Published in final edited form as:
J Neurosci. 2008 August 27; 28(35): 8765–8771. doi:10.1523/JNEUROSCI.1953-08.2008.

Temporary Activation of Long-Term Memory Supports Working Memory

Jarrod A. Lewis-Peacock and **Bradley R. Postle**
Department of Psychology, University of Wisconsin-Madison, 1202 W Johnson St, Madison, WI 53706, USA

Abstract

This paper describes a fMRI study of humans engaged in long-term memory (LTM) and working memory tasks. A pattern classifier learned to identify patterns of brain activity associated with viewing and making judgments about three categories of pictures (famous people, famous locations, and common objects). The evaluation of these stimuli relied on perception and long-term semantic and/or episodic memories. We investigated whether this classifier could successfully decode brain activity from a subsequent delayed paired-associate recognition working memory task that required the short-term retention of the same stimuli. We reasoned that the LTM-trained classifier would be able to decode delay-period activity only if that activity reflected, to some extent, the temporary activation of LTM. Our results demonstrated successful decoding: delay-period activity from a distributed network of brain regions matched learned patterns of activity for task-relevant stimuli to a greater extent than task-irrelevant stimuli. In varying degrees throughout the delay, activity reflected the target (a retrospective code) and its associate (a prospective code), with considerable variability among subjects. Although PFC demonstrated category-specific patterns of activity during the LTM task, these patterns were not reinstated in PFC during the working memory task. We conclude that the short-term retention of information can be supported by the temporary reactivation of LTM representations.

Keywords

memory; long-term memory; working memory; paired association; fMRI; pattern classification

INTRODUCTION

Working memory refers to the retention of information in conscious awareness when this information is not present in the environment, to its manipulation, and to its use in guiding behavior. For decades, an influential view has held that working memory functions are supported by the operation of specialized systems that act as buffers for the storage and manipulation of information (Baddeley and Hitch, 1974; Goldman-Rakic, 1987). Neurobiological evidence consistent with this view includes evidence suggesting that prefrontal cortex (PFC) may be a neural substrate for storage buffers of this model (e.g., Haxby et al., 2000; Constantinides et al., 2001; Leung et al., 2002; Courtney, 2004; Narayanan et al., 2005; Zarahn et al., 2005). Further, many neurobiological models of working memory function assume a critical role for PFC neurons whose specialized properties support sustained delay-period activity (e.g., Machens et al., 2005; Durstewitz and Seamans, 2006; Mongillo et al.,

Corresponding author with complete address, including an email address Jarrod A. Lewis-Peacock, Department of Psychology, University of Wisconsin-Madison, 1202 W Johnson St, Madison, WI 53706 USA, Tel: +1 608-265-3888, Fax: +1 608-262-4029, Email: E-mail: lewispeacock@wisc.edu.

2008). However, there is also a large body of results from neuropsychological, neurophysiological, and neuroimaging studies that is difficult to reconcile with the depiction of the PFC as a critical substrate for working memory storage (reviewed in Postle, 2006). This points to an alternative view that depicts working memory as emergent from the coordinated recruitment, via attention, of brain systems that have evolved to accomplish sensory-, representation-, and action-related functions. This alternative, therefore, emphasizes the *temporary activation of LTM representations* (e.g., Anderson, 1983; Cowan, 1995; Oberauer, 2002; Ruchkin et al., 2003). This *activated LTM* model, however, also remains controversial (e.g., Baddeley, 2003; Düzel, 2003; Kroger, 2003; Logie and Della Salla, 2003; Majerus et al., 2003; Vallar, 2003).

Several previous studies have produced data that are consistent with the *activated LTM* model by demonstrating sustained delay-period activity in regions that are associated with LTM representation of the stimulus domain being tested. For example, face-specific delay-period activity has been localized to regions of inferior temporal cortex that are believed to support the perception and long-term retention of faces (e.g., Druzgal and D'Esposito, 2003; Postle et al., 2003; Ranganath et al., 2004a; Postle, 2005). Such results cannot be interpreted as direct tests of this model, however, because to do so would be to commit the logical fallacy of 'reverse inference' (in formal terms, affirming the consequent) (Poldrack, 2006). The present experiment was designed to support stronger inference than these earlier studies by applying multivariate pattern classification to fMRI data: First, we recorded distinct patterns of neural activity corresponding to the engagement of LTM processes; Second, we tested for the reinstatement of these patterns during the short-term retention of the same stimuli. Thereby, this study effected a decisive test of the hypothesis that the temporary activation of LTM representations contributes to the short-term retention of information in working memory.

MATERIALS AND METHODS

Subjects

Ten subjects (all right-handed; 7 men and 3 women; ages 19–32 years) were recruited from the undergraduate and medical campuses of the University of Wisconsin-Madison. None reported any medical, neurological, or psychiatric illness, and all gave informed consent.

Overview of Study

The experiment proceeded in three phases, each implemented with E-Prime software. First, subjects performed a LTM task for three categories of visual stimuli – famous people, famous locations, and common objects (Fig. 1A). The evaluation of these stimuli required perception and long-term semantic and/or episodic memories. (For example, to render a judgment of like or dislike for the stimulus "John Wayne," one would have to retrieve knowledge about the famous actor John Wayne, and perhaps also memories of scenes from his movies.) This procedure replicated that of a recent study of free recall which demonstrated that category-specific brain activity patterns arising from this stimulus-judgment task reliably reappeared just prior to the verbal recall of an item from a category (Polyn et al., 2005). Second, outside the scanner, subjects learned six pairings of 12 stimuli that were selected at random from the training set. Finally, subjects performed delayed paired-associate recognition (a "working memory task") on these stimulus pairs (Fig. 1B). A pattern classifier was trained to distinguish category-specific patterns of brain activity from the first task, and was then used to decode brain activity from the delay-period of the second task. We reasoned that the only way delay-period activity from the working memory task could be classified by a LTM-trained pattern classifier would be if delay-period activity significantly matched LTM patterns of activity, i.e., if the working memory task produced the temporary reinstatement of the LTM activity. In this

way, successful decoding of working memory activity would provide conclusive evidence for the *activated LTM* model.

Behavioral Procedures

1. Stimulus Judgment Task ("LTM task")—Subjects viewed a total of 90 stimuli drawn from three categories – 30 famous people, 30 famous locations, and 30 common objects – and indicated (on a 4-point Likert scale, using a stimulus-response box) how much they liked the celebrity, how much they would like to visit the location, or how often they encountered the object in everyday life. Each stimulus was presented one time only, for a total of 90 randomly ordered stimulus presentations. Each LTM trial consisted of a cue period (2 s), a stimulus period (5 s), and a judgment period (3 s). After each trial, subjects performed an arithmetic task (16 s) to reduce interference between trials (Polyn et al., 2005).

2. Paired-Associate Learning—Twelve stimuli (four people, four locations, and four objects) were then selected at random from the training set and paired arbitrarily so that one stimulus from each category was paired with each of the three categories. This produced three pairs with stimuli from the same category (e.g., $Person_1$-$Person_2$) and three pairs with stimuli from different categories (e.g., $Person_3$-$Location_1$). Outside the scanner, subjects learned these within-category and between-category pairs through repeated three-alternative forced choice testing (with foils drawn from the set of 12) until they achieved a criterion-level of performance of 24 consecutive correct trials.

