*Cerebral Cortex*, January 2017;27: 162–172

**doi: 10.1093/cercor/bhw380**
Advance Access Publication Date: 2 December 2016
Original Article

ORIGINAL ARTICLE

# Cortical Circuit for Binding Object Identity and Location During Multiple-Object Tracking

Lauri Nummenmaa[1,2], Lauri Oksama[3], Erico Glerean[4,5] and Jukka Hyönä[2]

[1]Turku PET Centre, University of Turku, Turku, Finland, [2]Department of Psychology, University of Turku, Turku, Finland, [3]National Defence University, Helsinki, Finland, [4]Department of Neuroscience and Biomedical Engineering, School of Science, Aalto University, Espoo, Finland and [5]Advanced Magnetic Imaging Centre, Aalto Neuroimaging, School of Science, Aalto University, Espoo, Finland

Address correspondence to Lauri Nummenmaa, Turku PET Centre, Turku University Hospital, FI-20520 Turku, Finland. Email: latanu@utu.fi

## Abstract

Sustained multifocal attention for moving targets requires binding object identities with their locations. The brain mechanisms of identity-location binding during attentive tracking have remained unresolved. In 2 functional magnetic resonance imaging experiments, we measured participants' hemodynamic activity during attentive tracking of multiple objects with equivalent (multiple-object tracking) versus distinct (multiple identity tracking, MIT) identities. Task load was manipulated parametrically. Both tasks activated large frontoparietal circuits. MIT led to significantly increased activity in frontoparietal and temporal systems subserving object recognition and working memory. These effects were replicated when eye movements were prohibited. MIT was associated with significantly increased functional connectivity between lateral temporal and frontal and parietal regions. We propose that coordinated activity of this network subserves identity-location binding during attentive tracking.

**Key words:** attention, eye movements, fMRI, object tracking

## Introduction

Many real-world tasks require sustained attention for moving targets. For example, following a basketball match or overseeing a group of children on a playground requires that we constantly track, index and update the location of each moving target in our working memory. Behavioral studies using the multiple-object-tracking (MOT) paradigm suggest that attentive indexing and tracking of multiple identical objects may occur in parallel (Pylyshyn and Storm 1988; Cavanagh and Alvarez 2005). Yet, when the targets have distinct multiple identities—such as individual children on the playground—target identity needs to be accessed and bound with location information, thus likely introducing a serial component to the multiple identity tracking (MIT) process (Oksama and Hyönä 2004).

Neuroimaging studies have established that a cortical circuit spanning superior and inferior parietal and frontal cortices is involved in attentive tracking of identical targets, and that responses within this circuit are amplified during tracking the locations of the objects versus merely attending to them (Culham et al. 2001; Jovicich et al. 2001; Howe et al. 2009). This circuit bears significant resemblance to those involved in working memory and control of overt and covert attention (see meta-analyses in Grosbras et al. 2005; Owen et al. 2005). Thus, it has been proposed that MOT involves interactions between systems distributing attention over the space and maintaining positions of unattended objects in the visual working memory (Cavanagh and Alvarez 2005). The brain basis of object-identity-location binding during attentive tracking of multiple

objects with distinct identities remains nevertheless poorly understood.

Cognitive models of MOT fall into 2 categories: those assuming parallel versus serial tracking. According to the multifocal attention model (Cavanagh and Alvarez 2005) and the indexing model (Pylyshyn and Storm 1988), tracking can be carried out in parallel. These models postulate independent tracking mechanisms that are capable of tracking object positions either preattentively (the indexing model) or postattentively (the multifocal model). The multifocal attention model proposes that attention can be allocated simultaneously to multiple higher-level foci, thus allowing independent tracking of the moving targets. The limited attentional resources can be allocated covertly between the foci as they track moving targets. Because the foci may also have access to feature information, parallel tracking of specific targets is presumably possible (see also Howe and Ferguson 2015). However, it is still debated whether or not object locations and their features can be accessed in parallel (Pylyshyn and Storm 1988; Kahneman et al. 1992).

In contrast to the parallel models, attentional switching models (Pylyshyn and Storm 1988; Yantis 1992; Oksama and Hyönä 2004, 2008) assume that tracking is based on a single focus of attention, which cycles rapidly between the targets to be tracked. These models predict that the tracking load (i.e. the number of to-be-tracked targets) is closely associated with spatial working memory capacity, reflected in increased number of overt eye movements, covert attention shifts, and cognitive load (as indexed by pupil size; Oksama and Hyönä 2016).

Thus, the issue whether or not MIT and MOT share a common tracking mechanism is still unresolved (Horowitz et al. 2007; Pinto et al. 2010; Cohen et al. 2011). The multifocal model and the object file theory assume a unitary parallel mechanism capable of tracking both objects' positions and features. In contrast, MOMIT (Oksama and Hyönä 2008) and Pylyshyn's indexing theory (Pylyshyn and Storm 1988) posit at least partially separate mechanisms for MOT and MIT. MOMIT's assumption about parallel access to peripheral location information suggests that position tracking during MOT can be carried out in parallel, but identity tracking in MIT should be based on serial attentional switching.

Behavioral studies suggest that position tracking and identity tracking may involve partially distinct cognitive systems, specifically, one for tracking positions and another for tracking identities (Horowitz et al. 2007; Oksama and Hyönä 2016). Yet, all previous functional imaging studies have focused on the MOT task with tracking of identical targets. Consequently, it remains unresolved 1) whether position (MOT) and identity (MIT) tracking share a common resource during attentive tracking and 2) how object identity and location are bound together during the MIT. If location-identity bindings are updated in parallel across the visual field based on the interactions of the dorsal and ventral visual streams (cf. parallel models), MIT and MOT would activate a similar set of frontopariatal regions involved in attentive tracking (Culham et al. 2001; Jovicich et al. 2001; Howe et al. 2009). Eye movement differences between the MIT and MOT tasks should be minimal, yet an increased number of distinct objects in the MIT task would lead to activity increases in the early visual regions, and possibly to some extent in the dorsal attention system due to the extra processing step involving the access of object identities (rather than their positions only). On the contrary, it is possible that identity-location binding requires focal attention which could be established serially and stored in the visual short-term memory when each target is under focal attention (cf. the attention-switching models). Thus, out-dated bindings of moving targets require constant updating under focal attention (Oksama and Hyönä 2004, 2008). Subsequently, in addition to mechanisms involved in attentive tracking, MIT would be manifested in significant activation increases in the lateral prefrontal (lPFC) regions involved in spatial working memory (Owen et al. 1999), and in increased functional connectivity between these systems and regions and those involved in object recognition.

### The current study

Here, we combined functional magnetic resonance imaging (fMRI) with concurrent eye movement recordings to reveal the cortical circuits binding object identity and location while tracking multiple moving objects with distinct (MIT) versus indistinguishable (MOT) identities (Fig. 1). Task load was manipulated by the number of the to-be-tracked objects. We show that a set of inferior and superior parietal, medial and lateral frontal as well as lateral occipital regions involved in working memory, attention control and object recognition are engaged while tracking multiple objects with distinct identities. Moreover, lateral frontal and ventral occipitotemporal regions were uniquely activated in MIT, likely providing the basis for identity-location binding. These effects were dissociated from eye movements, suggesting that they reflect activation of the working memory and covert attentional guidance systems.

## Materials and Methods

### Participants

The Aalto University Ethics Review Board approved the study protocol, and the study was conducted in accordance with the Declaration of Helsinki. Twenty-four adults (5 males, mean age = 28.13 years, SD = 8.07) with normal or corrected to normal vision participated in Experiment 1 and 15 adults (7 males, ages 19–28, mean age = 22.3 years, SD = 2.4) in Experiment 2. Formal power analysis for the experiment would be nontrivial, because our study involved a novel MIT-MOT paradigm not previously used in the fMRI research, and the main statistical approaches involved a novel type of network analysis for which formal a priori power analysis does not exist. Henceforth, we opted for an informal criterion for power by roughly doubling the sample size of similar behavioral experiments yielding robust effect sizes (Oksama and Hyönä 2016) and clearly exceeding the sample sizes in previous fMRI studies on MOT (Culham et al. 2001; Jovicich et al. 2001; Howe et al. 2009). Individuals with a history of neurological or psychiatric disease or current medication affecting the central nervous system were excluded. All participants were compensated for their time and travel costs, and they signed the ethics-committee-approved informed consent forms. In Experiment 2, two participants were excluded because the eye-tracker could not be calibrated prior to the fMRI experiment. Before the fMRI experiment, the participants were tested with a visuo-spatial complex span test of working memory (see Oksama and Hyönä 2004).

