Goal-dependent dissociation of visual and prefrontal cortices during working memory

Sue-Hyun Lee, **Dwight J. Kravitz**, and **Chris I. Baker**
Laboratory of Brain and Cognition, National Institute of Mental Health, National Institutes of Health, Bethesda, MD 20892

Abstract

To determine the specific contribution of brain regions to working memory, human participants performed two distinct tasks on the same visually-presented objects. During the maintenance of visual properties object identity could be decoded from extrastriate, but not prefrontal, cortex, whereas the opposite held for non-visual properties. Thus, the ability to maintain information during working memory is a general and flexible cortical property with the role of individual regions being goal-dependent.

The role of individual regions (including parietal, extrastriate and prefrontal cortex) during working memory has long been debated (e.g.[1]). Early proposals suggested that lateral prefrontal cortex (lPFC) both controls working memory and directly maintains information. In contrast, more recent studies of *visual* working memory, have suggested that while lPFC initiates and controls working memory, maintenance occurs in posterior cortices[2,3]. This new evidence has relied critically on multivariate analyses showing that the contents of visual working memory can be decoded from occipitotemporal[2,4–6] and parietal[3] (but see[2,6]) cortex but not the lPFC[2,3], despite activation in all regions. However, there is also evidence that lPFC plays a role in the maintainenance of *non-visual* (e.g. verbal, conceptual) content. For example, impairments have been reported with transcranial magnetic stimulation (TMS) to the lPFC during the maintenance of non-visual, but not visual information[7].

To reconcile these findings, we propose that the functional role of regions within the working memory network is flexible, dependent on the nature of maintained information. To test this hypothesis we designed an experiment in which the same visual stimuli were presented but the maintained information differed due to task demands. Participants saw two sample objects (Fig. 1a) before being cued as to which object was relevant for that trial (see also[4]). In visual task runs, participants had to indicate whether the probe (an object fragment) belonged to the cued object or not, requiring the maintenance of visual features.

Corresponding author: Lee, S-H (sue-hyun.lee@nih.gov).

No conflicts of interest.

Author contributions
S.L., D.J.K., and C.I.B. designed the research. S.L. performed the research and analyzed the data. D.J.K. contributed analytic tools. C.I.B. supervised the project. S.L., D.J.K., and C.I.B. wrote the manuscript.

In non-visual task runs, participants had to indicate whether the probe (a whole object) was from the same subcategory as the cued object or not (Fig. 1b) requiring the maintenance of the name or subcategory of the object. This equated the stimulus presentation across tasks. We also utilized ultra-high field(7T)functional magenetic resonance imaging (fMRI) in order to collect high resolution data (1.8mm isotropic voxels) across the whole brain to test decoding across the entire working memory network simultaneously (see[8] for a discussion of the importance of decoding in these investigations).

To verify the nature of the information maintained during the delay period of each task, we first performed a separate behavioral study outside the scanner (Fig. 2a). During the delay, we presented either visual object or word distractors and measured their impact on task performance. To ensure that participants actively processed the distractors, they had to respond whenever they saw a nonsense object or word in the sequence. The visual object distractors impaired performance on the visual more than the non-visual task, whereas the word distractors showed the opposite pattern yielding a highly significant interaction in a two-way ANOVA between Task (visual, non-visual) and Distractor Type (object, word) ($F_{(1,13)} = 15.05$, $p < 0.01$). These results establish the maintenance of visual information during the visual task and non-visual information during the non-visual task (Fig. 2b).

In the fMRI experiment, a separate group of participants performed either the visual or non-visual task (with no distractors) in each of eight interleaved runs(Fig. 1b, Supplementary Table 1). The primary analyses focused on two regions-of-interest (ROIs), lateral prefrontal cortex (lPFC) and posterior fusiform cortex (pFs), identified in independent localizer scans (Fig. 3a ; see Supplemental Materials for whole-brain analyses). lPFC is commonly activated in working memory studies[9,10], and pFs is thought to be involved in the high-level processing and memory for visual objects[11].