3. Delayed Paired-Associate Recognition ("working memory task")—Subjects then performed delayed paired-associate recognition in the scanner. Each working memory trial consisted of a target stimulus (1 s), a delay period (11 s), a probe stimulus (1 s), and an ITI (13 s). Subjects indicated with a Yes/No button press whether the probe stimulus was the correct associate of the target stimulus. The trial depicted in Fig. 1B is an example of a *Person-Location* trial: the target is a person (John Wayne) and the probe is a location (Vatican City). Each stimulus pair was presented 12 times (six times in each direction, i.e. the person-location pair was presented six times with the person as target, and six times with the location as target). Therefore there were 36 (three pairs x 12 exposures) within-category trials, and 36 between-category trials. Trials were configured such that there was a probability of 0.5 that the probe stimulus was the correct associate of the target, with foils drawn equally from all categories; thus the probe stimulus belonged to the correct *category* on 2/3 of all trials. In addition, subjects performed 24 *no-memory* trials in which they were presented with a target stimulus not from the learned set. They were instructed ahead of time that these trials did not require memory because the target stimulus would always reappear as the probe stimulus; simply a "No" button press was required when the probe stimulus appeared.

Validation Subjects and Experimental Subjects

To validate our method we instructed two of the subjects (1 male, 1 female) to solve the working memory task using a prospective strategy (i.e. "as soon as you see the first picture, quickly recall its associate and hold the associate in mind for the duration of the delay period"). This created a situation in which we would know the contents of the subjects' working memory independent of the experimental data, and could thereby assess the validity of the classification of the delay-period activity. After successful validation, we recruited eight new subjects who received no instruction about performance strategy, allowing us to interrogate delay-period activity in an unbiased manner. The data from these "experimental" subjects were used to test our hypotheses. Statistical analyses in validation subjects focused on the within-subject differences in classifier estimates, and all between-category working memory trials were pooled together and treated as independent replications (n=2, df_{error}=71). For experimental subjects, statistical comparisons also focused on the within-subject differences in classifier

estimates, but trial-averaging was used, yielding a single set of data for each subject (n=8, df_{error}=7).

fMRI Acquisition and Preprocessing

Whole-brain images were acquired with a 3-T scanner (GE Signa VH/I). For all volunteers, we acquired high-resolution T1-weighted images (30 axial slices, 0.9375 × 0.9375 × 4 mm). We used a gradient-echo, echo-planar sequence (time repetition = 2000 ms, echo time = 50 ms) to acquire data sensitive to the blood oxygen level-dependent (BOLD) signal (Kwong et al. 1992; Ogawa et al. 1992) within a 64 × 64 matrix (30 axial slices coplanar with the T1 acquisition, 3.75 × 3.75 × 4 mm). Six scans of both the LTM and working memory tasks were obtained for each subject, each scan lasting 6 min 50 s (LTM) and 7 min 16 s (working memory). All task runs were preceded by 20 s of dummy pulses to achieve a steady state of tissue magnetization. Preprocessing of the functional data was done with the AFNI software package (Cox, 1996) using the following preprocessing steps, in order: correction for slice time acquisition and rigid-body realignment to the first volume from the experimental task with 3dvolreg; removal of signal spikes with 3dDespike; removal of the mean from each voxel and linear and quadratic trends from within each run with 3dDetrend; and correction for magnetic field inhomogeneities (using in-house software). Finally, functional data from the working memory task were aligned to data from the LTM task using 3dAllineate. Note that neither was spatial smoothing imposed nor were the data spatially transformed into a common atlas space prior to hypothesis testing. Rather, the data from each subject were analyzed in that subject's unsmoothed, native space. For classification analyses, a feature selection analysis of variance (ANOVA) was applied to the preprocessed images to select those voxels whose activity varied significantly (p<0.05) between people, location, and object categories over the course of the LTM experiment. These feature-selected voxels served as input nodes to the pattern classifiers (mean 8,489 voxels, SD 2,107).

Classifier Training

The Princeton Multi-Voxel Pattern Analysis Toolbox (MVPA) (http://www.csbmb.princeton.edu/mvpa), in conjunction with the Matlab Neural Network Toolbox, was used for all pattern classification analyses. Preprocessed fMRI signal from the initial 10 s of each LTM trial associated with the cue, stimulus, and judgment periods were used to train a two-layer feedforward neural network (via backpropagation) to distinguish three patterns of brain activity corresponding to the study of people, locations, and objects (Fig. 1A). To reduce prediction error in analyses involving the pattern classifier, the reported classifier outputs were the average of 50 backpropagation networks, each initialized with a different set of random weights (Bishop, 1995). Data for all classification analyses were shifted in time by 6 s to account for hemodynamic lag. We evaluated classification accuracy by training on five blocks of data (fMRI task runs) and testing on the novel sixth block. The blocks used for training were then rotated and a new block of data was tested until all six blocks of data had been classified.

Because interpretation of activity from the LTM task is central to the logic of this experiment, further procedural and theoretical discussion will be useful. Procedurally, inclusion of the cue and judgment periods in training (when the visual stimulus was absent) replicated the procedures of Polyn et al., 2005. To assess empirically whether these data should be included, we calculated the classification accuracy at each time point of the 10 s training window and found that category discrimination was well above chance throughout the entire period, with marginally enhanced discrimination during stimulus presentation. Thus, we are confident that comparable stimulus category-specific activity was being evoked during all three trial phases. At a theoretical level, it is important to establish whether or not working memory was engaged during the LTM task. (After all, successful decoding of working memory task data would not

be surprising if the LTM task also engaged working memory.) In this regard, it is helpful to emphasize that the *activated LTM* hypothesis posits a mechanism for the short-term retention of stimulus-specific information. From this perspective, it is clear that the LTM task did not require subjects to retain a stimulus representation across a delay period when the stimulus was not present in the environment. What *was* required in the LTM task, in addition to perception and retrieval from LTM, was the maintenance of the rules that guided behavior on this task. Although some definitions of working memory incorporate the retention of rules and/or of behavioral set, the *activated LTM* model does not relate to these constructs. Therefore, the LTM task was suitable for testing the *activated LTM* hypothesis.

Voxel Discrimination Maps

To assess the relative importance of different brain areas to the classification of the stimulus categories, we calculated, from the trained pattern classifier, which voxels discriminated between the patterns of activity corresponding to each stimulus category. We started with the voxel importance formula from Polyn et al.(2005), $imp_{ij} = w_{ij} * avg_{ij}$; where w_{ij} is the weight between input unit *i* and output unit *j* and avg_{ij} is the average activity of input *i* during study of category *j*. To quantify the discriminability of each voxel, we then computed this value for all output categories and created their ratio, yielding $discrim_{ij} = imp_{ij} / (imp_{ij} + imp_{ik} + imp_{il})$; where imp_{ij} is as before, and imp_{ik} *and* imp_{il} are the voxel importance values for categories *k* and *l*, respectively. These discrimination scores for each voxel gave an indication of how much that voxel activated a given output over and above the other outputs (e.g., the ratio indicated how much a voxel activated the “person” output in comparison to the “location” and “object” outputs). This analysis produced three separate distributions (centered on 0.33) of voxel-wise discrimination scores corresponding to each category. We used two standard deviations above the mean of each distribution as the threshold for including a voxel in the “discrimination map” for a given category.

Decoding Delay-Period Activity

A trained pattern classifier for each subject, trained on all six blocks of LTM data, was used to assess the extent to which category-specific patterns of brain activity could be identified in the delay-period of the subsequent delayed paired-associate recognition task. Preprocessed fMRI signal at intervals of TR = 2 s was classified from the initial 16 s of each working memory trial (Fig. 1B) associated with target presentation (1 s), delay period (11 s), probe presentation (1 s), and the first 3 s of ITI. Possible contamination of delay-period estimates from probe stimulus processing (after the 6 s hemodynamic adjustment) was not a concern, as this processing would be expected to introduce noise, not coherent activity. This follows from the fact that the stimulus presented as the probe was from the same category as the associate of the target on only 2/3 of the trials, the remaining trials presenting foils, of which were drawn from a different category than the expected associate (see *Delayed Paired-Associate Recognition* above).

