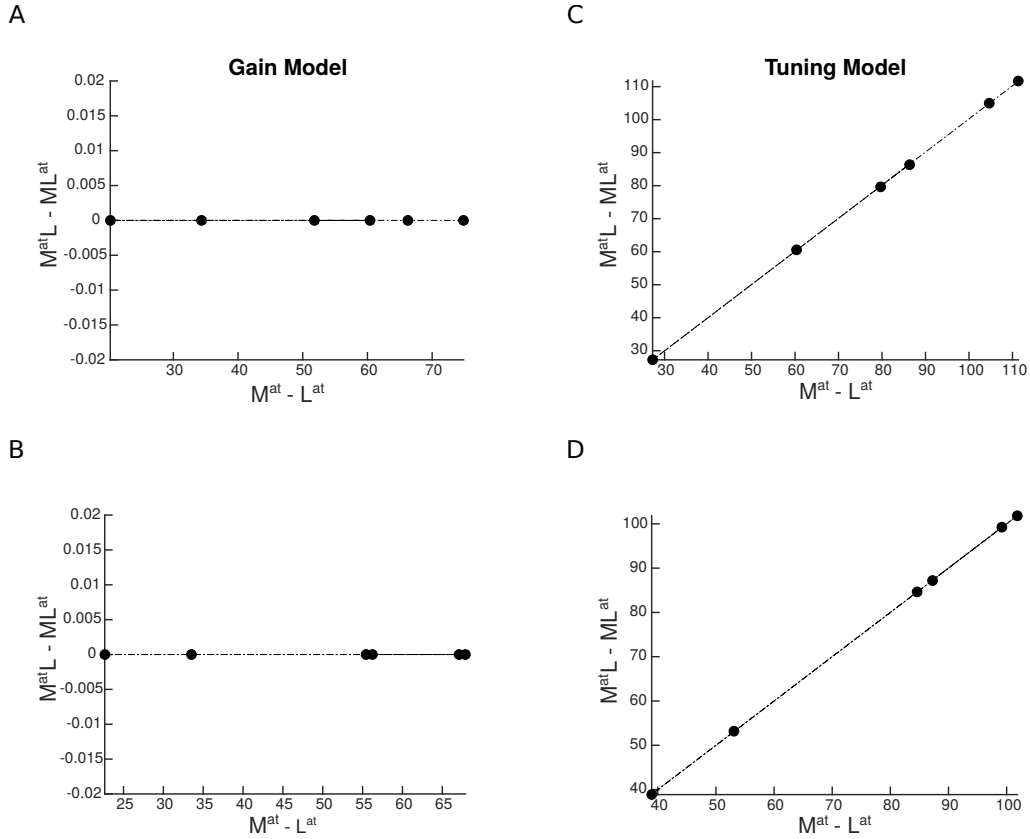

Figure 7: Attentional modulation as a function of category distance, as predicted by the two attentional mechanisms. $M^{at}L$ and ML^{at} denote the two paired conditions with attention directed to the more preferred (M) or the less preferred (L) stimulus, respectively. M^{at} and L^{at} represent the isolated conditions, respectively with the M or the L stimulus presented in isolation. Top panels represent predictions in a region with strong preference for a specific category, and bottom panels illustrate predictions in an object-selective region. Each circle represents a pair of categories. (A) Predicted attentional modulation based on the gain model in a region with strong preference for a specific category. (B) Predicted attentional modulation based on the gain model in an object-selective region. (C) Predicted attentional modulation based on the tuning model in a region with strong preference for a specific category. (D) Predicted attentional modulation based on the tuning model in an object-selective region.

Next, for the multivariate analysis, we assessed the attentional weight shift for each category pair as attention shifted from one stimulus to the other (Equation 5), and examined its relationship with the multivariate category distance (Equation 3). Here, too, we find that the response gain model predicted no relationship between attentional weight shift and category distance (Figure 8A-B). In contrast, the tuning sharpening model predicted a positive relationship between weight shift and category distance, providing further evidence for tuning sharpening as the underlying mechanism for attentional enhancement.

