Tuning attention to object categories:
Spatially global effects of attention to faces in visual processing

Viola S. Störmer[1], Michael A. Cohen[2], George A. Alvarez[3]

[1] Department of Psychology, University of California San Diego
[2] Department of Psychology, Program in Neuroscience, Amherst College
[3] Department of Psychology, Harvard University

Please address correspondence to:

Viola Störmer
Department of Psychology
University of California, San Diego
McGill Hall, Room 5314
San Diego, CA 92093
Email: vstoermer@ucsd.edu

ABSTRACT

Feature-based attention is known to enhance visual processing globally across the visual field, even at task-irrelevant locations. Here, we asked whether attention to object categories, in particular faces, shows similar location-independent tuning. Using electroencephalography (EEG), we measured the face-selective N170 component of the EEG signal to examine neural responses to faces at task-irrelevant locations while participants attended to faces at another task-relevant location. Across two experiments, we found that visual processing of faces was amplified at task-irrelevant locations when participants attended to faces relative to when participants attended to either buildings or scrambled face parts. The fact that we see this enhancement with the N170 suggests that these attentional effects occur at the earliest stage of face processing. Two additional behavioral experiments showed that it is easier to attend to the same object category across the visual field relative to two distinct categories, consistent with object-based attention spreading globally. Together, these results suggest that attention to high-level object categories shows similar spatially global effects on visual processing as attention to simple individual, low-level features.

Introduction

One critical aspect of human visual cognition is the ability to rapidly detect task-relevant objects in cluttered visual environments. For example, when looking for a person in a busy street scene, the ability to selectively focus on faces or bodies while disregarding other objects would enhance pedestrian detection. It has been shown that attention to visual objects modulates neural activity in category-selective regions of higher visual cortex. For example, when attending to a face, neural activity increases in brain regions that are sensitive to faces (e.g., the fusiform face area, FFA), relative to when attending to other objects (Wojciulik et al., 1998; Serences et al., 2004). These effects of object-based attention on visual processing have often been studied by asking participants to attend to one of two superimposed objects (e.g., a face and a house) presented on top of one another so that they compete at the same location (O'Craven et al., 1999; Cohen & Tong, 2015; Baldauf & Desimone, 2014). Thus, any observed attentional effects cannot be attributed to spatial attention and instead must be driven by object-based attention. Of course, attention not only modulates object-selective regions; it also alters neural processing of lower level regions (e.g., MT, V4, etc.) that are sensitive to basic features such as motion and color. Critically, it has been repeatedly shown that neural responses in these lower-level regions are not only enhanced at the attended location, but also in other unattended regions of space (Andersen et al., 2013; Saenz et al., 2002; Serences & Boynton, 2007; Störmer & Alvarez, 2014; Treue & Martinez-Trujillo, 1999; Zhang & Luck, 2009). It should be noted that these previous studies have demonstrated this automatic spread of attention only in cases when observers are attending to single

basic features like color, orientation, or motion direction. No such effects have been observed with more complex objects. This may be due to the fact that object-based attention requires selection processes that encompass multiple features that are organized in a specific configuration while still allowing for some degree of variation of these features, because low-level properties of objects differ substantially even within a category (e.g., for faces: hairstyle, race, or viewpoint). Given this complexity of attending to an object category relative to a single feature, it cannot be assumed that high-level attentional tuning processes would be spatially global in the same way as attention to single features.

Here, we test whether, and at what point in time, attention spreads globally across the visual field for high-level object categories. We focus on the category of faces, which are processed holistically, are highly familiar, and provide an established neural marker in the electroencephalogram (EEG) signal: the N170 (Bentin et al., 1996; McCarthy et al., 1999; Rossion & Jacques 2008; Rossion, 2014). We asked participants to attend to different object categories in rapidly presented image streams and measured the N170 component to stimuli that were presented at a location outside the focus of attention (i.e., the hemifield opposite of the attended stream). In the first experiment, participants attended to either faces amongst buildings or buildings amongst faces. In the second experiment, we examined the extent to which the particular configuration of face parts (i.e., eyes, mouth, and nose) mattered by having participants attend to scrambled face parts amongst buildings. Across both experiments, we found that the face-sensitive N170 elicited by stimuli at task-irrelevant locations was boosted when

participants attended to intact faces, but not when attending to either buildings or scrambled face parts. In two subsequent behavioral experiments we then examined the behavioral consequences of this effects and found that it is more difficult to attend to two categories across the visual field than one category, consistent with the account of global spreading. Overall, these results indicate that attention to faces modulates face processing across the visual field, suggesting that some object- and category-based attention possibly share similar global enhancement mechanisms like attention to basic visual features.

EXPERIMENT 1

Materials and Methods

Participants. The final data set of Experiment 1 includes 12 participants. The data of two participants were excluded from the analysis; one because of excessive artifacts in the EEG (> 30% rejected due to eye movements, blinks, and muscle tension) and one participant did not finish the experiment (the person had to leave earlier than the scheduled time). All participants had normal or corrected-to-normal vision, were between the ages of 18-28 years old, and gave written informed consent prior to the experiment. All experimental procedures were approved by the Committee on the Use of Human Subjects in Research under the Institutional Review Board for the Faculty of Sciences of Harvard University.

Experimental design. Participants attended to a stream of rapidly presented images and detected pictures of a particular category (either faces or buildings) within the stream.

The stimuli were adjusted so that it was equally difficult to detect either a face or a building in the stream. First, the distractor images looked like random noise, but each one was the average of a building image and a face image whose phase had been 100% randomized. Thus, the power spectrum of the noise images equally resembled that of faces and buildings, but the noise images did not look like either a face or a building. Second, to increase task difficulty and match performance across conditions, we randomized the phase of the face and building *target images* using a thresholding procedure (QUEST; Watson & Pelli, 1983). Prior to the EEG session, participants performed 108 trials, and from trial-to-trial we adjusted the level of phase randomization separately for each image set and participant to obtain a performance level of about 80% correct for each category.

On each trial, the display consisted of a central fixation cross (0.3° x 0.3°) and outlined boxes on the left and right side of the screen that served as placeholders (4° x 4°, midpoint at 5° eccentricity). Participants were instructed to keep their gaze in the center of the screen throughout each trial. The stimulus set contained 30 grayscale images of different faces and 30 grayscale images of different buildings with high within-category diversity for each set (stimuli from Cohen et al., 2016). — The faces all varied in viewpoint, hairstyle, race, and age, and the buildings included castles, skyscrapers, lighthouses, and huts. Noise stimuli were created from these images by randomizing the phase of all images (100% randomized), and taking the pixel average of a randomly selected face scramble and a randomly selected building scramble. This resulted in 30 noise images that matched the overall power of the face and building images.

