Parietal Cortex Integrates Contextual and Saliency Signals During the Encoding of Natural Scenes in Working Memory

Valerio Santangelo,[1,2]* Simona Arianna Di Francesco,[1,3] Serena Mastroberardino,[1] and Emiliano Macaluso[1]

[1]Neuroimaging Laboratory, Santa Lucia Foundation, Rome, Italy
[2]Department of Philosophy, Social Sciences and Education, University of Perugia, Perugia, Italy
[3]Department of Psychology, "Sapienza" University of Rome, Rome, Italy

◆ ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ ◆

Abstract: The Brief presentation of a complex scene entails that only a few objects can be selected, processed indepth, and stored in memory. Both low-level sensory salience and high-level context-related factors (e.g., the conceptual match/mismatch between objects and scene context) contribute to this selection process, but how the interplay between these factors affects memory encoding is largely unexplored. Here, during fMRI we presented participants with pictures of everyday scenes. After a short retention interval, participants judged the position of a target object extracted from the initial scene. The target object could be either congruent or incongruent with the context of the scene, and could be located in a region of the image with maximal or minimal salience. Behaviourally, we found a reduced impact of saliency on visuospatial working memory performance when the target was out-of-context. Encoding-related fMRI results showed that context–congruent targets activated dorsoparietal regions, while context–incongruent targets de-activated the ventroparietal cortex. Saliency modulated activity both in dorsal and ventral regions, with larger context-related effects for salient targets. These findings demonstrate the joint contribution of knowledge-based and saliency-driven attention for memory encoding, highlighting a dissociation between dorsal and ventral parietal regions. *Hum Brain Mapp 36:5003–5017, 2015.* © 2015 Wiley Periodicals, Inc.

Key words: context; saliency; attention; working memory; natural scenes; parietal cortex; fMRI

◆ ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ ◆

INTRODUCTION

A central issue in human cognition is to understand how we encode and maintain a reliable internal representation of the external world. When we look at a scene in the real world, which would typically involve many objects and high levels of complexity, we will later remember only some of the information included in the scene, as a consequence of our limited working memory (WM) capacity [see, e.g., Luck and Vogel, 2013]. The internal

representation of the external environment is crucial for a number of other cognitive functions, such as long-term memory storage and learning, mental imagery, reasoning, and decisional processes [see, e.g., Fuster, 2006]. Real world scenes are characterised by objects and background elements that are in relation one to each other according to a semantically coherent schema [e.g., Henderson and Hollingworth, 1999]. When we look at an everyday life scene we do not see a "bag" of isolated items, but rather we make use of contextual information and represent the scene as a whole [Pinker, 2007]. Previous literature highlighted the key role played by contextual information during the encoding of natural scenes [e.g., see also Preston et al., 2013; Stirk and Underwood, 2007]. Here, we asked how context-related information contributes to the encoding of objects in WM and how these high-level knowledge-based signals interact with low-level sensory salience for this selection process.

The role of contextual information can be studied by comparing the probability to encode and then remember an object that has a consistent vs. inconsistent relationship with the scene context: e.g., a cup of tea on a breakfast tray vs. a vacuum cleaner in a box shower. Context congruency can improve memory performance in free-recall paradigms [Goodman, 1980]; but out-of-context objects lead to better recognition performance, when the participants are simply asked to report whether an item was present or absent in the scene [Hollingworth and Henderson, 1998; Pezdek et al., 1989]. The latter effect has been interpreted in an "attention framework": out-of-context objects grab attention, which in turn increases the probability of encoding these objects in WM [Silva et al., 2006].

These behavioural data provide us with some initial evidence about the involvement of visual spatial attention during encoding of out-of-context objects, but previous studies did not considers how these "internal" knowledge-based signals interact with stimulus-driven sensory signals that also contribute to attention control. An extensive body of work highlighted the role of visual salience, defined by stimulus intensity, colour, orientation, when viewing naturalistic scenes. Salient locations in the image attract subjects' gaze and attention [Itti and Koch, 2001; Yoshida et al., 2012]. The sensory salience of objects presented during the encoding of complex images has been also associated with enhanced WM performance [Fine and Minnery, 2009], and with the activation of attention controlling areas in the posterior parietal cortex [Santangelo and Macaluso, 2013]. However, in the attention/perception domain, several studies have demonstrated that these stimulus-driven signals interact with "internal", top-down mechanisms of attention control [e.g., Doricchi et al., 2010; Hopfinger and West, 2006; Indovina and Macaluso, 2007; Kincade et al., 2005; Natale et al., 2009; Santangelo et al., 2009; Serences et al., 2005; Serences and Yantis, 2007]. For instance, during the exploration of naturalistic scenes the effect of vis-

ual salience is dramatically reduced when subjects are instructed to perform a specific task, such as searching for a predefined target object [Torralba et al., 2006; Wolfe, 1994]. The specification of top-down goals appears to override the effect of bottom-up salience, with subjects prioritizing search at locations that are most compatible with the target object [Einhauser et al., 2008; Foulsham and Underwood, 2008; Mack and Eckstein, 2011].

Here we used functional imaging to investigate the interactions between contextual knowledge and stimulus-driven signals during the encoding of naturalistic scenes in visuospatial WM. Participants were presented with pictures of naturalistic scenes and, following a short retention interval, they judged the location of a target object extracted from the initial scene. We used a factorial design to manipulate independently whether the target object was extracted from a location of maximal or minimal salience, and whether the target identity was congruent or incongruent with the context of the scene. Stimulus salience was indexed using saliency maps computed with a biologically plausible model [Itti et al., 1998]. The congruence or incongruence of the target was manipulated by including one object that did not belong to the scene context, and object incongruence was confirmed with a separate behavioural test (see "Materials and Methods").

Behaviourally, we predicted that both contextual knowledge (i.e., out-of-context objects) and stimulus-driven signals (i.e., salient objects) would increase the probability that participants would successfully remember the position of target object at retrieval [Goodman, 1980; Fine and Minnery, 2009; Pezdek et al., 1989; Santangelo and Macaluso, 2013; Silva et al. 2006]. In line with the literature postulating an interplay between contextual factors and low-level sensory signals for attention control [e.g., Torralba et al., 2006; Wolfe et al., 2011], we hypothesised that the effect of salience would be reduced when the target object is "out-of-context". This would demonstrate that the presence of knowledge-related, top-down constraints modulates the effects of stimulus-driven signalling during selection for encoding of spatial information in WM.

Functionally, we examined the pattern of activation in regions belonging to the frontoparietal (FP) attention control networks [Corbetta and Shulman, 2002], where mechanisms of selection should contribute to encoding in WM [Gazzaley and Nobre, 2012]. Many studies have highlighted the role of the FP cortex both in attention control and WM functions [e.g., Ikkai and Curtis, 2011; Naghavi and Nyberg, 2005]. The FP cortex has been associated with several aspects of WM, including a role of the parietal cortex in the maintenance of items in memory [Todd and Marois, 2004; Xu and Chun, 2006], as well as selection processes during encoding [Majerus et al., 2007]. On the other hand, the engagement of prefrontal regions has been associated primarily with the monitoring of information stored in WM [e.g., Champod and Petrides, 2007]. Concerning the role of the FP cortex in attention control, the

influential model proposed by Corbetta and Shulman [2002] posits a dissociation between "dorsal" vs. "ventral" components/nodes. The dorsal FP regions include the intraparietal sulcus (IPS) and the superior premotor cortex and are involved in top-down voluntary control of attention [e.g., Corbetta et al., 2000; Hahn et al., 2006]. In contrast, the ventral FP regions include the temporoparietal junction (TPJ) and the middle and the inferior frontal gyrus (MFG/IFG) and are associated with stimulus-driven control [see Corbetta et al., 2008]. In particular, these ventral areas have been activated in tasks requiring spatial re-orienting towards behaviourally relevant stimuli, such as targets presented outside the current focus of attention [e.g., Corbetta et al., 2000; Indovina and Macaluso, 2007; Kincade et al., 2005].

While these dorsal–ventral dissociations have been widely investigated in perceptual and attentional tasks, it is still largely unexplored whether analogous effects applies also to memory functions [but see, Cabeza et al., 2008, for a proposal concerning episodic memory retrieval]. With the current manipulation of context and salience we were able to test the hypothesis that these two types of signals would impact differentially on the activity of regions belonging to the dorsal and ventral FP networks during WM encoding. We predicted that in scenes without any out-of-context object, saliency should contribute to WM encoding via activation of the dorsal parietal cortex [see Santangelo and Macaluso, 2013]. In contrast, the encoding of "attention grabbing" out-of-context targets should involve ventral regions that previous studies have associated with detection and filtering mechanisms [Corbetta et al., 2008] and the violation of context-related expectations [Geng and Vossel, 2013].

