

Localization of load sensitivity of working memory storage: Quantitatively and qualitatively discrepant results yielded by single-subject and group-averaged approaches to fMRI group analysis

Eva Feredoes* and Bradley R. Postle

Department of Psychology, University of Wisconsin-Madison, 1202 West Johnson St., Madison, WI 53706, USA

Received 21 April 2006; revised 7 December 2006; accepted 11 December 2006
Available online 3 January 2007

The impetus for the present report is the evaluation of competing claims of two classes of working memory models: *Memory systems* models hold working memory to be supported by a network of prefrontal cortex (PFC)-based domain-specific buffers that act as workspaces for the storage and manipulation of information; *emergent processes* models, in contrast, hold that the contributions of PFC to working memory do *not* include the temporary storage of information. Empirically, each of these perspectives is supported by seemingly mutually incompatible results from functional magnetic resonance imaging (fMRI) studies that either do or do not find evidence for delay-period sensitivity to memory load, an index of storage, in PFC. We hypothesized that these empirical discrepancies may be due, at least in part, to methodological factors, because studies reporting delay-period load sensitivity in PFC typically employ spatially normalized group averaged analyses, whereas studies that do not find PFC load sensitivity typically use a single-subject “case-study” approach. Experiment 1 performed these two approaches to analysis on the same data set, and the results were consistent with our hypothesis. Experiment 2 evaluated one characteristic of the single-subject results from Experiment 1 – considerable topographical variability across subjects – by evaluating its test–retest reliability with a new group of subjects. Each subject was scanned twice, and the results indicated that, for each of several contrasts, test–retest reliability was significantly greater than chance. Together, these results raise the possibility that the brain bases of delay-period load sensitivity may be characterized by considerable intersubject topographical variability. Our results

highlight how the selection of fMRI analysis methods can produce discrepant results, each of which is consistent with different, incompatible theoretical interpretations.

© 2006 Elsevier Inc. All rights reserved.

Introduction

Working memory refers to the ability to retain information in an active, accessible state when it is not present in the environment, to manipulate this information when necessary, and to use it to guide behavior (Baddeley and Hitch, 1974; Miller et al., 1960). Because individual differences in working memory predict a remarkable breadth of behavioral outcomes, from general fluid intelligence and Stroop task performance to reading ability and standardized test performance (e.g., Baddeley, 1986; Daneman and Carpenter, 1980; Engle et al., 1999; Gathercole and Pickering, 2000; Jonides, 1995), its cognitive and neural organization are of central interest to cognitive neuroscience. The impetus for the present report is the evaluation of competing claims of two classes of working memory models. *Memory systems* models hold working memory to be supported by a network of domain-specific buffers that act as workspaces for the storage and manipulation of information. In what is arguably the most influential of these models, these buffers include the phonological store/articulatory loop, the visuospatial sketchpad, and the episodic buffer (Baddeley, 2000; Baddeley and Logie, 1999). Many memory systems accounts also stipulate that regions of the frontal cortex, particularly the prefrontal cortex (PFC), are the critical neural substrates of these working memory storage buffers (e.g., Courtney, 2004; Davachi et al., 2004; Goldman-Rakic and Leung, 2002; Haxby et al., 2000; Leung et al., 2002, 2004; Logie and Della Salla, 2003; Mottaghy et al., 2002; Munk et al., 2002; Pessoa et al., 2002; Sala et al., 2003; Slotnick, 2005; Tek et al., 2002). An alternative to the systems framework portrays working memory as a property that arises through the coordinated recruitment, via attention, of brain systems

Abbreviations: BOLD; blood oxygen level-dependent signal; BA; Brodmann area; fMRI; functional magnetic resonance imaging; GLM; general linear model; PFC; prefrontal cortex; HRF; hemodynamic response function; M.N.I.; Montreal Neurological Institute; ROI; region of interest; RT; reaction-time; SS; single-subject; SNGA; spatially-normalized group-averaged analysis; SPL; Superior Parietal Lobule.

* Corresponding author. Fax: +1 608 262 4029.

E-mail address: feredoes@wisc.edu (E. Feredoes).

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

that have evolved to accomplish sensory-, representation-, or action-related functions. One corollary of this *emergent processes* view is that the contributions of PFC to working memory do *not* include the temporary storage of information (for a review, see Postle, 2006).

Dating at least to the seminal publications of Goldman-Rakic and her colleagues in the late 1980s and early 1990s, evidence for PFC delay-period activity that is consistent with a storage function has been viewed as a cornerstone of evidence for systems models. One variable that has received a great deal of scrutiny is that of the domain of information to be remembered, and whether or not the topographical organization of PFC delay-period activity associated with different domains is segregated. (For example, does working memory for the location vs. the identity of a stimulus preferentially recruit dorsolateral vs. ventrolateral regions of PFC, respectively?) Indeed, the question of the organization-by-domain of working memory function in PFC has become a conceptual battleground between adherents to systems models vs. advocates of alternative views (e.g., Duncan and Owen, 2000; Fuster, 2002; Goldman-Rakic, 2000; Goldman-Rakic and Leung, 2002; Haxby et al., 2000; Miller, 2000; Muller et al., 2002; Owen et al., 1999; Passingham and Rowe, 2002; Petrides, 2000; Postle and D'Esposito, 2000; Postle et al., 2003; Sala et al., 2003; Slotnick, 2005; Ungerleider et al., 1998).

This report focuses on another factor that can shed light on the neural bases of memory storage processes, the factor of load. Memory load refers the number of items that must be retained during a working memory task, and a region whose delay-period activity varies systematically with load is a candidate substrate for memory storage functions.¹ As is the case with domain specificity, evidence for load sensitivity of PFC is mixed. One study has investigated the effects of varying memory load on delay-period activity in PFC with a task that compared the effects of varying load with those of varying manipulation demands on delayed recognition of the ordinal position of letter stimuli (Postle et al., 1999). Load was operationalized with the contrast of delay-period activity from trials with a 5-letter memory set vs. delay-period activity from trials with a 2-letter memory set. This study also operationalized the concept of *manipulation* in working memory by requiring subjects, on some trials, to reposition into alphabetical order the 5 randomly ordered letters of a memory set. The results revealed a significant manipulation effect (i.e., [$\text{Delay}_{\text{Alphabetize } 5} - \text{Delay}_{\text{Forward } 5}$]) in dorsolateral PFC in 5 of 5 subjects, and load effects in dorsolateral PFC in only 2 subjects. Reliable load effects were, however, seen in all subjects in left posterior perisylvian cortex.

Corroborating results came from a functional magnetic resonance imaging (fMRI) study by Rypma and D'Esposito (1999) that found load effects in PFC during the encoding period, but not the delay period, and delay-period load effects in left inferior parietal cortex. Subsequent research, however, has painted a more complex picture of the effects of varying load on delay-period activity in PFC. One study that varied load parametrically between 1 and 8 letters revealed no encoding-related load effects,

but significant delay- and probe-related load effects in dorsolateral PFC, and trends in this direction in ventrolateral PFC. An individual-differences analysis indicated that these patterns of load-dependent effects by trial epoch varied between high- and low-performing groups, leading the authors to interpret the delay-period load effects in PFC as evidence for strategic reorganization of information, rather than for storage per se (Rypma et al., 2002).

Another study reported load effects in PFC with an n-back task and an item recognition task (Veltman et al., 2003), but because the design of neither task permitted isolation of delay-period activity, the implications of this study for storage processes is unclear. Finally, three very recent studies using verbal material have produced inconsistent results: Narayanan et al. (2005) and Zarahn et al. (2005) found delay-period sensitivity to load in PFC, whereas Postle et al. (2006), in a replication and extension of the Postle et al. (1999) study that used both fMRI and fMRI-guided repetitive transcranial magnetic stimulation (rTMS), did not. This latter study also found that delay-period rTMS of PFC disrupted performance on *Alphabetize 5*, but not *Forward 5*, trials, whereas delay-period rTMS of the superior parietal lobule (SPL) disrupted performance on both types of trials.

What might account for these discrepant findings in the “load” literature? Consideration of the studies reviewed here, plus others (reviewed in Postle, 2006), suggests that one factor might be methodological: Narayanan et al. (2005) and Zarahn et al. (2005) performed spatially-normalized group-average (SNGA) analyses; whereas Postle and colleagues (1999, 2006) used a single-subject (SS) approach. The empirical portions of this paper, therefore, will be devoted to evaluating the hypothesis that group analyses employing a SNGA approach are likely to find evidence for delay-period load sensitivity in PFC, whereas group analyses employing the SS approach are not.

The development of these two approaches to fMRI data analysis can be traced, in part, to two different traditions of brain research. The SNGA approach was pioneered by researchers with previous H_2O^{15} positron emission tomography experience (e.g., Buckner et al., 1995; Corbetta et al., 1990; Courtney et al., 1996; Evans et al., 1988; Fox et al., 1988; Friston et al., 1990; Haxby et al., 1995; Petersen et al., 1988; Raichle et al., 1983), a method whose relatively poor signal-to-noise characteristics required the averaging across data from multiple subjects in order to maximize sensitivity. With the advent of fMRI, the SNGA approach has remained popular, in large part because of the population-level inference that it supports when subject is treated as a random factor. Because SNGA analyses will only detect effects that are topographically overlapping in the brains of the majority of subjects, their use entails the assumption that there exist single structure–function mappings that are characteristic of the “average” person (i.e., the prototype that is representative of the population from which the sample was drawn).

SS approaches, in contrast, were developed to analyze fMRI data, often with the goal of emulating extracellular electrophysiology experiments in awake, behaving monkeys. Thus, for example, SS analyses have been used for retinotopic mapping of the human visual system and other experiments explicitly designed to replicate and extend what is known about the functional neuroanatomy of the monkey visual system (e.g., Engel et al., 1994; Levy et al., 2001; Sereno et al., 1995; Tootell et al., 1993, 1995). The SS approach used in the Postle et al. (1999, 2006) studies was developed to permit analysis of delay-period activity during delayed response and delayed recognition (Zarahn et al.,

¹ There are, of course, other processes that may be sensitive to load, such as chunking (Bor et al., 2003; Rypma and D'Esposito, 1999) or other types of top-down control (Miller and Cohen, 2001). Nonetheless, load-sensitivity has been viewed as a necessary, although not sufficient, property of a brain area hypothesized to support storage functions (Glahn et al., 2002; Manoach et al., 1997; Narayanan et al., 2005; Veltman et al., 2003; Zarahn et al., 2005). For recent reviews see (Postle, 2006; Rypma, 2006).

1997a) in a manner analogous to studies in the monkey (e.g., Funahashi et al., 1989; Fuster and Alexander, 1971; Niki and Watanabe, 1976). SS analyses have also been used to study basic physiological processes, such as the hemodynamic response function (Boynton et al., 1996) and adaptation (Grill-Spector and Malach, 2001; Murray et al., 2006). SS analyses are often used to identify and/or interrogate functionally defined regions whose topographical location can vary considerably across subjects, due to differences in gross anatomy or other factors (e.g., Aguirre et al., 1998a; Kanwisher et al., 1997; Peelen and Downing, 2005; Ranganath et al., 2004). Finally, SS analyses have been adopted by many investigators employing relatively recently developed multivariate analysis approaches (e.g., Haynes and Rees, 2005; Kamitani and Tong, 2005; Polyn et al., 2005). One study, for example, used a neural-network pattern classifier to identify patterns of whole-brain activity, at the SS level, associated with the processing of face, location, and object stimuli. In a subsequent scan, these category-specific patterns were shown to be reinstated in advance of free recall of items from the category in question (Polyn et al., 2005). The rationale for analyzing SS data was twofold: first, incorporating “all of the voxels that show[ed] differential activity across conditions in a given subject” provided “increase[d] sensitivity” over “focusing on ROIs (regions of interest) obtained using a group analysis” (p. 22 of Supporting Online Material for Polyn et al., 2005); second, the analysis was premised on associating fMRI activity with subjects’ free recall behavior, and because this behavior was almost certain to differ markedly across individual subjects, brain–behavior correlations could only be investigated on a subject by subject basis.

