Attentional Modulation of Object Representations in Working Memory

Jöran Lepsien and Anna C. Nobre

Brain & Cognition Laboratory, Department of Experimental Psychology, University of Oxford, Oxford OX1 3UD, UK

We investigated the role of object-based attention in modulating the maintenance of faces and scenes held online in working memory (WM). Participants had to remember a face and a scene, while cues presented during the delay instructed them to orient their attention to one or the other item. Event-related functional magnetic resonance imaging revealed that orienting attention in WM modulated the activity in fusiform and parahippocampal gyri, involved in maintaining representations of faces and scenes respectively. Measures from complementary behavioral studies indicated that this increase in activity corresponded to improved WM performance. The results show that directed attention can modulate maintenance of specific representations in WM, and help define the interplay between the domains of attention and WM.

Keywords: attention, fMRI, maintenance, mental representations, working memory

Introduction

Current models of working memory (WM) differ in their views of the role of posterior cortices in maintaining representations to guide future action. Whereas some researchers emphasize the role of prefrontal areas in maintaining task-relevant information (Goldman-Rakic 1987; Smith and Jonides 1999; Courtney 2004), others stress the sustained activation of representations in posterior cortices under the control of prefrontal areas (D’Esposito et al. 2000; Fuster 2000; Petrides 2000; Passingham and Sakai 2004). Prefrontal areas are thought to support maintenance, monitoring, and/or manipulation of information by biasing representations in posterior brain regions in a top-down fashion (Desimone and Duncan 1995; Awh and Jonides 2001; Curtis and D’Esposito 2003; Ranganath and D’Esposito 2005; Postle 2006).

This view of WM is closely related to views of attention in the human brain (Desimone and Duncan 1995; Kastner and Ungerleider 2001). Numerous studies have shown enhanced cortical responses in brain areas representing attended locations, features or objects, including both increases in background neural activity and stimulus-evoked responses (Corbetta et al. 1991; Chelazzi et al. 1993; Heinze et al. 1994; Luck et al. 1997; Chawla et al. 1999; Kastner et al. 1999; O’Craven et al. 1999). Similarly, during WM delays, attention could be directed toward relevant information, selecting and enhancing internal representations, so that the information survives passive decay and distraction.

Studies have supported the participation of posterior areas in WM representations showing persistent activity in inferior temporal areas during maintenance of visual objects (e.g., Druzgal and D’Esposito 2001, 2003; Postle et al. 2003; Ranganath, Cohen, et al. 2004; Ranganath, DeGutis, et al. 2004). Furthermore, these representations can be flexibly controlled. Using a paired-associate task, Ranganath, Cohen, et al. (2004) showed that activity in posterior occipital-temporal areas that are sensitive to stimulus categories was modulated according to the category of the stimulus that had to be recalled and maintained for the task goals. Behaviorally, it has also been shown that visual spatial WM facilitates behavioral performance to probes appearing at relevant locations, and enhances activity in visual cortex elicited by task-irrelevant probes within memorized locations, consistent with the idea that spatial attention can support rehearsal within spatial WM (Awh et al. 1998, 1999; Postle 2006).

However, the ability of attention to bias WM representations dynamically according to changing task goals remains, as yet, untested. The possibility that attention can modulate maintenance-related activity in WM has recently been raised by demonstrations that attention can be voluntarily oriented to locations of representations held in WM, in an analogous way to orienting attention in the perceptual domain (Griffin and Nobre 2003; Nobre et al. 2004; Lepsien et al. 2005). The present experiments specifically tested the ability of attention to bias activity in functionally specialized posterior brain areas during WM maintenance. Retro-cues presented during the delay oriented attention selectively to a specific object held online in WM (Fig. 1). Unlike previous studies on WM, the items to be probed changed during the course of the trial, allowing us to investigate the effect of attentional deployment on maintenance-related activity. The neural bases of the effects were measured using event-related functional magnetic resonance imaging (fMRI). Attentional biasing of object-based representations was expected to result in relative increases in activity of brain areas coding attended versus ignored objects. On the behavioral level, advantages in retrieval for cued objects were predicted, as already demonstrated for attentional orienting toward locations in WM (Griffin and Nobre 2003).

Materials and Methods

Participants

Fourteen healthy volunteers participated in each of 2 behavioral experiments (Experiment 1: aged 21–51, 5 males; Experiment 2: aged 21–31, 5 males). Another 17 (aged 22–32, 6 males) performed the main experiment while being scanned with fMRI. Participants gave written informed consents. The study protocol was approved by the local ethics committee.

Stimuli

Pictures of faces and scenes were chosen as stimuli because of the many demonstrations that they differentially engage processing in the posterior fusiform gyrus (FG) (Puce et al. 1995; Kanwisher et al. 1997) and parahippocampal gyrus (PHG) (Aguirre et al. 1998; Epstein and Kanwisher 1998). It should be noted that both areas may also respond
to other visual stimuli (Gauthier et al. 1999; Haxby et al. 2001) but show a relative selectivity for faces and scenes. Importantly, the FG and PHG exhibit opposite response properties, that is, the PHG responds only weakly to faces and vice versa. These areas also show sustained and load-dependent activity during the delay in WM tasks (Druzgal and D’Esposito 2001, 2003; Postle et al. 2003; Sala et al. 2003; Ranganath, Cohen, et al. 2004; Ranganath, DeGutis, et al. 2004; Rama and Courtney 2005).

