

Published in final edited form as:

*J Cogn Neurosci*. 2011 December ; 23(12): 4094–4105. doi:10.1162/jocn\_a\_00045.

## The Temporal Dynamics of Object Processing in Visual Cortex During the Transition from Distributed to Focused Spatial Attention

Chien-Te Wu<sup>1,2,3,4</sup>, Melissa E. Libertus<sup>1,2,5</sup>, Karen L. Meyerhoff<sup>1</sup>, and Marty G. Woldorff<sup>1,2,3</sup>

<sup>1</sup>Center for Cognitive Neuroscience, Duke University, Durham, NC 27708

<sup>2</sup>Department of Psychology and Neuroscience, Duke University, Durham, NC 27708

<sup>3</sup>Department of Psychiatry, Duke University, Durham, NC 27708

<sup>4</sup>School of Occupational Therapy, National Taiwan University College of Medicine, Taipei, Taiwan

<sup>5</sup>Department of Psychological and Brain Sciences, Johns Hopkins University

### Abstract

Several major cognitive neuroscience models have posited that focal spatial attention is required to integrate different features of an object to form a coherent perception of it within a complex visual scene. Although many behavioral studies have supported this view, some have suggested that complex perceptual discrimination can be performed even with substantially reduced focal spatial attention, calling into question the complexity of object representation that can be achieved without focused spatial attention. In the present study, we took a cognitive neuroscience approach to this problem by recording cognition-related brain activity both to help resolve the questions about the role of focal spatial attention in object-categorization processes and to investigate the underlying neural mechanisms, focusing particularly on the temporal cascade of these attentional and perceptual processes in visual cortex. More specifically, we recorded electrical brain activity in humans engaged in a specially designed cued-visual-search paradigm to probe the object-related visual processing before and during the transition from distributed to focal spatial attention. The onset times of the color-popout cueing information, indicating where within an object array the subject was to shift attention, was parametrically varied relative to the presentation of the array (i.e., either occurring simultaneously or being delayed by 50 or 100 ms). The electrophysiological results demonstrate that some level of object-specific representation can be formed in parallel for multiple items across the visual field under spatially distributed attention, before focal spatial attention is allocated to any of them. The object-discrimination process appears to be subsequently amplified as soon as focal spatial attention is directed to a specific location and object. This set of novel neurophysiological findings thus provides important new insights on fundamental issues that have been long-debated in cognitive neuroscience concerning both object-related processing and the role of attention.

### Introduction

Searching for an Aerosmith CD among other items scattered on your desk is just one example of the challenging tasks our visual system performs regularly: extracting information about the presence and location of a specific object in a crowded visual scene. At lower levels of the visual processing streams, distributed subregions of visual cortex

extract basic visual features from across the visual scene, such as orientation (Hubel & Wiesel, 1968), color (Zeki, 1973), or motion (Maunsell & Van Essen, 1983). In general, however, we perceive the world as composed of objects, not “features”. Given the deluge of sensory inputs we receive continually from the environment, to identify an object at a location, the different features of an object need to be integrated at some level within the visual pathway to form a coherent percept (Roskies, 1999; Treisman, 1999; Wolfe & Cave, 1999).

The influential Feature Integration Theory (FIT) proposed by Treisman and colleagues (Treisman & Gelade, 1980) posited that focal spatial attention facilitates the integration of simple features (e.g., colors, orientation) at the attended location to form the perception of an intact object. Various behavioral studies have provided empirical data supporting the basic validity of this theory (reviewed in Treisman, 1999; Wolfe & Cave, 1999). For example, in visual-search experiments, when a target item is defined by a combination of two or more feature values that it shares with a background of distractors (i.e., a conjunction search, such as searching for a red circle within an array of green circles and red triangles), participants usually exhibit a search pattern function that is distinguished by the averaged search time increasing linearly as the number of distractors increases. Such a pattern suggests that participants need to serially shift their focal spatial attention to process each item in turn to determine whether it has the target combination of features. In contrast, in a “popout” search, when the to-be-searched-for item is defined by a unique feature value in a single feature dimension with respect to the distractors (e.g., one red circle in an array of green circles), the average search time does not increase very much as the number of distractors increases (a flat search-pattern function), suggesting a mostly parallel search process that does not significantly depend on serially shifting focal spatial attention (e.g., Treisman & Gelade, 1980). Another piece of evidence comes from “illusory conjunctions”, a perceptual phenomenon in which features from different objects are incorrectly combined to form an “illusory” perception of a non-existent object. Such conjunctions occur more often when objects are presented outside the focus of spatial attention, suggesting that focal spatial attention is necessary for correct feature binding (e.g., Cohen & Ivry, 1989; Prinzmetal, 1981; Treisman & Schmidt, 1982; Wolfe & Cave, 1999).

Some other behavioral studies, however, have reported that complex perceptual categorization is possible even when focal spatial attention is highly limited. For example, in certain types of conjunction search, the search-pattern function is considerably shallower in slope than might be predicted by a serial search process (Nakayama & Silverman, 1986; Treisman & Sato, 1990; Wolfe, Cave, & Franzel, 1989), implying that the conjoining of features into a multifeatureal object may not depend as strongly on focal attention as had been suggested. Furthermore, when the availability of focal attention is heavily reduced, participants can still perform reasonably well in complicated perceptual tasks, such as face-gender discrimination (Reddy, Wilken, & Koch, 2004) or animal/non-animal categorization of natural scenes (Li, VanRullen, Koch, & Perona, 2002). These behavioral results argue that certain perceptual categorization tasks that presumably require high-level visual processing can be performed even with a highly reduced level of focal attention (Koch & Tsuchiya, 2007). Therefore, it is unclear what complexity of perceptual processes can be achieved with and without focal attention when multiple items are presented simultaneously.

In the present study we took a neurophysiological approach to address these questions with two primary goals. First, we aimed to investigate the perceptual categorization of objects along the visual pathway under *distributed* versus *focal spatial attention*. By using specific key neurophysiological markers of corresponding perceptual and attentional processes, we could study the different characteristics of object representation under distributed spatial attention and under focused spatial attention when multiple items are competing to be

processed by the visual cortices. Second, we aimed to explore the temporal dynamics of the neural activity reflecting object categorization processing *during the transition* from a distributed-attention state to a focused-attention state as focal attention gets shifted to a particular object of interest. The high temporal resolution of the neurophysiological recordings was expected to help us understand the underlying neural mechanisms for these processes, particularly by helping to reveal the temporal cascades of these attentional and perceptual processes within the visual cortices.

More specifically, we recorded electroencephalographic brain activity (EEG) while participants were performing a color-popout-cued visual-search task in which there were variable temporal delays of the color-popout cueing information (Fig. 1). In this task, the visual-search array for each trial always contained three faces and three houses, randomly distributed amongst six spatial locations. Participants were instructed to covertly shift their focal spatial attention to the uniquely colored image location (i.e., the popout) within the array. Their task was to press a button if that image was an infrequently occurring target, defined as a slightly blurred image of either category type (i.e., face or house). Targets were equally likely to be a face or house, thereby eliminating any strategic advantage of the participants attending selectively to faces or houses before cue onset. Critically, the onset time of the attention-shifting color-popout cue relative to the onset of the search array was varied from trial to trial, occurring either 0, 50, or 100 ms after the array onset. The aim was to reveal the processing of object-specific perceptual categorization before and during the transition time from a distributed-attention state to a focused-attention state. Of note in this paradigm is that the discrimination task itself (i.e., whether the popout-cue image was blurry or not) was orthogonal in all cases as to whether that image was a face or a house.