Our paired-associate recognition procedure, modeled after prior studies in monkeys (Takeda et al., 2005), permitted us to investigate retrospective and prospective coding in working memory. Assuming that subjects solved the task by actively representing task-relevant information across the delay period, the delay-period brain activity would be expected to reflect either retention of a representation of the target stimulus (a “retrospective” code), or the recall (from LTM) and retention of a representation of the paired-associate stimulus (a “prospective code”), or some combination thereof. There was an important distinction in our design between the two types of delayed-recognition memory trials: *within-category* trials and *between-category* trials. For within-category trials, regardless of whether a subject maintained a retrospective representation of the target or a prospective representation of its associate, the delay-period activity patterns would be expected to reflect the same stimulus category.

Between-category trials were the critical trials, however, because they could distinguish retrospective from prospective neural representations. If, for example, a *person*-like delay-period activity pattern was identified on a *Person-Object* trial, this would indicate that the subject was remembering the item presented at the beginning of the trial. Prospective delay-period activity could only occur if, upon seeing the target stimulus (a person in this example), the subject retrieved from LTM the representation of its associate, an object, and retained this representation in working memory for the remainder of the trial. Only data for between-category trials are presented here. Results for within-category and no-memory control trials are available in the Supplemental Material.

It is important to note that support for the *activated LTM* hypothesis would come in the form of successful decoding of delay-period activity as *task-relevant* vs. *task-irrelevant*. Whether this task-relevant activity is in the form of prospective or retrospective activity is not critical for this principle hypothesis. Rather, the prospective vs. retrospective distinction was included 1) for validation of the methodology (see Fig. 4A), and Fig. 2) to explore at a finer grain of detail the nature of the short-term retention of information across individuals (see Fig. 4C).

Prefrontal Cortex Activity

In monkeys, neural activity in lateral PFC has been found to initially represent the target object (a retrospective code), but towards the end of the delay period, begin to reflect the anticipated associate object (a prospective code) (Rainer et al., 1999). In a different study, robust delay-period neuronal activity in inferior temporal cortex was found to actively maintain prospective information, while retrospective representations were attenuated and often replaced by distractor information (Takeda et al., 2005). Much of the work on human prospective memory has emphasized the role of the PFC in accounts that emphasize the planning and cognitive control aspects of prospective memory (e.g., D'Esposito et al., 2000; Curtis et al., 2004; Burgess et al., in press). However, there has also been evidence for prospective coding outside PFC, e.g. in caudate nucleus (Postle and D'Esposito, 2003), and in inferior temporal cortex (Ranganath et al., 2004b).

In the present study, we interrogated the role of PFC activity in working memory retention by attempting to decode delay-period activity from 1) only voxels located inside PFC (*PFC-Only* condition), and 2) all voxels posterior to PFC (*No-PFC* condition). Anatomically-derived PFC masks were generated for each subject in AFNI by backwards transforming a TT_Daemon atlas mask (consisting of Brodmann areas 8–11, 44–46) into that subject's native space. The number of voxels passing feature selection in each condition was 1,069 ± 283 for PFC-Only and 7,421 ± 1,907 for No-PFC. So that the number of voxels interrogated in each condition was equivalent, for each subject we sampled randomly from the No-PFC voxels and created 50 new masks that contained the same number of voxels as in the PFC-Only condition. These PFC-Only and (equivalently-sized) No-PFC masks were selectively applied during classification to constrain the voxel populations used for analyses.

RESULTS

Classifier Training

For all subjects, activity from the LTM task was reliably classified as consistent with the appropriate category of the trial (Fig. 2). Based on one-tailed *t* tests, classifier prediction accuracy was significantly above chance (0.333) for each trial type, with $p<0.001$, People: $t(9)=21.2$; Locations: $t(9)=11.0$; Objects: $t(9)=14.6$. We performed further analyses by 1) masking out all voxels in PFC (*No-PFC* condition), and 2) masking out all voxels outside PFC (*PFC-Only* condition). Classifier training on the LTM data, when restricted to these voxel populations, revealed successful discrimination of category-specific activity, significant at

$p<0.001$ (Fig. 2): (No-PFC): People: $t(9)=19.7$; Locations: $t(9)=10.7$; Objects: $t(9)=12.6$; (PFC-Only): People: $t(9)=9.5$; Locations: $t(9)=5.5$; Objects: $t(9)=8.8$. Although classification accuracy was greater in the No-PFC compared to the PFC-Only condition, the prediction accuracy in both conditions was reliably above chance. That is, activity both in PFC and in more posterior brain regions demonstrated reliable category-specific patterns of activity during the LTM task.

Distributed Representations

Average voxel activity and network weights from the pattern classifier were analyzed to estimate the extent to which each voxel discriminated between the three categories (see Materials and Methods). Fig. 3 illustrates the voxels that exerted the strongest influence in discriminating each of the three stimulus categories (for 1 representative subject). Although canonical category-selective areas contributed to the classification of the three categories (e.g. the mid-fusiform gyrus for person stimuli, parahippocampal gyrus for location stimuli, and lateral occipital cortex for object stimuli), these regions did not solely drive classification. Each was a component of a distributed network of brain regions involved in the classification of a particular category. This replicates the findings of Polyn et al., 2005. Data summarizing the category discrimination analysis in specific anatomical regions across all 10 subjects are presented in Table 1.

Decoding Delay-Period Activity

Subjects performed the delayed paired-associate recognition trials with near perfect accuracy (98% ± 0.05). All between-category trials (correct and incorrect) were included in the analyses.

Validation Subjects

Classification results for the two subjects whose data served as procedural validation are shown in Fig. 4A. These subjects were instructed to concentrate on the anticipated associate stimulus during the delay period – i.e., to solve the working memory task using a prospective strategy. The classifier reliably identified their delay-period activity as reflecting a prospective representation of the associate throughout the delay-period, paired $t(71)=4.9$, $p<0.001$ (Fig. 4B).

Experimental Subjects

Having validated our method by demonstrating successful decoding of brain activity in subjects for whom we had independent knowledge of their mental state, we now consider the eight experimental subjects in Fig. 4C. Critically, the pattern classifier reliably decoded delay-period activity as *task-relevant*. Classifier estimates of task-relevant retrospective and prospective activity together were reliably higher throughout the delay period than estimates of task-irrelevant activity, paired $t(7)=3.1$, $p<0.01$ (Fig. 4D). Had delay-period activity not reinstated the activity patterns from the LTM task, in contrast, the classifier would have been expected to produce indistinguishable estimates for each category. (This was verified empirically by randomizing the labels of delay-period data at each time point and re-classifying, which produced statistically inseparable classification estimates.) For these eight subjects, prospective activity was not separable from retrospective activity until the very end of the delay period (10–12 s), when the patterns of activity were reliably more prospective than retrospective, paired $t(7)=2.8$, $p<0.025$. This prospective coding was likely a neural reflection of the cognitive anticipation of the probe stimulus that appeared after the delay. As a result of the prospective bias at this time point, the task relevant (i.e., average of prospective and retrospective estimates) v. irrelevant distinction was not significant (Fig. 4D). Beginning with probe onset (time=12 s), the retrospective and prospective categories were no longer separable, and these task-relevant categories continued to be indistinguishable from the task-irrelevant

category. It is important to emphasize that it is not only the recognition of prospective coding but also retrospective coding that implicates the temporary activation of LTM for working memory retention. Assuming that delay-period activity reflects the active retention of information in working memory, the observation of category-specific activity corresponding either to the target or its associate during the delay indicates that the short-term retention of information was accomplished, at least in part, by the temporary activation of LTM patterns for these stimuli.