The images were presented in a rapid serial visual presentation (RSVP) either on the left or right side of the screen and lasted for 4 seconds (see Figure 1). Each stream consisted of either zero, one, or two faces, as well as zero, one, or two buildings, and participants were required to count the number of target stimuli (either faces or buildings) within each stream, while ignoring the other images and noise patches. All stimuli were presented in random order with the exception that the first two and last two images were always noise patches, and that each face or building image was followed by at least one noise patch. Side of presentation (left, right) was varied on a trial-by-trial basis and prior to each trial a central arrow cue indicated which side to attend to. At the end of each trial, a question mark appeared in the center of the screen and participants used the number pad on the keyboard to indicate how many targets they had detected.

The target stimuli were always presented for 117ms, but the presentation times of the nontarget stimuli were jittered across each trial between 93 to 300ms (uniform distribution) to avoid eliciting oscillatory responses in the visual cortex (i.e., steady-state-visual evoked responses; Regan, 1989; Störmer et al., 2013) due to a rhythmically flickering image stream. Participants performed 28 blocks with 18 trials each. Before each block, participants were told which category to attend to during the next block, and participants alternated between attended category from one block to the other. Half of the participants started with an “attend-to-faces” block, the remaining half started with an “attend-to-buildings” block. After each block, participants received

feedback on their performance in terms of how many points they made (20 points per correct trial).

While participants performed the task, task-irrelevant probe stimuli were presented on the unattended side at random time intervals. These stimuli were taken from a new set of 10 face images and 10 house images. Note that for the probe stimuli, we used images that showed upfront faces and stereotypical houses (not buildings; see Supplementary Material for all stimuli used). In each trial, 2 faces and 2 houses were presented in random order, each stimulus for 100ms. These probe stimuli were presented at random times with the constraint that they were never presented before 210ms or after 3,200ms post RSVP onset. Furthermore, the minimum interval between each of the probes was set to 500ms to avoid overlap in the event-related potentials (ERPs). Although these probe stimuli were entirely task-irrelevant, they were the main focus of the EEG analysis. To make the probe stimuli less disruptive to the participants, these images were presented at a smaller size than the attended RSVP stream (2° x 2°) and also at a lower contrast level (dimmed about 20%). Overall contrast was matched across the two stimulus types (faces and houses).

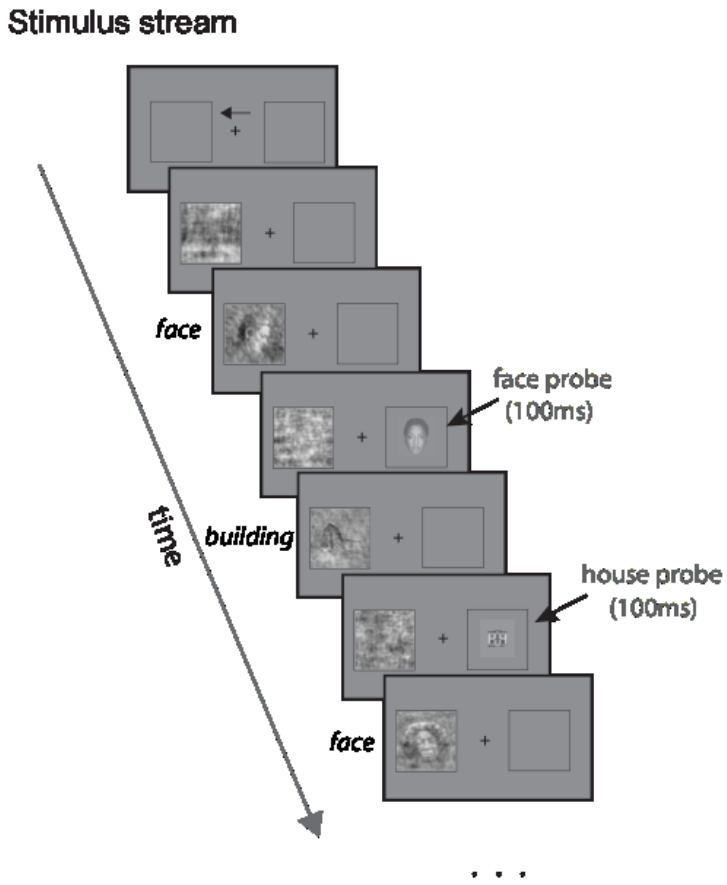


Figure 1. Example of the stimulus stream in Experiment 1. The initial arrow indicates which side to attend to (in this case left) to count images from one particular category (either faces or buildings) on that side only. In this example, two faces and one building are present. At random times, task-irrelevant probe stimuli were presented on the unattended side.

Electrophysiological recordings and analysis. To check whether behavioral performance differed between the two attention conditions, a paired t-test with attention condition as a within-subject factor (faces vs. buildings) was conducted. EEG was recorded continuously from 32 Ag/AgCI electrodes arranged according to the 10-20 system, mounted in an elastic cap and amplified by an ActiCHamp amplifier (BrainVision LLC). All

scalp electrodes were referenced to an electrode on the right mastoid online, and were digitized at a rate of 500Hz. Signal processing was performed with MATLAB (The MathWorks) using the EEGLAB and ERPLAB toolboxes (Delorme & Makeig, 2004; Lopez-Calderon & Luck, 2014). Continuous EEG data was filtered offline with a bandpass of 0.01-112 Hz. Trials with horizontal eye movements, blinks, or excessive muscle movements were excluded from the analysis (cf., Störmer et al., 2014). Artifact-free data was re-referenced to the average of the left and right mastoids. Event-related potentials (ERPs) were time-locked to the onset of the probe stimulus and averaged separately for face and house probes and attended category (faces, buildings), separately for each participant. ERPs were digitally low-pass filtered (-3dB cutoff at 25 Hz) and the mean amplitude of the N170 component was measured between 170 to 200ms at two posterior electrode sites (PO7/PO8, P7/8) over the hemisphere contralateral to the probe stimuli, with respect to a 200-ms pre-stimulus period. The mean amplitudes were subjected to a repeated-measures analysis of variance (ANOVA) with probe type and attention condition as within-subject factors. Planned pairwise comparisons were conducted to examine which conditions were driving any differences in N170 amplitude.