Two event-related potential (ERP) effects, known as the **N2pc** and the **N170**, were extracted from the EEG and used as high-temporal-resolution neurophysiological markers for attentional allocation and object-specific perceptual categorization (i.e., face vs. non-face), respectively. The **N2pc** derives its name from being the second negative potential (“N2”) peaking approximately 200–300ms after the onset of a stimulus array containing a lateral goal-relevant popout item, with an amplitude typically largest over posterior (“p”) parieto-occipital scalp areas contralateral (“c”) to the goal-relevant popout item. These functional characteristics have thus indicated that this component is sensitive to the location of the shifting of attentional focus onto a goal-relevant popout item in a visual search array (Hopf, Boelmans, Schoenfeld, Luck, & Heinze, 2004; Luck & Hillyard, 1994; Woodman & Luck, 1999).

The **N170** component, on the other hand, is a negative ERP component that is largest for face stimuli as compared to other object categories, typically peaking approximately 170 ms after stimulus onset (Bentin, Allison, Puce, Perez, & McCarthy, 1996; Carmel & Bentin, 2002; Rossion et al., 2000). It is therefore thought to reflect face-selective perceptual processing. In addition, we have recently demonstrated that the N170 face-selective effect (i.e., face vs. other object categories) is robustly modulated by spatial attention, including showing that it is absent in unattended locations when spatial attention is strongly focused elsewhere (Crist, Wu, Karp, & Woldorff, 2007). Accordingly, this component could be used as a functional marker for attention-modulated categorization between face and non-face visual stimuli.

By investigating the temporal modulation of these two ERP components as a function of the onset delay of an attention-shifting color-popout cue, we could gain insight into the interaction between attentional allocation and object-specific perceptual categorization processing, from the standpoints of both the underlying cognitive and neural mechanisms. There were two primary predictions. First, we hypothesized that the peak latencies for the

N2pc component, which is sensitive to task-relevant covert shifts of attention, would vary linearly across the three cue-delay conditions. Second, because the image array for each trial always contained equal numbers of faces and houses, when contrasting the array type in which the image at the popout-cued location was a face with the array type in which the image at the popout-cued location was a house, there should be no net face-specific activity from the array as a whole under distributed spatial attention, before the arrival of focused spatial attention as cued by the color popout. In contrast, if focal attention is required to enable the perceptual discrimination of a complex object, attention would need to be directed to the cued location to be able to produce an N170 effect to the image. Indeed, our previous results showing that no N170 effects were observed for images presented at spatially unattended locations, at least when attention was strongly focused elsewhere (Crist et al., 2007). Accordingly, in the present paradigm, it would be expected that this face-selective neurophysiological effect would only occur *after* focal attention had been directed toward the image location. Thus, we expected that the latencies of the object-specific (i.e., face-specific) categorization effects would also increase linearly as a function of cue-onset delays, following in time the shifting of attention as indexed by the N2pc. We expected that our high temporal resolution recordings of brain activity would both help inform questions concerning the role of spatial attention on object processing, as well as help delineate the temporal cascade of neural activity in visual cortex that underlies the key perceptual and attention processes during the transition from distributed attention across a field of objects to focused attention on one of them.

## Methods

### Participants

Sixteen healthy volunteers with normal or corrected-to-normal vision participated in the study (age 19–45 yrs, mean 27.4 yrs; 9 females). Data from 9 additional participants were excluded either due to electrical noise problems (high-frequency noise apparently due to some internal instability of the head boxes and amplifiers during these recording sessions, N=6) or high trial-rejection rates caused by eye blinks (N=3). All participants were paid \$10/hr for their participation. Informed consent was obtained from each participant according to a protocol approved by the Duke University Institutional Review Board.

### Stimuli and Tasks

Figure 1 illustrates the design of the experiment. The default image maintained on the screen during the inter-trial intervals contained six image-item place holders located in a circular configuration with a radius of  $5.2^\circ$  from the center fixation (measured from the center of each image-item box). Each trial started with an array of six black-and-white images (width:  $2.1^\circ$ ; height:  $2.9^\circ$ ) presented in the place holder locations. Half of each array consisted of face images and half of house images, randomly arranged across the six place holders. The face images were modified from the images of the AR face database (Martinez & Benavente, 1998), and the house images were randomly selected from the internet. Either immediately or after specific time delays, each image became colored in either cyan or yellow, with five of the image items on each trial being of the same color and one of a different color, thereby forming a color pop-out in the display. The stimulus onset asynchrony (SOA) between the initial black-and-white image array of the trial and the color transition of the array was varied on a trial-by-trial basis. These SOAs served as attention-shifting cue delays, leading to three different cue-delay conditions (no\_delay, 50ms\_delay, or 100ms\_delay; Fig. 1). The presentation duration for the array for each trial was kept consistent at 250 ms for all the delay conditions. The trial onset asynchronies were varied between 1100 ms and 1400 ms.

Participants were instructed to maintain fixation on the central fixation cross throughout the entire run. For each trial, when the image array was presented, the subjects' task was to shift their spatial attention as quickly as possible to the location of the color pop-out appeared and press a button if the image at that location was slightly blurred. The pop-out location occurred randomly in one of the four corners in the array, and 20% of trials contained the blurred-image targets at that location. The target images were rendered blurred by spatially convolving regular images with a 2-D Gaussian kernel (radius 10 pixels).

Each participant completed 12 runs, six with cyan as the pop-out color (and yellow for the distractors) and six with yellow as the pop-out color (and cyan as the distractors). Each run consisted of 160 trials, approximately one third for each of the three delay conditions.

### Data Acquisition

Stimulus presentation was controlled by a personal computer running the "Presentation" software package (Neurobehavioral Systems, Inc., Albany, CA). EEG was recorded from 64 tin electrodes, mounted in an elastic custom cap (Electro-Cap International, Inc., Eaton, OH) and referenced to the right mastoid during recording. Electrode impedances were kept below 2 k $\Omega$  for the two mastoid electrodes and the ground electrode, 10 k $\Omega$  for the eye electrodes, and 5 k $\Omega$  for the remaining electrodes. Eye blinks and eye movements were monitored by horizontal and vertical EOG electrodes for later rejection of trials containing such artifacts. Vertical eye movements and eye blinks were detected by two electrodes placed below the orbital ridge of each eye, each referenced to the electrodes above the corresponding eye. Horizontal eye movements were monitored by two electrodes located at the outer canthi of the eyes. During recording, eye movements were also monitored using a closed-circuit video monitoring system. Both the EEG and EOG channels were recorded with a band-pass filter of 0.01 to 100 Hz and a gain setting of 1000 using two SynAmps amplifiers (Neuroscan, Inc., Charlotte, NC). Raw signals were continuously digitized with a sampling rate of 500 Hz and digitally stored for later off-line analysis. Recordings took place in a sound-attenuated, dimly lit, electrically shielded chamber.

### Data Analyses

**Behavioral Data**—The hit rates, false-alarm rates, and response times (RTs) were computed across each trial type. These measures were subsequently analyzed using repeated-measure analyses of variance (ANOVA) and paired t-tests.

#### Event-Related Potential (ERP) Data

**EEG signal preprocessing:** Artifact rejection was performed off-line by discarding epochs of the EEG that were contaminated by eye movements, eye blinks, excessive muscle-related potentials, drifts, or amplifier blocking. ERP averages were calculated for the different cue-delay trials types from 800 ms before to 1500 ms after stimulus presentation. The averages were digitally filtered with a non-causal, zero-phase running-average filter of 9 points, which strongly reduces signal frequencies at and above 56 Hz at our sample frequency of 500 Hz. To increase the sensitivity of detecting the pattern of temporal onsets for different delay conditions, two additional non-causal filters were used to filter out frequencies higher than 25 Hz and lower than 0.1 Hz. All the EEG data were re-referenced to the algebraic average of the two mastoid electrodes.