SNGA and SS analyses differ in the nature of effects that each can detect. SS analyses are less well suited to detect subtle effects that may be reliable at the group level but that do not exceed threshold in some or all of the SS data sets in the study. (This caveat is less applicable, however, when region of interest (ROI)-based analyses are used.) For SNGA analyses, constraints arise because this approach entails the warping of data (either the fMRI data themselves or volumes of parameter estimates computed from the fMRI data) from each subject into a common atlas space (typically Montreal Neurological Institute (M.N.I.) or another Talairach space), then evaluating the reliability of the effect in question, across subjects, at each voxel. As a result, this approach will only detect effects that occur in the same location in atlas space in the majority of subjects. Indeed, because of this feature, an assumption that (either implicitly or explicitly) underlies the SNGA approach is that there exists a single mapping from anatomy to function in the human brain, invariant across the population (e.g., healthy young adults) at a certain spatial scale. It follows from this that the SNGA approach is a priori less well suited to detect activity associated with a function whose intersubject anatomical variability exceeds this spatial scale. (This is for the same reason that, more generally, the arithmetic mean is ill-suited to summarize a multimodal distribution).

One empirical example of this latter situation comes from a recent study of face selectivity in the monkey visual system. Whereas classical electrophysiological studies have found local regions of cortex to contain at most 20–30% face selective neurons, a combined fMRI/electrophysiology study found that fMRI-identified face-selective regions contained >90% face-selective cells. The anatomical variability of these regions across animals makes it unlikely, however, that they would be identified by a SNGA analysis (Tootell, personal communication; Tsao et al.,

2006). Despite this limitation, a clear advantage of the SNGA approach is that when it identifies effects of interest, assuming that the factor of *subject* has been treated as a random effect, the inference from this result to the population from which the sample of subjects is drawn is straightforward.

Population inference from SS analyses can be decidedly more complicated. This is because of the ambiguity in deciding whether or not particular effects from multiple SS data sets can be considered sufficiently comparable that they can be averaged across subjects. In some instances, this is relatively straightforward, such as when first identifying face-selective ROIs in SS data sets, then interrogating these ROIs with an independent contrast (e.g., sensitivity to encoding and maintenance of face vs. “place” information, Ranganath et al., 2004). (Although for recent debate on the use of such “functional localizers,” see (Friston and Henson, 2006; Friston et al., 2006; Saxe et al., 2006).) In cases such as this, effects of interest can be computed from the ROI(s) of each SS data set, and the reliability of these effects across the sample evaluated with a standard tool such as a *t*-test or analysis of variance (which would typically treat the factor of *subject* as a random effect). In other cases, as we shall see in this report, the intersubject topographical variability of the effects of interest is sufficiently great (e.g., spanning lobes and/or hemispheres) that the assumption of “comparability” across effects drawn from SS data sets is less obvious. To foreshadow our results, the SS analyses in Experiment 1 will indeed reveal large intersubject topographical variability of load sensitivity. This will lead to Experiment 2, an empirical investigation of the test–retest reliability of this and other effects at the SS level, and to discussion of how to address the problem of population inference with SS analyses.

In the present study we reasoned that the “true” topographical pattern of working memory load sensitivity that underlies all of the studies summarized up to this point may be one of relatively small intensity, but topographically relatively invariant, effects in PFC, and of relatively large intensity, but topographically highly variable, effects in non-PFC regions. By this reasoning, small effects not detectable by the SS approach but detected by SNGA analyses, by virtue of the topographic overlay of many small effects, occur in PFC. Large, topographically variable effects can, however, be detected in the SS, but not in SNGA maps. Consistent with this prediction is the fact that effects detected by SNGA analyses tend to be smaller than the mean of the peak magnitudes of the same effects within any single subject, or averaged across SS analyses. For example, Narayanan et al. (2005) and Zarahn et al. (2005) studies that used a SNGA analysis approach produced PFC delay-period load effects that, from their figures, appear to be of approximately 0.2% signal change. The SS group analyses of Postle et al. (1999, 2006), in contrast, yielded group load effect sizes that are an order of magnitude larger.²

² Note that across fMRI analysis packages, differences exist in the way that baseline, from which the percentage signal change is measured, is calculated. Postle et al. (1999, 2006) and Zarahn et al. (2005) used the same method, in which intensity was normalized via voxel-wise division of the time series mean. Narayanan et al. (2005), on the other hand, determined a baseline specific to each ROI by firstly calculating a session-specific grand mean from each ROI, then dividing the values within each session, within each ROI, by the mean and multiplying this value by 100 (Christoff et al., 2001). Thus, this baseline measure was similar to those of the other experiments, but calculated within each ROI rather than within each voxel.

Table 1
Behavioral results from Experiment 1

Instruction	Mean accuracy (% correct)	Mean RT (ms)
Alphabetize	74.84 (18.22)	1722.85 (553.18)
Forward 5	79.61 (14.30)	1705.27 (535.47)
Forward 2	89.59 (16.20)	1381.59 (517.43)

Numbers in parentheses indicate standard deviations.

In Experiment 1 we analyzed the same data set with the SNGA and SS methods. We predicted that these two analyses would yield two discrepant sets of results. The SS analysis would identify load-sensitive voxels in regions that would vary from subject to subject, but that would not reliably localize to the PFC. The SNGA analysis, in contrast, would identify a region of load-sensitivity in the PFC. Furthermore, we predicted that the effect size identified by the SNGA analysis would be markedly smaller (i.e., of lower intensity) than the mean effect size identified by the SS analysis.

Experiment 1

To test our hypothesis, we analyzed the working memory data from 24 subjects with SNGA and with SS analyses. Subjects performed a delayed-recognition of letter order task that varied memory load and executive control.

Methods

Subjects

Twenty-four right-handed subjects (mean age = 22.7 years [S.D. = 3.18], males = 12) were recruited in total from the University of Wisconsin-Madison community. The data from 12 of these are also reported elsewhere (Postle et al., 2006). All subjects were paid for their participation and the study was approved by the local institutional review board.

Behavioral task

The task, requiring delayed recognition of item ordinal position, was identical to that used by Postle et al. (1999, 2006). Each trial began with the simultaneous presentation of two or five consonant letters in a single row for 1 s, followed by instructions (“Forward” or “Alphabetize”) for 500 ms, followed by an 8 s delay period, followed by a memory probe consisting of an item from the memory set and a digit, followed by a 13 s intertrial interval. On Forward trials, subjects were to retain a memory of the two or five letters in the order in which they were presented, and indicate with a YES/NO button press whether or not the probe digit represented the ordinal position in which the probed letter had appeared in the initial

stimulus display (it did so with $p=0.5$). On Alphabetize trials, subjects were to reorder the letters into alphabetical order. On these trials they were to indicate with a YES/NO button press whether the probe digit represented the alphabetical position of the probed letter with respect to the other four letters in the memory set (it did so with $p=0.5$). Trial type was pseudorandomized, with the constraints that an equal number of each trial type appeared in each 12-trial block, as did an equal number of valid and invalid probes. Alphabetization trials were included as part of another experiment (Postle et al., 2006) and will not be considered further in Experiment 1.

fMRI

Data acquisition and preprocessing

Whole-brain images were acquired with a 3T scanner (GE Signa VH/I). High resolution T1-weighted images (30 axial slices, $0.9375\text{ mm} \times 0.9375\text{ mm} \times 4\text{ mm}$) were obtained in all participants, and a gradient echo, echoplanar sequence (TR = 2000 ms, TE = 50 ms) was used 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, $3.75\text{ mm} \times 3.75\text{ mm} \times 4\text{ mm}$, no skip). Scans of the delayed-recognition task were preceded by a scan used to derive an estimate of the hemodynamic response function (HRF) for each participant. The HRF characterizes the fMRI response resulting from a brief impulse of neural activity (Boynton et al., 1996), and can vary markedly across participants (Aguirre et al., 1998b; Handwerker et al., 2004). During this 7 min scan, each participant performed a simple reaction-time (RT) task that required a bimanual button press once every 20 s in response to a brief change in shape of the fixation stimulus. The eight scans of the working memory task each lasted 6 min 20 s (6 min of task preceded by 20 s of discarded acquisitions to achieve a steady state of tissue magnetization).

Data analysis

A partial F -test associated with a Fourier basis covariate set (Josephs et al., 1997) was applied to the HRF derivation data from each subject in order to determine the significance of task correlated activity in each voxel of primary somatosensory and motor cortex. An HRF estimate was extracted from suprathreshold voxels by spatially averaging their time series, filtering the resultant averaged fMRI time series to remove high ($>0.244\text{ Hz}$ $<1\text{ Hz}$) frequencies, adjusting it to remove the effects of nuisance covariates (Friston et al., 1995), and trial averaging. Our application of the general linear model (GLM) takes into account an empirical estimate of the temporal autocorrelation of BOLD data and has been demonstrated empirically to produce a mapwise false positive rate of ≤ 0.05 with unsmoothed data when Bonferroni correction is applied (Zarahn et al., 1997b). The subject-specific HRF was used to convolve the delta (i.e., “stick”) function covariates of the modified GLM

Notes to Table 2:

Brodman areas (BA) are approximated, and listed for descriptive purposes. Effect sizes are collapsed across all voxels within a region. If more than one region displayed a reliable load effect, effect sizes are shown separately for each in this table. For all summary statistics relating to the SS group analysis, however, a single load-sensitivity effect size was computed for each subject by collapsing across all load-sensitive voxels. A = Subject; B = Hemisphere; C = number of voxels; D = Load effect size (%); E = Euclidian distance between the SS- and SNGA-identified regions (mm). For each SS data set, if there was more than one voxel for a SS load-sensitive region then distance was calculated from the local maximum of the region; if there was more than one local maximum within a region, the distance between each maximum and the SNGA PFC ROI was calculated, and the average of these distances reported. CS = central sulcus; PoCG = postcentral gyrus; PoCS = postcentral sulcus; PreCG = precentral gyrus; PreCS = precentral sulcus; SFS = superior frontal sulcus; SMG = supramarginal gyrus; AG = angular gyrus; IPS = intraparietal sulcus; SFG = superior frontal gyrus; STS = superior temporal sulcus; STG = superior temporal gyrus; MTG = middle temporal gyrus; ACC = anterior cingulate cortex; FG = fusiform gyrus; ParaHG = parahippocampal gyrus.

