The Normalization Model Captures the Effects of Object-based Attention in the Human Visual Cortex

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

Gholam-Ali Hossein-Zadeh
School of Electrical Engineering, University of Tehran

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

Corresponding Author: Maryam Vaziri-Pashkam
Address: Laboratory of Brain and Cognition, National Institute of Mental Health, 10 Center Dr., Bethesda MD, 20814
Email: maryam.vaziri-pashkam@nih.gov
Keywords: normalization model, attention, object-based attention, human visual cortex, object category

Authors declare no conflict of interest.

Abstract

Here, we report that normalization model can capture the effects of object-based attention across the visual hierarchy in the human brain. We used superimposed pairs of objects and asked participants to attend to different targets. Modeling voxel responses, we demonstrated that the normalization model outperforms other models in predicting voxel responses in the presence of attention. Our results propose normalization as a canonical computation operating in the primate brain.

With the limited resources of the brain for processing the incoming visual input, attention plays an important role in selecting relevant information for further processing. When directed towards a certain stimulus, attention modulates neural responses in a way that the response related to the attended stimulus is enhanced relative to the other parts of the visual input, showing features of response or contrast gain depending on the task settings[1–4] . Previous studies have demonstrated that the amount of this modulation in response differs across neurons[5] , and can vary based on the number of presented stimuli[6] . The normalization model of attention successfully explains these various attentional effects within a single computational framework[7–9] . Several monkey electrophysiology studies have used this model to account for attentional effects in the visual cortex[5,6,10,11] . However, it is not obvious if these results could be generalized to the human brain. Here, we used functional MRI to record the BOLD response during object-based attention in humans and investigated if the normalization model can capture attentional effects in the human visual cortex.

In a blocked-design fMRI paradigm, human participants (n=19) viewed semi-transparent grey-scale stimuli from the two categories of houses and human bodies (Figure 1A). Each experimental run consisted of one-stimulus (isolated) and two-stimulus (paired) blocks, with attention directed either to one of the two objects or to the color of the fixation point. The experiment, therefore, had a total number of 7 conditions (four isolated and three paired conditions, see Figure 1C). In paired blocks, we superimposed the two stimuli to minimize the effect of spatial attention and force participants to use object-based attention (Figure 1B,C). Independent localizer runs were used to localize the primary visual cortex (V1), the object-selective regions in the lateral occipital cortex (LO) and posterior fusiform gyrus (pFs), the extrastriate body area (EBA), and the parahippocampal place area (PPA) for each participant (Figure 1D).

To examine the effects of attention, we fit a general linear model and estimated the coefficients for each voxel in each task condition. We then defined the preferred (P) and null (N) stimulus categories for each voxel in the five regions of interest (ROIs), including V1, LO, pFs, EBA, and PPA, according to the voxel's response to isolated body and isolated house conditions. For each voxel, we then rearranged the seven task conditions according to the voxel's preferences. The conditions are hereafter referred to as: P^{at}, $P^{at}N$, PN^{at}, N^{at}, P, PN, N, with P and N denoting the presence of the preferred and null stimuli, respectively, and the superscript *at* denoting the attended category. Mean voxel responses in the five ROIs for all task conditions are illustrated by navy lines in Figure 2A-E. The change in BOLD response induced by the shift in attention across task conditions demonstrates how attention affects voxel responses.