In sum, the tuning model predicts the empirically-observed effect of target-distractor similarity on attentional modulation both at the univariate and at the multivariate level, while the response gain model does not.

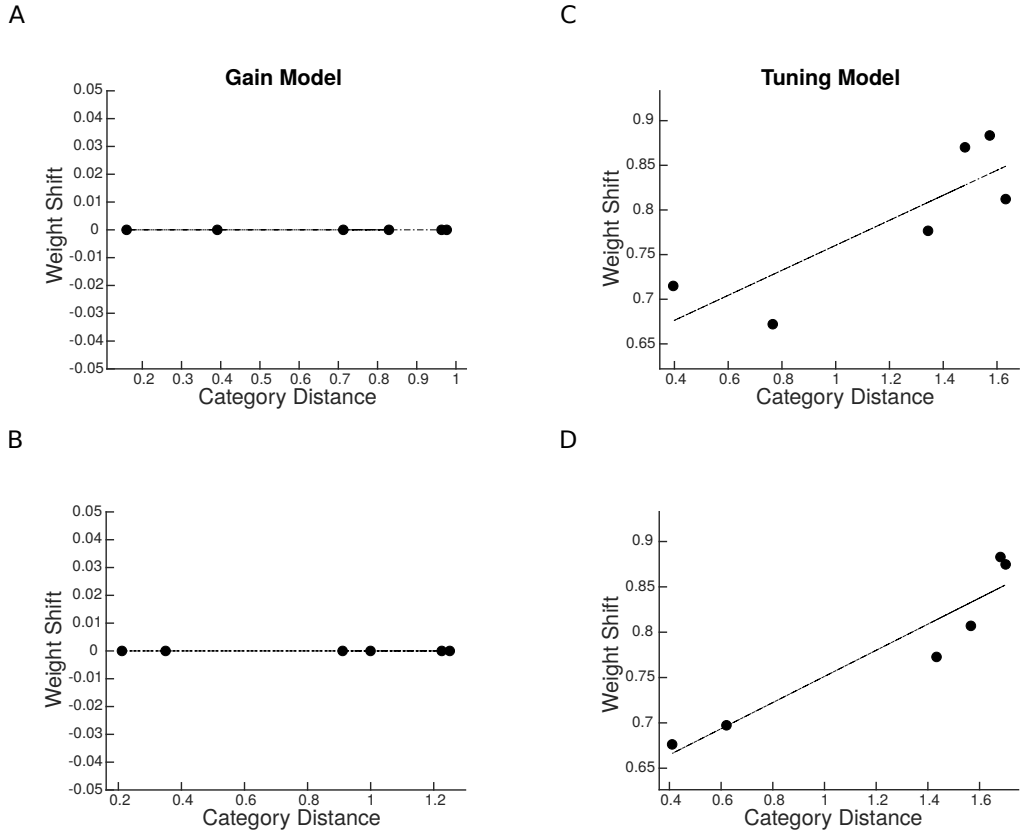

Figure 8: Predicted weight shift as a function of category distance.
Weight shift for each pair is calculated using Equation 5. Category distance represents the difference in multi-voxel representation between responses to the two isolated stimuli, calculated by Equation 3. Top panels are related to predictions in a region with strong preference for a specific category and bottom panels illustrate predictions in an object-selective region. (A) Weight shift predicted by the gain model in a region with strong preference for a specific category. (B) Weight shift predicted by the gain model in an object-selective region. (C) Weight shift predicted by the tuning model in a region with strong preference for a specific category. (D) Weight shift predicted by the tuning model in an object-selective region.

Discussion

Visual stimuli compete for resources in the brain. The biased competition model posits that attention to a stimulus biases this competition in favor of the attended stimulus [1,3,7]. Here, we examined the change in this attentional bias by systematically varying the target and distractors. Using fMRI, we showed that rather than being a constant top-down bias, attentional modulation depends on the similarity between the target and the distractor in their cortical representation, both at the univariate level and at the multivariate level. Using simulations, we arbitrated between the response gain model and the tuning sharpening model as mechanisms of attention for the observed effect, and showed that the empirical results were explained by the latter and not the former.

