

## Distinct spatial patterns of brain activity associated with memory storage and search

Eric Zarahn,<sup>a,c,\*</sup> Brian C. Rakitin,<sup>a,b</sup> Diane Abela,<sup>a</sup> Joseph Flynn,<sup>a</sup> and Yaakov Stern<sup>a,b,c</sup>

<sup>a</sup>The Cognitive Neuroscience Division of the Taub Institute, Columbia University College of Physicians and Surgeons, 630 West 168th Street, Columbia University, New York, NY 10032, USA

<sup>b</sup>Department of Neurology, Columbia University College of Physicians and Surgeons, 630 West 168th Street, Columbia University, New York, NY 10032, USA

<sup>c</sup>Department of Psychiatry, Columbia University College of Physicians and Surgeons, 630 West 168th Street, Columbia University, New York, NY 10032, USA

Received 15 March 2006; revised 18 July 2006; accepted 21 July 2006

Available online 8 September 2006

**The time it takes for a human participant to decide whether a given stimulus is an element of a remembered set increases approximately linearly with the number of elements in the set. Here we tested for and detected a spatial pattern of brain activity whose magnitude of expression during this memory search process correlates with set size. We then tested the idea that memory search simply involves a re-activation of neurons involved in remembering the set by statistically comparing the patterns of brain activity corresponding to memory search and set size dependent working memory maintenance. These patterns were significantly different, suggesting that memory search and working memory maintenance are mediated by distinct neural mechanisms.**

© 2006 Elsevier Inc. All rights reserved.

*Keywords:* fMRI; Memory search; Set size; Parietal; Prefrontal; Premotor; Working memory; Multivariate linear model; Canonical variates analysis

### Introduction

In a Delayed Item Recognition (DIR) task, the participant is presented with a set of stimuli, maintains information about this set throughout the duration of a retention delay, and then compares this remembered information to a probe stimulus in order to make a recognition judgment (i.e., “does the probe match any of the remembered stimuli?”). It has been consistently observed that the amount of time required to decide if the probe belongs to the memorized set, depends nearly linearly on the set size (Sternberg, 1966; Burle and Bonnet, 2000; Rypma et al., 2002; Veltman et al., 2003). This has been taken as evidence in support of an exhaustive, serial (Sternberg, 1966) or parallel, limited capacity memory search (Townsend, 1990; Burle and Bonnet, 2000).

Aside from whether memory search is serial or parallel, the relationship between it and working memory (WM) maintenance is not well understood. Electrophysiological data from non-human primates suggest that recognition of a probe involves to some extent the same neurons activated during WM maintenance (Wilson et al., 1993; Desimone, 1996; Miller et al., 1996; Rainer et al., 1998). But it is not clear if the neural substrates of memory search are identical to those of either recognition or WM maintenance (Pollmann et al., 2000). Some physiological models posit that activity in the same neuronal ensembles underlies WM maintenance as well as memory search (Jensen and Lisman, 1998; Burle and Bonnet, 2000). Starting with the assumption, shared by these models, that the neural mechanisms underlying both memory search and WM maintenance should depend linearly on set size, we tested the hypothesis that the brain systems involved in memory search are identical to those involved in WM maintenance via canonical variates analysis (CVA) with sequential latent root testing.

CVA is a statistical method based on singular value decomposition (SVD) of a  $k \times m$  matrix of covariances between  $k$  predictors and  $m$  data variables (Worsley et al., 1997). In neuroimaging data,  $m$  is the number of spatial observations (e.g., the number of voxels). CVA will therefore represent each of the spatial patterns corresponding to the relationships between voxel-wise neuroimaging data and predictors (which we will refer to generally as activation patterns) as a weighted sum of orthonormal latent spatial patterns. A special case relevant to the current work is if all  $k$  activation patterns are identical to within a scaling factor, then CVA would demonstrate this by showing that there is only one latent pattern required to generate each of the original patterns. This is of course ignoring noise, which will always distribute itself to some extent across  $k$  estimated latent spatial patterns, even if only some smaller number contains signal. Thus a statistical procedure, sequential latent root testing, is required to decide if an estimated spatial pattern contains some signal or only noise. A linear algebraic interpretation of sequential latent root testing is that it assesses the rank of the set of true activation patterns.

\* Corresponding author. Department of Psychiatry, Columbia University College of Physicians and Surgeons, 630 West 168th Street, Columbia University, New York, NY 10032, USA. Fax: +1 212 342 1838.

E-mail address: ez84@columbia.edu (E. Zarahn).

Available online on ScienceDirect (www.sciencedirect.com).

CVA has different variants, depending on the assumptions made about the distribution of the data and the transformations applied to the covariance matrix (Worsley et al., 1997). The particular variant of CVA we use in this paper involves transformation of predictors such that the covariances of interest are whitened, unit variance contrasts (a contrast being a linear combination of regression coefficients), and was developed by Worsley and colleagues for neuroimaging data with errors correlated across space and repeated measures, such as fMRI time series (Worsley et al., 1997) or summarized measures of condition-wise fMRI responses within participant (Zarahn et al., 2005). Worsley and colleagues referred to this version of CVA with sequential latent root testing as a multivariate linear model (MLM).

One might ask why CVA rather than statistical parametric mapping (SPM) is being used to address our main hypothesis. We are interested in determining whether two activation patterns (linear fMRI set size effects associated with (1) WM maintenance and (2) memory search) are identical to within a scaling factor, and this is precisely what CVA does. In contrast, the analogous SPM analysis would involve voxel-wise subtractions of effect intensities between patterns (1) and (2), and hence could not disambiguate between their being distinct versus their being scaled versions of one another (as both cases would lead to a positive SPM result). Therefore, CVA is a more valid test than SPM to assess the null hypothesis of patterns being identical to within a scaling factor.

By the same pattern identity logic, CVA also allowed us to explicitly test for within-laboratory replication across independent experiments. In this framework, replication is satisfied if the same effect assessed in independent samples yields spatial activation patterns that are identical to within a scaling factor. This is a more inclusive definition of replication than requiring the activation patterns to be exactly the same (i.e., identical with a scaling factor of unity). Replication in neuroimaging is rarely assessed, and most methods that have been employed to assess reliability of results across experiments lack a clear null-hypothesis. CVA provides an explicit test for replication of a brain activation pattern across samples with the null hypothesis being pattern identity, i.e., activation maps being scaled versions of one another.

**Materials and methods**

*Participants*

Young, right-handed participants were recruited through flyers posted at the Columbia University campus and advertisements placed in local newspapers [experiment A:  $n=40$ , 30 M and 10 F, age= $25.1 \pm 3.9$ , years of education= $15.7 \pm 1.4$ ; experiment B:  $n=20$ , 14 M and 6 F, age= $24.6 \pm 2.8$ , years of education= $15.7 \pm 1.2$ ]. All participants supplied informed consent as approved by the Columbia University Institutional Review Board. Volunteers were screened for psychiatric and neurologic illness via a questionnaire. Previous reports examined load dependent patterns of retention delay fMRI signal from experiment A (Habeck et al., 2004; Habeck et al., 2005; Zarahn et al., 2005).

*Behavioral task*

Experiments A and B had nearly identical designs, and were treated as replications. The behavioral task used in both experiments A and B was a DIR task for letters (Sternberg, 1966). Each trial lasted 13 s. Participants were instructed to respond as accurately as possible. No feedback about their performance was given during the scanning session. The sequence of events within a DIR trial is provided in Fig. 1. The stimuli to be remembered were presented for 3 s. The geometry of the stimuli in the memory set was a 2 by 3 array, regardless of set size, with asterisks acting as non-letter placeholders for set sizes 1 and 3. With the offset of the letter array, participants were instructed to hold the memory set in mind for a 7-s maintenance interval (retention delay). Finally, a probe letter (lowercase, centered in the field of view) appeared for 3 s. In response to the probe, participants indicated by a button press whether or not the probe matched a letter in the memory set (right index finger button press to indicate “yes”, left index finger button press to indicate “no”).