Importantly, we expected that the processing of the two types of signals—context and salience—would interact with each other and yield to the modulation of activity in both attention control systems. The latter would fit with recent proposals that emphasise the role of both dorsal and ventral FP regions for the integration of stimulus-driven and endogenous signals in attention control [e.g., see Geng and Vossel, 2013; Macaluso and Doricchi, 2013]. Understanding these interactions in the framework of WM encoding will help specifying the relationship between control mechanisms in attention and visuospatial WM [Gazzaley and Nobre, 2012; Naghavi and Nyberg, 2005], and will contribute to current theoretical [e.g., Oliva and Torralba, 2007; Wolfe et al., 2004, 2011] and computational [e.g., Torralba et al., 2006] models of attention control in real-life situations [see also Peelen and Kastner, 2014].

MATERIALS AND METHODS

Participants

Sixteen right-handed healthy volunteers took part in the fMRI experiment. Two participants were excluded from data analysis because of within-fMRI-run head-movements larger than 3 mm or 3°, leaving 14 participants for the final analyses (4 males and 10 females, mean age: 22.0 years, range: 20–32 years). All participants gave written consent to the study, which was approved by the independent Ethics Committee of the Santa Lucia Foundation.

Stimuli and Task

Each trial consisted of a "cue" display (2 s), an encoding phase (4 s), a maintenance phase (8 s delay), a retrieval phase (3 s) and, finally, a memory confidence judgment (3 s, see Fig. 1A). In the encoding phase, subjects were presented with pictures depicting scenes of everyday life. Depending on the "context" condition (see below), the picture sometime included an out-of-context object that served as memory target in the incongruent conditions. Following the 8 s delay involving the presentation of a blank screen, subjects were shown a single target object cut out from the original picture and presented on a gray background (retrieval phase). Subjects had to report whether the target object was in the "same vs. different" location with respect to the position at encoding. For the confidence judgment, a display with the question "Are you sure? (y/n)" was presented and participants pressed one of the two response buttons. It should be noted that with the current two-alternative forced-choice task (same vs. different target location), subjects may provide quite a few "correct responses" even when failing to actually recollect the target object location, because they were just "guessing" (50% chance level). Because of this, all our analyses involving accuracy and latencies of retrieval performance, fixation-patterns and neuroimaging data considered only the trials that included a "sure" response on the confidence judgment [see Supporting Information Table SI; cf. also Santangelo and Macaluso, 2013].

The experimental design included three "context" conditions: congruent (CON), incongruent (INC) and "cued" control condition (CUED). In the CON trials, the pictures included only objects with a congruent semantic relation with the scene, resulting in a defined visual space [Henderson, 2003]. In the INC trials, the pictures contained an out-of-context object (e.g., a vroom in a shower box). It should be noted that when the context was congruent the identity of the target object was unknown until retrieval. In contrast during the encoding phase of the incongruent trials, the participants could guess what object would be tested at retrieval, i.e. the out-of-context object. In order to control for this difference in the level of un-/certainty about the target identity during encoding, the design also included a CUED context condition. In this control condition, the picture contained only in-context objects, as in the CON condition, but now the identity of the target object was disclosed before the beginning of the encoding phase. In the CUED trials, the "cue phase" consisted in the presentation of a written word indicating the target
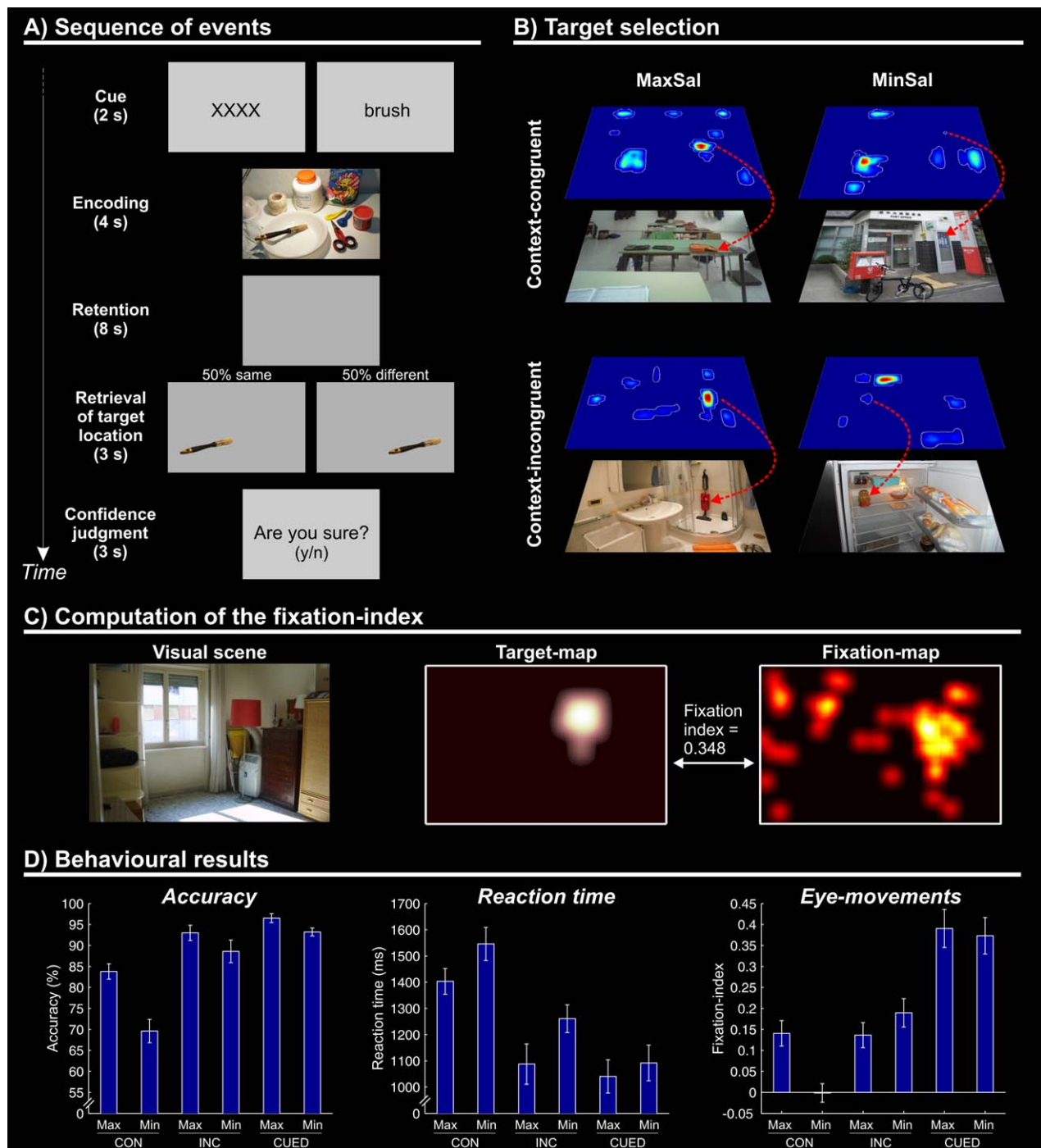
Figure 1.

(A) Schematic diagram showing the sequence of events during one trial. The trial began with the presentation for 2 s of either a cue word (indicating the target object for the cued condition) or a mask made by four "X" letters. A picture was then presented for 4 s (encoding), followed by a blank display for 8 s (retention). At retrieval, a single target object was presented either in the same or different location compared with the object's location in the original picture. Participants had to judge the "same vs. different" location of the target (3 s) and then express a confidence judgment requiring a "yes/no" response (3 s). (B) Examples of pictures in the congruent and incongruent condition, including target objects selected at the point of maximal (left panel) or minimal salience (right panel). (C) Schematic illustration showing the computation of the fixation index. For each picture, we constructed a target map including the location/shape of the target object within the original image (the red lamp, in this example). This was then correlated with the fixation map associated with the same image, thus producing a numerical index for each picture. (D) Mean accuracy, mean RT, and fixation-index (±standard errors) plotted according to context condition (CON, INC, and CUED) and target saliency (max vs. min).

that would be tested during retrieval, while in the CON and INC trials a string of "xxxx" was presented instead (see Fig. 1A). Areas involved specifically in the processing of the out-of-context objects should respond in INC trials, but not in the control CUED trials. In contrast, areas merely modulated by target un-/certainty should show similar patterns of activation both in INC and CUED conditions. Because the CUED condition served only as a control, the design did not include any CUED image with an out-of-context object. Such trials would allow testing for possible interactions between target un-/certainty and the presence of the out-of-context object, but this was beyond the aim of the current study.

The experiment included 144 trials, each making use of a different picture. The pictures included internal (e.g., a kitchen, a bathroom, etc.) and external scenes (e.g., a garden, a street, etc.), but no single-object photo or living things such as people or animals. The pictures were displayed at 18° × 12° of visual angle (resolution in pixels: 680 × 448). For each picture, we designated a target object that would then be presented at retrieval. The target objects were extracted either at the location of minimal (min) or maximal (max) salience. Salience was defined according to the model proposed by Itti et al. [1998]. For each picture, the saliency map was computed with the "Saliency Toolbox 2.2" (http://www.saliencytoolbox.net/), which uses local discontinuities in line orientation, intensity contrast and colour opponency (see also Fig. 1B). The final set of 144 pictures was selected from a larger set that comprised images downloaded from World Wide Web and pictures shot on purpose for this study. We selected pictures that included a well-defined and recognisable object that was located either at the position of minimal or maximal salience in the image. This was then used as the target object for the memory retrieval test. All the pictures with an out-of-context object were shot on purpose.