For No-PFC analyses, the decoding of subsequent delay-period activity was successful (Fig. 5A,B), with task-relevant activity statistically separated from task-irrelevant activity, paired $t(7)=4.3$, $p<0.005$. Note that the qualitative pattern of delay-period estimates from the No-PFC analyses was very similar to those from the original whole brain analysis (Fig. 4C). For PFC-Only analyses, the decoding of delay period activity failed (Fig. 5C,D). The numerical classifier match values for all categories decreased relative to No-PFC, and task-relevant activity was not distinguished from task-irrelevant activity, paired $t(7)= -0.7$, *n.s*. Indeed, the latter was numerically higher than the former for much of the delay period.

DISCUSSION

The analyses presented here tested directly the hypothesis that patterns of brain activity evoked when subjects perceived and evaluated visual stimuli (drawing upon semantic and episodic LTM) would be reinstated when representations of these stimuli were held in working memory. Consistent with this hypothesis, we decoded delay-period activity with a pattern classifier that was trained on LTM activity. Thus, these data support the idea that the short-term retention of information can be supported by the temporary activation of LTM representations. This model does not necessarily rule out a parallel contribution to delay-task performance by specialized working memory systems. However, in this study we were unable to find evidence for category-specific delay-period activity in the PFC, a region proposed to support such systems. Although activity patterns in PFC during the LTM task were dissociable for people, location, and object stimuli, these category-specific activity patterns were not reinstated during the delay period of the working memory task. This contrasts with more posterior brain regions in which category-specific patterns of LTM activity re-emerged during the delay period (Fig. 5).

Similar to findings from the monkey (e.g., Takeda et al., 2005), we observed that sustained delay-period activity in human brains corresponded to both retrospective and prospective representations. Across subjects, there were qualitatively different classification results for delay-period activity. Fig. 6 contrasts individual results for two subjects. On *Location-Person* trials, the delay-period activity for Subject 1 was classified as most consistent with *Location* activity for the first 4 s of the delay. For the remainder of the delay period, the brain activity was classified as most resembling patterns of *Person* activity. We interpret that this subject maintained a representation of the target (a location), then retrieved from LTM a representation of its associate (a person) and actively maintained this representation for the remainder of the delay period. That is, the subject temporarily activated and maintained a LTM representation in the same brain regions where those representations were initially observed. The crossover from a representation of a location to that of a person suggests a transition from retrospective to prospective coding during the delay period of these trials. This coding strategy was evident across all between-category trials for this subject (Fig. 6B). Subject 6, in contrast, maintained *Person* activity throughout the delay period of *Person-Location* trials (Fig. 6C). This demonstrates a retrospective trace of the target stimulus. This retrospective strategy was evident across all between-category trials for this subject (Fig. 6D). These data suggest inter-subject variability in the cognitive strategy employed to solve the working memory task. We also observed substantial intra-subject variability, in that subjects displayed retrospective coding for some trial types and prospective coding for others. Averaging classifier estimates

across trials for all subjects diminished these opposing trends and produced results that demonstrated no reliable bias across our sample for retrospective vs. prospective coding, but, importantly, captured a reliable dissociation between task-relevant and task-irrelevant patterns of activity.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by NIH grant MH064498 (B.R.P.). The authors would like to thank Sean Polyn for his helpful suggestions and stimuli, and Ken Norman and his group at Princeton for providing insightful feedback on our analyses.

REFERENCES

Anderson, JR. Cambridge, MA: Harvard University Press; 1983. The Architecture of Cognition.

Baddeley, AD. London: Oxford University Press; 1986. Working Memory.

Baddeley AD. New data: Old pitfalls. Behavioral and Brain Sciences 2003;26:729–730.

Baddeley, AD.; Hitch, GJ. Working Memory. In: Bower, GH., editor. The Psychology of Learning and Motivation. New York: Academic Press; 1974. p. 47-89.

Bishop, CM. Oxford: Oxford University Press; 1995. Neural Networks for Pattern Recognition.

Burgess, PW.; Dumontheil, I.; Gilbert, SJ.; Okuda, J.; Schölvinck, ML.; Simons, JS. On the role of rostral prefrontal cortex (area 10) in prospective memory. In: Kliegel, M.; McDaniel, MA.; Einstein, GO., editors. Prospective memory: Cognitive, neuroscience, developmental, and applied perspectives. Mahwah NJ: Erlbaum; in press.

Constantinides C, Franowicz MN, Goldman-Rakic PS. The sensory nature of mnemonic representation in the primate prefrontal cortex. Nature Neuroscience 2001;4:311–316.

Courtney SM. Attention and cognitive control as emergent properties of information representation in working memory. Cognitive, Affective, & Behavioral Neuroscience 2004;4:501–516.

Cowan, N. New York: Oxford University Press; 1995. Attention and Memory: An Integrated Framework.

Cox R. Software for analysis and visualization of functional magnetic resonance neuroimages. Computers and Biomedical Research 1996;29:162–173. [PubMed: 8812068]

Curtis CE, Rao VY, D'Esposito M. Maintenance of spatial and motor codes during oculomotor delayed response tasks. The Journal of Neuroscience 2004;24:3944–3952. [PubMed: 15102910]

D'Esposito M, Ballard D, Zarahn E, Aguirre GK. The role of prefrontal cortex in sensory memory and motor preparation: an event-related fMRI study. NeuroImage 2000;11:400–408. [PubMed: 10806027]

Druzgal TJ, D'Esposito M. Dissecting contributions of prefrontal cortex and fusiform face area to face working memory. Journal of Cognitive Neuroscience 2003;15:771–784. [PubMed: 14511531]

Durstewitz D, Seamans JK. Beyond bistability: Biophysics and temporal dynamics of working memory. Neuroscience 2006;139:119–133. [PubMed: 16326020]

Düzel E. Some mechanisms of working memory may not be evident in the human EEG. Behavioral and Brain Sciences 2003;26:732.

Engle, RW.; Kane, MJ.; Tuholski, SW. Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In: Miyake, A.; Shah, P., editors. Models of Working Memory. Cambridge, U.K: Cambridge University Press; 1999. p. 102-134.

Goldman-Rakic, PS. Circuitry of the prefrontal cortex and the regulation of behavior by representational memory. In: Mountcastle, VB.; Plum, F.; Geiger, SR., editors. Handbook of Neurobiology. Bethesda: American Physiological Society; 1987. p. 373-417.

Haxby JV, Petit L, Ungerleider LG, Courtney SM. Distinguishing the functional roles of multiple regions in distributed neural systems for visual working memory. NeuroImage 2000;11:380–391. [PubMed: 10806025]

Jonides, J. Working memory and thinking. In: Smith, EE.; Osherson, DN., editors. An Invitation to Cognitive Science. Cambridge, MA: MIT Press; 1995. p. 215-265.

Kroger JV. Long-term memory, features, and novelty. Behavioral and Brain Sciences 2003;26:744–745.

Leung H-C, Gore JC, Goldman-Rakic PS. Sustained mnemonic response in the human middle frontal gyrus during on-line storage of spatial memoranda. Journal of Cognitive Neuroscience 2002;14:659–671. [PubMed: 12126506]

Logie RH, Della Salla S. Working memory as a mental workspace: Why activated long-term memory is not enough. Behavioral and Brain Sciences 2003;26:745–746.

Machens CK, Romo R, Brody CD. Flexible Control of Mutual Inhibition: A Neural Model of Two-Interval Discrimination. Science 2005;307:1121–1124. [PubMed: 15718474]

Majerus S, Van der Linden M, Collette F, Salmon E. Does sustained ERP activity in posterior lexico-semantic processing areas during short-term memory tasks only reflect activated long-term memory? Behavioral and Brain Sciences 2003;26:746–747.