During the delay period of each task, both lPFC and pFs were active but there was no difference in the BOLD response magnitude in either ROI (Fig. 3b, Supplementary Figs. 1 and 2). A two-way ANOVA with Task and ROI (lPFC, pFs) as within-subject factors revealed only a main effect of ROI ($F_{(1,21)}=23.75$, $p<0.01$) and no main effects or interactions involving Task (all $F_{(1,21)}<2.72$, $p > 0.11$).

We next trained support vector machines (SVMs) on the patterns of response during the delay period to test whether the identity of the maintained object could be decoded[8]. Stronger decoding was observed during the maintenance of visual than non-visual properties in pFs, whereas the opposite was true in lPFC (Fig. 3c, Supplementary Figs. 3 and 4). A two-way ANOVA with ROI and Task as within-subject factors revealed a highly significant interaction between ROI and Task ($F_{(1,21)}=11.44$, $p<0.01$), but no main effects (both $F_{(1,21)}<0.04$, $p>0.84$). In pFs, decoding during the visual ($t_{(21)}=2.67$, $p < 0.01$) but not non-visual ($t_{(21)}=-0.12$, $p>0.45$) task was significantly greater than chance and there was a significant difference between them ($t_{(21)}=2.07$, $p<0.05$) (Fig. 3c). In lPFC, the opposite pattern was observed (Fig. 3c), with significant decoding during the non-visual ($t_{(21)}=1.73$, $p<0.05$) but not visual task ($t_{(21)}=-0.37$, $p>0.35$) and a significant difference between them ($t_{(21)}=2.31$, $p<0.05$). These results demonstrate a task-dependent functional dissociation between prefrontal and posterior cortices, with decoding during the delay determined by the

nature of the maintained information. Thus, both prefrontal and posterior cortices are capable of maintaining information during working memory, but are directly involved only in the maintenance of the types of information they generally process: visual object identity information in pFs (e.g.[12]) and verbal or conceptual information in lPFC[7].

This framework may help reconcile the results of prior studies of lateral prefrontal cortex. Our results for the visual task are consistent with recent visual working memory studies emphasizing the role of prefrontal cortex only in the control of working memory[3,6]. Further, our results for the non-visual task are consistent with prior TMS data[7] and single cell recording results in monkeys[13] that strongly suggest information maintenance in lateral prefrontal areas for abstract information. Importantly, prefrontal cortex appears capable of both control (e.g.[14]) and maintenance in certain contexts, necessitating a reformulation of its general role in working memory. Future studies will need to shed light on whether these two functions arise from the same neuronal populations or from nearby but distinct populations.

Although we did not find any significant difference between tasks in the posterior parietal cortex (Supplementary Fig. 5), our formulation of working memory might also resolve a debate[15] about whether the parietal cortex is directly involved in maintenance[3] or only in control[2]. Parietal involvement in maintenance has been found for complex visuospatial patterns[3], which it is generally involved in processing (see[16] for a review), but not for simple direction or speed which are maintained in more posterior areas[2].

Our results add to growing evidence for the importance of internally generated top-down signals in defining representations in visual areas (e.g imagery[17], color knowledge[18]). Further research is needed to better understand both how these top-down signals interact with stimulus-driven processing and whether and how phenomenologically distinct processes, such as working memory, imagery, and attention, differ from one another within these cortical areas.

In summary, our results suggest that working memory is dependent on behavioral goals, with this context determining both the nature of the maintained information and the cortical regions involved. These results imply that the ability to maintain information is a general property of cortex, flexibly engaged in accordance with current behavioral goals. This model of working memory suggests a highly adaptable mechanism with implications for the function of many regions throughout the cortex.

Online Methods

Participants

22 neurologically intact and right-handed participants (age 25.50 ± 1.21 years, 11 females) took part in the fMRI experiment (2 additional participants were excluded because they did not respond in either task for more than 40% of trials). This sample size was justified based on other prior working memory studies using MVPA[3]. 14 additional neurologically intact and right-handed participants (age 24.57 ± 1.16 years, 7 females) took part in the behavioral experiment with distractors. All participants provided written informed consent for the procedure in accordance with protocols approved by the NIH Institutional Review Board.