ERP averages were extracted by selectively averaging the EEG epochs as a function of the various conditions (side, time, delay, and image content of popout cue). We focused on two ERP components: the N2pc and the face-versus-house occipital negativity (FH\_N<sub>diff</sub>). The former served as the index of lateralized attentional allocation to task-relevant popout items in an array (Hopf et al., 2004; Luck & Hillyard, 1994; Woodman & Luck, 1999), while the

latter served as an index for neural differentiation between face and house images related to object-related processing. These two components were computed separately for the different cue-delay conditions, using only the non-target trials. Detailed steps to extract both N2pc and FH\_N<sub>diff</sub> effects are described below in the corresponding Results section.

**Analysis of the temporal characteristics of the N2pc and FH\_N<sub>diff</sub> effects:** Several different approaches were used to assess the temporal characteristics of the N2pc and FH\_N<sub>diff</sub> effects for the different cue-onset delay conditions. First, the ERPs were statistically analyzed, using two different methods (see below), to establish the latency windows of significant activity, within which the peak latencies in the grand averages were measured. Second, in order to statistically analyze and compare the effect latencies for the different delay conditions, the peak latencies of the difference-wave effects for the individual subjects were measured and analyzed.

The first method to establish the windows of significant activity for each of the three delay conditions was by analyzing the effects within a series of 6ms-moving windows applied from 0 ms to 600 ms (non-overlapping windows). For the purpose of probing the temporal range of significant activity, a significance alpha-level of 0.05 for was used for each of these windows, combined with an additional criterion of requiring at least three contiguous significant windows (i.e., to correct for type-I error from multiple comparisons). It is of value to note that if the two methods used to establish the latency windows in the first step had employed lenient criteria for statistical significance, the subsequent peak latency analysis would have been hampered by spurious non-significant activity. However, as described below, the two methods are complimentary with each other and together provide legitimate criteria for determining corresponding latency windows.

As a second method to correct for type-I errors from multiple comparisons in establishing the windows of significant activity, we also used a permutation test (for more details, see Appelbaum, Wade, Vildavski, Pettet, & Norcia, 2006; Blair & Karniski, 1993). In this approach, paired t-tests for the contrast of interest for each time point provided information about the time segments in which these continuous t-scores exceeded a preset alpha-level. To obtain a reference noise distribution, the original time courses were then randomly shuffled and the same paired t-test procedures executed repeatedly. We could then determine the “critical length” of continuous t-scores that was longer than 95% of this reference distribution (i.e.,  $p < 0.05$ ) to derive our time windows of interest.

Within the windows of significant activity established for the N2pc and FH\_N<sub>diff</sub> in the three delay conditions, the peak latencies of the effects in the grand averages were determined. However, as another way to assess the peak latencies of these effects, and as a way to be able to statistically compare them between conditions, for each component the peak latency for each participant and each condition was extracted by a computer algorithm from the appropriate difference waves averaged across the three corresponding channels. For the N2pc, the search window for the negative peak was 180–350 ms post-stimulus. For the FH\_N<sub>diff</sub>, there turned out to be two distinguishable phases, which were termed the early FH\_N<sub>diff</sub> and the late FH\_N<sub>diff</sub>. The search window for the FH\_N<sub>diff</sub> was 100–180 ms, and for the late FH\_N<sub>diff</sub> it was 190–400 ms. These measures of peak latencies for each effect were then subjected to statistical tests using within-subject repeated-measures ANOVAs with DELAY (No\_delay, 50ms\_delay, and 100ms\_delay) as the single factor.

## Results

### Behavior

The behavioral results show that participants could accurately detect the presence of the infrequent blurry targets [hit rate: 97.2%; false alarm (FA): 1.4%; mean response time (RT) for correct responses 511.8 ms]. Furthermore, the delay of attentional allocation was reflected in the relative delays of the mean RT (where the RTs were measured relative to the onset of the whole search array) from the no\_delay (468 ms) to the 50ms\_delay (507 ms) to the 100ms\_delay (565 ms) conditions. These RT differences, which showed a regular sequential pattern of ~50 ms across the three delay conditions, were highly significant ( $F(2,30)=438$ ,  $p<0.000001$ ). The behavioral data also indicated that the difficulty level was very similar between discriminating blurred-face versus blurred-house targets [hit rates: blurred-face targets (97.3%), blurred-house targets (97.1%), n.s.], suggesting that neither image category (faces or houses) possessed an advantage for detection over the other category in these arrays<sup>1</sup>.

### Electrophysiology

As mentioned earlier, the N2pc effect served as the marker for the shifting of focal spatial attention and the FH\_N<sub>diff</sub> as a marker for face-selective perceptual processing. Figure 4 illustrates the averaged N2pc and FH\_N<sub>diff</sub> effects for the different delay conditions, with the transparent colored shading indicating the corresponding statistically significant time intervals established by the 6ms-moving-window analyses.

#### N2pc

**Extracting the N2pc activity:** First, ERPs were extracted by averaging separately across trials in which the color pop-out was presented on the left visual field (left-popout) and across trials in which the color pop-out was presented on the right visual field (right-popout), separately for each cue-onset delay (Fig. 2a & 2b). Second, as is typical for the N2pc (Luck & Hillyard, 1994), these data were combined over left-popout and right-popout trials, by collapsing together the contralateral ERP traces for the left-popout trials (i.e., right hemisphere sites) and the contralateral traces for the right-popout trials (i.e., left hemisphere sites), while also collapsing together the corresponding ipsilateral traces for the left-popout and right-popout trials (Fig. 2c). The N2pc activity was then obtained from the difference waves derived by subtracting the averaged ERPs of the ipsilateral channels from the averaged ERPs of the corresponding contralateral channels (Fig. 2d). Amplitudes of the N2pc components were statistically tested over the averaged data prior to taking the difference waves from three lateral occipito-parietal electrodes (Fig. 2c, e.g., Eimer, 1996; Woodman & Luck, 1999) using a within-subject repeated-measure ANOVA with the factor CONTRALATERALITY (Contralateral vs. Ipsilateral).

**N2pc findings:** In line with our prediction, participants shifted their spatial attention according to the three delays of the color-popout attentional shifting cues, as evidenced by the corresponding delays of the N2pc effects. These delays can be clearly seen in Figure 4a, in which the shaded areas show the time periods with significant N2pc activity for each cue-delay as revealed by the applied series of 6ms-moving-window ANOVA analyses. Within

<sup>1</sup>An analysis of the RTs separately for the different target image categories indicated that participants responded slightly faster to the blurred-face than to the blurred-house targets (face RT<sub>mean</sub>=506 ms, house RT<sub>mean</sub>=517 ms,  $t_{15}=3.98$ ,  $p<0.005$ ); however, they were also more likely to falsely identify a non-blurred face as being blurred than a non-blurred house as being blurred (false alarm rate for faces: 1.9%, for houses: 0.8%,  $t_{15}=3.17$ ,  $p<0.01$ ). Therefore, it appears that there may have been a small speed-accuracy trade-off for the blurred-face targets, reflected in the slightly faster reaction time but slightly higher average false-alarm rate. On the other hand, considering that the hit rates and false alarm rates were arguably near ceiling and floor, respectively, and considering the close similarity of behavior patterns across categories, these slight differences do not otherwise seem particularly informative.

the time ranges in which significant activity was demonstrated by these analyses, the peak latencies of the N2pc in the grand-average ERP difference waves were measured: 228ms for the no\_delay condition, consistent with previous N2pc reports (Hopf et al., 2004; Luck & Hillyard, 1994; Woodman & Luck, 1999), 268ms for the 50ms\_delay condition, and 308ms for the 100ms\_delay condition. The same peak latency patterns were also verified by a waveform-permutation test (see Methods).

