

Early Top–Down Control of Visual Processing Predicts Working Memory Performance

Aaron M. Rutman, Wesley C. Clapp, James Z. Chadick,
and Adam Gazzaley

Abstract

■ Selective attention confers a behavioral benefit on both perceptual and working memory (WM) performance, often attributed to top–down modulation of sensory neural processing. However, the direct relationship between early activity modulation in sensory cortices during selective encoding and subsequent WM performance has not been established. To explore the influence of selective attention on WM recognition, we used electroencephalography to study the temporal dynamics of top–down modulation in a selective, delayed-recognition para-

digm. Participants were presented with overlapped, “double-exposed” images of faces and natural scenes, and were instructed to either remember the face or the scene while simultaneously ignoring the other stimulus. Here, we present evidence that the degree to which participants modulate the early P100 (97–129 msec) event-related potential during selective stimulus encoding significantly correlates with their subsequent WM recognition. These results contribute to our evolving understanding of the mechanistic overlap between attention and memory. ■

INTRODUCTION

Goal-directed selective attention influences the magnitude and speed of neural processing in cortical regions where sensory information is actively represented, via a process known as top–down modulation (Gazzaley, Cooney, McEvoy, Knight, & D’Esposito, 2005; Kastner & Ungerleider, 2000; Luck, Chelazzi, Hillyard, & Desimone, 1997; Desimone & Duncan, 1995). Many studies have capitalized on the high temporal resolution of electroencephalography (EEG) to reveal early influences of top–down control on visual processing in humans (Hillyard & Anllo-Vento, 1998), and more recently to establish a direct relationship between neural measures of modulation and indicators of behavioral performance, such as the speed of stimulus detection (Talsma, Mulckhuysse, Slagter, & Theeuwes, 2007; Thut, Nietzel, Brandt, & Pascual-Leone, 2006). Furthermore, evidence has emerged that demonstrates a mechanistic overlap between the processes of selective attention and working memory (WM). Several studies have revealed a major role of WM in the control of visual selective attention (Awh & Jonides, 2001; de Fockert, Rees, Frith, & Lavie, 2001; Desimone, 1996), whereas others have shown that selective attention is a key component of WM (Awh & Jonides, 2001). Recent studies utilizing EEG have investigated the time course of attentional involvement in WM, presenting a model in which attention is utilized throughout the WM maintenance period (Sreenivasan,

Katz, & Jha, 2007; Jha, 2002), likely by biasing cortical processing of relevant sensory representations and activity modulation of distractors (Sreenivasan & Jha, 2007). Although data have revealed that WM maintenance may depend on temporally early attentional factors (Sreenivasan et al., 2007), notably for distracting information (Zanto & Gazzaley, 2009), a direct correlation between early neural measures of selective activity modulation during encoding and subsequent WM performance has not yet been described.

Selective attention results in activity modulation at very early stages of visual processing (Schoenfeld, Hopf, Martinez, & Mai, 2007; Martinez et al., 2006; Khoe, Mitchell, Reynolds, & Hillyard, 2005; López, Rodríguez, & Valdés-Sosa, 2004; Pinilla, Cobo, Torres, & Valdes-Sosa, 2001; Valdes-Sosa, Bobes, Rodriguez, & Pinilla, 1998), including amplitude modulations of the P100 (~100 msec) and N170 (~170 msec) event-related potential (ERP) components (see Hillyard & Anllo-Vento, 1998), which have been localized to visual cortical areas in lateral extrastriate cortex (Di Russo, Martínez, Sereno, Pitzalis, & Hillyard, 2002; Gomez Gonzalez, Clark, Fan, Luck, & Hillyard, 1994). We hypothesize that such early top–down modulation of cortical activity reflects the fidelity of sensory representations of relevant information in such a manner that it confers a behavioral benefit on maintaining that information in mind.

Here we explore how early markers of visual processing that are modulated when attention is selectively directed to complex, real-world visual objects (i.e., human faces or natural scenes) relate to subsequent WM recognition

performance. Our study utilized a delayed recognition task in which participants were instructed to remember two stimuli (800 msec each) over the course of a 4-sec delay period (Figure 1). We used overlapping transparent images of faces and scenes, with either the face or the scene relevant (and the other irrelevant) for the WM task, in a design similar to previous studies of object-based attention (Furey et al., 2006; Yi & Chun, 2005; Serences, Schwarzbach, Courtney, Golay, & Yantis, 2004; O'Craven, Downing, & Kanwisher, 1999). Recording posterior EEG measures while participants viewed the overlapped stimuli during the encoding period (equivalent bottom-up input with variations only in instructions) enabled us to evaluate the timing of top-down modulation and correlate these measures with recognition accuracy recorded after the delay period.

METHODS

Participants

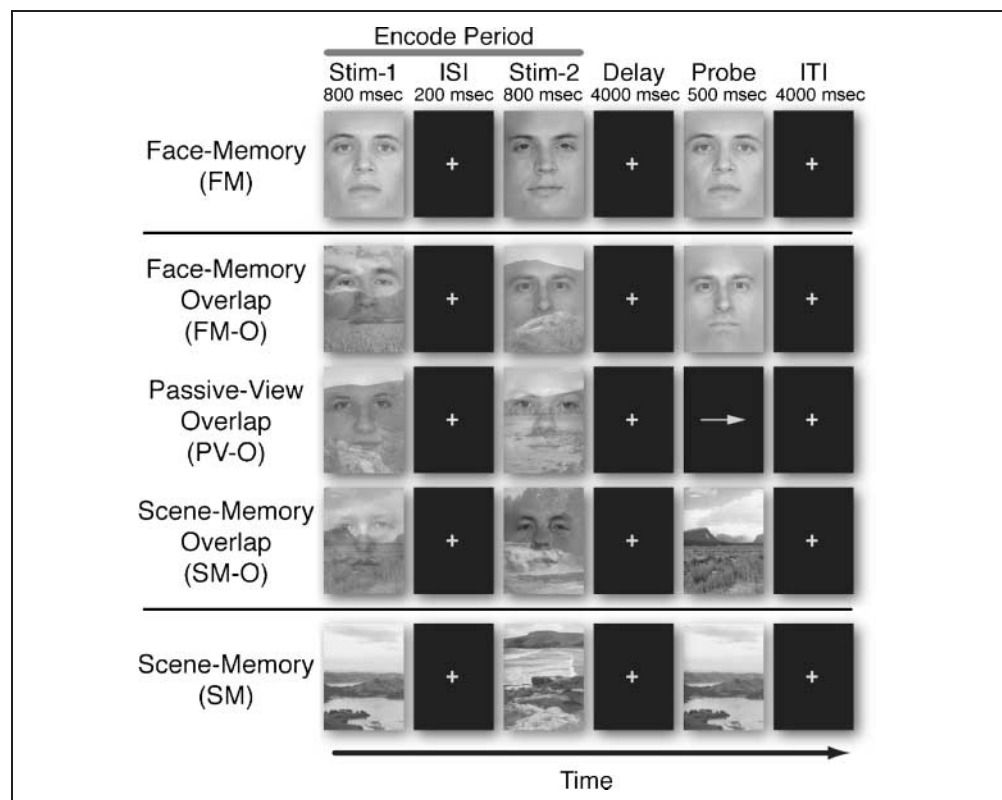
Nineteen healthy, right-handed individuals (mean age = 22.9 years; range = 18–34 years; 10 men) with normal or corrected-to-normal vision volunteered, gave consent,

and were monetarily compensated to participate in the study. Participants were prescreened, and none used any medication known to affect cognitive state.