### Design for Experiment 1

Figure 1 shows the overview of the task design. The experimental design was a 2 (MIT vs. MOT) × 3 (0 vs. 2 targets vs. 4 targets) fully within-participants design. The stimuli were 8 familiar objects drawn with black outline against a white background. They were selected from a standardized set of black-and-white line drawings (Snodgrass and Vanderwart 1980). Each picture was shown equally often as the target.

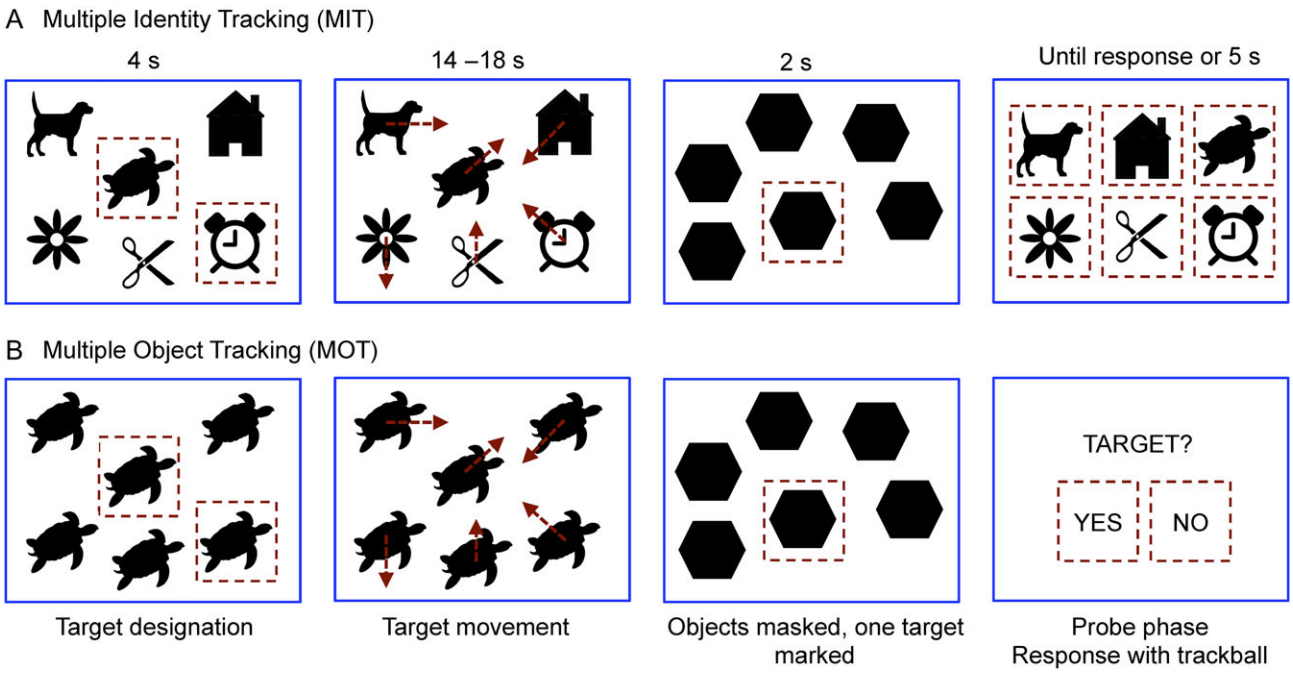


**Figure 1.** Experimental design for MIT (*A*) and MOT (*B*). Participants were first shown an array of items, of which the to-be-tracked items (0, 2, or 4) were flashed. Subsequently, the targets moved around the screen for from 14 to 18 s, after which they were replaced with visual masks, of which one was flashed. In the MIT probe phase, participants were shown all the items from the tracking array and asked to designate which one of them occupied the location of the flashed mask. In the MOT probe phase, participants responded whether or not the location occupied by the flashed mask contained one of the targets. In Experiment 2, a fixation cross stayed on the screen throughout the trials, and the color of the cross indicated whether or not participants were allowed to move their eyes. *Note*: Items shown here are for visualization purposes only and were not used in the experiments; the real task involved 8 different objects.

Participants performed the MIT and MOT tasks with 2 or 4 targets. In the MIT trials, all 8 stimuli were different objects, whereas in the MOT trials all stimuli were identical. A separate control condition with 0 (identical or different) targets with no response demand was also included in the design, thus resulting in a 2 (MIT vs. MOT) × 3 (0, 2, 4 targets) fully within-participants design. At the beginning of each trial, all objects were shown statically on their randomly chosen starting positions for 1 s. Next, 0, 2 or 4 objects were designated as the targets by flashing a frame around them for 3 s. The participants' task was to track these objects during the subsequent from 14 to 18 s tracking phase, or follow the display passively if no targets were designated.

The objects moved around on the screen along linear trajectories (each chosen randomly from cardinal and intercardinal directions) with an average speed of 6.3°/s, the velocity ranging from 2.6 to 10.9°/s. The objects never overlapped with each other or disappeared from the screen; collisions with other objects or screen borders resulted in a "bump" where the object (s) changed its direction of movement. After the tracking phase, the objects were replaced with a mask for 2 s and one object was probed by a flashing frame around it. In the MIT task, the participants were shown all the 8 possible objects, and they had to choose the probed object. In the MOT task, participants had to decide whether or not the probed object was one of the tracked targets. In the 0-target control condition, the participants were simply asked to press the response button. The responses were given with a MRI compatible trackball. The next trial was initiated after the constant 5-s response interval expired. The participants were familiarized with the tasks prior to scanning. The experiment was split up into 3 consecutive 16-min runs. Each run consisted of 30 trials, with 5 trials per condition presented in a random order.

## Design for Experiment 2

As successful MIT performance may require foveal fixations on the targets, differences in eye movement patterns may confound comparisons between MIT and MOT. In Experiment 2, we thus addressed the role of eye movements in multiple-object versus identity tracking. The participants performed the MIT and MOT tasks with 4 targets, while either maintaining a fixation or moving the eyes freely. This resulted in a 2 (MIT vs. MOT) × 2 (Fixate vs. Move) fully within-participants design. The experimental design and trial order were similar to that in Experiment 1 with the following exceptions: The participants always tracked 4 objects. The fixation cross stayed on the screen throughout the trials. At the target designation phase, the fixation cross changed to red or blue to indicate whether participants should maintain a fixation or if they would be allowed to move their eyes during the next trial. The experiment was split up into four 11-min runs, with 5 trials per condition presented in random order in each run.

## Design for Localizer Scans

In Experiment 1, separate localizers were run for working memory and object recognition. In the *n*-back working memory localizer task participants were shown 16-s blocks of single digits at the center of the screen for 1 s each, and their task was to press the response button if the currently shown digit matched the one shown *n* positions back. Before each block, the *n* (0, 2 or 4) for the next trial was shown on the screen. There were altogether 6 blocks with each *n*-condition presented in a random order. In the object recognition, lateral occipital cortex (LOC) localizer, pictures of objects and scrambled objects were both shown in 8 separate 16-s blocks. Each stimulus was shown for 1 s without breaks between the stimuli, and the blocks were separated by a 8-s rest period.