Each experimental block contained 10 trials at each of the 3 set sizes, with 5 true negative and 5 true positive probes per set size. BOLD fMRI data were acquired for three experimental blocks per participant, yielding a total of 30 experimental trials per set size per

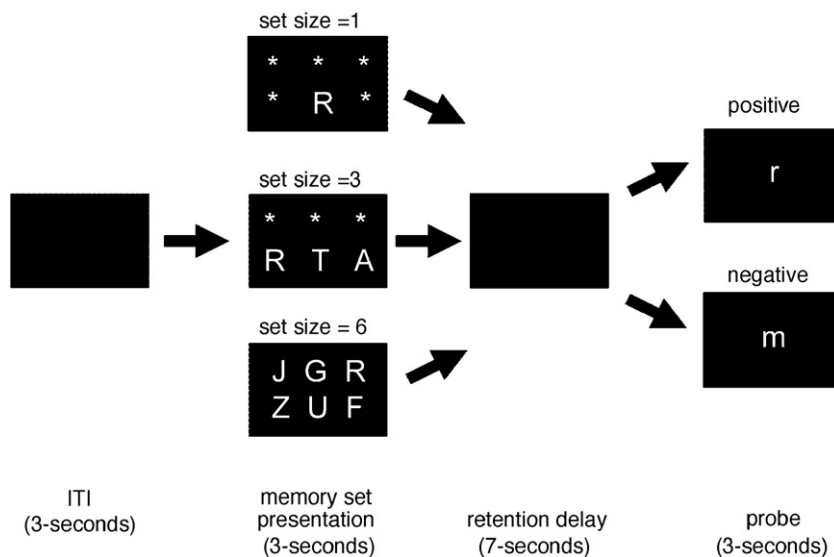


Fig. 1. The DIR task is schematized. ITI: inter-trial interval.

participant. Experiments A and B differed only in the manner in which the stimulus sets were constructed: In experiment A each successive crossing of set size and probe type was associated with the same stimulus set in each of the three blocks (even though the sequence in which the factor crossings were traversed was randomized in each block); in experiment B each successive crossing of set size and probe type was associated with a different stimulus set in each of the three blocks. These stimulus sets were fixed across participants in both experiments A and B.

DIR trials were separated by a minimum 3-s inter-trial interval. In addition, blank intervals (presentation of a blank screen for 2 s, requiring no behavioral output) were pseudo-randomly interspersed between DIR trials to both provide a baseline condition for positive control purposes and reduce the likelihood of neurophysiological responses predictive of the beginning of trials. The pseudo-randomization of these blank intervals was via a random-without-replacement scheme (thus, more than one blank trial could occur sequentially, leading to an effectively jittered inter-trial interval), with a total of 70 blank intervals per block. The presentation of DIR trials of different set sizes was also pseudo-randomly sequenced via a random-without-replacement scheme. The duration of each block was 620 s. There were approximate 1-min breaks between blocks.

Participants were trained on 7 blocks of DIR trials prior (range of 1 day to 2 weeks) to the acquisition of fMRI data, the first 6 of which were administered with feedback. The training session was conducted to reduce task-related skill learning during the course of fMRI search (Kirschen et al., 2005).

During fMRI scanning, task stimuli were back-projected onto a screen located at the foot of the MRI bed using an LCD projector. Participants viewed the screen via a mirror system located in the head coil. Responses were made on a LUMItouch response system (Photon Control Company). Task onset was electronically synchronized with the MRI console. Task administration and data collection (reaction time and accuracy) were controlled using PsyScope (Cohen et al., 1993).

#### *Statistical analysis of behavioral data*

Behavioral data were analyzed from all trials on which the participant responded. Analyses of memory search rates involved *t*-tests based on least-squares slopes of reaction time with respect to set size computed within each participant. Least-squares *y*-intercepts of reaction time with respect to set size were also computed within each participant for ancillary analyses. Statistical analyses of effects of set size on the accuracy measure *d'* (Green, 1988) were performed via the repeated measures general linear model (GLM) module in SPSS for Windows, Release 11.0.1.  $\alpha=0.05$  per comparison was used.

#### *fMRI data acquisition*

During the performance of each block of the DIR task, 207 BOLD images (Kwong et al., 1992; Ogawa et al., 1993), were acquired with an Inera 1.5 T Phillips MR scanner equipped with a standard quadrature head coil, using a gradient echo echo-planar (GE-EPI) pulse sequence [TE/TR=50 ms/3000 ms; flip angle=90 degrees;  $64 \times 64$  matrix, in-plane voxel size=3.124 mm  $\times$  3.124 mm; slice thickness=8 mm (no gap); 17 trans-axial slices per volume]. Four additional GE-EPI excitations were performed before the task began, at the beginning of each run, to allow transverse magnetiza-

tion immediately after radio-frequency excitation to approach its steady-state value; the images corresponding to these excitations were discarded. A T2-weighted, fast spin echo image was also acquired from each participant for spatial normalization purposes [TE/TR=100 ms/2000 ms; flip angle=90 degrees,  $256 \times 256$  matrix; in-plane voxel size=0.781 mm  $\times$  0.781 mm; slice thickness=8 mm (no gap); 17 trans-axial slices per volume].

#### *fMRI data pre-processing*

All image pre-processing was implemented using the SPM99 program (Wellcome Department of Cognitive Neurology). The following steps were taken in turn for each participant's GE-EPI dataset: Data were temporally interpolated and shifted to correct for the order of slice acquisition using the first slice acquired in the TR as the reference. All GE-EPI images were realigned to the first volume of the first session. The T2-weighted (structural) image was co-registered to the first GE-EPI volume using the mutual information co-registration algorithm implemented in SPM99. This co-registered high-resolution image was then used to determine parameters ( $7 \times 8 \times 7$  non-linear basis functions) for transformation into a Talairach standard space (Talairach and Tournoux, 1988) defined by the Montreal Neurologic Institute (MNI) template brain supplied with SPM99. This transformation was then applied to the GE-EPI data, which were re-sliced using sinc-interpolation to  $2 \text{ mm} \times 2 \text{ mm} \times 2 \text{ mm}$ .

#### *fMRI statistical analysis*

All statistical analysis was implemented using the SPM99 program and other code written in MATLAB 5.3 (Mathworks, Natick, MA). The fMRI data analysis comprised two levels of voxel-wise GLMs (Friston et al., 2005), and CVA (Worsley et al., 1997). In the first-level GLM, the individual participants' fMRI time series were modeled with predictors representing the expected BOLD fMRI response (implicitly, relative to the blank intervals) to the three DIR trial components of memory set presentation, retention delay, and probe periods, separately for each block and crossing of the set size and probe type factors. Trials on which there was no motor response from the participant during the probe period were modeled separately and were not included for analysis in the second-level GLM. Each time series predictor was constructed by convolution of a sequence discrete-time delta functions representing the onsets of a given DIR trial component, a rectangular function of duration dictated by the duration of the assumed neural response for that trial component (Zarahn, 2000), and an assumed BOLD impulse response function (as represented by default in SPM99). Based in part on prior knowledge as well as regression diagnostics, two rectangular functions were used for the trial component of memory set presentation: one modeling a relatively brief (400 ms) neural response at the beginning of that trial component, and another modeling a neural response lasting throughout that entire component (3000 ms); the same two rectangular functions were used for probe presentation. A single rectangular function of 7000 ms duration was used for the retention delay.

Single-participant contrast images for both the memory set presentation and probe presentation trial components were computed from a linear combination of the two parameter estimates that estimated the area under the curve of the neural response for that trial component. For the retention delay trial component, the analogous contrast was simply the coefficient of its

single basis function. After collapsing over block, this yielded 18 contrast estimate images per participant: (3 trial components)  $\times$  (3 set sizes)  $\times$  (2 probe types). These contrast estimate images were intensity normalized via voxel-wise division by the time series mean, masked with an image that represented the intersection of useable data from all participants and had a gray matter prior probability  $>.25$  in standard space using the tissue-type prior probability images supplied with the SPM99 program (this eliminated the bulk of the ventricles and surrounding white matter from the search volume), and spatially smoothed with an isotropic Gaussian kernel, full-width-at-half-maximum=8 mm. The purpose of masking prior to smoothing is to avoid differences between participants in partial volume contributions to data within the mask after smoothing. The resulting images were used as the dependent data in a second-level, voxel-wise GLM (Friston et al., 2005). This second-level GLM thus had a single between-participants factor (experiment) and 3 repeated measure factors (trial component, set size, and probe type) leading to 18 repeated measures per participant per voxel, and a total of 36 columns in the full rank design matrix. The repeated measures covariance matrix of the second-level data was estimated at each voxel, and spatially averaged. This matrix was used to approximate the known observation error covariance matrix ( $\Sigma$  in (Worsley et al., 1997)).