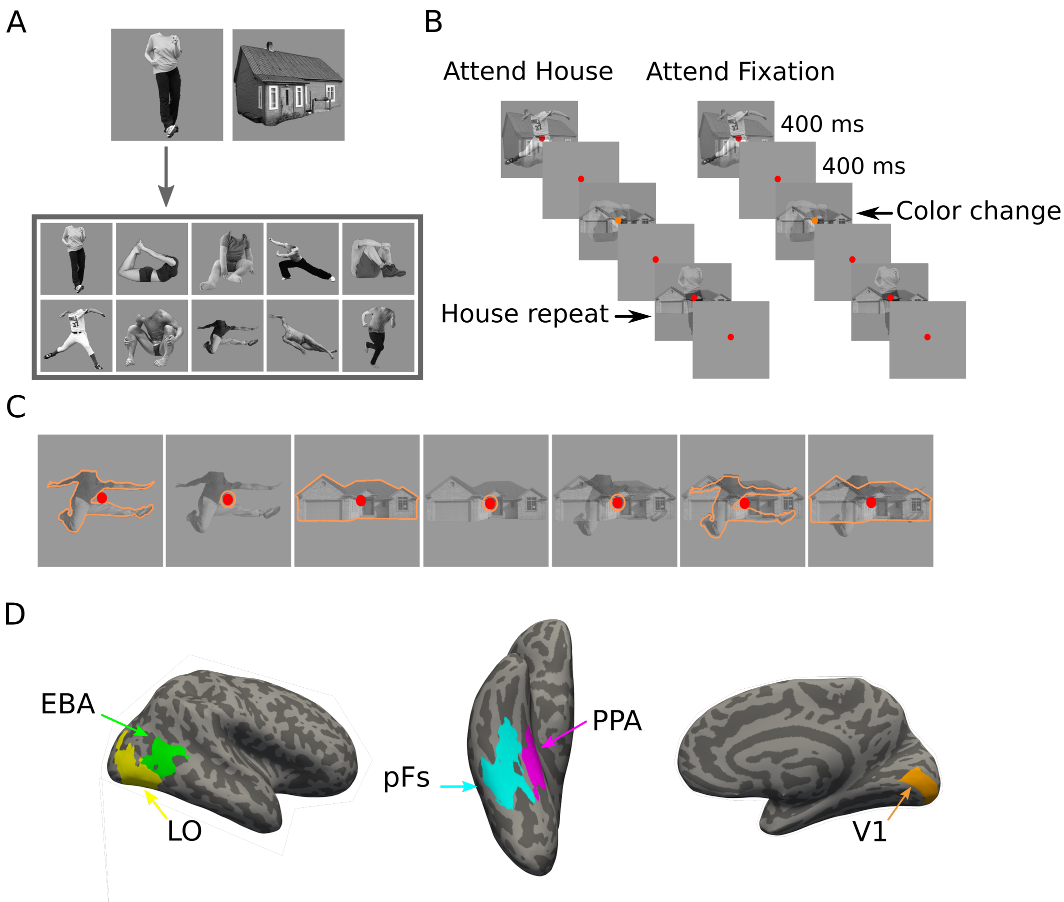

Figure 1: (A) The two stimulus categories (body and house), with the ten exemplars of the body category. (B) Experimental paradigm including the timing of the trials and the inter-stimulus interval. In the example block depicted on the left, both stimulus categories were presented, and the participant was cued to attend to the house category. The two stimuli were superimposed in each trial, and the participant had to respond when the same stimulus from the house category appeared in two successive trials. The color of the fixation point randomly changed in some trials from red to orange, but the participants were asked to ignore the color change. The example block depicted on the right illustrates the condition in which stimuli were ignored and viewers were asked to attend to the fixation point color, and respond when they detected a color change. (C) The 7 task conditions in each experimental run. For illustration purposes, we have shown the attended category in each block by orange outlines. The outlines were not present in the actual experiment. (D) Regions of interest for an example participant, including the primary visual cortex V1, the object-selective regions LO and pFs, the body-selective region EBA, and the scene-selective region PPA.