In addition to the ERP analysis, we also ran a time-frequency analysis to check whether participants were continuously and reliably attending to the cued location throughout each trial. In particular, we assessed occipital alpha activity over the hemisphere ipsilateral and contralateral to the cued location as a marker of attentional allocation. Alpha activity is known to be decreased over the hemisphere contralateral to the

attended location relative to ipsilateral (for a review, see Marshall et al, 2015; Kelly, Gomez-Ramirez, & Foxe, 2009; Worden et al., 2000). For this analysis, EEG data was segmented into epochs -400 to 4,0000ms with respect to the onset of each trial and analyzed on a single-trial basis via complex morlet wavelets (Störmer et al., 2016). Single-trial spectral amplitudes were calculated via 6 cycle wavelets at 76 different frequencies separately for each electrode, time point, spatial attention condition (left vs. right), and participant. Then, single-trial spectral amplitudes were averaged separately for left and right-cue trials, and a mean baseline (-350 to -100ms) was subtracted from each time point for each frequency separately. Finally, conditions were collapsed across left and right cue and left and right hemisphere to reveal activity ipsilateral and contralateral to the attended side. Alpha-band amplitude was measured over the range 8-14Hz at parietal-occipital electrode sites PO7/PO8 throughout the entire time interval. Paired-t-tests were performed to test for reliable difference with respect to the cued location.

Results

Behavior: Based on the thresholding procedure prior to the EEG session (see Methods), on average, face images were presented at a higher phase randomization rate (58%) than buildings (43%). This resulted in equal performance for both conditions in the main EEG behavioral task (see Figure 2A; 83.6% correct for faces vs. 83.0% correct for buildings, p=0.83, paired t-test).

ERPS: *Attention modulated N170 responses to faces, but not buildings.* Visual inspection of the ERP waveforms elicited by the probes reveals a clear N170 component which was focused over the hemisphere contralateral to the probe presentation (see Figure 2B & C): a larger amplitude to face relative to house probes in the time interval 170 to 200ms post stimulus onset. An ANOVA with the factor probe type (face, house) and attention condition (faces, buildings) confirmed this difference, revealing a main effect of stimulus type, $F(1,11) = 11.63$, $p = 0.005$, $\eta^2 = 0.13$. There was no main effect of attention (p=0.49), but there was a significant interaction between stimulus type and attention, $F(1,11) = 18.77$, p=0.001, $\eta^2 = 0.06$. Follow-up pairwise comparisons revealed that for the ERPs elicited by face stimuli, the N170 was larger when participants attended to faces relative to buildings ($t(11) = 2.63$, p=0.02, η^2= 0.39). For the ERPs elicited by house probes, the waveform tended to be larger when participants attended to buildings relative to faces, however, this effect did not quite reach significance ($t(11) = 2.16$, p= 0.06, η^2=0.20).

Time-frequency analysis: Alpha power (8-14Hz) was measured over the hemisphere contralateral and ipsilateral to the cued location across all trial types, revealing a clear decrease in alpha activity over the hemisphere contralateral to the attended location relative to ipsilateral ($t(11) = 2.35$; $p = 0.03$; , η^2= 0.33). This control analysis shows that participants maintained spatial attention at the cued location as instructed.