Tasks and Stimuli

Two specific memory tasks (visual, non-visual) were performed in both the fMRI and behavioral experiments in separate runs. Each task required the maintenance of either visual or non-visual properties of the same set of sample objects. There were six sample object images: clock, watch, motorcycle, scooter, teapot, and kettle, consisting of three pairs of objects from the same general category but with distinct subcategory names (e.g. clock, watch) (Fig. 1a).

Each run contained one task, and before each run participants were instructed what task they would perform (Fig. 1b). On each trial of both tasks, a participant saw two unrelated sample objects presented sequentially, followed by a cue indicating which object had to be remembered during delay period (see also[4]) (Fig. 1b). After the delay period, a probe image was presented. Participants indicated their match/non-match judgment via a button box, and the button that indicated each response type was counterbalanced across participants.

In the visual task, the probe was an object fragment either from the cued image or from the other image in its general category. Participants were instructed to press the 'yes' button if the fragment image belonged to the cued object and otherwise press the 'no' button as quickly as possible. The size of the object fragment was approximately 1/12 of the whole object. This fragment size was chosen to force participants to maintain detailed visual information.

In the non-visual task, the probe was a whole object image either from the same or different but matched subcategory as the cued object (e.g. watch and clock). Participants were instructed to press 'yes' for the probe that was the same subcategory as the cued object and otherwise press 'no' as quickly as possible. Every trial of this non-visual task contained a new probe image, which was never presented as a sample image and never re-used on other trials. This probe required the participants to maintain the subcategory name of the cued object and rendered visual information ineffective in making the matching judgment. No specific strategy for performing each task was provided to the participants.

fMRI Experiment

Each task was presented in 4 scan runs consisting of 24 trials each. In each trial, two sample object images were presented sequentially with 1s duration for each and 750ms inter-sample interval. A roman numeral cue (I or II) was then presented for 500ms, followed by a 8s delay period and a 300ms probe. This cue indicated whether the first or second sample object was to be remembered across the delay. The length of each trial was 14s and the inter-trial interval (ITI) varied with an average ITI of 6s. The order of the sample objects and cues were randomized and counterbalanced across runs. 2 scan runs of each task (thus 48 trials) included the full set of the presentation cases for the all possible first and second objects and cues.

Behavioral experiment with distractors

The behavioral paradigm closely mirrored the fMRI paradigm (Fig. 2a) with the following exceptions. During the delay period, whose length varied (3.0, 3.8 or 4.8 s), a sequence of

either visual object or word distractors was presented. Object distractors were one of the four sample objects not used in that trial. Word distractors were the subcategory names of those same four images. These two types of distractors were presented during each of the two types of task, leading to a 2 × 2 factorial design: Distractor Type (object, word) x Task (visual, non-visual). This paradigm allowed us to test what information is being maintained in each task, by testing for a selective effect of each type of distractor in each task.

In order to further increase the salience and effectiveness of the distractors, targets requiring a response were embedded in the sequence (object: abstract embryos[19], words: nonwords (http://www.neuro.mcw.edu/mcword/)(Fig. 2a; green squares). During each sequence, 1 or 2 target embryos or nonwords were presented, and participants pressed a button whenever they saw a target. In both word or object distractor sequences, each distractor was presented for 170ms, with a 30ms interval preceding the next distractor during which only fixation cross was presented. The first distractor was presented 500ms after the initiation of the delay period, and the last distractor was presented 900ms before the termination of the delay period. When the length of the delay period was 3.0, 3.8 or 4.8 s, each potential distractor was repeated two, three, or four times, respectively. The number of targets embedded in the distracter sequence also scaled with the duration of the delay period (3.0, 3.8s: 1 target, 4.8s: 2 targets).

For the analyses of reaction time, median values were used to limit the influence of outliers. As in the fMRI experiment each task was presented in separate runs, with 4 runs of 48 trials each for each task with distractor type balanced across trials in each run. The order of the distractor sequences, distractor types (object, word), cues, and objects were randomized and counterbalanced.