Effect of target-distractor similarity on attentional modulation

Using stimuli from four object categories, our study reveals the neural basis of the attentional effect graded by target-distractor similarity in the human brain both at the univariate level and at the multivariate level. This finding has two important implications:

First, our results show that in the competition between multiple stimuli, the attentional bias is not constant. Previous studies have shown attentional modulation in the human brain as an average value without considering its variance for different pairings of targets and distractors [6]. These previous accounts of attention cannot explain the variance in performance for the same number of stimuli from different categories. Assessing the role of stimulus content in the bias caused by attention, we confirm that attention enhances the response related to the target. We refine our understanding by showing that however the attentional bias offers less advantage for a more similar target-distractor pair.

Second, this finding provides direct neural evidence for the adverse effects of target-distractor similarity on performance, as previously reported in behavioral studies [12,14]. While behavioral data have suggested that this effect is due to limitation in processing, no investigation has been made to determine the underlying reason or find a mechanistic explanation. Our results demonstrate that this reduction in performance is because the representation of the target (relative to the distractor) is less effectively enhanced by attention when the target becomes more similar to the distractor.

We observed a significant attentional modulation in higher-level regions of the occipito-temporal cortex, but not in V1. Evidence on the effect of attention on V1 responses are divergent, with some previous neuroimaging studies showing a significant effect of attention on neural responses [35,36], while others reporting no significant effect of attention [29,37]. We believe that this apparent discrepancy results from the form of attention under study. Here, we study object-based attention with a superimposed design that excludes response modulation by space-based attention. Previous reports of significant attentional modulation in V1 include studies of space-based attention with stimuli presented at different locations [35,36]. Considering the high reliance of V1 responses to location, the effect of attention is less pronounced when the two stimuli are presented at the same location, as is the case in the present study.

A model for object-based attentional enhancement

Using a simulation approach, we provide a mechanistic explanation for the observed graded attentional effect. Our modeling results have two implications:

First, we demonstrate that tuning sharpening, but not response gain, predicts the observed reduction in the effect of attention for more similar target-distractor pairs both at the univariate and at the multivariate level. Previous research has shown that a change in the tuning function improves attentional selection at high external noise levels [32]. Our results indicate that a change in tuning function could also lead to behavioral disadvantage in an environment where the target is not very different from the surrounding items. When attention is directed towards the target, the response to non-target objects that are more similar to the target is also enhanced, albeit to a lesser amount, leading to an overall weaker effect of attention for a more similar target-distractor pair.

Second, providing evidence from the human brain in favor of tuning sharpening, we suggest tuning sharpening as the underlying mechanism in the domain of object-based attention. By comparing the response gain model and the tuning sharpening model directly in a single study, we provide strong evidence that arbitrates between the theories. The effects of attention have generally been explained by attention acting through increasing the contrast or response gain, especially for space-based attention [15,16,38]. However, a simple increase in gain cannot explain all reported effects of attention, and a change in the shape of the tuning curves has been observed during visual search [39], and feature-based attention [17,40,32].

It is important to note that our speculation on the role of tuning sharpening in object-based attention is based on simulations and not neural data. To ascertain tuning sharpening as the underlying mechanism for object-based attention, intracranial recordings from the human brain are needed.

Conclusion

In sum, our results unravel the cortical basis by which target-distractor similarity affects attentional modulation, and indicate tuning sharpening as the underlying mechanism for response enhancement

during object-based attention.

Acknowledgement

We thank Sajad Aghapour for helpful discussions. We thank Kiarash Farahmandrad for help with the graphical illustration of the vector plot. Maryam Vaziri-Pashkam was supported by NIH Intramural Research Program ZIA-MH002035.

References

[1] Robert Desimone and John Duncan. Neural mechanisms of selective visual attention. *Annual review of neuroscience*, 18(1):193–222, 1995.