In total, 140 unique faces and scenes were used. Face stimuli were selected from those developed by the Max-Planck-Institute for Biological Cybernetics (Tuebingen, Germany) (Troje and Bulthoff 1996). Only front-view images were used, which were converted to gray scale. In addition, the ears and neck were removed from every face using a graphics program for the second behavioral and the fMRI experiment. Each face was then centered on a gray box (450 $\times$ 337 pixels), which matched the size of the scene stimuli. Scene stimuli were photographs either taken by the experimenters or obtained from the Internet. They depicted outdoor views of different types of landscapes with a clear spatial layout. Scenes comprised views of mountain landscapes, forests, deserts, fields, rivers/lakes, or coasts/beaches. As for face stimuli, all images were converted to gray scale, and scaled to 450 $\times$ 337 pixels.

Object-Based Retro-cueing Tasks
In order to investigate whether the neural representation of a face or a scene held online in WM was altered by object-based attentional orienting, a double retro-cueing procedure was used. The tasks shared commonalities with previous retro-cueing tasks (Griffin and Nobre 2003; Nobre et al. 2004; Lepsien et al. 2005). Three complementary experiments were conducted to investigate the behavioral effects of object-based orienting in WM (Experiments 1 and 2) and its neural consequences and control (fMRI experiment). All tasks were programmed and presented using the Presentation software package (Neurobehavioral Systems, Albany, CA).

The basic task design is illustrated in Figure 1. A trial started with the presentation of a fixation cross for 500 ms. This cross was slightly bigger than all other fixation crosses used, and served as a warning signal indicating the start of the trial. Subsequently, an array of 2 stimuli was presented. Each stimulus was presented for 1 s, separated by a 1-s interstimulus interval. After a randomized and variable delay (delay-1, 4-9.5 s), a series of 2 cues appeared (each for 100-ms duration in Experiment 1; 250-ms duration in Experiment 2 and FMRI experiment), separated by a fixed delay (delay-2) (5-s duration in Experiments 1 and 2, 4.75-s duration in FMRI experiment). The first cue (cue-1) oriented the participant’s attention to one of the 2 stimuli encoded in the initial array. The second cue (cue-2) instructed the participant either to shift attention from the currently attended object to the other object (switch cue: figure 8 symbol, rotated by 90°) or to maintain the currently attended object (stay cue: figure 8 symbol). A long and variable delay (delay-3, 4-9.5 s) followed the switch/stay cue, after which a probe stimulus occurred (500-ms duration) to which the participant had to make a response. The interval between trials was fixed at 2 s (Experiment 1) or 1.75 s (Experiment 2 and FMRI experiment). Each experiment was composed of 120 trials.

In all experiments, participants performed a variant of the delayed-match-to-sample task but the nature of the decision required or the stimulus types encoded differed slightly, as explained for each experiment.
below. They indicated their answer with one of 2 buttons (right hand: index finger: yes, middle finger: no).

**Behavioral Experiments**

Two behavioral experiments were performed outside the fMRI scanner to investigate whether participants were able to follow the double-cues accurately, and to optimize information processing based on their predictive information (Griffin and Nobre 2003; Nobre et al. 2004; Lepsien et al. 2005; see also Posner 1980). In both behavioral experiments, participants were asked to decide if the probe stimulus matched one of the stimuli presented during the beginning of the trial and to make a speeded-choice response accordingly. To test for behavioral benefits and costs of object-based orienting in WM, valid (50%), invalid (30%), and nonmatch trials (20%) were introduced. In valid trials, the combination of first and second cue correctly predicted which stimulus would be probed at the end of the trial. In invalid trials, the uncued stimulus appeared as the probe. In nonmatch trials, a new stimulus appeared as the probe, which was from the same category as the cued item. Participants completed the 120 trials presented in one block, and had the opportunity to pause after every 20 trials.

**Experiment 1**

In the first behavioral experiment, face and scene stimuli were always presented in the initial array, in an unpredictable and balanced order. The first cue oriented attention to either the face stimulus ("F") or the scene stimulus ("S"), and the second cue instructed participants to shift or maintain the focus of attention. This led to 4 combinations of retro-cues: face/switch, face/stay, scene/switch, and scene/stay; each repeated 30 times. There were 15 valid, 9 invalid, and 6 nonmatch trials for each combination of retro-cues.

To equate for task difficulty in responding to face and scene stimuli, similar types of scenes were used for the nonmatch probes as in the initial array, drawing from within the same category of place (mountain, field, coast/beach, forest, river/lake, or desert). This compensated for the fact that face stimuli were more similar in general to one another than scene stimuli.

**Experiment 2**

In the second behavioral experiment, each trial contained stimuli from only one of the 2 categories (faces or scenes). This ensured that any benefits or costs of object-based attentional orienting were specific to the objects maintained in WM, and did not result simply from more general response biases toward a particular stimulus category. In Experiment 1, cueing effects could, in principle, result from participants expecting to respond to stimuli from a particular category, and being disproportionately slowed to respond to stimuli for the other category (in invalid trials).

In order to distinguish unequivocally between the 2 stimuli from the same category for subsequent retro-cueing, one of the stimuli was tinted red and the other green. Red and green stimuli appeared in unpredictable and balanced order. Accordingly, the first cue instructed the participants to orient their attention to the red or the green stimulus (red or green figure 8 symbol). The second cue was a switch or stay cue, identical to the one used in Experiment 1. Finally, a black-and-white probe was presented, and participants decided whether it had appeared during the initial array. As for Experiment 1, nonmatch scene stimuli were drawn from the same category of place.

Trials with only-face stimuli and trials with only-scene stimuli were randomly intermixed. As in Experiment 1, participants completed valid (15), invalid (9), or nonmatch (6) trials in switch and stay conditions using face-only or scene-only stimuli.

