

## RESEARCH ARTICLE

WILEY

# Distinct neural substrates for visual short-term memory of actions

Ying Cai<sup>1,2</sup> | Zhisen Urgolites<sup>3</sup> | Justin Wood<sup>4</sup> | Chuansheng Chen<sup>5</sup> |  
Siyao Li<sup>1,2</sup> | Antao Chen<sup>6</sup> | Gui Xue<sup>1,2</sup>

<sup>1</sup>State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute of Brain Research, Beijing Normal University, Beijing 100875, People's Republic of China

<sup>2</sup>Center for Collaboration and Innovation in Brain and Learning Sciences Beijing Normal University, Beijing 100875, People's Republic of China

<sup>3</sup>Department of Psychiatry, University of California, San Diego La Jolla, California 92093

<sup>4</sup>Department of Psychology, University of Southern California, Los Angeles, California 90089

<sup>5</sup>Department of Psychology and Social, Behavior University of California, Irvine, California 92697

<sup>6</sup>School of Psychology, Southeast University, Chongqing, 400700, People's Republic of China

## Correspondence

Gui Xue, State Key Laboratory of Cognitive Neuroscience and Learning & IDG/McGovern Institute of Brain Research, Beijing Normal University, Beijing 100875, People's Republic of China.  
Email: guixue@gmail.com

## Funding information

National Science Foundation of China, Grant/Award Number: 31730038; 973 Program, Grant/Award Number: 2014CB846102; 111 Project, Grant/Award Number: B07008; NSFC and the German Research Foundation (DFG) Joint Project, Grant/Award Number: NSFC 61621136008/DFG TRR-169; Guangdong Pearl River Talents Plan Innovative and Entrepreneurial Team, Grant/Award Number: #2016ZT06S220

## Abstract

Fundamental theories of human cognition have long posited that the short-term maintenance of actions is supported by one of the “core knowledge” systems of human visual cognition, yet its neural substrates are still not well understood. In particular, it is unclear whether the visual short-term memory (VSTM) of actions has distinct neural substrates or, as proposed by the spatio-object architecture of VSTM, shares them with VSTM of objects and spatial locations. In two experiments, we tested these two competing hypotheses by directly contrasting the neural substrates for VSTM of actions with those for objects and locations. Our results showed that the bilateral middle temporal cortex (MT) was specifically involved in VSTM of actions because its activation and its functional connectivity with the frontal-parietal network (FPN) were only modulated by the memory load of actions, but not by that of objects/agents or locations. Moreover, the brain regions involved in the maintenance of spatial location information (i.e., superior parietal lobule, SPL) was also recruited during the maintenance of actions, consistent with the temporal-spatial nature of actions. Meanwhile, the frontoparietal network (FPN) was commonly involved in all types of VSTM and showed flexible functional connectivity with the domain-specific regions, depending on the current working memory tasks. Together, our results provide clear evidence for a distinct neural system for maintaining actions in VSTM, which supports the core knowledge system theory and the domain-specific and domain-general architectures of VSTM.

## KEYWORDS

visual short-term memory, actions, fMRI, functional connectivity

## 1 | INTRODUCTION

Many types of information (e.g., objects, locations, and actions) are encountered in everyday life and their processing and memory are critical for individuals' survival and well-being. Researchers have historically focused on the processing and memory of objects and locations but paid less attention to actions. Recent studies, however, revealed that

the processing and memory of actions may be linked to mental disorders and normal variations in social interactions. For example, difficulties in perceiving actions have been found to be related to multiple mental disorders including schizophrenia (Thakkar, Peterman, & Park, 2014) and autism (Pokorny et al., 2015; Von Hofsten & Rosander, 2012). Among normal subjects, visual short-term memory (VSTM) of actions was significantly associated with social emotions such as

empathy (Gao, Ye, Shen, & Perry, 2016). Therefore, it is important to explore the neural mechanisms of VSTM of actions, which unfortunately are poorly understood. In particular, it is not clear whether VSTM of actions shares the same neural substrates as those of objects or spatial locations.

Many existing studies have revealed robust dissociations between spatial information and object identity information in VSTM. For instance, the working memory capacities of objects and spatial information are independent of each other, and there are significant costs for binding visual and spatial information together in memory (Hollingworth, 2007; Klauer & Zhao, 2004; Smyth & Scholey, 1994; Wood, 2011b). Moreover, object and spatial representations in VSTM are supported by different neural substrates, that is, the ventral and dorsal visual pathways, respectively (Haxby et al., 1991; Postle, Stern, Rosen, & Corkin, 2000). Because actions consisted of an agent (or agents, which are objects) performing a series of biological movements, conveying continuous temporal-spatial change (Alibali, 2005), it is thus possible that VSTM of actions is supported by neural regions underlying the VSTM of objects and spatial locations.

Unlike the spatio-object architecture of VSTM, an alternative perspective comes from studies of types of core knowledge in adults, infants, and nonhuman animals (Carey, 2009; Hauser & Spelke, 2004; Spelke, 2000). The “core knowledge” perspective has a rich research tradition that seeks to characterize the fundamental psychological mechanisms for human cognition. According to Spelke and Kinzler (2007), a core knowledge system is domain-specific (representing a particular kind of entity), task-specific (addressing specific questions about the world), and encapsulated (showing no interference across different systems). They have identified five core knowledge systems: objects and their interactions, agents and their actions, sets and their numerical relations, locations and their geometric relations, and ingroup/outgroup recognition. Recent studies on VSTM have focused on three separate systems: an object/agent recognition system that maintains identity information (what/who is involved in the event), a place recognition system that maintains location information (where the event takes place), and an object tracking system that maintains action information (what happens; Wood, 2011a). Preliminary support for this “core knowledge” architecture of VSTM has come from experiments that found independent storage capacity limits for actions, objects, and locations (Shen, Gao, Ding, Zhou, & Huang, 2014; Wood, 2007, 2008, 2011a) and from studies that showed impaired performance of VSTM of actions when they were bound with agents or locations (Ding et al., 2015; Wood, 2008). This core knowledge architecture in VSTM is also in line with the sensory-recruitment theory of VSTM, which posits that the stimulus-specific sensorimotor cortices (i.e., regions involved in perception) are also engaged in the temporary maintenance of the perceived representations, whereas the fronto-parietal network (FPN) is involved in allocating attention to the maintained representations (D'Esposito & Postle, 2015). Taken together the “core knowledge theory” and the “sensorimotor recruitment” model, one could hypothesize that there would be domain-specific neural bases for the maintenance of VSTM of actions, objects, and locations, but a domain-general attention control mechanism for all three types of VSTM.

To date, however, this hypothesis has not been directly tested at the neural level. Neuroimaging studies have implicated several brain regions in the processing of actions. The middle temporal cortex (MT) is involved in the processing of both meaningful and meaningless actions (Grèzes, Costes, & Decety, 1999; Rumiati et al., 2005). The bilateral presupplementary motor areas (SMC) and inferior frontal gyrus (IFG) have also been implicated in action understanding (Iacoboni et al., 2005), action imitation, and motor planning (Buccino et al., 2004; Rizzolatti, 2005). An extended frontoparietal network forms the “mirror neuron system” (MNS) or, more broadly, the “action observation network” (Cross, Kraemer, Hamilton, Kelley, & Grafton, 2009; Lingnau & Downing, 2015; Pokorný et al., 2015). Whether and to what extent these regions are engaged in the maintenance of actions in VSTM is less clear. Given that actions contain continuous biological motion of agents (or parts of agents), the MT might play a role in maintaining actions in VSTM. Supporting the role of MT in maintaining motion information, transcranial magnetic stimulation on MT impairs the VSTM of motion of a moving dot (Silvanto & Cattaneo, 2010). Other studies using multivariate pattern analysis (MVPA) have successfully decoded the specific motion direction during the delay period in MT areas (Emrich, Riggall, LaRocque, & Postle, 2013; LaRocque, Riggall, Emrich, & Postle, 2017). Nevertheless, in another study of action VSTM using point-light motion animations, it has been found that the frontal and parietal lobules were involved in maintaining action during the delay (Lu et al., 2016). As this study did not include the VSTM task for objects and spatial locations, it remains unclear if these regions are specific to VSTM of actions or they are shared by VSTM of other stimuli (or aspects of stimuli) such as objects/agents or spatial locations.

In this study, we conducted two fMRI experiments to directly compare the neural substrates for VSTM of actions with those of objects/agents and locations. In Experiment 1, participants were asked to perform change-detection tasks for three types of visual stimuli (human actions represented by mini-movies, irregular 3D geometrical objects, and locations marked by white dots) to identify action-specific brain regions and functional connectivity. To further elucidate the neural substrates underlying VSTM encoding and maintenance, we performed Experiment 2 in which the participants were asked to encode exactly the same action, agent, and location information (mini movies of various agents playing a series of actions in different locations). A retro-cue was used to specify which type of information to maintain during a long delay period. We hypothesized that the brain regions involved in the processing of actions in Experiment 1 would be specifically involved in the maintenance (not the encoding) of actions in VSTM in Experiment 2. We further predicted that action maintenance would share some neural substrates with the maintenance of spatial locations because actions involve spatial information. Finally, the frontoparietal network, as the region for attention control, should be involved in the maintenance of all three types of information in VSTM.

## 2 | MATERIALS AND METHODS

### 2.1 | Participants

Fourteen subjects (7 females; mean age = 22.1 years) participated in Experiment 1. Two additional subjects were excluded due to the

misunderstanding of instructions (more than 70% trials with no response). The first run from another subject was also discarded for the same reason. Twenty-one subjects (14 females; mean age = 22.57 years old) participated in Experiment 2. Five additional subjects were excluded due to poor behavioral performance (the memory capacities were  $<0.8$  in two or more conditions). The last run from another subject was discarded due to a technical problem. All subjects had no history of neurological or psychiatric problems. Informed written consents were obtained from the subjects before the experiments. The fMRI studies were approved by the institutional review boards of the School of Psychology at Southwest University and the State Key Laboratory of Cognitive Neuroscience and Learning at Beijing Normal University.

## 2.2 | Materials and procedures

A change-detection task was used in Experiment 1 using mini-movies of actions of human figures, irregular 3D objects, and spatial locations marked by white dots (Wood, 2011a; Figure 1a). For each kind of material, eight highly discriminable items were used. Three-dimensional objects were irregular geometric objects ( $6^\circ \times 6^\circ$ ) in the same color (dark green). The locations were marked by white dots ( $1^\circ$  in radius) and were set in an invisible  $5 \times 5$  grid ( $15^\circ \times 15^\circ$ ). Actions were performed by an agent generated by Poser 6 software from Smith Micro (<http://my.smithmicro.com/poser-3d-animation-software.html>), subtending  $10.5^\circ$  (height)  $\times$   $4^\circ$  (width). The movements included heading to side/front, bending to side/front, raising hand to side/front, and raising leg to side/front. The agents are identical across trials and performed each action using the left or right side of the body with an equal chance. Each action lasted 250 ms, plus additional 250 ms with the agent standing still on the screen.