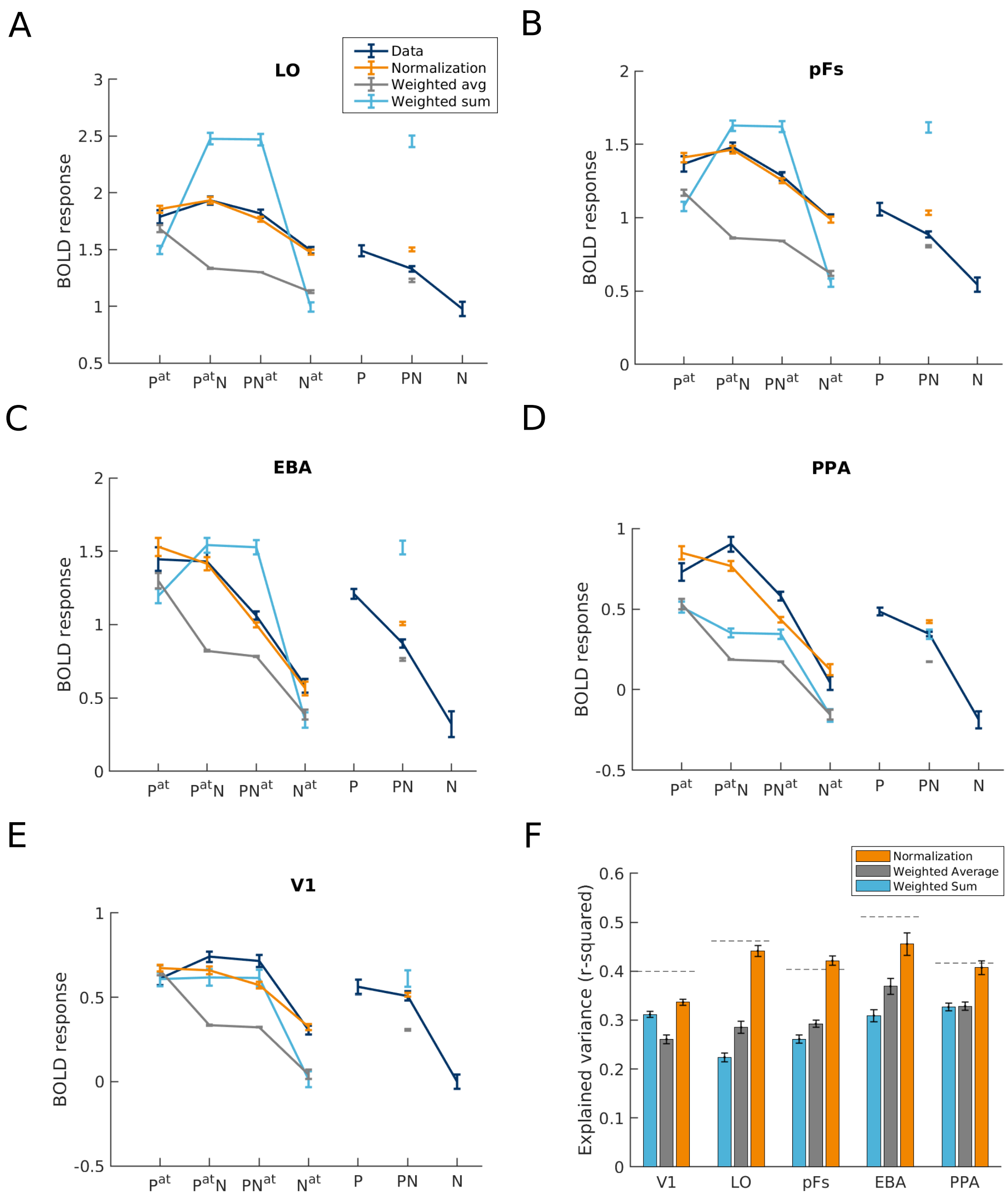

Figure 2: (A-E) Average fMRI responses and model predictions in the five regions of interest. Navy lines represent average responses. Light blue, grey, and orange lines show the predictions of the weighted sum, the weighted average, and the normalization models, respectively. The x-axis labels represent the 7 task conditions, P^{at}, $P^{at}N$, PN^{at}, N^{at}, P, PN, N, with P and N denoting the presence of the preferred and null stimuli and the superscript *at* denoting the attended category. For instance, P refers to the condition in which the unattended preferred stimulus was presented in isolation, and $P^{at}N$ refers to the paired condition with the attended preferred and unattended null stimuli. Error bars represent standard errors of mean. (F) Mean explained variance, averaged over voxels in each region of interest for the 5 conditions calculated by the three models. Light blue, grey, and orange bars show the average variance explained by the weighted sum, weighted average, and normalization models, respectively, with the dashed lines above each set of lines indicating the noise ceiling in each ROI. Error bars represent the standard errors of mean.

Next, we used the three models of weighted sum, weighted average, and normalization to predict voxel responses in different task conditions. The weighted sum model benefits from simplicity and has succeeded in qualitatively explaining the neural responses in previous monkey electrophysiology studies[12] . According to this model, the response to multiple stimuli is the sum of the responses to each individual stimulus presented in isolation, and attention to each stimulus increases the part of the response associated with the attended stimulus:

$$R_{P,N} = R_P + R_N \tag{1a}$$
$$R_{P^{at},N} = \beta R_P + R_N \tag{1b}$$
$$R_{P,N^{at}} = R_P + \beta R_N \tag{1c}$$

Here, $\mathrm{R_{P,N}}$ denotes the response elicited with the preferred and null stimuli present in the receptive field, and $\mathrm{R_P}$ and $\mathrm{R_N}$ denote the response to isolated preferred and null stimuli, respectively. The superscript *at* specifies the attended stimulus, and the stimulus is ignored otherwise. The response related to the attended stimulus is weighted by β, which is the parameter related to attention.