**fMRI Experiment**

The purpose of the fMRI experiment was to investigate how neural activity in posterior brain areas involved in encoding and maintaining specific types of stimuli in WM was modulated by object-based attentional orienting. The task was therefore designed to maximize the deployment of attention triggered by the double retro-cues, and to enable measurement of brain activity in posterior areas during the retro-cueing period in a way that was uncontaminated by any activity related to the preceding encoding array of subsequent probe stimuli.

In the fMRI experiment, the decision required by participants differed from that of the initial behavioral experiments by forcing the obligatory use of the cueing stimuli. Imperative cues were used, with 100% validity. Participants had to decide whether the probe stimulus matched the cued item. The rationale for using imperative cues was to maximize the use of and reliance upon the cues, and therefore to optimize the magnitude of the putative neural effects for fMRI measures. However, by using this design we sacrificed the ability to measure simultaneously the behavioral advantages associated with the cues. The stimulus presentation parameters were similar to those in Experiment 1. The initial array always contained a face and a scene stimulus, presented in unpredictable and balanced order. The first cue instructed participants to orient attention to either the face stimulus ("F") or to the scene stimulus ("S"), and the second cue instructed them to shift or maintain the focus of attention.

There were equal numbers of match and nonmatch trials (50% each). In nonmatch trials, the probe was either a new item from the same category (66.7%) or the other (uncued) item from the initially presented stimuli (33.3%). This manipulation ensured that the participants could not base their decision on familiarity alone but had to follow both cues in order to know which item was cued. Participants were instructed to emphasize accuracy.

There were 30 trials for each of the 4 combinations of retro-cues—face/switch, face/stay, scene/switch, and scene/stay (15 match and 15 nonmatch). In addition, 30 null events were introduced, during which a fixation cross was presented between 6 and 17 s. The durations of the null events followed a logarithmic distribution (50%: 6-9 s, 33.3%: 10-13 s, 16.7%: 14-17 s; mean = 10.23 s). The 4 experimental trial types and the null events were presented in a constrained randomized and unpredictable order in a single block. The randomization procedure ensured that there was approximately the same number of transitions of the different levels of the following factors between successive trial types: first stimulus, second stimulus, cue-1, cue-2, response, condition (cue-1 \* cue-2), condition \* response, condition \* null events. Importantly, this meant that the different types of encoding arrays leading up to the double cues as well as the types of probes following the retro-cues were fully equated and did not contribute to any putative modulation of brain activity in posterior areas by the retro-cues.

Stimulus durations and delays were also optimized for the fMRI procedure. The double-cue period was separated from the preceding stimulus array (delay-1) and the subsequent probe stimulus (delay-3) by long and variable delays (4-9.5 s), whose duration varied pseudorandomly in a logarithmic fashion (50%: 4-5.5 s, 33.3%: 6-7.5 s, 16.7%: 8-9.5 s; mean = 6.06 s). This kind of distribution has several advantages: the skewing of the intervals toward the shorter duration helps keep the task to an endurable length; temporal expectations about the presentation of the next stimuli are kept constant; and the length and variability of the delay enable good separation of hemodynamic responses to individual events within trials (Friston et al. 1998; cf. Nobre et al. 2004). In contrast, the separation between the 2 cues (delay-2) was kept constant in order to allow us to plot the activation within the relevant posterior areas as continuous time-series and to compare these directly across the 4 experimental conditions. In total, the length of a trial varied between 19 and 30 s, with a mean of 23.13 s.

In addition, the onset asynchrony between successive stimuli was always a multiple of 500 ms, and stimulus onsets varied systematically in relation to the beginning of the image acquisition (see fMRI methods below). Using a constrained randomization procedure, an approximately equal distribution of all events of interest over the whole time range of a TR was achieved (oversampling). This resulted in a "virtual resolution" for data sampling of 500 ms.

**Localizer Task**

A separate localizer experiment was also conducted during the fMRI session to define functionally the region in posterior FG responsive to faces and the region along the PHG responsive to places. Participants viewed pairs of faces, scenes, or checkerboard-like grid-stimuli, and indicated whether 2 stimuli in a pair matched with a speeded-choice response (right hand; index finger: yes, middle finger: no). The stimulus types were presented in a blocked fashion. Stimuli were presented for 500 ms, separated by 2 s, and followed by an interval of 1 s to give the


Cerebral Cortex
Page 3 of 12
response. Each block contained 4 trials, leading to a block length of 16 s. In addition, a baseline period, in which just a fixation cross was presented for 16 s, was introduced as a fourth condition. The 3 stimulus blocks were always presented in a randomized order, and separated by a fixation period, and were then followed by one baseline block. This sequence was repeated 6 times. In total, the localizer experiment lasted 7.8 min.

**Image Acquisition**

Subjects were equipped with mirror glasses (and corrective lenses, if necessary), allowing them to view a screen mounted in front of the scanner. Stimuli were projected onto this screen by a projector placed outside the scanning room. Subjects gave responses using a custom-made MRI-compatible button box.