Stimuli

The stimuli consisted of grayscale images of faces and natural scenes. All face and scene stimuli were novel across all tasks, across all runs, and across all trials of the experiment. Images were 225 pixels wide and 300 pixels tall (14 × 18 cm), and were presented foveally, subtending a visual angle of 3° from a small cross at the center of the image. The face stimuli consisted of a variety of neutral-expression male and female faces across a large age range. Hair and ears were removed digitally, and a blur was applied along the contours of the face as to remove any potential non-face-specific cues. The sex of the face stimuli was held constant within each trial. Images of scenes were not digitally modified beyond re-sizing and gray-scaling. For the tasks consisting of overlapped faces and scenes, one face and one scene were randomly paired, made transparent, and digitally overlapped using Adobe Photoshop CS2 such that both the

Figure 1. Experimental paradigm. Five different tasks were presented in a delayed-recognition task design. All trials involved viewing two images (Stim-1, Stim-2) (encode), followed by a 4-sec period (delay), and concluded with a third image (probe). Encoding stimuli in FM and SM consisted of isolated pictures of faces and natural scenes, whereas encoding stimuli in the three overlap tasks (FM-O, PV-O, and SM-O) consisted of an overlapped scene and a face picture. For the four memory tasks, participants were instructed to remember the relevant encoding stimuli, maintain the images in mind over the delay period, and respond with a button press whether or not the probe image matched one of the two encoding period images. For the passive view task (PV-O), participants were instructed to relax and view the double-exposed images without trying to remember them, after which they responded to the direction of an arrow with a button press.



face and the scene were equally visible. Overlapped and isolated images were randomly assigned to the different tasks.

Experimental Procedures

The experimental paradigm was comprised of five different tasks in a delayed-recognition WM task design (Figure 1). Each task consisted of the same temporal sequence with only the instructions differing across tasks. All tasks involved viewing two images (Stim-1, Stim-2), each being displayed for 800 msec (with a 200-msec ISI). These images were followed by a 4-sec period (delay) in which the images were to be held in mind (mentally rehearsed). After the delay, a third image appeared (probe). The participant was instructed to respond with a button press (as quickly as possible without sacrificing accuracy) whether or not the probe image matched one of the previous two images (Stim-1, Stim-2). This was followed by an intertrial interval (ITI) lasting 4 sec.

For three of the five tasks, the Stim-1 and Stim-2 images were composed of both a scene and a face superimposed upon each other. For these double-exposed images, the participants were instructed to focus their attention on and hold in mind either the face or the scene, while ignoring the other. In the face memory-overlap task (FM-O), the faces were held in mind while the scenes were ignored, and vice versa in the scene memory-overlap task (SM-O). When the probe image appeared, it was composed of an isolated face in the FM-O task, or an isolated scene in the SM-O task. For the passive view (PV-O) task, participants were instructed to relax and view the double-exposed images without trying to hold them in mind, after which they responded to an arrow direction with a button press. For the other two tasks, the Stim-1 and Stim-2 images were each composed of a single stimulus without any distracting information: a face in the face memory task (FM) and a scene in the scene memory task (SM). The task was presented in three separate runs, each run consisting of each of the five task sets presented in blocks and counterbalanced in random order across all participants. Each task set consisted of a block of 20 trials of that task (60 total trials per task condition for all 3 runs, 120 total encode period images). Each blocked task set was preceded by an instruction screen cueing the subject to the specific memory goal of the task (i.e., “remember the faces”).

Following the main experiment, participants performed a surprise postexperiment recognition test in which they viewed 320 nonoverlapped images, including 160 faces and 160 scenes. Eighty of the faces and 80 of the scenes were novel stimuli that were not included in the main experiment. There were 20 faces each from the FM, FM-O, SM-O, and PV-O tasks, and 20 scenes each from the SM, SM-O, FM-O, and PV-O tasks. No encoded stimulus was included that was also a match during a trial of the main

experiment, so that no stimuli in the postexperiment test were seen more than once before. All included face and scene stimuli (both novel images and images from the experiment) were randomly ordered, and participants were asked to rate their confidence of recognition of each image as follows: 1 = definitely did not see the image during the course of the experiment; 2 = think that the image was not seen during the experiment; 3 = think that the image was seen during the course of the experiment; and 4 = definitely saw the image during the experiment. An incidental long-term memory recognition index for each stimulus was calculated by subtracting the rating of novel stimuli for each participant.

Eye-movement Control Experiment

Eye tracking was performed on five participants (recruited with the same exclusionary criteria) while they performed the main experiment with identical instructions. Data were collected on an ASL EYE-TRAC6 (Applied Science Laboratories, Bedford, MA) sampled at 60 Hz. Eye blinks were removed and data were high-pass filtered at 0.5 Hz using a fifth-order Butterworth filter to remove drift using MATLAB (MathWorks, Natick, MA). Across-condition time-series analysis was performed using paired *t* tests with an uncorrected alpha value of .05. Analyses of variance (ANOVAs) were calculated using a two-way repeated measures ANOVA, and post hoc *t* tests were performed for eye-position differences between conditions, using an alpha value of .05 with Tukey–Kramer correction.

Electrophysiological Recordings

Neural data were recorded at 1024 Hz through a 24-bit BioSemi ActiveTwo 64-channel Ag–AgCl active electrode EEG acquisition system in conjunction with BioSemi ActiView software (CortechSolutions, LLC, Wilmington, NC). Electrode offsets were maintained between ± 20 mV. Precise markers of stimulus presentation were acquired using a photodiode. Trials with excessive peak-to-peak deflections, amplifier clipping, or excessive high-frequency (EMG) activity were excluded prior to analysis.

Electrophysiological Data Analysis

Preprocessing was conducted through the EEGLAB toolbox (Swartz Center for Computational Neuroscience, UCSD, La Jolla, CA) for MATLAB. Off-line, the raw EEG data were high-pass filtered (0.5 Hz), referenced to an average reference, and segmented into epochs beginning 200 msec before stimulus onset and ending 800 msec after stimulus onset. Single epochs were baseline-corrected using an average from -200 to 0 msec before stimulus appearance. Eye movements and artifacts were removed through an independent

component analysis by excluding components consistent with topographies for blinks and eye movements and electrooculogram time series. Artifact-free data epochs were then split by task, filtered (1–30 Hz), and averaged, to create stimulus-locked ERPs.