On incongruent trials (48 pictures), the target was always the out-of-context object. In half of these trials, the out-of-context object was at the point of maximal salience (i.e., 24 context–incongruent, high-saliency trials: INC max), and in the other half at the location of minimal salience (24 context–incongruent, low-saliency trials: INC min). For the congruent trials (48 pictures), we designated as the memory target an in-context object, such as a pencil case in a classroom, again located at either the point of maximal or minimal salience. There were 24 trials with context–congruent, high-saliency targets (CON max), and 24 trials with context–congruent, low-saliency targets (CON min). For the control trials (48 pictures), we used congruent pictures only, again with the target located at either the point of maxima or minimal salience (CUED max, CUED min; 24 trials each).

We performed several quantitative evaluations of the pictures to control for possible differences between the context conditions. First, we made sure that the size of the target objects did not significantly differ between the six main experimental conditions [$F(5,138) = 1.6$; $p = 0.162$]

(see Supporting Information Table SI). Second, we checked the eccentricity of the target objects in the six conditions. Separately for each picture, we computed the centre of mass of the target object as the average of the horizontal and vertical coordinates of each pixel belonging to the target object. We extracted the horizontal eccentricity considering the horizontal position of the centre of mass and converted this to degrees of visual angle. The absolute horizontal eccentricity values were submitted to a one-way ANOVA that did not reveal any significant difference between the six conditions: [$F(5,138) < 1$; n.s.] (see Supporting Information Table SI). Finally, we made sure that the distances between the target objects (centre of mass) in the minimal saliency conditions (i.e., CON min, INC min and CUED min) and the peaks of maximal saliency in the corresponding scene did no differ across conditions [mean distance ± standard error of the mean: 7.2 ± 0.9, 6.5 ± 0.6, and $6.7 \pm 0.8°$ of visual angle, respectively; $F(2,69) < 1$; n.s.]. These quantitative measures helped us to rule out that factors such as the target size or eccentricity affected our comparisons between conditions. Nonetheless, we acknowledge that other factors/parameters may contribute to the processing of natural scenes, including the overall number of objects in each scene, which is a common limitation of studies using naturalistic material.

During the WM retrieval phase, in half of the trials the target object was presented at the same location as in the original image, while in the other half of the trials the target object was presented at the mirror location in the opposite hemifield (see Fig. 1A). It should be noticed that the object location was flipped, but the stimulus vertical axis remained unchanged. Moreover, because the eccentricity of the targets in the original images was matched across conditions (see above and Supporting Information Table SI), also the distance between the original object location and the location of the target at retrieval in the "flipped trials" was matched across the experimental conditions. Participant reported whether the target was at the same vs. different location by pressing one of two buttons with their right hand.

Participants underwent three fMRI runs, lasting approximately 18 min each and including 48 trials each. The order of the trials within and across runs was randomised with three constraints: (1) each run included the same number of congruent, incongruent and cued targets; (2) there was an equal number of high and low saliency targets, balanced for each context condition; (3) at retrieval, half of the targets were at the "same" and half at the "different" location, balanced for each combination of context and saliency condition. Between each trial, there was a variable intertrial interval of 1–3 s, uniformly distributed.

Postscanning Test of the in-Congruence Manipulation

To confirm that subjects could reliably recognise the out-of-context objects in incongruent trials, after the fMRI

experiment each subject was presented again with the set of 48 incongruent pictures. For each picture, subjects were asked to indicate the most out-of-context element using a mouse click ("Which is in your opinion the most odd object/element in this scene?"), and then rate its "oddity" using a 10-point scale (1 = minimum level of oddity; 10 = maximum level of oddity). Subjects had no temporal constraint to provide their response. In almost all of the cases (98.7%) the object selected as "odd" was the one used as the target object during fMRI, with an average rating of 7.5 points. The average oddity rating was equivalent in the minimal (7.2 ± 0.3) and maximal-salience conditions (7.8 ± 0.2), as revealed by a two-sample t test [$t(46) = -1.5$, $p = 0.135$].

Eye Movements and Fixation Indexes

Together with the memory performance data we also acquired the subjects' gaze-position during fMRI. This was done to provide us with additional information about the interaction between salience and context-related signals for the allocation of spatial attention resources, indexed here with fixation patterns [see also Santangelo and Macaluso, 2013; Silva et al., 2006]. The eye-movement data were recorded with an ASL eye-tracking system, adapted for use in the scanner (Applied Science Laboratories, Bedford, MA; Model 504, sampling rate 60 Hz). For each picture we computed a fixation-index that was derived from two-dimensional "fixation-maps" [see Fig. 1C; cf. also Santangelo and Macaluso, 2013]. We considered the fixations in a 4 s window, corresponding to the duration of the encoding phase, starting from the picture onset. Fixations were identified as time-windows with a minimum duration of 100 ms where the velocity of the eye-traces did not exceed 50°/s. For each picture, we generated a 680 × 448 matrix of zeros and we assigned value 1 at the coordinates of the corresponding fixation positions represented by the median of the x/y-position within each fixation window. The procedure was repeated for each fixation and for all 14 participants. This matrix was then scaled by its sum, which ensured that all maps had a total value equal to 1, while retaining the information about the relative spatial distribution of the fixation peaks. Finally, each map was smoothed with a Gaussian filter (FWHM = 1°) to generate the final fixation map. The fixation maps have the highest values at locations that were fixated more frequently [see Henderson, 2003, for review].

For each picture, the target object fixation index was computed using the correlation (Fisher transformed to Z) between the fixation map and the "target map" (see Fig. 1C). The target map consisted of a 680 × 448 matrix with values set equal to 1 at the coordinates of the target object and zeros everywhere else, scaled by its sum and smoothed with a Gaussian filter (FWHM = 1°). The resulting fixation-index will be high when the designated target object corresponds to an highly fixated location in the image. Fixation indexes were used to further characterise the distribution of spatial attention in the six main experimental conditions (context × saliency) considering trials in which the target object was successfully remembered (i.e., correct and "sure" responses; see Fig. 1D).

Magnetic Resonance Imaging

A Siemens Allegra (Siemens Medical Systems, Erlangen, Germany) operating at 3T and equipped for echo-planar imaging (EPI) was used to acquire the functional magnetic resonance images. A quadrature volume head coil was used for radio frequency transmission and reception. Head movement was minimised by mild restraint and cushioning. Thirty-two slices of functional MR images were acquired using blood oxygenation level-dependent imaging (3 × 3 mm, 2.5 mm thick, 50% distance factor, repetition time = 2.08 s, time echo = 30 ms), covering the entirety of the cortex.

fMRI Data Analysis

We used SPM8 (Wellcome Department of Cognitive Neurology) implemented in MATLAB 7.4 (The Math-Works Inc., Natick, MA) for data preprocessing and statistical analyses. Each participant underwent three fMRI-runs, each comprising 515 volumes. After having discarded the first four volumes of each run, all images were corrected for head movements. Slice-acquisition delays were corrected using the middle slice as reference. All images were normalised to the standard SPM8 EPI template, resampled to 2 mm isotropic voxel size, and spatially smoothed using an isotropic Gaussian kernel of 8 mm FWHM. Time series at each voxel for each participant were high-pass filtered at 220 s and prewhitened by means of autoregressive model AR (1).