The MLM version of CVA (hence, simply CVA) involves SVD of a  $k \times m$  ( $k$  being referred to as the number of effects of interest) matrix  $Z$  of  $k$  whitened contrasts and  $m$  voxels

$$\mathbf{Z} = \text{Var} < \mathbf{C}^T \hat{\mathbf{B}} >^{-1/2} \mathbf{C}^T \hat{\mathbf{B}} \mathbf{D}_\sigma^{-1},$$

where  $\mathbf{D}_\sigma$  is an  $m \times m$  matrix of estimated voxel-specific error standard deviations,  $\mathbf{C}$  is an  $l \times k$  contrast weight matrix defining the  $k$ -dimensional effects of interest,  $\hat{\mathbf{B}}$  is an  $l \times m$  matrix of ordinary least-squares estimators (indicated by the hat) of the coefficients of (in this paper) the second-level  $n \times l$  design matrix.  $\text{Var} < \mathbf{C}^T \hat{\mathbf{B}} >$  is the variance of a column of  $\mathbf{C}^T \hat{\mathbf{B}}$  (ignoring voxel-specific error variance magnitude), and it is the presence of the  $\text{Var} < \mathbf{C}^T \hat{\mathbf{B}} >^{-1/2}$  and  $\mathbf{D}_\sigma^{-1}$  terms in  $Z$  that whiten the contrasts  $\mathbf{C}^T \hat{\mathbf{B}}$  (Worsley et al., 1997). For any given dataset and design matrix,  $\hat{\mathbf{B}}$  is the same regardless of how the effects of interest are defined. Rather, it is  $\mathbf{C}$  (in relation to the design matrix) that defines the effects of interest. When  $k=2$ , for example,  $\mathbf{C}$  has 2 columns, each of which contains weights that define one of the effects of interest (see below for the method of computing the contrast weights corresponding to linear effects of set size).

SVD of  $\mathbf{Z}$  yields latent spatial patterns and latent predictor variables (Worsley et al., 1997). Sequential latent root testing is used to determine how many of the estimated latent spatial patterns of  $\mathbf{Z}$  contain signal (Worsley et al., 1997). In the  $q$ th step of sequential latent root testing ( $q$  runs from 1 to  $k$ ),  $F$ -statistics are used to compare the mean of the last  $k - q + 1$  eigenvalues to its distribution under the null hypothesis (Worsley et al., 1997). If the  $q$ th step is significant,  $q$  is incremented by 1 and the next step is performed; if the  $q$ th step is not significant, then the test stops and  $q - 1$  roots are inferred; if all  $k$  steps are significant, then  $k$  roots are inferred (Worsley et al., 1997).

The degrees of freedom of the sequential latent root testing  $F$ -statistics depend on  $q$  (so that the first and second latent root tests will have different degrees of freedom),  $k$ , the degrees of freedom at each voxel, and the number of spatial degrees of freedom, which depends on the search volume and the spatial smoothness of the errors (Worsley et al., 1997). Though the

numerator and denominator degrees of freedom might seem large compared to what is typically seen in the behavioral sciences, they are approximately correct under the same assumptions as SPM. See Worsley et al. (1997) for formulae and full details.

Here, selected effects of interest from the second level GLM were subjected to sequential latent root testing to test hypotheses at the spatially omnibus level concerning identity (i.e., equality to within a multiplicative scaling factor) of pairs of spatial patterns of brain activation. In our study there were two different primary purposes underlying our choices of effects of interest. One was to test for replication of the slope of fMRI response with respect to set size (within task period) across experiments by assessing the identity of the spatial patterns corresponding to this effect in each experiment. In these analyses  $k=2$ , and each dimension corresponded to the linear set size effect within task period from one of the experiments (the non-zero contrast weights yielding the linear effect of set size within a task period and session were obtained from the first row of the pseudo-inverse of the matrix  $[[1 \ 3 \ 6]^T [1 \ 1 \ 1]^T]$ ). The other goal was to assess the identity of the spatial patterns of slope of fMRI response with respect to set size during the retention delay and probe periods. In these analyses as well,  $k=2$ , and each dimension corresponded to one of the effects averaged over the two experiments. Additionally, there were also tests in which  $k=1$ , and these correspond simply to spatially omnibus tests of a single effect (Worsley et al., 1997).

Statistical inference in sequential latent root testing is at the spatially omnibus level, i.e., specified effects at all loci in the brain are tested simultaneously. Thus, it affords no inference about localization. However, when  $k=2$ , it does provide the correct type of test required to assess a null hypothesis of patterns being identical to within a scaling factor (or, when  $k=1$ , a null hypothesis of zero effect everywhere in the brain). Furthermore, sequential latent root testing inference is at the population-level if the design itself uses an appropriate population-level error variance estimate (which is true for our second-level model).

We specified the false positive rate of all sequential latent root tests to be  $\alpha=0.05$ . When  $k=1$ , latent spatial patterns scaled by their singular values are distributed as SPM $\{t\}$ . An aside is that these SPM $\{t\}$ , which relied on an estimated  $\Sigma$ , are different from what one would obtain if simply using regression computations as implemented within SPM99 (which assumes  $\Sigma$  is an identity matrix). For display these SPM $\{t\}$  were thresholded at a  $t$ -value corresponding to  $\alpha=0.001$  uncorrected for multiple comparisons and a cluster size of 50 voxels. After a coordinate correction using a MATLAB routine made publicly available by Matthew Brett (<ftp://ftp.mrc-cbu.cam.ac.uk/pub/imaging/MNI2tal/mni2tal.m>), likely anatomic labels of cluster maxima were assigned with the Talairach Daemon (Lancaster et al., 2000).

The observed expression of a latent spatial pattern in a given participant and condition is the inner product of that pattern with the appropriately normalized data image from that participant and condition (after adjustment of that image for effects of no interest). The units of observed expression are voxel-wise error standard deviations (Worsley et al., 1997). The values of observed expression in these units can be misleadingly small as in this design voxel-wise error standard deviation is much larger than the standard deviation of the effects of interest. This is because the latter depends not only on the former but also on  $\Sigma$  and sample size. Sequential latent root testing uses valid variance estimators that depend on, but are not simply equal to voxel-wise error standard deviations. As the effects of interest in the current study

represented linear combinations of effects over conditions (i.e., the slope of fMRI response with respect to set size), the observed expressions were summarized to a single number per participant (the units of which are still voxel-wise error standard deviations) through computation within each participant of the inner product of observed expression as a function of condition with the appropriate condition weights.

## Results

We first present behavioral results indicating that, in the healthy population studied, reaction time manifests an expected linear component in its relationship with memory set size. This result makes it possible to consider linear fMRI set size effects associated with the probe period as correlates of memory search. We then demonstrate that accuracy does not vary as a function of set size, making it possible to hypothesize that memory load increased linearly as a function of set size and thus supporting interpretation of linear fMRI set size effects associated with the retention delay as correlates of verbal WM maintenance. We then proceed to compare the neural correlates of memory search and verbal WM maintenance.

### Reaction time

As we were only interested in the existence of linear set size effects on reaction time, reaction time data within each participant were summarized as a least-squares slope. As expected (Sternberg, 1966), there was a linear component to the relationship of reaction time and set size in both experiments [mean slope  $\pm$  s.d. in experiment A:  $59.1 \pm 30.8$  ms/item,  $t(39)=12.1$ , two-tailed  $p < 10^{-6}$ ; experiment B:  $63.6 \pm 27.7$  ms/item,  $t(19)=10.3$ , two-tailed  $p < 10^{-6}$ ]. The magnitude of the slopes did not differ significantly across the two experiments [ $t(58)=0.55$ , two-tailed  $p=0.58$ ].