Localizer Task

An independent functional localizer task was used to define electrodes of interest (EOIs) for each participant (Liu, Harris, & Kanwisher, 2002). The localizer task consisted of a 1-back design in which participants attended to seven blocks of 20 faces and seven blocks of 20 scenes. Participants were instructed to attend to the stimuli and to indicate when each 1-back match occurred by pressing a button with both forefingers. Face and scene blocks were randomly intermixed. Face and scene trials were then segmented separately and averaged. Epochs to repeated stimuli were not included in the average in order to prevent motor contamination in the ERP. The P100 component was identified at lateral posterior electrodes as the first positive deflection appearing between 50 and 150 msec after stimulus onset. The N170 component was identified at posterior sites as the maximal negative peak between 120 and 220 msec after stimulus onset. As revealed in previous studies, we found a significant preference for faces at both 100 msec (Herrmann, Ehlis, Ellgring, & Fallgatter, 2005; Liu et al., 2002), and 170 msec after stimulus onset (Herrmann et al., 2005; Liu et al., 2002; Bentin, Allison, Puce, Perez, & McCarthy, 1996) in components for all posterior-lateral electrodes, such that they revealed significantly larger amplitudes for faces versus scenes (electrodes P10, PO8, P8, O2, P9, PO7, P7, O1; all p values < .02). The lateral posterior electrode that showed the largest P100 and N170 amplitude difference between faces and scenes was defined as that participant's P100 EOI and N170 EOI, respectively. EOIs included the following electrodes: P8, P10, PO4, PO8, O2, P7, P9, PO7, and O1.

Event-related Potentials

Epochs from each task of the main experimental task were separately segmented, baselined at –200 to 0 msec relative to stimulus onset, and then averaged. Only encoding-period segments (Stim-1, Stim-2) from correct trials were included. ERPs from each of the tasks included a mean of 116 averaged epochs per participant per task (range 80–120). The peak of the P100 ERP component for each posterolateral electrode was defined as the maximal positive voltage of the first positive deflection appearing between 50 and 150 msec after stimulus onset, whereas the peak of the N170 component was defined as the maximal negative voltage between 120 and 220 msec after stimulus onset. After the peak was identified for each individual,

ERP amplitudes were then calculated as the area ± 4 msec from the peak latency. Across-participant ERP ANOVA and t test statistics were calculated using amplitudes and latencies from each participant's EOI.

Statistical Analysis

Behavioral and ERP data were each subjected to a repeated measures 2×2 ANOVA (with stimuli type and overlap as factors) and checked against a normal distribution using a Lilliefors test. Post hoc two-tailed t tests were corrected for multiple comparisons using Tukey's honestly significant difference criterion and an alpha of .05. Time windows for significant divergence of face and scene localizer data were calculated using paired t tests for each time point. These were not corrected for multiple comparisons under the assumption that time-dependent measures are not independent comparisons.

RESULTS

Behavioral Results

WM accuracy and response time (RT) data were subjected to separate, repeated measures 2×2 ANOVA with the type of stimulus attended (face vs. scene) and overlap status (overlapped vs. nonoverlapped) as factors. WM accuracy revealed a main effect of overlap [$F(1, 18) = 55.05, p < .0001$], such that accuracy was significantly reduced in tasks with overlapped stimuli relative to tasks with face and scene stimuli presented in isolation (FM-O: 82.7% vs. FM: 89.5%, $p < .01$; SM-O: 83.9% vs. SM: 92.9%, $p < .01$; Figure 2A). This WM performance reduction for the overlapping stimuli was also evident as an increased RT for overlap tasks [$F(1, 18) = 15.09, p < .001$] (FM-O: 1096 msec vs. FM: 1055 msec, $p = .09$; SM-O: 1103 msec vs. SM: 1029 msec, $p < .01$; Figure 2B).

There was a main effect of stimulus for WM accuracy [$F(1, 18) = 4.8, p < .05$], but no interaction between stimulus and overlap [$F(1, 18) = 1.17, p < .287$]; post hoc comparisons revealed that accuracy was reduced for faces compared to scenes, only in the nonoverlapped tasks (SM: 92.9%, FM: 89.5%, $p < .01$). There was no main effect of stimulus for RT, and no interaction between stimulus and overlap for RT. Accuracy in the passive view (PV-O) task was 99.3%; RTs to arrow direction averaged 593 msec.

Results of the surprise postexperiment recognition test revealed that participants remembered the previously seen stimuli in the long term (d' : nonoverlap = 0.58, $SE = \pm 0.08$; relevant overlap = 0.39, $SE = \pm 0.08$; irrelevant overlap = 0.35, $SE = \pm 0.06$). The recognition strength reported by the participants (indexed by confidence ratings 1 through 4) revealed that relevant stimuli from both nonoverlapped and overlapped tasks were rated significantly higher than irrelevant stimuli from overlapped tasks ($p < .05$ and $p < .05$, respectively) and stimuli from the