Localizer Design

The Regions-of-Interest (ROIs) used for analysis were determined by two localizer scans (Fig. 3a). The object-selective localizer was used to identify object-selective regions of cortex. Participants viewed alternating 16 second blocks of grayscale object images and retinotopically matched scrambled images[17]. The resulting object-selective lateral occipital complex (LOC) can be divided into an anterior ventral (pFs) (Fig. 3a; left) and a posterior dorsal part (LO) (Supplementary Fig. 5d). The category task localizer was used to identify areas involved in processing the object category and/or name information required during the non-visual task. During this localizer scan, participants performed alternating 24s blocks of an object category and a color detection task. In the object category blocks, participants were presented object images and pressed the button whenever they saw two of the same object category in a row. In color detection blocks, participants were presented with just a fixation cross, periodically changing in color and pressed the button whenever they saw a red fixation cross. During each 24 second block participants had to respond three times. The resulting regions were found in the frontal, parietal, and visual cortices (Fig. 3a; right). Lateral prefrontal regions were used as lPFC in the later analyses (Fig. 3b, c, Supplementary Figs. 1 and 4), as was the activated region in the posterior parietal cortex (Supplementary Fig. 5a–c). Because there was no significant difference of the results between the left and right hemisphere in each ROI, analyses were collapsed across hemisphere.

fMRI data acquisition

Participants were scanned on the 7T Siemens scanner at the fMRI facility on the NIH campus in Bethesda. Images were acquired using a 32-channel head coil with an in-plane resolution of 1.78×1.78mm, and 52 1.8 mm slices (0.18 mm inter-slice gap, repetition time [TR] = 2s, echo time [TE] = 25ms, matrix size = 108×108, field of view (FOV) = 192mm). Whole brain volumes were scanned, and our slices were oriented approximately parallel to the base of the temporal lobe. All functional localizer and main task runs were interleaved. Standard MPRAGE (magnetization-prepared rapid-acquisition gradient echo) and corresponding GE-PD (gradient echo – proton density) images were collected before the experimental runs in each participant, and the MPRAGE images were then normalized by the GE-PD images for use as a high-resolution anatomical data for the following fMRI data analysis[20].

fMRI Data Analysis

Data analysis was conducted using AFNI (http://afni.nimh.nih.gov/afni)[21], SUMA (AFNI surface mapper), FreeSurfer, and custom MATLAB scripts. Data preprocessing included slice-time correction, motion-correction, and smoothing (smoothing was performed only for the localizer data, not the event-related data, with Gaussian blur of 5mm full-width half-maximum (fwhm)).

To derive the BOLD response magnitudes during the visual and nonvisual tasks, we deconvolved the responses using the AFNI software package (3dDeconvolve using TENT function with motion parameters and up to 4th order polynomials treated as regressors of no interest) and derived the β-value of each voxel for each of the 10 TRs 0–18s after the initiation of each trial. The response magnitudes during the delay period were produced by averaging the β-values of individual voxels across the TRs 6–10s after the onset of each trial (Fig. 3b and Supplementary Figs. 2 and 5b,e). To visualize ROI-specific response, the response magnitudes were averaged across all voxels within each ROI at each TR (Supplementary Fig. 1) or averaged across the delay period (Fig. 3b, Supplementary Fig. 5b, e).

To decode maintained information, we used multi-class classification using a linear support vector machine (SVM) approach. We used the LIBSVM package developed by Chang and Lin (http://www.csie.ntu.edu.tw/~cjlin/libsvm/). The 4 runs of each task were divided into two halves, each of which included full set of the presentation cases (see above '*fMRI Experiment*' section), in two possible ways. For each of the splits, β-values were extracted from the voxels within each ROI (the ROI analyses) or a sphere (radius = 6.3mm, corresponding to ~123 voxels) across the whole brain (the searchlight analysis[22]), and then normalized in each voxel by subtracting the mean value across all object conditions[17,23]. To decode the information during the delay period of each task, the patterns of the response averaged across time points 6–10s (after the onset of each trial) from half of the total 4 runs of each task, were used to train a classifier for cued object identity and the patterns averaged across time points 6–10s from the other half was used to determine the classification accuracy, exhaustively (Fig. 3c, Supplementary Fig. 3, Supplementary Fig. 5c, f). For the decoding analysis at each TR, the linear classifier was trained using the response from TRs