### fMRI Acquisition

MRI was performed with Siemens MAGNETOM Skyra 3-T MRI scanner at the Advanced Magnetic Imaging Centre (Aalto NeuroImaging, Aalto University). Whole-brain data were acquired with $T_2$*-weighted echo-planar imaging (EPI), sensitive to blood-oxygen-level-dependent (BOLD) signal contrast with the following parameters: 33 axial slices, 4-mm slice thickness, TR = 1700 ms, TE = 24 ms, flip angle = 70°, FOV = 256 mm, voxel size 3 × 3 × 4 $mm^3$, ascending interleaved acquisition with no gaps between slices. In Experiment 1, EPI data were acquired from a total of 3 runs with 553 volumes in each and in Experiment 2 from a total of 4 runs with 377 volumes in each. LOC localizer consisted of a single run with 210 volumes and the WM localizer of a single run with 232 volumes. Each run was preceded by 5 dummy volumes to allow for equilibration effects. $T_1$-weighted structural images were acquired at a resolution of 1 × 1 × 1 $mm^3$. Stimuli were delivered using E-prime software (Neurobehavioral Systems, Inc.). They were back-projected on a semitransparent screen using a 3-micromirror data projector (Christie X3, Christie Digital Systems Ltd) and from there via a mirror to the participant.

Data were preprocessed using SPM12 software (http://www.fil.ion.ucl.ac.uk/spm/). The EPI images were realigned to the first scan by rigid body transformations to correct for head movements. EPI and structural images were coregistered and normalized to the $T_1$ standard template in Montreal Neurological Institute (MNI) space (Evans et al. 1994) using linear and nonlinear transformations, and smoothed with a Gaussian kernel of FWHM 8-mm.

### Analysis of Task-Evoked BOLD Responses

Data were analyzed using the conventional 2-stage random effects model. Low-frequency signal drift was removed using a high-pass filter (cutoff 128 s), and AR(1) modeling of temporal autocorrelations was applied. Motion parameters were included in the first-level models to account for motion-related variance.

In Experiment 1, we first modeled the MIT and MOT trials with boxcar functions (only tracking phase was analyzed) in general linear model (GLM), and entered the number of the to-be-tracked objects (2 or 4) as parametric modulators. Subsequently, modeling the parametric effects of the number of targets (i.e. attentional load; Culham et al. 2001) separately for the MIT and MOT conditions revealed brain regions whose hemodynamic responses increase as a function of task demands (i.e. main effects of MIT and MOT). By contrasting these parametric responses to the number of targets in the MIT versus MOT condition, we could thus reveal brain regions showing stronger responses during the MIT and MOT conditions. Because Experiment 2 involved only trials with 4 targets, we used boxcar regressors to model the different task conditions of the 2 × 2 design (MIT vs. MOT × fixate vs. move eyes freely). This enabled us to contrast directly the hemodynamic responses to the MIT versus MOT tasks in the fixate versus move-freely conditions.

In all second-level analyses, participant-wise contrast images were first generated for the contrasts of interest and the second-level analysis used these contrast images in a new GLM and generated statistical images, that is, SPM-t maps. With balanced designs at first level (i.e. similar events for each participant, in similar numbers), this second-level analysis closely approximates a true mixed-effects design, with both within- and between-participant variance. Statistical threshold was set at $P < 0.01$, false discovery rate (FDR) corrected at cluster level.

In Experiment 1, we also conducted complementary region-of-interest (ROI) analyses. The ROIs were defined functionally on the basis of the *n*-back working memory localizers. For the *n*-back working memory localizer, we modeled the onsets of the working memory task trials and entered the *n* of the *n*-back task as a parametric modulator. This yielded a functional definition of ROIs whose activity was linearly dependent on the working memory load. For the LOC localizer, we modeled the object and scrambled object blocks with boxcar regressors, and computed the contrast between these conditions to define ROIs whose voxels were sensitive to the presence of objects. For each ROI, we then computed condition-wise signal changes (%) in the main object tracking experiment using the MarsBaR toolbox (Brett et al. 2002).

### Functional Connectivity

For the functional connectivity analysis, we considered 12 ROIs derived from the aforementioned localizer scans as nodes and computed the connectivity between each pair of nodes (i.e. a link; total number of links = 66) for the 2 tasks (MIT, MOT) with different levels of task load (2, 4, and 2 & 4 combined). To control for head motion confounds on connectivity, we preprocessed the data as described in Power et al. (2012). Briefly, for each participant and each run we extracted a ROI time series as the average of the voxels time series within each ROI. These time series were then z-scored and de-trended linearly. We then regressed out the 6 motion parameters with Volterra expansion and bandpass filtered the data between 0.01 and 0.08 Hz. Quality control was performed using the frame-wise displacement (FD) parameter: all runs for all participants had less than 7% of time points over the recommended threshold of 0.5 mm. All time points were thus retained to ensure that the connectivity values for all tasks/participants/runs were computed over the same number of time points. In all group level connectivity, the average individual FDs were used as nuisance regressors as recommended in Yan et al. (2013). Importantly, differences in FD were not correlated with the task condition and there were no significant differences in FD between the conditions.

To compute the mean task-dependent connectivity for the MIT and MOT tasks (averaged over 2 and 4 targets) for each participant, time points not related to the task of interest were first masked out. Next, a Pearson correlation between pairs of masked ROIs time series was used as the task-dependent connectivity value. Task regressors were shifted by 5 s to account for hemodynamic delay. We tested for the statistical significance of each link by Fisher transforming individual connectivity values averaged across the participants and runs. Statistical significance was assessed with bootstrap resampling over 50 000 iterations. For each iteration, individual surrogate networks were computed by generating surrogate ROI time series with identical magnitude of frequency spectrum and scrambled spectral phase (Laird et al. 2004). At each iteration, a group level surrogate network was computed and the maximum value of correlation was stored to build the maximal statistic distribution (Nichols and Holmes 2002). This effectively controlled for the multiple comparison problem as the link with highest group-average correlation value was stored at each iteration. The final significance level was set by identifying the 95th percentile of the maximal statistic distribution (0.63 for MIT, 0.62 for MOT).

Finally, we compared link strengths between the MIT and MOT conditions for each task load (2 and 4) using paired sample *t* tests. Statistical significance was assessed with permutation testing (Glerean et al. 2016). For each iteration and for each

link, the labels of the task conditions were shuffled randomly and a surrogate *t*-value was computed. Link-wise *P*-values were then estimated from the distribution of the surrogate *t*-values. To control for multiple comparisons, we used the Benjamini-Hochberg False Discovery Rate with $q < 0.05$.

### Eye Movement Recordings

During fMRI, eye movements were recorded with an MRI compatible Eyelink 1000 eye-tracker (SR Research; sampling rate 1000 Hz, spatial accuracy better than 0.5°, with a 0.01° resolution in the pupil-tracking mode). A 9-point calibration and validation was completed at the beginning of each run. Saccade detection was performed using a velocity threshold of 30°/s and an acceleration threshold of 4000°/s$^2$. Eye movement data were manually screened for artifacts. For each experimental condition, the mean number and duration of fixations, saccade amplitude and velocity, number of blinks and pupil size were extracted. To spatially visualize, the fixation distributions across the MIT and MOT tasks, we also constructed fixation heatmaps (Lahnakoski et al. 2014).

## Results

### Experiment 1: Behavioral Performance

Behavioral results are summarized in Figure 2; see Supplementary Table S1 for full results and statistical tests. Tracking was in general accurate (mean = 94%) and the performance between MIT and MOT differed profoundly only in the more difficult 4-item-tracking condition. A higher cognitive load for tracking multiple items was confirmed by a linear increase in pupil size as a function of the number of the to-be-tracked objects. Moreover, pupil size increased more steeply in MIT than in MOT as a function of load. MIT and MOT resulted in different eye fixation patterns, as evidenced by the heatmaps (Fig. 2). Participants made more and longer saccades in MIT than in MOT; this effect was more profound in the 4- versus 2-object-tracking condition. These effects were driven by load-dependent increases in the MIT condition, whereas load did not influence the number of fixations or saccade amplitudes in the MOT condition.