### Memory performance

Averaged over set sizes, memory performance was significantly greater than chance in both experiments [ $d'$  in experiment A:  $3.35 \pm 0.35$ ,  $t(39)=59.7$ , two-tailed  $p < 10^{-6}$ ; experiment B:  $3.15 \pm 0.47$ ,  $t(19)=29.6$ , two-tailed  $p < 10^{-6}$ ]. Unlike for reaction time, we were interested not only in linear but also quadratic set size effects (this exhausts the 3-level set size factor). The logic was that if accuracy did not change at all with set size, then memory load (i.e., the amount of information being maintained in WM) could be reasonably hypothesized to increase linearly with set size. There was no significant main effect of set size [ $F(2,116)=0.57$ ,  $p=0.57$ ]. Nor was there a significant main effect of experiment [ $F(1,58)=3.41$ ,  $p=0.07$ ] or set size  $\times$  experiment interaction [ $F(2,116)=1.04$ ,  $p=0.36$ ].

### Brain correlates of memory search and verbal WM maintenance

During the probe period, participants are presented with the probe stimulus and decide whether it is part of the memorized set. Averaging the data of experiments A and B, the linear relationship between brain activity during the probe period and set size was significant at the spatially omnibus level [ $F(318,48177)=1.36$ ,  $p=0.00025$ ]. The thresholded SPM{t} is displayed in Fig. 2 (in red). The spatial weights comprising the pattern reflect the magnitude of the linear relationship of set size and probe period fMRI activity at those locations. Only the largest positive weights are displayed in the figure, even though all of the weights formally contribute to the spatially omnibus test and the pattern expression. The largest estimated weights are present in left inferior frontal gyrus, bilateral anterior insula, left middle frontal gyrus, and medial frontal gyrus (Table 1). The medial frontal gyrus (Pollmann et al., 2000; Cairo et al., 2004), right middle frontal gyrus and right posterior lobe of the cerebellum (our dataset did not include this area of the cerebellum) (Cairo et al., 2004), and anterior insula

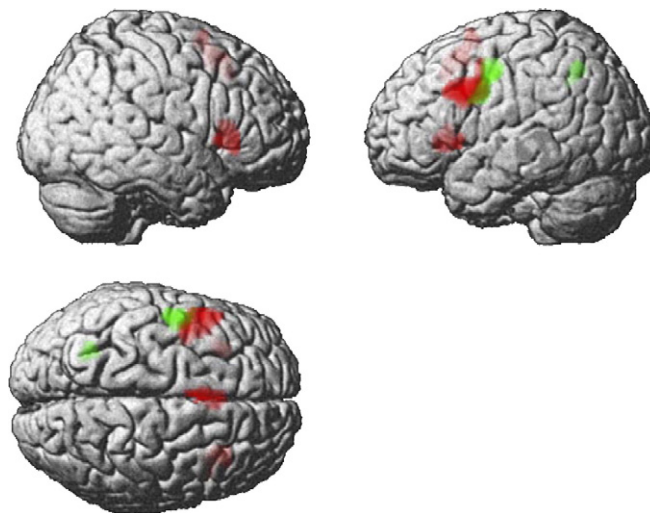


Fig. 2. Memory search and WM maintenance brain activity patterns. Shown are the thresholded spatial patterns corresponding to the slope (only voxels with positive slopes are shown) of fMRI activity with respect to set size associated with the retention delay (green) and the probe period (red). The retention delay pattern presumably reflects WM maintenance; the probe period pattern presumably reflects memory search. Both patterns were obtained from the average of effects across experiments A and B (and so  $k=1$  for both). For display purposes, these spatial patterns (which are distributed as SPM{t}) have been thresholded at  $p < 0.001$  uncorrected for multiple comparisons and a cluster size of 50 voxels ( $0.4 \text{ cm}^3$ ). The intensity of color on the brain surface is the integral, along a path normal to the brain surface, of  $t$ -values which have been exponentially decayed (space constant = 14 mm) based on their depth from the brain surface.

Table 1  
Coordinates and anatomic labels of local maxima

Talairach coordinates (mm)			Anatomic label	Brodmann area	% fMRI signal change/letter, averaged over appropriate trial component		Pearson correlation with RT slope	
x	y	z			Experiment A	Experiment B	Experiment A	Experiment B
<i>Positive linear component of relationship between set size and fMRI signal associated with probe period</i>								
32	22	8	Insula	13	0.045	0.139	-0.080	0.141
-42	5	31	IFG	9	0.075	0.149	0.018	0.063
-53	13	32	MFG	9	0.083	0.090	0.126	-0.004
-50	21	30	MFG	9	0.086	0.129	0.100	0.158
-30	23	1	Insula	13	0.094	0.085	0.130	-0.013
-2	14	45	Medial FG	6	0.084	0.105	0.142	0.261
6	16	45	Medial FG	6	0.024	0.097	-0.023	-0.041
8	10	51	SFG	6	0.036	0.099	0.035	-0.055
0	9	62	SFG	6	0.070	0.183	0.016	0.234
<i>Positive linear component of relationship between set size and fMRI signal associated with retention delay</i>								
-51	-4	43	PrCG	4	0.016	0.014	0.061	0.113
-50	3	27	PrCG	6	0.020	0.014	0.091	-0.050
-30	-54	45	SPL	7	0.020	0.019	0.219	-0.183

Shown are Talairach coordinates and anatomical labels of local maxima of spatial patterns representing linear set size effects during retention delay or probe periods (shown in Fig. 2). The spatial patterns were thresholded at  $\alpha=0.001$  uncorrected for multiple comparisons, cluster size=50 voxels. Both the probe and retention delay patterns were obtained from the average of effects across experiments A and B. None of the correlations with RT slope were significant at  $\alpha=0.05$ , uncorrected for multiple comparisons. IFG: inferior frontal gyrus, MFG: middle frontal gyrus, PrCG: precentral gyrus, SFG: superior frontal gyrus, SPL: superior parietal lobule.

(Pollmann et al., 2000) have been reported to be involved in memory search. However, neither Pollmann and colleagues nor Cairo and colleagues formally modeled both retention delay and probe period set size effects.

To assess within-laboratory replication of this effect, CVA was used to test whether the (true) spatial patterns of the linear fMRI set size effect associated with the probe period from experiments A and B were identical to within a scaling factor. This null hypothesis was not rejected [ $F(318,48177)=0.85$ ,  $p=0.98$ ]. Moreover, the expression (Fig. 3) of this common spatial pattern was not significantly different across experiments [ $t(58)=0.15$ , two-tailed  $p=0.89$ ], i.e., the scaling factor was not significantly different from unity.

Table 1 shows the estimated fMRI % change values associated with experiments A and B at each local maximum. One might wonder whether the variation in these values between experiments does not suggest some systematic difference between experiments in linear fMRI set size effects, despite the formal statements of identical patterns and equivalent pattern expressions in the two experiments. We take up this issue in Discussion.