corresponding with 6–10s following trial onset, but then tested on all 10 TRs (Supplementary Fig. 4)[4]. All the main analyses of the fMRI data included data from all trials, as an incorrect response does not necessarily indicate that the wrong object was held in memory. Identical analyses performed on only the correct trials revealed nearly identical results. A two-way ANOVA with ROI and Task as within-subject factors on decoding results based only on the correct trials, revealed a highly significant interaction between ROI and Task ($F_{(1,21)} = 13.52$, $p < 0.01$), but no main effects of ROI or Task (both $F_{(1,21)} < 0.02$, $p > 0.89$).

We also performed the same decoding analyses based on the response of an anatomically defined lateral prefrontal cortex. For this analysis, we used rostral-middle-frontal region defined by the cortical parcellation of FreeSurfer[24], and found nearly identical results to those from functionally defined lPFC. While the BOLD response magnitudes did not show any significant difference between tasks ($t_{(21)} = 0.74$, $p > 0.46$), decoding during the non-visual task(but not visual task) was significantly greater than chance($t_{(21)} = 3.11$, $p < 0.01$) and greater than that during the visual task ($t_{(21)} = 3.18$, $p < 0.01$).

Before mapping data from the activation or searchlights to the standard-mesh surfaces reconstructed by AFNI and SUMA (Supplementary Figs. 2 and 3)[25], we smoothed the analyzed data with Gaussian blur of 5mm fwhm.

Statistical Analyses

As all comparisons were planned, to compare activation or decoding results with basal level (zero or chance level, respectively), we used one-sample t-tests (one-tailed). When we compared behavior, activation, or decoding results between tasks, we mainly used two-way ANOVAs (tests of within-subjects effects)to examine the effects of ROI (or Distractor), Task, and the interaction between them. If any interaction effect was revealed, we used one-tailed paired t-tests to examine detailed effects between tasks with the assumption of predicted direction. Otherwise we used two-tailed paired t-tests. To reveal the effects of ROI, Task, timepoints, and their interactions, we used three-way ANOVAs (tests of within-subjects effects). Greenhouse-Geisser corrections were used for all ANOVAs with factors with more than 2 levels. Error bars in all plots indicate one between-subject standard error.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. D'Esposito M. Philos Trans R Soc Lond B Biol Sci. 2007; 362:761–72. [PubMed: 17400538]