Mongillo G, Barak O, Ysodyks M. Synaptic Theory of Working Memory. Science 2008;319:1543–1546. [PubMed: 18339943]

Narayanan N, Prabhakaran V, Bunge SA, Christoff K, Fine EM, Gabrieli JD. The role of prefrontal cortex in the maintenance of verbal working memory information: an event-related fMRI analysis. Neuropsychology 2005;19:223–232. [PubMed: 15769206]

Oberauer K. Access to information in working memory: exploring the focus of attention. Journal of Experimental Psychology: Learning, Memory, and Cognition 2002;28:411–421.

Poldrack RA. Can cognitive processes be inferred from neuroimaging data? Trends in Cognitive Sciences 2006;10:64–69. [PubMed: 16406759]

Polyn SM, Natu VS, Cohen JD, Norman KA. Category-specific cortical activity precedes retrieval during memory search. Science 2005;310:1963–1966. [PubMed: 16373577]

Postle BR. Delay-period activity in prefrontal cortex: one function is sensory gating. Journal of Cognitive Neuroscience 2005;17:1679–1690. [PubMed: 16269105]

Postle BR. Working memory as an emergent property of the mind and brain. Neuroscience 2006;139:23–38. [PubMed: 16324795]

Postle BR, D'Esposito M. Spatial working memory activity of the caudate nucleus is sensitive to frame of reference. Cognitive, Affective, and Behavioral Neuroscience 2003;3:133–144.

Postle BR, Druzgal TJ, D'Esposito M. Seeking the neural substrates of working memory storage. Cortex 2003;39:927–946. [PubMed: 14584560]

Rainer G, Rao SC, Miller EK. Prospective coding for objects in the primate prefrontal cortex. Journal of Neuroscience 1999;19:5493–5505. [PubMed: 10377358]

Ranganath C. Working memory for visual objects: Complementary roles of inferior temporal, medial temporal, and prefrontal cortex. Neuroscience 2006;139:277–289. [PubMed: 16343785]

Ranganath C, DeGutis J, D'Esposito M. Category-specific modulation of inferior temporal activity during working memory encoding and maintenance. Cognitive Brain Research 2004a;20:37–45. [PubMed: 15130587]

Ranganath C, Cohen MX, Dam C, D'Esposito M. Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory recall. The Journal of Neuroscience 2004b;24:3917–3925. [PubMed: 15102907]

Ruchkin DS, Grafman J, Cameron K, Berndt RS. Working memory retention systems: a state of activated long-term memory. Behavioral and Brain Sciences 2003;26:709–777. [PubMed: 15377128]

Takeda M, Naya Y, Fujimichi R, Takeuchi D, Miyashita Y. Active maintenance of associative mnemonic signal in monkey inferior temporal cortex. Neuron 2005;48:839–848. [PubMed: 16337920]

Vallar G. The short-term/long-term memory distinction: Back to the past? Behavioral and Brain Sciences 2003;26:757.

Zarahn E, Rakitin B, Abela D, Flynn J, Stern Y. Positive evidence against human hippocampal involvement in working memory maintenance of familiar stimuli. Cerebral Cortex 2005;15:303–316. [PubMed: 15342440]

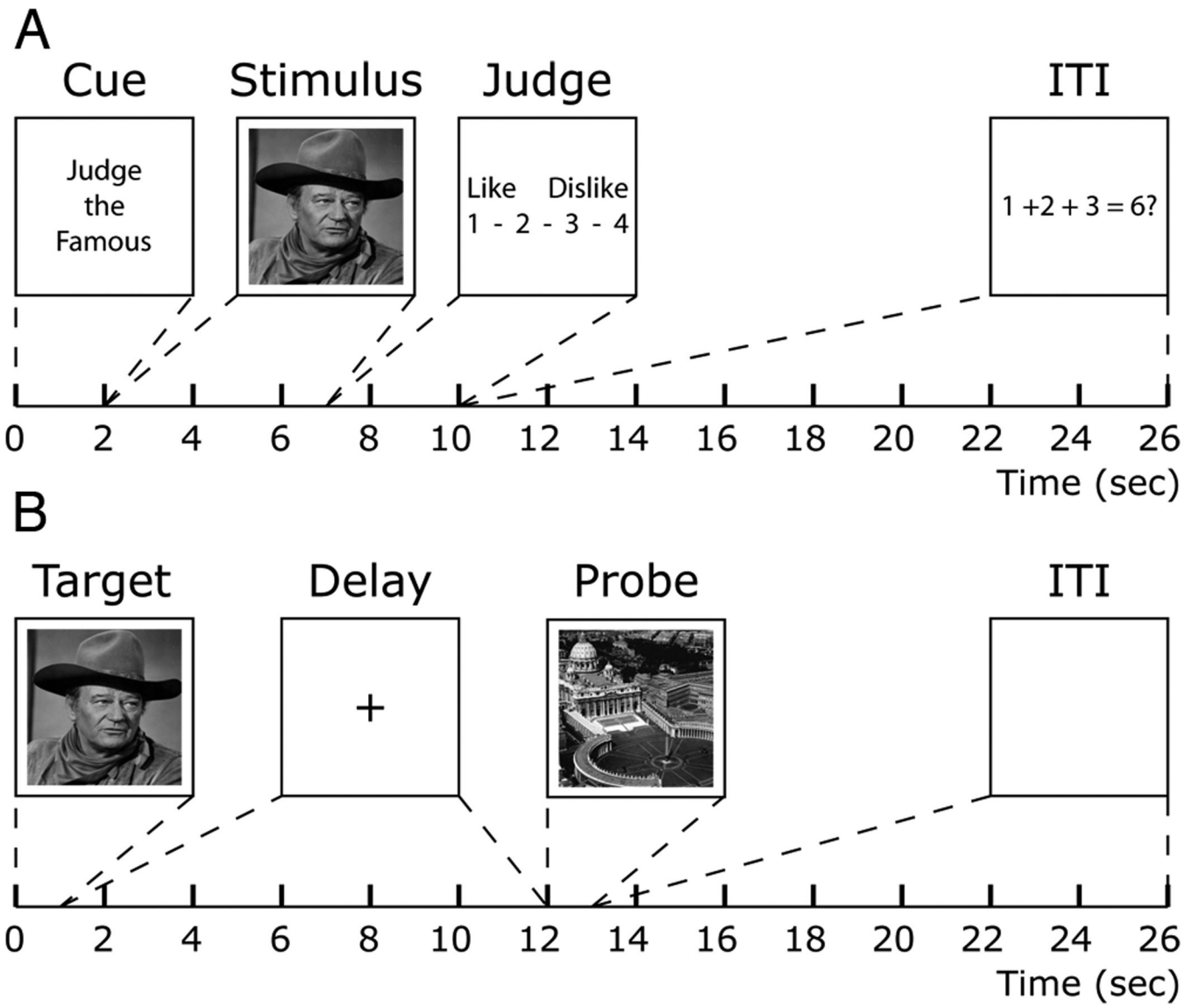

Figure 1. Long-term and working memory trials

The timelines are shown for A) LTM trials and B) delayed paired-associate working memory trials. The dashed lines connecting the boxes to the timelines indicate the onset and duration of each trial phase.