### Experiment 1: Full-Volume Analysis of Task-Evoked Responses

We first assessed the load-dependent activations in MIT and MOT. Both tasks reliably activated wide areas (Fig. 3, top and middle row), including inferior (IPL) and superior (SPL) parietal lobules, precuneus, middle frontal gyrus (MFG), and in precentral gyrys overlapping the typical location of the frontal eye fields (FEF; Paus 1996). When MIT and MOT were contrasted against each other in an interaction analysis (MIT vs. MOT × 2 vs. 4 targets), significant activations were observed in inferior and superior parietal cortices (IPL, SPL), medial and lPFC (MFG, superior frontal gyrus [SFG], supplementary motor area [SMA]) as well as lateral occipital cortices (MT/V5; Fig. 3, bottom row). These activation clusters overlapped almost completely with those observed in the *n*-back working memory task and the LOC localizer. These results were essentially replicated with the conventional *t*-contrasts where the MIT and MOT tasks were contrasted directly against each other in the 2 and 4 object conditions. No effects were observed in the opposite contrasts. Finally, participants' spatial working memory capacity predicted linearly the responses to the MIT versus MOT task in the same regions as observed in the main MIT versus MOT analysis, including IPL/SPL, PREC, middle cingulum, MFG, SFG, and SMA. Additional associations were observed in bilateral postcentral gyri and anterior cingulum; yet, effects in occipital cortex and ventral visual regions were markedly absent.

### Experiment 1: ROI Analysis of Task-Evoked Responses

ROI analyses on the functionally defined working memory (IPL, MFG, SFG, precentral gyrus, SMA) and object recognition (MT/V5) ROIs (Fig. 4) revealed clear regional differences in response profiles. First, only object-recognition regions (FG, MT/V5) showed greater responses during passive MIT versus MOT performance (i.e. attending but not tracking objects with different versus identical objects; see also Supplementary Fig. S2). In the low-load condition (2 targets), responses were stronger during MIT than MOT in all regions except bilateral IPL and SFG. In the high-load condition (4 targets), responses were

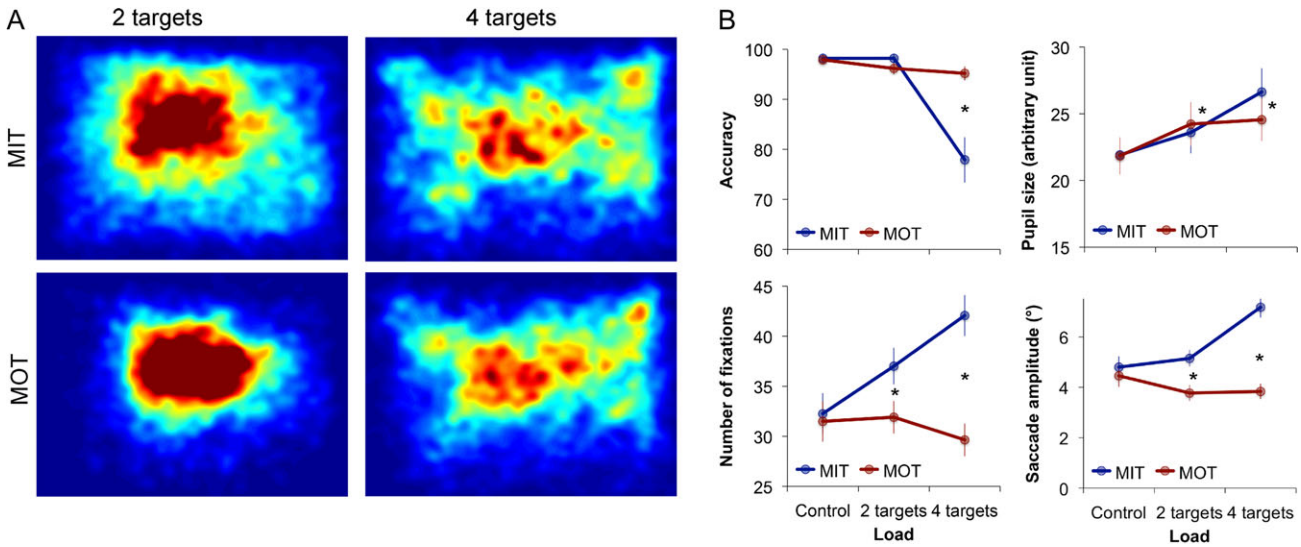


**Figure 2.** Tracking performance and eye movements in Experiment 1. Heatmaps (*A*) show that participants made more fixations in MIT than MOT and in the 4-target than the 2 target condition. This was accompanied with (*B*) a better performance in the 4-target MOT condition, increased cognitive load (indexed by pupil size) as a function of number of targets, and a greater number of fixations and longer saccades in MIT than MOT in the 2 versus 4-target condition.

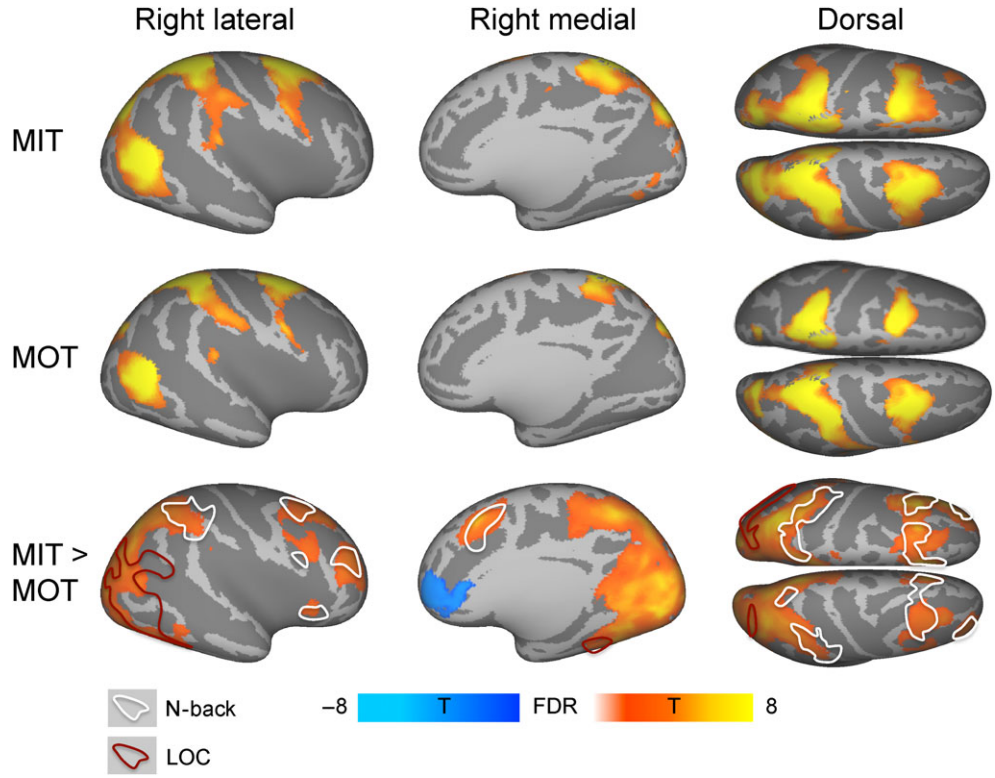


**Figure 3.** Brain regions whose activity increased as a function of the number of targets in MIT (top), MOT (middle), and MIT versus MOT (bottom). Colorbar denotes the *t*-statistic range. The data are plotted at $P < 0.01$, FDR corrected at cluster level (except for MOT vs. MIT which is shown at 0.005, uncorrected). White outline shows areas activated in *N*-back working memory localizer and red outline areas activated in LOC localizer at $P < 0.05$ FWE corrected.