Each trial began with a 1 s fixation on a white cross, followed by a 4 s learning sequence consisting of 1–5 randomly chosen actions, objects, or locations without replacement. To match the duration of the study period, each item was randomly assigned to one of the 8 time slots of 500 ms, with the constraint that the first and the last slot were used first. After a 2 s delay with a blank screen, there was a 1 s presentation of the word “Test,” followed by the presentation of the probe that lasted for 250 ms. Participants indicated within 3 s whether the probe had been presented in the earlier learning sequence, which was the case for 50% of the trials. An event-related design was used with the three types of memory tasks pseudo-randomly mixed. There were 30 trials for each category and set size combination (450 trials in total), which were equally assigned to 10 runs that were finished in two consecutive one-hour scan sessions.

In Experiment 2, we made two changes to clearly separate the encoding and maintenance stages of the VSTM. First, we asked subjects to encode all information but then provided a retro-cue to indicate which type of information to maintain. Second, we extended the maintenance period from 2 to 7.5 s (Figure 1c). The materials included a set of eight highly discriminable agents (identified by gender, face, hair type, and clothing color), subtending  $7^\circ$  (height)  $\times$   $2.5^\circ$  (width). These agents occupied one of the eight equally distributed locations in a circle ( $5^\circ$  away from the center). To make the task difficulty of the

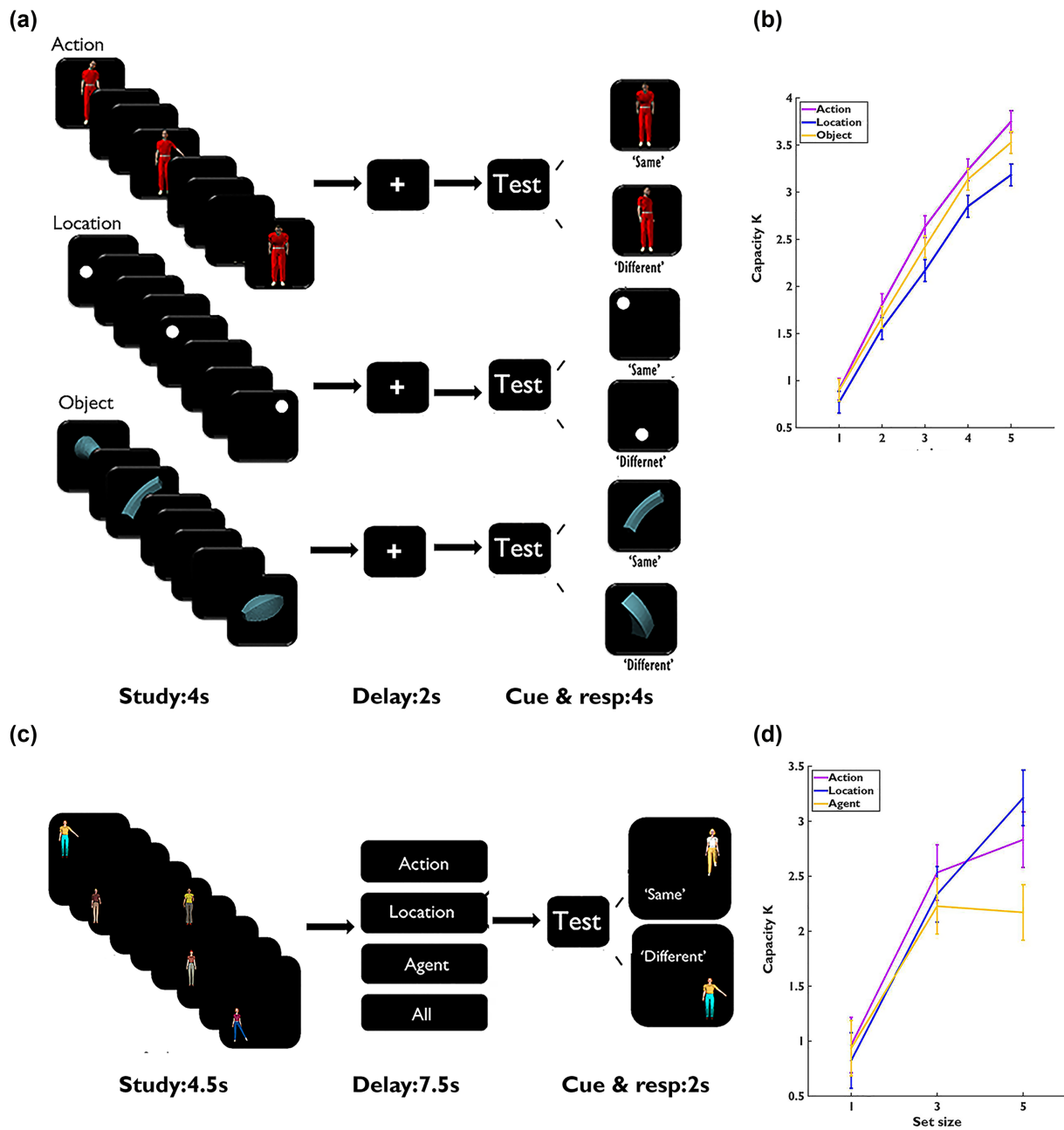
spatial location condition comparable to that of the other two conditions, a random angle (ranging from  $0^\circ$  to  $180^\circ$ ) was added to each trial so that a different set of 8 locations were used in each trial. The movements were the same as those in Experiment 1. The new study sequences consisted of 1, 3, or 5 items that combined different agents, actions, and locations (i.e., no two items shared the same feature in any of the three dimensions). Subjects were instructed to remember all the information in each item. To match the duration of study for each trial, each item was randomly assigned to one of the 9 time slots of 500 ms, with the constraint that the first and the last slots were used first. To allow sufficient time to encode the stimuli, each pair of two items were separated by at least one empty time slot. After the encoding period, a memory cue was presented to instruct the participants which dimension of the information to maintain. For the domain-specific conditions, the cue was “Action,” “Location,” or “Agent”; for the domain-general condition, the cue was “All [information].” The cue was shown in white against black background for the whole 7.5 s maintenance period. A 1 s “Test” cue was then presented, followed by a probe. Subjects were required to judge within 2 s whether they had seen the same action, location, or agent (regardless of whether the uncued dimensions were the same or not) during learning. For the “All [information]” condition, participants were asked to judge whether they had learned exactly the same agent doing the same action in the same location. The four VSTM tasks (“action,” “agent,” “location,” and “all”) were randomly mixed. There were 15 trials for each task and set-size combination (180 trials in total), which were equally assigned to 5 runs of about 10 min each.

## 2.3 | Behavioral analysis

First,  $K$  was calculated for each task and each set size using the following formula:  $K = S \times (H - F)/(1 - F)$ , where  $S$  is the set size of the encoded array,  $H$  stands for the hit rate, and  $F$  is for the false alarm rate. This estimation was considered to be more appropriate when all studied items were simultaneously probed, whereas the Cowan'  $K$  was suitable when only a single cued item was probed during test (Rouder, Morey, Morey, & Cowan, 2011). Pair-wise comparisons between two adjacent set size conditions (e.g., SS1 vs SS2) were conducted to examine how the  $K$  was modulated by task load. Also, we used an iterative method to estimate each individual's short-term capacity in a given task across set sizes (Alvarez & Cavanagh, 2008). In this method, we first averaged the  $K$ s across all set sizes. Starting from the lowest set size, the set sizes whose  $K$  were smaller than the averaged  $K$  were dropped, and the  $K$  estimates from the remaining set sizes were averaged again. This procedure was iterated until the averaged  $K$  value was stabilized. One-way ANOVA was then used to compare WM capacity across tasks.

## 2.4 | fMRI data collection

The imaging data were collected on Siemens 3 T Trio scanners (Siemens Medical Systems, Erlangen, Germany) in the MRI Centers at Southwest University (Experiment 1) and Beijing Normal University



**FIGURE 1** Experimental task and behavioral results. (a) Schematic depiction of the trial structure in Experiment 1 (set size = 3). Each trial lasted for 13 s, including fixation (1 s), encoding (4 s), delay (2 s), test cue (1 s), probe and response period (3 s), and intertrial interval (mean 2 s, jitter from 0 to 6 s). (b) The VSTM capacities as a function of set size in Experiment 1. (c) Procedure of Experiment 2 (set size = 5). Each trial lasted for 16 s, including encoding (4.5 s), maintenance (7.5 s), probe and response (2 s), and intertrial interval (2 s). (d) The VSTM capacities as a function of set size in Experiment 2. The “same” and “different” probes are presented here for a location task. Error bars represent standardized errors of the mean (SEM)

(Experiment 2). For both experiments, anatomical MRI was acquired using a T1-weighted, 3D sequence. The parameters for this sequence were: TR/TE/FA/TI = 1,900 ms/3.39 ms/7°/900 ms, FOV = 256 × 256 mm, voxel size = 1.33 mm × 1 mm × 1.33 mm. One hundred and forty-four sagittal slices were acquired to provide high-resolution structural images of the whole brain. A single-shot T2\*-weighted gradient-echo, EPI sequence was used for functional imaging acquisition with the following parameters: TR/TE/FA = 2,000 ms/30 ms/90°, FOV = 192 × 192 mm, matrix = 64 × 64, and slice thickness = 3 mm.

Forty-one contiguous axial slices parallel to the AC-PC line were obtained to cover the whole cerebrum and partial cerebellum.

## 2.5 | Preprocessing procedure and statistical analysis

The preprocessing and statistical analysis of fMRI data were carried out using FEAT (fMRI Expert Analysis Tool) version 5.98, part of the FSL (FMRIB software library, version 4.1, [www.fmrib.ox.ac.uk/fsl](http://www.fmrib.ox.ac.uk/fsl)). The first 2 volumes before the task were automatically discarded by the scanner

to allow for T1 equilibrium. The remaining images were then realigned to correct for head movements using MCFLIRT, a tool for affine inter- and intermodal brain image registration (Jenkinson et al., 2002; Jenkinson & Smith, 2001). Translational movement parameters never exceeded 1 voxel in any direction for any subject or session. EPI images were registered to the MPAGE structural image, and into standard (i.e., MNI) space (resampled in to  $2\text{ mm} \times 2\text{ mm} \times 2\text{ mm}$  resolution), by using 12-parameter affine transformations (Jenkinson & Smith, 2001). Registration from MPAGE structural image to standard space was further refined by using FNIRT non-linear registration (Andersson, Jenkinson, & Smith, 2007). Data were spatially smoothed by using a 5-mm full-width at half-maximum Gaussian kernel, and filtered in the temporal domain by using a nonlinear high-pass filter with a 90 s cut-off. The preprocessing procedures were exactly the same for both experiments.