2. Riggall AC, Postle BR. J Neurosci. 2012; 32:12990–8. [PubMed: 22993416]

3. Christophel TB, Hebart MN, Haynes JD. J Neurosci. 2012; 32:12983–9. [PubMed: 22993415]

4. Harrison SA, Tong F. Nature. 2009; 458:632–635. [PubMed: 19225460]

5. Serences JT, Ester EF, Vogel EK, Awh E. Psychol Sci. 2009; 20:207–14. [PubMed: 19170936]

6. Linden DEJ, Oosterhof NN, Klein C, Downing PE. J Neurophysiol. 2012; 107:628–39. [PubMed: 22013235]

7. Romero L, Walsh V, Papagno C. J Cogn Neurosci. 2006; 18:1147–55. [PubMed: 16839288]

8. Lewis-Peacock JA, Postle BR. Neuropsychologia. 2012; 50:470–8. [PubMed: 22108440]

9. Curtis CE, D'Esposito M. Trends Cogn Sci. 2003; 7:415–423. [PubMed: 12963473]

10. Buckner RL, Wheeler ME. Nat Rev Neurosci. 2001; 2:624–634. [PubMed: 11533730]

11. Kravitz DJ, Saleem KS, Baker CI, Ungerleider LG, Mishkin M. Trends Cogn Sci. 2013; 17:26–49. [PubMed: 23265839]

12. Kravitz DJ, Kriegeskorte N, Baker CI. Cereb Cortex. 2010; 20:2916–2925. [PubMed: 20351021]

13. Funahashi S, Bruce CJ, Goldman-Rakic PS. J Neurophysiol. 1989; 61:331–349. [PubMed: 2918358]

14. Lee TG, D'Esposito M. J Neurosci. 2012; 32:15458–15466. [PubMed: 23115183]

15. Sligte IG, Van Moorselaar D, Vandenbroucke ARE. J Neurosci. 2013; 33:1293–4. [PubMed: 23345204]

16. Kravitz DJ, Saleem KS, Baker CI, Mishkin M. Nat Rev Neurosci. 2011; 12:217–230. [PubMed: 21415848]

17. Lee SH, Kravitz DJ, Baker CI. NeuroImage. 2012; 59:4064–4073. [PubMed: 22040738]

18. Simmons WK, et al. Neuropsychologia. 2007; 45:2802–2810. [PubMed: 17575989]

19. Brady MJ, Kersten D. J Vis. 2003; 3:413–422. [PubMed: 12901712]

20. Van De Moortele PF, et al. NeuroImage. 2009; 46:432–446. [PubMed: 19233292]

21. Cox RW. Comput Biomed Res. 1996; 29:162–173. [PubMed: 8812068]

22. Kriegeskorte N, Goebel R, Bandettini P. Proc Natl Acad Sci U S A. 2006; 103:3863–3868. [PubMed: 16537458]

23. Haxby JV, et al. Science. 2001; 293:2425–2430. [PubMed: 11577229]

24. Desikan RS, et al. NeuroImage. 2006; 31:968–980. [PubMed: 16530430]

25. Saad ZS, Reynolds RC. NeuroImage. 2012; 62:768–73. [PubMed: 21945692]
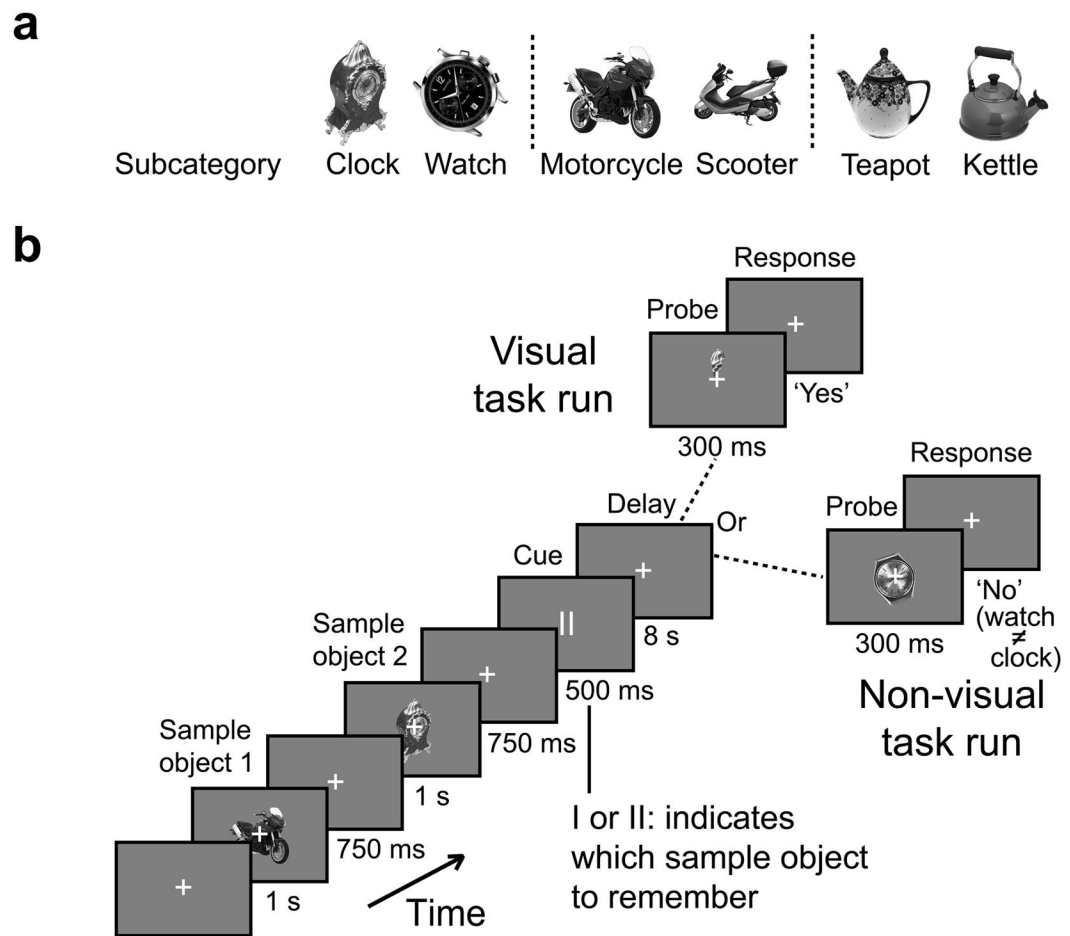
Figure 1. Experimental design