Table 2
Results from the SS analyses, including the SNGA-derived PFC ROI

A	Results from SS analyses					SNGA derived load masks								
	B	Region (BA)	C	D	E	Load effect sizes (%) by ROI voxel size (3,7,50 or 100 voxels)								
						3	7	50	100					
1	L	Caudal bank preCS (4/6)	1	6.639	11.9	0.460	0.033	0.150	0.181					
	L	Thalamus	1	2.683	52.7									
2	R	Cerebellum	1	2.032	96.2	0.991	0.686	0.266	−0.006					
3	L	Border of SMG (40)/STG(22)	2	1.459	26.5	−0.043	−0.143	−0.076	−0.046					
	L	ACC (24)	3	0.925	50.6									
4	L	SPL (7)	2	0.890	81.7	1.774	1.006	0.732	0.143					
	R	SPL (7)	1	0.896	95.0									
	L	SMG (40)	8	0.969	54.3									
	R	SMG (40)	2	0.747	67.5									
	R	AG (39)	2	0.834	109.9									
	L	AG (39)	13	1.668	47.2									
	L	MFG (9/46)	6	1.127	30.5									
	R	MFG (9/46)	8	1.131	99.1									
	L	IFG (44/45)	12	1.084	24.1									
	L	SFG (6)	4	1.573	48.1									
	L	ACC (24)	2	0.783	49.1									
	R	ACC (24)	6	0.919	48.8									
	L	MTG (37)	2	1.387	54.5									
	L	Occipital gyrus (19)	4	1.150	85.3									
	L	STG/ParaHG (22)	1	1.865	43.4									
	R	FG (20)	4	0.906	115.5									
	5	L	mid-CS (1/4)	1	2.903					17.2	−0.245	−0.047	−0.016	−0.027
		R	SFS/PreCS (4/6/46)	2	2.343					64.7				
	6	L	PoCG (3)	1	1.892					37.5	0.693	0.504	0.254	−0.033
		L	Thalamus	2	1.776					74.8				
7	L	PostCG (2/1)	1	1.173	30.7	0.346	0.145	0.14	−0.165					
	R	PostCG (2/1)	1	0.822	97.9									
	R	PreCG (4)	2	0.941	101.4									
8	L	Anterior bank PreCS (4/6)	1	0.898	25.4	0.349	0.209	0.139	0.089					
9	R	FG (20)	1	1.659	106.6	0.094	0.028	−0.084	−0.223					
10	–	–	–	–	–	−0.126	−0.172	−0.221	−0.245					
11	L	Anterior bank PoCS (3)	1	2.589	36.5	0.602	0.317	0.237	0.250					
12	L	PreCS (6)	1	1.222	24.3	0.105	0.068	0.090	−0.017					
13	R	PreCG (6)	3	6.644	88.8	0.001	−0.196	−0.297	−0.119					
	R	SFG (6)	1	2.047	95.3									
14	L	Cuneus (18/19)	1	1.501	57.7	0.198	0.145	0.186	0.208					
15	R	Cerebellum	1	1.339	75.9	0.507	0.358	0.297	0.311					
	R	MFG (9/46)	1	0.849	78.0									
	R	IPS/SMG (40)	1	0.878	63.7									
	L	IPS (40)	1	1.390	81.1									
	L	SPL (7)	4	1.472	66.1									
16	–	–	–	–	–	0.209	−0.010	−0.036	−0.056					
17	–	–	–	–	–	0.121	−0.067	−0.226	−0.235					
18	–	–	–	–	–	1.067	0.698	0.077	0.159					
19	–	–	–	–	–	−0.007	−0.037	−0.039	0.032					
20	L	PreCS (4/6)	1	5.085	32.2	0.031	−0.096	−0.156	−0.199					
21	L	PreCG/CS (4/6)	2	2.075	13.8	−0.039	−0.315	0.076	0.075					
	L	CS	2	2.459	16.0									
	R	STS (22)	1	1.260	104.4									
	R	IPS (40)	1	1.035	106.7									
22	–	–	–	–	–	−0.069	−0.091	−0.114	−0.118					
23	–	–	–	–	–	−0.213	−0.186	−0.041	−0.054					
24	L	FG (20)	1	1.387	85.3	−0.163	0.281	0.207	0.113					
	L	Occipital Gyrus (18)	1	1.065	72.2									
	R	SFG (6/8)	2	1.082	65.6									
Mean (%)				2.106%	62.74 mm	0.290%	0.143%	0.064%	0.001%					
SD				1.341%	28.83 mm	0.470%	0.329%	0.220%	0.158%					
95% CI				0.637%		0.188%	0.132%	0.089%	0.063%					

(Worsley and Friston, 1995) that were used to analyze the data from the working memory scans. The principle of the fMRI time series analysis was to model the signal change associated with each discrete epoch of the trial with covariates comprised of BOLD HRFs positioned along the timeline of the task in order to best model the trial epoch in question (Postle et al., 2000; Zarahn et al., 1997a). Because the smoothness of the fMRI response to neural activity allows fMRI-evoked responses that arise from temporally dependent events to be resolved on the order of 4 s (Zarahn et al., 1997a), delay-period covariates were positioned in the middle of the delay (i.e., 4 s after the offset of the Instruction). Although fundamentally conservative, this approach ensured that estimates of delay-period activity would not be contaminated by variance attributable to the stimulus-presentation epochs that preceded the delay period. (Because of the temporal dependencies of the delayed-recognition task, such contamination would necessarily occur if we were to model the delay period with, say, a boxcar covariate that spanned the entire 8 s of the delay period (Postle, 2005).) The least-squares solution of the GLM of the fMRI data yielded parameter estimates associated with each covariate of interest, and the parameter estimates were scaled such that effect sizes were expressed as mean percentage change from baseline. (Baseline was defined as unmodeled portions of the trial plus the intertrial interval.) Load-sensitive voxels were identified with the contrast $[\text{Delay}_{\text{Forward } 5} - \text{Delay}_{\text{Forward } 2}]$.

SS analyses. Load-sensitive voxels were identified in a three-step process. First, we identified voxels showing a main effect of delay-period activity with the contrast $[\text{Delay}_{\text{Forward } 5} + \text{Delay}_{\text{Forward } 2}]$, thresholded at a mapwise α of 0.05 (Bonferroni corrected for multiple comparisons). From this main effect map, a binary mask was created in which suprathreshold voxels were given a value of 1 and subthreshold voxels were given a value of 0. Second, we identified delay-period load-sensitive voxels with the orthogonal contrast $[\text{Delay}_{\text{Forward } 5} - \text{Delay}_{\text{Forward } 2}]$, and the resultant whole-brain load map was masked with the mask from step 1 and thresholded to an α of 0.05 (again, Bonferroni corrected, this time for the number of voxels in the mask). Third, for each voxel or cluster of voxels that survived step 2, we inspected the trial-averaged time series from *Forward 5* and *Forward 2* trials to verify that they displayed trial-related activity. “Trial related” was operationalized as a positive delay effect for *Forward 5* trials and a positive effect for at least one epoch of *Forward 2* trials. (This would rule out, for example, voxels showing no *Forward 5* delay-period activity, but negative *Forward 2* delay-period activity). A load effect was determined for each SS data set by collapsing across all load-sensitive voxels. These values were averaged across SS data sets to compute the group-averaged load effect from the SS analysis.

SNGA analysis. (i) the GLM (as described in the SS analysis procedure) was solved for each SS data set; (ii) the contrast $[\text{Delay}_{\text{Forward } 5} - \text{Delay}_{\text{Forward } 2}]$ was applied to each SS data set, and the resultant volumes of parameter estimates were warped to a template in M.N.I. space using the FLIRT routine in the FSL analysis package, then entered into a second-level GLM that treated subject as a random effect. This process was repeated two more times with exogenous spatial smoothing with Gaussian kernels of 5 mm and 8 mm FWHM applied to the data prior to the solution of the second-level GLM. These three levels of smoothness were chosen because we did not know, a priori, the spatial dispersion of delay period load effects. The load map computed from the second-level GLM was thresholded at $p \leq 0.001$ (uncorrected for multiple

comparisons). Confirmation of anatomical regions was done via conversion of M.N.I. coordinates to Talairach coordinates with the *mn2tal* Matlab routine of Matthew Brett (<http://www.mrc-cbu.cam.ac.uk/Imaging/mnispace.html>), and then consultation of the atlas of Talairach and Tournoux (Talairach and Tournoux, 1988). The SNGA load effect sizes were determined by collapsing across all voxels within each PFC ROI demonstrating a reliable effect and averaging across the resultant effect sizes from each subject.

If, as predicted, the SNGA group analysis produced a reliable load effect in PFC, we planned to further evaluate the results produced by the two analysis approaches by “tracing backward” to the data of each SS data set to identify the voxels that had contributed to the effect identified in the SNGA analysis. We would do this by converting the SNGA load-effect map into a binary mask in which supra-threshold voxels from the PFC would be given a value of 1 and all other voxels a value of 0, then applying this mask to the spatially normalized data from each SS data set. The results from this procedure would be evaluated in two ways: (i) the masked statistical map would be thresholded (two-tailed) at the critical t -value equal to that determined for the $[\text{Delay}_{\text{Forward } 5} + \text{Delay}_{\text{Forward } 2}]$ map, and the number of suprathreshold voxels tabulated; (ii) the threshold would be lowered to zero (two-tailed) and the load effect, collapsed across all the voxels in the SNGA-derived ROI, calculated. Additionally, to allow for some non-overlap between the activation peaks from the SNGA-derived load mask and SS data sets, the volume of the SNGA-derived load mask would be grown over a range of sizes and the load effect extracted from each iteration of the ROI.

Results

Behavioral

The behavioral results (Table 1) indicated that accuracy for the retention of 5 letters was lower, and RTs longer, than for *Forward 2* trials.

fMRI

SS analyses

Load-sensitive voxels were identified in 17 subjects and there was considerable variability in the topographical localization of these effects across subjects (Table 2, Fig. 1). Load sensitivity was evident in the left posterior perisylvian cortex (i.e. the area including the supramarginal, angular and superior temporal gyri) in 3 subjects, in right posterior perisylvian cortex in 5 subjects, in the left SPL in 4 subjects, and in the right SPL in 1 subject, in the anterior cingulate cortex in 2 subjects (1 in left hemisphere, 1 bilateral), in the region from the precentral sulcus to the postcentral sulcus in 10 subjects (8 of these were in the left hemisphere, 1 in the right hemisphere and 1 bilateral), in 2 subjects in dorsolateral PFC (Brodmann area (BA) 9/46; one of these was bilateral and the other was in the right hemisphere); in the right frontal eye fields in 1 subject, in the inferior temporal gyrus in 3 subjects (2 were in the right hemisphere), in the left thalamus in 2 subjects, in the left occipital lobe in 3 subjects, in the right fusiform gyrus in 2 subjects and in the right cerebellum in 2 subjects.

The mean load effect size from the 17 subjects showing an effect was 2.106% (95% confidence interval (CI)=0.637). Including all subjects and using values of 0 for the seven SS data

sets that did not produce a supra-threshold load effect, the mean load effect was 1.492% (95% CI=0.594).

SNGA group analyses

The load-sensitivity results produced by the SNGA based analyses performed on the unsmoothed data were consistent with our hypothesis, revealing a significant load effect in 3 contiguous voxels in left frontal cortex, at the boundary of premotor, dorsolateral PFC, and ventrolateral PFC (i.e., the boundary of BAs 6 and 9; M.N.I. coordinates $-52.2, -5.4, 43$). The maximum *t*-statistic for these voxels was 3.901. Individual voxels outside of the PFC also displayed delay-period load-sensitivity, including at the border of the left central sulcus and postcentral gyrus and in the left precentral gyrus, left and right thalamus, left middle occipital gyrus, left inferior temporal gyrus and left and right cerebellum. The mean load effect size of the significant PFC load-sensitive voxels was 0.290% (95% CI=0.188). The same PFC foci of delay-period load sensitivity were observed at the 5 mm level of exogenous smoothing (and none were observed at the 8 mm level) so all further consideration of the SNGA analyses will be drawn from the data to which no exogenous smoothing was applied.

Having thusly confirmed our prediction that SNGA analysis results would differ from the SS analysis results in that only the former would produce reliable load sensitivity in PFC, we proceeded with our plan to evaluate the contribution of each SS data set to the SNGA result. First we applied the SNGA mask to the normalized statistical map from each SS data set, and then set the threshold to the same *t*-value that had been used to identify load sensitive voxels in the SS analysis (i.e., at $\alpha=0.05$ within the $[\text{Delay}_{\text{Forward } 5} + \text{Delay}_{\text{Forward } 2}]$ mask). At this threshold, we did not identify any supra-threshold voxels in any SS data sets in the SNGA PFC ROI. We next set the threshold to 0 (two-tailed) and extracted, for each subject, trial-averaged time series data and delay effects collapsed across the SNGA-derived load-sensitive ROI. Qualitatively, visual inspection of the trial-averaged activity from the SNGA PFC ROIs revealed that it was, for all but 3 SS data sets (2, 4 and 15), very noisy, demonstrating very little evidence of task-related structure (Fig. 1). Nonetheless, the group average of these time series data clearly contained task-related structure (Fig. 2). Finally, we grew the SNGA PFC ROI to volumes of 7, 50 and 100 voxels and extracted the load effect from each. The mean load effect size was progressively smaller at each larger ROI, and not different from zero at volumes of 50 and 100 voxels (Table 2).