In Experiment 1, the data were modeled at the first level by using a general linear model (GLM) within FSL. Fifteen event types, including the combination of three memory tasks (action, location, and object) and five set sizes (SS1 to SS5), were separately modeled. The event onset was defined as the onset of the maintenance period, the duration of the event was set to 2 s (the length of delay), and the onset of the event was convolved with the canonical hemodynamic response function (double-gamma) to generate the regressors. It should be noted that due to the short and fixed delay between encoding, maintenance, and response, we did not model the encoding and response stages in the same model as this could introduce nonorthogonalization-related issues. So the result in Experimental 1 might have reflected a load effect in all three stages. The load-sensitive regions for each memory task were obtained by comparing the BOLD responses for SS4 and SS5 with those for SS1 and SS2. A higher level analysis created cross-run contrasts for each subject for a set of contrast images using a fixed effect model. They were then input into a random-effect model for group analysis using FMRIB's Local Analysis of Mixed Effect stage 1 only (Beckmann, Jenkinson, & Smith, 2003; Woolrich, Behrens, Beckmann, Jenkinson, & Smith, 2004).

In Experiment 2, 12 events, including the combination of four memory tasks (action, location, agent, and all) and three set sizes (SS1, SS3, and SS5), were modeled. In addition, we used finite impulsive response function (FIR) to model the shape of the BOLD response, allowing us to specify the BOLD response in different stages (Cohen et al., 1997; Courtney, Ungerleider, Keil, & Haxby, 1997; Xu & Chun, 2006). For each type of event, 14 time points (28 s in total), starting from the onset of the learning sequences were modeled, covering the whole trial (encoding, 4.5 s; maintenance, 7.5 s; retrieval, 2 s; and inter-trial interval (ITI) periods, 2 s). Due to the hemodynamic delay, the BOLD signal usually peaked 4–6 s after task onset. Therefore, the BOLD response from 12 to 16 s after stimulus onset, which mainly reflected the activity during maintenance, were averaged to represent the maintenance response (Todd & Marois, 2004; Xu & Chun, 2006). The load-sensitive regions for each memory task were obtained by comparing the BOLD responses for SS3 and SS5 with those for SS1. Meanwhile, to test whether the subjects processed all the same information during the study period, we compared the activations at 6–8 s

after the stimulus onset and we expected no differences across memory tasks during this period. For both experiments, the group images were thresholded using the cluster detection statistics with a height threshold of  $z > 2.7$  and a cluster probability of  $p < .05$ , corrected for whole-brain multiple comparisons (FWE) using Gaussian random field theory (GRFT). The reproducibility of the results was further examined by a conjunction analysis across two experiments. All the neural imaging figures in the Results section were generated using BrainNet software (Xia, Wang, & He, 2013).

## 2.6 | Regions of interest (ROIs) analysis

We defined the action-specific ROIs as regions that showed greater load sensitivity for the action memory task than the other two tasks. This was achieved in three steps. First, we identified load-sensitive areas for each of the three tasks (action, object, and location) (Supporting Information, Figure S1a–c for Experiment 1 and Figure S3a–c for Experiment 2). Second, data from the action task were compared to those of the object task and the location task separately (Supporting Information, Figure S1d,e for Experiment 1 and Figure S3d,e for Experiment 2). Finally, we did conjunction analyses on the three contrasts: action (high load) versus action (low load), action (high vs low load) > object/agent (high vs low), and action (high vs low) > location (high vs low). The conjunction analyses were conducted to identify the voxels showing significant effects in all three contrasts (Friston, Penny, & Glaser, 2005; Nichols, Brett, Andersson, Wager, & Poline, 2005). Statistical thresholds were set with a height threshold of  $z > 2.7$  and a cluster probability of  $p < .05$ , corrected for whole-brain multiple comparisons (FWE) using Gaussian random field theory (GRFT). It should be noted that whether or not to include action (high load) versus action (low load) contrast in the conjunction analysis did not affect the result. We also defined object-specific and location-specific ROIs using the same method. The domain-general ROIs were defined as regions showing load sensitivity for all three memory tasks, using the conjunction of three contrasts including action (high vs low), location (high vs low), and object/agent (high vs low). Pair-wise comparisons between two adjacent set size conditions (e.g., SS1 vs SS2) were conducted for each ROI to examine how BOLD signal changes were modulated by memory load and task. All the  $p$  values reported for the ROIs analysis were FDR-corrected across different set sizes, tasks, and ROIs.

## 2.7 | Functional connectivity analysis

Previous studies suggested that the frontal-parietal network (FPN) played a critical role in VSTM (Li et al., 2017; Postle, 2015). Cognitive adaptability is achieved by rapidly updating the patterns of functional connectivity between the FPN hubs and the modules for specialized functions (Cole et al., 2013; Fuster, Bauer, & Jervey, 1985; Gazzaley et al., 2007; Repovs & Barch, 2012). In this study, the generalized form of context-dependent psychophysiological interaction analysis (gPPI) (McLaren, Ries, Xu, & Johnson, 2012) was conducted to test if there was any action-specific functional connectivity with the FPN during the maintenance of actions. Compared to traditional PPI analysis

(Friston et al., 1997), the gPPI analysis reduces false positives and false negatives and allows us to assess connectivity using more than two task conditions. To guarantee the independency from univariate analysis, we did whole-brain search analysis to identify any voxels that showed action-specific change. We also conducted the same connectivity analysis after removing the mean activities in each voxel to minimize contributions from the simultaneous BOLD responses in different areas. In Experiment 1, the frontal and parietal domain-general ROIs (i.e., the left MFG or bilateral superior IPS) were chosen as the seeds. The time course of each seed region was defined as the physiological variable. Its interactions with different memory task (action vs object vs location) were defined as the psychophysiological interaction variables. The three task regressors were included as nuisance covariates. Similar to activation analysis, we first set up six separate contrasts: the connectivity change for each task versus baseline, and the three pair-wise comparisons between tasks. The same conjunction analysis as that used in the activity analysis was done to obtain brain regions showing action-specific changes in functional connectivity. The location-specific and object/agent-specific regions were identified in the same way. The same gPPI analysis was also done in Experiment 2, focusing on the connectivity during the long delay period (8 s). Group images were thresholded with a height threshold of  $z > 2.7$  and a cluster probability of  $p < .05$ , corrected for whole-brain multiple comparisons (FWE) using Gaussian random field theory (GRFT).

### 3 | EXPERIMENT 1 RESULTS

#### 3.1 | Behavioral results

The overall VSTM capacity for actions was  $2.856$ ,  $SD = 0.461$ , which was not significantly different from capacities for locations ( $M = 2.439$ ,  $SD = 0.629$ ) or objects ( $M = 2.689$ ,  $SD = 0.557$ ),  $F(2,39) = 1.370$ ,  $p = .266$ . The  $K$  for the action task increased monotonically from SS1 to SS5 (two-tailed paired-samples  $t$  tests,  $ps < .012$ ), but the  $K$ s for the location and object tasks increased from SS1 to SS4 ( $ps < .001$ ) and then plateaued at SS4 (SS4 vs SS5:  $ps > .101$ ) (Figure 1b). All the  $p$  values were corrected by Benjamini–Hochberg false discovery rate (FDR) method, across different set sizes and tasks.

#### 3.2 | Action-specific regions

Action-specific ROIs were defined as the regions showing greater load sensitivity for actions than the other two memory tasks via conjunction analysis. Our results revealed that the bilateral middle temporal areas (MT: center of gravity (COG) in Montreal Neurological Institute (MNI) coordinate system,  $x/y/z: -49/49, -67/-65, 8/6$ ) were selectively activated in maintaining action information. For the action-specific regions, we further examined whether their activities tracked behavioral measures of VSTM capacity based on the action task (Figure 2a and Supporting Information, Table S1). Bilateral MT activities increased from SS1 to SS4 ( $ps < .012$ ), and reached a plateau at SS4 (SS4 vs SS5,  $ps > .125$ ). Activities in these regions were not consistently modulated by memory load for the object ( $ps > .904$ ) or location tasks (from SS2

to SS4,  $ps > .48$ ; except an increase from SS1 to SS2 in left MT:  $t(13) = -2.901$ ,  $p = .049$ , and from SS4 to SS5 in right MT:  $t(13) = -4.180$ ,  $p = .009$ ).