The weighted average model has been proposed as a model that can explain responses to multiple stimuli more accurately than the weighted sum model[13–15] . According to this model, the response to paired stimuli is the average of the response to each stimulus presented alone, weighted by β as the attention-related parameter:

$$R_{P,N} = \frac{R_P + R_N}{2} \tag{2a}$$
$$R_{P^{at},N} = \frac{\beta R_P + R_N}{2} \tag{2b}$$
$$R_{P,N^{at}} = \frac{R_P + \beta R_N}{2} \tag{2c}$$

The normalization model of attention is a more elaborate model that can be described using divisive normalization with a saturation term in the denominator[5,7–9] :

$$R_{P,N} = \frac{c_P L_P + c_N L_N}{c_P + c_N + \sigma} \tag{3}$$

Here, $\mathrm{L_P}$ and $\mathrm{L_N}$ denote the excitatory drive induced by the preferred or the null stimulus, respectively and σ represents the semi-saturation constant. $\mathrm{c_P}$ and $\mathrm{c_N}$ are the respective contrasts of the stimuli. Zero contrast denotes that the respective stimulus is not present in the visual field. When attention is directed towards one of the stimuli, we can rewrite the equation (3) as:

$$R_{P,N} = \frac{c_P L_P + c_N L_N}{c_P + c_N + \sigma} \tag{3a}$$
$$R_{P^{at},N} = \frac{\beta c_P L_P + c_N L_N}{\beta c_P + c_N + \sigma} \tag{3b}$$
$$R_{P,N^{at}} = \frac{c_P L_P + \beta c_N L_N}{c_P + \beta c_N + \sigma} \tag{3c}$$

To compare these models, we split the fMRI data into two halves (odd and even runs) and estimated the model parameters separately for each voxel of each participant using the first half of the data. All comparisons of data with model predictions were made using the left-out second half of the data. Figure 2A-E shows average BOLD responses for the five modeled task conditions in all ROIs. We compared predicted responses from each model with the observed responses in the left-out half of the data to calculate the goodness of fit for each model. The goodness of fit was calculated for each voxel by taking the square of the correlation coefficient between the predicted model response and the respective fMRI response across the five conditions with paired stimuli. Noise ceiling was calculated by taking the r-square of the correlation between voxel responses in the two halves of data. As illustrated in Figure 2F, the normalization model explained the variance in the responses significantly better than the weighted average and the weighted sum models in all regions of interest. Moreover, the explained variance by the normalization model predictions was not significantly different from the noise ceiling in LO, pFs and PPA ($ts < 1.7$, $ps > 0.11$).

Interestingly, just focusing on the paired condition in which none of the stimuli were attended (the PN condition), the normalization model was not necessarily better than the weighted average model. For this condition, the normalization model was better than the weighted average model in V1 and PPA ($ts > 2.7$, $ps < 0.013$) but the two models showed no difference in predicting the data in all other regions ($ts < 1.7$, $ps > 0.11$). These results are in line with previous studies that have shown that the weighted average model can predict neural and voxel responses in the absence of attentional modulations[13, 14, 16] . However, we show that in the presence of attentional modulations, the normalization model has a significant advantage.

Next, comparing the responses in different conditions, we observed two features in the data. First, for the paired conditions, shifting attention from the preferred to the null stimulus caused a reduction in voxel responses. We calculated this reduction in response for each voxel by $(P^{at}N - PN^{at})$ (Figure 3A). This response change was significantly greater than zero in all ROIs ($ts > 5.9$, $ps < 1.5 \times 10^{-5}$) except V1 ($t(18) = 0.18$, $p = 0.43$). Because the same stimuli were presented in both conditions but the attentional target changed from one category to the other, this change in response could only be related to the shift in attention. Second, the effect of the unattended stimulus on the response depended on voxel selectivity for that stimulus, with the unattended preferred stimulus having larger effects on the response than the unattended null stimulus. Attending to the preferred stimulus in the presence of the null stimulus caused the response to approach the response elicited when attending to the isolated preferred stimulus. Therefore, attention effectively removed the effect of the null stimulus. However, attending to the null stimulus in the presence of the preferred stimulus did not eliminate the effect of the preferred stimulus and yielded a response well above the response elicited by attending to the isolated null stimulus. This is the first time such asymmetry has been reported in human fMRI studies, but similar results have been found in monkey electrophysiology studies[5, 6, 17] . To quantify the observed asymmetry we calculated an asymmetry index for each voxel by $(PN^{at} - N^{at}) - (P^{at} - P^{at}N)$ which is illustrated in Figure 3B. This index was significantly greater than zero in all regions ($ts > 6$, $ps < 10^{-5}$).