Discussion

The results from Experiment 1 were consistent with our prediction that SS and SNGA group analyses of the same data sets would produce qualitatively and quantitatively different results. The qualitative difference is that whereas the SNGA analysis produced a reliable load effect in left lateral PFC (at the border of BAs 6 and 9), the SS analysis did not. Instead, the SS analyses produced load effects that were topographically heterogeneous across subjects, with the majority occurring in sensorimotor and posterior temporo-parietal regions. Only 2 subjects showed dorsolateral PFC load-sensitivity and in neither of these was load-sensitivity exclusive to this region. The quantitative difference was that the load effect aggregated across subjects in the SS analysis was, at 2.106% signal change, an

order of magnitude greater than that produced by the SNGA analysis (0.290% signal change).³ These results are therefore consistent with the idea that choice of analysis method influences whether or not a data set will yield evidence of sensitivity to working memory load in the PFC.

What we observed in Experiment 1 is that the anatomical location of the large, suprathreshold load effects observed in the data of most SS data sets was sufficiently variable across subjects that it was not detected in the SNGA analysis. (Whether this variability is best classified as “true” or as “noise” will be taken up in Experiment 2.) Thus, it is important to note that it was not simply the case that the “peaks” of the SS effects were non-overlapping, and so the SNGA analysis identified the locus of intersection of many “tails” of the SS effects.⁴ Rather, as is illustrated in Fig. 1, the SNGA analysis identified a region that, in the SS data sets, was not only subthreshold, but also clearly distinct from the region(s) showing anatomically suprathreshold effects.

The SNGA load effect in PFC was, by definition, reliable across subjects, but did it reflect a working memory storage function? It is difficult to conclude this for the majority of subjects when one inspects Fig. 1. With the exception of subjects 2, 4 and 15, the activity in these voxels does not appear to be task related. (Note that this does not change when the SNGA PFC ROI is grown to accommodate some topographical imprecision of the ROI). This question will be taken up again in General discussion.

Turning now to the SS group analysis, its implementation ensured that the regions selected would demonstrate activity that appeared to be task related. But what are we to make of the topographic variability of load effects across subjects? One possibility is that the variability reflects an inherent “instability” of the effect, and thereby calls into question the interpretability of data from any single subject. (i.e., data from any SS are expected to be statistically “noisy”.) A second possibility is that these results reflect true interindividual variability in the functional neuroanatomy of working memory storage processes. One way to begin evaluating these two possibilities is to evaluate empirically the test–retest reliability of SS statistical maps. This was undertaken in Experiment 2.

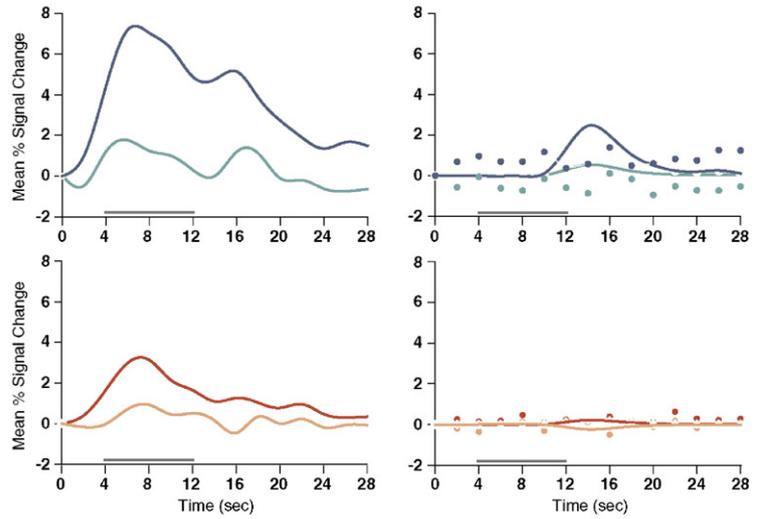
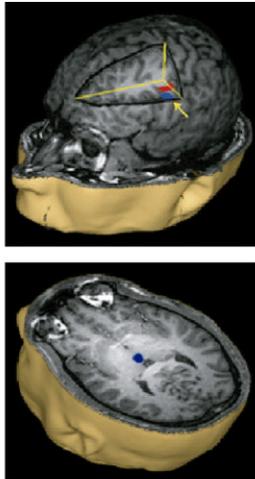
Experiment 2

Is a large portion of the topographical heterogeneity of the load effect identified by the SS analysis due to measurement error, and thus best considered statistical noise (Narayanan et al., 2005; Veltman et al., 2003)? We evaluated this possibility by measuring empirically the scan–rescan reliability of topographical patterns of activity produced by the SS analysis procedure. Six new subjects underwent the entire procedure from Experiment 1 on two separate occasions (Session 1 and Session 2). Statistical maps from the two sessions were then compared to determine the precise voxel-to-voxel overlap in thresholded maps across sessions, as well as the

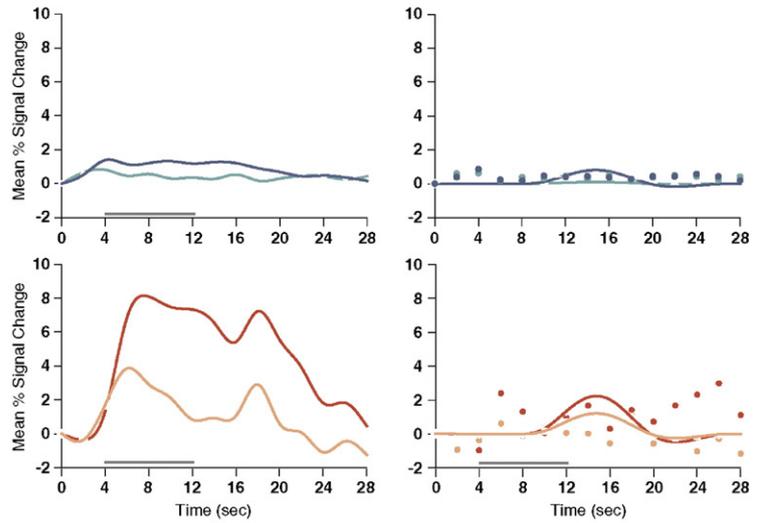
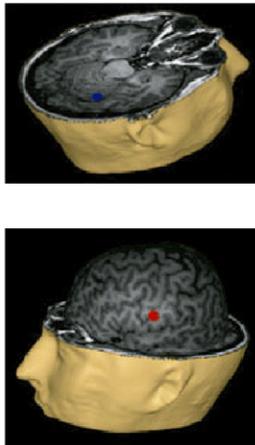
³ A caveat when interpreting these differences in the effect size is that the aggregated SS value was computed by averaging across all subjects regardless of where the hotspots were spatially located. This value therefore assumes that all of these brain regions have the same neural effect size, resulting in comparable BOLD signal changes.

⁴ Note that in SNGA analyses of functions that are expected to be more topographically uniform across subjects, such as motion perception or oculomotor control (Swallow et al., 2003), the quantitative difference in effect sizes between SNGA results and SS results can often be attributed to this “intersection of tails” phenomenon.

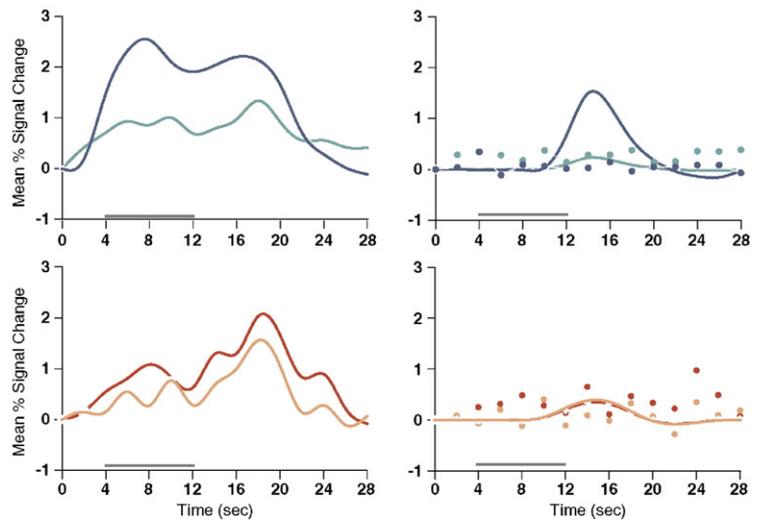
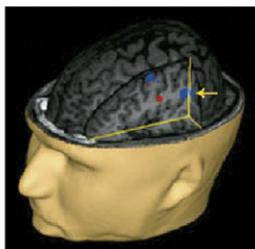
Subject 1



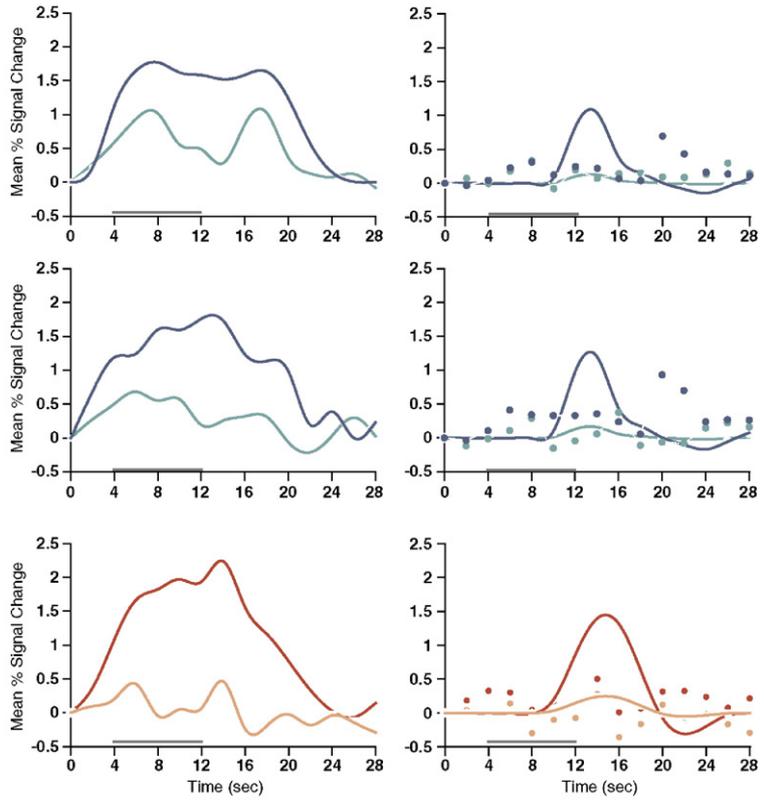
Subject 2



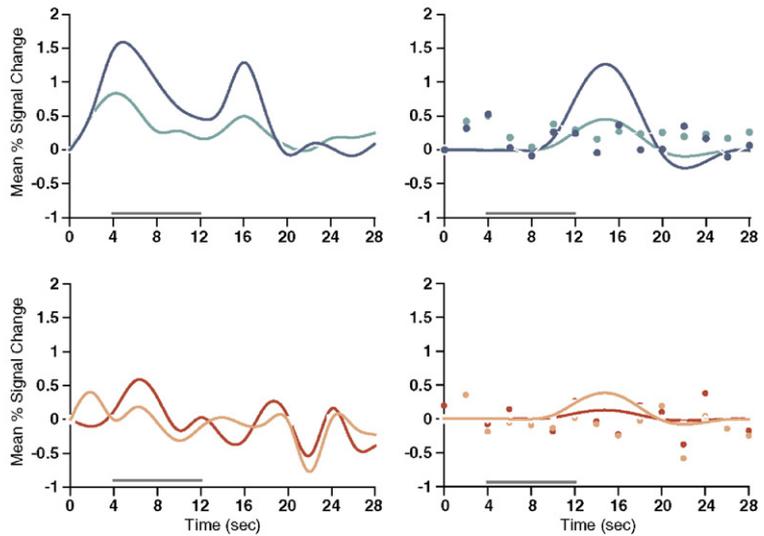
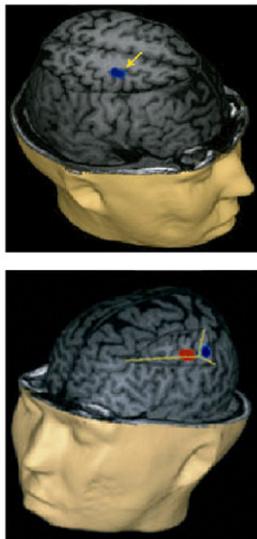
Subject 3