#### 3.3 | Activation of the location- or object-specific brain areas

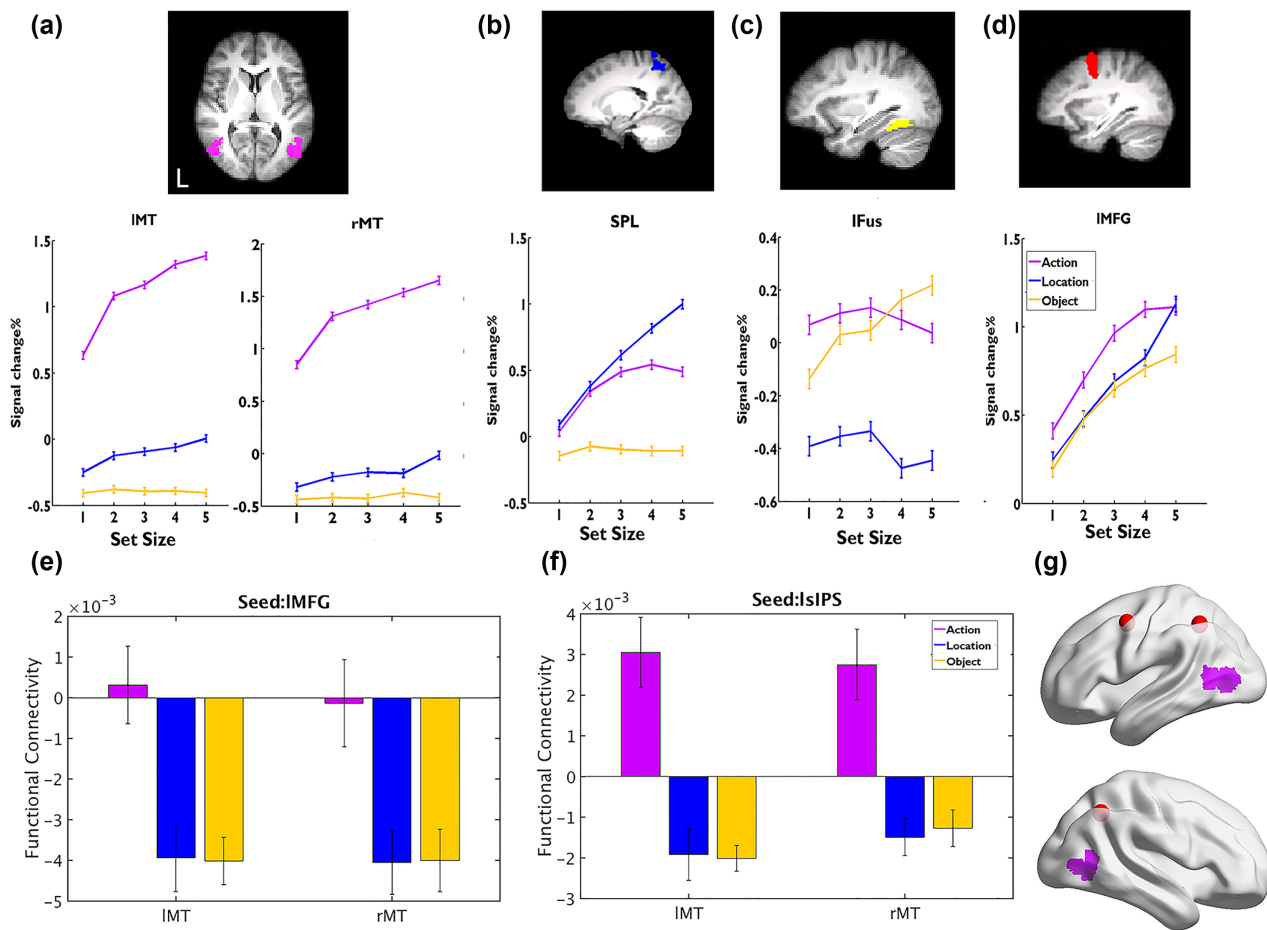
Using a similar approach as above, we found several location-specific brain regions, including the superior parietal lobule (SPL) extending to the precuneus ( $x/y/z: -5, -53, 59$ ), right postcentral cortex (PostCG,  $53, -24, 43$ ), and bilateral superior frontal gyrus (SFG,  $x/y/z: -22/27, -9/-7, 58/58$ ). Object-specific brain regions were found in the bilateral fusiform gyrus (Fus,  $x/y/z: -32/34, -56/-50, -16/-19$ ) and occipital pole (OP,  $x/y/z: -31/32, -94/-94, -4/-1$ ) (see details in Supporting Information, Figure S2 and Table S1). We then examined whether these location- and object-specific regions were also involved in the maintenance of actions. Our results revealed that the activities of location-specific regions (i.e., the SPL, PostCG, and bilateral SFG) were also modulated by the set size of the action task, where the activities increased from SS1 to SS3 ( $ps < .049$ , except for the change from SS1 to SS2 in right SFG,  $p = .108$ ) and peaked at SS3 ( $ps > .288$ ). These regions were not modulated by the set size of the object task ( $ps > .160$ ) (Figure 2b and Supporting Information, Figure S2). In contrast, the activities of object-specific regions (bilateral fusiform and OP) were not modulated by the set size of either the action or the location task ( $ps > .732$ ) (Figure 2c and Supporting Information, Figure S2).

#### 3.4 | Domain-general regions

The domain-general regions were defined as the regions showing load sensitivity for all three tasks (action, location, and object). They included the left middle frontal gyrus (MFG,  $x/y/z: -39, 2, 45$ ), bilateral superior intraparietal sulcus (sIPS,  $x/y/z: -37/37, -50/-54, 44/42$ ), paracingulate gyrus (ParaCG,  $x/y/z: -1, 17, 48$ ), and bilateral LOC ( $x/y/z: -45/53, -66/-64, -7/-8$ ) (Supporting Information, Table S1). For the action and object tasks, the activations of all these ROIs increased from SS1 to SS4 [ $ps < .050$ , except that there were no increases from SS1 to SS2 in ParaCG and right sIPS ( $ps > .124$ )] and then peaked at SS4 [ $ps > .167$ , except for an increase from SS4 to SS5 in the object task in ParaCG,  $t(13) = -3.145$ ,  $p = .017$ ]. For the location task, the activations of all these ROIs increased monotonically from SS1 to SS3 ( $ps < .032$ ), then from SS3 to SS4, the activations did not increase ( $ps > .075$ , except in left MFG,  $t(13) = -2.645$ ,  $p = .029$ ), while all the activations increased further from SS4 to SS5 again ( $ps < .001$ ) (Figure 2d and Supporting Information, Figure S2).

#### 3.5 | Action-specific functional connectivity change

Our gPPI analysis revealed that using the left MFG as the seed, its functional connectivity was significantly stronger with bilateral MT ( $x/y/z: -47/49, -69/-67, 10/4$ ) during the action task than during the other two tasks (Figure 2e). Moreover, this pattern was replicated by using the bilateral sIPS as seeds as well (mean  $x/y/z: -47/49, -69/-65, 8/5$ ) (Figure 2f). These gPPI analyses with three different seeds



**FIGURE 2** Results of Experiment 1. (a–d) The BOLD signal changes as a function of set size for different tasks in the action-specific brain area (a), location-specific area (b), object-specific area (c), and domain-general brain area (d); see Supporting Information, Table S1 and Figure S2 for results for other ROIs. (e) The functional connectivity between bilateral MT and IMFG, as a function of different VSTM task. (f) The purple regions are the overlapping regions (by conjunction analysis) showing action-specific increases in functional connectivity using IMFG and bilateral sIPS as seeds (red dots). Error bars represent SEM [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

yielded highly overlapping action-specific functional connectivity (Figure 2g). Meanwhile, functional connectivity was strongest between the left MFG and left SPL ( $x/y/z$ :  $-8, -60, 50$ ), between the left sIPS and right PostCG ( $x/y/z$ :  $42, -40, 36$ ), and between the right sIPS and the right SPL ( $x/y/z$ :  $26, -56, 60$ ) during the location task, and the functional connectivity was strongest between these general ROIs and the left fusiform gyrus (mean  $x/y/z$ :  $-32, -36, -26$ ) during the object task. However, the object/location-specific functional connectivity was statistically significant only when using a relatively low height threshold of  $z > 2.3$  and a cluster probability of  $p < .05$  (see more details in Supporting Information, Table S2).

### 3.6 | Summary of Experiment 1

Experiment 1 revealed action-specific activations in the bilateral MT and functional connectivity between MT and FPN, suggesting distinct neural substrates for VSTM of actions. However, these results were somewhat confounded by the differences in visual stimuli (videos of human actions, 3D shapes as objects, and dots for locations). Furthermore, Experiment 1 used a short delay design and hence could not

clearly differentiate the encoding and maintenance processes. To overcome the above limitations, Experiment 2 used the same materials for all three tasks to eliminate the differences in stimuli and used a retrocue and a prolonged delay period (from the original delay of 2 s to the new delay of 7.5 s) to clearly identify the neural substrates of working memory maintenance.

## 4 | EXPERIMENT 2 RESULTS

### 4.1 | Behavioral results

The mean VSTM capacity for actions was 2.883,  $SD = 0.856$ , which was not significantly different from that for either locations (2.774,  $SD = 0.633$ ) or objects/agents (2.306,  $SD = 0.893$ ),  $F(2,40) = 1.475$ ,  $p = .241$ . Across the three set sizes (SS1, SS3, and SS5), the VSTM capacity for actions increased from SS1 to SS3 ( $t(20) = 12.919$ ,  $p < .001$ ) and plateaued (SS3 vs SS5:  $t(20) = -0.609$ ,  $p = .549$ ). The same pattern was observed for VSTM of agents (SS1 vs SS3:  $t(20) = 9.298$ ,  $p < .001$ ; SS3 vs SS5:  $t(20) = 0.157$ ,  $p = .877$ ). The VSTM of locations increased monotonically ( $ps < .006$  from SS1 to SS5; Figure

1d). The  $K$  in the "All" condition ( $1.47$ ,  $SD = 0.7$ ) was significantly lower than those in the other three task conditions ( $ps < .001$ ), and the  $K$  in SS5 decreased significantly from SS3 ( $p < .001$ ), due to overwhelming task load (See more details in Supporting Information, Table S3). As a result, we excluded the "All" condition from further analysis.

## 4.2 | Action-specific regions

The neural responses from 12 to 16 s after stimulus onset were averaged to represent the maintenance-related response. Conjunction analysis found that the left MT ( $x/y/z$ :  $-55, -59, 5$ ), the supplementary motor cortex (SMC,  $x/y/z$ :  $-3, 11, 57$ ), and the left inferior frontal gyrus (IFG:  $x/y/z$ :  $-47, 8, 31$ ) showed action-specific increases in activations as a function of VSTM load (Figure 3a and Supporting Information, Table S4). Moreover, for all these action-specific regions, the activities increased from SS1 to SS3 ( $ps < .001$ ) then reached a plateau (SS3 vs SS5,  $ps > .195$ ). However, in left MT, the activations were not modulated by the memory load of locations or agents ( $ps > .096$ , except that the activities decreased in the agent task from SS3 to SS5,  $t(20) = 4.062$ ,  $p = .003$ ). The activations in SMC and left IFG ROIs were also not modulated by memory load of locations ( $ps > .096$ ), but both of the activities showed an increase from SS1 to SS3 (SS1 vs SS3,  $ps < .003$ ) and reached the plateau (SS3 vs SS5,  $ps > .092$ ) in the agent task (Figure 3b).

## 4.3 | Activation of the location- or object-specific brain areas

The bilateral SPL ( $x/y/z$ :  $-11/9, -49/-53, 71/70$ ) were identified as location-specific areas. For both left and right SPL, the BOLD peaked at SS3 (SS1 vs SS3,  $ps < .034$ , SS3 vs SS5,  $ps > .534$ ). However, during the action task, the BOLD signal increased from SS1 to SS3 ( $ps < .034$ ) but then decreased afterward (SS3 vs SS5,  $ps < .047$ ). For the agent task, the BOLD signal did not change from SS1 to SS5 ( $ps > .587$ ), but decreased significantly at SS5 ( $ps < .034$ ). No differences in BOLD signals were found across the bilateral SPL ( $ps > .265$ ) and we averaged the bilateral SPL signals in Figure 3d.