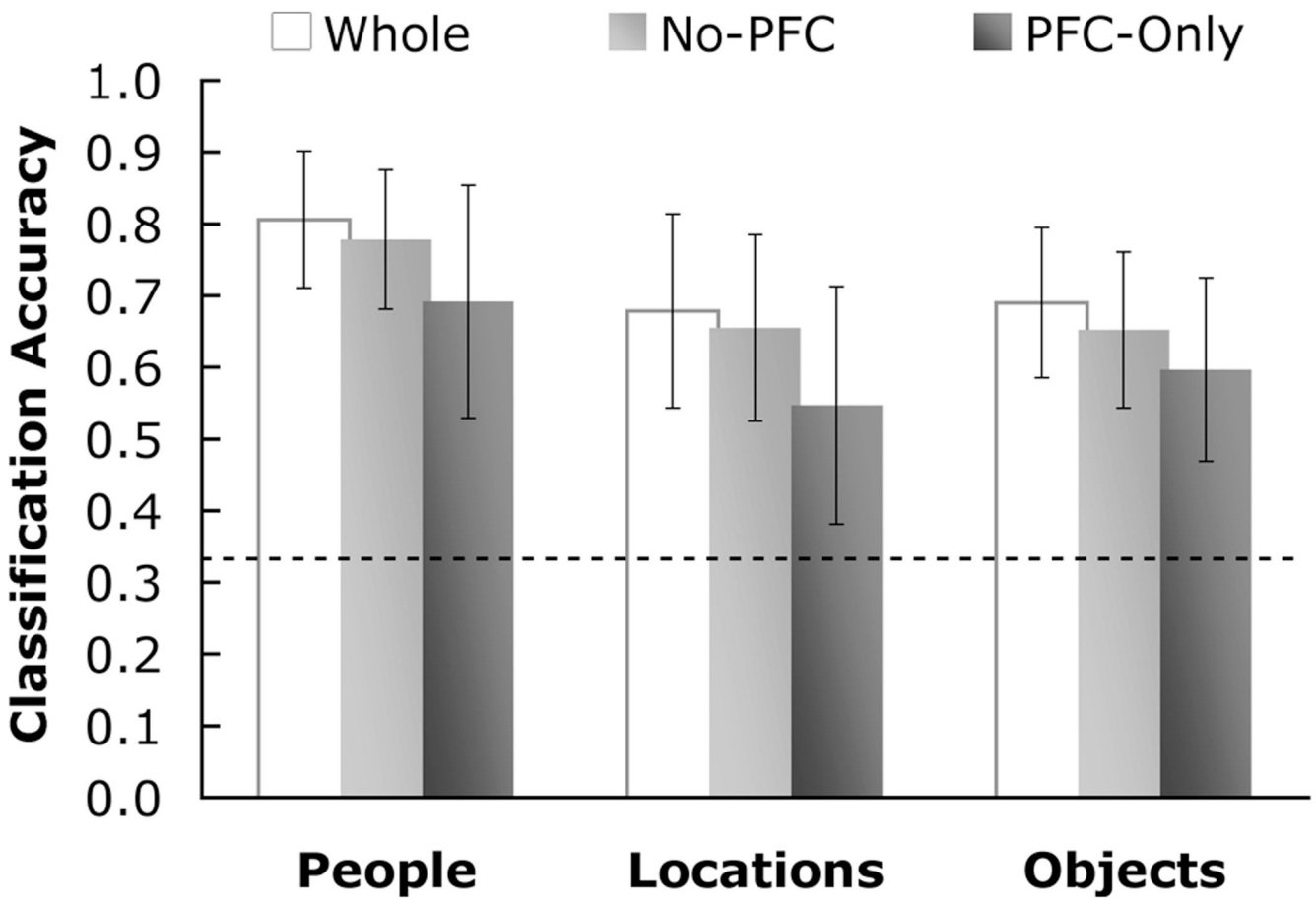

Figure 2. Classifier training accuracy on LTM trials

The prediction accuracy of the classifier for identifying the correct category of brain activity in the LTM trials is shown for three analysis conditions: whole-brain ("Whole"), with PFC voxels masked out ("No-PFC"), and with all voxels outside of PFC masked out ("PFC-Only"). The accuracy scores represent the mean prediction accuracy for the initial 10 s of all LTM trials across all subjects. Accuracy is shown separately for each trial type: People, Locations, and Objects. The dashed line represents the chance level of prediction at 0.333 (the classifier could predict *Person*, *Location*, or *Object* at each time point). Error bars are 95% confidence intervals for the one-tailed *t* tests at $p<0.001$.

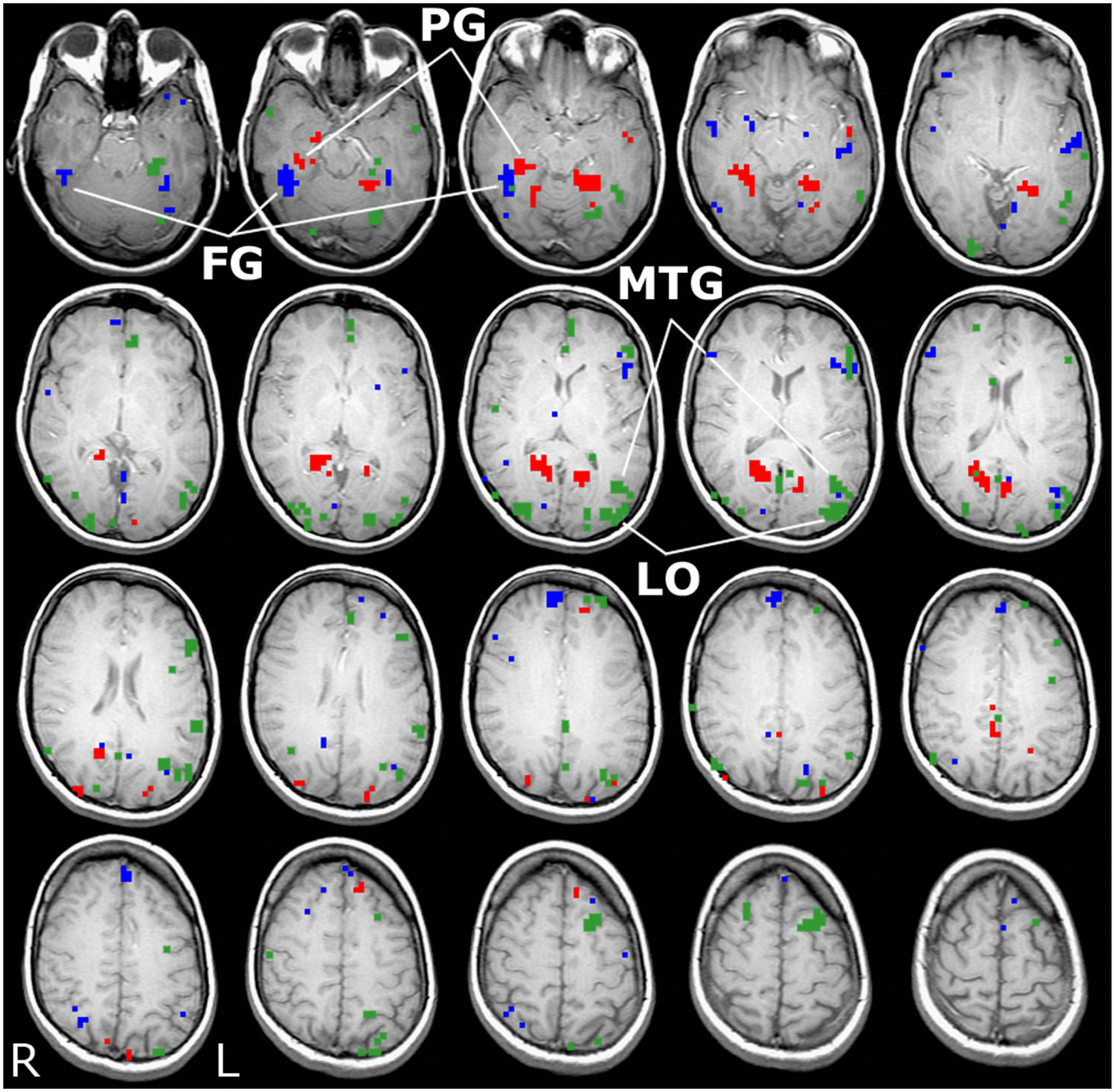

Figure 3. Classifier-derived voxel discrimination maps

Classifier-derived voxel discrimination maps (from the LTM task) from one representative subject (Subject 4). Voxels are colored that were important for detecting patterns of brain activity corresponding to the perception and LTM of people (blue), locations (red), and objects (green). *FG*, fusiform gyrus; *PG*, parahippocampal gyrus; *MTG*, middle temporal gyrus; *LO*, lateral occipital cortex.