MRIs were acquired using a Varian-Inova 3-T scanner at the Oxford Centre for Functional Magnetic Resonance Imaging of the Brain. Functional images were obtained with a single-shot T2*-weighted echo-planar imaging (EPI) sequence (time echo [TE] = 30 ms, flip angle = 90°, time repetition [TR] = 3 s). A Magnex head-dedicated gradient insert coil was used in conjunction with a birdcage head radiofrequency coil tuned to 127.4 MHz. Twenty-four slices with a thickness of 5 mm covered the entire cortex and parts of the cerebellum (64 × 64 matrix with a field of view of 192 × 192 mm, resulting in a notional voxel size of 3 × 3 × 5 mm). An automated shimming algorithm was used to reduce inhomogeneities in the magnetic field (Wilson et al. 2002). The main experiment was performed in one run consisting of 1037 images. The localizer task was performed in a separate run consisting of 156 images. The first 4 images of each run contained no experimental manipulation, in order to allow the signal intensities to saturate, and were subsequently deleted.

A structural image of each participant was acquired at the end of the session, using a high-resolution T1-weighted sequence (3D Turbo fast, low-angle shot sequence; TR = 15 ms; TE = 6.9 ms; 1 × 1 mm in-plane resolution and 1.5-mm slice thickness).

**Image Processing and Analysis**

Data were analyzed using statistical parametric mapping (SPM2, Wellcome Department of Cognitive Neurology, London, UK). Data from the main experiment and from the localizer task were processed separately. Images were corrected for slice timing, and subsequently realigned and unwarped to correct for movement artifacts (Andersson et al. 2001). High-resolution anatomical T1 images were first skull-stripped using the Brain Extraction Tool, part of the FMRI Software Library (Smith 2002), and then coregistered with the realigned functional images (Friston et al. 1995). Structural and functional images were spatially normalized into a standardized anatomical framework (Montreal Neurological Institute [MNI] space; Collins et al. 1994) using the T1 template provided by the Medical Research Council—Cognition and Brain Sciences Unit in Cambridge, UK (http://www.mrc-cbu.cam.ac.uk/Imaging/Common/no_skull_norm.shtml). The normalization procedure failed for 3 data sets. In these cases, the images were normalized using the default EPI template provided in SPM2. Functional images were spatially smoothed using a Gaussian kernel with 8 mm3 full width at half maximum, to account for anatomical variability between subjects and to conform the data to a Gaussian model (Hopfinger et al. 2000). In addition the data were temporally filtered to eliminate slow signal drifts (high-pass filter: 256 s).

Data from individual subjects were analyzed using a General Linear Model with correction for serial autocorrelations (using the AR(1) model). The model included explanatory variables for all phases of a trial (see Fig. 1), which were convolved with the canonical hemodynamic response function (HRF) provided within SPM2. Specifically, the onsets for the following events were entered into the model: warning signal, stimulus-1, stimulus-2, delay-1, cue-1, delay-2, cue-2, delay-3, probe, and error trials. Two types of stimulus-1 and stimulus-2 were modeled, depending on whether they were face or scene stimuli. Two types of cue-1 and delay-2 were modeled, depending on the object instructed by the cue (face or scene). Four types of cue-2 and delay-3 were modeled, depending on the nature of each cue: face/switch, face/stay, scene/switch, and scene/stay. Error trials included trials with incorrect, late, or no response.
were observed, suggesting that RTs and cueing effects were similar for the different types of objects cued, and for switch and stay cues. The mean overall accuracy in Experiment 1 was very high (94.7%), and did not differ as a function of validity, cue-1, or cue-2 (all $P > 0.399$). The absence of main effects or interactions between the factors cue-1 and cue-2 indicates that task difficulty was well matched between the different types of cue combinations used.

RTs and accuracies during the match trials in Experiment 2 were compared with repeated-measures ANOVAs including the factors of trial validity (valid, invalid), stimulus type in the trial (face, scene), and cue-2 instruction (switch, stay). Analysis of mean correct RTs revealed a significant main effect of validity ($F_{1,10} = 5.04, P = 0.049$), again indicating significantly faster responses during valid trials than during invalid trials. In addition, a trend toward a main effect of stimulus type was observed, reflecting slower RTs in face-trials than in scene-trials ($F_{1,10} = 4.62, P = 0.057$). Neither the main effect of cue-2 nor any interactions reached significance (all $P > 0.286$). The mean overall accuracy was 85.45%. The accuracy analysis revealed significant main effects of validity ($F_{1,10} = 7.73, P = 0.019$) and cue-2 ($F_{1,10} = 6.12, P = 0.033$), as well as a trend toward an effect of stimulus type ($F_{1,10} = 4.44, P = 0.061$). Accuracies were higher during valid trials as compared with invalid trials, and were higher during switch trials as compared with stay trials. Accuracy tended to be lower for face-trials than for scene-trials. None of the interactions reached significance (all $P > 0.261$).

fMRI Experiment

In the fMRI experiment, participants indicated whether the probe matched the cued item. In contrast to the behavioral experiment, the use of the cues was obligatory for correct task performance, and accuracy was emphasized over speed. Hence only accuracy data were analyzed.

Prior to statistical analysis, data from 3 participants were excluded because they made extensive head movements during scanning or did not perform above chance across all conditions. Trials with RTs longer than 2250 ms were considered errors. The remaining 14 participants reached a mean accuracy of 91.2%. A repeated-measures ANOVA with the factors response (match, nonmatch), cue-1 (face, scene), and cue-2 (switch, stay) did not reveal any significant main effects (all $P > 0.126$) or interactions (all $P > 0.250$), indicating that the conditions and stimuli were well balanced for task difficulty, and that the subjects followed the cues adequately.

The analyses of the fMRI data addressed 2 issues. Firstly, modulation of WM maintenance-related activity in the fusiform and parahippocampal gyri by object-based attentional orienting was analyzed using the ROIs defined by the localizer task. Secondly, brain areas controlling object-based orienting of attention within WM and reweighting of WM content were identified.