A similar conjunction method did not find any agent-sensitive region with sustained activations during the maintenance period in Experiment 2. The object-specific ROIs (bilateral fusiform) found in Experiment 1 showed increased activations during the encoding period but the activities decayed during the delay across all tasks (Figure 3e).

## 4.4 | Domain-general regions

In the conjunction analysis of the load-sensitive areas across all three tasks, we replicated the findings of Experiment 1, with the left MFG ( $x/y/z$ :  $-28, 8, 58$ ) and bilateral sIPS ( $x/y/z$ :  $-19/39, -61/-56, 49/46$ ) being involved in the maintenance of all three types of stimuli (Figure 3a). The activations in the bilateral sIPS and left MFG reached a plateau at SS3 in the action and agent tasks (SS1 vs SS3:  $ps < .001$ ; SS3 vs SS5:  $ps > .183$ , except in left MFG whose activities decreased at SS5 in the agent task,  $t(20) = 2.394$ ,  $p = .048$ ). In contrast, all these regions increased monotonously for the location task (SS3 vs SS5:  $ps < .045$ ) (Figure 3c).

## 4.5 | Activations during the encoding period

For both domain-specific and domain-general ROIs, two-way ANOVA revealed a significant main effect of set size ( $ps < .014$ ), but no effect of task condition or condition by set size interaction ( $ps > .167$ ) during the encoding period (6~8 s after stimulus onset). These results suggested that participants indeed encoded all information during the encoding period of our retro-cue design.

## 4.6 | Action-specific functional connectivity change

The whole-brain gPPI analysis did not reveal any significant clusters showing action-specific increases in functional connectivity with the domain-general seeds (left MFG and bilateral sIPS). Focusing on regions showing action-specific increases in activities (i.e., left MT, SMC, and left IFG), we found that these regions showed greater functional connectivity with the parietal cortex in the action task than those in the location or agent task (Figure 3g,  $ps < .001$ ), which was consistent with the findings of Experiment 1.

## 4.7 | Summary of Experiment 2

Experiment 2 replicated the two critical findings of Experiment 1. The first finding was that there were action-specific activations as well as functional connectivity within the MT area, which provided clear evidence for the distinct neural substrate for maintaining actions during the VSTM delay. The second finding was that, across both experiments, the location-specific areas were always recruited for the action task during the delay, which revealed that action maintenance shared the neural substrates supporting the maintenance of spatial information.

# 5 | CONJUNCTION ANALYSIS OF EXPERIMENTS 1 AND 2

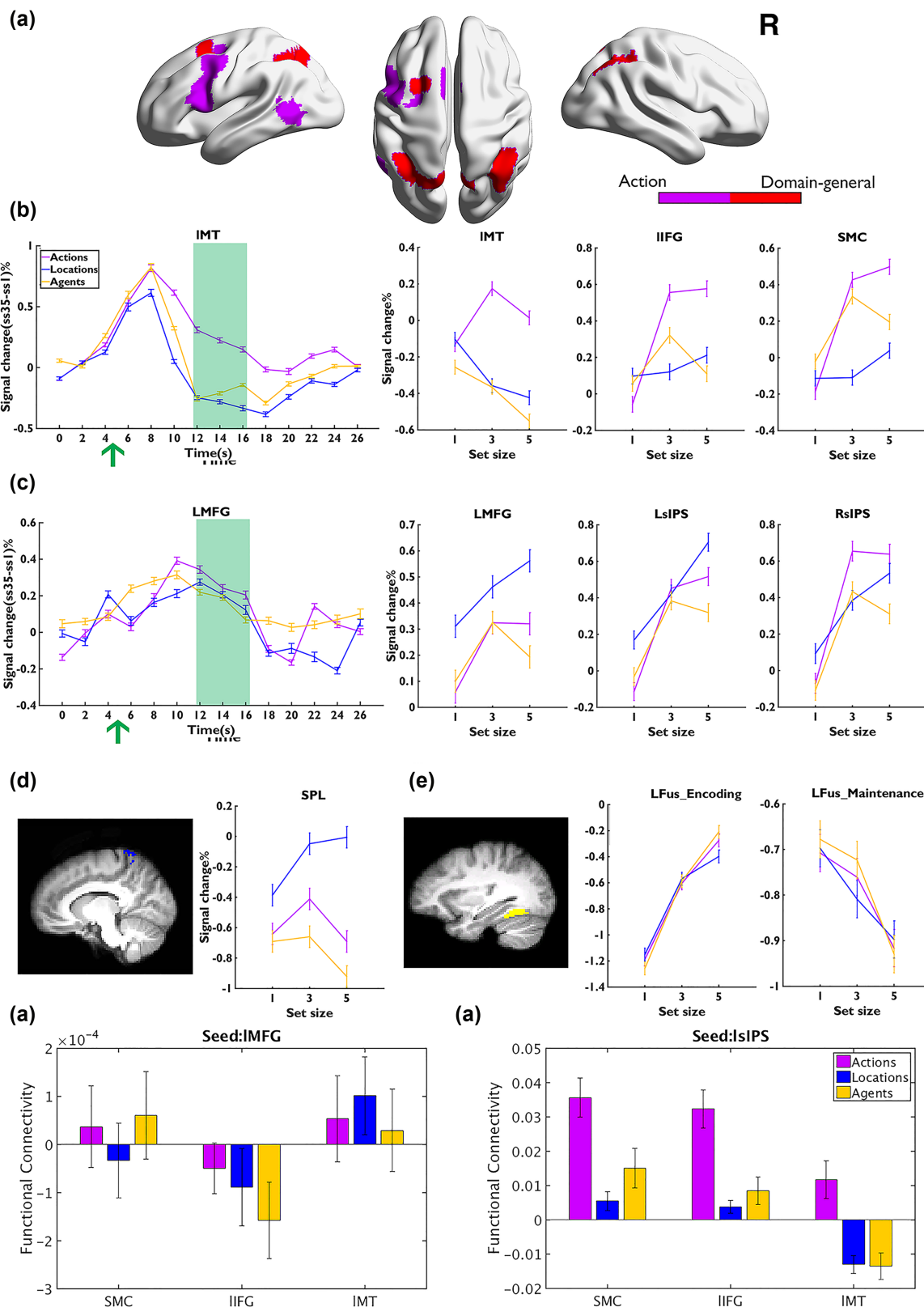
## 5.1 | Overlap of action-specific and domain-general areas

To examine the reproducibility of the results, we directly examined the overlap of the action-specific and domain-general ROIs from Experiments 1 and 2. The results showed consistent activation for action- and location-specific areas, as well as for domain-general areas (Figure 4 and Supporting Information, Table S5). Because no significant object-specific region was found in Experiment 2, the conjunction analysis revealed no object-specific region.

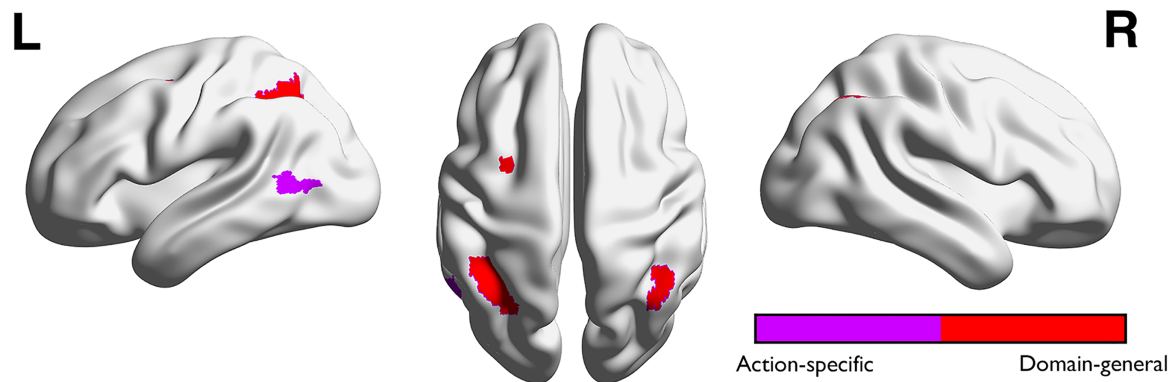
## 5.2 | Reidentification of location-specific areas

Given that the location-specific ROIs were also modulated by actions, our pair-wise contrasts between locations and actions might have excluded some important location-related regions. If the actions could not be maintained independently from the spatial information, we would expect that all the location-related regions should be recruited in the action task as well. We thus redefined the location-specific areas by a conjunction of locations versus objects in Experiments 1 and 2, which would guarantee that all location-sensitive regions were





**FIGURE 3** Results of Experiment 2. (a) The action-specific (purple) and domain-general (red) areas. (b,c) The time courses of BOLD signal changes as a function of three VSTM tasks in action-specific/domain-general areas (left, the green arrow indicates the onset of the retrocue) and the signal changes as a function of set size across tasks during maintenance (right, 12~16 s within the green square). (d) The activities in location-specific area during maintenance. (e) Activities in the fusiform during encoding and maintenance. Error bars represent SEM [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

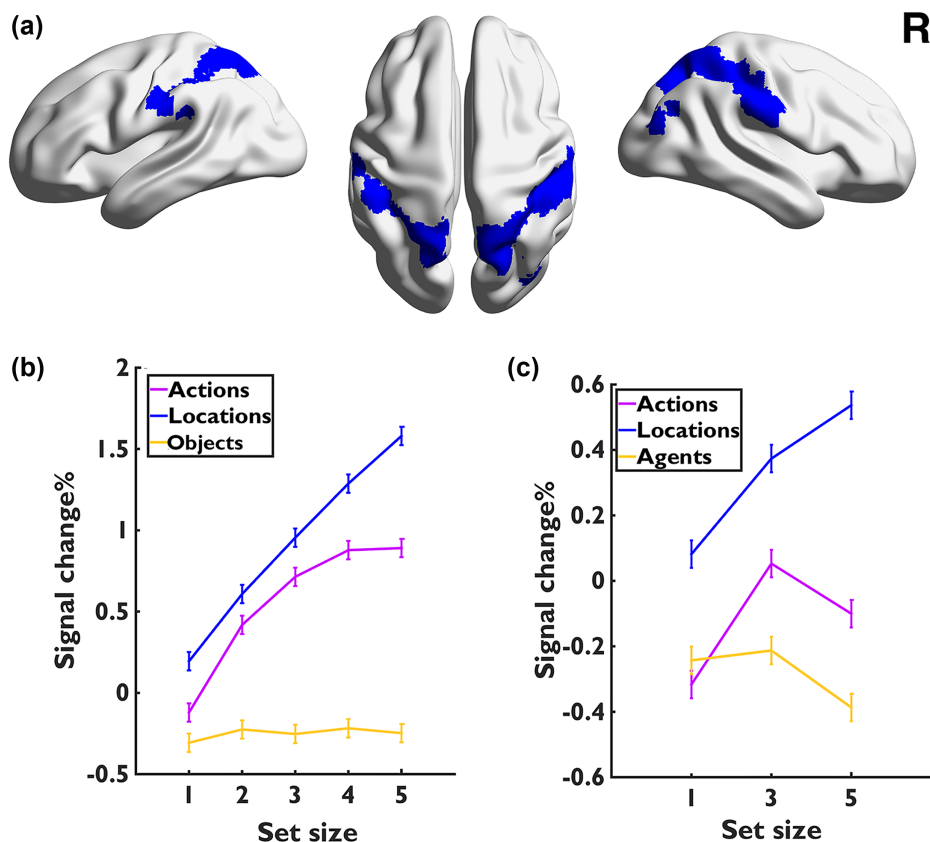


**FIGURE 4** The conjunction results of Experiments 1 and 2. Action-specific (purple) and domain-general regions (red) were obtained by conjunction analysis across the two experiments [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