To further assess the peak latencies of the N2pc for each delay condition and to be able to compare them statistically, we used a computer algorithm to extract the peak latency of the N2pc from each individual participant for each condition, extracted from the appropriate difference waves. These latency values were then subjected to a one-way ANOVA with CONDITION (no\_delay, 50ms\_delay, and 100ms\_delay) as a factor. The results (Table 1; also see Fig. 4a) revealed significant peak latency differences of the N2pc activity across the three delay conditions ( $F_{2,30}=139.4$ ,  $p<0.00001$ ), as well as between all pairs of conditions (no-delay vs. 50ms-delay conditions [ $t_{1,15}=5.3$ ,  $p<0.00001$ ]; 50ms-delay vs. 100ms-delay [ $t_{1,15}=18.7$ ,  $p<0.00001$ ]; and no-delay vs. 100ms-delay [ $t_{1,15}=16.3$ ,  $p<0.00001$ ]).

### Face-house difference negativity (FH\_N<sub>diff</sub>)

**Extracting the Face-House difference negativity (FH\_N<sub>diff</sub>):** Due to the cue-delay manipulation of attentional allocation, expected to result in a corresponding delay of the face-processing related occipital negativity, the term “N170” did not seem to be appropriate nomenclature for the face-selective ERP effect. Therefore, for the remainder of the paper, we shall refer to the face-selective occipital ERP effect as the face-house difference negativity (FH\_N<sub>diff</sub>). By the nature of the current design, this neural effect reflects a categorization in the brain between faces and houses.

To extract this effect, the averaged ERPs were first generated by averaging across trials in which the color pop-out image was a face and was presented in the left visual field (pop-out left face), across trials in which the color pop-out image was a house and was presented on the left visual field (pop-out left house), and analogously across trials in which the color pop-out was a face or house on the right. Second, the face-house difference (FH\_N<sub>diff</sub>) effects were generated by subtracting the house-ERPs from the face-ERPs for the same pop-out side (i.e., both on the left or both on the right, Fig. 3a). As is clear in Figure 3a, the face-house negativity difference was strongly contralateral to the laterally occurring object item of interest. This highly contralateral distribution for the face/house difference to a lateral object was consistent with some other reports in which the top-down (e.g., spatial attention) or the bottom-up (e.g., perception adaptation) effects on faces were stronger contralaterally when faces were presented bilaterally (Carlson & Reinke, 2010; Kovacs, Zimmer, Harza, Antal, & Vidnyanszky, 2005). Accordingly, in order to best capture the relevant face-selective processing effects here, while combining across conditions in which the object of interest was on the left and when it was on the right, the left and right face/house difference effects were combined together in a contralateral/ipsilateral fashion (which thereby also facilitated comparison to the activity extracted from the classic contralateral/ipsilateral N2pc analyses, described above). More specifically, the contralateral FH\_N<sub>diff</sub> effects for the left-popout trials (i.e., right hemisphere sites) and the contralateral FH\_N<sub>diff</sub> effects for the right-popout trials (i.e., left hemisphere sites) were collapsed together, and the corresponding ipsilateral FH\_N<sub>diff</sub> effects for the left-popout and right-popout trials collapsed together (Fig. 3b). The contralateral FH\_N<sub>diff</sub> effects were then obtained by subtracting the averaged ipsilateral ERPs from the corresponding contralateral ones (Fig. 3c & 3d). Amplitudes of this contralateral FH\_N<sub>diff</sub> effect were statistically tested over the averaged data from three occipito-temporal electrodes (Fig. 3d) using within-subject repeated-measures ANOVAs with the factor CATEGORY (Face vs. House).



**Face-specific activity findings:** Interestingly, and in clear contrast, the temporal patterns of the  $FH\_N_{diff}$  effects as a function of the different delay conditions differed from the linear pattern of the N2pc and deviated from our initial predictions in some key ways (Fig. 4b). For all three cue-delay conditions, the  $FH\_N_{diff}$  showed a biphasic deflection, with a negative deflection of smaller amplitude in the early time window from 130–180 ms (approx. the typical N170 latency), followed by a significantly larger ( $F_{1,15}=3.42$ ,  $p<0.005$ ) negative deflection in a later time window from 220–400 ms. The grand-average peak latencies measured over windows of interest established by both the 6ms-moving-window ANOVA and the permutation test are listed in Table 1.

Again, we extracted and analyzed the individual peak latencies for statistical testing across conditions. There was a significant difference of the latency patterns for the  $FH\_N_{diff}$  effects between the early phase and the late phase, as revealed by a significant interaction between Phase and Cue-delay ( $F_{2,26}=10.11$ ,  $p<0.005$ ). Moreover, specific comparisons of the effect of cue-delay for the different phases showed that the peak latency of the *early phase* of the  $FH\_N_{diff}$  effects *did not* vary as a function of cue delay ( $F_{2,28}=0.51$ , n.s.) whereas the peak latencies for the *later phase* of the  $FH\_N_{diff}$  effects showed a robust, significant, *temporal shift* across the delay conditions ( $F_{2,28}=14.63$ ,  $p<0.0005$ )<sup>2</sup>. For the late phase, follow-up pair-wise t-tests further revealed a significant difference between the no-delay and 100ms-delay conditions ( $t_{1,14}=3.6$ ,  $p<0.005$ ) and between the 50ms-delay and 100ms-delay conditions ( $t_{1,14}=4.5$ ,  $p<0.0005$ ), but not between the no\_delay and 50ms\_delay conditions ( $t_{1,14}=1.3$ ,  $p=0.2$ ).

**Individual Latency Comparison between N2pc and  $FH\_N_{diff}$** —A direct comparison of the individual peak latencies in each delay condition showed that the early phase of the  $FH\_N_{diff}$  (reflecting face-house selective processing) was significantly earlier than the peak latency of the N2pc (reflecting the attention shifting in each condition), including for the no-delay condition (all  $p$ 's<.0001; see Fig. 4). Given that participants could not predict where a target may appear, their spatial attention was presumably distributed across the whole image array until the information from the color-popout cue became available to trigger the attentional shifting. This result therefore suggests that at least some degree of the object-specific categorization process -- as reflected by the early phase of the face-specific  $FH\_N_{diff}$  - was occurring without (i.e., before) focal attention.

In contrast to the early phase of the  $FH\_N_{diff}$ , the peak latencies for the larger late phase of the  $FH\_N_{diff}$  showed significant temporal dispersion across the three delay conditions, suggesting that they likely reflect amplified object-specific categorization following the shifting and focusing of spatial attention onto the popout item. Notably, however, the latency difference between the no-delay and 50ms-delay conditions for the  $FH\_N_{diff}$  effect was not significant (12ms, n.s.) but between the 50ms-delay and 100ms-delay conditions it was (59ms;  $t_{14}=4.53$ ,  $p<0.0005$ ). This pattern of results suggests that the onset of the attentional enhancement following the attentional shift might have been accelerated by the initial parallel processing of the image array (see Discussion).

## Discussion

In the current study, results from both the behavioral and electrophysiological data confirm that participants' spatial attentional shift was delayed in a regular manner according to the delay of the occurrence of the attentional-shifting color-popout cues, as predicted. However,

<sup>2</sup>Note that one subject's data was excluded from the early-window analysis due to the absence of a peak in that time window, and another subject's data was excluded from the late-window analysis because it contained extreme outliers [greater than 2 SD from the mean] in two of the three delay conditions.

contrary to our predictions, the presence of an early time-invariant phase of the biphasic pattern observed in the FH\_N<sub>diff</sub> effect for each cue-delay, along with its latency being significantly earlier than that of the attentional-shift-sensitive N2pc effect, suggests that object-specific categorization processes were occurring without focal spatial attention (i.e., under distributed spatial attention prior to the shift of attention), at least to a sufficient degree to generate some face-specific neurophysiological activity. In addition, however, once focused spatial attention was shifted toward the cued location and its influence brought to bear, it amplified the object-specific processing of that item, as evidenced by the higher amplitudes, extended duration, and distinctive shifting pattern across the different cue-delay conditions of the second phase of the FH\_N<sub>diff</sub>.