Modulation of Activity in Fusiform and Parahippocampal Regions

Parameter estimates. The signal changes based on parameter estimates within the ROIs in left and right FG and PHG were used to assess the involvement of these regions in object-related WM maintenance, and subsequently to test for modulation of maintenance-related activity by attentional orienting following retro-cues.

The presence of maintenance-related activity was tested by comparing the level of activity during the long and jittered
interval immediately following the initial face and scene stimuli (delay 1) against zero using a t-test (planned, one-tailed) (see Fig. 3a,b). Activity levels were significantly above baseline for the FG bilaterally [right FG: \( t(13) = 3.18, P = 0.0035 \); left FG: \( t(13) = 3.87, P = 0.001 \)]. Activity in the right PHG showed a strong trend toward significant activation [\( t(13) = 1.53, P = 0.075 \), one-tailed] but the left PHG was not activated significantly above the baseline [\( t(13) = 0.64, P = 0.27 \)].

To investigate whether maintenance-related activity in FG and PHG was modulated by orienting attention to specific objects in WM, parameter estimates in each ROI were compared during the long and jittered interval following the double cues (delay-3) using a repeated-measures ANOVA that tested the factors of cue-1 (face, scene) and cue-2 (switch, stay) (see Fig. 3a,b). Use of this interval for the analysis ensured that all preceding events were balanced, and therefore could not contribute to the results; equated the physical appearance of the cue stimulus across the comparisons of interest; and provided the necessary temporal separation and jitter to separate the parameter estimates from subsequent events.

Significant interactions between cue-1 and cue-2, which indicated that the level of activation increased when the relevant object matched the preferred type of stimulus in the ROI were observed in the right FG (\( F_{1,13} = 13.08, P = 0.003 \)) and in the PHG bilaterally (right PHG: \( F_{1,13} = 23.85, P < 0.001 \); left PHG: \( F_{1,13} = 23.28, P < 0.001 \)). In these 3 regions, the main effects of cue-1 and cue-2 were not significant (all \( P > 0.203 \)). To investigate the interactions further, paired-sample t-tests (planned, one-tailed) were calculated between switch and stay trials within the same category (face/switch vs. face/stay, and scene/switch vs. scene/stay). The tests were significant in all 3 ROIs [right PHG, face/switch vs. face/stay: \( t(13) = 3.35, P = 0.0025 \); right PHG, scene/switch vs. scene/stay: \( t(13) = 5.23, P < 0.001 \); left PHG, face/switch vs. face/stay: \( t(13) = 3.31, P = 0.003 \); left PHG, scene/switch vs. scene/stay: \( t(13) = 4.66, P < 0.001 \); right FG, face/switch vs. face/stay: \( t(13) = 2.55, P = 0.012 \); right FG scene/switch vs. scene/stay: \( t(13) = 2.22, P < 0.0225 \)].

In the left FG, the ANOVA did not reveal a significant interaction (\( F_{1,13} = 2.43, P = 0.143 \)) but instead revealed a significant main effect of cue-1 (\( F_{1,13} = 8.87, P = 0.011 \)). Activity in this ROI was higher after initial face cues than scene cues. The t-tests also did not reach significance in this ROI, though a trend was observed toward significance for the combination of first and second cues (\( F_{1,13} = 1.85, P = 0.089 \); scene/switch vs. scene/stay: \( t(13) = 0.52, P = 0.306 \)).

**Time courses.** Analysis of the time courses during the double-cue period revealed how the modulation of activity in the posterior ROIs unfolded during object-based orienting according to the different cues. Time courses were extracted from the preprocessed fMRI signal, and thus reflect the fMRI signal rather than the fit of the data to the model (as given by beta values/parameter estimates). Accordingly, the time courses were not corrected for the overlap of activity from preceding events.

The fixed interval between the 2 cues (delay-2) enabled the direct comparison of the time courses between the 4 conditions of interest (face/switch, face/stay, scene/switch, and scene/stay) over the intervals after the first (face, scene) and second (switch, stay) cues. The tight control of the types of preceding and of the subsequent events, as well as the long and jittered intervals between double-cue period and preceding and subsequent events, ensured that the activity during this time period was uncontaminated from systematic differences from surrounding events.

As shown in the Figure 3a,b, the activity following the first cue overlaps, on average, with the undershoot in the hemodynamic response triggered by the encoding of the initial face and scene stimulus. However, because the events preceding the first cue were fully equated across trials, it was nevertheless possible to investigate changes in the hemodynamic signal over time as a function of object-based orienting. To test for modulation of the time courses statistically, the average activity corresponding to the period after cue-1 (3.0—7.5 s after cue-1) and to the period after cue-2 (8.0—12.5 s after cue-1) was compared across conditions.

A clear separation in the time courses occurred after the first cue depending on whether it signaled orienting attention to the face or scene stimulus. Activity in the PHG increased following a scene cue, whereas activity further decreased following a face cue. The opposite pattern occurred in the FG. As expected, activity from switch and stay trials overlapped during the interval following the first cue because the trials are equivalent up to that point. Accordingly, modulation of activity from the first time-window by face versus scene cues was analyzed with paired-sample t-tests (planned, one-tailed; averaged over switch and stay trials). The tests were significant for the right and left PHG [right PHG: \( t(13) = 2.88, P = 0.0065 \); left PHG: \( t(13) = 6.2, P < 0.001 \)], and a trend was observed in the right FG [\( t(13) = 1.6, P = 0.067 \)]. Brain activity within these regions was modulated in the predicted fashion. Orienting attention toward the representation of the scene stored in WM led to a systematic increase of activity in the left and right PHG (as compared with being cued toward the face), and the inverse occurred for the right FG. Modulation in the left FG was not statistically significant [\( t(13) = 0.69, P = 0.25 \)].