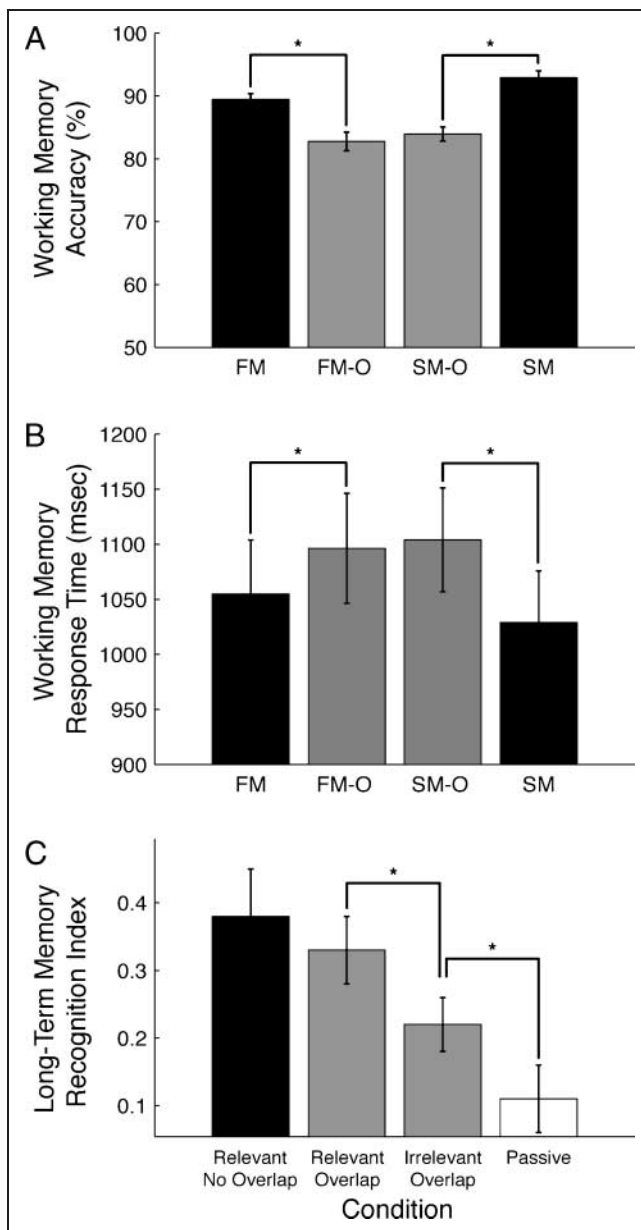


Figure 2. Behavioral results. (A) WM accuracy. Tasks utilizing overlapped images showed significantly reduced WM recognition accuracy. (B) WM RT. Overlapped tasks showed significant increases in RT relative to the nonoverlapped task counterparts. (C) Long-term memory recognition index. A postexperiment recognition test revealed significantly better recognition of relevant images in the overlap tasks (faces in FM-O and scenes in SM-O) than irrelevant images from the overlap tasks (faces from SM-O and scenes from FM-O), as well as images from the passive view task (PV-O). Error bars represent standard error of the mean. Asterisks denote significant differences ($p < .05$). FM-O = face memory-overlap; SM-O = scene memory-overlap; FM = face memory; SM = scene memory.

passive view task ($p < .01$ and $p < .01$ respectively) (Figure 2C). These data confirm that participants were performing the experiment as instructed, such that they were selectively directing their attention to the relevant stimuli and ignoring the irrelevant stimuli.

EEG Results

P100 Component

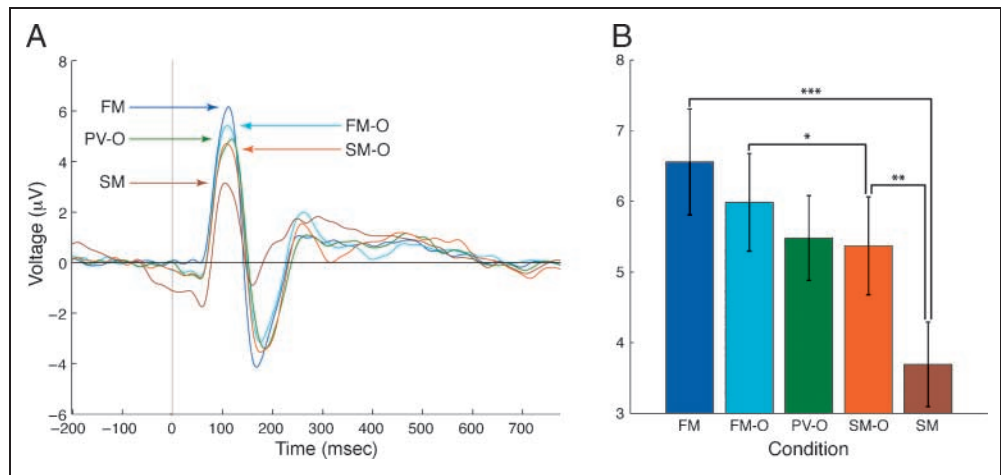
P100 peak latency and amplitude from posterior EOIs were subjected to separate 2×2 ANOVA with the type of stimulus attended (face vs. scene) and overlap status (overlapped vs. nonoverlapped) as factors. P100 measures of peak latency were not significantly different between stimulus type or overlap [ANOVA: main effect of stimulus, $F(1, 18) = 0.86, p = .34$; overlap, $F(1, 18) = 0.21, p = .66$; mean latency across participants: FM, 110 msec; FM-O, 113 msec; PV-O, 113 msec; SM-O, 114 msec; SM, 115 msec; all $p > .17$ for all two-tailed comparisons]. However, measures of P100 amplitude showed significant differences [main effect of overlap: $F(1, 18) = 11.36, p < .005$; main effect of stimulus type: $F(1, 18) = 32.28, p < .0001$; and an interaction between overlap and stimulus type: $F(2, 18) = 16.06, p < .001$]. Post hoc comparisons revealed that the amplitude of the P100 was significantly greater for the FM task than for the SM task (FM vs. SM, $p < .0001$; all participants exhibited greater P100 amplitude in FM vs. SM) (Figure 3A and B), revealing a differential response in the P100 component for faces compared to scenes, as reported by others (i.e., bottom-up effect) (Herrmann et al., 2005; Liu et al., 2002). Importantly, we report that for spatially overlapped images of faces and scenes with equivalent bottom-up information, attention to one stimulus while ignoring the other resulted in significant attentional modulation at this early time point in visual processing (i.e., top-down effect) (FM-O vs. SM-O, $p < .01$; 15 of 19 participants exhibited greater P100 amplitude in FM-O vs. SM-O) (Figure 3A and B). The P100 component of the FM-O task was significantly different from the SM-O task at 97–129 msec (paired two-tailed t tests across time points, $p < .05$). P100 amplitude in the FM-O task was closer to that of the FM task, whereas the P100 amplitude in the SM-O task was closer to that of the SM task (FM vs. FM-O, $p = .10$; SM-O vs. SM, $p < .01$). Although P100 amplitude in the passive view task (PV-O) was between FM-O and SM-O, it was not significantly different from either overlap task (PV-O vs. FM-O, $p = .11$; PV vs. SM-O, $p = .72$).

Topography maps of the P100 difference between pairs of tasks are shown in Figure 4. The lateralized posterior topography of the nonoverlapped face and scene difference (FM vs. SM: bottom-up contrast) is comparable to the overlapped face and scene difference (FM-O vs. SM-O: top-down contrast), revealing that top-down modulation occurs in approximately the same visual cortical regions that distinguish the stimuli based on bottom-up stimulus-driven differences.

N170 Component

An ANOVA showed a significant effect of overlap and stimulus type for N170 latency [main effect of stimulus type: $F(1, 18) = 10.93, p < .005$; main effect of overlap: $F(1, 18) = 21.97, p < .0005$]. Post hoc t tests revealed

Figure 3. Top-down modulation of the P1 component. (A) Grand-average waveform of P1 EOIs ($n = 19$). (B) P100 peak amplitudes ($n = 19$). All peak amplitudes of memory tasks show significant differences across tasks (PV-O is not significantly different than FM-O or SM-O). Error bars represent standard error of the mean. Asterisks denote significant difference (single, $p < .05$; double, $p < .01$; triple, $p < .0001$). FM-O = face memory-overlap; SM-O = scene memory-overlap; PV-O = passive view-overlap; FM = face memory; SM = scene memory.



that the mean N170 latencies significantly differ between isolated faces and scenes (FM, 174 msec vs. SM, 157 msec, $p < .01$), but were not significantly different for overlapped tasks (FM-O, 184 msec vs. SM-O, 176 msec, $p = .13$). The N170 peaked significantly later in the presence of distraction (FM-O later than FM, $p < .01$; SM-O later than SM, $p < .01$). However, there was no interaction between stimulus type and overlap [$F(1, 1) = 0.73, p = .40$].