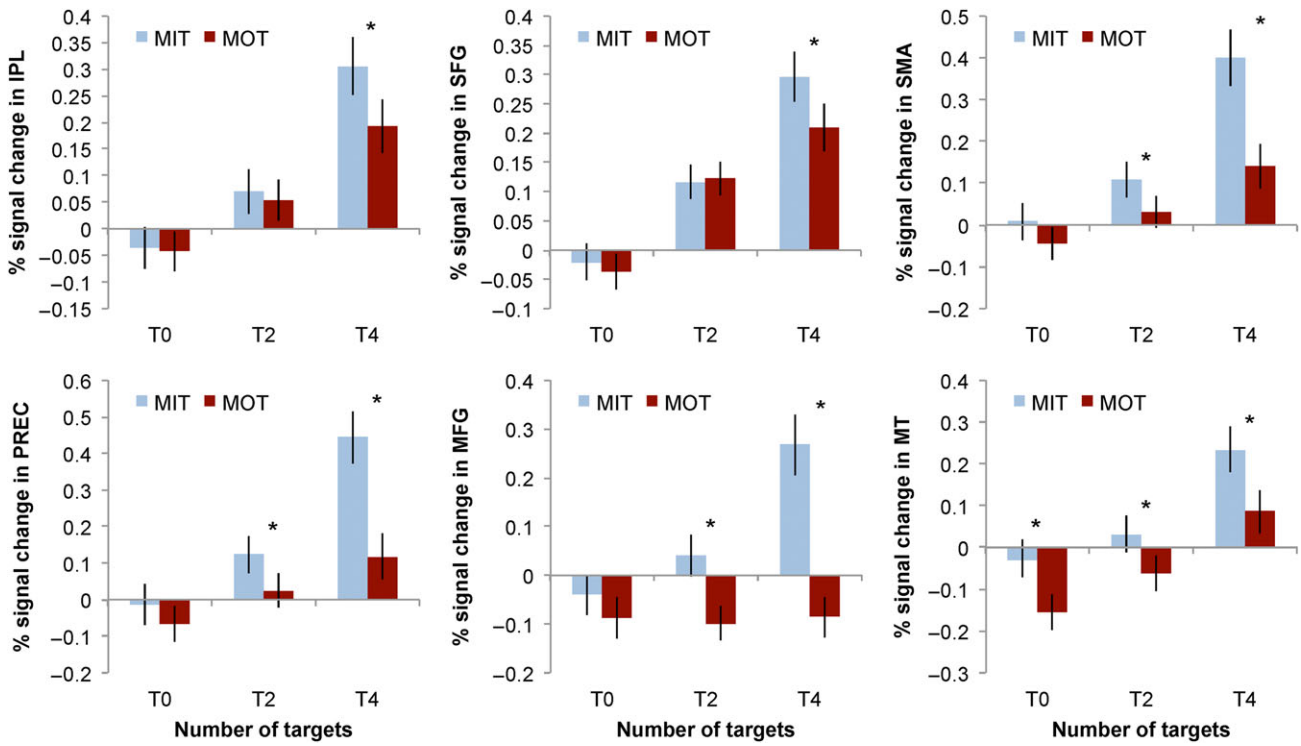


**Figure 4.** Mean (+SEM) signal changes for MIT and MOT in ROIs for *n*-back working memory (IPL, SFG, SMA, PREC, MFG) and object localizer (MT) tasks. Statistically significant ($P < 0.05$, Bonferronni corrected) differences between the MIT and MOT conditions are denoted with asterisks. IPL, inferior parietal lobule; MFG, middle frontal gyrus; SFG, superior frontal gyrus, PREC, precentral gyrus; SMA, supplementary motor area; MT, middle temporal gyrus (area MT).

stronger during MIT than MOT in all regions (all Ps < 0.05, Bonferronni corrected). SMA and bilateral IPL and SFG showed both stronger responses to tracking versus attending (0 vs. 2 targets) and to load-dependent increase in tracking-related activity (from 2 to 4 targets) during both MIT and MOT. In the right-hemispheric MFG, effects of attention versus tracking and load-dependent responses were only observed in MIT but not in MOT. In the left-hemispheric MFG, the effect of load was significant for MIT but not for MOT. As expected, MT and FG responses to simply visually attending to objects compared with object tracking were indistinguishable in the MOT task (except in right MT), whereas in MIT tracking significantly increased activity in all these regions in comparison to merely visually attending to the objects.

### Experiment 1: Functional Connectivity

Figure 5 summarizes the task-dependent connectivity results. During MIT, connectivity was strongest between contralateral ROIs (MT, SFG, IPL) as well as between IPL-SFG and IPL-MFG. Left precentral gyrus (and to some extent also SMA and RSFG) was connected to almost all other ROIs in the network. During MOT, the number of significant links was lower, with strongest connectivity between RIP-RMFG as well as between contralateral areas (SFG, MT, FG, MFG). Again, left precentral gyrus served as the hub of the network. When comparing MIT directly against MOT, no differences were observed at the low-load (2 targets) condition. However, the high-load condition (4 targets) led to significantly stronger connectivity in

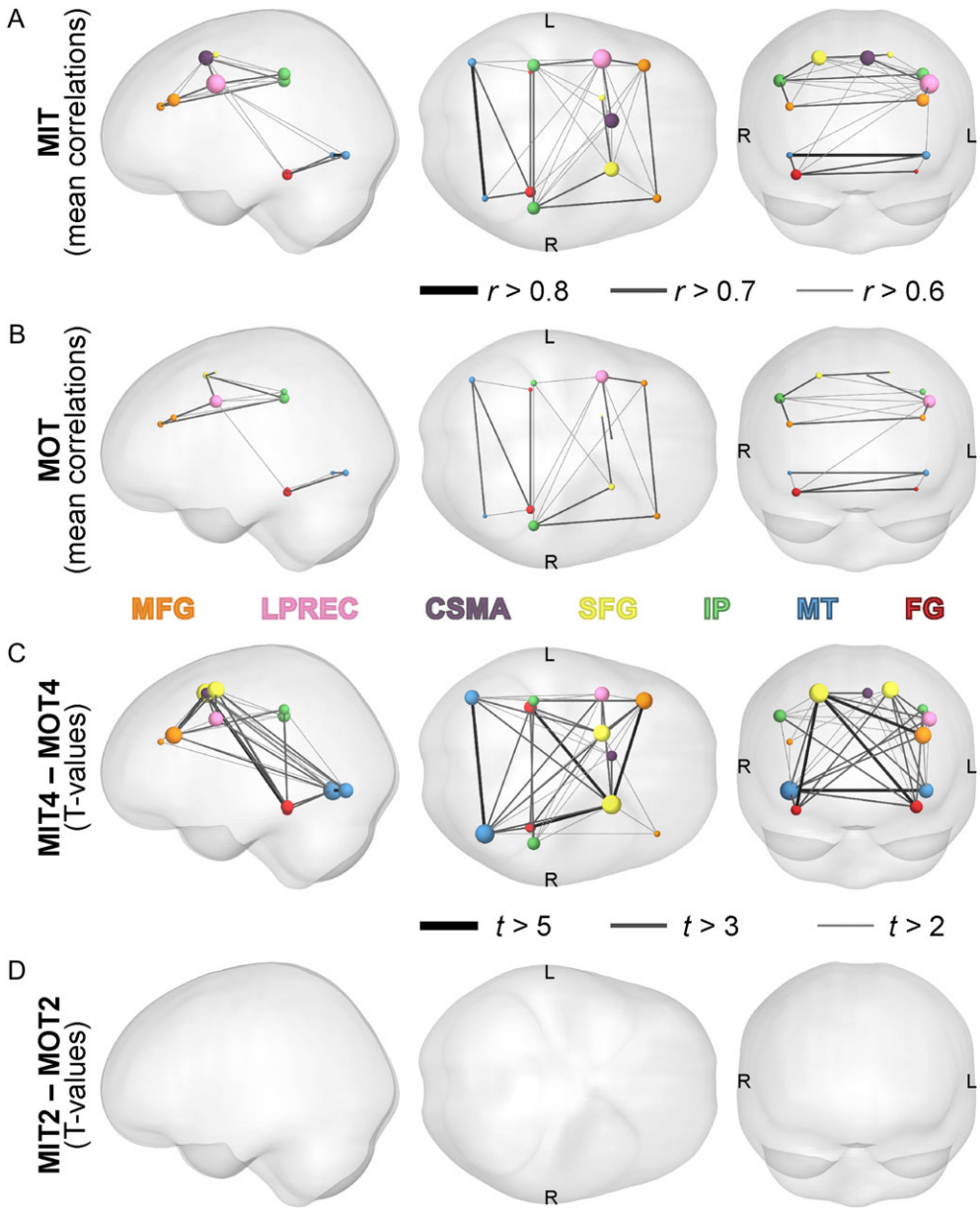


**Figure 5.** Functional connectivity. Mean functional connectivity values across all MIT and MOT trials (averaged over 2 and 4 targets) are shown in panels *A* and *B*. For these plots, line thickness defines the link strength; node color shows the node region and node diameter is proportional to the significant links attached to each node. Panels *C* and *D* show the links whose strength was significantly stronger in MIT than MOT with 4 and 2 targets. For these plots, line thickness defines the *T* value for each link (i.e. link strength comparison between MIT and MOT). All data are thresholded at $P < 0.05$, FDR corrected. MFG, middle frontal gyrus; LPREC, left precentral gyrus; SMA, supplementary motor area; SFG, superior frontal gyrus, IPL, inferior parietal lobule; MT, area MT, FG, fusiform gyrus.