After the second, switch or stay, cues, maintenance-related brain activity in the posterior ROIs was further modulated according to the attended object. Activity from the second time-window was entered into repeated-measures ANOVAs testing the factors cue-1 (face, scene) and cue-2 (switch, stay) for every ROI separately. Significant interactions between cue-1 and cue-2 were observed in the right FG and right and left PHG (right FG: \( F_{1,13} = 6.60, P = 0.023 \); right PHG: \( F_{1,13} = 15.00, P = 0.002 \); left PHG: \( F_{1,13} = 24.14, P < 0.001 \)). In these areas, the levels of activation after the second cue were relatively higher if attention was shifted toward the preferred object, after switch cues, or was maintained on the preferred object, after stay cues. The levels of activation decreased when attention was shifted away from the preferred object. The interaction between cue-1 and cue-2 was not significant in the left FG (\( F_{1,13} = 1.89, P = 0.193 \)). In none of the regions was the main effect of cue-2 (switch or stay) significant (all \( P > 0.142 \)) pointing out that the second cue itself did not change the activity in these ROI but that the combination of first and second cue drove the changes in the time courses. A main effect for cue-1 occurred in the left PHG (\( F_{1,13} = 5.98, P = 0.03 \)) only, reflecting more activation following the initial scene cues.

**Areas Controlling the Orienting of Attention to Objects in WM** To identify the brain areas controlling the orienting of attention to faces or scenes held online in WM, and presumably triggering modulation in FG and PHG, activity elicited by the retro-cues was analyzed at the whole-brain level. Because of the fixed
interval between the 2 retro-cues, the analysis combined information from the 2 cues. Brain areas activated by each type of cue-1 (face or scene) relative to the implicit baseline were identified, and the conjunction of both contrasts was calculated (Nichols et al. 2005). The result of the conjunction analysis was further masked by the main effect of switch versus stay for the second cues ([face/switch + scene/switch] – [face/stay + scene/stay]; P = 0.05, uncorrected). The analysis therefore used the comparisons with the most power (cue-1) to reveal brain areas activated by retro-cues, and used the tightly controlled comparison between switch and stay cues (cue-2) to rule out general effects of no interest. The second cues (switch or stay) triggered attentional orienting only when they were switch cues but not when they were stay cues (see also Yantis et al. 2002 for a comparable logic). The direct comparison between switch versus stay cues therefore provides a clean view of brain areas related to the control of object-based orienting in WM, controlling for spurious activity related to visual transients, decoding of cue meaning, phase in the trial, etc.

The results are summarized in Table 1 and Figure 4. Object-based orienting cues activated a network of parietal and frontal cortices. The cortical activations were more extensive and more numerous in the left hemisphere. Parietal activations included the left intraparietal sulcus and medial superior parietal lobule/precuneus. Frontal activations occurred medially around the cingulate sulcus, and laterally in the posterior part of inferior frontal sulcus, close to the junction with the precentral sulcus, extending into middle and inferior frontal gyri.

Discussion

A recent view of WM holds that stimulus representations are stored in posterior brain regions, and that maintenance is achieved by directing attention to those representations, under the control of the dorsolateral prefrontal cortex (see Curtis and D’Esposito 2003; Postle 2006). The present experiments investigated this notion by directly manipulating the allocation of attention toward the representations of either faces or scenes held within WM, and investigating whether the activity in brain areas representing the memory trace of these objects was altered dynamically. Using a novel WM object-based retro-cuing task, we demonstrated that orienting attention toward a representation of a face or a scene held within WM increased the maintenance-related activity in fusiform and parahippocampal gyri, respectively. Findings from complementary behavioral experiment revealed a corresponding performance advantage when participants were cued toward an item which was probed at the end of a trial. Together, the findings reiterate the extensive interplay between attention and WM, and, consistent with the model described above, clearly demonstrate the crucial role of attentional orienting for WM maintenance.

Behavioral Effects of Object-Based Orienting in WM Representations

Both behavioral experiments demonstrated an enhancement of WM performance as the result of orienting attention to relevant object representations held in WM. Valid retro-cues were accompanied by shorter RTs as compared with invalid retro-cues, indicating a beneficial effect of cueing for matching WM content to the probe stimulus (see Fig. 2a,c). This effect was independent of whether attention was oriented between or within stimulus categories, that is, whether participants had to memorize both a face and a scene, or only faces or only scenes.

The replication using stimuli from the same category in Experiment 2 excluded the possibility that the observed RT differences in Experiment 1 could be accounted for by differential response biases toward expected versus unexpected stimulus categories.

Significant cueing effects also became significant for accuracy measures in Experiment 2, where attention was oriented within stimulus categories. Accuracies were significantly higher following valid cues than invalid cues (see Fig. 2d), further strengthening the notion that the cueing manipulation changed the quality of the WM representation. For Experiment 1 no such effects were observed, probably because the participants performed near ceiling (see Fig. 2b).