Because the effects of interest entered into this CVA were the slopes with respect to set size of the fMRI responses associated with the probe period, this spatial pattern is necessarily expressed during the probe period to a greater extent at larger set sizes. It is also possible, though not a necessity, that this pattern's expression would reflect individual differences in memory search rates. Based on the idea that fMRI signal is a temporal integral of neural activity (Boynton et al., 1996; Logothetis et al., 2001), the simplest reasoning would predict that participants with larger memory search durations (i.e., slower memory search rates) would evidence larger effects of set size on the fMRI signal. Such a positive relationship between memory search duration and pattern expression was detected [ $t(56)=2.11$ , one-tailed  $p=0.02$ ], and was not significantly different between experiments [ $t(56)=-.23$ , two-tailed  $p=0.82$ ]. Although statistically significant, the estimated correlation is weak

( $R^2=0.10$ ; Fig. 4); there is the suggestion that a linear relationship is inappropriate, but as this finding is not central to our main hypothesis we do not expand on this issue here. The analogous correlations at the local maxima (Table 1) were all not significant at  $\alpha=0.05$ , suggesting that the implicit averaging manifested in spatial pattern expression is required to detect this subtle effect. Nonetheless, this result does support that expression of this pattern bears the expected direction of relationship to individual differences in memory search rates. In contrast, the relationship between expression of this pattern and individual differences in the reaction time  $y$ -intercept was not significant [ $t(56)=0.41$ , two-tailed

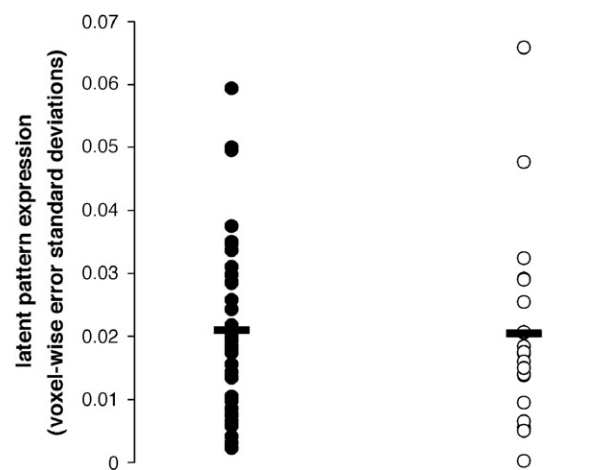


Fig. 3. Expression of brain activity related to memory search. Shown are one-dimensional scatterplots of the observed expression of the spatial pattern representing the slope of brain activity with respect to set size during probe presentation in experiments A (●) and B (○). Each datum represents the observed expression of this pattern in a given participant. The horizontal bars are the predicted (i.e., the mean) expressions of this pattern in each group.

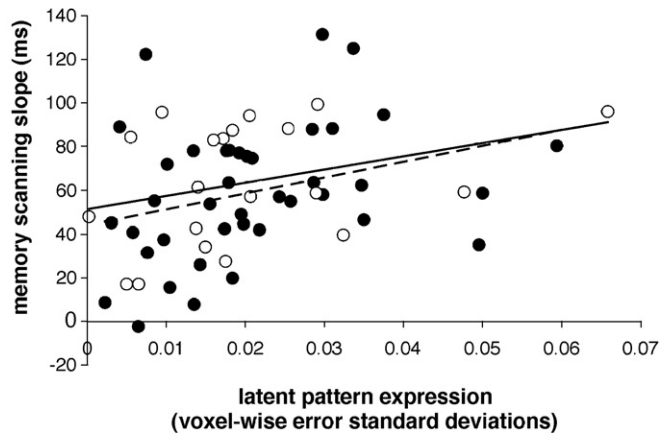


Fig. 4. Observed expressions of the spatial pattern corresponding to the slope of brain activity with respect to set size during the probe period are plotted versus the reaction time slope in experiments A (●) and B (○), along with least-squares lines [expt. A (–), expt. B (–)]. Each datum corresponds to a single participant.

$p=0.68$ ]. The reaction time  $y$ -intercept reflects processes related to perceptual processing of the probe and motor response execution, but not memory search (Sternberg, 1966).

During the retention delay, brain activity related to WM maintenance must occur. Furthermore, activity during the retention delay correlating with memory load (i.e., the amount of

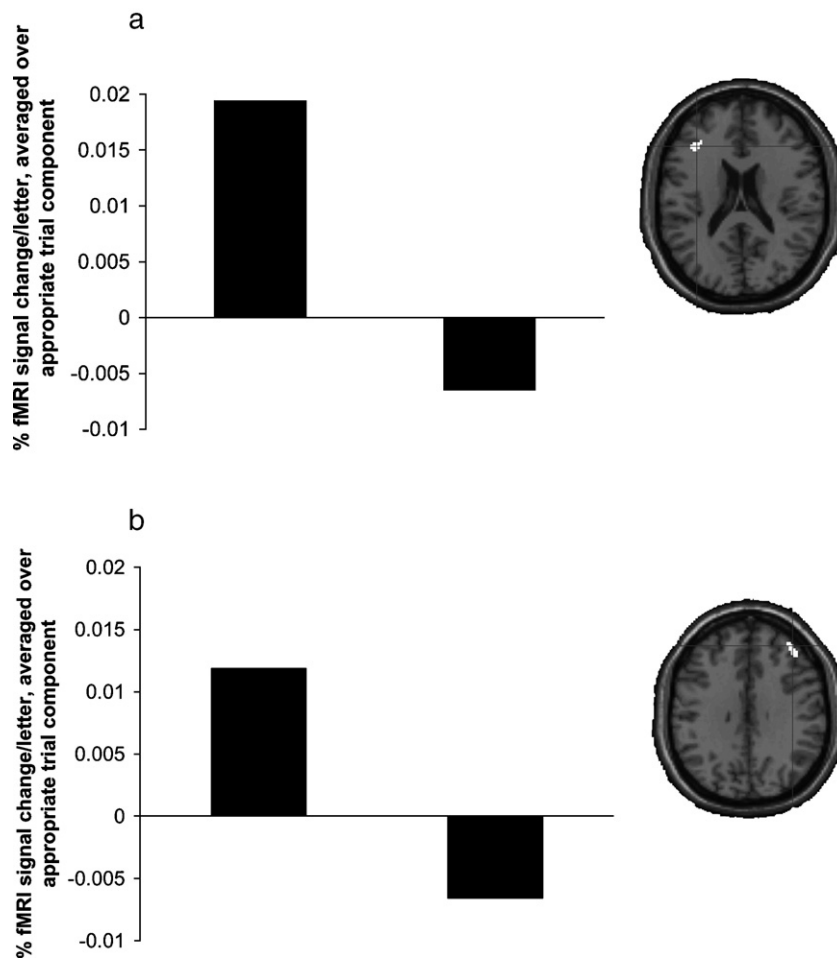


Fig. 5. Two locations in prefrontal cortex are shown that had substantially different retention delay linear fMRI set size effects in experiments A (left bar) and B (right bar). The raw effect sizes are plotted for the local maxima (crosshairs) of each suprathreshold cluster (in white; display thresholding same as in Fig. 2). These axial slices are from the second latent pattern of the retention delay linear fMRI set size effects in the two experiments. This latent pattern was expressed with different signs in, and so tended to reflect areas that had differential responses with respect to, the two experiments. (a) Left middle frontal gyrus, Brodmann area 46, Talairach coordinates [–36, 30, 17]. (b) Right middle frontal gyrus, Brodmann area 9, [40, 44, 27].

information maintained in WM) is more reasonably attributable to WM maintenance than activity that does not increase with memory load. As accuracy did not change as a function of set size, we take set size as a proxy for memory load. Averaging over both experiments, there was a significant linear fMRI set size effect associated with the retention delay [ $F(318,48177)=2.84, p<10^{-6}$ ]. The largest positive spatial weights were limited to the left precentral gyrus and left superior parietal lobule (shown in green in Fig. 2 and listed in Table 1). Both of these brain regions have been reported to be associated with verbal WM maintenance (Wager and Smith, 2003).

To assess within-laboratory replication of this effect, CVA was used to test whether the spatial patterns of the linear fMRI set size effect associated with the retention delay from experiments A and B were identical to within a scaling factor. This was not so [ $F(318,48177)=1.63, p<10^{-6}$ ]. Qualitatively, there were many regions with large estimated set size dependent retention delay effects in experiment A that were much weaker in experiment B (data not shown). These included bilateral middle frontal gyrus (Fig. 5). The effects in left precentral gyrus and left superior parietal lobule were consistent across experiments (Table 1).

#### *Comparison of brain correlates of memory search and verbal WM maintenance*

If memory search and verbal WM maintenance are mediated by the same neural mechanisms, then our logic would have their corresponding activation patterns being scaled versions of one another. Conversely, if their activation patterns are not scaled versions of one another, then we would conclude that memory search and verbal WM maintenance are mediated by separate neural mechanisms. This was tested by specifying these two activation patterns (averaged over experiments) as effects of interest. If sequential latent root testing yielded two significant latent components, this would reject the null hypothesis of pattern identity to within a scaling factor.