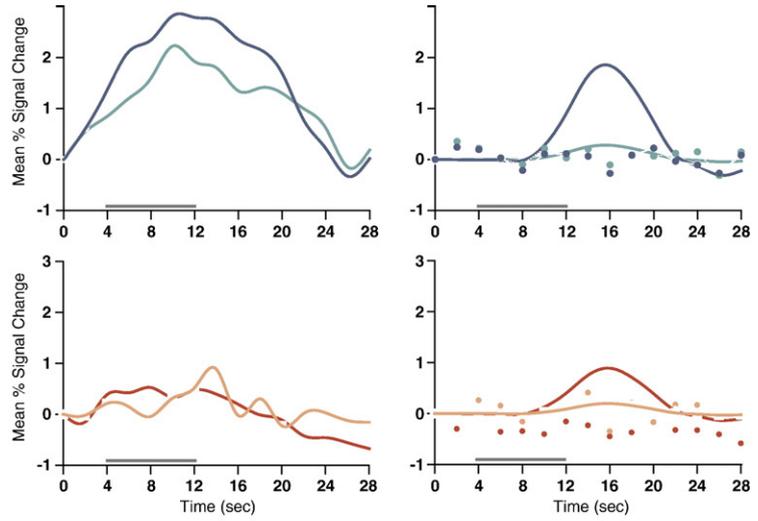
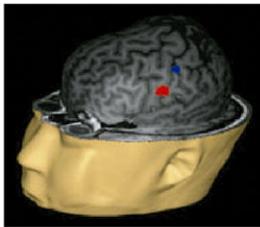
Subject 4



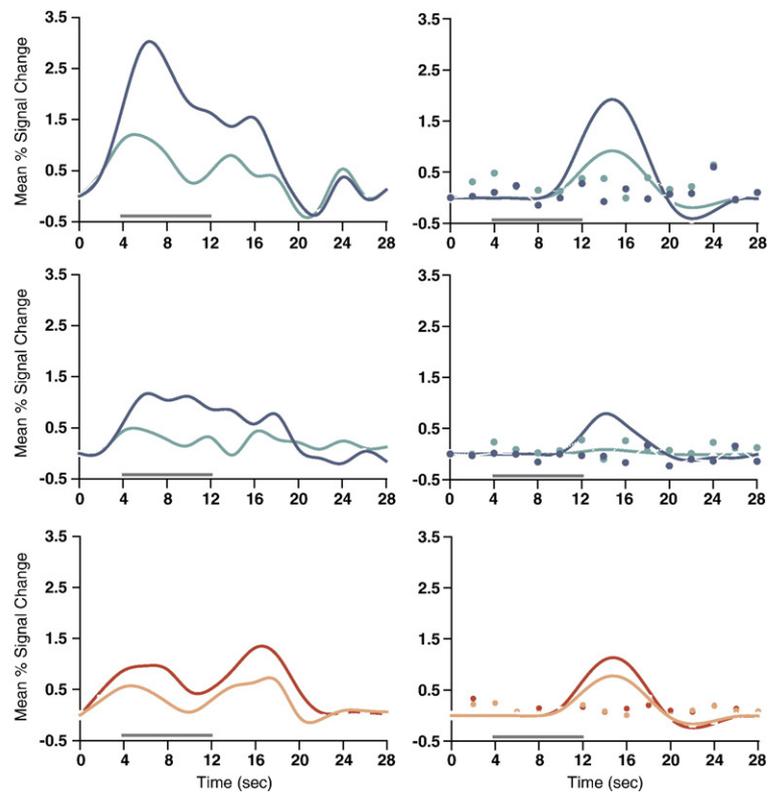
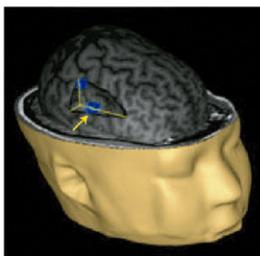
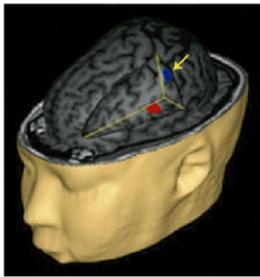
Subject 5



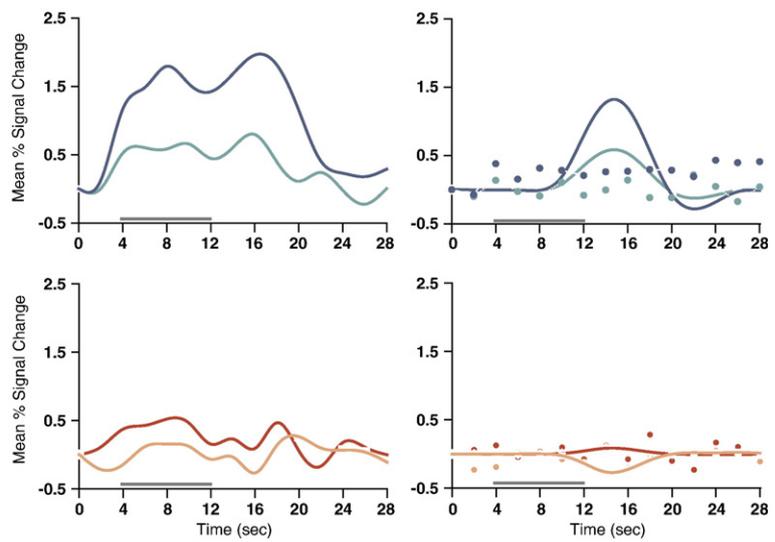
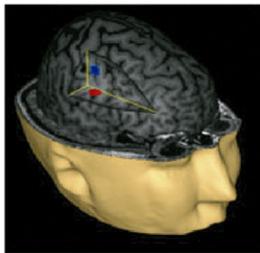
Subject 6



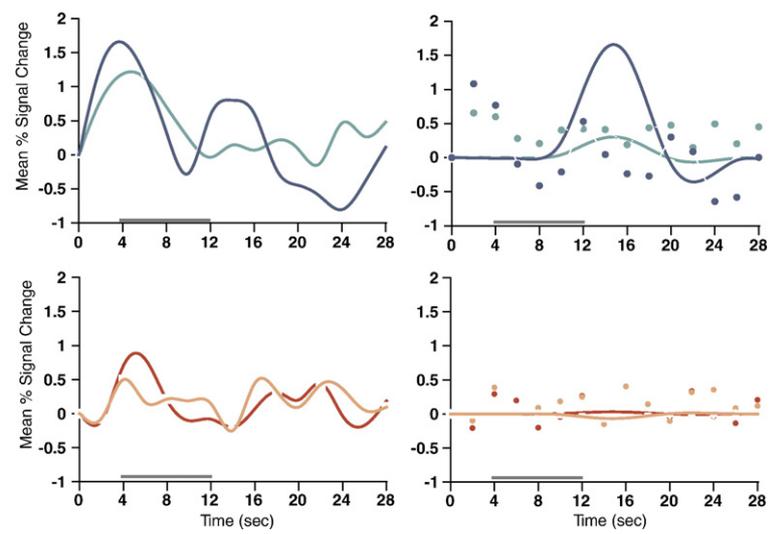
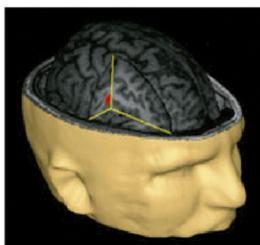
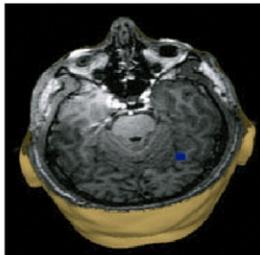
Subject 7



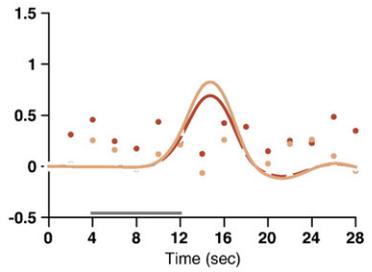
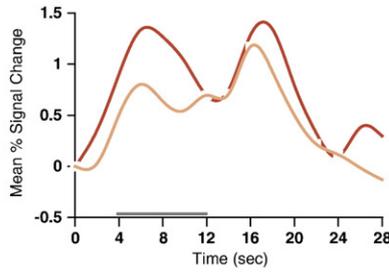
Subject 8



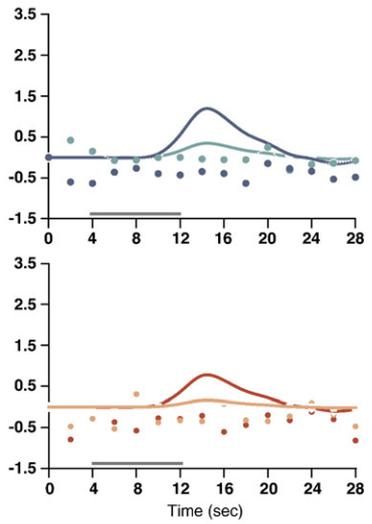
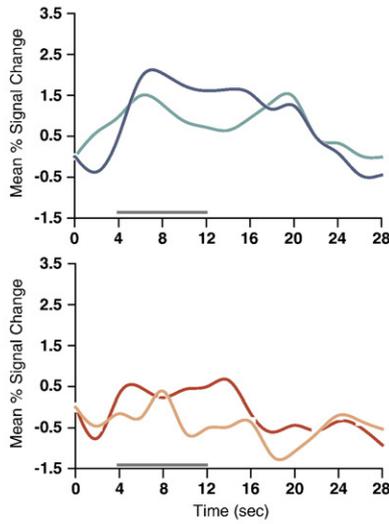
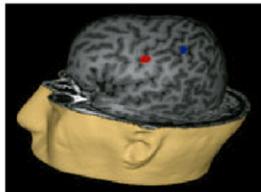
Subject 9



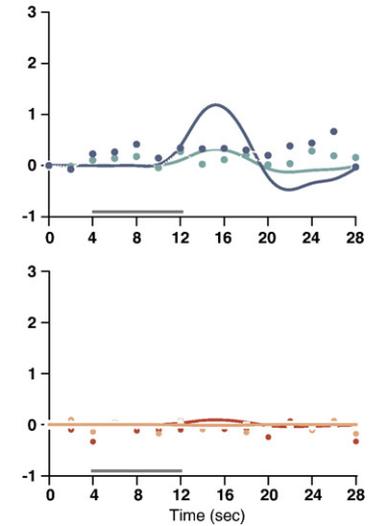
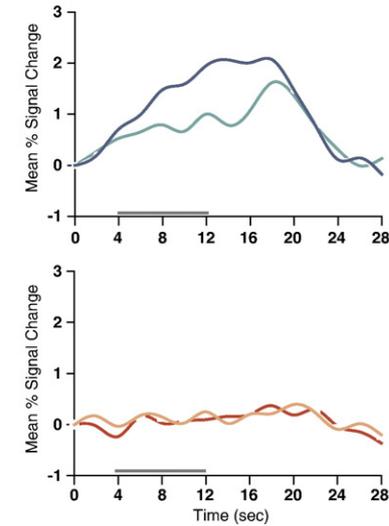
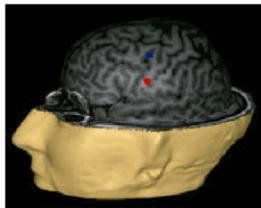
Subject 10