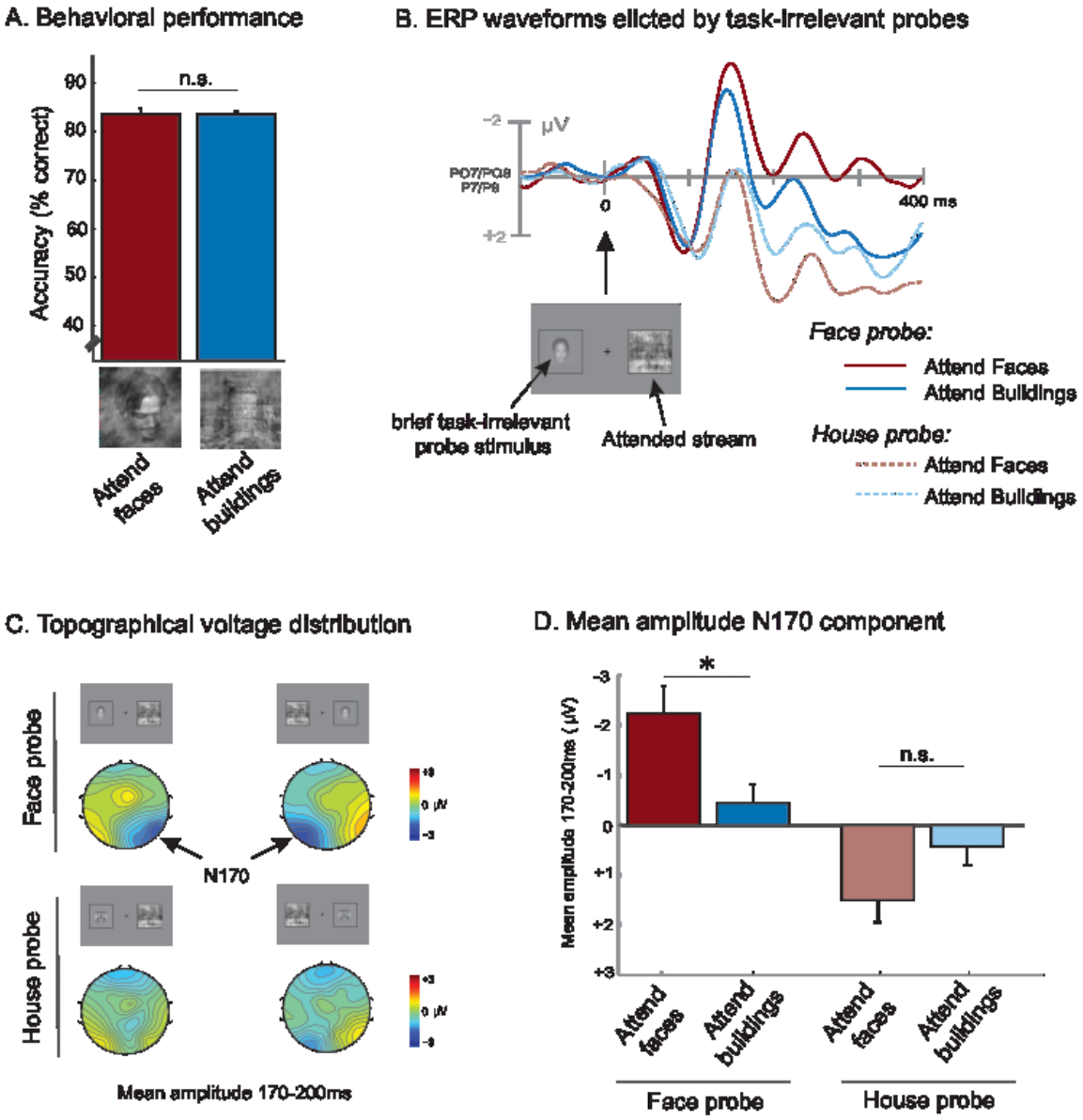


Figure 2: Results of Experiment 1. A) After the thresholding procedure, accuracy was matched across both attention conditions in the main EEG experiment. B) ERP waves elicited by the probe stimuli on the unattended side show a clear face-selective N170 component, such that face probes elicited a larger N170 than houses (dark, solid lines vs. light, dashed lines). Furthermore, starting at about 170ms after probe onset, ERPs elicited by faces show a larger amplitude when participants attended to faces relative to buildings (red solid vs. blue solid line). C) The topographical maps (top view across all attention conditions) show a clear focus of the N170 component over the posterior hemisphere contralateral to the

face probe stimulus (left). There was no hint of a N170 for house probes (right). D) Mean amplitudes of the N170 component (170-200ms) reveal a clear attention effect such that the N170 elicited by faces was larger when participants were attending to faces relative to buildings.

EXPERIMENT 2

Experiment 1 showed that when participants attend to faces on one side of the visual field, neural responses to faces that appear in the opposite side of the visual field are amplified starting 170ms after the face appears. This suggests that tuning attention to complex object categories, such as faces, spreads globally across the visual field and enhances processing of category-specific responses even at unattended locations. However, it remains unclear whether this results from tuning to individual low-level features of the faces, or a more holistic, face-specific attentional template. To address this question, we asked participants to attend to face parts vs. intact faces in a second experiment. If attending to face parts drives face-selective responses across the visual field, we would conclude that the global spread of basic feature-based attention gives rise to higher-level face selectivity. Alternatively, if face-selective responses only spread across the visual field when attending to intact faces, we would conclude that the attentional tuning appears to occur at a higher level of representation (i.e., holistic face representation).

Materials and Methods

Participants. The final sample of Experiment 2 consisted of 24 participants; data of three participants had to be excluded from the data analysis because of excessive

artifacts in the EEG signal (> 30% of trials rejected due to eye movements, blinks, and muscle tension) and one additional participant had to be excluded because of electrode failure (P7 died during the experimental session).

Experimental design. Stimulus presentation and task parameters were the same as in Experiment 1, except that different stimulus sets were used and additional attention conditions were included. One stimulus set contained 30 grayscale images of upfront faces that only included the inner parts of the faces (no neck or hair); the second stimulus set contained 30 grayscale images of scrambled face parts – so images that contained the two eyes, nose, and mouth of a face, but in which each part would appear at random locations, not forming an intact face. These images were cropped ovally so that the outer contour matched the contour of the intact face images. The third stimulus set contained 30 grayscale images of houses (3 houses overlapped with houses used in the building stimulus set of Experiment 1; for all stimuli see Supplementary Figure S1). Two stimulus sets were presented within the same RSVP stream just like in Experiment 1, with either intact faces and houses together, or scrambled faces and houses together. Thus, across the different blocks, participants either attended to intact faces among houses, houses among intact faces (similar to the Experiment 1), scrambled face parts among houses, or houses among scrambled face parts. Just like in Experiment 1, performance was individually matched across conditions prior to the EEG session by using a thresholding procedure to adjust phase randomization of the target images with overall 128 trials. In the main EEG task, participants completed 16 blocks with 32 trials each. Probe stimuli were the same as in Experiment 1.

Electrophysiological recordings and analysis. Behavioral and EEG data were collected and analyzed as in Experiment 1. Behavioral performance was analyzed using a repeated measures ANOVA with attention condition as a within-subject factor. For the statistical analysis of the ERP data, an ANOVA with factors stimulus type (face vs. house) and attention condition (attend faces among houses, attend scrambled face parts among houses, attend houses among faces, attend houses among scrambled face parts) was carried out. If a stimulus type X attention condition interaction was to be found, we planned to conduct follow-up ANOVAs separately for each stimulus type (face and house) with the within-subject factor attention condition. Finally, if this ANOVA showed significant effects for the attention condition, paired t-tests were planned to test which attention conditions reliably differed from one another.

Results

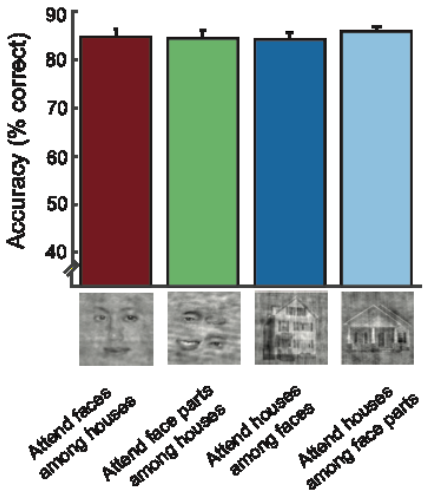
Behavior: Similar to Experiment 1, based on the thresholding procedure prior to the EEG session, on average the image categories were presented at different phase randomization rates with 71% for intact faces, 65% for scrambled face parts, and 58% for houses among intact faces, and 61% for houses among scrambled face parts. This resulted in equal performance across all conditions in the main EEG task (attend intact faces among houses: 84.8%; attend scrambled face parts among houses: 84.4%; attend houses among intact faces: 84.2%; attend houses among scrambled face parts: 85.8%; p=0.91, repeated-measures ANOVA; see Figure 3A).

ERPs: Attention modulated N170 responses to intact faces, but not face parts or buildings. The main ANOVA revealed a main effect of stimulus type, $F(1,23)=78.63$, $p<0.0001$, $\eta^2 = 0.23$, a main effect of attention condition, $F(3, 23) = 3.08$, $p=0.03$, $\eta^2 = 0.11$, and a stimulus type by attention interaction, $F(3,23)$, $p=0.03$, $\eta^2 = 0.06$ (see Supplementary Figure S2 for ERP waveforms to all conditions).