Various models have posited that focal spatial attention is required to integrate different features of an object to form a coherent perception of it within a complex visual scene (e.g., Treisman, 1999; Treisman & Gelade, 1980; Wolfe & Cave, 1999). However, a number of behavioral studies have suggested a need for some modification of this theoretical view (e.g., Li et al., 2002; Nakayama & Silverman, 1986; Reddy et al., 2004; Treisman & Sato, 1990; Wolfe et al., 1989). In the current study, we used a delayed-cuing visual search paradigm combined with neurophysiological recordings to provide new insights into this issue from a cognitive neuroscience standpoint. In particular, using the variable temporal delays of the popout cue onsets, our results showed the different characteristics for the neuronal representation of object-specific categorization under broadly distributed spatial attention versus focal spatial attention. Furthermore, the high temporal resolution of the EEG recordings helped reveal the temporal cascades of object-specific visual processing within the visual cortices during the period of transition from a distributed-attention state to a focal-attention state within the visual cortices.

The first major finding is the time-invariant nature of object-specific categorization activity observed in the early phase of the face-selective brain activity (the FH\_N<sub>diff</sub> effect peaking at 140–150 ms), occurring even before focal attention had been shifted to the cued location, indicating that some degree of high-level perceptual categorization can occur in parallel across the visual scene without focal spatial attention. Three lines of evidence suggest that the early FH\_N<sub>diff</sub> is functionally equivalent to the typical N170 face-specific effect, which reflects a categorical discrimination between face and non-face stimuli (Bentin et al., 1996; Carmel & Bentin, 2002; Rossion et al., 2000; Rousselet, Husk, Bennett, & Sekuler, 2008). First, the FH\_N<sub>diff</sub> was extracted by the same type of face-versus-nonface contrast used to extract the N170 component. Second, the peak time for the early FH\_N<sub>diff</sub> is in the latency range of 140–150 ms (Fig. 4), which is in the range of previous findings in the N170 literature (Bentin et al., 1996; Carmel & Bentin, 2002; Rossion et al., 2000; Rousselet et al., 2008). Third, the topographic distribution of this effect on the scalp (largest at ventrolateral occipital sites, Fig. 3) is highly consistent with that of previous studies of the N170. Accordingly, if focal spatial attention were necessary for correct high-level object categorization, and if this early face/non-face categorization effect were due to attention-dependent processing at the cued location, the peak latency of the early FH\_N<sub>diff</sub> effect should not only be delayed relative to a typical time of occurrence for a no\_delay condition, but it should have also varied progressively with the different cue-delay conditions. However, such a temporal pattern was *not* observed for the peak latencies of the early face-specific FH\_N<sub>diff</sub> effect, but only for the latencies of the later phase of the FH\_N<sub>diff</sub> (Table 1). These results together therefore neurophysiologically support that at least some degree of object-specific categorization processing must have occurred before focal spatial attention arrived. The results also show that the categorization-related activity was then boosted and extended once focused attention was directed to that location.

It should be noted that the early  $FH_{N_{diff}}$  effect emerged under a *distributed* attention state over the whole array. Accordingly, this condition is actually different from the *ignore* condition in our previous study (Crist et al., 2007) that showed no significant N170 effect at an unattended location when attention was strongly focused elsewhere. These findings therefore also echo previous findings reporting that participants can perform certain higher-level perceptual categorization tasks in parallel (e.g., Rousselet, Fabre-Thorpe, & Thorpe, 2002) or under substantially reduced levels of focal spatial attention (Li et al., 2002; Reddy et al., 2004; VanRullen, Reddy, & Koch, 2004). In the present case, the findings support the view that some degree of such processing can be performed in parallel on multiple items across the visual field under circumstances when attention is distributed across that field and thus across those items.

As noted above, the pattern of results suggests that at least some degree of object-specific categorization processing was able to occur before focal spatial attention arrived. To the extent that one takes the view that feature binding of some sort needs to occur in order for object-specific categorization to take place, then this also would imply that some degree of feature binding must have occurred before focal spatial attention arrived. Although this result would appear to contradict a key prediction of Feature Integration Theory, namely that focused spatial attention is necessary for object feature integration, it is important to consider that this might depend on what perceptual process the early face-selective  $FH_{N_{diff}}$  activity reflects. A useful model here may be the concept of “feature bundling” vs. “feature binding” proposed by Wolfe and Cave (Wolfe & Cave, 1999). According to this model, features from objects are “bundled” together without correct arrangement under distributed spatial attention (“pre-attentively”) (Gajewski & Brockmole, 2006; Quinlan, 2003; Takegata et al., 2005). Furthermore, according to this view, focal attention is necessary to then correctly “bind” features together to form a coherent perception for identification. Accordingly, the early  $FH_{N_{diff}}$  that occurred before the allocation of directed focal attention may reflect such a hypothesized feature-bundling effect that can be performed under distributed attention, whereas the later  $FH_{N_{diff}}$  activity may reflect a focal-attention-requiring feature-binding effect. Thus, it may be the case that the feature-bundling processing is enough for differentiating face vs. non-face stimuli, but not enough for higher-level object-related processing, such as identifying a specific face. Several previous findings using ERPs reported the processing of familiar versus unfamiliar faces showed differential effects on an N250 (Schweinberger, Pickering, Jentzsch, Burton, & Kaufmann, 2002) or an N400 component (Eimer, 2000), but not on the N170 component, suggesting that the N170 component may represent a structure coding process for faces but not for identification (Eimer, 2000; Marzi & Viggiano, 2007; Schweinberger et al., 2002).

The emergence of the early phase of the  $FH_{N_{diff}}$  effect may seem to be surprising considering that each search array contained equal numbers of faces and houses, which we hypothesized would result in no net face-selective effect from the array. However, due to the design of three images in each hemifield, when the data were sorted based on the image category of the popout item, there is an inherent imbalance of the “category” distribution between the hemifield that the popout is on relative to the other side. More specifically, trials in which the color popout was a face would be statistically slightly more likely to have more faces on that side than on the other side (i.e., two on that side and only one on the other), and vice versa for the “color-popout house” trials. Consequently, assuming there is a face-selective processing difference at any of the locations under distributed attention (which there appears to be) the contrast between these two types of trials would yield a face-selective effect on the popout side (even though the *net* face-selective activity from the whole array across the entire visual field would be zero). The early phase therefore appeared to result from the interplay of distributed attention and the inherent imbalance of the search array between hemifields in this study. Regardless, however, the temporally invariant nature

of the early phase suggests that distributed attention is not only sufficient for object-specific discrimination across multiple items, but that it is also sensitive to the statistical difference of object categories across the field (Treisman, 2006).

One could argue that the early  $FH\_N_{diff}$  component might be due to the popout characteristic of face stimuli because they are particularly ecologically salient (Hershler & Hochstein, 2005; but see VanRullen, 2006). That is, it could be argued that the imbalanced stimuli could have served as a peripheral cue to rapidly shift some attention to the hemifield with the greater number of faces. There are several lines of evidence in the data, however, that would argue against this possibility. First, the behavioral results showed no significant advantage for correct blurred-target identification for faces as compared to houses. Second, every 6-item image array had an equal number of faces and houses, and thus there should not have been any *a priori* reason for a singleton popout effect to any particular face. In contrast, there was only one color singleton, which was task-relevant and induced the later color-cue  $N2pc$ . Third, if the inherent imbalance of the category distribution could have served as a peripheral cue to trigger an attentional shift toward the face-dominant side, it would have been expected to elicit a very fast and early  $N2pc$  contralateral to that side. Such an early neural shift should then have appeared with the lateralized parietal-occipital distribution characteristic of the  $N2pc$ , an effect which was not observed in the data. Rather, the first effect, the early face-selective negativity effect, had the classic ventro-lateral occipital distribution of a face-selective  $N170$  and occurred before any indication of any attentional shifting. It was then after this that we see the classic  $N2pc$  parietal-occipital reflection of a spatial shift of attention being elicited by the extremely salient, and highly task-relevant, color-popout cues. Accordingly, it seems rather unlikely that the emergence of the early  $FH\_N_{diff}$  was due to shift of spatial attention triggered by a face pop-out effect.