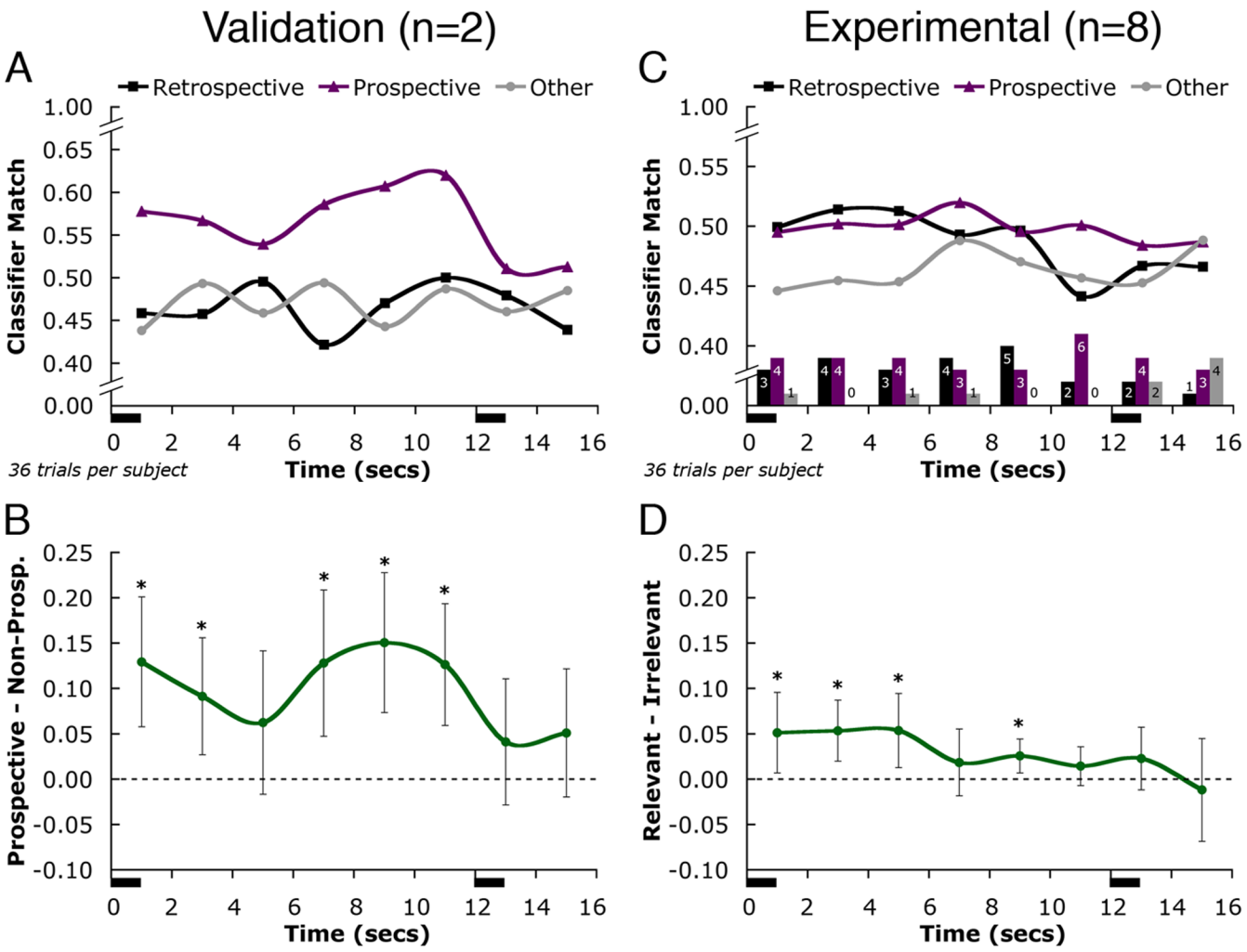

Figure 4. Classification of between-category trials

Mean classifier estimates of delay-period activity are shown in A) for the two validation subjects who were instructed to solve the working memory task using a prospective strategy, and in C) for the eight experimental subjects who received no instruction on which coding strategy to prioritize during the task. Note that A) validates the method, in that delay-period activity classification matches the cognitive strategy that these two subjects were instructed to adopt. The vertical axis shows the classifier estimates of the correspondence between the delay-period brain activity and the learned patterns of category-specific activity (for people, locations, and objects) associated with the stimuli from each trial. Estimates are collapsed across all 36 between-category trials for each subject, into a *Retrospective* (the category of the target), a *Prospective* (the category of its associate), and an *Other* (task-irrelevant) category. The horizontal axis shows the timing of the working memory trial. The black bars along this axis indicate the presentations of target and probe stimuli (target: 0–1 s, probe: 12–13 s). The bar graphs at the bottom of C indicate the number of subjects that demonstrated each type of category selectivity at each time point (black: *Retrospective*, purple: *Prospective*, grey: *Other*). These selectivity statistics were calculated by assigning each subject to a particular category based on the largest average estimate of the three categories for each point. Statistical comparisons for validation subjects in B) focused on the within-subject difference between the Prospective classifier estimate and the average of the non-prospective estimates (Retrospective

and Other). For experimental subjects, statistical comparisons in D) focused on the within-subject difference between the average of the task-relevant estimates (Retrospective and Prospective) and the task-irrelevant estimate (Other). Points marked with stars are significant at $p<.01$ in B and $p<0.05$ in D. Error bars represent a 95% confidence interval around the within-subject difference scores.