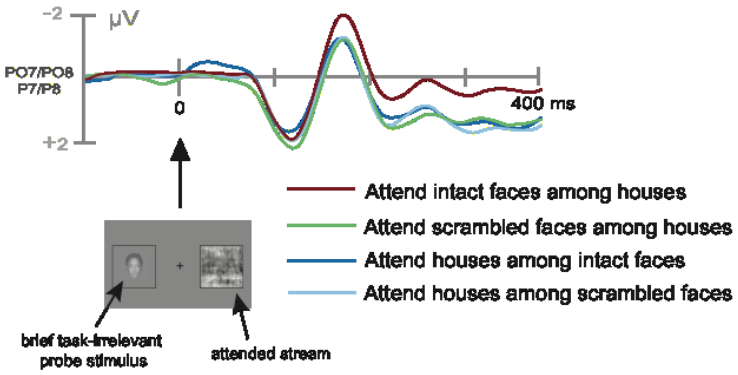
This omnibus ANOVA was followed by planned ANOVAs focusing on each stimulus type separately. For the ERPs elicited by house stimuli, there was no effect of attention, $F(3, 23) = 1.29$, $p=0.29$, as expected. In contrast, for ERPs elicited by face stimuli, there was a reliable effect of attention, $F(3,23) = 5.5$, $p = 0.0019$, $\eta^2 = 0.04$. To examine which conditions drove this effect, planned follow-up paired t-tests were performed. As depicted in Figure 3C, the face-sensitive N170 was largest when participants attended to intact faces among houses, relative to all other conditions (vs. attend houses among faces, $t(23) = 4.04$; $p < 0.00001$, $\eta^2 = 0.42$; vs. attend scrambled face parts among houses, $t(23) = 2.45$; $p = 0.02$, $\eta^2 = 0.20$; vs. attend houses among scrambled face parts, $t(23) = 3.71$; $p = 0.001$, $\eta^2 = 0.38$). None of the other conditions showed any differences (all ps > 0.89).

Time-frequency analysis: As expected and consistent with Experiment 1, occipital alpha activity showed a decrease over the hemisphere contralateral vs. ipsilateral to the cued location ($t(23) = 2.82$, $p =0.01$; $\eta^2= 0.26$), indicating that participants were attending to the cued visual half-field.

A. Behavioral performance



B. ERP waveforms elicted by task-irrelevant probes



C. Mean face-sensitive N170 component

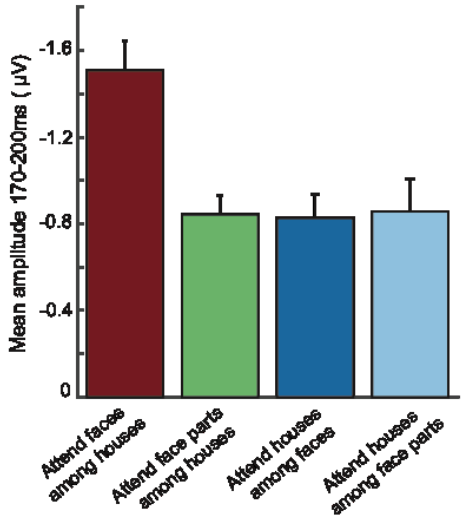


Figure 3*.* Results of Experiment 2. A) Performance during the EEG session was matched across conditions after the thresholding procedure. B) ERP waveforms elicited by the face probes show an enlarged amplitude starting at about 170ms when participants are attending to faces relative to all other conditions. C) Mean amplitudes (170-200ms) of the N170 component show clear modulations, such that when participants attend to intact faces, the N170 is enhanced. Attending to face parts elicits a N170 just as large as when attending to houses.

EXPERIMENTS 3 and 4

The first two experiments show that when attention is tuned to a high-level category, such as faces, category-selective processing is enhanced across the visual field. However, it is unclear whether category-based attention obligatorily spreads, regardless of task demands, or whether it simply tends to do so as long as it is not detrimental to task performance. To test this, we conducted two behavioral experiments in which we asked participants to attend to one category (either faces or buildings) at two locations, or to attend to different categories across two locations (e.g., faces on the left, buildings on the right). Thus, it would be beneficial if object-based attention spread when attending to the same category but detrimental when attending to different categories at two locations. If participants can control the spatial spreading of high-level attention this should result in equal performance across the conditions. Conversely, if attention to object-categories obligatorily spreads even when this spreading is disadvantageous for the task, this would be reflected in lower performance when attending to distinct categories relative to the same category.

Materials and Methods

Participants. Twelve participants completed Experiment 3 and another 12 participants

completed Experiment 4. All experimental procedures were approved by the Committee on the Use of Human Subjects in Research under the Institutional Review Board for the Faculty of Sciences of Harvard University (Exp. 4) or the Institutional Review Board at the University of California, San Diego (Exp. 3).

Experimental design. We designed two tasks that required participants to covertly monitor two image streams left and right of fixation at the same time. Participants were instructed to either attend to the same category of images on both sides (faces or buildings), or to attend to different categories on both sides (faces *and* buildings), and to detect the simultaneous presentation of two images. For Experiment 3, we instructed participant to attend to one category at a specific location (i.e., faces left, buildings right or vice versa), and in Experiment 4, we instructed participants to count any simultaneous presentation of a face and a building as a target, regardless of location (i.e., a face on the left and building on the right, or vice versa, would both count as a target stimulus). This latter design was an attempt to ensure that participants were not making mistakes because of confusing which category at which location to attend to. It simply required them to detect any two faces, any two buildings (same category), or any face and any building (different category) appearing at the same time, regardless of location.

In both experiments the stimuli were the same as in Experiment 1. The number of single face stimuli (on just one side) or single building stimuli would vary from 0, 1, or 2 on every trial, and likewise, the number of simultaneously presented images (i.e., potential targets) was balanced such that either 0, 1, or 2 face targets (face + face), 0, 1

or 2 building targets (building + building), or 0, 1 or 2 mixed targets (face + building) could appear.

Each stimulus was presented for 160ms and 20 images were presented on each trial (each RSVP stream lasted 3.2 seconds). All stimuli were presented in random order with the exception that the first two and last two images were always noise patches, and that each face/building image was followed by at least one noise patch. At the end of each trial, participants had to indicate how many targets they had detected by pressing that number on the number pad of a keyboard. Which category to attend to was varied between blocks and participants were instructed before the start of each block which category to attend to. In Experiment 3, which category to attend to was written on top of each image stream throughout each trial to facilitate matching the to-be-attended category to each location (i.e., in the case of two distinct categories). To match performance across categories, the faces were presented at a higher rate of phase randomization (64% phase-randomized) than buildings (50% phase-randomized) across all conditions. Participants completed 6 blocks with 54 trials each in each experiment. They received feedback after every 27 trials.