Analysis of N170 amplitude reveals the classic finding of face selectivity (N170 face-selective effect; Bentin et al., 1996), with an ANOVA across tasks showing a main effect of stimulus type [$F(1, 18) = 13.9, p < .005$], and post hoc t tests revealing a significantly more negative N170 component for isolated faces than scenes (FM vs. SM, $p < .01$). However, the N170 amplitude was not modulated by top-down attention in this experiment; that is, N170 amplitudes in the overlapped tasks were not significantly different from each other [main effect of overlap: $F(1, 18) = 0.31, p = .58$; FM-O vs. SM-O, $p = .81$]. N170 amplitude in the PV-O task was not significantly different from the other overlap tasks (vs. FM-O, $p = .58$; vs. SM-O, $p = .23$).

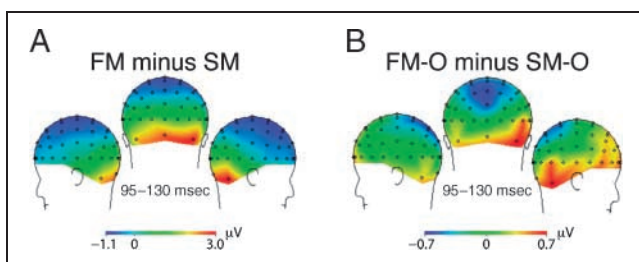


Figure 4. Topographic ERP difference maps at 95–130 msec (P100 component). (A) The lateralized posterior scalp topography of the nonoverlapped face and scene difference (FM minus SM: bottom-up contrast) is comparable to (B) the topography of the overlapped face and scene difference (FM-O minus SM-O: top-down contrast).

Neural-Behavioral Correlations

We report a significant across-participant correlation between the P100 modulation index and WM accuracy index ($r = .45, p < .05$; Figure 5); that is, the degree to which a participant selectively modulates activity in the first 100 msec of encoding a stimulus is a significant predictor of their ability to accurately recognize the stimulus after a 4-sec delay. We found this critical correlation by

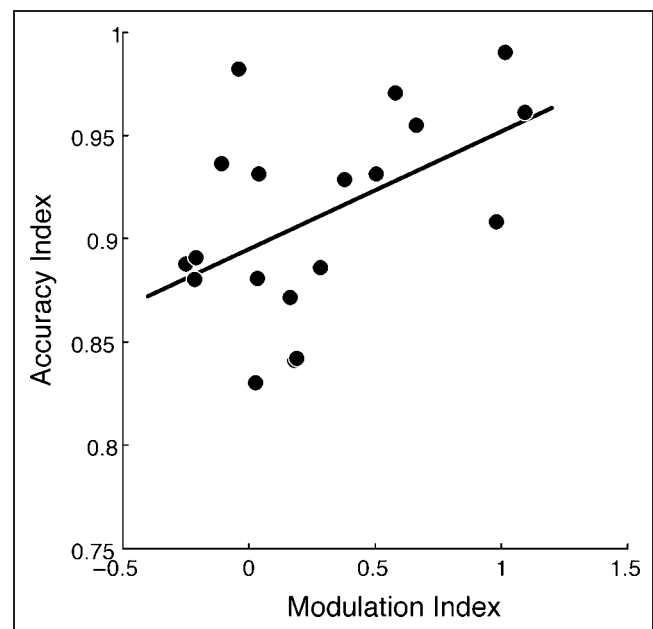


Figure 5. Neural-behavioral correlation. Measures of attentional modulation (P1 modulation index) significantly correlate with WM recognition (accuracy index). Subjects with greater attentional modulation of P100 amplitude (~100 msec poststimulus presentation) show greater ability to subsequently remember encoded stimuli after a delay period of WM maintenance (4 sec poststimulus presentation), $r = .45, p < .05$.

developing indices that allowed for a comparison of neural activity and behavior. First, to generate an index of top-down modulation of the P100 amplitude, we computed a P100 modulation index for each participant as the difference in P100 amplitude in the overlap tasks, corrected by the difference in P100 amplitude in non-overlap tasks:

$$\text{Modulation index} = \frac{\text{FM-O}_{\text{amp}} - \text{SM-O}_{\text{amp}}}{\text{FM}_{\text{amp}} - \text{SM}_{\text{amp}}}$$

This index allowed us to normalize for individual differences in bottom-up sensory processing. Second, to generate an index of WM recognition performance, we computed a WM accuracy index for each participant, composed of the participant's average accuracy in overlap tasks, corrected by their average accuracy in nonoverlap tasks:

$$\text{Accuracy index} = \frac{(\text{FM-O}_{\text{ACC}} + \text{SM-O}_{\text{ACC}})/2}{(\text{FM}_{\text{ACC}} + \text{SM}_{\text{ACC}})/2}$$

This index allowed us to normalize for individual differences in WM abilities. There was no comparable finding of significant correlation between P100 modulation and long-term memory measures, perhaps due to sparse sampling of long-term memory measures.

Eye-movement Control

To investigate the possibility that a condition-dependent shift in eye position either before or within 100 msec after stimulus presentation may have resulted in the reported P1 effect (as opposed to covert selective attention), we performed an additional experiment with eye tracking alone under identical conditions and instructions to the EEG experiment. Analysis revealed that there were no condition-specific differences in eye position at any time point. Furthermore, the median eye position prior to stimulus onset (−200 to 0 msec) and immediately after stimulus onset (0 to 100 msec) showed no dependence on condition in the vertical or horizontal directions [two-way repeated measures ANOVA—*vertical-pre*: $F(3, 4) = 1.35, p = .26$; *vertical-post*: $F(3, 4) = 2.17, p = .09$; *horizontal-pre*: $F(3, 4) = 2.08, p = .11$; *horizontal-post*: $F(3, 4) = 0.09, p = .96$; post hoc *t* tests—*vertical-pre*: FM vs. SM, $p = .85$; FM-O vs. SM-O, $p = .59$; *vertical-post*: FM vs. SM, $p = .28$; FM-O vs. SM-O, $p = .49$; *horizontal-pre*: FM vs. SM, $p = .41$; FM-O vs. SM-O, $p = .79$; *horizontal-post*: FM vs. SM, $p = .99$; FM-O vs. SM-O, $p = .97$]. In addition, measures of WM accuracy for each participant in the eye-tracking experiment were within 2 standard deviations of the mean WM accuracy measures for participants in the main experiment.

Although this experiment cannot definitively demonstrate that eye position was not an influence on the reported P1 effect and behavioral correlation (because

eye-tracking data were not obtained for the EEG sessions), these results reveal that participants do not seem to rely on a consistent and differential shift in eye gaze to perform the experiment. Furthermore, reports from participants in the EEG experiment do not suggest that a strategy of fixating their eyes at a particular location was utilized (e.g., repositioning gaze above the center of the screen prior to stimulus onset to more easily detect featural information from the faces, such as the eyes).