Attentional Modulation of Maintenance-Related Activity in FG and PHG

Attentional top-down signals that bias information processing toward relevant perceptual representations are an important prerequisite for goal-directed behavior. Numerous studies have shown that responses in specialized cortices are enhanced if attention is directed toward a location or stimulus feature which they represent (e.g., Moran and Desimone 1985; Desimone and Duncan 1995; Tootell et al. 1998; Chawla et al. 1999; O’Craven et al. 1999; Giesbrecht et al. 2003; Serences et al. 2004; Gazzaley et al. 2005). Intuition suggests that our ability to focus on representations of relevant objects and locations held in memory is equally important. This notion is being borne out by recent research showing that it is possible to orient attention toward selective locations within mental representations (Griffin and Nobre 2003; Nobre et al. 2004; Lepsien et al. 2005). The present study extends this notion by showing the ability to orient attention to objects within WM and by demonstrating enhancement of the relevant neural activations when mental representations of objects are brought into the “spotlight of attention.” To our knowledge this is the first demonstration of attentional top-down biasing of task-relevant representations in the WM domain.

The finding that representations of objects in WM can be the target of attentional top-down modulation has strong implications for recent views on WM. It has been proposed that WM maintenance is accomplished by keeping the relevant to-be-remembered internal representations of sensory stimuli within the focus of attention (Desimone and Duncan 1995; Awh and Jonides 2001; Curtis and D’Esposito 2003; Ranganath and D’Esposito 2003; Postle 2006). Partial corroborative evidence comes from reports of sustained activity in inferior temporal regions for the maintenance of faces and scenes in WM (e.g., Drugal and D’Esposito 2001, 2003; Postle et al. 2003; Sala et al. 2003; Ranganath, Cohen, et al. 2004; Ranganath, DeGutis, et al. 2004; Rama and Courtney 2005), and reports showing the ability to orient attention to spatial locations in WM (Griffin and Nobre 2003; Nobre et al. 2004; Lepsien et al. 2005). However, the direct demonstration of the influence of attention on maintenance-related activity and WM performance remained lacking.

The present experiments establish a direct link between the deployment of attention in WM and the quality of WM maintenance, both in terms of WM performance and of enhancement of maintenance-related activity in relevant posterior brain regions. Orienting attention toward the representations of a face or a scene led to a selective increase of the sustained activity in FG and PHG.
Figure 3. Attentional modulation of maintenance-related activity in the left and right PHG (a) and in the left and right FG (b). At the top of each figure are activations of the relevant brain areas within the localizer task shown for one representative subject. The coordinates shown are the means of the individual subjects’ peak coordinates, which were obtained individually using an independent localizer task. The bar-plot graphs show the mean parameter (± SEM) estimates were extracted from the third delay period (following the second cue) for the relevant posterior area in the left hemisphere (left plot) and right hemisphere (right plot) for the 4 experimental conditions: face/switch, face/stay, scene/switch, scene/stay.
The present study contributes to characterizing the extensive interplay between attention and WM (e.g., Cowan 1988; Baddeley 1993; Downing 2000; de Fockert et al. 2001), and builds especially on the work of Awh, Postle, and their colleagues. They have suggested that spatial attention supports rehearsal in spatial WM (Awh et al. 1998, 1999; Awh and Jonides 2001; Postle et al. 2004), and showed the persistence of spatial attention at memorized locations during the delay. Visual responses to task-irrelevant probe stimuli at these memorized locations were amplified and WM performance decreased, when the focus of spatial attention was disrupted. Here, besides gaining new knowledge by investigating WM for objects, we demonstrate that the focus of attention within WM (i.e., within the maintenance period) can be changed dynamically, pointing to a more flexible mechanism than previously proposed. In addition, we show directly the modulation of maintenance-related activity, and facilitation of task-relevant WM retrieval. Thus, in addition to attention being necessary for accurate WM, it can also dynamically alter WM maintenance and performance.

The task design used ensured that attentional biasing of representations occurred only within the WM maintenance period. There was no systematic biasing of attention before the maintenance period because the relevant-to-be-remembered objects were only cued during the delay period. The modulations in activity can also not be explained by long-term retrieval processes because the neural activity remained elevated also after stay cues, which just instructed participants to keep the already-selected object in the attentional focus, without any additional retrieval being necessary. Likewise, the level of activity dropped if attention was orientated away. Findings from the behavioral experiments further suggest that the attention-related biasing of FG and PHG activity is relevant to improvements in WM maintenance.

The shifts in baseline activity of specialized perceptual areas in our experiment were similar to those previously described in perceptual attention studies, in anticipation of relevant location or features (Chawla et al. 1999; Kastner et al. 1999). It could be argued that the present data could similarly be accounted for by the anticipation of a specific type of probe stimulus. However, it is unlikely that the present modulation can be fully accounted by the expectations of an upcoming object because the analysis of the time course indicated an enhancement already during the second delay before the second switch/stay cue. At this phase of a trial, the participants were not actually expecting a particular stimulus. Our findings therefore suggest modulation of WM maintenance-related activity over and above any contribution of modulation related to stimulus expectation.

The present experiment also bears similarities with investigations of visual imagery, in which participants are instructed to generate vivid, percept-like images from their memory. These studies have reported functional similarities between visual imagery and visual perception, such as category-related activation for imagery of faces and houses in FG and PHG (Ishai et al. 2000, 2002; O’Craven and Kanwisher 2000). Imagery requires memory retrieval to construct mental images, and is likely to be mediated by maintenance in WM (Ishai et al. 2002). Thus, both the present experiment and visual imagery may represent top-down biasing of memory representations for different purposes.