Both the first [ $F(635,72037)=2.19, p<10^{-6}$ ] and second [ $F(318,48177)=1.30, p=0.00024$ ] steps of sequential latent root testing yielded significant results. This result is inconsistent with memory search and WM maintenance being mediated by identical neural mechanisms. However, this does not imply that there are no brain regions where there is a substantial correlation of fMRI activity with set size during both WM maintenance and memory search. There was a single cluster of voxels that satisfied  $p<0.001$  (uncorrected for multiple comparisons) for both effects (each averaged across experiments). The cluster maximum Talairach coordinate was  $[-44, 4, 31]$ ; its anatomical label was inferior frontal gyrus, Brodmann area 9. This could indicate a true co-localization of these processes at the neuronal level at this locus, or could instead reflect the existence of neuronal populations separately manifesting these effects, which are spaced closely relative to the resolution of the processed imaging data (estimated full-width-at-half-maximum = 1.3 cm).

## **Discussion**

### *Simultaneous modeling of brain activity associated with memory search and WM maintenance*

In the current study we demonstrated via CVA in conjunction with sequential latent root testing (1) the existence of linear

relationships between set size and fMRI signal associated with memory search, (2) the reliability of the corresponding activation pattern across experiments, (3) the distinction of this activation pattern from that corresponding to set size dependent WM maintenance, and (4) a failure to replicate activation patterns corresponding to set size dependent WM maintenance across experiments. Results 2, 3, and 4 involved explicitly assessing whether two patterns of activation were qualitatively distinct or identical to within a scaling factor. CVA can exactly test such hypotheses. In contrast, SPM cannot in general disambiguate these two hypotheses, and so would not have been an appropriate test.

In contrast to other whole brain fMRI studies during verbal WM maintenance (Rypma et al., 1999; Pollmann et al., 2000; Veltman et al., 2003; Cairo et al., 2004; Kirschen et al., 2005), set size effects during the retention delay were properly accounted for when estimating such effects during the probe period (and vice-versa). This is critical as the neurophysiological component of the fMRI signal is hypothesized to be a temporally lagged and smoothed version of neural activity (Boynton et al., 1996; Logothetis et al., 2001), which would make the fMRI signal temporally associated with the probe period have a dependence on neural activity during the retention delay (this may be thought of as a temporal “partial volume effect”). Simultaneous modeling of the filtered neural responses from all DIR trial components was used to correct for this dependence (Zarahn, 2000). A study by Pollmann and colleagues was mindful of this issue, and used variation of set size and the delay interposed between stimulus set and probe presentation to classify fMRI responses as being related to either retrieval from long term memory, WM maintenance, memory search, or motor response execution (Pollmann et al., 2000). While meticulously devised, their qualitative approach can nevertheless yield ambiguous inference about memory load sensitivity during the retention delay versus the probe period. Though our conclusions are in accord with theirs regarding the involvement of anterior insula and supplementary motor area (medial frontal gyrus) in memory search, they differ concerning parietal cortex, in which their method indicated fMRI responses associated with memory search but ours detected fMRI responses associated with WM maintenance.

In the current study, there were also relatively large probe period set size effects at the border of the left inferior and middle frontal gyrus in the vicinity of Brodmann area 9 and in the frontal opercular/anterior insula region near (but not formally labeled as) Brodmann area 47. In another study (Rypma et al., 2002), a correlation of probe period fMRI signal with set size was significant in dorsolateral (including Brodmann area 9), but not ventrolateral (including Brodmann area 47) prefrontal regions of interest. The spatial resolution difference between region of interest and voxel-wise estimation is a possible explanation for this discrepancy. Also, the maxima near Brodmann area 47 in our study are very close to (and formally labeled as) anterior insula, such that they might not have been included in the ventral prefrontal region of interest of the other study. There were also differences in the time series modeling approaches of that study (Zarahn et al., 1997) and the current one (Zarahn, 2000). Finally, with regard to power, the sample size of that study was small ( $n=8$ ), suggesting the possibility of a false negative.

### *The neural correlates of WM maintenance*

WM is a psychological construct used to describe the limited capacity maintenance (which we have referred to as WM



maintenance) and manipulation of information on a time scale of seconds (Baddeley, 1986; Cowan, 2001). WM seems to be divided into verbal, spatial, and object sub-systems (Tresch et al., 1993; Shah and Miyake, 1996; Hecker and Mapperson, 1997; Barnes et al., 2001). Verbal WM is thought to be critical for language comprehension and reasoning (Baddeley, 1992). Based on neuropsychological dissociations (Warrington and Shallice, 1969; Vallar and Baddeley, 1984) and word length, phonemic similarity, irrelevant speech, and articulatory suppression effects (Belleville et al., 1992; Longoni et al., 1993; Cowan et al., 2000; Baddeley et al., 2002), verbal WM maintenance has been modeled as an articulatory loop in which sub-vocal rehearsal refreshes a phonological store. An early  $O^{15}$  positron emission tomography study identified the left supramarginal gyrus (BA 40) and the left frontal operculum of the inferior frontal gyrus (i.e., presumably Broca's area, BA 44/45) as the neural substrates of the phonological store and rehearsal components, respectively (Paulesu et al., 1993). Subsequent studies have also detected bilateral posterior parietal, left lateral premotor, and supplementary motor area involvement during articulatory loop functioning (Schumacher et al., 1996; D'Esposito et al., 1998; Jonides et al., 1998; Veltman et al., 2003), with the latter two areas seemingly involved in sub-vocal rehearsal (Smith and Jonides, 1999).

In the current study, which used parametric set size manipulation and separate modeling of fMRI signal attributable to the different trial components, retention delay fMRI signal in left premotor and left parietal cortex showed consistent linear set size relationships across two independent experiments; such effects were seen in other areas including prefrontal cortex and anterior insula in only one of these experiments. Aside from a seemingly trivial difference in stimulus selection (see Materials and methods), we as of yet have no explanation for this systematic difference between experiments. For example, primary behavioral measures from the DIR task and basic demographics (i.e., age, IQ, and education) were indistinguishable between the two experiments. Nevertheless, our net finding is consistent with previous data that portions of left premotor and left parietal cortex in part mediate the articulatory loop/verbal WM maintenance (Smith and Jonides, 1999).

Data from a transcranial magnetic stimulation (TMS) study using a DIR task for letters support a necessary role for premotor, but not parietal or prefrontal, cortex in verbal WM maintenance (Herwig et al., 2003), even though all three areas were selected for TMS based on their being observed to have set size effects associated with the retention delay in a companion fMRI experiment. The authors reasoned from their findings that (1) there is essentially a one-way flow of input from premotor to parietal cortex during the retention delay, and (2) prefrontal cortex is involved not in WM maintenance but in encoding, retrieval, manipulation and monitoring (and since the DIR task did not require manipulation or monitoring, there was no affect of temporarily lesioning prefrontal cortex during the retention delay). With regard to (1), reversible lesion via cooling of parietal cortex in the macaque during spatial and non-spatial WM tasks led to changes in neuronal firing in prefrontal areas and vice-versa, implying a reciprocal flow of information between frontal and parietal cortex (Quintana et al., 1989; Chafee and Goldman-Rakic, 2000). Despite this effective connectivity in non-spatial WM tasks, cooling of parietal cortex did not cause decrements in accuracy, while cooling of frontal cortex did (Bauer and Fuster, 1976; Quintana and Fuster, 1993; Quintana et al., 1989). However, cooling of parietal cortex did cause slowing of reaction time (these

studies did not decompose reaction time into memory search dependent and independent components) and reaching inaccuracies (but not so much to cause an increase in qualitative errors). Thus the neurophysiological hypothesis of Herwig and colleagues that information flow is insubstantial from parietal to premotor cortex seems to be incorrect, though their negative behavioral findings obtained using TMS on parietal cortex in humans during a DIR task seem consistent with findings obtained using cortical cooling in non-human primates. With regard to (2), this supposition seems difficult to reconcile with the very fact that the authors reported a set size dependent increase in fMRI signal in prefrontal cortex (that is, if prefrontal cortex is not involved in maintenance, why was there a set size effect during maintenance?). However, in our experiment B, there was essentially nil set size dependent fMRI signal in dorsolateral prefrontal cortex in the presence of preserved performance, which could indirectly point to an epiphenomenal role of prefrontal cortex in verbal WM maintenance in humans.