included. The results revealed a set of areas along the dorsal visual pathway, including the bilateral SPL ( $x/y/z$ :  $-30/32$ ,  $-51/-49$ ,  $62/59$ ), PostCG ( $x/y/z$ :  $-47/46$ ,  $-26/-29$ ,  $40/51$ ) and dorsal LOC ( $x/y/z$ :  $-18/27$ ,  $-64/-66$ ,  $60/57$ ) (Figure 5a and Supporting Information, Table S6). As expected, we found that within all these regions, the activations tracked the memory load increases in both location and action tasks and in both experiments. These regions showed no increases in the object or agent task ( $p_s > .583$ ) except a decrease at SS5 in SPL in Experiment 2 ( $t(20) = 2.543$ ,  $p = .019$ ) (Figure 5b,c).

## 6 | DISCUSSION

In this study, we conducted two experiments to systematically examine the neural substrates for VSTM of actions. Experiment 1 revealed that compared to VSTM of objects and locations, VSTM of actions were specifically linked to activations in the bilateral MT and functional connectivity between MT and FPN. Experiment 2 replicated the results of Experiment 1 after controlling for potential differences in visual stimuli with a retro-cue design and separating the encoding and maintenance



**FIGURE 5** (a) Location-specific areas defined by the contrast between locations versus objects/agents, to account for the fact that actions also involved the location-specific regions. (b,c) The BOLD response changes as a function of set size across the three tasks in Experiments 1 and 2 (c) [Color figure can be viewed at [wileyonlinelibrary.com](http://wileyonlinelibrary.com)]

stages by adding a longer delay period. Furthermore, we also found that the brain areas that processed location information (bilateral SPL and SFG) were also involved in the processing of actions. In contrast, brain areas that processed objects/agents were not involved in the processing of actions.

Although previous behavioral studies already suggested separate cognitive processes involved in the VSTM of actions, objects/agents, and locations (Shen et al., 2014; Wood, 2007, 2008, 2011a), this study for the first time provided clear evidence for separate neural mechanisms. Consistent with previous evidence of the role of MT in observing actions (Decety et al., 1997; Jacquet & Avenanti, 2013; Rizzolatti & Craighero, 2004), we found that the MT was uniquely involved in the short-term maintenance of actions in both experiments. Experiment 2 also revealed that VSTM involved the bilateral IFG and SMC, two areas that are suggested to be critical for high-level action cognition, such as action understanding, imitation, and goal-directed motor planning processes (Buccino et al., 2001; Iacoboni et al., 2005; Petzschner & Krüger, 2012). Together, these findings support the sensorimotor recruitment theory, which posits that the stimulus-specific sensorimotor cortices, which are involved in sensory perception, play a role in the temporal maintenance of actions (D'Esposito & Postle, 2015).

One interesting finding is that the MT showed sustained activity during the whole maintenance period (Figure 3a). This pattern was evident in the other two location-specific regions (i.e., SPL and SFG). In contrast, the object-specific regions (i.e., bilateral fusiform) did not show sustained activation during maintenance, but instead were modulated by the set size of objects during the encoding stage. Although spatial working memory studies have reported sustained activations in the posterior parietal cortex and SFG during a long delay (maximum 6~9 s) (Courtney, Petit, Maisog, Ungerleider, & Haxby, 1998; Malhotra, Coulthard, & Husain, 2009; Xu & Chun, 2006), no sustained activation during maintenance has been found for dot movements (Emrich, Riggall, LaRocque, & Postle, 2013; Riggall & Postle, 2012). Meanwhile, mixed results have been reported for VSTM of objects (colors, shapes, etc.): sustained activation was found in some studies (Xu & Chun, 2006) but not others (Nelissen, Stokes, Nobre, & Rushworth, 2013). These results suggest that different neuronal and synaptic mechanisms might be involved in the maintenance of different types of visual information. For example, the transiently modified synaptic weights have been found to help to maintain object information (Barak & Tsodyks, 2014; Stokes et al., 2013; Sugase-Miyamoto, Liu, Wiener, Optican, & Richmond, 2008). Further studies are required to examine this important question.

Consistent with previous studies (Courtney et al., 1998; Malhotra et al., 2009; Xu & Chun, 2006), the SPL and SFG were involved in the VSTM for location information in our current location tasks. Interestingly, these two areas were also modulated significantly by the memory load of actions even though the location information was irrelevant in the action task. This is consistent with the fact that actions are characterized by continuous temporal-spatial changes (Alibali, 2005). However, this conjecture of shared neural resources for actions and spatial locations does not seem to be consistent with behavioral studies that

showed no behavioral interference between spatial locations and actions in the dual task condition (Wood, 2007). One possible explanation is that the SPL might not be functionally necessary for action maintenance. Further lesion or virtual lesion studies are needed to address this issue. It has also been suggested that spatial working memory is involved in the integration of various visual features (e.g., color, shape, and movement) into a coherent representation in working memory (Wood, 2011b). Consistently, a recent neural imaging study used multivariate analysis and revealed that the location context was obligatorily kept during the VSTM delay regardless of its task relevance (Foster, Bsaies, Jaffe, & Awh, 2017). However, this cannot explain why only actions but not objects/agents recruited the location-specific regions.

Unlike the two location-related regions, the object-related regions (i.e., the fusiform) were not involved in VSTM of actions. These regions were not responsive to memory load in the action task and showed different functional connectivity patterns for actions versus objects/agents. These results provided neural evidence for the dissociation between actions and objects/agents found in behavioral studies. For example, one previous behavioral study found that the dual task for actions and colored dots did not affect behavioral performance (Wood, 2011a). A more recent VSTM study showed that changing the color of biological motion stimuli did not affect the behavioral performance for actions, and changing the motions did not affect the performance for colors either (Ding et al., 2015).

This study also found two brain regions (the left MFG and bilateral superior IPS) that were domain-general, namely, involved in all three types of VSTM. The superior IPS region overlaps with the region found to be modulated by both the number of spatial locations and object complexity (Xu & Chun, 2006). Consistently, sustained activations in the FPN have been reported by a large number of fMRI studies (Curtis & D'Esposito, 2003; Eriksson, Vogel, Lansner, Bergstrom, & Nyberg, 2015; Riggall & Postle, 2012) and have been suggested to reflect either the central executive function according to the multicomponent working memory model (Baddeley, 2003; Smith & Jonides, 1997) or the general attention process (Corbetta & Shulman, 2002; Huettel, Mieuek, Jurkowski, & McCarthy, 2004). Our results further revealed that these two domain-general regions had functional connectivities with domain-specific sensory regions and that these functional connectivities were responsive to domain-specific task demands. This finding extends the previous fMRI finding of significant functional connectivity increases between FFA and the nodes in the FPN during VSTM for faces (Gazzaley, Rissman, & D'Esposito, 2004), and the finding of significant associations between such connectivity increases and memory performance for color squares during VSTM (Kuo, Yeh, Chen, & D'Esposito, 2011; Lee & D'Esposito, 2012). Taken together, all these results supported the flexible hub theory (Cole et al., 2013; Fuster et al., 1985; Gazzaley et al., 2007; Repovs & Barch, 2012) that cognitive adaptability is achieved by rapidly updating the patterns of global functional connectivity between the FPN attentional control hubs and the modules for specialized functions (such as the action, location, and object subsystems). Our results thus provided support to the sensory-recruitment hypothesis, but were inconsistent with the role of the

parietal lobe in VSTM storage (Leavitt, Mendoza-Halliday, & Martinez-Trujillo, 2017; Xu, 2017).

Although this study obtained convergent evidence for neural substrates dedicated to VSTM of actions, several important questions remain to be answered. First, this study found elevated, load-sensitive activities during the maintenance period. However, it is unclear whether these activities are functionally necessary for the maintenance of actions. Our results showed that BOLD signal pattern did not always track the working memory capacity either across task loads or across individuals, which is consistent with recent model-fitting evidence showing that the BOLD activity followed a saturation function rather than reflecting strict capacity limits (Bays, 2018). Future studies should use multiple voxel pattern analysis to examine whether the activation patterns in these regions contain item-specific information, and how the fidelity of the representations is modulated by VSTM load and related to memory performance. In addition, (virtual) lesion approaches are also useful to establish the functional necessity of these activations/representations in working memory. Second, this study cannot distinguish the different neural substrates that supported simple movements (dots movements or rigid motion of an agent) and human actions (an avatar conducting an action). To further explore this question, future research needs to directly compare these two types of VSTM.

More importantly, existing theoretic models have suggested that the “core knowledge” systems emerge early in human development and thus are common to infants, children, and adults. To further test this hypothesis, future studies should examine the developmental changes of this domain-specific and domain-general architecture of VSTM. In particular, future studies should examine whether there are developmental changes in activation overlap and dual-task interference among different domains. In the visual object domain, mounting evidence has suggested that early childhood is a period for developmental differentiation (Deen et al., 2017; Golarai et al., 2007), whereas aging is associated with dedifferentiation (i.e., less functional specificity/selectivity) (Burianová, Lee, Grady, & Moscovitch, 2013; Carp, Park, Polk, & Park, 2011; Li, Lindenberger, & Sikström, 2001; Park et al., 2004, 2012; Payer et al., 2006; Voss et al., 2008; Zheng et al., 2017). A previous functional imaging study showed that the neural overlap in visual representations affected the VSTM capacity when stimuli from different visual categories were presented (Cohen, Konkle, Rhee, Nakayama, & Alvarez, 2014).