Statistical inference was based on a random effects approach, which comprised two steps: first-level multiple regression models estimating contrasts of interest for each subject, followed by the second-level analyses for statistical inference at the group-level. The aim of the study was to test for the effects of context (knowledge-based signals), salience (stimulus-driven signals), and the interaction between these two factors, in trials when participants successfully encoded the position of the target object. Accordingly, the first-level models considered as events-of-interest only correctly remembered trials followed by a "sure" confidence judgment. The trials were modelled separately according to the 3 contexts (CON, INC, and CUED) and the two levels of salience (max and min): that is, CON max, CON min, INC max, INC min, CUED max, and CUED min. The average number of these trials varied in the six different conditions (see Supporting Information Table SI). Correction for non-sphericity [Friston et al., 2002] was then used to account for possible differences in error variance across conditions, arising—e.g.,—because of the different number of trials in each

TABLE I. MNI coordinates (*x, y, z*), Z values, and *p* values for areas showing a significant overall activation or de-activation, across the six main experimental conditions: CON max, CON min, INC max, INC min, CUED max, CUED min

	x y z	Z value	p-FWE-corr
Overall activation			
Left MFG*	−50 8 30	4.93	0.047
Left IPS*	−24 −70 44	5.20	0.017
Right IPS*	28 −68 42	4.81	0.073
Left LinG	−22 −78 −10	5.86	<0.001
Right LinG	20 −82 0	6.14	<0.001
Left FG	−28 −68 −12	6.12	<0.001
Right FG	28 −60 −12	5.56	0.003
Overall de-activation			
Right AG*	52 −56 46	5.77	0.001
Right TPJ*	54 −28 20	5.57	0.003
Right SMG*	64 −28 36	5.54	0.003
Left TPJ*	−56 −30 22	5.47	0.005
Right ITG	42 −8 −20	6.13	<0.001
PCC	6 −34 18	6.01	<0.001
Left FO	−56 0 2	5.55	0.003
Right MTG	64 −42 −2	5.29	0.012

Note: *Area used for the regions of interest (ROI) analyses. The significance level was set to p-FWE-corr < 0.05 at the voxel level, considering the whole brain as the search volume. For the "de-activation" contrast, an additional constraint about the minimum cluster size (20 voxels) was applied. MFG: middle frontal gyrus; IPS: intraparietal sulcus; LinG: lingual gyrus; FG: fusiform gyrus; AG: angular gyrus; TPJ: temporoparietal junction; SMG: supramarginal gyrus; ITG: inferior temporal gyrus; PCC: posterior cingulated cortex; FO: frontal operculum; MTG: middle temporal gyrus.

of the six conditions of interests and/or any nonindependent error terms for the repeated measures. All the other trials (i.e., incorrect responses followed by "sure" confidence judgment; plus all trials with "unsure" confidence judgment, irrespective of accuracy; see Supporting Information Table SI) were modelled as a separate event-type and were not entered in the group analyses. The BOLD response was modelled using a finite impulse response (FIR; time window = 20 s; number of bins = 10), time-locked to the onset of the pictures.

For second-level group analyses, we considered the parameter estimates associated the 3rd bin of the FIR model, corresponding to the expected peak of encoding-related activity, approximately 6 s after the picture onset. We carried out two second-level analyses. First, we sought to localise regions of the dorsal and ventral frontoparietal systems that responded in the current experimental settings; that is, during the WM encoding of complex and naturalistic pictures. Accordingly, we tested for the overall activation or de-activation, irrespective of context and saliency condition. For each subject we computed a single contrast that averaged activity across the six conditions of

interest. The resulting contrast-images (one for each subject) were submitted to a one-sample *t* test, wherein we tested for "overall activation" (contrast: +1) and "overall de-activation" (contrast: −1). The resulting activations/de-activations were used to define regions of interest (ROIs) that were then used to test for condition-specific effects, see below. The statistical threshold was set to p-FWE-corrected <0.05 at the voxel level, considering the whole brain as the volume of interest (see Table I).

Given our main hypothesis about the role of the fronto-parietal attention systems for the integration of salience and context signals during WM encoding (cf. Introduction), we selected the seven peaks of the activation/de-activation maps that belonged to the dorsal or the ventral FP cortex (see Table I, areas marked with an asterisk; and Fig. 2). These included two regions in the superior parietal cortex (the left and the right intraparietal sulcus, IPS); three regions in the right inferior parietal cortex (the temporo-parietal junction, the supramarginal gyrus, and the angular gyrus; TPJ, SMG, and AG); plus the TPJ and the middle frontal gyrus (MFG) in the left hemisphere. Unexpectedly, these initial contrasts did not reveal any significant effect in the dorsal premotor cortex that therefore was not included in any of the following analyses. We examined any condition-specific effect in the selected regions using seven ROIs, each consisting of a sphere with a radius of 10 mm centred on the peak of activation/de-activation map. It should be noticed that the activation and de-activation contrasts are orthogonal to the condition-specific effects and will not bias our main condition-specific tests toward any specific pattern of activation [see, e.g., Buchel et al., 1998][1].

The main analysis tested for the effects of context, saliency, and the interaction between these two factors. Using the same first-level models, we now computed six separate contrast images corresponding to the six conditions of interest (context × salience), again considering the 3rd bin of the FIR model. The contrast images were modelled at the group level with a repeated-measure ANOVA with nonsphericity correction. This model was used to extract the data of the seven frontoparietal ROIs (cf. Table I; see also Fig. 2), using the eigenvariate data extraction function of SPM. The response of each ROI was then analysed with SPSS (Statistical Package for Social Science, version 21.0) using a 3 × 2 repeated-measures ANOVA with the factors of context and salience. When appropriate, the Huynh–

---

[1]An alternative approach to localize relevant voxels in the current task settings would have involved the use of a localizer protocol with passive presentation of naturalistic pictures. However, this approach might have been problematic as passive viewing may not activate all the fronto-parietal areas involved in attention and WM that we actually sought to investigate here. Alternatively, we could have used some traditional attention or WM tasks with simple stimuli (e.g., spatial cueing or n-back tasks) but again possibly triggering activity in regions slightly different to those that engaged during the encoding of rich and complex naturalistic pictures.
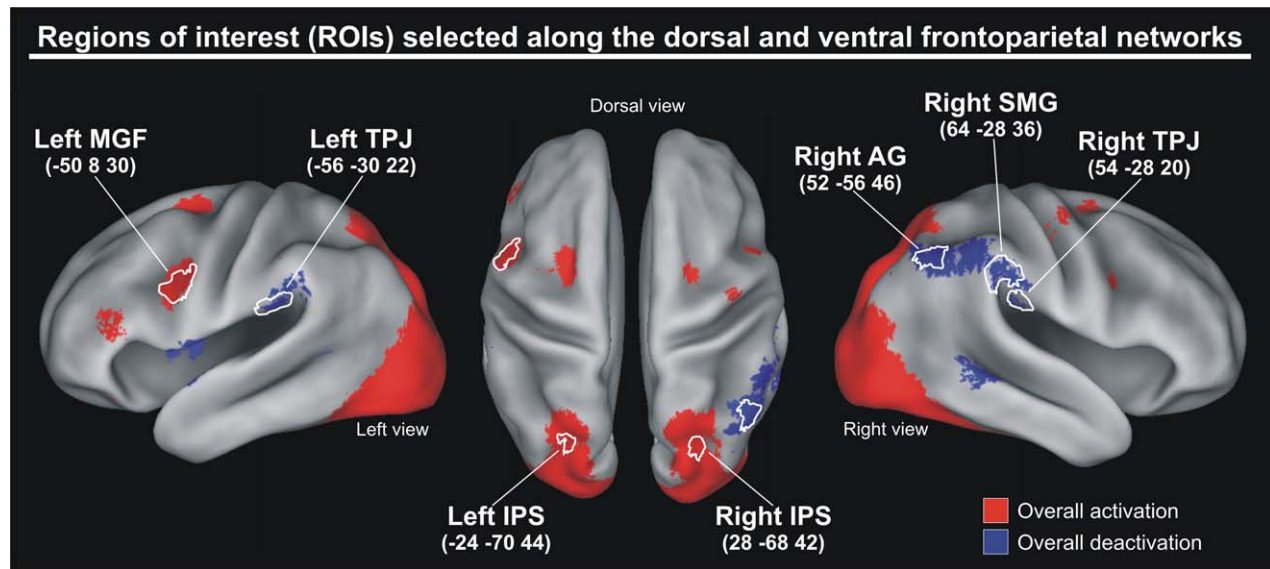
Figure 2.
Encoding-related overall activation (red map; threshold at p-unc = 0.001, for display purposes) and de-activation (blue map; displayed at a threshold of p-FWE-corr. = 0.05) overlaid on an inflated template. The figure also displays the location of the seven regions of interest (ROIs) selected along the dorsal and ventral frontoparietal attention systems (see also Table I).

Feldt procedure was used to correct for violations of sphericity.

RESULTS

Behavioural Data

Behaviourally, we predicted that both high salience and out-of-context objects would be better remembered than low salience and incontext objects [see e.g., Fine and Minnery, 2009; Silva et al. 2006]. We also predicted a reduced effect of salience in the presence of knowledge-related factors, that is, when the target object was also "out-of-context" [Torralba et al., 2006; Wolfe et al., 2011].

To test these predictions we carried out three different 3×2 repeated-measures ANOVAs on the accuracy, reaction times (RT), and eye-movements data (see Fig. 1D and Table II). These ANOVAs included the factors of "context" (CON, INC, CUED), and "saliency" (max, min). Concerning the accuracy and RT data, we found that INC target objects were better remembered that CON target objects, as demonstrated by the significant main effect of context (see Table II). *Post hoc* comparisons confirmed that subjects were more accurate and faster in the INC than CON context (91% and 1175 ms, vs. 77% and 1474 ms; both ps < 0.001), highlighting the expected "consistency effect" [e.g., Pezdek et al., 1989]. The retrieval performance in the INC context was still lower than in the control CUED condition, where subjects were fastest and most accurate as a consequence of their knowledge of the target identity (95% and 1067 ms; INC vs. CUED, both ps < 0.030). As predicted, the ANOVAs also revealed that maximal-saliency targets were better remembered than minimal-saliency targets, as demonstrated by the significant main effect of target saliency (Table II). Specifically, subjects were faster and more accurate in judging the target location at retrieval when, at encoding, the target object was at the point of maximal saliency (91%, 1177 ms) compared with the point of minimal saliency (84%, 1300 ms).