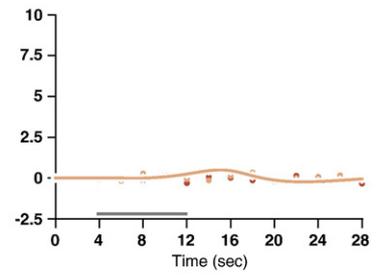
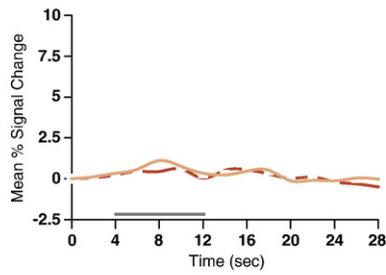
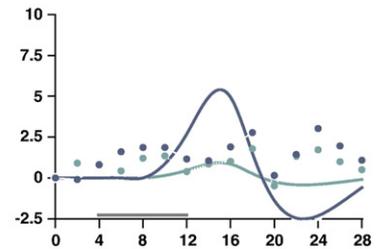
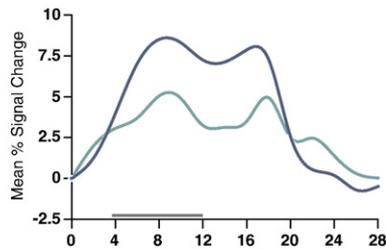
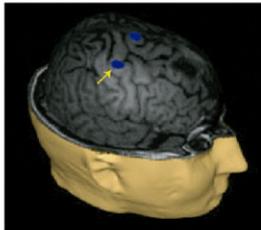
Subject 11



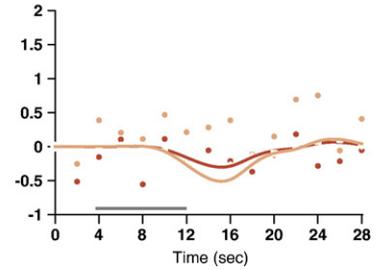
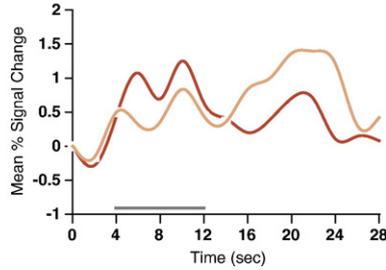
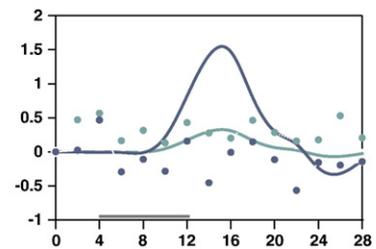
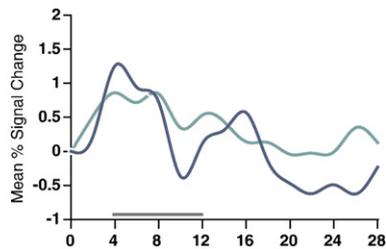
Subject 12



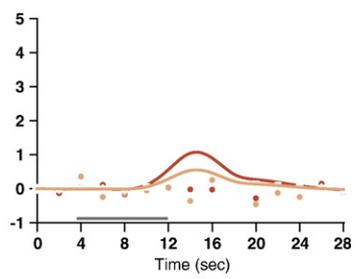
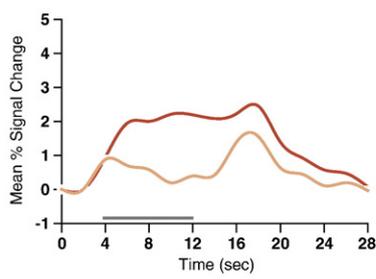
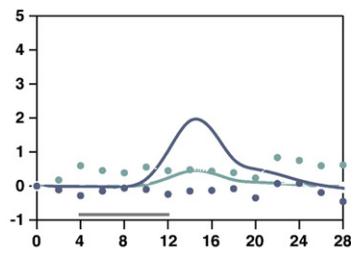
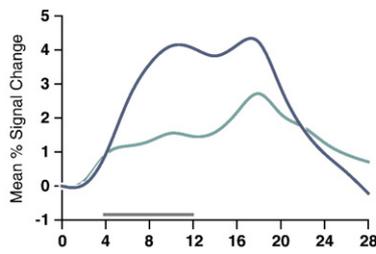
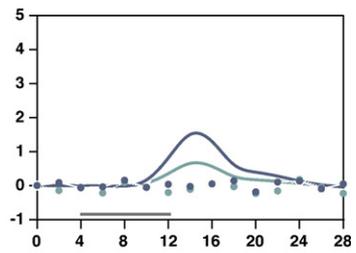
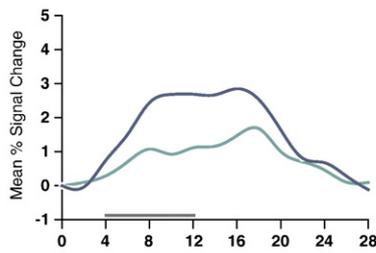
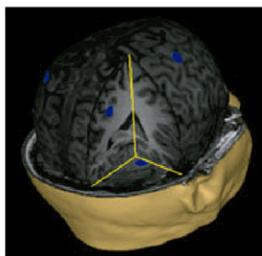
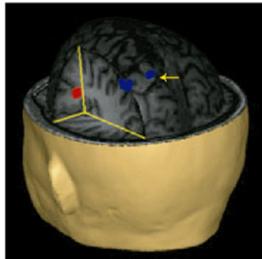
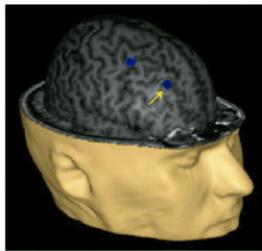
Subject 13



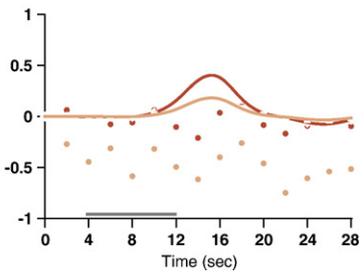
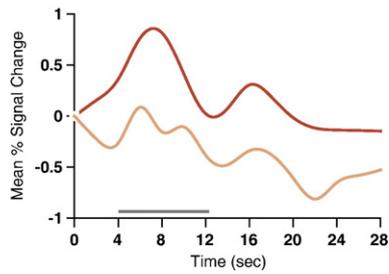
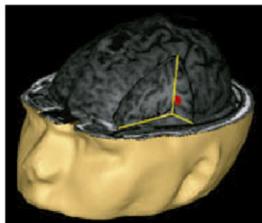
Subject 14



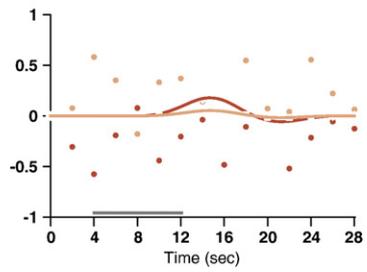
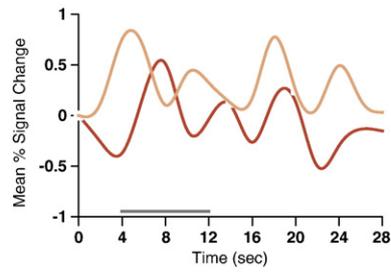
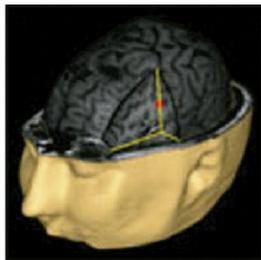
Subject 15



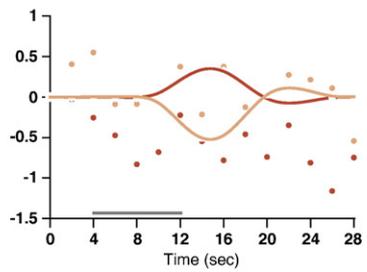
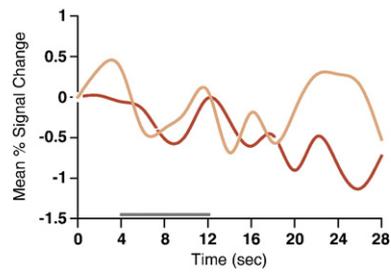
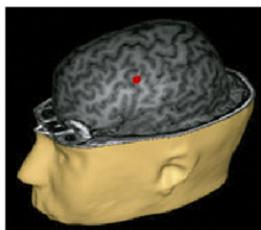
Subject 16



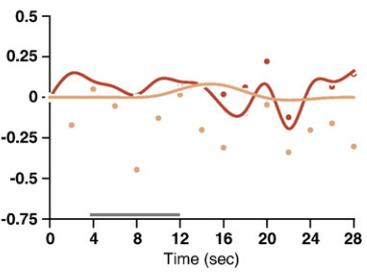
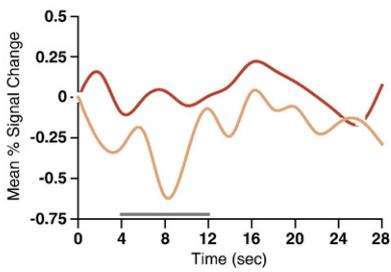
Subject 17



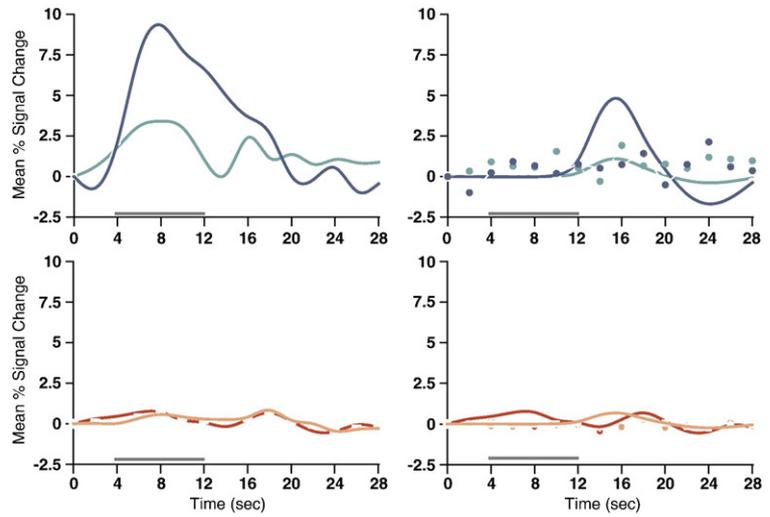
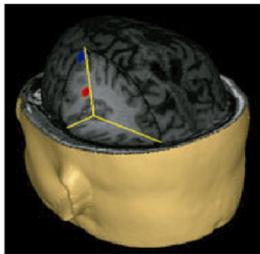
Subject 18