To summarize, this study systematically compared the neural substrates for the maintenance of actions, agents, and locations in VSTM. Combining the evidence from both activation and functional connectivity analyses, we found that the bilateral MT was specifically involved in the action task, suggesting a separate system for VSTM of actions. We further found that the action task also involved the location-specific regions but not the object/agent-specific regions. Finally, we found domain-general regions that were involved in the processing of all three types of visual stimuli. These findings for the first time revealed the neural substrates of short-term maintenance of actions, supporting the “core knowledge” system theory during VSTM, and furthered our understanding of the neural architecture of human VSTM.

## ACKNOWLEDGMENT

This work was sponsored by the National Science Foundation of China (31730038), 973 Program (2014CB846102), the 111 Project (B07008), the NSFC and German Research Foundation (DFG) Joint Project NSFC 61621136008/DFG TRR-169, and the Guangdong Pearl River Talents Plan Innovative and Entrepreneurial Team grant #2016ZT06S220.

## AUTHOR CONTRIBUTIONS

YC, ZU, JW, and GX designed research; YC, SL, and AC performed research; YC and GX analyzed data; and YC, ZU, JW, CC, and GX wrote the article.

The authors declare no conflict of interest.

## ORCID

Ying Cai  <http://orcid.org/0000-0001-5125-2556>

Antao Chen  <http://orcid.org/0000-0001-9321-681X>

## REFERENCES

- Alibali, M. W. (2005). Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information. *Spatial Cognition and Computation*, 5(4), 307–331.
- Alvarez, G. A., & Cavanagh, P. (2008). Visual short-term memory operates more efficiently on boundary features than on surface features. *Perception & Psychophysics*, 70(2), 346–364.
- Andersson, J. L., Jenkinson, M., & Smith, S. (2007). *Non-linear registration, aka spatial normalisation FMRIB technical report TR07JA2*. FMRIB Analysis Group of the University of Oxford.
- Bays, P. M. (2018). Reassessing the evidence for capacity limits in neural signals related to working memory. *Cerebral Cortex (New York, N.Y. : 1991)*, 28(4), 1432–1438.
- Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews. Neuroscience*, 4(10), 829–839.
- Barak, O., & Tsodyks, M. (2014). Working models of working memory. *Current Opinion in Neurobiology*, 25, 20–24. doi:<https://doi.org/10.1016/j.conb.2013.10.008>
- Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2003). General multilevel linear modeling for group analysis in FMRI. *NeuroImage*, 20(2), 1052–1063. [https://doi.org/10.1016/s1053-8119\(03\)00435-x](https://doi.org/10.1016/s1053-8119(03)00435-x)
- Buccino, G., Binkofski, F., Fink, G. R., Fadiga, L., Fogassi, L., Gallese, V., ... Freund, H.-J. (2001). Action observation activates premotor and parietal areas in a somatotopic manner: An fMRI study. *European Journal of Neuroscience*, 13(2), 400–404.
- Buccino, G., Lui, F., Canessa, N., Patteri, I., Lagravinese, G., Benuzzi, F., ... Rizzolatti, G. (2004). Neural circuits involved in the recognition of actions performed by nonconspecifics: An FMRI study. *Journal of Cognitive Neuroscience*, 16(1), 114–126. <https://doi.org/10.1162/089892904322755601>
- Burianová, H., Lee, Y., Grady, C. L., & Moscovitch, M. (2013). Age-related dedifferentiation and compensatory changes in the functional network underlying face processing. *Neurobiology of Aging*, 34(12), 2759–2767.
- Carey, S. (2009). *The origin of concepts*. Oxford University Press.
- Carp, J., Park, J., Polk, T. A., & Park, D. C. (2011). Age differences in neural distinctiveness revealed by multi-voxel pattern analysis.

- NeuroImage*, 56(2), 736–743. Neuroimage, <https://doi.org/10.1016/j.neuroimage.2010.04.267>
- Cohen, J. D., Perlstein, W. M., Braver, T. S., Nystrom, L. E., Noll, D. C., Jonides, J., & Smith, E. E. (1997). Temporal dynamics of brain activation during a working memory task. *Nature*, 386(6625), 604–608.
- Cohen, M. A., Konkle, T., Rhee, J. Y., Nakayama, K., & Alvarez, G. A. (2014). Processing multiple visual objects is limited by overlap in neural channels. *Proceedings of the National Academy of Sciences*, 111(24), 8955–8960. <https://doi.org/10.1073/pnas.1317860111>
- Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., & Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience*, 16(9), 1348–1355. <https://doi.org/10.1038/nn.3470>
- Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews. Neuroscience*, 3(3), 201–215. <https://doi.org/10.1038/nrn755>
- Courtney, S. M., Petit, L., Maisog, J. M., Ungerleider, L. G., & Haxby, J. V. (1998). An area specialized for spatial working memory in human frontal cortex. *Science*, 279(5355), 1347–1351. <https://doi.org/10.1126/science.279.5355.1347>
- Courtney, S. M., Ungerleider, L. G., Keil, K., & Haxby, J. V. (1997). Transient and sustained activity in a distributed neural system for human working memory. *Nature*, 386(6625), 608–611. <https://doi.org/10.1038/386608a0>
- Cross, E. S., Kraemer, D. J. M., Hamilton, A. F. D. C., Kelley, W. M., & Grafton, S. T. (2009). Sensitivity of the action observation network to physical and observational learning. *Cerebral Cortex (New York, N.Y. : 1991)*, 19(2), 315–326. <https://doi.org/10.1093/cercor/bhn083>
- Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, 7(9), 415–423. [https://doi.org/10.1016/S1364-6613\(03\)00197-9](https://doi.org/10.1016/S1364-6613(03)00197-9)
- D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, 66(1), 115–142. <https://doi.org/10.1146/annurev-psych-010814-015031>
- Decety, J., Grezes, J., Costes, N., Perani, D., Jeannerod, M., Procyk, E., ... Fazio, F. (1997). Brain activity during observation of actions. Influence of action content and subject's strategy. *Brain*, 120(10), 1763–1777.
- Deen, B., Richardson, H., Dilks, D. D., Takahashi, A., Keil, B., Wald, L. L., ... Saxe, R. (2017). Organization of high-level visual cortex in human infants. *Nature Communications*, 8, 13995.
- Ding, X., Zhao, Y., Wu, F., Lu, X., Gao, Z., & Shen, M. (2015). Binding biological motion and visual features in working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 41(3), 850.
- Emrich, S. M., Riggall, A. C., LaRocque, J. J., & Postle, B. R. (2013). Distributed patterns of activity in sensory cortex reflect the precision of multiple items maintained in visual short-term memory. *Journal of Neuroscience*, 33(15), 6516–6523. <https://doi.org/10.1523/JNEUROSCI.5732-12.2013>
- Eriksson, J., Vogel, E. K., Lansner, A., Bergstrom, F., & Nyberg, L. (2015). Neurocognitive architecture of working memory. *Neuron*, 88(1), 33–46. <https://doi.org/10.1016/j.neuron.2015.09.020>
- Foster, J. J., Bsales, E. M., Jaffe, R. J., & Awh, E. (2017). Alpha-band activity reveals spontaneous representations of spatial position in visual working memory. *Current Biology*, 27(20), 3216–3223. e3216. <https://doi.org/10.1016/j.cub.2017.09.031>
- Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage*, 6(3), 218–229. <https://doi.org/10.1006/nimg.1997.0291>
- Friston, K. J., Penny, W. D., & Glaser, D. E. (2005). Conjunction revisited. *NeuroImage*, 25(3), 661–667. Neuroimage, <https://doi.org/10.1016/j.neuroimage.2005.01.013>
- Fuster, J. M., Bauer, R. H., & Jervey, J. P. (1985). Functional interactions between inferotemporal and prefrontal cortex in a cognitive task. *Brain Research*, 330(2), 299–307.
- Gao, Z., Ye, T., Shen, M., & Perry, A. (2016). Working memory capacity of biological movements predicts empathy traits. *Psychonomic Bulletin & Review*, 23(2), 468–475. <https://doi.org/10.3758/s13423-015-0896-2>
- Gazzaley, A., Rissman, J., Cooney, J., Rutman, A., Seibert, T., Clapp, W., & D'Esposito, M. (2007). Functional interactions between prefrontal and visual association cortex contribute to top-down modulation of visual processing. *Cerebral Cortex (New York, N.Y. : 1991)*, 17(suppl 1), i125. 1, 135. <https://doi.org/10.1093/cercor/bhm113>
- Gazzaley, A., Rissman, J., & D'Esposito, M. (2004). Functional connectivity during working memory maintenance. *Cognitive, Affective, & Behavioral Neuroscience*, 4(4), 580–599.
- Golarai, G., Ghahremani, D. G., Whitfield-Gabrieli, S., Reiss, A., Eberhardt, J. L., Gabrieli, J. D., & Grill-Spector, K. (2007). Differential development of high-level visual cortex correlates with category-specific recognition memory. *Nature Neuroscience*, 10(4), 512.
- Grèzes, J., Costes, N., & Decety, J. (1999). The effects of learning and intention on the neural network involved in the perception of meaningless actions. *Brain*, 122(10), 1875–1887. Neurology, <https://doi.org/10.1093/brain/122.10.1875>
- Hauser, M. D., & Spelke, E. (2004). Evolutionary and developmental foundations of human knowledge. *The cognitive neurosciences*, 3, 853–864.
- Haxby, J. V., Grady, C. L., Horwitz, B., Ungerleider, L. G., Mishkin, M., Carson, R. E., ... Rapoport, S. I. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proceedings of the National Academy of Sciences of the United States of America*, 88(5), 1621–1625.
- Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology Human Perception & Performance*, 33(1), 31–47. <https://doi.org/10.1037/0096-1523.33.1.31>
- Huettel, S. A., Misiurek, J., Jurkowski, A. J., & McCarthy, G. (2004). Dynamic and strategic aspects of executive processing. *Brain Research*, 1000(1–2), 78–84. <https://doi.org/10.1016/j.brainres.2003.11.041>
- Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., & Rizzolatti, G. (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology*, 3(3), e79. <https://doi.org/10.1371/journal.pbio.0030079>
- Jacquet, P. O., & Avenanti, A. (2013). Perturbing the action observation network during perception and categorization of actions' goals and grips: State-dependency and virtual lesion TMS effects. *Cerebral Cortex (New York, N.Y. : 1991)*. <https://doi.org/10.1093/cercor/bht242>
- Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2), 825–841. <https://doi.org/10.1006/nimg.2002.1132>
- Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, 5(2), 143–156.
- Klauer, K. C., & Zhao, Z. (2004). Double dissociations in visual and spatial short-term memory. *Journal of Experimental Psychology: General*, 133(3), 355–381. <https://doi.org/10.1037/0096-3445.133.3.355>