Finally, we found a modulation of the effect of saliency according to context factor. This was evidenced by a significant interaction between the two factors on the accuracy data (Table II): Subjects were more accurate in judging maximal- than minimal-saliency targets in the CON context (84 vs. 70%; $p < 0.001$). A related statistical trend was present also in the INC context (93 vs. 89%; $p = 0.064$), but not in the CUED context (97 vs. 93%; $p = 0.155$). A two-tailed paired t test confirmed that the magnitude of the saliency effect, targets at max- vs. min saliency, was larger for CON than INC trials (max- minus min saliency: CON = 14% vs. INC = 4%; $t(13) = 2.6$; $p = 0.021$][2]. These results are in line with our predictions and demonstrate an effect of bottom-up salience for

---

[2]Note that although not significant, the RT data showed an analogous interaction pattern as the accuracy data (see Table 2), with a larger magnitude of the "max vs. min saliency" difference in the CON (1403 vs. 1546 ms) than in the INC (1088 vs. 1261 ms) and CUED (1041 vs. 1092 ms) contexts.

TABLE II. Statistical results for the ANOVA models including the factors of context (CON, INC, and CUED) and salience (maximal vs. minimal) on the accuracy, reaction time (RT), and eye-movement (EM) data

	Main effect of context	Main effect of salience	Context × salience interaction
Accuracy	$F(2, 26) = 57.0; p < 0.001; \eta^2 = 0.814$	$F(1, 13) = 39.2; p < 0.001; \eta^2 = 0.751$	$F(2, 26) = 7.1; p^* = 0.007; \eta^2 = 0.353$
RT	$F(2, 26) = 59.9; p < 0.001; \eta^2 = 0.822$	$F(1, 13) = 23.2; p < 0.001; \eta^2 = 0.641$	$F(2, 26) = 2.0; p = 0.153; \eta^2 = 0.134$
EM	$F(2, 46) = 35.7; p < 0.001; \eta^2 = 0.608$	$F(1, 23) = 1.9; p = 0.180; \eta^2 = 0.077$	$F(2, 46) = 4.8; p^* = 0.015; \eta^2 = 0.172$

Note: *Sphericity corrected p value.

successful memory performance, but also that this was reduced or absent in the presence of top-down selection signals in the INC and CUED contexts.

These findings were supported further by the analyses of the fixation patterns during encoding. The 3 × 2 ANOVA on the eye-movements data revealed that during encoding the target object was fixated progressively less from CUED (mean = 0.38) to INC (mean = 0.16) to CON (mean = 0.07) context, cf. main effect of context in Table II. On average there was no difference between the fixations of maximal vs. minimal-saliency targets, but the effect of saliency was modulated according to context (see corresponding main effects and interactions in Table II). In the CON context, target objects that were later successfully remembered were fixated more, when they were presented at the position of maximal compared to minimal saliency ($p < 0.001$). In contrast, the effect of salience on fixations was not significant in the INC and CUED contexts ($p = 0.246$ and $p = 0.783$, respectively). This was again consistent with the prediction that the presence of top-down selection signals related to the out-of-context and the pre-cued targets reduced the contribution of bottom-up saliency during WM encoding.

fMRI Data

The analysis of the imaging data aimed to assess the impact of externally driven factors (salience) and top-down, internally driven signals (context) during encoding in WM. First, we identified regions that engaged during the encoding phase of the trial. We performed t tests to localise areas that activated or de-activated during encoding, irrespective of context and saliency (see Table I and Fig. 2). The test for positive activations revealed significant effects in the intraparietal sulcus bilaterally (IPS), the left middle frontal gyrus (MFG), plus several clusters in the occipital visual cortex, including the fusiform and the lingual gyri bilaterally (see red map in Fig. 2, plus Fig. S1, Supporting Information, and Table I). The activation of the occipital cortex was not surprising given that this contrast identified also the sensory responses associated with the visual presentation of the pictures [e.g., Hasson et al., 2004]. However, we later focused on any condition-specific response in the IPS and MFG that belong to the attention control systems, see also below. The opposite contrast, testing for de-activations, revealed several significant clusters in the right inferior parietal cortex, including the right temporoparietal junction (TPJ), the supramarginal gyrus (SMG) and the right angular gyrus (AG). Significant de-activations were also detected in the left TPJ, the right inferior and middle temporal gyri, the left frontal operculum and the posterior cingulate cortex (see blue map in Fig. 2, Fig. S1 in Supporting Information, and Table I).

Next, we turned to the main analysis that investigated the condition-specific effects of context, saliency and the interaction between these two factors. Because of our primary interest in the responses of the nodes of the fronto-parietal attention networks (see Introduction), we selected the areas showing "overall activation or de-activation" that were located in the dorsal or the ventral fronto-

TABLE III. Statistical results for the ANOVA models including the factors of context (CON, INC, and CUED) and salience (maximal vs. minimal) on the fMRI data extracted from the seven regions of interest (ROIs)

	Main effect of context	Main effect of salience	Context × salience interaction
Left MFG	$F(2, 26) = 3.6; p^* = 0.048; \eta^2 = 0.215$	$F(1, 13) = 0.9; p = 0.772; \eta^2 = 0.007$	$F(2, 26) = 4.8; p^* = 0.022; \eta^2 = 0.268$
Left IPS	$F(2, 26) = 9.8; p = 0.002; \eta^2 = 0.429$	$F(1, 13) = 6.7; p = 0.023; \eta^2 = 0.339$	$F(2, 26) = 18.9; p < 0.001; \eta^2 = 0.593$
Right IPS	$F(2, 26) = 22.6; p < 0.001; \eta^2 = 0.635$	$F(1, 13) = 0.3; p = 0.599; \eta^2 = 0.022$	$F(2, 26) = 12.7; p^* = 0.001; \eta^2 = 0.494$
Left TPJ	$F(2, 26) = 40.1; p^* = 0.002; \eta^2 = 0.755$	$F(1, 13) = 1.8; p = 0.199; \eta^2 = 0.124$	$F(2, 26) = 29.8; p < 0.001; \eta^2 = 0.696$
Right TPJ	$F(2, 26) = 21.6; p < 0.001; \eta^2 = 0.625$	$F(1, 13) = 0.8; p = 0.782; \eta^2 = 0.006$	$F(2, 26) = 16.2; p < 0.001; \eta^2 = 0.555$
Right SMG	$F(2, 26) = 23.3; p < 0.001; \eta^2 = 0.641$	$F(1, 13) = 2.8; p = 0.119; \eta^2 = 0.176$	$F(2, 26) = 8.2; p = 0.002; \eta^2 = 0.386$
Right AG	$F(2, 26) = 8.1; p^* = 0.003; \eta^2 = 0.385$	$F(1, 13) = 18.6; p = 0.001; \eta^2 = 0.588$	$F(2, 26) = 10.4; p^* = 0.003; \eta^2 = 0.443$

Note: *Sphericity corrected p value. MFG: middle frontal gyrus; IPS: intraparietal sulcus; TPJ: temporoparietal junction; SMG: supramarginal gyrus; AG: angular gyrus.

parietal cortex. These included two regions in the superior parietal cortex (left and right IPS); three regions in the right inferior parietal cortex (rTPJ, rSMG, and rAG); plus the TPJ and the middle frontal gyrus (MFG) in the left hemisphere (see Table I, regions marked with an asterisk, and Fig. 2). For each of these regions, we examined any condition-specific effect using a 3×2 repeated-measure ANOVA including the factors of context and saliency.

The ANOVAs revealed a main effect of context in all the seven regions (see Table III for details). *Post hoc* analyses indicated that in all ROIs the pattern of activity in the CON and INC context was different compared to the control CUED condition: CON vs. CUED, all $ps < 0.046$; INC vs. CUED, all $ps < 0.027$. This confirmed the involvement of the selected regions in the current task, over and above any trivial effect of visual stimulation or WM task-performance. However, please note that the main effect of context reflects a difference in mean activity across maximal and minimal saliency conditions and should be interpreted cautiously, because all ROIs also showed a context x salience interaction. For each ROI, Table III reports the main effects of context and salience, plus the context by salience interaction that was significant in all ROIs. The latter highlights that the patterns of activity during the encoding of max/min-saliency objects were different according to the context condition.