[2] Sabine Kastner, Peter De Weerd, Robert Desimone, and Leslie G Ungerleider. Mechanisms of directed attention in the human extrastriate cortex as revealed by functional mri. *science*, 282(5386):108–111, 1998.

[3] John H Reynolds, Leonardo Chelazzi, and Robert Desimone. Competitive mechanisms subserve attention in macaque areas v2 and v4. *Journal of Neuroscience*, 19(5):1736–1753, 1999.

[4] Diane M Beck and Sabine Kastner. Stimulus context modulates competition in human extrastriate cortex. *Nature neuroscience*, 8(8):1110–1116, 2005.

[5] Stephanie McMains and Sabine Kastner. Interactions of top-down and bottom-up mechanisms in human visual cortex. *Journal of Neuroscience*, 31(2):587–597, 2011.

[6] Leila Reddy, Nancy G Kanwisher, and Rufin VanRullen. Attention and biased competition in multi-voxel object representations. *Proceedings of the National Academy of Sciences*, 106(50):21447–21452, 2009.

[7] Jeffrey Moran and Robert Desimone. Selective attention gates visual processing in the extrastriate cortex. *Science*, 229(4715):782–784, 1985.

[8] Stefan Treue and John HR Maunsell. Attentional modulation of visual motion processing in cortical areas mt and mst. *Nature*, 382(6591):539–541, 1996.

[9] John H Reynolds and Robert Desimone. Interacting roles of attention and visual salience in v4. *Neuron*, 37(5):853–863, 2003.

[10] Stefan Treue and John HR Maunsell. Effects of attention on the processing of motion in macaque middle temporal and medial superior temporal visual cortical areas. *Journal of Neuroscience*, 19(17):7591–7602, 1999.

[11] Mazyar Fallah, Gene R Stoner, and John H Reynolds. Stimulus-specific competitive selection in macaque extrastriate visual area v4. *Proceedings of the National Academy of Sciences*, 104(10):4165–4169, 2007.

[12] Michael A Cohen, Talia Konkle, Juliana Y Rhee, Ken Nakayama, and George A Alvarez. Processing multiple visual objects is limited by overlap in neural channels. *Proceedings of the National Academy of Sciences*, 111(24):8955–8960, 2014.

[13] Steven L Franconeri, George A Alvarez, and Patrick Cavanagh. Flexible cognitive resources: competitive content maps for attention and memory. *Trends in cognitive sciences*, 17(3):134–141, 2013.

[14] Michael A Cohen, George A Alvarez, Ken Nakayama, and Talia Konkle. Visual search for object categories is predicted by the representational architecture of high-level visual cortex. *Journal of neurophysiology*, 117(1):388–402, 2017.

[15] Carrie J McAdams and John HR Maunsell. Effects of attention on orientation-tuning functions of single neurons in macaque cortical area v4. *Journal of Neuroscience*, 19(1):431–441, 1999.

[16] John H Reynolds and Leonardo Chelazzi. Attentional modulation of visual processing. *Annu. Rev. Neurosci.*, 27:611–647, 2004.

[17] Julio C Martinez-Trujillo and Stefan Treue. Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current biology*, 14(9):744–751, 2004.

[18] Maryam Vaziri-Pashkam and Yaoda Xu. Goal-directed visual processing differentially impacts human ventral and dorsal visual representations. *Journal of Neuroscience*, 37(36):8767–8782, 2017.

[19] Maryam Vaziri-Pashkam and Yaoda Xu. An information-driven 2-pathway characterization of occipitotemporal and posterior parietal visual object representations. *Cerebral Cortex*, 29(5):2034–2050, 2019.

[20] Yaoda Xu and Maryam Vaziri-Pashkam. Task modulation of the 2-pathway characterization of occipitotemporal and posterior parietal visual object representations. *Neuropsychologia*, 132:107140, 2019.

[21] Anders M Dale, Bruce Fischl, and Martin I Sereno. Cortical surface-based analysis: I. segmentation and surface reconstruction. *Neuroimage*, 9(2):179–194, 1999.