DISCUSSION

This study investigated top-down modulation of early visual processing and the influence of such modulation on subsequent WM recognition performance. We capitalized on the presence of well-described EEG signal differences associated with bottom-up processing of isolated face and scene stimuli (Herrmann et al., 2005; Liu et al., 2002; Bentin et al., 1996) to explore attentional influences on sensory cortical processing in the context of interfering information (i.e., overlapped stimuli). By maintaining bottom-up, sensory information constant and manipulating task goals, we were able to isolate the influence of top-down modulation on visual processing. We found that significant modulation of visual cortical activity begins as early as 97 msec after stimulus presentation (P100 component). Importantly, we found that at this early time point the extent to which participants selectively modulate neural representations of task-relevant information, when distracted by irrelevant information, correlates with their ability to successfully recognize the relevant stimuli after a period of WM maintenance. This provides a direct correlative link between neural activity in early visual cortex during selective encoding and behavioral measures of WM performance.

Early Visual Cortex Modulation

Modulation of early ERP components have been well documented during covert spatial-based attention (Hillyard, Vogel, & Luck, 1998), and more recently in feature-based attention tasks (Schoenfeld et al., 2007). In contrast to spatial- and feature-based attention, object-based attention involves the integration of spatial and feature aspects of an object to yield a holistic representation. In the current study, the use of spatially superimposed faces and scenes minimizes spatial-based mechanisms (Furey et al., 2006; Yi & Chun, 2005; Serences et al., 2004; O'Craven et al., 1999), and the task goals of successfully recognizing the relevant object after a delay period reduces reliance solely on feature information. Although the task design in the current study minimizes both spatial- and feature-based attentional mechanisms, there may still be an influence of feature and spatial information during WM encoding. For example, a shift in covert spatial attention to an anticipated location, such as that containing salient facial features, may occur during or prior to the cue period, although

none of the participants reported relying on a consistent feature or spatial strategy. Moreover, the eye-tracking control experiment revealed that overt eye movements were not likely a confounding factor in the reported neural results.

We report significant modulation of the P100 component in a selective attention task for complex real-world objects. This finding is consistent with several previously published reports of object-based attention, but is at odds with others. Object-based studies using illusory surface paradigms have documented significant modulation of the P100 (Valdes-Sosa et al., 1998), and even the earlier C1 component (Khoe et al., 2005; Valdes-Sosa et al., 1998). However, some studies have found modulatory changes that begin slightly later in the time course of visual processing, at the N170 component, ~170 msec (Martinez, Ramanathan, Foxe, Javitt, & Hillyard, 2007; Martinez, Teder-Salejarvi, Vazquez, et al., 2006; He, Fan, Zhou, & Chen, 2004; Pinilla et al., 2001); these studies utilized either the discrimination of illusory surfaces defined by transparent motion or the detection of luminance/shape changes at one end of an object.

Also, the current findings are in contrast to the results of an MEG study that utilized similar stimuli (superimposed faces and houses), but in a 1-back repetition detection task. This study showed modulation only at later time points (>190 msec) (Furey et al., 2006). Our results may have revealed earlier modulation due to greater task demands imposed by a two-item delayed-recognition task; it has been shown that increasing task difficulty results in enhanced activity modulation (Spitzer, Desimone, & Moran, 1988).

It is important to note that unlike several other EEG studies that did not find P100 selectivity for faces, we observed a P100 amplitude preference for faces versus scenes both in the main experiment and in an independent localizer task where faces and scenes were presented in separate blocks. Although the current study and several others (Herrmann et al., 2005; Itier & Taylor, 2002; Linkenkaer-Hansen et al., 1998) have revealed P100 selectivity to faces, others that have used face stimuli have found the P100 to reflect more domain-general aspects of visual processing (Rossion, Joyce, Cottrell, & Tarr, 2003; Rossion et al., 1999). Although all P100 findings likely represent early visual processing, it is possible that our results and those of studies that did not reveal P100 face selectivity may not reflect exactly the same type of processing, potentially as a result of differences in task design. However, the current study was intended to capitalize on the observed face selectivity of the P100 in the functional localizer task only to serve as an early marker of attentional control processes.

In a recently published study, we utilized face and scene stimuli in a similar two-item delayed-recognition task, but instead of using simultaneously presented overlapped stimuli, the face and scene images were presented sequentially, without overlap (Gazzaley et al., 2005). Interestingly,

the study revealed significant N170 modulation, but not significant P100 amplitude modulation by attentional goals. However, we recently increased the number of research participants in the sequential design version of this task and revealed significant top-down modulation of the P100 amplitude for sequentially presented relevant versus irrelevant faces (Gazzaley et al., 2008), thus paralleling the current study findings of very early object-based modulation. Because it has been postulated that early bottom-up face processing is rapid and largely automatic (Heisz, Watter, & Shedden, 2006), it is especially significant that top-down modulation can occur at such an early phase in processing these stimuli.

Several studies have suggested that early face processing (P100/M100 component) is a reflection of face categorization/holistic perception (Itier & Taylor, 2004; Liu et al., 2002), whereas later processing (N170 component) reflects configural information of faces (Latinus & Taylor, 2006; Goffaux, Gauthier, & Rossion, 2003; Liu et al., 2002; Rossion et al., 2000). If so, it follows that P100 modulation observed in the overlap tasks might represent early successful categorization of a face as being distinct from a scene, perhaps based on low-level feature analysis (Latinus & Taylor, 2006). However, this raises the question as to why the N170 component was not modulated by attention in the current study (i.e., no significant difference between FM-O and SM-O). One potential reason is based on previous findings that configural face processing requires extraction of low spatial frequency (LSF) information (Goffaux, Hault, Michel, Vuong, & Rossion, 2005). In the current study, the application of a transparency filter and an overlapped image obscures LSF information, while largely preserving high spatial frequency (HSF) information. When Goffaux et al. (2003) applied a filter to face stimuli that eliminated LSF and retained HSF information, face-selective N170 perceptual effects were abolished. It is thus possible that the bottom-up perceptual modifications to the faces, introduced by our experimental design, resulted in less LSF information and interfered with top-down influences at this stage. In support of this notion, it has also been revealed that the projection of LSF information to prefrontal cortex influences top-down modulation of visual cortical areas at ~180 msec (Bar et al., 2006). This may thus explain why the present results differ from those previously reported using the sequential version of this paradigm, that is, the same task with preserved LSF information resulted in significant top-down modulation of the N170 (Gazzaley et al., 2005).

However, this explanation does not account for the fact that several studies in which LSF information was present have also not revealed N170 modulation as a function of attention (Carmel & Bentin, 2002; Cauquil, Edmonds, & Taylor, 2000). It is possible that N170 modulation was not observed in the current study and these other studies because the salience of face stimuli was already too high to benefit from additional perceptual modulation at this

stage of encoding. Indeed, it has been argued that relative to stimuli with high salience, stimuli with low salience are more likely to benefit from additional attentional modulation (Hawkins, Shafto, & Richardson, 1988).