The *post hoc* analyses showed that the level of activity in the right TPJ and left MFG was lower for targets at max saliency than targets at min saliency, but only in the INC context: both $ps < 0.040$; compare bar 3 vs. bar 4, in the signal plots of Figure 3A. In contrast, in the CON context there was no significant difference between max- and min-saliency targets: both $ps > 0.079$; compare bar 1 vs. bar 2 in plots of Figure 3A. In the CUED context, the activity of the left MFG did not vary according to saliency: $p = 0.111$; while in the right TPJ there was a significant effect of saliency, but with a positive response in the max-saliency condition: max > min; $p = 0.002$; compare bar 5 vs. bar 6 in the right plot in Figure 3A. The finding of opposite patterns of activation in the INC and CUED contexts rules out the explanation that the de-activation for max-saliency targets in the INC condition merely reflected the easy/fast detection of the target stimulus during encoding. The latter would predict analogous patterns in the INC and CUED conditions, with an early detection of the target object because this was either out-of-context or precued. *Post hoc* analyses did not revealed any related pattern of activity in the other ventral ROIs (i.e., the left TPJ, the right SMG and AG; see Supporting Information).

In the left and right IPS, the *post hoc* analyses revealed that activity was higher for targets at max saliency than targets at min saliency, specifically in the CON context. In the right IPS, this effect was significant ($p = 0.018$; compare bar 1 vs. bar 2 in the right plot on Fig. 3B), while in the left IPS there was a related statistical trend ($p = 0.072$). In the INC context, there was no difference between min-

and max saliency (both $ps > 0.091$; compare bar 3 vs. bar 4, in the plots of Fig. 3B); but the postdoc tests revealed an effect of saliency in the CUED context (both $ps < 0.044$). Importantly, however, the pattern of activation in the CUED context was different from that observed in the CON context, now with a de-activation of both left and right IPS in the max-saliency conditions (see bar 5 in the plots of Fig. 3B).

In sum, the fMRI analyses revealed significant interactions between context and saliency both in dorsal and ventral regions. The patterns of activity were markedly different in the two sets of areas: saliency was found to activate the dorsal regions during encoding in congruent contexts, while in the ventral regions we found that saliency lead to larger de-activation during encoding within incongruent contexts.

DISCUSSION

Both high-level contextual information and low-level sensory salience can bias information encoding into WM [Fine and Minnery, 2009; Foulsham and Underwood, 2008; Hollingworth and Henderson, 1998; Melcher and Piazza, 2011; Pooresmaeili et al., 2014]. However, how these factors interact to determine the selection of relevant information in the external world and the formation of internal memory representations is still largely unexplored. Here, we hypothesised that these two types of signals would modulate the activity of regions in the frontal and parietal cortex, consistent with the existence of common cognitive control processes for attention and WM [Ikkai and Curtis, 2011; Naghavi and Nyberg, 2005].

We investigated the interplay between high-level contextual information and low-level sensory salience at the behavioural and the neural level using a task involving WM encoding of natural scenes. Our behavioural findings showed that both salience and context-related signals can affect WM performance, in line with previous studies that investigated the contribution of externally/internally driven signals during WM encoding [Botta et al., 2010; Gazzaley et al., 2005; Gazzaley and Nobre, 2012; Schmidt et al., 2002]. We found a reduced impact of target saliency on WM performance when the display included an out-of-context object. This indicates that internal *a priori* knowledge about real-world scenes leads to the detection and orienting of attention towards the location of the out-of-context object [Hollingworth and Henderson, 1998; Silva et al., 2006]. Concurrently, the presence of these high-level signals reduced the impact of low-level sensory salience on WM performance. The latter results fit with related findings in visual search paradigms that showed diminished effects of stimulus-driven saliency when participants utilise internal, task-related knowledge to guide spatial orienting [Oliva and Torralba, 2007; Wolfe and Horowitz, 2004]. We demonstrate analogous effects in the context of WM encoding and infer that perceptual and contextual/
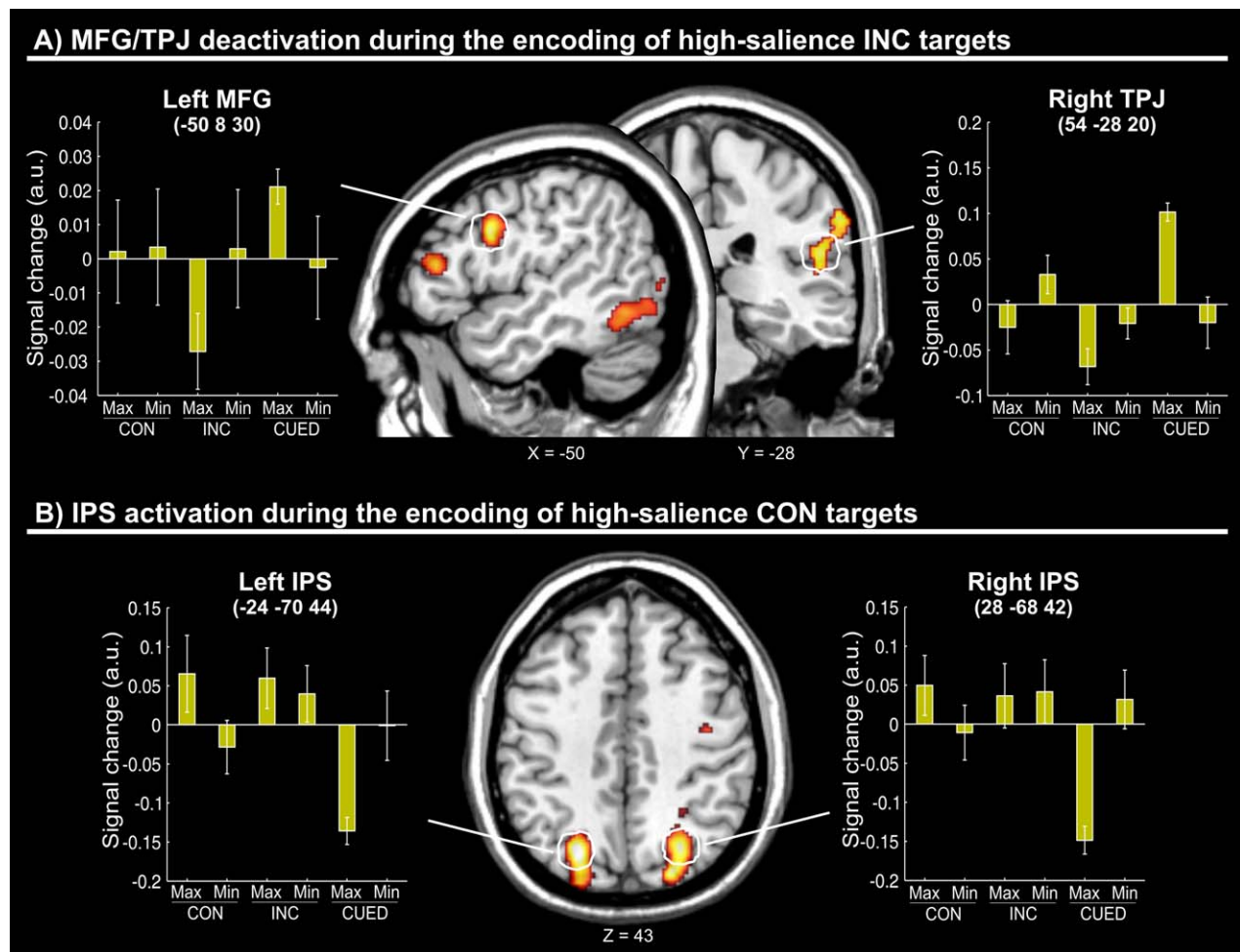
Figure 3.

(**A**) Central panel: sagittal and coronal sections on a standard MNI template showing the anatomical location of the left MFG and the right TPJ, with the corresponding "overall activation map" (p-unc = 0.001, for display purposes) and the "overall de-activation map" (p-FWE-corr. = 0.05); cf. also Figure 2. Left/right panels: bar plots summarizing the activity at encoding of the right TPJ and left MFG that showed a significant context × saliency interaction. These plots show that in these regions the activity decreased during the encoding of INC targets that were later successfully remembered, but only when they were at maximal salience (compare bar 3 vs. bar 4). (**B**) Central panel: axial section showing the location of the right and left IPS clusters that showed a significant context x saliency interaction (p-unc = 0.001, for display purposes). Left/right panels: signal plots showing the activity of the left and right IPS for all the conditions. These plots indicate a selective activation of the right IPS during the encoding of CON targets that were later successfully remembered, but only when they were at maximal salience (compare bar 1 vs. bar 2). All signal plots show activity within 10 mm spheres (ROIs) centred on the relevant regions (cf. Table I). The error bars represent the standard error of mean. CON/INC/CUED: congruent, incongruent, cued context; min/max: maximal, minimal salience.

semantics factors contribute to a common processes of spatial selection, here jointly allowing the formation of a reliable internal representation of naturalistic scenes.

Most importantly, the current findings revealed for the first time the neural underpinnings of the interplay between contextual- and salience-related factors during WM encoding of natural scenes. Our imaging data showed that the encoding of context–incongruent targets that were later confidently remembered de-activated the inferior parietal cortex, with a significant interaction between context and salience signals in the right TPJ and in the left MFG. In contrast, the encoding of salient targets in context–congruent conditions was found to activate the IPS in the dorsal parietal cortex.