Statistical analysis. A repeated measures ANOVA with within-subject factor attention condition (faces, buildings, both) was carried out. If this ANOVA was to show a significant result, we planned on following up with pairwise comparisons (paired t-tests) to see which conditions drove the effect.

Results

Figure 4 shows the results for Experiments 3 and 4. In both cases, there was a clear advantage for attending to the same category relative to attending to two distinct categories when monitoring a rapid stream of images. For Experiment 3, there was a main effect of attention condition ($F(2,11) = 13.64$, $p < 0.0001$, $\eta^2 = 0.22$. Pairwise comparisons revealed that participants performed lowest when attending to two categories at different locations (e.g., faces left, buildings right, or vice versa), relative to attending to faces alone ($t(11) = 6.17$, $p < 0.00001$, $\eta^2 = 0.77$), or buildings alone ($t(11) = 4.46$, $p < 0.00001$, $\eta^2 = 0.64$), and there was no reliable difference between attending to faces vs. buildings($p = 0.48$, paired t-test). Similar results were obtained in Experiment 4: We observed a main effect of attention condition, $F(2,11) = 24.51$, $p < 0.00001$, $\eta^2 = 0.36$, and pairwise comparisons showed that participants performed lowest when attending to both faces and buildings at the same time, relative to when attending to faces alone ($t(11) = 5.48$, $p < 0.000001$, $\eta^2 = 0.73$) or buildings alone ($t(11) = 6.72$, $p < 0.00001$, $\eta^2 = 0.80$); there was no difference between attending to faces vs. attending to buildings (p=0.65, paired t-test).

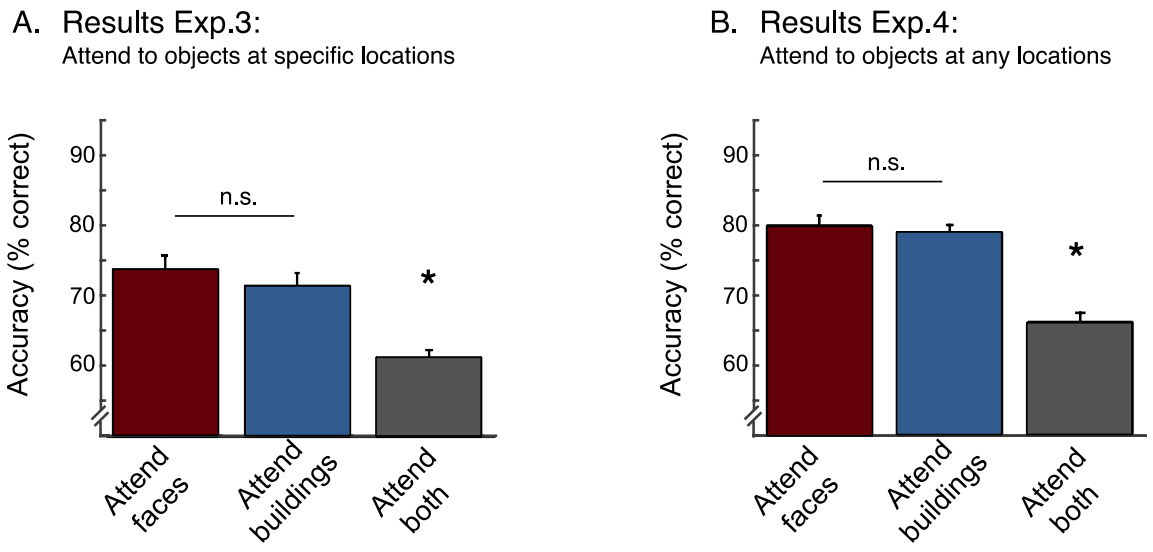


Figure 4. Results of Experiments 3 and 4. Accuracy was lower when attending to two distinct high-level categories at the same time across the visual field relative to when attending to a single category (either faces or buildings).

Discussion

We found that attention to high-level object categories influences the feedforward sweep of category-selective neural activity across the visual field. A rapid image stream was presented on one side of the visual field and observers attended to either faces, buildings, or scrambled face parts. To probe the selectivity of the visual system for the attended vs. ignored category in regions outside the focus of attention, face and house images were flashed on the opposite side of the visual field, and the N170 component – the earliest neural marker of face processing – was examined. The N170 was amplified when participants attended to faces relative to both buildings and scrambled face parts. Together, these results suggest that the selection of high-level object categories

increases the response of neurons tuned to the attended category throughout the visual field.

This spatially global spreading of attention is similar to what has been observed for attention to simple features. For example, attending to the color red among other colors in the left visual field enhances neural signals in both the left (attended) and right (unattended) visual field and vice versa (Saenz et al., 2002; Serences & Boynton, 2007; Andersen, Hillyard, Müller, 2013; Störmer & Alvarez, 2014). Here, we find that attention to a high-level category, specifically faces, enhances face processing at unattended, task-irrelevant locations. Could the effects observed here be simply driven by global spreading of feature-based attention? Schoenfeld and others (2014) have shown that object-based attention involves the sequential activation of feature-specific cortical modules, suggesting that when attending an object, lower-level visual features of that object are also enhanced, and the enhancement of these simple features could spread globally, ultimately resulting in enhanced processing of faces at the unattended location. We here argue that this explanation of our data is unlikely for three reasons. First, the fact that the modulation emerges exactly in the time window of the N170 makes is unlikely that this is a low-level feature spreading effect since those modulations have been shown to occur earlier at around 100ms (Moher et al., 2014; Zhang & Luck, 2009). Second, we observe the effects at the level of the category-specific N170, an EEG marker that has been shown to be selective to face processing across many different studies (Rossion, 2014), including ours. Specifically, the N170 is not observed when observers see basic visual features that generally comprise a face (e.g.,

eyes, a nose, a mouth, etc.) unless those visual features are put together in a specific configuration that creates a specific object: a face. Finally, we directly tested this last point by having participants attend to scrambled faces (Exp. 2). In that case, the N170 enhancement disappears. Thus, at a minimum, participants needed to attend to the particular configuration of features, which is more than what simple feature-based experiments have shown.