In considering how activity modulation can occur so early in the processing of the overlapped visual stimuli (i.e., 100 msec after stimulus presentation), it is important to recognize that participants were cued to the relevant information, such that they were aware of the stimulus to be remembered prior to presentation. This aspect of the current study parallels that used in most spatial attention tasks, which also report modulation of the P100 amplitude. In other words, anticipatory gain modulation may preactivate sensory cortical areas to enhance the efficiency of subsequent sensory processing, as described by others (Kastner, Pinsk, De Weerd, Desimone, & Ungerleider, 1999; Luck et al., 1997).

Neural–Behavioral Correlation

It is well established that selective attention confers a behavioral performance advantage for a variety of perceptual tasks, such as visual detection (Posner, Snyder, & Davidson, 1980), discrimination (Carrasco & McElree, 2001), and categorization (Heekeren, Marrett, Bandettini, & Ungerleider, 2004). In a comparable manner, failure to selectively direct attentional resources negatively impacts memory performance in both young (Zanto & Gazzaley, 2009) and older adults (Gazzaley et al., 2008; Gazzaley, Cooney, Rissman, & D’Esposito, 2005). The behavioral advantage mediated by selective attention is presumably the result of reduced interference from irrelevant information in a system with limited capacity (Hasher, Lustig, & Zacks, 2008; Vogel, McCollough, & Machizawa, 2005), likely mediated via top–down control mechanisms originating from prefrontal cortex (for a review, see Gazzaley & D’Esposito, 2007). However, only recently have direct correlations between the magnitude of visual cortex activity modulation and behavioral measures of perceptual and memory performance been established (Gazzaley, Cooney, Rissman, et al., 2005; Vogel & Machizawa, 2004; Pessoa, Kastner, & Ungerleider, 2002; Rees, Friston, & Koch, 2000; Brewer, Zhao, Desmond, Glover, & Gabrieli, 1998).

By revealing a significant correlation between very early measures of visual cortex activity during selective stimulus encoding and subsequent WM recognition accuracy, our results contribute to a growing literature describing the relationship between visual activity modulation and behavioral performance. Specifically, the degree to which participants modulate the P100 amplitude in overlap tasks predicts their subsequent recognition accuracy. This finding suggests that robust and early modulation generates higher fidelity stimulus representations, which translates to improved maintenance of relevant information across a delay period, resulting in superior recognition ability.

Conclusion

Consistency of goal-directed activity modulation occurring so early in the processing of spatial-, feature- and object-based information suggests that domain-general mechanisms of top–down modulation are targeted on early cortical regions of the visual processing stream. The influence of such early top–down modulation of neural representations for real-world objects on WM recognition performance is consistent with a growing appreciation of the dynamic relationship of attention and WM (Awh & Jonides, 2001; de Fockert et al., 2001).

Acknowledgments

This work was supported by National Institutes of Health Grant K08-AG025221, R01-AG030395, and the American Federation of Aging Research (AFAR). We thank Nick Planet and Derek Wu for their assistance in EEG data acquisition and pre-processing.

Reprint requests should be sent to Adam Gazzaley, University of California, San Francisco, 600 16th Street, Genentech Hall, Room N472J, San Francisco, CA 94158, or via e-mail: adam.gazzaley@ucsf.edu.

REFERENCES

- Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*, 119–126.
- Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top–down facilitation of visual recognition. *Proceedings of the National Academy of Sciences, U.S.A.*, *103*, 449–454.
- Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience*, *8*, 551–565.
- Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1998). Making memories: Brain activity that predicts how well visual experience will be remembered. *Science*, *281*, 1185.
- Carmel, D., & Bentin, S. (2002). Domain specificity versus expertise: Factors influencing distinct processing of faces. *Cognition*, *83*, 1–29.
- Carrasco, M., & McElree, B. (2001). Covert attention accelerates the rate of visual information processing. *Proceedings of the National Academy of Sciences, U.S.A.*, *98*, 5363–5367.
- 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*, 2167–2171.
- de Fockert, J. W., Rees, G., Frith, C. D., & Lavie, N. (2001). The role of working memory in visual selective attention. *Science*, *291*, 1803–1806.
- Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences, U.S.A.*, *93*, 13494–13499.
- Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*, 193–222.
- Di Russo, F., Martínez, A., Sereno, M., Pitzalis, S., & Hillyard, S. (2002). Cortical sources of the early components of the visual evoked potential. *Human Brain Mapping*, *15*, 95–111.
- Furey, M. L., Tanskanen, T., Beauchamp, M. S., Avikainen, S., Uutela, K., Hari, R., et al. (2006). Dissociation of