A second major finding was the absence of a significant difference for the peak latencies of the late  $FH\_N_{diff}$  effect between the no\_delay and 50ms\_delay conditions ( $\sim 15\text{--}20$  ms, n.s.), despite the fact that focal attention was shifted 50 ms later in the second case (Fig. 4b; Table 1). At first view, these results may seem surprising. Based on the logic discussed above, the late  $FH\_N_{diff}$  component represents a delayed attentional modulation effect on object-categorization-specific perceptual processing once spatial attention was directed toward that object. Therefore, a reasonable prediction was that the peak latency of this late  $FH\_N_{diff}$  would parallel the temporal change patterns of the  $N2pc$ , following it shortly after in time and thus spreading out in time accordingly. A possible explanation for the observed somewhat shorter-latency pattern may be that the search array had been presented and processed (i.e., up through the retina and the low-level visual cortices) before attention was shifted, which may have subsequently enabled a faster influence of focused attention on the object-specific visual processing once it was selectively directed. The presence of a significant delay of  $\sim 70$  ms of the late-phase face-selective activity for the 100ms\_delay condition, as compared to the no\_delay condition, suggests that such a pre-processing advantage could adjust for only part of the 100-ms attentional shift delay (i.e.,  $\sim 30$  ms). Thus, the patterns of these effects indicate a temporal processing advantage for shifting spatial attention toward visual information that is already in the visual system, as compared to the situation in which spatial attention is shifted to newly presented stimuli on the visual field.

In conclusion, our study reveals several important new findings concerning attention and perceptual categorization processes in the brain. First, it shows that some degree of object-specific representation can be performed in parallel across multiple objects in the visual field when spatial attention is distributed across those objects (i.e., before spatial attention is focused on one of them), and our neural measures indicate when that occurs. Secondly, once spatial attention is shifted to a specific location, it quickly results in a robust enhancement

and extension of the neural activity related to the object-categorization discrimination process of the object at that location. Third, initial parallel processing of the input, due to the presence of the stimulus information in the visual pathways prior to attentional shifting, appears to speed up the attentional enhancement of the neural processing underlying object-related discrimination of an item once focused attention is brought to bear on that item.

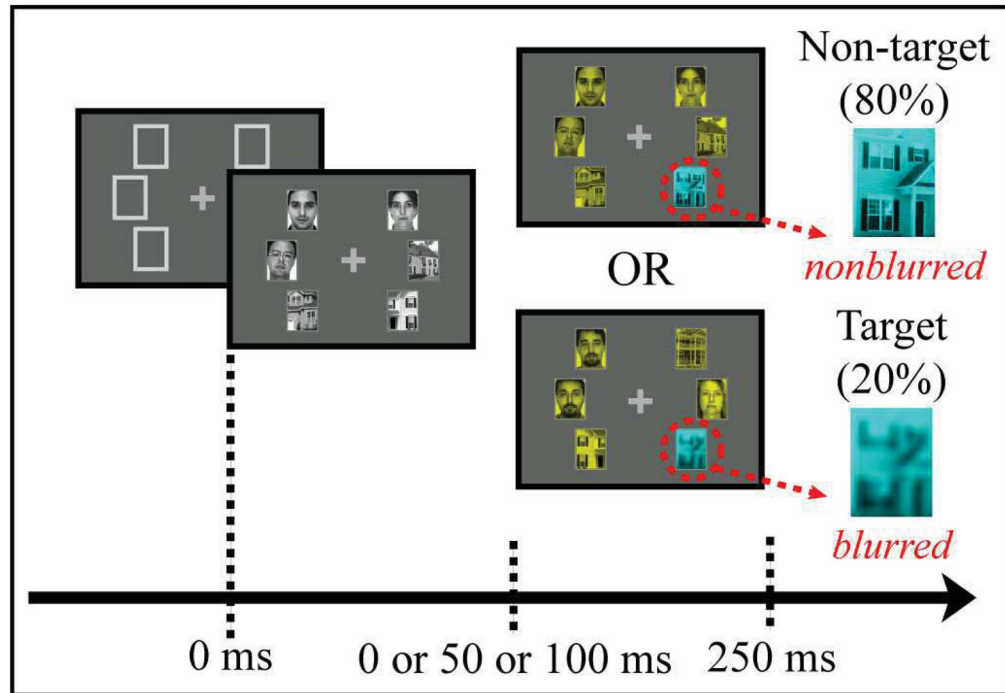
## Acknowledgments

The authors would like thank Nico Boehler, Tineke Grent-'t-Jong, Shih-Chieh Lin, and Rufin VanRullen for providing valuable comments and suggestions on earlier versions of the manuscript, and Greg Appelbaum and Mark Pettet for discussion on the permutation tests. This research was supported by NIH grants R01-MH060415 and R01-NS051428 to M.G.W.

## Reference List

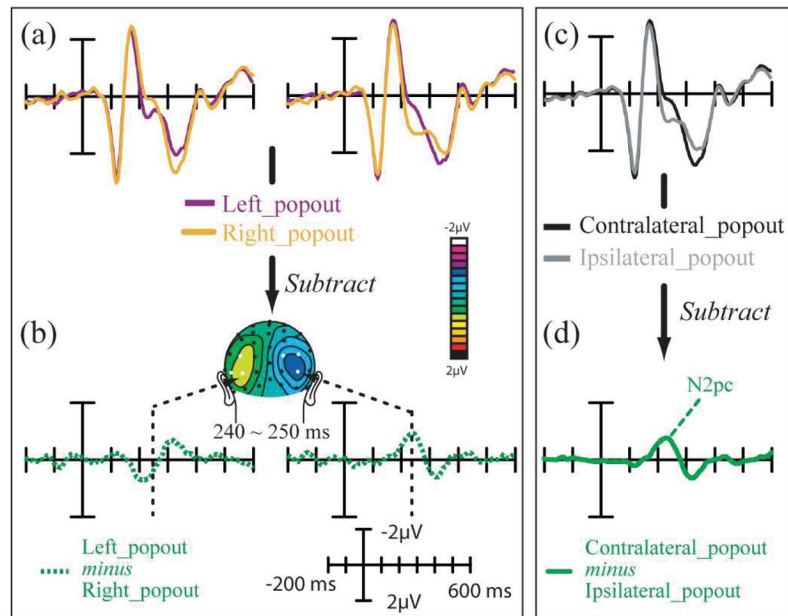
- Appelbaum LG, Wade AR, Vildavski VY, Pettet MW, Norcia AM. Cue-invariant networks for figure and background processing in human visual cortex. *Journal of Neuroscience*. 2006; 26:11695–11708. [PubMed: 17093091]
- Bentin S, Allison T, Puce A, Perez A, McCarthy G. Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*. 1996; 8:551–565. [PubMed: 20740065]
- Blair RC, Karniski W. An Alternative Method for Significance Testing of Wave-Form Difference Potentials. *Psychophysiology*. 1993; 30:518–524. [PubMed: 8416078]
- Carlson JM, Reinke KS. Spatial attention-related modulation of the N170 by backward masked fearful faces. *Brain Cogn*. 2010; 73:20–27. [PubMed: 20207467]
- Carmel D, Bentin S. Domain specificity versus expertise: factors influencing distinct processing of faces. *Cognition*. 2002; 83:1–29. [PubMed: 11814484]
- Cohen A, Ivry R. Illusory Conjunctions Inside and Outside the Focus of Attention. *Journal of Experimental Psychology-Human Perception and Performance*. 1989; 15:650–663. [PubMed: 2531202]
- Crist RE, Wu CT, Karp C, Woldorff MG. Face Processing is Gated by Visual Spatial Attention. *Front Hum Neurosci*. 2007; 1:10. [PubMed: 18958224]
- Eimer M. The N2pc component as an indicator of attentional selectivity. *Electroencephalogr Clin Neurophysiol*. 1996; 99:225–234. [PubMed: 8862112]
- Eimer M. Effects of face inversion on the structural encoding and recognition of faces - Evidence from event-related brain potentials. *Cognitive Brain Research*. 2000; 10:145–158. [PubMed: 10978702]
- Gajewski DA, Brockmole JR. Feature bindings endure without attention: Evidence from an explicit recall task. *Psychonomic Bulletin & Review*. 2006; 13:581–587. [PubMed: 17201355]
- Hershler O, Hochstein S. At first sight: a high-level pop out effect for faces. *Vision Res*. 2005; 45:1707–1724. [PubMed: 15792845]
- Hopf JM, Boelmans K, Schoenfeld MA, Luck SJ, Heinze HJ. Attention to features precedes attention to locations in visual search: Evidence from electromagnetic brain responses in humans. *Journal of Neuroscience*. 2004; 24:1822–1832. [PubMed: 14985422]
- Hubel DH, Wiesel TN. Receptive fields and functional architecture of monkey striate cortex. *J Physiol*. 1968; 195:215–243. [PubMed: 4966457]
- Koch C, Tsuchiya N. Attention and consciousness: two distinct brain processes. *Trends in Cognitive Sciences*. 2007; 11:16–22. [PubMed: 17129748]
- Kovacs G, Zimmer M, Harza I, Antal A, Vidnyanszky Z. Position-specificity of facial adaptation. *Neuroreport*. 2005; 16:1945–1949. [PubMed: 16272884]
- Li FF, VanRullen R, Koch C, Perona P. Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences of the United States of America*. 2002; 99:9596–9601. [PubMed: 12077298]
- Luck SJ, Hillyard SA. Spatial-Filtering During Visual-Search - Evidence from Human Electrophysiology. *Journal of Experimental Psychology-Human Perception and Performance*. 1994; 20:1000–1014. [PubMed: 7964526]