MIT than MOT. Largest shifts were observed in connections from ventral temporal to frontal (FG and MT to SFG) and ventral temporal to parietal (MT and FG to IPL). MFG, SFG and MT were the strongest hubs in this network.

### Experiment 2: Do Eye Movements Explain Differences Between MIT and MOT?

Cortical regions involved in eye movement control (IPS, FEF; see Grosbras et al. 2005) were significantly more active during MIT than MOT; this effect was coupled with an increased fixation rate in the MIT condition. Thus, we ran a control analysis where BOLD responses were predicted with the trialwise fixation count. The results overlaid as white border in Supplementary Fig. S1 reveal significant eye movement dependent responses in the areas activated by MIT versus MOT. Thus, even though the brain basis of overt and covert attention shifts are shared (Awh et al. 2006), and saccadic eye movements are naturally coupled during MIT (Oksama and Hyönä 2016), the current design cannot disentangle the contribution of location-identity binding and eye movements to the hemodynamic responses during MIT. Therefore, we ran Experiment 2 where we directly assessed the contribution of eye movements to 1) behavioral performance and 2) hemodynamic responses in the MIT and MOT tasks. The design was similar to that in Experiment 1 with the exception that the participants always tracked 4 targets in MIT and MOT. On half of the trials, they were instructed to keep fixating at the central fixation cross, whereas on the other half of the trials they were allowed to move their eyes freely. This resulted in a 2 × 2 fully within-participants design with comparable length and statistical power as in Experiment 1.

### Experiment 2: Behavioral Performance

Behavioral results are summarized in Figure 6; see Supplementary Table S2 for full results and statistical tests. Tracking accuracy was again high ($M$ = 85% correct responses), yet restricting eye movements impaired performance and more strongly so in the MIT task. Pupil size was larger in the MIT than MOT task, both when freely moving the eyes and maintaining a fixation. As expected, the instruction to maintain a fixation markedly reduced the number of fixations and the amplitude of saccades. Fixation count was, however, not brought down to zero because maintaining a steady fixation for 15 s is physiologically practically impossible. These fixations resulted mainly from blinks and subsequent corrective saccades whose mean amplitude was 1.3°. Importantly, number of fixations in the MIT and MOT tasks differed only when the participants were able to move their eyes freely, but not when fixation was maintained.

### Experiment 2: fMRI

When the move-freely versus fixation conditions were contrasted with each other, significant bilateral activations were observed in visual cortex and fusiform gyrus as well as in IPL and SPL and precentral gyri. We next validated the finding of Experiment 1 that MIT increased frontoparietal and lateral occipitotemporal activation more than MOT both in the fixation and free viewing condition (Fig. 7); for both conditions this analysis yielded results consistent with those in Experiment 1 (Fig. 3). Next, an interaction analysis (MIT vs. MOT > Free viewing vs. Maintain fixation) was used to reveal areas where increased activation in the MIT versus MOT task would be dependent on eye movements. This revealed only one activation cluster spanning early visual cortex and higher-order associative areas; yet importantly, no effects were observed in the frontal and parietal regions where most profound differences were observed in the MIT versus MOT contrasts in Experiments 1 and 2.

## Discussion

Our results revealed that a fronto-temporo-parietal circuit supports attentive tracking of multiple objects when binding object identity and location. Both MIT and MOT engaged a shared circuit involving inferior and superior parietal and medial and lateral prefrontal cortices. These regions likely support attentive tracking and short-term storage of target positions, which is the shared requirement between the tasks. Accordingly, this set of brain areas had remarkable overlap with those engaged by nonspatial working memory ($n$-back; Fig. 3). Activation in additional lateral frontal and ventral visual regions was,

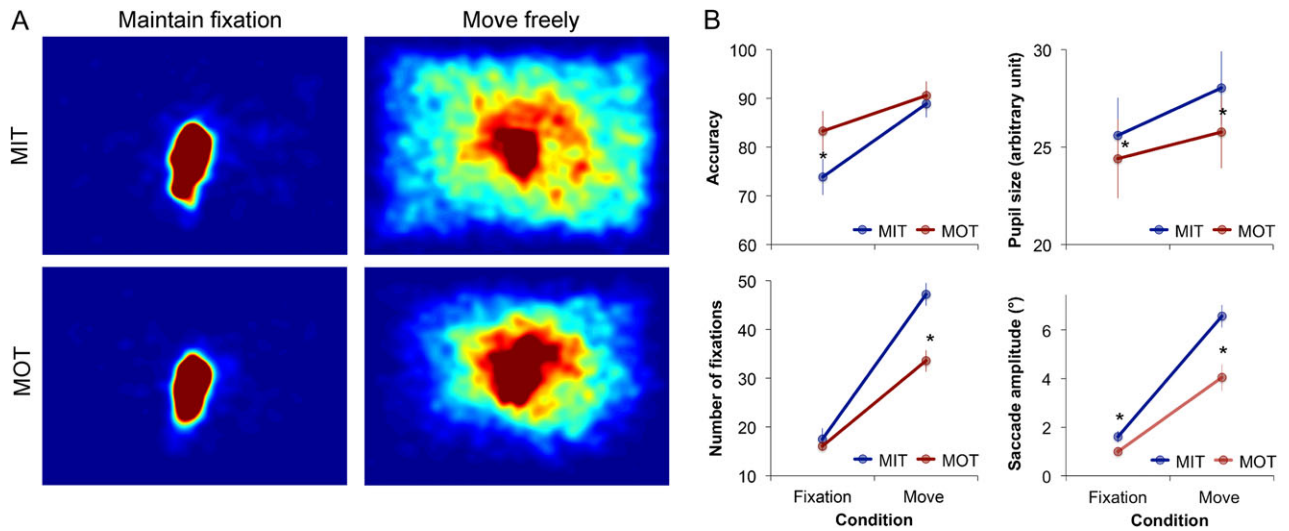


Figure 6. Tracking performance and eye movements in Experiment 2. Heatmaps (A) reveal that the fixation patterns during unconstrained viewing are more widely distributed in MIT than MOT. This effect is abolished when participants are instructed to maintain a fixation. Direct comparisons (B) revealed better performance in MOT than MIT, and increased cognitive load (indexed by pupil size), a greater number of fixations and longer saccades in MIT than MOT.