(a) Object images were from three broad categories (indicated by dotted lines) with two subcategory exemplars each. (b) Visual and non-visual tasks were performed in separate runs and required maintenance of either visual or non-visual information during the delay period. In both tasks, two objects were presented sequentially followed by a cue, a delay, and a probe image. In the visual task, the probe was a partial object image, and participants indicated whether it came from the cued object. In the non-visual task, the probe was a whole object image, and participants determined whether it was from the same subcategory as the cued object.
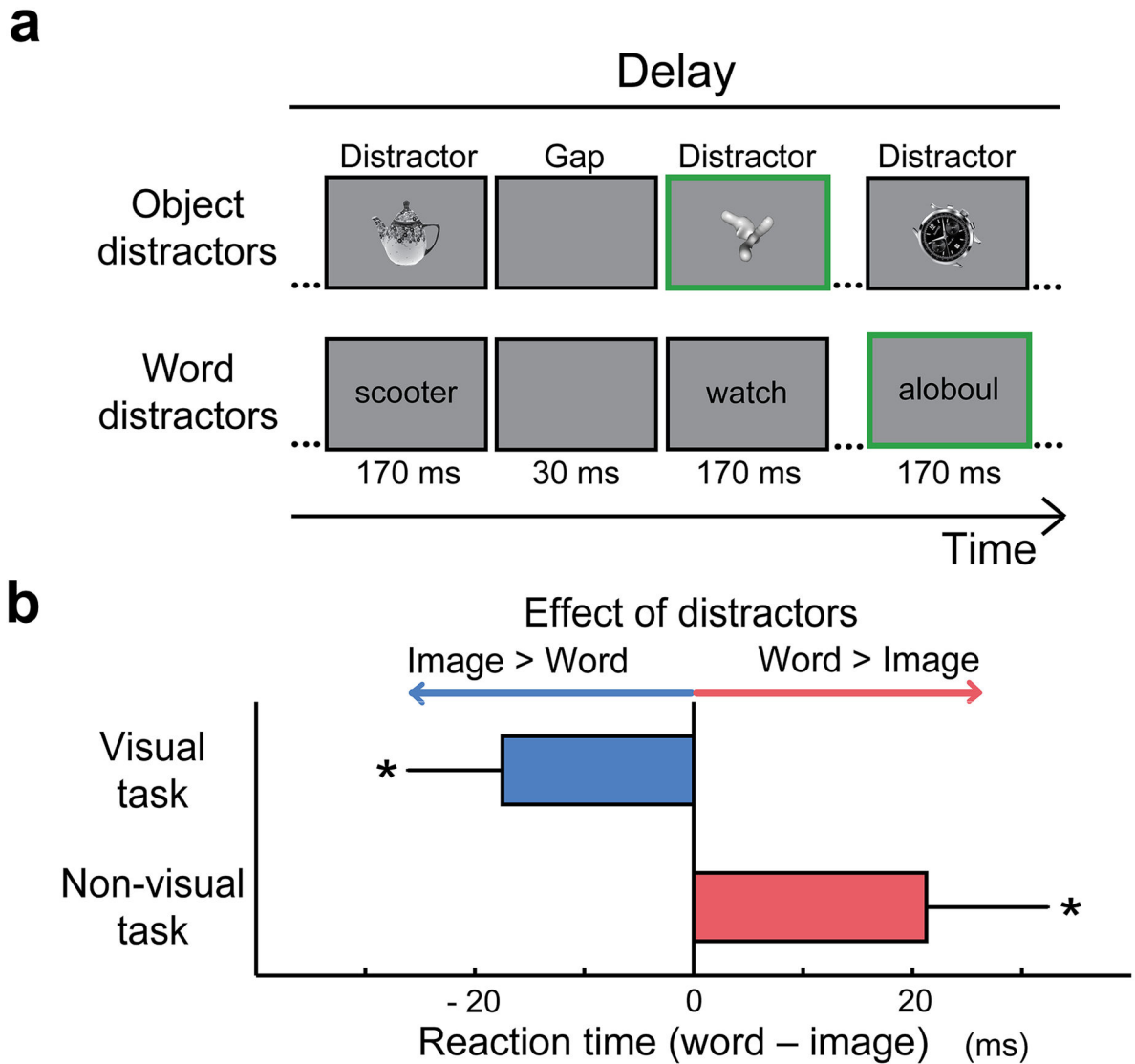
Figure 2. Nature of maintained information