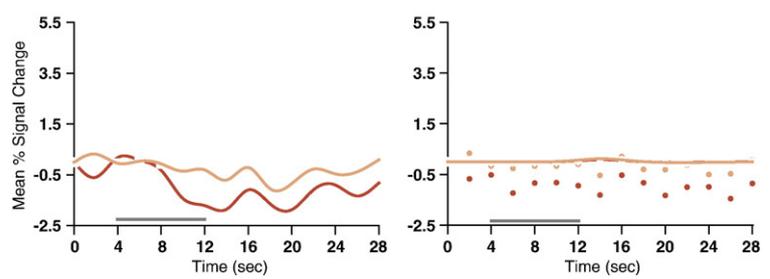
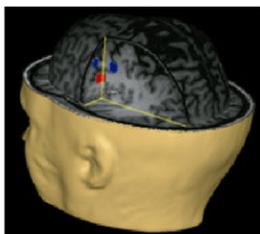
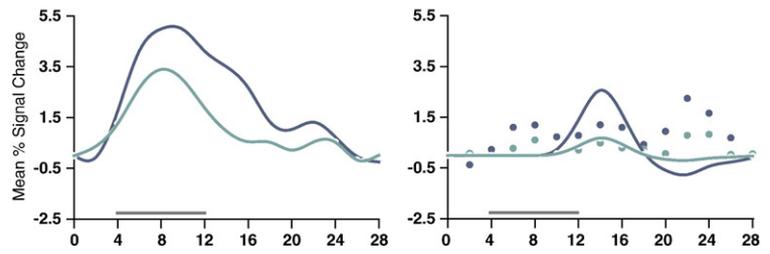
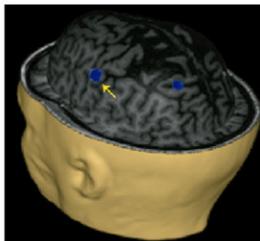
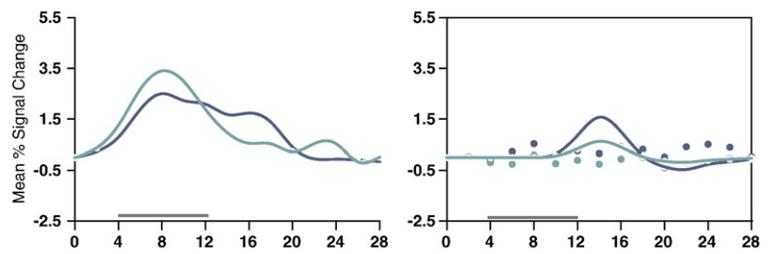
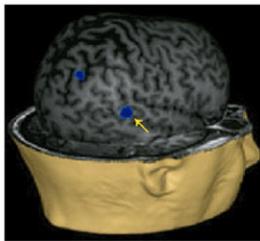
Subject 19



Subject 20



Subject 21



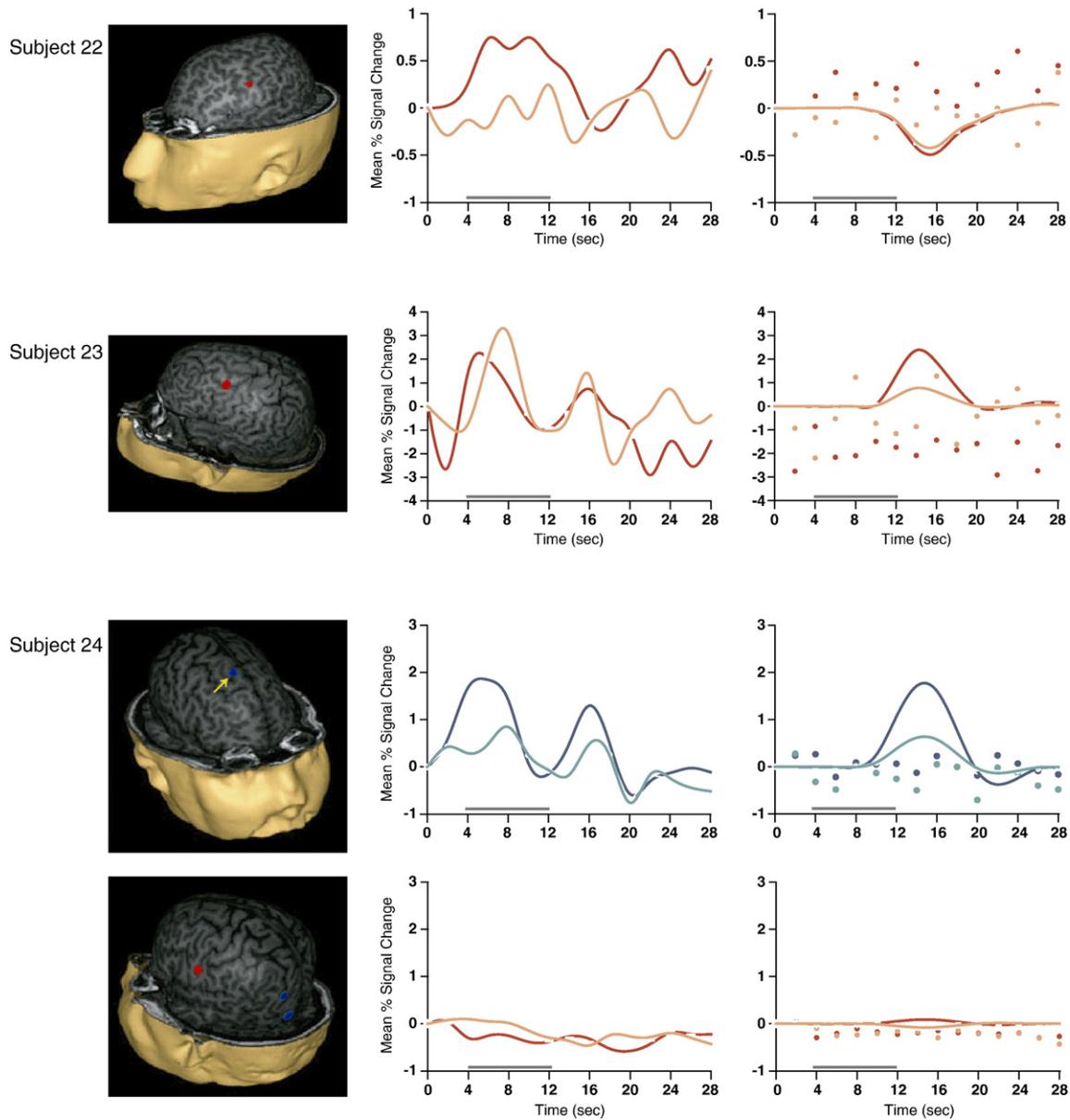


Fig. 1. SS and SNGA results from Experiment 1. Regions with load sensitivity detected with SS analysis are shown in blue, along with, plots of the trial-averaged time series extracted from these load-sensitive regions (left column; blue=*Forward 5*; green=*Forward 2*), and corresponding relevant information from the solution of the GLM (delay-period covariates scaled by their parameter estimates, and residual error at each time point, right column). The same is shown for the SNGA-derived load regions masks (red brain regions, and corresponding times series and statistical data; red=*Forward 5*; orange=*Forward 2*). Note that in the plots of delay effects calculated from SNGA-derived load regions, the maximum excursion of the $\text{Delay}_{\text{Forward 5}}$ covariates are often not greater than the residual error term from the GLM at the same time point. For clarity, the Euclidean distances between SS and SNGA-derived regions are reported in Table 2.

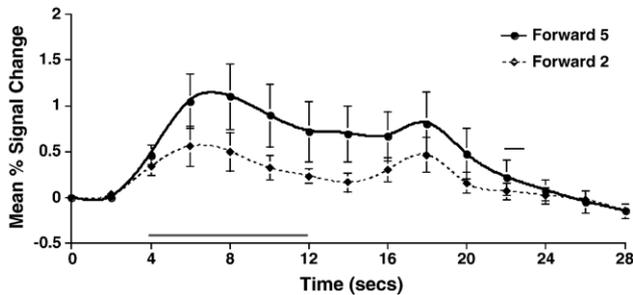


Fig. 2. Group average time series for *Forward 5* and *Forward 2* trials extracted from the SNGA PFC ROI.

intersession similarity in trends in functionally defined regions. Empirical estimates of the amount of overlap expected by chance were obtained by arbitrary pairing of SS data sets from Experiment 1.

Methods

Subjects

Six right-handed subjects who did not participate in Experiment 1 (mean age = 25.83 years [SD = 3.92], 3 males) were recruited from the University of Wisconsin-Madison community. All subjects provided informed consent and were paid for their participation.

Procedure

Experiment 2 used the same behavioral and fMRI procedures as Experiment 1, with the difference that each subject went through the entire procedure twice, on two separate days (median number of days separating Session 1 from Session 2 = 7; range = 1–17). The second session differed from the first only in that it featured a different set of stimuli; the order of stimulus-set presentation was counterbalanced across subjects.

Analyses

Data from the two sessions were first analyzed separately with the methods used for SS analyses in Experiment 1. We then calculated two indices of overlap, one measuring the proportion of (suprathreshold) voxels identified by a contrast in one session that were also identified by the same contrast applied to the data from the second session, and a second indexing the intersession stability of activity within functionally defined voxels. In each case, these indices were calculated for four statistical contrasts. Three of these involved working memory: [$\text{Delay}_{\text{Forward 5}} - \text{Delay}_{\text{Forward 2}}$], [$\text{Delay}_{\text{Alphabetize 5}} - \text{Delay}_{\text{Forward 5}}$], and [$\text{Delay}_{\text{Forward 5}} - \text{baseline}$].⁵ The fourth identified activity in a simple sensorimotor task, the HRF derivation scan, with

⁵ We added this contrast in the event that any of the SS data sets from Experiment 2 did not show suprathreshold load effects. The relative sparseness of suprathreshold voxels for the load contrast is not surprising when one considers that there was no “load” task being performed by subjects in the scanner. Rather, this contrast isolated the difference between two trial epochs that would each be expected, a priori, to feature relatively low levels of task-evoked signal, because each corresponded to a portion of the task with neither sensory nor motoric elements. Although not specific to load, the [$\text{Delay}_{\text{Forward 5}} - \text{baseline}$] contrast was expected to produce more robust statistical maps.

the contrast [button-press cue–baseline]. This task, known to produce robust activation, was also assumed to be less susceptible to strategic, attentional, and learning-related factors than might be the working memory-related measures. Thus, we reasoned that it would provide an empirical estimate of the highest overlap one might expect for both intersession and intersubject comparisons. Having generated these four statistical maps, we next compared the results from the two sessions by obtaining two measures of between-session reliability: overlap in thresholded maps and stability of activity within functionally defined voxels. These are detailed in the next section. To obtain estimates of chance-level values of these indices, we repeated these steps with SS data from Experiment 1, after randomly pairing subjects. This process produced measures of intersubject reliability. Finally, we evaluated quantitatively whether our estimates of intersession reliability (i.e., “intrasubject”) differed from what would be expected by chance, by statistically comparing them against the empirically derived estimates of intersubject reliability.

Overlap in thresholded maps

We begin by describing the procedure followed for the 6 new subjects participating in Experiment 2. The first step in this analysis was to generate whole brain-thresholded statistical maps for each of the four contrasts described above in Session 1 and Session 2 (load-sensitive maps were produced using the same method described in Experiment 1). Second, for each subject and for each contrast, the thresholded map containing the largest number of voxels was designated “Session A”, and the corresponding map with fewer voxels “Session B”. Third, the thresholded contrast maps from Session A were converted into binary masks with a value of 1 at each suprathreshold voxel and a value of 0 for all other voxels. Fourth, the Session A mask volumes were coregistered (using a rigid-body transformation implemented in SPAMALIZE software (http://brainimaging.waisman.wisc.edu/~oakes/spam/spam_frames.htm)) to their corresponding volumes from Session B. Fifth, the Session A masks were applied to the corresponding (whole-brain, thresholded) statistical maps of the same contrast from Session B to produce “Session A^B” maps. Finally, the thresholded overlap index (ω_t) was determined with the formula

$$\omega_t = \text{voxels}_{\text{Session A} \wedge \text{B}} / \text{voxels}_{\text{Session A}}$$

The same procedure was followed for subject pairs drawn from Experiment 1, with the exception that all analyses were performed in M.N.I. space, thereby precluding the need for coregistration between Session A and Session B subjects. Finally, the reliability of the intersession overlap indices for each of the four contrasts was determined by comparing, with Mann–Whitney *U*-tests, intersession vs. intersubject group trends in ω_t .