- face-selective cortical responses by attention. *Proceedings of the National Academy of Sciences, U.S.A.*, *103*, 1065–1070.
- Gazzaley, A., Clapp, W., Kelley, J., McEvoy, K., Knight, R. T., & D'Esposito, M. (2008). Age-related top-down suppression deficit in the early stages of cortical visual memory processing. *Proceedings of the National Academy of Sciences, U.S.A.*, *105*, 13122–13126.
- Gazzaley, A., Cooney, J. W., McEvoy, K., Knight, R. T., & D'Esposito, M. (2005). Top-down enhancement and suppression of the magnitude and speed of neural activity. *Journal of Cognitive Neuroscience*, *17*, 507–517.
- Gazzaley, A., Cooney, J. W., Rissman, J., & D'Esposito, M. (2005). Top-down suppression deficit underlies working memory impairment in normal aging. *Nature Neuroscience*, *8*, 1298–1300.
- Gazzaley, A., & D'Esposito, M. (2007). Unifying prefrontal cortex function: Executive control, neural networks and top-down modulation. In J. Cummings & B. Miller (Eds.), *The human frontal lobes* (2nd ed., pp. 187–206). New York: The Guildford Press.
- Goffaux, V., Gauthier, I., & Rossion, B. (2003). Spatial scale contribution to early visual differences between face and object processing. *Cognitive Brain Research*, *16*, 416–424.
- Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception*, *34*, 77–86.
- Gomez Gonzalez, C. M., Clark, V. P., Fan, S., Luck, S. J., & Hillyard, S. A. (1994). Sources of attention-sensitive visual event-related potentials. *Brain Topography*, *7*, 41–51.
- Hasher, L., Lustig, C., & Zacks, J. M. (2008). Inhibitory mechanisms and the control of attention. In A. Conway, C. Jarrold, M. Kane, A. Miyake, & J. Towse (Eds.), *Variation in working memory* (pp. 227–249). New York: Oxford University Press.
- Hawkins, H. L., Shafto, M. G., & Richardson, K. (1988). Effects of target luminance and cue validity on the latency of visual detection. *Perception & Psychophysics*, *44*, 484–492.
- He, X., Fan, S., Zhou, K., & Chen, L. (2004). Cue validity and object-based attention. *Journal of Cognitive Neuroscience*, *16*, 1085–1097.
- Heekeren, H. R., Marrett, S., Bandettini, P. A., & Ungerleider, L. G. (2004). A general mechanism for perceptual decision-making in the human brain. *Nature*, *431*, 859–862.
- Heisz, J. J., Watter, S., & Shedden, J. M. (2006). Progressive N170 habituation to unattended repeated faces. *Vision Research*, *46*, 47–56.
- Herrmann, M. J., Ehlis, A. C., Ellgring, H., & Fallgatter, A. J. (2005). Early stages (P100) of face perception in humans as measured with event-related potentials (ERPs). *Journal of Neural Transmission*, *112*, 1073–1081.
- Hillyard, S. A., & Anllo-Vento, L. (1998). Event-related brain potentials in the study of visual selective attention. *Proceedings of the National Academy of Sciences, U.S.A.*, *95*, 781–787.
- Hillyard, S. A., Vogel, E. K., & Luck, S. J. (1998). Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences*, *353*, 1257–1270.
- Itier, R. J., & Taylor, M. J. (2002). Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: A repetition study using ERPs. *Neuroimage*, *15*, 353–372.
- Itier, R. J., & Taylor, M. J. (2004). N170 or N1? Spatiotemporal differences between object and face processing using ERPs. *Cerebral Cortex*, *14*, 132–142.
- Jha, A. P. (2002). Tracking the time-course of attentional involvement in spatial working memory: An event-related potential investigation. *Brain Research, Cognitive Brain Research*, *15*, 61–69.
- Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased activity in human visual cortex during directed attention in the absence of visual stimulation. *Neuron*, *22*, 751–761.
- Kastner, S., & Ungerleider, L. G. (2000). Mechanisms of visual attention in the human cortex. *Annual Review of Neuroscience*, *23*, 315–341.
- Khoe, W., Mitchell, J. F., Reynolds, J. H., & Hillyard, S. A. (2005). Exogenous attentional selection of transparent superimposed surfaces modulates early event-related potentials. *Vision Research*, *45*, 3004–3014.
- Latinus, M., & Taylor, M. J. (2006). Face processing stages: Impact of difficulty and the separation of effects. *Brain Research*, *1123*, 179–187.
- Linkenkaer-Hansen, K., Palva, J. M., Sams, M., Hietanen, J. K., Aronen, H. J., & Ilmoniemi, R. J. (1998). Face-selective processing in human extrastriate cortex around 120 ms after stimulus onset revealed by magneto- and electroencephalography. *Neuroscience Letters*, *253*, 147–150.
- Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience*, *5*, 910–916.
- López, M., Rodríguez, V., & Valdés-Sosa, M. (2004). Two-object attentional interference depends on attentional set. *International Journal of Psychophysiology*, *53*, 127–134.
- Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology*, *77*, 24–42.
- Martinez, A., Ramanathan, D. S., Foxe, J. J., Javitt, D. C., & Hillyard, S. A. (2007). The role of spatial attention in the selection of real and illusory objects. *Journal of Neuroscience*, *27*, 7963–7973.
- Martinez, A., Teder-Salejarvi, W., Vazquez, M., Molholm, S., Foxe, J. J., Javitt, D. C., et al. (2006). Objects are highlighted by spatial attention. *Journal of Cognitive Neuroscience*, *18*, 298–310.
- O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, *401*, 584–587.
- Pessoa, L., Kastner, S., & Ungerleider, L. G. (2002). Attentional control of the processing of neural and emotional stimuli. *Brain Research, Cognitive Brain Research*, *15*, 31–45.
- Pinilla, T., Cobo, A., Torres, K., & Valdés-Sosa, M. (2001). Attentional shifts between surfaces: Effects on detection and early brain potentials. *Vision Research*, *41*, 1619–1630.
- Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology*, *109*, 160–174.
- Rees, G., Friston, K., & Koch, C. (2000). A direct quantitative relationship between the functional properties of human and macaque V5. *Nature Neuroscience*, *3*, 716–723.
- Rossion, B., Delvenne, J. F., Debatisse, D., Goffaux, V., Bruyer, R., Crommelinck, M., et al. (1999). Spatio-temporal localization of the face inversion effect: An event-related potentials study. *Biological Psychology*, *50*, 173–189.
- Rossion, B., Gauthier, I., Tarr, M. J., Despland, P., Bruyer, R., Linotte, S., et al. (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*, 69–74.
- Rossion, B., Joyce, C. A., Cottrell, G. W., & Tarr, M. J. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *Neuroimage*, *20*, 1609–1624.
- Schoenfeld, M. A., Hopf, J. M., Martinez, A., & Mai, H. M. (2007). Spatio-temporal analysis of feature-based attention. *Cerebral Cortex*, *17*, 2468–2477.
- Serences, J. T., Schwarzbach, J., Courtney, S. M., Golay, X., & Yantis, S. (2004). Control of object-based attention in human cortex. *Cerebral Cortex*, *14*, 1346–1357.
- Spitzer, H., Desimone, R., & Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science*, *240*, 338–340.
- Sreenivasan, K. K., & Jha, A. P. (2007). Selective attention supports working memory maintenance by modulating perceptual processing of distractors. *Journal of Cognitive Neuroscience*, *19*, 32–41.
- Sreenivasan, K. K., Katz, J., & Jha, A. P. (2007). Temporal characteristics of top-down modulations during working memory maintenance: An event-related potential study of the N170 component. *Journal of Cognitive Neuroscience*, *19*, 1836–1844.
- Talsma, D., Mulckhuyse, M., Slagter, H. A., & Theeuwes, J. (2007). Faster, more intense! The relation between electrophysiological reflections of attentional orienting, sensory gain control, and speed of responding. *Brain Research*, *1178*, 92–105.
- Thut, G., Nietzel, A., Brandt, S. A., & Pascual-Leone, A. (2006). Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *Journal of Neuroscience*, *26*, 9494–9502.
- Valdes-Sosa, M., Bobes, M. A., Rodriguez, V., & Pinilla, T. (1998). Switching attention without shifting the spotlight object-based attentional modulation of brain potentials. *Journal of Cognitive Neuroscience*, *10*, 137–151.
- Vogel, E. K., & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, *428*, 748–751.
- Vogel, E. K., McCollough, A. W., & Machizawa, M. G. (2005). Neural measures reveal individual differences in controlling access to working memory. *Nature*, *438*, 500–503.
- Yi, D. J., & Chun, M. M. (2005). Attentional modulation of learning-related repetition attenuation effects in human parahippocampal cortex. *Journal of Neuroscience*, *25*, 3593–3600.
- Zanto, T. P., & Gazzaley, A. (2009). Neural suppression of irrelevant information underlies optimal working memory performance. *Journal of Neuroscience*, *29*, 3059–3066.