#### *CVA as a test for replication in neuroimaging*

A failure to replicate a finding within-laboratory using standard meta-analytic techniques can be due to low power (i.e., a false negative in one of the experiments), an unclear understanding of the phenomenon, and/or poor consistency in experimental methodology, including participant sampling. A significant methodological strength of CVA with sequential latent root testing is that it provides a formal test for replication in neuroimaging data, either between- or within-laboratory. Here, it detected a failure to replicate within-laboratory set size dependent effects associated with the retention delay, while supporting replication of set size effects during the probe period. Other failures to replicate similar types of results within-laboratory [e.g., (Rypma and D'Esposito, 1999, 2000; Rypma et al., 2002)] have not involved a formal hypothesis test comparing the studies, and so could be due to low power having caused a false negative in one of the studies. In contrast, as failures to replicate detected with CVA are due to a deviation from the null hypothesis of pattern identity, they cannot be due to low power.

The null hypothesis of the replication test using CVA is that the spatial patterns from the various experiments are identical to within a scaling factor. A failure to reject this null hypothesis is considered a replication, and a rejection of this null hypothesis is taken to refute replication. The test statistic for assessing this null hypothesis is a summation of effects across all the voxels in the dataset (Worsley et al., 1997). This spatial summation of effects implies that spatially diffuse effects that might be too weak to discern at a voxel-wise level will summate. Therefore, it is not a contradiction to detect a failure to replicate without having any particularly large voxel-wise effects evident in the corresponding latent spatial pattern. Also, it is certainly possible to find individual voxels that subjectively suggest the presence of a failure to replicate even when sequential latent root testing fails to detect this at a spatially omnibus level. There are two possible reasons for this: (a) even under the null hypothesis of replication there are likely to be some voxels with apparently disparate (even if not significantly different) estimated effects between studies, and (b) while spatially omnibus tests outperform SPM for detecting diffuse effects, the opposite is true for spatially concentrated effects (Worsley et al., 1995).

In terms of spatial pattern expression, sequential latent root testing failed to detect different spatial patterns associated with

probe period fMRI slope in experiments A and B, supporting within-laboratory replication of this neuroimaging effect. Moreover, expression of the common spatial pattern did not significantly differ between experiments. However, at the local maxima of the probe period fMRI slope spatial pattern, one might feel there is a suggestion that the effects are weaker in experiment A than experiment B (though there were no map-wise significant differences; data not shown), and perhaps in a spatially dependent way. The points made above about local versus global effects need to be used when reconciling these data.

#### *A neural distinction between WM maintenance and memory search*

We tested the idea that memory search, which occurs during the probe period, involves the same neural mechanisms used in WM maintenance. This hypothesis was motivated by certain physiological models that have memory search and WM maintenance involving the same neuronal ensembles (Jensen and Lisman, 1998; Burle and Bonnet, 2000). Also, there are observations that some neurons in prefrontal cortex show stimulus-selective responses during retention delays as well as during decision periods in delayed response or delayed match-to-sample task contexts (Funahashi et al., 1989; Wilson et al., 1993; Desimone, 1996; Miller et al., 1996; Rainer et al., 1998). However, we found that the spatial pattern of brain activity that correlated with set size during memory search was significantly different from the analogous spatial pattern of brain activity during the retention delay. As the test concerned how many latent spatial patterns were required to summarize the data, this finding cannot be explained simply by a difference in the overall intensity of these two effects (which would have still led to a single latent spatial pattern sufficing to explain them both). Furthermore, that the spatial pattern comparison was within-participant, within-task, and within-trial rules out many other alternative explanations. This result implies that memory search and WM maintenance are not mediated by identical brain systems, and hence are not relying on identical cognitive mechanisms.

One dispute on which this dissociation might shed some indirect light is the discrepancy (Jou, 2001) between the capacity of WM, which on average seems to be approximately 4 items, and the near linearity of the reaction time versus set size relationship for memory search, which appears to extend at least up to 7 items (Rypma et al., 2002), as there is no a priori need for different neural mechanisms to have the same capacity constraints. These results also speak against those physiological models that posit the same neuronal ensembles maintain items in WM and are sequentially activated during memory search (Jensen and Lisman, 1998; Burle and Bonnet, 2000), unless those ensembles are restricted to the junction of left premotor cortex and left inferior frontal gyrus where both memory search and WM maintenance effects were present.

#### **Acknowledgments**

This research was supported by NIA grants RO1 AG026158 and RR00645. Eric Zarahn was supported by a NARSAD Young Investigator Award.

#### **References**

Baddeley, A.D., 1986. Working Memory. Oxford Univ. Press, New York.  
Baddeley, A., 1992. Working memory. *Science* 255, 556–559.

- Baddeley, A., Chincotta, D., Stafford, L., Turk, D., 2002. Is the word length effect in STM entirely attributable to output delay? Evidence from serial recognition. *Q. J. Exp. Psychol.*, A 55A, 353–369.
- Barnes, L.L., Nelson, J.K., Reuter-Lorenz, P.A., 2001. Object-based attention and object working memory: overlapping processes revealed by selective interference effects in humans. *Prog. Brain Res.* 134, 471–481.
- Bauer, R.H., Fuster, J.M., 1976. Delayed-matching and delayed-response deficit from cooling dorsolateral prefrontal cortex in monkeys. *J. Comp. Physiol. Psychol.* 90, 293–302.
- Belleville, S., Peretz, I., Arguin, M., 1992. Contribution of articulatory rehearsal to short-term memory: evidence from a case of selective disruption. *Brain Lang.* 43, 713–746.
- Boynton, G.M., Engel, S.A., Glover, G.H., Heeger, D.J., 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.* 16, 4207–4221.
- Burle, B., Bonnet, M., 2000. High-speed memory scanning: a behavioral argument for a serial oscillatory model. *Cogn. Brain Res.* 9, 327–337.
- Cairo, T.A., Liddle, P.F., Woodward, T.S., Ngan, E.T., 2004. The influence of working memory load on phase specific patterns of cortical activity. *Cogn. Brain Res.* 21, 377–387.
- Chafee, M., Goldman-Rakic, P., 2000. Inactivation of parietal and prefrontal cortex reveals interdependence of neural activity during memory-guided saccades. *J. Neurophysiol.* 550–1566.
- Cohen, J., MacWhinney, B., Flatt, M., Provost, J., 1993. PsyScope: an interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behav. Res. Methods Instrum. Comput.* 25, 257–271.
- Cowan, N., 2001. The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behav. Brain Sci.* 24, 87–185.
- Cowan, N., Nugent, L.D., Elliott, E.M., Geer, T., 2000. Memory-search and rehearsal processes and the word length effect in immediate recall: a synthesis in reply to service. *Q. J. Exp. Psychol.*, A 53A, 666–670.
- Desimone, R., 1996. Neural mechanisms for visual memory and their role in attention. *Proc. Natl. Acad. Sci. U. S. A.* 93, 13494–13499.
- D'Esposito, M., Aguirre, G.K., Zarahn, E., Ballard, D., Shin, R.K., Lease, J., 1998. Functional MRI studies of spatial and nonspatial working memory. *Cogn. Brain Res.* 7, 1–13.
- Friston, K.J., Stephan, K.E., Lund, T.E., Morcom, A., Kiebel, S., 2005. Mixed-effects and fMRI studies. *NeuroImage* 24, 244–252.
- Funahashi, S., Bruce, C.J., Goldman-Rakic, P.S., 1989. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* 61, 331–349.
- Green, D., 1988. Signal Detection Theory and Psychophysics, Peninsula Publishing, Los Altos, CA.
- Habeck, C., Rakitin, B.C., Moeller, J., Scarmeas, N., Zarahn, E., Brown, T., Stern, Y., 2004. An event-related fMRI study of the neurobehavioral impact of sleep deprivation on performance of a delayed-match-to-sample task. *Cogn. Brain Res.* 18, 306–321.
- Habeck, C., Rakitin, B.C., Moeller, J., Scarmeas, N., Zarahn, E., Brown, T., Stern, Y., 2005. An event-related fMRI study of the neural networks underlying the encoding, maintenance, and retrieval phase in a delayed-match-to-sample task. *Cogn. Brain Res.* 23, 207–220.
- Hecker, R., Mapperson, B., 1997. Dissociation of visual and spatial processing in working memory. *Neuropsychologia* 35, 599–603.
- Herwig, U., Abler, B., Schonfeldt-Lecuona, C., Wunderlich, A., Grothe, J., Spitzer, M., Walter, H., 2003. Verbal storage in a premotor-parietal network: evidence from fMRI-guided magnetic stimulation. *NeuroImage* 20, 1032–1041.
- Jensen, O., Lisman, J.E., 1998. An oscillatory short-term memory buffer model can account for data on the Sternberg task. *J. Neurosci.* 18, 10688–10699.
- Jonides, J., Schumacher, E.H., Smith, E.E., Koeppel, R.A., Awh, E., Reuter-Lorenz, P.A., Marshuetz, C., Willis, C.R., 1998. The role of parietal cortex in verbal working memory. *J. Neurosci.* 18, 5026–5034.