The selection of high-level object categories differs in many ways from the selection of simple features. First, high-level object categories comprise multiple parts that need to occur in a specific configuration to render an object. For example, to see a face, the eyes need to be aligned next to each other, with the nose centered below them, and the mouth at the bottom. Second, the appearance of real-world objects varies substantially from one another even within a category. They often appear at different angles, with different low-level details that need to be ignored when looking for the broad object category (e.g., such as different viewpoints; DiCarlo et al., 2012). Thus, unlike feature-based attention, object-based attention must rely on tuning mechanisms that encompass the general features and feature configurations of object categories, while also allowing for variations within a category. Thus, it seems particularly surprising that such complex tuning processes spread across locations and enhance visual processing of object categories across the entire visual field. It needs to be noted, however, that the present study showed such high-level tuning for faces only, a particularly well-learned object category. Thus, it remains an open question whether

such high-level tuning would generalize to other object categories or are unique to the processing of faces.

In Experiment 1, we chose the stimuli such that they would vary substantially in their appearance within category in an attempt to discourage participants from simply attending to low-level visual features while encouraging them to tune their attention to a complex feature configuration at the level of object categories. However, it remains possible that participants attended to some lower-level aspects of the faces (or buildings) to perform the task. If this were the case, the attention system would not be tuned to the specific high-level feature configuration of a face, but possibly only to parts of the object category (e.g., the mouth). It has previously been shown that attending to an object attribute does not only enhance processing of that attribute, but also other features of the same object (O'Craven et al., 1999; Chapman & Störmer, 2018). This would mean that when attending to a mouth, not only would processing of the mouth be enhanced, but possibly the entire face. Accordingly, it could be the case that the enhancement of stimuli on the unattended side was driven by attention to lower-level face parts rather than the entire object. We tested this possibility in Experiment 2 by asking participants to attend to scrambled face parts. The fact that the spatially global enhancement only occurred when participants attended to intact, complete faces, but was absent when participants attended to scrambled face parts suggests that category-selective neural activity can be facilitated when attention is allocated to full-fledged object configurations, but that attending to arbitrarily configured parts of object categories is not sufficient to drive these high-level category-selective modulations.

The present results indicate that the selection of complex object categories is implemented via feedback signals to higher levels of the visual processing hierarchy that receive inputs from the whole visual field. This raises the question of whether such high-level tuning is accomplished via direct input to higher-level representations, or the accumulation of modulation to a particular constellation of basic features. On either account, these findings show that attentional selection on the basis of object categories spreads globally throughout the visual field. Why would attention to high-level object categories operate in a spatially global way? In many situations, such global facilitation at the level of object categories can be beneficial. For example, when searching for an object (with no knowledge about its location), feedback signals that modulate the gain of neurons with receptive fields across the whole visual field would accelerate finding that object through parallel enhancement across locations. However, in other cases, for example when selecting two distinct objects at different locations concurrently, spatially global modulations would cause interference between these object categories, imposing severe limits on the ability to attend to two objects at the same time. The fact that participants' performance was lower when attending to two distinct categories at different locations relative to the same category (Exps. 3 & 4) is consistent with this interference account, and suggest that the global spreading may be – at least to some degree – obligatory.

Of particular interest was at which processing stage these spatially global effects of attention would arise. Previous ERP studies investigating attentional modulations of simple features typically found modulations in between 150 to 300ms, which is

relatively late for simple features such as color or form; thus, these rather late modulations were attributed to delayed feedback signals (Anllo-Vento & Hillyard, 1996; Hillyard & Münte, 1984; Eimer, 1995). In our study, attention modulated the first reliable neural index of face-selective processing – namely the N170 component. While some studies have reported even earlier effects of face processing in the EEG signal (~100ms), it is unclear whether these earlier modulations are truly due to face processing per se, or instead driven by low-level differences between the stimulus sets (Ganis, Smith, & Schendan, 2012; Desjardins, & Segalowitz, 2013). At this point, it appears that the N170 is the first reliable marker of face processing (Rossion, 2014). Thus, we believe it is appropriate to interpret the N170 modulations in our study as an early effect of attention on face processing. Such early modulations may seem surprising in light of previous findings on feature-based attention which often occurred around the same time or even later. However, there is some evidence that in situations of high competition, feature-based attention can modulate sensory processing as early as 100ms post stimulus onset (Zhang & Luck, 2009; Moher et al, 2015). This suggests that the task we used here provided sufficient competition between high-level categories for attention to influence the earliest stages of face processing. Another possibility is that faces are "special" and can more easily be modulated by attention at an early processing stage relative to other object categories. This seems unlikely though, as many studies actually fail to find attentional modulations of the N170 component (Cauquil et al., 2000; Lueschow et al., 2004; but see Crist et al., 2008). Nonetheless, to expand the present findings and test their generalizability, future studies should examine different

types of high-level object categories to see whether similar early attentional modulations can be found.

The results reported here are thematically consistent with a previous study that suggested the presence of spatially global modulations for intermediate visual processing stages. Peelen & Kastner (2009) examined activity patterns in object-selective (LO) cortex while participants attended to bodies or cars in a real-world visual search task. When participants were asked to search for cars in the left and right visual hemifield (but not at the top and bottom on the vertical meridian), neural activation patterns in LO carried information about cars, even when the cars were not presented at the relevant positions, but only at the task-irrelevant positions. While this study showed spatially global spreading of attention in LO, a brain region known to be sensitive to basic shape features, our data build upon these previous results by demonstrating spatially global spreading of attention to even higher levels of visual processing – namely regions that are sensitive to the processing of object categories (i.e., faces). More importantly, the previous study was not able to address the time course of these effects due to the sluggish response of the BOLD signal in functional magnetic resonance imaging (fMRI). Thus, it is unclear at what processing stage the spatially global effects in this previous study arose. Using EEG, we were here able to show that attention influences the feedforward sweep of face processing across the whole visual field.

Overall, the present data show that when searching for complex, high-level object categories, such as faces, attention influences category-selective responses in a

spatially global manner. Critically, this global boost of high-level object representations happens at the earliest stage of category-selective processing – within 170ms after stimulus onset. Such an early influence of attention on visual processing seems particularly beneficial because it can help with the rapid detection of high-level object categories across the visual field.

Acknowledgements

This work was supported by a Marie-Curie fellowship (EU Grant PIOF-GA-2012-329920 to VSS) and a National Science Foundation grant (BCS-1829434 to VSS), a National Science Foundation Graduate Research Fellowship and a National Eye Institute NRSA (F32 EY-024483 to MAC), and an NSF-CAREER award (BCS-0953730 to GAA). We thank Anna Riley-Shepard for help with data collection and stimulus preparation.