(a) Distractors presented during the delay period of the visual or non-visual task. Participants responded whenever they saw a nonsense object or non-word (green boxes) (b) The visual task was more impaired with object than word distractors, whereas the opposite was true in the non-visual task. Bars show the mean RT difference between the two distractor types in the two tasks across participants (n=14). Error bars in this and all subsequent figures indicate the between-subject standard error. (*: p<0.05, $t_{(13)}$> 1.98)
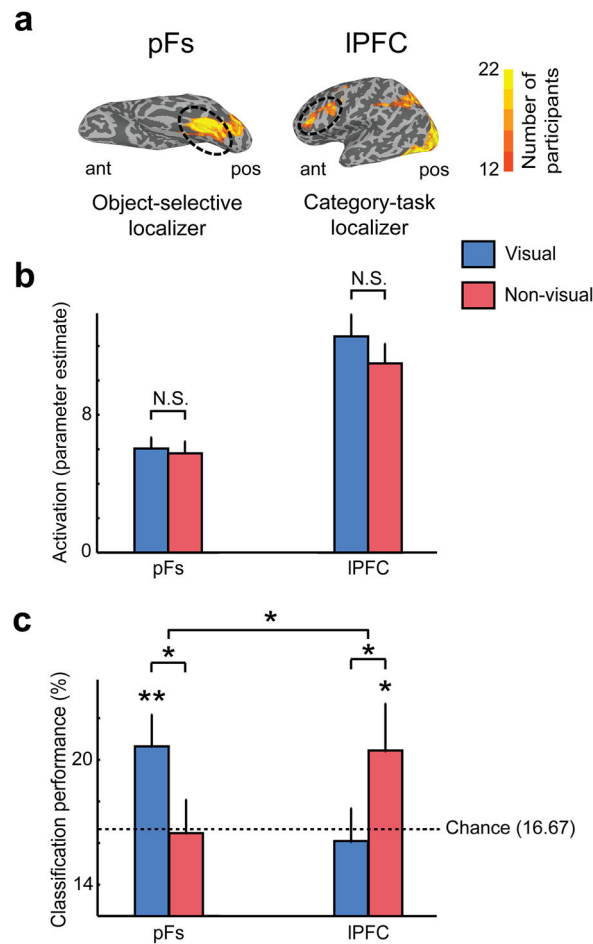
Figure 3. Decoding of object identity
(a) Overlap map across participants (n=22) for object-selective and category-task localizers (see Online Methods) showing location of pFs and lPFC ROIs (dotted circles). (b) No effect of task on response magnitude. (c) Decoding of object identity was significantly greater for the visual than non-visual task in pFs while the opposite was true in lPFC.(*: p<0.05; **: p<0.01)