The current findings in the dorsoparietal cortex are consistent with our previous work demonstrating the role of

sensory salience during WM encoding using naturalistic scenes [Santangelo and Macaluso, 2013]. Here, the engagement of the intraparietal sulcus was specific for the context–congruent condition that required search for potential target objects without any additional contextual information about the possible target object. In the absence of such high-level signals, we suggest that the dorsal parietal cortex assigns attentional priorities based on low-level sensory features [Gottlieb et al., 2007; Nardo et al., 2011], which in turns increases the likelihood of the salient/ attended object to be stored in visuospatial WM. This is also consistent with the current eye-movements results, showing a memory advantage at retrieval for high-salience congruent targets that were fixated more during the encoding phase.

The main novelty of this study concerns the finding that the ventroparietal cortex also contributes to the encoding of object-related spatial information in WM by making use of both top-down, context-related and bottom-up, saliency-related signals. The right TPJ de-activated more during the encoding of objects in the incongruent than congruent context. Moreover, these regions showed an interaction between context and saliency with the largest de-activation when the out-of-context object that was successfully encoded was also at a salient location (Fig. 3A).

The link between the ventral parietal cortex and WM encoding has been suggested by a few previous imaging studies. Anticevic et al. [2010] reported that the level of the right TPJ de-activation is critical during memory encoding, as it was found to predict performance in a delayed WM task that included the presentation of distracting stimuli during the maintenance phase. The right inferior parietal cortex has been traditionally associated with stimulus-driven spatial orienting [Corbetta and Shulman, 2002] and with the filtering-out of distractors during search tasks [Corbetta et al., 2008; Shulman et al., 2003, 2007]. The relationship between these processes is still not fully understood. Shulman et al. [2003], [2007] asked subjects to monitor streams of rapid serial visually presented (RSVP) distracting letters for a to-be-detected target. They reported right TPJ de-activation until target detection. De-activation was also found in the prefrontal cortex, including the middle and inferior frontal gyri, bilaterally, though more extensive in the right hemisphere [Shulman et al., 2007]. This was interpreted as a filtering mechanism that would act during search: when attention is fully focused via top-down control, the right TPJ (together with other regions of the "ventral attention system") would de-activate to minimise stimulus-driven reorienting toward irrelevant distractors that compete for processing resources. Here, we did not manipulate competition in the time domain using an RSVP task, but rather we presented natural scenes that entailed high-levels of "spatial" competition between multiple objects presented simultaneously. The de-activation pattern of the right TPJ may reflect selective attention to pools of

objects initially evaluated as potential targets for WM encoding. After the "identification"/selection of these possible targets, the right TPJ de-activation would prevent spatial reorienting toward other stimuli [Shulman et al., 2003], hence preventing overload of WM capacity [Luck and Vogel, 2013].

We should note that in the current study we could not address to what extent the spatial nature of the memory retrieval task, that required "same vs. different location" judgments, contributed to our results during encoding. This would require adding an additional "task" factor to the experimental design, e.g., by changing the target identity rather than the target position during retrieval. While the functioning of frontoparietal attention networks has been primarily addressed in the context of spatial tasks, there is extensive evidence that these regions engage also in paradigms that do not explicitly require any spatial orienting. The dorsal parietal cortex has been found to activate when participants are asked to direct top-down attention in nonspatial tasks [Wojciulik and Kanwisher, 1999], and the ventral parietal cortex has been shown to activate also in tasks requiring temporal rather than spatial orienting [Nobre et al., 1999]. Analogously, studies that specifically investigated WM for space- and object-related information revealed activation of prefrontal and parietal regions irrespective of the specific type of information that participants were asked to select and store [e.g., Corbetta et al., 2005; McCarthy et al., 1996]. Accordingly, we suggest that our findings here may replicate also using non-spatial tasks or memory judgments.

While the right TPJ de-activation during the encoding of incongruent targets can be related to attention filtering mechanisms, the activity in the prefrontal cortex (the left MFG) may reflect the processing of unexpected and novel (i.e., never-experienced) object-context associations in the context–incongruent scenes. Consistently, a recent study [Remy et al., 2014] reported that the processing of out-of-context objects as compared to the processing of incontext objects during a categorisation task induced increased activation in a network of regions, involving both the medial temporal lobe and the prefrontal cortex, including the MFG. This was interpreted as an evidence for additional processing required by novel contextual associations during the categorisation task. We propose that here the activation of the left MFG reflected the encoding novel object–context associations into WM during the viewing of scenes including out-of-context targets.

In addition to these overall effects of context, our results revealed an interplay between context and low-level sensory salience. Both the right TPJ and the left MFG showed a significant interaction between context and saliency, with maximal de-activation when subjects encoded an out-of-context object that was also at a salient location (Fig. 3A). The finding of a joint effect of both knowledge-based and stimulus-driven signals in the right TPJ is in agreement with previous results that linked the activation of this

region with the integration of high-level attention factors and low-level sensory salience. Using simple visual stimuli, Weidner et al. [2008] asked participants to search for a target defined by a singleton feature, i.e., either color or orientation that was semantically precued. Target saliency was varied parametrically in each dimension. They found that cue validity modulated saliency processing in the TPJ, though in the left hemisphere, suggesting a central role played by the inferior parietal cortex in integrating endogenous control related to the "attentional-set" with bottom-up saliency signals [see also Corbetta et al., 2008, for a review; Serences et al., 2005]. Here we extended these findings in terms of memory-related effects during the encoding of naturalistic scenes.

The involvement of the ventral parietal cortex in the encoding of out-of-context objects is also consistent with recent proposals linking the ventral attention system with contextual updating of expectations [Geng and Vossel, 2013; Macaluso and Doricchi, 2013]. Here, context–incongruent targets involved a breach of high-level expectations related to *a priori* knowledge about plausible objects layouts in natural scenes. In the right TPJ, the modulation by bottom-up salience was specific for the encoding of out-of-context targets (INC context) and was not found when the target identity was precued (CUED context). This rules out that these effects merely reflected the detection of a task-relevant stimulus during encoding (note that the out-of-context was always used as the memory target). If this was the case, one would expect a similar pattern of activity in the control CUED trials, where the task-relevance of one specific object—the target—was disclosed before encoding.

To summarise, we showed that the encoding of objects in natural scenes depends on the interplay between context-related information and sensory salience. The imaging findings highlighted a possible neural substrate of this behavioural interaction: when target identification/ selection was driven by contextual information we found encoding-related effects in the ventroparietal cortex, with a modulatory effect of stimulus-driven salience in the right TPJ, plus the MFG in the left prefrontal cortex. We propose that the right TPJ makes use of contextual information and sensory salience to identify, and select against the many co-occurring distractor objects, the most likely target object (i.e., the out-of-context object); while the MFG would support WM encoding via processing of novel object/context associations. In contrast, the encoding of salient targets in the context–congruent condition was found to activate the IPS in the dorsal parietal cortex, which we link to processing-priority biases represented in these regions. We conclude that both context-related information and sensory salience contribute to the selection of stimuli to be encoded in visuospatial WM, and that this occurs via saliency-related modulation of activity in distinct dorsal and ventral nodes of the frontoparietal attention controlling networks.

REFERENCES

Anticevic A, Repovs G, Shulman GL, Barch DM (2010): When less is more: TPJ and default network deactivation during encoding predicts working memory performance. Neuroimage 49:2638–2648.

Botta F, Santangelo V, Raffone A, Olivetti Belardinelli M, Lupianez J (2010): Exogenous and endogenous spatial attention effects on visuo-spatial working memory. Q J Exp Psychol 63: 1590–1602.

Buchel C, Holmes AP, Rees G, Friston KJ (1998): Characterizing stimulus-response functions using nonlinear regressors in parametric fMRI experiments. Neuroimage 8:140–148.

Cabeza R, Ciaramelli E, Olson IR, Moscovitch M (2008): The parietal cortex and episodic memory: an attentional account. Nat Rev Neurosci 9:613–625.

Champod AS, Petrides M (2007): Dissociable roles of the posterior parietal and the prefrontal cortex in manipulation and monitoring processes. Proc Natl Acad Sci USA 104:14837–14842.

Corbetta M, Kincade JM, Ollinger JM, McAvoy MP, Shulman GL (2000): Voluntary orienting is dissociated from target detection in human posterior parietal cortex. Nat Neurosci 3:292–297.

Corbetta M, Patel G, Shulman GL (2008): The reorienting system of the human brain: From environment to theory of mind. Neuron 58:307–324.

Corbetta M, Shulman GL (2002): Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci 3: 201–215.

Corbetta M, Tansy AP, Stanley CM, Astafiev SV, Snyder AZ, Shulman GL (2005): A functional MRI study of preparatory signals for spatial location and objects. Neuropsychologia 43: 2041–2056.

Doricchi F, Macci E, Silvetti M, Macaluso E (2010): Neural correlates of the spatial expectancy components of endogenous and stimulus-driven orienting of attention in the Posner task. Cereb Cortex 20:1574–1585.