References

Allison, T., Puce, A., Spencer, D. D., & McCarthy, G. (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebral cortex*, *9*(5), 415-430.

Andersen, S. K., Hillyard, S. A., & Müller, M. M. (2013). Global facilitation of attended features is obligatory and restricts divided attention. *The Journal of Neuroscience*, *33*(46), 18200-18207.

Anllo-Vento, L., & Hillyard, S. A. (1996). Selective attention to the color and direction of moving stimuli: electrophysiological correlates of hierarchical feature selection. *Perception & psychophysics*, *58*(2), 191-206.

Baldauf, D., & Desimone, R. (2014). Neural mechanisms of object-based attention. *Science*, *344*(6182), 424-427.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of cognitive neuroscience*, *8*(6), 551-565.

Cauquil, A. S., Edmonds, G. E., & Taylor, M. J. (2000). Is the face-sensitive N170 the only ERP not affected by selective attention?. *Neuroreport*, *11*(10), 2167-2171.

Chapman, A. F., & Störmer, V. S. (2018). Feature-based attention is not confined by object boundaries: spatially global enhancement of irrelevant features. PsyArxiv Preprint. doi:10.31234/osf.io/356vk

Crist, R. E., Wu, C. T., Karp, C., & Woldorff, M. G. (2008). Face processing is gated by visual spatial attention. *Frontiers in Human Neuroscience*, *2*, 10.

Cohen, E. H., & Tong, F. (2015). Neural mechanisms of object-based attention. *Cerebral Cortex*, *25*(4), 1080-1092.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21.

Desjardins, J. A., & Segalowitz, S. J. (2013). Deconstructing the early visual electrocortical responses to face and house stimuli. *Journal of Vision*, *13*(5), 22-22.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition?. *Neuron*, *73*(3), 415-434.

Eimer, M. (2011). The face-sensitive N170 component of the event-related brain potential. *The Oxford handbook of face perception*, 329-344.

Eimer, M. (1995). Event-related potential correlates of transient attention shifts to color and location. *Biological psychology*, *41*(2), 167-182.

Ganis, G., Smith, D., & Schendan, H. E. (2012). The N170, not the P1, indexes the earliest time for categorical perception of faces, regardless of interstimulus variance. *Neuroimage*, *62*(3), 1563-1574.

George, N., Jemel, B., Fiori, N., Chaby, L., & Renault, B. (2005). Electrophysiological correlates of facial decision: insights from upright and upside-down Mooney-face perception. *Cognitive Brain Research*, *24*(3), 663-673.

Hillyard, S. A., & Münte, T. F. (1984). Selective attention to color and location: An analysis with event-related brain potentials. *Perception & psychophysics*, *36*(2), 185-198.

Jeffreys, D. A. (1996). Evoked potential studies of face and object processing. *Visual Cognition*, *3*(1), 1-38.

Kelly, S. P., Gomez-Ramirez, M., & Foxe, J. J. (2009). The strength of anticipatory spatial biasing predicts target discrimination at attended locations: a high-density EEG study. *European Journal of Neuroscience*, *30*(11), 2224-2234.

Liu, J., Higuchi, M., Marantz, A., & Kanwisher, N. (2000). The selectivity of the occipitotemporal M170 for faces. *Neuroreport*, *11*(2), 337-341.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in human neuroscience*, *8*, 213.

Lueschow, A., Sander, T., Boehm, S. G., Nolte, G., Trahms, L., & Curio, G. (2004). Looking for faces: Attention modulates early occipitotemporal object processing. *Psychophysiology*, *41*(3), 350-360.

Marshall, T. R., O'Shea, J., Jensen, O., & Bergmann, T. O. (2015). Frontal eye fields control attentional modulation of alpha and gamma oscillations in contralateral occipitoparietal cortex. *Journal of Neuroscience*, *35*(4), 1638-1647.

Maunsell, J. H., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in neurosciences*, *29*(6), 317-322.

McCarthy, G., Puce, A., Belger, A., & Allison, T. (1999). Electrophysiological studies of human face perception. II: Response properties of face-specific potentials generated in occipitotemporal cortex. *Cerebral Cortex*, *9*(5), 431-444.

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport*, *11*(1), 69-72.

Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *Neuroimage*, *39*(4), 1959-1979.

Rossion, B. (2014). Understanding face perception by means of human electrophysiology. *Trends in cognitive sciences*, *18*(6), 310-318.

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature neuroscience*, *5*(7), 631-632.

Schoenfeld, M. A., Hopf, J. M., Merkel, C., Heinze, H. J., & Hillyard, S. A. (2014). Object-based attention involves the sequential activation of feature-specific cortical modules. *Nature Neuroscience*, *17*(4), 619.

Serences, J. T., Schwarzbach, J., Courtney, S. M., Golay, X., & Yantis, S. (2004). Control of object-based attention in human cortex. *Cerebral Cortex*, *14*(12), 1346-1357.

Serences, J. T., & Boynton, G. M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron*, *55*(2), 301-312.

Störmer, V. S., Winther, G. N., Li, S. C., & Andersen, S. K. (2013). Sustained multifocal attentional enhancement of stimulus processing in early visual areas predicts tracking performance. *Journal of Neuroscience*, *33*(12), 5346-5351.

Störmer, V. S., & Alvarez, G. A. (2014). Feature-based attention elicits surround suppression in feature space. *Current Biology*, *24*(17), 1985-1988.

Störmer, V. S., Feng, W., Martinez, A., McDonald, J. J., & Hillyard, S. A. (2016). Salient, irrelevant sounds reflexively induce alpha rhythm desynchronization in parallel with slow potential shifts in visual cortex. *Journal of Cognitive Neuroscience*, *28*(3), 433-445.

Treue, S., & Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575-579.

Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & psychophysics*, *33*(2), 113-120.

Wojciulik, E., Kanwisher, N., & Driver, J. (1998). Covert visual attention modulates face-specific activity in the human fusiform gyrus: fMRI study. *Journal of Neurophysiology*, *79*(3), 1574-1578.

Worden, M. S., Foxe, J. J., Wang, N., & Simpson, G. V. (2000). Anticipatory biasing of visuospatial attention indexed by retinotopically specific-band electroencephalography increases over occipital cortex. *J Neurosci*, *20*(RC63), 1-6.

Zhang, W., & Luck, S. J. (2009). Feature-based attention modulates feedforward visual processing. *Nature neuroscience*, *12*(1), 24-25.