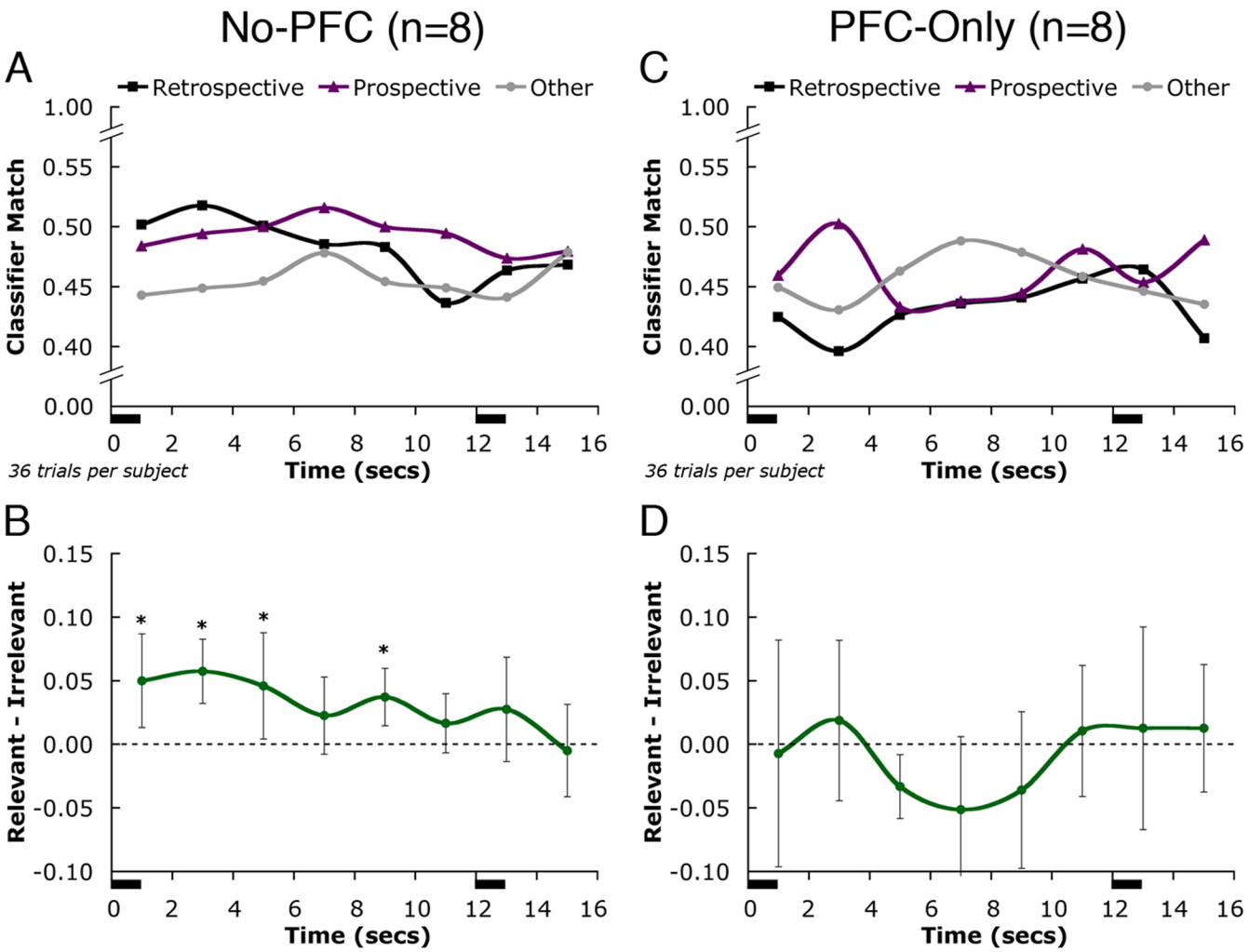

Figure 5. Classification results for PFC

Mean classifier estimates of delay-period activity averaged across all eight experimental subjects are shown in A) when all PFC voxels were removed from analysis, and in C) when only those voxels in PFC were included. Statistical comparisons of task-relevant vs. task-irrelevant estimates are shown for both conditions, in B) and D), respectively. Graph conventions are as described in Fig. 4.

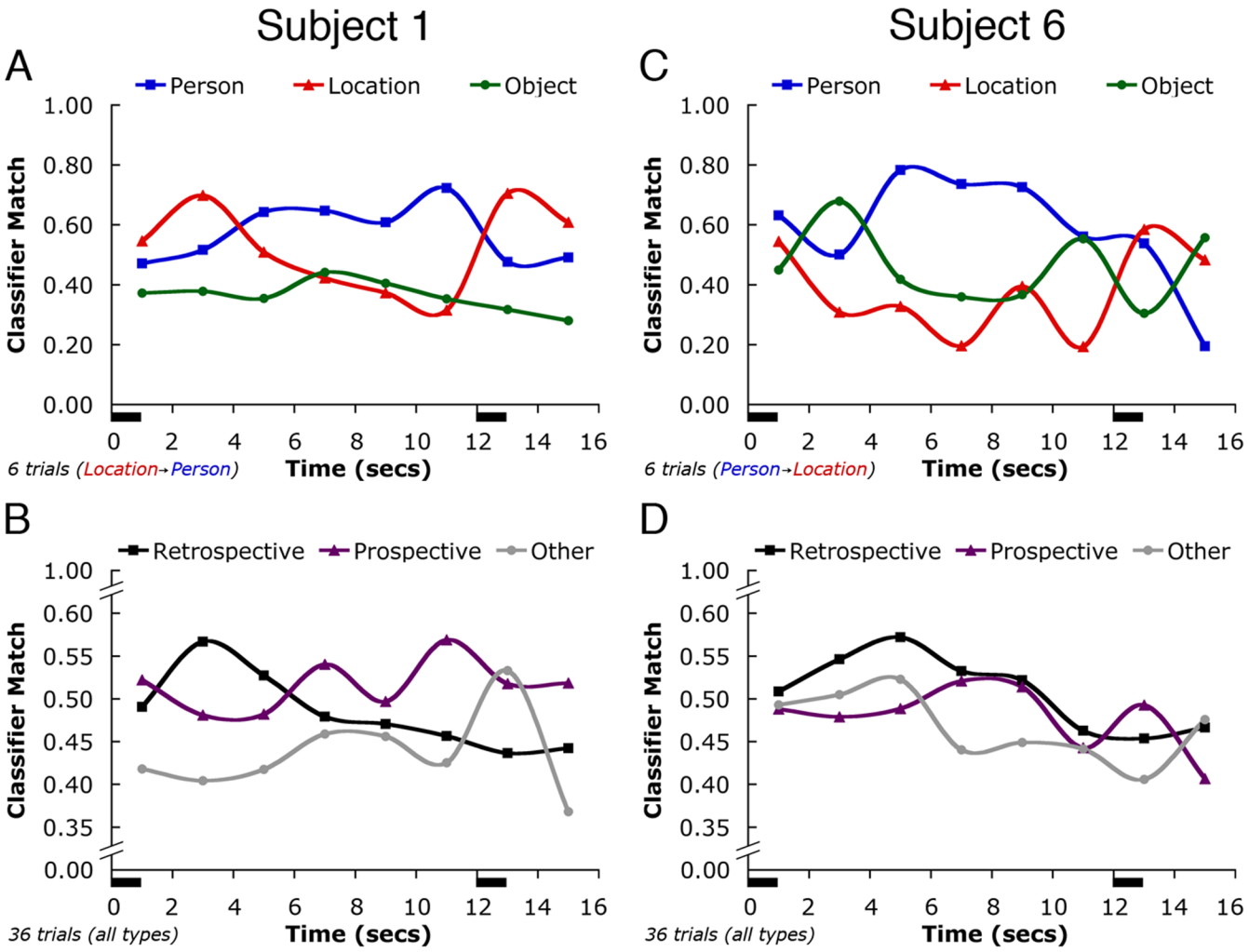

Figure 6. Individual variability in classification results

Classifier estimates of delay-period activity is shown from between-category trials for two representative subjects, illustrating individual differences in task strategy. A) Results for a specific between-category trial type (*Location-Person*) are shown for Subject 1 and B) for the complementary trial type (*Person-Location*) for Subject 6. The classifier estimates for all 36 between-category trials from each subject were recoded into *Retrospective*, *Prospective*, and task-irrelevant *Other* categories in C and D, respectively. Graph conventions are as described in Fig. 4.

Table 1

Distribution of classifier-derived discriminant voxels

Data indicate the number of subjects (out of 10) for whom category-discriminating voxels were found in particular anatomical regions. For example, important voxels for the discrimination of *Location* representations were found in the left parahippocampal gyrus (PG) in 9 subjects, and in the right PG in all 10 subjects. *FG*: mid fusiform gyrus; *PG*: parahippocampal gyrus; *LO*: lateral occipital cortex; *V1*: cuneus, calcarine sulcus, caudal lingual gyrus; *msFP*: medial superior Frontal Pole; *MTG*: middle temporal gyrus; *MFG*: middle frontal gyrus; *SFS*: superior frontal sulcus; *PreCG*: precentral gyrus; *PC*: posterior cingulate/precuneus.

	FG	**PG**	**LO**	**V1**	**msFP**	**MTG**	**MFG**	**SFS**	**PrCG**	**PC**
Category	*L/R*	*L/R*	*L/R*	*L/R*	*Midline*	*L/R*	*L/R*	*L/R*	*L/R*	*L/R*
People	8/10	1/1	7/3	7/9	8	6/8	6/8	5/2	1/2	8/10
Locations	5/5	9/10	6/8	9/9	3	3/2	4/6	1/3	1/0	5/8
Objects	10/10	9/2	10/8	9/8	10	9/5	7/7	10/7	6/4	7/6

J Neurosci. Author manuscript; available in PMC 2009 June 19.