[22] Martin I Sereno, AM Dale, JB Reppas, KK Kwong, JW Belliveau, TJ Brady, BR Rosen, and RB Tootell. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science*, 268(5212):889–893, 1995.

[23] Roger BH Tootell, Nouchine K Hadjikhani, Wim Vanduffel, Arthur K Liu, Janine D Mendola, Martin I Sereno, and Anders M Dale. Functional analysis of primary visual cortex (v1) in humans. *Proceedings of the National Academy of Sciences*, 95(3):811–817, 1998.

[24] Rafael Malach, JB Reppas, RR Benson, KK Kwong, H Jiang, WA Kennedy, PJ Ledden, TJ Brady, BR Rosen, and RB Tootell. Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences*, 92(18):8135–8139, 1995.

[25] Kalanit Grill-Spector, Tammar Kushnir, Talma Hendler, Shimon Edelman, Yacov Itzchak, and Rafael Malach. A sequence of object-processing stages revealed by fmri in the human occipital lobe. *Human brain mapping*, 6(4):316–328, 1998.

[26] Zoe Kourtzi and Nancy Kanwisher. Cortical regions involved in perceiving object shape. *Journal of Neuroscience*, 20(9):3310–3318, 2000.

[27] Russell Epstein, Alison Harris, Damian Stanley, and Nancy Kanwisher. The parahippocampal place area: recognition, navigation, or encoding? *Neuron*, 23(1):115–125, 1999.

[28] Paul E Downing, Yuhong Jiang, Miles Shuman, and Nancy Kanwisher. A cortical area selective for visual processing of the human body. *Science*, 293(5539):2470–2473, 2001.

[29] Narges Doostani, Gholam-Ali Hossein-Zadeh, and Maryam Vaziri-Pashkam. The normalization model predicts responses during object-based attention in the human visual cortex. *eLife*, 12:e75726, 2023.

[30] Amy M. Ni, Supratim Ray, and John H.R. Maunsell. Tuned normalization explains the size of attention modulations. *Neuron*, 73(4):803 – 813, 2012.

[31] Kathleen M O'Craven, Paul E Downing, and Nancy Kanwisher. fmri evidence for objects as the units of attentional selection. *Nature*, 401(6753):584–587, 1999.

[32] Sam Ling, Taosheng Liu, and Marisa Carrasco. How spatial and feature-based attention affect the gain and tuning of population responses. *Vision research*, 49(10):1194–1204, 2009.

[33] Stefan Treue and Julio C Martinez Trujillo. Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, 399(6736):575–579, 1999.

[34] Pinglei Bao and Doris Y Tsao. Representation of multiple objects in macaque category-selective areas. *Nature communications*, 9(1):1774, 2018.

[35] David C Somers, Anders M Dale, Adriane E Seiffert, and Roger BH Tootell. Functional mri reveals spatially specific attentional modulation in human primary visual cortex. *Proceedings of the National Academy of Sciences*, 96(4):1663–1668, 1999.

[36] Sunil P Gandhi, David J Heeger, and Geoffrey M Boynton. Spatial attention affects brain activity in human primary visual cortex. *Proceedings of the National Academy of Sciences*, 96(6):3314–3319, 1999.

[37] Maurizio Corbetta, Francis M Miezin, Susan Dobmeyer, Gordon L Shulman, and Steven E Petersen. Attentional modulation of neural processing of shape, color, and velocity in humans. *Science*, 248(4962):1556–1559, 1990.

[38] Kai J Fox, Daniel Birman, and Justin L Gardner. Gain, not concomitant changes in spatial receptive field properties, improves task performance in a neural network attention model. *Elife*, 12:e78392, 2023.

[39] Tolga Çukur, Shinji Nishimoto, Alexander G Huth, and Jack L Gallant. Attention during natural vision warps semantic representation across the human brain. *Nature neuroscience*, 16(6):763–770, 2013.

[40] Stephen V David, Benjamin Y Hayden, James A Mazer, and Jack L Gallant. Attention to stimulus features shifts spectral tuning of v4 neurons during natural vision. *Neuron*, 59(3):509–521, 2008.