- Martinez, AM.; Benavente, R. The AR Face Database (Rep No CVC Technical Report #24). 1998.
- Marzi T, Viggiano MP. Interplay between familiarity and orientation in face processing: An ERP study. *Int J Psychophysiol.* 2007
- Maunsell JH, Van Essen DC. Functional properties of neurons in middle temporal visual area of the macaque monkey. I Selectivity for stimulus direction, speed, and orientation. *J Neurophysiol.* 1983; 49:1127–1147. [PubMed: 6864242]
- Nakayama K, Silverman GH. Serial and Parallel Processing of Visual Feature Conjunctions. *Nature.* 1986; 320:264–265. [PubMed: 3960106]
- Prinzmetal W. Principles of Feature Integration in Visual-Perception. *Perception & Psychophysics.* 1981; 30:330–340. [PubMed: 7322810]
- Quinlan PT. Visual feature integration theory: Past, present, and future. *Psychological Bulletin.* 2003; 129:643–673. [PubMed: 12956538]
- Reddy L, Wilken P, Koch C. Face-gender discrimination is possible in the nearabsence of attention. *Journal of Vision.* 2004; 4:106–117. [PubMed: 15005651]
- Roskies AL. The binding problem. *Neuron.* 1999; 24:7–25. [PubMed: 10677022]
- Rossion B, Gauthier I, Tarr MJ, Despland P, Bruyer R, Linotte S, et al. 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.* 2000; 11:69–74. [PubMed: 10683832]
- Rousselet GA, Fabre-Thorpe M, Thorpe SJ. Parallel processing in high-level categorization of natural images. *Nat Neurosci.* 2002; 5:629–630. [PubMed: 12032544]
- Rousselet GA, Husk JS, Bennett PJ, Sekuler AB. Time course and robustness of ERP object and face differences. *J Vis.* 2008; 8:3–18. [PubMed: 18831616]
- Schweinberger SR, Pickering EC, Jentsch I, Burton AM, Kaufmann JM. Event-related brain potential evidence for a response of inferior temporal cortex to familiar face repetitions. *Brain Res Cogn Brain Res.* 2002; 14:398–409. [PubMed: 12421663]
- Takegata R, Brattico E, Tervaniemi M, Varyagina O, Naatanen R, Winkler L. Preattentive representation of feature conjunctions for concurrent spatially distributed auditory objects. *Cognitive Brain Research.* 2005; 25:169–179. [PubMed: 15953710]
- Treisman A. Solutions to the binding problem: progress through controversy and convergence. *Neuron.* 1999; 24:105–125. [PubMed: 10677031]
- Treisman A. How the deployment of attention determines what we see. *Vis cogn.* 2006; 14:411–443. [PubMed: 17387378]
- Treisman A, Sato S. Conjunction search revisited. *J Exp Psychol Hum Percept Perform.* 1990; 16:459–478. [PubMed: 2144564]
- Treisman A, Schmidt H. Illusory Conjunctions in the Perception of Objects. *Cognitive Psychology.* 1982; 14:107–141. [PubMed: 7053925]
- Treisman AM, Gelade G. A feature-integration theory of attention. *Cognit Psychol.* 1980; 12:97–136. [PubMed: 7351125]
- VanRullen R. On second glance: Still no high-level pop-out effect for faces. *Vision Res.* 2006; 46:3017–3027. [PubMed: 16125749]
- VanRullen R, Reddy L, Koch C. Visual search and dual tasks reveal two distinct attentional resources. *Journal of Cognitive Neuroscience.* 2004; 16:4–14. [PubMed: 15006031]
- Wolfe JM, Cave KR. The psychophysical evidence for a binding problem in human vision. *Neuron.* 1999; 24:11–25. [PubMed: 10677023]
- Wolfe JM, Cave KR, Franzel SL. Guided search: an alternative to the feature integration model for visual search. *J Exp Psychol Hum Percept Perform.* 1989; 15:419–433. [PubMed: 2527952]
- Woodman GF, Luck SJ. Electrophysiological measurement of rapid shifts of attention during visual search. *Nature.* 1999; 400:867–869. [PubMed: 10476964]
- Zeki SM. Colour coding in rhesus monkey prestriate cortex. *Brain Res.* 1973; 53:422–427. [PubMed: 4196224]



**Figure 1. Trial Structure**

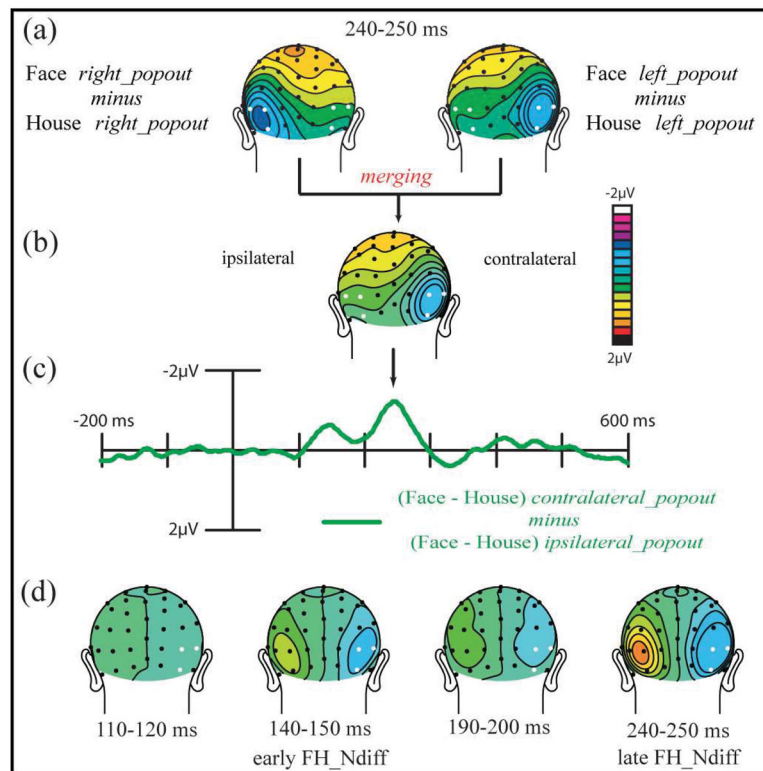
Each trial began with a circular image array containing three gray-scale faces and three gray-scale houses whose locations were randomly shuffled from trial to trial. After a delay of either 0 ms, 50 ms, or 100 ms, the same image array was colored to form a color-popout display, with five turning one color and one turning another color. (Note that for the 0 ms delay trials, participants would immediately see the color-popout display). Participants were instructed to covertly shift their focal spatial attention to the color-popout location immediately after the onset of the color-popout display and to press a button if the image at the popout location was blurred (i.e., a target), regardless of the image type (i.e., regardless of whether it was a face or a house). Thus, the image-blurriness detection task was orthogonal to any categorization of the images. Only 20% of trials contained targets. Each image array was presented for 250 ms, with a trial onset asynchrony varied from 1100–1400 ms.