We compared the observed features of the data (Figure 3A,B) with those predicted by the three models to investigate model results in more detail. We observed that the normalization model predicted both data features significantly better than the two other models, as illustrated in Figure 3C,D. First, the normalization model predicted that shifting attention from the preferred to the null stimulus in the paired conditions reduces voxel responses. The amount of this reduction in response was also the same as what we observed in the data in all regions ($ts < 1.94$, $ps > 0.068$) except for LO ($t(18) = 2.87$, $p = 0.01$). The weighted average model predicted very little response reduction, significantly different from what was observed in the data in all regions ($ts > 4.2$, $ps < 0.0005$) except for V1 ($t(18) = 0.32$, $p = 0.75$). The weighted sum model predicted no significant reduction in response. Figure 3C illustrates the observed and predicted reductions in response in different ROIs. The normalization model was also better at predicting the observed asymmetry in the data. Figure 3D illustrates the asymmetry indices for the data and the three models in all regions. The normalization model predicted the asymmetry index more accurately than the weighted sum model in all regions ($ts > 5.7$, $ps < 10^{-5}$) except for V1 and PPA ($ts < 1.9$, $ps > 0.07$). Notably, the predicted index by the normalization model in LO and pFS was not significantly different from the observed index ($ts < 1.7$, $ps > 0.11$). The weighted average model predicted no significant asymmetry in attentional modulation.

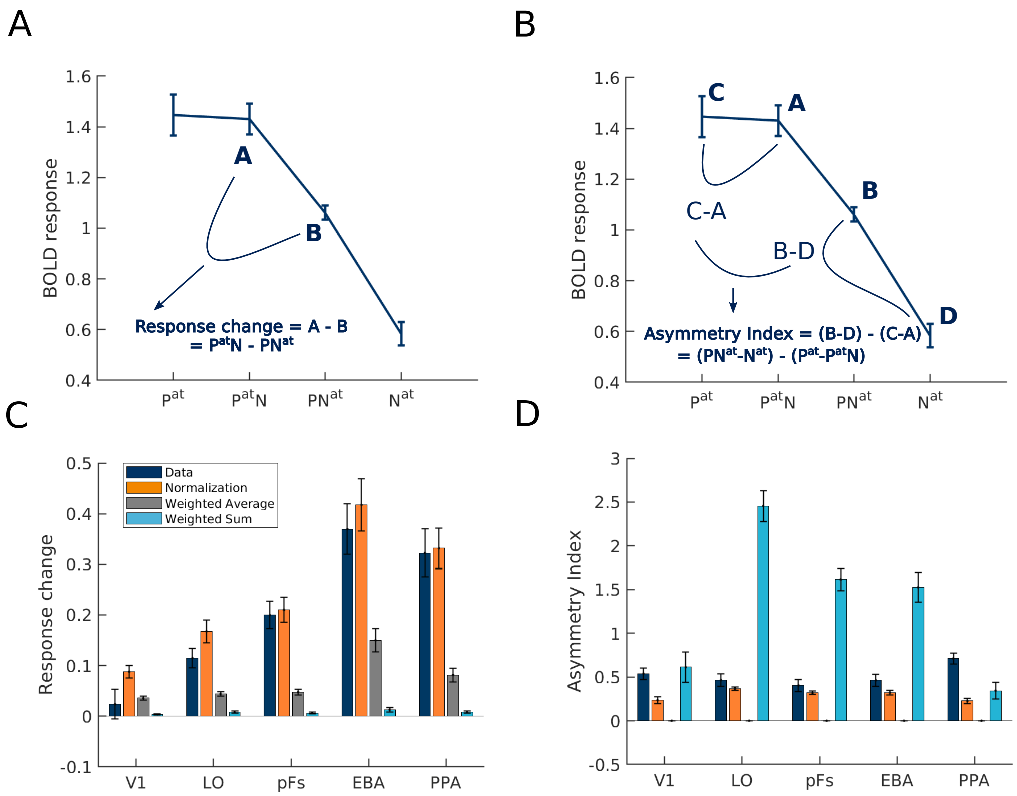

Figure 3: (A) Change in BOLD response when attention shifts from the preferred to the null stimulus in the presence of two stimuli, illustrated here for EBA. (B) The observed asymmetry in attentional modulation for attending to the preferred versus the null stimulus, depicted for EBA. (C) The observed response change and the corresponding amount predicted by different models in different regions, calculated as illustrated in plot A. Error bars represent the standard errors of mean. (D) The observed and predicted asymmetries in attentional modulation in different regions, calculated as illustrated in plot B. Error bars represent the standard errors of mean.