- Kuo, B.-C., Yeh, Y.-Y., Chen, A. J. W., & D'Esposito, M. (2011). Functional connectivity during top-down modulation of visual short-term memory representations. *Neuropsychologia*, *49*(6), 1589–1596. doi:<https://doi.org/10.1016/j.neuropsychologia.2010.12.043>
- Leavitt, M. L., Mendoza-Halliday, D., & Martinez-Trujillo, J. C. (2017). Sustained activity encoding working memories: Not fully distributed. *Trends in Neurosciences*, *40*(6), 328–346. <https://doi.org/10.1016/j.tins.04.004>
- LaRocque, J. J., Riggall, A. C., Emrich, S. M., & Postle, B. R. (2017). Within-category decoding of information in different attentional states in short-term memory. *Cerebral Cortex*, *27*(10), 4881–4890. <https://doi.org/10.1093/cercor/bhw283>
- Lee, T. G., & D'Esposito, M. (2012). The dynamic nature of top-down signals originating from prefrontal cortex: A combined fMRI-TMS study. *Journal of Neuroscience*, *32*(44), 15458–15466.
- Li, S., Cai, Y., Liu, J., Li, D., Feng, Z., Chen, C., & Xue, G. (2017). Dissociated roles of the parietal and frontal cortices in the scope and control of attention during visual working memory. *NeuroImage*, *149*, 210–219. doi:<https://doi.org/10.1016/j.neuroimage.2017.01.061>
- Li, S.-C., Lindenberger, U., & Sikström, S. (2001). Aging cognition: From neuromodulation to representation. *Trends in Cognitive Sciences*, *5*(11), 479–486.
- Lingnau, A., & Downing, P. E. (2015). The lateral occipitotemporal cortex in action. *Trends in Cognitive Sciences*, *19*(5), 268–277. doi:<https://doi.org/10.1016/j.tics.2015.03.006>
- Lu, X., Huang, J., Yi, Y., Shen, M., Weng, X., & Gao, Z. (2016). Holding biological motion in working memory: An fMRI study. *Frontiers in Human Neuroscience*, *10*, (251). <https://doi.org/10.3389/fnhum.2016.00251>
- Malhotra, P., Coulthard, E. J., & Husain, M. (2009). Role of right posterior parietal cortex in maintaining attention to spatial locations over time. *Brain*, *132*(3), 645–660. <https://doi.org/10.1093/brain/awn350>
- McLaren, D. G., Ries, M. L., Xu, G., & Johnson, S. C. (2012). A generalized form of context-dependent psychophysiological interactions (gPPI): A comparison to standard approaches. *NeuroImage*, *61*(4), 1277–1286. <https://doi.org/10.1016/j.neuroimage.2012.03.068>
- Nelissen, N., Stokes, M., Nobre, A. C., & Rushworth, M. F. (2013). Frontal and parietal cortical interactions with distributed visual representations during selective attention and action selection. *Journal of Neuroscience*, *33*(42), 16443–16458. <https://doi.org/10.1523/JNEUROSCI.2625-13.2013>
- Nichols, T., Brett, M., Andersson, J., Wager, T., & Poline, J. B. (2005). Valid conjunction inference with the minimum statistic. *NeuroImage*, *25*(3), 653–660. <https://doi.org/10.1016/j.neuroimage.2004.12.005>
- Park, D. C., Polk, T. A., Park, R., Minear, M., Savage, A., & Smith, M. R. (2004). Aging reduces neural specialization in ventral visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(35), 13091–13095. <https://doi.org/10.1073/pnas.0405148101>
- Park, J., Carp, J., Kennedy, K. M., Rodrigue, K. M., Bischof, G. N., Huang, C.-M., ... Park, D. C. (2012). Neural broadening or neural attenuation? Investigating age-related dedifferentiation in the face network in a large lifespan sample. *Journal of Neuroscience*, *32*(6), 2154–2158. <https://doi.org/10.1523/JNEUROSCI.4494-11.2012>
- Payer, D., Marshuetz, C., Sutton, B., Hebrank, A., Welsh, R. C., & Park, D. C. (2006). Decreased neural specialization in old adults on a working memory task. *Neuroreport*, *17*(5), 487–491. <https://doi.org/10.1097/01.wnr.0000209005.40481.31>
- Petzschner, F. H., & Krüger, M. (2012). How to reach: Movement planning in the posterior parietal cortex. *Journal of Neuroscience*, *32*(14), 4703–4704. <https://doi.org/10.1523/jneurosci.0566-12.2012>
- Pokorny, J. J., Hatt, N. V., Colombi, C., Vivanti, G., Rogers, S. J., & Rivera, S. M. (2015). The action observation system when observing hand actions in autism and typical development. *Autism Research*, *8*(3), 284–296. <https://doi.org/10.1002/aur.1445>
- Postle, B. R., Stern, C. E., Rosen, B. R., & Corkin, S. (2000). An fMRI investigation of cortical contributions to spatial and nonspatial visual working memory. *NeuroImage*, *11*(5), 409–423. doi:<https://doi.org/10.1006/nimg.2000.0570>
- Postle, B. R. (2015). The cognitive neuroscience of visual short-term memory. *Current Opinion in Behavioral Sciences*, *1*, 40–46. <https://doi.org/10.1016/j.cobeha.2014.08.004>
- Repovs, G., & Barch, D. M. (2012). Working memory related brain network connectivity in individuals with schizophrenia and their siblings. *Frontiers in Human Neuroscience*, *6*, 137. <https://doi.org/10.3389/fnhum.2012.00137>
- Riggall, A. C., & Postle, B. R. (2012). The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *Journal of Neuroscience*, *32*(38), 12990–12998. <https://doi.org/10.1523/jneurosci.1892-12.2012>
- Rizzolatti, G. (2005). The mirror neuron system and its function in humans. *Anatomy and Embryology*, *210*(5–6), 419–421.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*(1), 169–192. <https://doi.org/10.1146/annurev.neuro.27.070203.144230>
- Rouder, J. N., Morey, R. D., Morey, C. C., & Cowan, N. (2011). How to measure working memory capacity in the change detection paradigm. *Psychonomic Bulletin & Review*, *18*(2), 324–330. <https://doi.org/10.3758/s13423-011-0055-3>
- Rumiati, R. I., Weiss, P. H., Tessari, A., Assmus, A., Zilles, K., Herzog, H., & Fink, G. R. (2005). Common and differential neural mechanisms supporting imitation of meaningful and meaningless actions. *Journal of Cognitive Neuroscience*, *17*(9), 1420–1431. <https://doi.org/10.1162/0898929054985374>
- Shen, M., Gao, Z., Ding, X., Zhou, B., & Huang, X. (2014). Holding biological motion information in working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *40*(4), 1332.
- Silvanto, J., & Cattaneo, Z. (2010). Transcranial magnetic stimulation reveals the content of visual short-term memory in the visual cortex. *NeuroImage*, *50*(4), 1683–1689. <https://doi.org/10.1016/j.neuroimage.2010.01.021>
- Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology*, *33*(1), 5–42. doi:<https://doi.org/10.1006/cogp.1997.0658>
- Smyth, M. M., & Scholey, K. A. (1994). Interference in immediate spatial memory. *Memory & Cognition*, *22*(1), 1–13.
- Spelke, E. S. (2000). Core knowledge. *American Psychologist*, *55*(11), 1233.
- Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental science*, *10*(1), 89–96.
- Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, *78*(2), 364–375.
- Sugase-Miyamoto, Y., Liu, Z., Wiener, M. C., Optican, L. M., & Richmond, B. J. (2008). Short-term memory trace in rapidly adapting synapses of inferior temporal cortex. *PLoS Computational Biology*, *4*(5), e1000073. <https://doi.org/10.1371/journal.pcbi.1000073>
- Thakkar, K. N., Peterman, J. S., & Park, S. (2014). Altered brain activation during action imitation and observation in schizophrenia: A translational approach to investigating social dysfunction in schizophrenia. *American Journal of Psychiatry*, *171*(5), 539–548. <https://doi.org/10.1176/appi.ajp.2013.13040498>

- Todd, J. J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, 428(6984), 751–754.
- Von Hofsten, C., & Rosander, K. (2012). Perception-action in children with ASD. *Frontiers in Integrative Neuroscience*, 6, (115). <https://doi.org/10.3389/fnint.2012.00115>
- Voss, M. W., Erickson, K. I., Chaddock, L., Prakash, R. S., Colcombe, S. J., Morris, K. S., ... Kramer, A. F. (2008). Dedifferentiation in the visual cortex: An fMRI investigation of individual differences in older adults. *Brain Research*, 1244, 121–131. <https://doi.org/10.1016/j.brainres.2008.09.051>
- Wood, J. N. (2007). Visual working memory for observed actions. *Journal of Experimental Psychology: General*, 136(4), 639–652. <https://doi.org/10.1037/0096-3445.136.4.639>
- Wood, J. N. (2008). Visual memory for agents and their actions. *Cognition*, 108(2), 522–532. <https://doi.org/10.1016/j.cognition.2008.02.012>
- Wood, J. N. (2011). A core knowledge architecture of visual working memory. *Journal of Experimental Psychology Human Perception & Performance*, 37(2), 357–381. <https://doi.org/10.1037/a0021935>
- Wood, J. N. (2011). When do spatial and visual working memory interact? *Attention, Perception, & Psychophysics*, 73(2), 420–439. <https://doi.org/10.3758/s13414-010-0048-8>
- Woolrich, M. W., Behrens, T. E., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for fMRI group analysis using Bayesian inference. *NeuroImage*, 21(4), 1732–1747. <https://doi.org/10.1016/j.neuroimage.2003.12.023>
- Xia, M., Wang, J., & He, Y. (2013). BrainNet viewer: A network visualization tool for human brain connectomics. *PLoS One*, 8(7), e68910. <https://doi.org/10.1371/journal.pone.0068910>
- Xu, Y. (2017). *Reevaluating the sensory account of visual working memory storage*. *Trends in Cognitive Sciences*, 21(10), 794. Vol. ).
- Xu, Y., & Chun, M. M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, 440(7080), 91–95. <https://doi.org/10.1038/nature04262>
- Zheng, L., Gao, Z., Xiao, X., Ye, Z., Chen, C., & Xue, G. (2017). Reduced fidelity of neural representation underlies episodic memory decline in normal aging. *Cerebral Cortex (New York, N.Y. : 1991)*, 1–14.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Cai Y, Urgolites Z, Wood J, et al. Distinct neural substrates for visual short-term memory of actions. *Hum Brain Mapp*. 2018;39:4119–4133. <https://doi.org/10.1002/hbm.24236>