Einhäuser W, Spain M, Perona P (2008): Objects predict fixation better than early saliency. J Vis 8:18.1–26.

Fine MS, Minnery BS (2009): Visual salience affects performance in a working memory task. J Neurosci 29:8016–8021.

Foulsham T, Underwood G (2008): What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. J Vis 8:6.1–17.

Friston KJ, Glaser DE, Henson RN, Kiebel S, Phillips C, Ashburner J (2002): Classical and Bayesian inference in neuroimaging: Applications. Neuroimage 16:484–512.

Fuster JM (2006): The cognit: A network model of cortical representation. Int J Psychophysiol 60:125–132.

Gazzaley A, Cooney JW, McEvoy K, Knight RT, D'Esposito M (2005): Top-down enhancement and suppression of the magnitude and speed of neural activity. J Cogn Neurosci 17:507–517.

Gazzaley A, Nobre AC (2012): Top-down modulation: Bridging selective attention and working memory. Trends Cogn Sci 16: 129–135.

Geng JJ, Vossel S (2013): Re-evaluating the role of TPJ in attentional control: Contextual updating? Neurosci Biobehav Rev 27:2608–2620.

Goodman GS (1980): Picture memory: How the action schema affects retention. Cogn Psychol 12:473–495.

Gottlieb J (2007): From thought to action: the parietal cortex as a bridge between perception, action, and cognition. Neuron 53:9–16.

Hahn B, Ross TJ, Stein EA (2006): Neuroanatomical dissociation between bottom-up and top-down processes of visuospatial selective attention. Neuroimage 32:842–853.

Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R (2004): Intersubject synchronization of cortical activity during natural vision. Science 303:1634–1640.

Henderson JM (2003): Human gaze control during real-world scene perception. Trends Cogn Sci 7:498–504.

Henderson JM, Hollingworth A (1999): High-level scene perception. Annu Rev Psychol 50:243–271.

Hollingworth A, Henderson JM (1998): Does consistent scene context facilitate object perception? J Exp Psychol Gen 127:398–415.

Hopfinger JB, West VM (2006): Interactions between endogenous and exogenous attention on cortical visual processing. Neuroimage 31:774–789.

Kincade JM, Abrams RA, Astafiev SV, Shulman GL, Corbetta M (2005): An event-related functional magnetic resonance imaging study of voluntary and stimulus-driven orienting of attention. J Neurosci 25:4593–4604.

Ikkai A, Curtis CE (2011): Common neural mechanisms supporting spatial working memory, attention and motor intention. Neuropsychologia 49:1428–1434.

Indovina I, Macaluso E (2007): Dissociation of stimulus relevance and saliency factors during shifts of visuospatial attention. Cereb Cortex 17:1701–1711.

Itti L, Koch C (2001): Computational modelling of visual attention. Nat Rev Neurosci 2:194–203.

Itti L, Koch C, Niebur E (1998): A model of saliency-based visual attention for rapid scene analysis. IEEE Trans Pattern Anal Mach Intell 20:1254–1259.

Luck SJ, Vogel EK (2013): Visual working memory capacity: From psychophysics and neurobiology to individual differences. Trends Cogn Sci 17:391–400.

Macaluso E, Doricchi F (2013): Attention and predictions: control of spatial attention beyond the endogenous–exogenous dichotomy. Front Hum Neurosci 7:685.

Mack SC, Eckstein MP (2011): Object co-occurrence serves as a contextual cue to guide and facilitate visual search in a natural viewing environment. J Vis 11:1–16.

Majerus S, Bastin C, Poncelet M, Van der Linden M, Salmon E, Collette F, Maquet P (2007): Short-term memory and the left intraparietal sulcus: Focus of attention? Further evidence from a face short-term memory paradigm. Neuroimage 35:353–367.

McCarthy G, Puce A, Constable RT, Krystal JH, Gore JC, Goldman-Rakic P (1996): Activation of human prefrontal cortex during spatial and nonspatial working memory tasks measured by functional MRI. Cereb Cortex 6:600–611.

Melcher D, Piazza M (2011): The role of attentional priority and saliency in determining capacity limits in enumeration and visual working memory. Plos One 6:e29296.

Nardo D, Santangelo V, Macaluso E (2011): Stimulus-driven orienting of visuo-spatial attention in complex dynamic environments. Neuron 69:1015–1028.

Natale E, Marzi CA, Macaluso E (2009): FMRI correlates of visuospatial re-orienting investigated with an attention shifting double-cue paradigm. Hum Brain Mapp 30:2367–2381.

Nobre AC, Coull JT, Frith CD, Mesulam MM (1999): Orbitofrontal cortex is activated during breaches of expectation in tasks of visual attention. Nat Neurosci 2:11–12.

Oliva A, Torralba A (2007): The role of context in object recognition. Trends Cogn Sci 11:520–527.

Peelen MV, Kastner S (2014): Attention in the real world: toward understanding its neural basis. Trends Cogn Sci 18:242–250.

Pezdek K, Whetstone T, Reynolds K, Askari N, Dougherty T (1989): Memory for real-world scenes: The role of consistency with schema expectation. J Exp Psychol Learn Mem Cogn 15:587–595.

Pinker S (2007): The Stuff of Thought: Language as a Window onto Human Nature. New York: Viking Press.

Pooresmaeili A, Bach DR, Dolan RL (2014): The effect of visual salience on memory-based choices. J Neurophysiol 111:481–487.

Preston TJ, Guo F, Das K, Giesbrecht B, Eckstein MP (2013): Neural representation of contextual guidance in visual search of real-world scenes. J Neurosci 33:7846–7855.

Remy F, Vayssiere N, Pins D, Boucart M, Fabre-Thorpe M (2014): Incongruent object/context relationship in visual scenes: Where are they processed in the brain? Brain Cogn 84:34–43.

Santangelo V, Macaluso E (2013): Visual salience improves spatial working memory via enhanced parieto-temporal functional connectivity. J Neurosci 33:4110–4117.

Santangelo V, Olivetti Belardinelli M, Spence C, Macaluso E (2009): Interactions between voluntary and stimulus-driven spatial attention mechanisms across sensory modalities. J Cogn Neurosci 21:2384–2397.

Schmidt BK, Vogel EK, Woodman GF, Luck SJ (2002): Voluntary and automatic attentional control of visual working memory. Percept Psychophys 64:754–763.

Serences JT, Shomstein S, Leber AB, Golay X, Egeth HE, Yantis S (2005): Coordination of voluntary and stimulus-driven attentional control in human cortex. Psychol Sci 16:114–122.

Serences JT, Yantis S (2007): Spatially selective representations of voluntary and stimulus-driven attentional priority in human occipital, parietal, and frontal cortex. Cereb Cortex 17:284–293.

Shulman GL, Astafiev SV, McAvoy MP, d'Avossa G, Corbetta M (2007): Right TPJ deactivation during visual search: functional significance and support for a filter hypothesis. Cereb Cortex 17:2625–2633.

Shulman GL, McAvoy MP, Cowan MC, Astafiev SV, Tansy AP, d'Avossa G, Corbetta M (2003): Quantitative analysis of attention and detection signals during visual search. J Neurophysiol 90:3384–3397.

Silva MM, Groeger JA, Bradshaw MF (2004): Attention–memory interactions in scene perception. Spat Vis 19:9–19.

Spotorno S, Tatler BW, Faure S (2013): Semantic versus perceptual salience in visual scenes: Findings from change detection. Acta Psychol 142:168–176.

Stirk JA, Underwood G (2007): Low-level visual saliency does not predict change detection in natural scenes. J Vis 7:1–10.

Todd JJ, Marois R (2004): Capacity limit of visual short-term memory in human posterior parietal cortex. Nature 428:751–754.

Torralba A, Oliva A, Castelhano MS, Henderson JM (2006): Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. Psychol Rev 113:766–786.

Weidner R, Krummenacher J, Reimann B, Muller HJ, Fink GR (2008): Sources of top-down control in visual search. J Cogn Neurosci 21:2100–2113.

Wojciulik E, Kanwisher N (1999): The generality of parietal involvement in visual attention. Neuron 23:747–764.

Wolfe JM (1994): Guided search 2.0. A revised model of visual search. Psychon Bull Rev 1:202–228.

Wolfe JM, Horowitz TS (2004): What attributes guide the deployment of visual attention and how do they do it? Nat Rev Neurosci 5:495–501.

Wolfe JM, Võ ML, Evans KK, Greene MR (2011): Visual search in scenes involves selective and non-selective pathways. Trends Cogn Sci 15:77–84.

Xu Y, Chun MM (2006): Dissociable neural mechanisms supporting visual short-term memory for objects. Nature 440:91–95.

Yoshida M, Itti L, Berg DJ, Ikeda T, Kato R, Takaura K, White BJ, Munoz DP, Isa T (2012): Residual attention guidance in blind sight monkeys watching complex natural scenes. Curr Biol 22: 1429–1434.