Taken together, our results provide evidence that the normalization model can explain responses at the voxel level beyond the primary visual cortex and across the visual hierarchy, especially in category-selective areas of the human visual cortex, with and without attention, and in conditions with isolated or cluttered stimuli. While the normalization model has been successful in predicting neural responses in electrophysiology studies of the monkey visual cortex in the presence[5,6,11] or absence[18] of attention, none of the studies on the human visual cortex[16,19] have directly tested this model for predicting responses in the human brain. Only one study tested a model related to normalization on responses in the human cortex[20] , but they have not investigated the effect of attention on cortical responses. We also demonstrate for the first time that the normalization model is superior to the weighted average model, which has often been used interchangeably with the normalization model[13,16] , in its ability to account for fMRI responses in the presence of attention. It is noteworthy that here, we are looking at the BOLD responses. We are aware of the limitations of the fMRI technique as the BOLD response is an indirect measure of the activity of neural populations. We hope that future research would directly test the effectiveness of the normalization model in predicting neural responses in the human brain. Our current results provide strong evidence for divisive normalization as a canonical computation operating in the primate brain.

Acknowledgement

Maryam Vaziri-Pashkam was supported by NIH Intramural Research Program ZIA-MH002035.

References

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

[2] 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.

[3] Julio C Martınez-Trujillo and Stefan Treue. Attentional modulation strength in cortical area mt depends on stimulus contrast. *Neuron*, 35(2):365–370, 2002.

[4] John H Reynolds, Tatiana Pasternak, and Robert Desimone. Attention increases sensitivity of v4 neurons. *Neuron*, 26(3):703–714, 2000.

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

[6] Joonyeol Lee and John HR Maunsell. Attentional modulation of mt neurons with single or multiple stimuli in their receptive fields. *Journal of Neuroscience*, 30(8):3058–3066, 2010.

[7] David J Heeger. Normalization of cell responses in cat striate cortex. *Visual neuroscience*, 9(2):181–197, 1992.

[8] Matteo Carandini, David J Heeger, and J Anthony Movshon. Linearity and normalization in simple cells of the macaque primary visual cortex. *Journal of Neuroscience*, 17(21):8621–8644, 1997.

[9] John H Reynolds and David J Heeger. The normalization model of attention. *Neuron*, 61(2):168–185, 2009.

[10] Amy M Ni and John HR Maunsell. Spatially tuned normalization explains attention modulation variance within neurons. *Journal of neurophysiology*, 118(3):1903–1913, 2017.

[11] Amy M Ni and John HR Maunsell. Neuronal effects of spatial and feature attention differ due to normalization. *Journal of Neuroscience*, 39(28):5493–5505, 2019.

[12] Douglas A Ruff and Marlene R Cohen. Stimulus dependence of correlated variability across cortical areas. *Journal of Neuroscience*, 36(28):7546–7556, 2016.

[13] Davide Zoccolan, David D Cox, and James J DiCarlo. Multiple object response normalization in monkey inferotemporal cortex. *Journal of Neuroscience*, 25(36):8150–8164, 2005.

[14] Sean P MacEvoy and Russell A Epstein. Decoding the representation of multiple simultaneous objects in human occipitotemporal cortex. *Current Biology*, 19(11):943–947, 2009.

[15] Annelies Baeck, Johan Wagemans, and Hans P Op de Beeck. The distributed representation of random and meaningful object pairs in human occipitotemporal cortex: the weighted average as a general rule. *Neuroimage*, 70:37–47, 2013.

[16] Libi Kliger and Galit Yovel. The functional organization of high-level visual cortex determines the representation of complex visual stimuli. *Journal of Neuroscience*, 40(39):7545–7558, 2020.

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

[18] Pinglei Bao and Doris Y Tsao. Representation of multiple objects in macaque category-selective areas. *Nature communications*, 9(1):1–16, 2018.

[19] Ilona M Bloem and Sam Ling. Normalization governs attentional modulation within human visual cortex. *Nature communications*, 10(1):1–10, 2019.

[20] Kendrick N Kay, Jonathan Winawer, Aviv Mezer, and Brian A Wandell. Compressive spatial summation in human visual cortex. *Journal of neurophysiology*, 110(2):481–494, 2013.