**Figure 2. Analysis sequence to extract the contralateral N2pc effect**

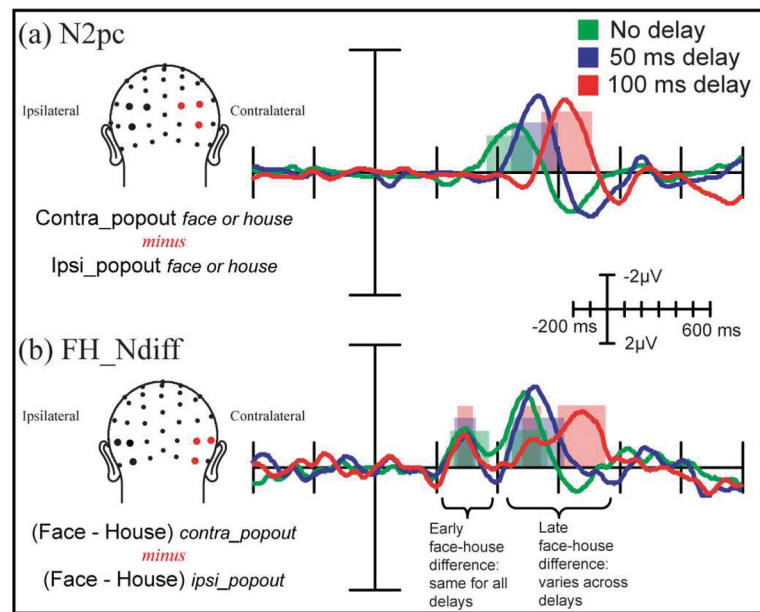
The extraction of this effect is illustrated in the No-delay condition using data collapsed across the three occipito-temporal channels marked in white on the scalp electrode-layout display. a) First, the averaged ERPs were generated separately for trials in which the color-popout was presented in the left visual field (left\_popout) and trials in which it was on the right (right\_popout). b) As shown previously in numerous EEG popout studies, the subtraction between the left\_popout and right\_popout conditions reveals a relative negative-polarity difference over parietal-occipital scalp contralateral to the attention-shifting direction (and a corresponding positive polarity difference ipsilaterally) at around 200–300 ms. c) To combine these data over left-popout and right-popout trials, the contralateral ERP traces for the left-popout trials (i.e., right hemisphere sites) and the contralateral traces for the right-popout trials (i.e., left hemisphere sites) were collapsed together, and the corresponding ipsilateral traces for the left-popout and right-popout trials collapsed together. d) The N2pc activity was then obtained by subtracting the averaged ERPs of the ipsilateral channels from the averaged ERPs of the contralateral channels.





**Figure 3. Analysis sequence to extract the contralateral face-house negativity difference effect (FH<sub>Ndiff</sub>)**

The extraction of this effect is illustrated in the No-delay condition using data collapsed across the three parieto-occipital channels marked in white. a) The face-house negativity difference (FH<sub>Ndiff</sub>) was first generated by subtracting the house ERPs from the face ERPs for the same popout side (i.e., both on the left or both on the right). Note that the FH<sub>Ndiff</sub> effect mainly emerged on the scalp region contralateral to the attended side. b) To further emphasize the attentional enhancement of the face/house categorization effect that is contralateral to the color popout locations (i.e., contralateral to the focus of spatial attention after the attentional shift), the contralateral FH<sub>Ndiff</sub> effects for the left-popout trials (i.e., right hemisphere sites) and the contralateral FH<sub>Ndiff</sub> effects for the right-popout trials (i.e., left hemisphere sites) were collapsed together, and the corresponding ipsilateral FH<sub>Ndiff</sub> effects for the left-popout and right-popout trials were collapsed together. c) The contralateral FH<sub>Ndiff</sub> effects were then obtained by subtracting the averaged ipsilateral ERPs from the averaged contralateral ERPs. Note the biphasic response structure (early phase and late phase). d) A similar topographic scalp potential distribution between the early phase (140~150 ms) and the larger-amplitude later phase (240~250 ms) can be observed. Note that the right hemisphere reflects contralateral minus ipsilateral FH<sub>Ndiff</sub> effects while the left hemisphere reflects the opposite subtraction.



**Figure 4. Different temporal dynamic patterns of spatial attentional shifting and object-specific categorization processing**

Averaged ERP waveforms from the three corresponding channels of interests (scalp locations marked in red) that showed the strongest effects for each component are plotted for the a) N2pc and b) the face-house negativity difference (FH\_N<sub>diff</sub>) effects. The temporal pattern of the N2pc effect, which reflects the shifting of focal spatial attention, showed a clear temporal dispersion, with a 40–50 ms step-delay across the no\_delay, 50ms\_delay and 100ms\_delay conditions. In contrast, the temporal patterns of the FH\_N<sub>diff</sub> effect, which reflects object-specific categorization processing, showed a biphasic deflection, with a small negative deflection in the early time window from 140–180 ms (approximately the typical latency of the conventional face-specific N170 component) that was time-invariant across the delay conditions, followed by a second, later, negative deflection from 220–400 ms that shifted temporally across the cue-delay conditions. In particular, for the second negativity, the no\_delay and 50ms\_delay conditions showed relatively close peak latencies of the face-selective FH\_N<sub>diff</sub>, whereas this later reflection of face-selective activity was significantly later in the 100ms-delay condition. The transparently shaded areas indicate the time windows with a statistically significant effect ( $p < 0.05$ ) based on the probing ANOVAs using 6ms-moving windows between 120~350 ms (see Methods). Note that both effects were extracted from the difference waveforms derived by subtracting the ERP waveforms ipsilateral to the attentional focus from those contralateral to the attentional focus, for each delay condition.

**Table 1**

The peak latencies of the N2pc and FH\_N<sub>diff</sub> component for three cue-delay conditions obtained by different latency tests

Testing Methods	N2pc peak latency		
	No_delay	50ms_delay	100ms_delay
<b>Grand-average peak latencies</b> *	228 ms	268 ms	308 ms
<b>Mean of individual-participant peak latencies</b>	227 ms	263 ms	314 ms

Testing Methods	FH_N <sub>diff</sub> peak latency		
	No_delay	50ms_delay	100ms_delay
<i>Early phase (0–180 ms)</i>			
<b>Grand-average peak latencies</b> *	148 ms	148 ms	146 ms
<b>Mean of individual-participant peak latencies</b>	141 ms	143 ms	145 ms
<i>Late phase (190–400 ms)</i>			
<b>Grand-average peak latencies</b> *	246 ms	262 ms	338 ms
<b>Mean of individual-participant peak latencies</b>	253 ms	265 ms	324 ms

\* Note: The grand-average peak latencies were the measured peak latencies from the grand-average ERPs in the latency windows of significant activity established by both the 6ms-moving-window ANOVA and the permutation test.