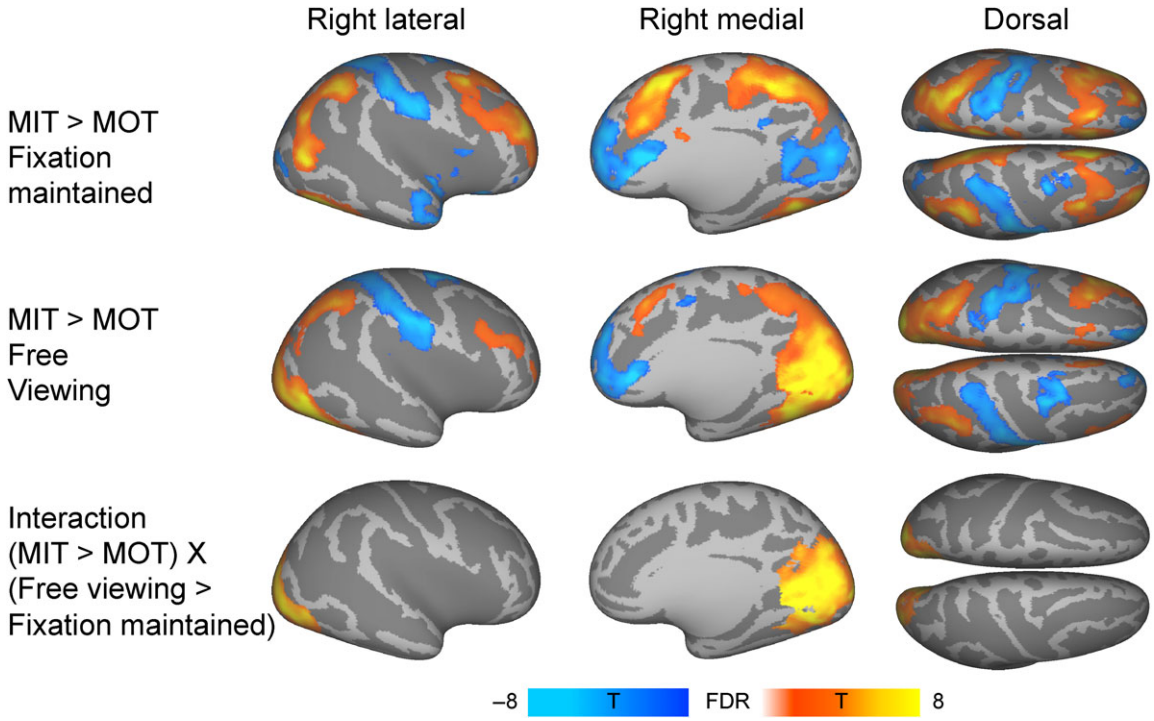


**Figure 7.** Brain regions showing larger response to MIT than MOT and vice versa while maintaining a fixation (top row) or moving eyes freely (middle row). Bottom row shows the regions whose increased activity in the MIT condition was contingent on eye movements. Colorbar denotes the *t*-statistic range. The data are plotted at $P < 0.01$, FDR corrected at cluster level.

however, significantly stronger during MIT than MOT, likely providing the basis for the location-identity binding, which is the additional process required in MIT and absent in MOT. These regional changes were also paralleled in functional connectivity, which was significantly increased during MIT versus MOT between ventral and lateral temporal (FG and MT) visual regions and dorsolateral prefrontal (SFG, MFG) and parietal (IPL) regions involved in attention and working memory.

These effects were replicated in 2 experiments and also when allowing and restricting eye movements. A direct comparison between MIT and MOT while moving eyes freely versus maintaining a fixation only revealed activation differences in the primary visual cortex and ventral temporal regions (FG) involved in object recognition. This confirms that the observed differences between MIT and MOT in the frontoparietal areas do not simply reflect eye movements naturally coupled with the MIT task (Grosbras et al. 2005). Instead, they likely reflect differential activation in the system maintaining target positions and identities in working memory, and deploying attention covertly.

## Independent and Overlapping Circuits for Visual Tracking and Identity-Location Binding

Although MIT and MOT differ from each other with respect to the requirement of identity-location binding, they both share task requirements related to refocusing of attention and updating object locations. Accordingly, MIT and MOT engaged a shared frontoparietal circuit corresponding with that observed in prior studies on attentive tracking (Culham et al. 2001; Jovicich et al. 2001; Howe et al. 2009) but also in those for working memory and eye movements (Grosbras et al. 2005; Owen et al. 2005). In this circuit, responses were generally stronger when tracking than simply attending to the objects (both for MIT and MOT), and activity differences between MIT and MOT became significant only in the high-load condition (4 targets). Despite an identical number of to-be-tracked targets and matched target trajectories, activity within these regions was significantly amplified during the MIT task. These results suggest that MIT and MOT also share a common load-dependent mechanism in the parietal cortices, which is likely to be involved in storing object positions and number of target objects during visual tracking (Culham et al. 2001; Jovicich et al. 2001), yet independent of processing object features (Howe et al. 2009).

While identity and location tracking shared a common frontoparial resource, identity tracking requiring binding object location with its identity resulted in load-dependent activity increase in the medial and lateral PFC during MIT but not during MOT. In particular, lPFC activity remained constant across all MOT conditions. Although the human working memory circuit spans over frontal and parietal cortices, both human and monkey studies have highlighted the role of lPFC in temporary retention of task-dependent information (D'Esposito 2007), particularly in spatial working memory (Owen et al. 1999). It is thus likely that this region supports the binding of object identities with their locations. In addition, ventral temporal (MT and fusiform gyrus) regions critical for object recognition (Bar et al. 2001; Amedi et al. 2005) also showed increased activation for tracking compared with mere attention primarily in the MIT but not in the MOT task, likely reflecting enhanced access to object identities when they are task-relevant.

These increases in regional activity were paralleled by functional connectivity changes. Both the MIT and MOT task increased connectivity between frontal and parietal regions involved in working memory control (MFG and LFG, PFC, IPS,

SFG) as well as connectivity from ventral visual (FG) to lPFC. These connections were, however, significantly stronger during MIT than MOT, and importantly, connections between ventral (FG) and lateral (MT) visual regions and inferior parietal (IPL) and lateral frontal (MFG) cortices were significantly stronger during MIT than MOT. However, this increase was only statistically significant in the high-load condition. The SFG and MFG were the central "hubs" of the object-identity-tracking network, suggesting that they play a key role in integrating the ventral and dorsal visual stream information for establishing identity-location bindings, which involves the integration of object-identity information from the ventral visual stream with location information available in the dorsal stream.

Altogether these results thus favor an intermediate position in the debate regarding unitary versus independent mechanisms for position and identity tracking (Pylyshyn and Storm 1988; Cavanagh and Alvarez 2005; Oksama and Hyönä 2016), in that tracking of multiple objects with identical versus distinct identities involve both shared and distinct neural systems. The shared component consists of frontoparietal areas involved in working memory and visual attention, likely supporting attentive tracking. The shared system is, however, loaded more heavily during MIT. On the other hand, medial frontal regions were uniquely activated during MIT, and their interconnectivity with the attentive tracking system and areas involved in object recognition likely supports identity-location binding during MIT.

Responses in the inferior parietal cortices during MIT versus MOT were linearly dependent on participants' working memory capacity, in line with prior studies on capacity-limited short-term information storage in the parietal cortex (Todd and Marois 2004). Previous studies, however, suggest that this limit is only seen for "what" and not for "where" information (Harrison et al. 2010). The load in the capacity-limited system supporting attentional refocusing and updating of location nevertheless increases significantly more when the location-identity bindings must be constantly updated during MIT. Higher cognitive load during MIT than MOT was also reflected in pupil size, a well-known index of cognitive effort (Granholm et al. 1996; Alnaes et al. 2014; Oksama and Hyönä 2016). Consequently, the inferior parietal and lateral frontal regions may act as the visuo-spatial sketchpad storing the number and the location of objects, while the episodic buffer is responsible for representing more complex features of objects, including the velocity of moving objects.

Cognitive models of MOT have proposed that MOT is carried out by a primarily parallel mechanism with limited resources (Pylyshyn and Storm 1988), whereas MIT would be strictly serial, requiring constant focal updating of object locations and identities (Oksama and Hyönä 2004, 2008). In support of the serial model, we found that 1) attentional load resulting from multiple targets increased eye movement rate more during MIT, 2) more fixations were made during MIT than MOT, and 3) restricting eye movements and increasing the number of tracked objects significantly impaired the MIT performance but also to some extent the MOT performance. Yet, both tasks could be accomplished in the absence of eye movements suggesting that both parallel and serial updating processes are in operation during MIT and MOT, and that even serial attention shifts during MIT may be executed covertly without moving the eyes.

Finally, we would like to note that the number of visually distinct identities varied between the MIT and MOT conditions, which could be considered a potential confound. We deliberately chose this approach because it simulates real-world identity tracking, which typically involves targets with distinct identities, such as different football players on the field. Yet, only passively attending to multiple distinct versus identical objects (i.e. the 0-target condition) only increased activity in the ventral visual cortex (see Supplementary Fig. S2), whereas attentive MIT and MOT tracking led to additional activation changes in the frontal and parietal regions. Thus, we argue that the increased ventral temporal responses in MIT reflect the increased number of distinct identities, whereas the frontoparietal activation changes are related to the actual tracking and updating process. It is noteworthy that human observers are capable of identity tracking even with visually identical objects whose identities are only established prior to the actual tracking (Pylyshyn 2004). Whether or not this type of task can be accomplished without significant contribution of the ventral visual cortex (Ganis et al. 2004; Reddy et al. 2010) needs to be established in future studies.