- Jou, J., 2001. The magic number four: can it explain Sternberg's serial memory scan data? *Behav. Brain Sci.* 24, 126–127.
- Kirschen, M.P., Chen, S.H., Schraedley-Desmond, P., Desmond, J.E., 2005. Load- and practice-dependent increases in cerebro-cerebellar activation in verbal working memory: an fMRI study. *NeuroImage* 24, 462–472.
- Kwong, K.K., Belliveau, J.W., Chesler, D.A., Goldberg, I.E., Weisskoff, R.M., Poncelet, B.P., Kennedy, D.N., Hoppel, B.E., Cohen, M.S., Turner, R., 1992. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc. Natl. Acad. Sci. U. S. A.* 89, 5675–5679.
- Lancaster, J.L., Woldorff, M.G., Parsons, L.M., Liotti, M., Freitas, E.S., Rainey, L., Kochunov, P.V., Nickerson, D., Mikiten, S.A., Fox, P.T., 2000. Automated Talairach Atlas labels for functional brain mapping. *Hum. Brain Mapp.* 10, 120–131.
- Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., Oeltermann, A., 2001. Neurophysiological investigation of the basis of the fMRI signal [see comment] *Nature* 412, 150–157.
- Longoni, A.M., Richardson, J.T., Aiello, A., 1993. Articulatory rehearsal and phonological storage in working memory. *Mem. Cogn.* 21, 11–22.
- Miller, E.K., Erickson, C.A., Desimone, R., 1996. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J. Neurosci.* 16, 5154–5167.
- Ogawa, S., Menon, R.S., Tank, D.W., Kim, S.G., Merkle, H., Ellermann, J.M., Ugurbil, K., 1993. Functional brain mapping by blood oxygenation level-dependent contrast magnetic resonance imaging. A comparison of signal characteristics with a biophysical model. *Biophys. J.* 64, 803–812.
- Paulesu, E., Frith, C.D., Frackowiak, R.S., 1993. The neural correlates of the verbal component of working memory. [see comment]. *Nature* 362, 342–345.
- Pollmann, S., Zysset, S., Wiggins, C.J., Von Cramon, D.Y., 2000. Dissociation of memory retrieval and search processes: an event-related fMRI study. *Microsc. Res. Tech.* 51, 29–38.
- Quintana, J., Fuster, J., 1993. Spatial and temporal factors in the role of prefrontal and parietal cortex in visuomotor integration. *Cereb. Cortex* 3, 122–132.
- Quintana, J., Fuster, J.M., Yajeya, J., 1989. Effects of cooling parietal cortex on prefrontal units in delay tasks. *Brain Res.* 503, 100–110.
- Rainer, G., Asaad, W.F., Miller, E.K., 1998. Memory fields of neurons in the primate prefrontal cortex. *Proc. Natl. Acad. Sci. U. S. A.* 95, 15008–15013.
- Rypma, B., D'Esposito, M., 1999. The roles of prefrontal brain regions in components of working memory: effects of memory load and individual differences. *Proc. Natl. Acad. Sci. U. S. A.* 96, 6558–6563.
- Rypma, B., D'Esposito, M., 2000. Isolating the neural mechanisms of age-related changes in human working memory. *Nat. Neurosci.* 3, 509–515.
- Rypma, B., Prabhakaran, V., Desmond, J.E., Glover, G.H., Gabrieli, J.D.E., 1999. Load-dependent roles of frontal brain regions in the maintenance of working memory. *NeuroImage* 9, 216–226.
- Rypma, B., Berger, J.S., D'Esposito, M., 2002. The influence of working-memory demand and subject performance on prefrontal cortical activity. *J. Cogn. Neurosci.* 14, 721–731.
- Schumacher, E.H., Lauber, E., Awh, E., Jonides, J., Smith, E.E., Koeppe, R.A., 1996. PET evidence for an amodal verbal working memory system. *NeuroImage* 3, 79–88.
- Shah, P., Miyake, A., 1996. The separability of working memory resources for spatial thinking and language processing: an individual differences approach. *J. Exp. Psychol. Gen.* 125, 4–27.
- Smith, E.E., Jonides, J., 1999. Storage and executive processes in the frontal lobes. *Science* 283, 1657–1661.
- Sternberg, S., 1966. High-speed scanning in human memory. *Science* 153, 652–654.
- Talairach, J., Tournoux, P., 1988. *Co-Planar Stereotactic Atlas of the Human Brain. Three-Dimensional Proportional System: An Approach of Cerebral Imaging.* Theime Medical Publisher, New York.
- Townsend, J.T., 1990. Serial vs. parallel processing—Sometimes they look like tweedledum and tweedledee but they can (and should) be distinguished. *Psychol. Sci.* 1, 46–54.
- Tresch, M.C., Sinnamon, H.M., Seamon, J.G., 1993. Double dissociation of spatial and object visual memory: evidence from selective interference in intact human subjects. *Neuropsychologia* 31, 211–219.
- Vallar, G., Baddeley, A.D., 1984. Fractionation of working memory—Neuropsychological evidence for a phonological short-term store. *J. Verbal Learn. Verbal Behav.* 23, 151–161.
- Veltman, D.J., Rombouts, S., Dolan, R.J., 2003. Maintenance versus manipulation in verbal working memory revisited: an fMRI study. *NeuroImage* 18, 247–256.
- Wager, T.D., Smith, E.E., 2003. Neuroimaging studies of working memory: a meta-analysis. *Cogn. Affect. Behav. Neurosci.* 3, 255–274.
- Warrington, E.K., Shallice, T., 1969. The selective impairment of auditory verbal short-term memory. *Brain* 92, 885–896.
- Wilson, F.A., Scalaidhe, S.P., Goldman-Rakic, P.S., 1993. Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science* 260, 1955–1958.
- Worsley, K.J., Poline, J.B., Vandal, A.C., Friston, K.J., 1995. Tests for distributed, nonfocal brain activations. *NeuroImage* 2, 183–194.
- Worsley, K.J., Poline, J.B., Friston, K.J., Evans, A.C., 1997. Characterizing the response of PET and fMRI data using multivariate linear models. *NeuroImage* 6, 305–319.
- Zarahn, E., 2000. Testing for neural responses during temporal components of trials with BOLD fMRI. *NeuroImage* 11, 783–796.
- Zarahn, E., Aguirre, G., D'Esposito, M., 1997. A trial-based experimental design for fMRI. *NeuroImage* 6, 122–138.
- Zarahn, E., Rakitin, B., Abela, D., Flynn, J., Stern, Y., 2005. Positive evidence against human hippocampal involvement in working memory maintenance of familiar stimuli. *Cereb. Cortex* 15, 303–316.