### The Orienting Network

Previous research on orienting spatial attention to mental representations has revealed partially overlapping networks for attentional orienting within WM and in the perceptual domain (Nobre et al. 2004; Lepsien et al. 2005). Accordingly, strong similarities in the networks of brain areas for shifting attention to objects in the internal and external world were expected. In particular, a recent study by Serences et al. (2004) was used for comparison, which investigated the neural basis of control of object-based attention in the perceptual domain, also controlling for many spurious factors in similar ways as the present experiment (e.g., visual transients, symbolic decoding, phase in trial). In addition, cueing stimuli in the present study were decoupled from target detection and responses, thus providing an arguably cleaner isolation of the control of object-based orienting.

Orienting attention to representations of faces and scenes held in WM was associated with activity in a left lateralized network of parietal and frontal cortices—including parts of the intraparietal sulcus, superior parietal lobule/precuneus; left dorsal prefrontal cortex around the inferior frontal sulcus; and cingulate gyrus. The brain areas found in the present study were similar to those observed by Serences et al. (see also Fink et al. 1997; Arrington et al. 2000 for investigations of object-based selection of space; and Le et al. 1998; Liu et al. 2003 for investigations of nonspatial orienting of attention to features). Minor differences were that the present network showed more pronounced left lateralization, and was missing activation in the frontal eye fields. Additional activations occurred around the cingulate sulcus and the left posterior inferior frontal sulcus.

---

**Table 1**  
Areas controlling the orienting of attention to objects in working memory

<table>
<thead>
<tr>
<th>H</th>
<th>Label</th>
<th>x, y, z (MNI)</th>
<th>Z</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parietal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>SPL/precun</td>
<td>−12, −72, 63</td>
<td>5.23</td>
<td>123</td>
</tr>
<tr>
<td>L</td>
<td>IPS</td>
<td>−39, −54, 54</td>
<td>5.14</td>
<td>Lo</td>
</tr>
<tr>
<td>L</td>
<td>IPS</td>
<td>−33, −63, 57</td>
<td>4.10</td>
<td>Lo</td>
</tr>
<tr>
<td>Frontal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>cingS</td>
<td>−3, 3, 60</td>
<td>4.85</td>
<td>108</td>
</tr>
<tr>
<td>L/R</td>
<td>cingS</td>
<td>0, 9, 51</td>
<td>4.79</td>
<td>Lo</td>
</tr>
<tr>
<td>L</td>
<td>IF5/FG/MFG</td>
<td>−51, 15, 33</td>
<td>4.48</td>
<td>59</td>
</tr>
<tr>
<td>L</td>
<td>IF5/FG/MFG</td>
<td>−48, 21, 27</td>
<td>4.08</td>
<td>Lo</td>
</tr>
</tbody>
</table>

Note: Coordinates, maximum z-values, spatial extent, and labels of significantly activated brain areas with the conjunction analysis of face and scene cues (threshold P = 0.05 [FDR]), masked by activity elicited by switch over stay cues (threshold P = 0.05, uncorrected). H = hemisphere; L = left; R = right; Lo = local maximum within a larger activation cluster; SPL = superior parietal lobule; IPS = intraparietal sulcus; precun = precuneus; IF5/FG = inferior frontal sulcus gyrus; MFG = middle frontal gyrus; cingS = cingulate sulcus.

The time course plots (lower panels) show the model-free fMRI signal evoked by the first cue (indicated by time zero on the x-axis) through the double-cue interval for the relevant posterior area in the left hemisphere (left panel) and right hemisphere (right panel). The onsets of the first and second cues are indicated by arrows. Brain activity over the preceding 4 s (the minimum interval between the preceding stimulus and cue-1) is shown to provide a view of the baseline level of activation. The interval between the 2 cues was fixed at 4.75 s. The interval after the second cue varied between 4 and 9.5 s. The sections of the curves before the onset of the first cues correspond to the first delay period, demonstrating that the activity following the first cues overlaps with the undershoot in the hemodynamic response triggered by these previous events (see Methods and Results sections for details) (f = face; s = scene; sw = switch; st = stay).
The posterior parietal cortex is especially prominent in the control of spatial attention (Mesulam 1981, 1999; Vandenberghe et al. 2001; Simon et al. 2002; Yantis et al. 2002). The present results, together with previous investigations of object-based attention, suggest that similar parietal areas also participate in controlling object-based attention to perceptual as well as WM representations. The left lateralization observed in the present experiment is consistent with neuropsychological demonstrations of object-based attention deficits following left parietal lesions (Egly et al. 1994). Evidence for lateralization from neuroimaging studies is inconclusive (e.g., not reported in Serences et al. 2004, but see Fink et al. 1997; Arrington et al. 2000), and merits further investigation.

An additional focus of activation observed around the left posterior inferior frontal sulcus has not been commonly observed in studies of object-based attentional orienting but instead is conspicuous in tasks involving executive control in WM and task switching. Several putative functions have been attributed to this region, for example: cognitive control (MacDonald et al. 2000), environmentally guided control when a stimulus does not give the full information needed to fulfill the task (Brass and von Cramon 2004; Forstmann et al. 2005), or relating a current input to WM contents (Monchi et al. 2001). All of these notions could apply to the present experiment, and are also consistent with the idea that this region supports manipulation of information in WM (Owen et al. 1999; Fletcher and Henson 2001), which probably best describes the situation in the present study.

Notes
The research was funded by research grant to A.C.N. by the James S. McDonnell Foundation. Conflict of Interest: None declared.

References

Address correspondence to Jörän Lepsien, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04105 Leipzig, Germany. Email: lepsien@cbs.mpg.de.

Figure 4. The orienting network. Conjunction of face- and scene-cue-related activity, masked by activity elicited by switch versus stay cues, mapped onto the mean brain of all subjects. Values indicate the z and x coordinates of the sections.