## Conclusions

We conclude that a frontoparietal cortical circuit supports attentive tracking of multiple objects with both shared and distinct identities. Identity-location binding requires focal (either covert or overt) attention, and object locations and identities are maintained in working memory when not under the attentional spotlight. When multiple identities need to be tracked, activity and interconnectivity of this network and additional ventral visual and lPFC areas are significantly increased, with lateral and medial prefrontal regions serving as "hubs" integrating the "what" and "where" information processed by the medial temporal and parietal cortices.

## Authors' contributions

L.N., L.O., and J.H. designed the experiments, L.O. acquired the data, L.N., L.O., and E.G. analyzed the data, L.N., L.O., E.G., and J. H. wrote the paper.

## Supplementary Material

Supplementary data are available at *Cerebral Cortex* online..

## Funding

Academy of Finland (MIND program grant #265917 to L.N. and grant #273413 to L.O.) and European Research Council (Starting Grant #313000).

## Notes

We thank Anna Anttalainen and Marita Kattelus for her help with the data acquisition and Satu Arola for her help in the data analysis. *Conflict of Interest*: The authors declare no competing financial interests.

## References

Alnaes D, Sneve MH, Espeseth T, Endestad T, de Pavert S, Laeng B. 2014. Pupil size signals mental effort deployed during multiple object tracking and predicts brain activity in the dorsal attention network and the locus coeruleus. J Vis. 14:20.

Amedi A, von Kriegstein K, van Atteveldt NM, Beauchamp MS, Naumer MJ. 2005. Functional imaging of human crossmodal identification and object recognition. Exp Brain Res. 166: 559–571.

Awh E, Armstrong KM, Moore T. 2006. Visual and oculomotor selection: links, causes and implications for spatial attention. Trends Cogn Sci. 10:124–130.

Bar M, Tootell RBH, Schacter DL, Greve DN, Fischl B, Mendola JD, Rosen BR, Dale AM. 2001. Cortical mechanisms specific to explicit visual object recognition. Neuron. 29:529–535.

Brett M, Anton J-L, Valabregue R, Poline J-B. 2002. Region of interest analysis using an SPM toolbox. 8th International Conference on Functional Mapping of the Human Brain; Sendai, Japan.

Cavanagh P, Alvarez GA. 2005. Tracking multiple targets with multifocal attention. Trends Cogn Sci. 9:349–354.

Cohen MA, Pinto Y, Howe PDL, Horowitz TS. 2011. The what-where trade-off in multiple-identity tracking. Atten Percept Psychophys. 73:1422–1434.

Culham JC, Cavanagh P, Kanwisher NG. 2001. Attention response functions: characterizing brain areas using fMRI activation during parametric variations of attentional load. Neuron. 32:737–745.

D'Esposito M. 2007. From cognitive to neural models of working memory. Phil Trans B. 362:761–772.

Evans AC, Collins DL, Mills SR, Brown ED, Kelly RL, Peters TM. 1993. 3D statistical neuroanatomical models from 305 MRI voumes. In Klaisner LA. ed. Nuclear Science Symposium & Medical Imaging Conference, Vols 1-3: 1993 Ieee Conference Record. New York: I E E E. p 1813–1817.

Ganis G, Thompson WL, Kosslyn SM. 2004. Brain areas underlying visual mental imagery and visual perception:an fMRI stydy. Cognit Brain Res. 20:226–241.

Glerean E, Pan RK, Salmi J, Kujala R, Lahnakoski JM, Roine U, Nummenmaa L, Leppämäki S, Nieminen-von Wendt T, Tani P, et al. 2016. Reorganization of functionally connected brain subnetworks in high-functioning autism. Hum Brain Mapp. 37:1066–1079.

Granholm E, Asarnow RF, Sarkin AJ, Dykes KL. 1996. Pupillary responses index cognitive resource limitations. Psychophysiology. 33:457–461.

Grosbras MN, Laird AR, Paus T. 2005. Cortical regions involved in eye movements, shifts of attention, and gaze perception. Hum Brain Mapp. 25:140–154.

Harrison A, Jolicoeur P, Marois R. 2010. "What" and "where" in the intraparietal sulcus: an fMRI study of object identity and location in visual short-term memory. Cereb Cortex. 20:2478–2485.

Horowitz TS, Klieger SB, Fencsik DE, Yang KK, Alvarez GA, Wolfe JM. 2007. Tracking unique objects. Percept Psychophys. 69: 172–184.

Howe PD, Horowitz TS, Morocz IA, Wolfe J, Livingstone MS. 2009. Using fMRI to distinguish components of the multiple object tracking task. J Vis. 9:11.

Howe PDL, Ferguson A. 2015. The identity-location binding problem. Cogn Sci. 39:1622–1645.

Jovicich J, Peters RJ, Koch C, Braun J, Chang L, Ernst T. 2001. Brain areas specific for attentional load in a motion-tracking task. J Cogn Neurosci. 13:1048–1058.

Kahneman D, Treisman A, Gibbs BJ. 1992. The reviewing of object files - object-specific integration of information. Cogn Psychol. 24:175–219.

Lahnakoski JM, Glerean E, Jääskeläinen IP, Hyönä J, Hari R, Sams M, Nummenmaa L. 2014. Synchronous brain activity across individuals underlies shared psychological perspectives. Neuroimage. 100:316–324.

Laird AR, Rogers BP, Meyerand ME. 2004. Comparison of Fourier and wavelet resampling methods. Magn Reson Med. 51: 418–422.

Nichols TE, Holmes AP. 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum Brain Mapp. 15:1–25.

Oksama L, Hyönä J. 2004. Is multiple object tracking carried out automatically by an early vision mechanism independent of higher-order cognition? An individual difference approach. Vis Cogn. 11:631–671.

Oksama L, Hyönä J. 2008. Dynamic binding of identity and location information: a serial model of multiple identity tracking. Cogn Psychol. 56:237–283.

Oksama L, Hyönä J. 2016. Position tracking and identity tracking are separate systems: evidence from eye movements. Cognition. 146:393–409.

Owen AM, Herrod NJ, Menon DK, Clark JC, Downey S, Carpenter TA, Minhas PS, Turkheimer FE, Williams EJ, Robbins TW, et al. 1999. Redefining the functional organization of working memory processes within human lateral prefrontal cortex. Eur J Neurosci. 11:567–574.

Owen AM, McMillan KM, Laird AR, Bullmore ET. 2005. N-back working memory paradigm: a meta-analysis of normative functional neuroimaging. Hum Brain Mapp. 25:46–59.

Paus T. 1996. Location and function of the human frontal eye-field: a selective review. Neuropsychologia. 34: 475–483.

Pinto Y, Howe PDL, Cohen MA, Horowitz TS. 2010. The more often you see an object, the easier it becomes to track it. J Vis. 10:15.

Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. 2012. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage. 59:2142–2154.

Pylyshyn ZW. 2004. Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. Vis Cogn. 11:801–822.

Pylyshyn ZW, Storm RW. 1988. Tracking multiple independent targets: evidence for a parallel tracking mechanism. Spat Vis. 3:179–197.

Reddy L, Tsuchiya N, Serre T. 2010. Reading the mind's eye: decoding category information during mental imagery. Neuroimage. 50:818–825.

Snodgrass JG, Vanderwart M. 1980. A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity and visual complexity. J Exp Psychol Hum Learn Memory. 6:174–215.

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

Yan C-G, Cheung B, Kelly C, Colcombe S, Craddock RC, Di Martino A, Li Q, Zuo X-N, Castellanos FX, Milham MP. 2013. A comprehensive assessment of regional variation in the impact of head micromovements on functional connectomics. NeuroImage. 76:183–201.

Yantis S. 1992. Multielement visual tracking - attention and perceptual organization. Cogn Psychol. 24:295–340.