Overlap in patterns of activity

This measure, which we will call ω_u (*u* for “unthresholded”) was similar to ω_t , except that it did not include some of the bias associated with excluding all subthreshold voxels from the analysis. In essence, it evaluated quantitatively the topographical stability of the effect in question, by evaluating the extent to which the voxels identified in Session A could be said to be performing the same function in Session B, regardless of where their level of activity fell with respect to the significance threshold. It also differed from ω_t in that it was a continuous, rather than a discrete measure. It was obtained for each scan pairing (either intersession or intersubject) after ω_t was

Table 3
Summary of behavioral results from Experiment 2

Instruction	Accuracy (% correct)		RT for correct responses (ms)	
	Session 1	Session 2	Session 1	Session 2
Alphabetize	82 (11.3)	84.3 (7.8)	1823.80 (207.72)	1647.21 (236.90)
Forward 5	88.2 (10.3)	88.5 (7.2)	1777.38 (265.26)	1578.55 (245.28)
Forward 2	96 (1.6)	88.7 (6.4)	1349.70 (213.66)	1349.82 (297.45)

Numbers in parentheses indicate standard deviations.

calculated by simply lowering the threshold of the Session A^B map to 0, two-tailed, and extracting the estimated effect size for the contrast in question, then normalizing by the effect size from Session A. As was done with ω_t , the reliability of the intersession estimates of ω_u was assessed for each of the four contrasts with Mann–Whitney *U*-tests by comparing intersession vs. intersubject group trends.

In the scan–rescan data set, overlap and effect sizes for the load, alphabetization and [Delay_{Forward 5}–baseline] contrasts were also extracted from ROIs of theoretical import to working memory: middle frontal gyrus, superior frontal gyrus, inferior frontal gyrus, anterior cingulate cortex, supramarginal gyrus, angular gyrus, SPL and cerebellum. (These operations were performed at ROI-wise thresholds). Both measures of ω were determined separately for left and right hemispheres of each ROI.

Results

Behavior

Performance across test and retest sessions was equivalent for all trial types (Table 3) and was comparable to Experiment 1.

fMRI

Overlap in thresholded maps

For the intersession comparisons, ω_t was highest for the [Delay_{Forward 5}–baseline] contrast (15.2%), followed by

[Delay_{Alphabetize 5}–Delay_{Forward 5}] (14.9%), then [HRF button press–baseline] (10.3%), and finally [Delay_{Forward 5}–Delay_{Forward 2}] (6.5%). The analogous values were lower across the board for the intersubject comparisons: [Delay_{Forward 5}–baseline] (2.8%); [Delay_{Alphabetize 5}–Delay_{Forward 5}] (1.8%); [HRF button-press cue–baseline] (3%); [Delay_{Forward 5}–Delay_{Forward 2}] contrast (0%) (Fig. 3). Pairwise tests indicated that, for each contrast, intersession ω_t was significantly greater than intersubject ω_t (all significant by two-tailed Mann–Whitney *U*-tests). Table 4 reports these same values from the intersession data, broken down by ROI.

Overlap in patterns of activity

For the intersession comparisons, ω_u was highest for [HRF button-press cue–baseline] contrast (58%), followed by [Delay_{Forward 5}–baseline] (52.2%), then [Delay_{Alphabetize 5}–Delay_{Forward 5}] (44.9%), then [Delay_{Forward 5}–Delay_{Forward 2}] (17.9%). As was the case with ω_t , ω_u values were uniformly smaller in the intersubject data: [HRF button-press cue–baseline] contrast (27.5%); [Delay_{Forward 5}–baseline] (–.2%); [Delay_{Alphabetize 5}–Delay_{Forward 5}] (7.1%); [Delay_{Forward 5}–Delay_{Forward 2}] (–2.4%) (Fig. 3). For the latter three contrasts, values of ω_u were not different from 0, indicating, for example, that knowledge of the network of voxels that is active during the retention of 5 items for one subject is of no value in predicting which voxels will be sensitive to this same contrast in a different subject. The between group differences in ω_u values were determined to be reliable for all contrasts by two-tailed Mann–Whitney *U*-tests.

Discussion

By two measures of overlap, the results of Experiment 2 indicate that SS test–retest reliability of the topography of delay-period activity is significantly higher than what would be expected by chance, as estimated by the arbitrary pairing of data from different subjects. Most importantly for the empirical and theoretical issue motivating these experiments, this was true for the load effect. An implication of this result is that a considerable portion of the

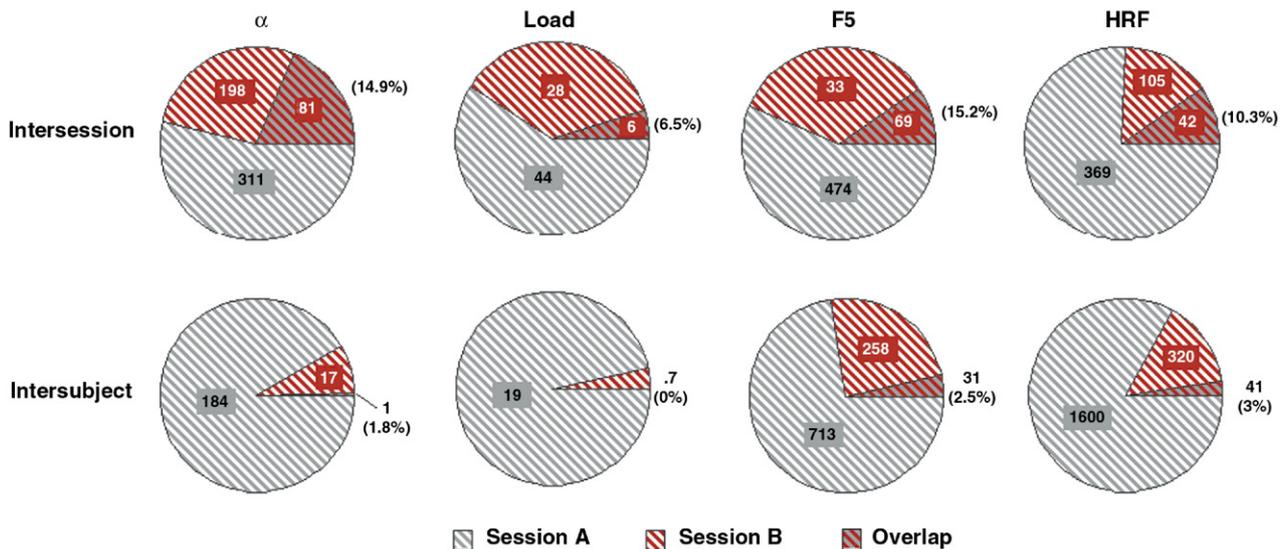


Fig. 3. Measure of overlap of thresholded maps from two separate scanning sessions for each of four contrasts. Top row=intersession data; bottom row=intersubject data. Values indicate the mean number of voxels significantly active at the whole-brain level. The overlapping portion of each chart indicates the proportion of voxels in Session B that were also identified from Session A.

Table 4
Intersession test–retest values from anatomical ROIs

ROI	Contrast					
	Alphabetization		Load		F5	
	ω_t (%)	ω_u (%)	ω_t (%)	ω_u (%)	ω_t (%)	ω_u (%)
lMFG	22.1 (28.1)	52.4 (33.9)	0	21.9 (17)	17.4 (12.4)	39.6 (46.4)
rMFG	35.4 (38)	43.1 (30)	0	26.3	12.6 (21.5)	56.9 (23.5)
lSPL	42.3 (26.8)	63.2 (29.3)	0	28.2	17.2 (31.1)	60.7 (23.3)
rSPL	11.3 (11.4)	68.7 (58.9)	0	34 (40.7)	0	16 (28.7)
lSFG	37.5 (12.6)	56.3 (37.1)	0	20.3 (4.5)	11.7 (15.6)	57.3 (28.5)
rSFG	41.7 (47.1)	48.1 (24.6)	0	30.2 (13.7)	29.7 (28.6)	51 (65.5)
lIFG	36.2 (28.8)	51.5 (11.9)	0	30.2 (5.3)	30.3 (15.9)	76.2 (18.1)
rIFG	18.5 (16.3)	53.4 (22.9)	14.3 (10.1)	47 (7.3)	8.3 (14.4)	78 (28.1)
lCingG	62.5 (53)	58 (31.6)	–	–	21.4 (20.2)	54.4 (26)
rCingG	38.3 (35.8)	62.2 (46.1)	0	22.7	13.5 (27)	50.2 (52.9)
lSMG	15.5 (16.2)	66.1 (17)	22.2 (11.1)	46.3	10.7 (15.2)	73.1 (23)
rSMG	19.3 (26)	50.6 (30.4)	12.5 (8.8)	94.3 (15.6)	5.9 (10.2)	60.4 (49.3)
lAG	54.4 (28.8)	51.1 (45.4)	0	–	0	24.7 (6.1)
rAG	18 (15)	87.1 (23.2)	0	15.9	29.2 (41.3)	21.8 (10.2)
lcerebellum	4.6 (6.4)	73.4 (34.7)	0	19.9 (36.8)	21.3	63.9 (7.9)
rcerebellum	41.1 (26.1)	50.3 (34.6)	0	36.4 (50)	2 (3.4)	51.4 (26.7)

between-subject variability of load sensitivity observed, for example, in Experiment 1, reflects true intersubject heterogeneity. That is, it cannot be dismissed as “noise,” or measurement error (Fig. 4).

Intersubject topographical variability, such as that described here, has also been described in neuroimaging studies of other domains of cognition. For example, in a previous systematic comparison of SS vs. SNGA analysis approaches, Miller and colleagues (2002) measured the correlation between fMRI signal from test and retest sessions of an episodic memory retrieval task. The highest degree of correlation was found between the two separate sessions from the same subjects. The correlation between subjects was found to be markedly lower. Furthermore, when the SNGA group analysis was performed, regions that emerged as significant were inconsistent with those seen in the SS analyses. Qualitatively similar results have been reported with visuo-perceptual and oculomotor tasks, in that the SNGA approach has been found to negatively affect the reliability of functional localization of the frontal eye fields and of area MT, compared to the SS approach (Swallow et al., 2003).

General discussion

The purpose of these experiments was to address an empirical impasse with theoretical implications. We did this by testing the hypothesis that different approaches to group analysis of fMRI data can produce different results. Experiment 1 demonstrated that, consistent with our hypothesis, whether one does or does not find reliable evidence of delay-period load sensitivity in PFC depends on the analysis method that one uses. This provides an answer to the empirical question of *why is there a discrepancy in the literature?* Somewhat paradoxically, however, the answer to this empirical question complicates the task of answering the underlying substantive question of *does the PFC contribute to the working memory storage of verbal information?* The reason for this is that, as this paper has made clear, the two types of analysis under scrutiny have different goals.

SS analyses may be best suited to address questions at the level of *how does the brain of this individual who is participating in my study support the cognitive process(es) in which I am interested?*

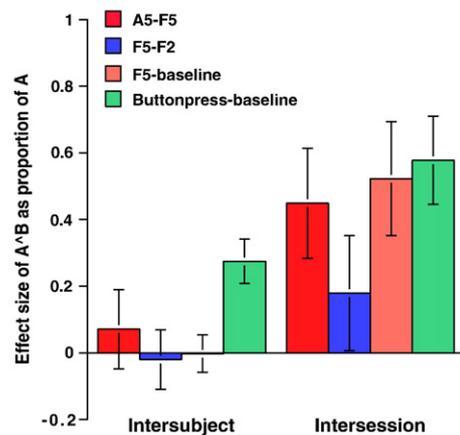
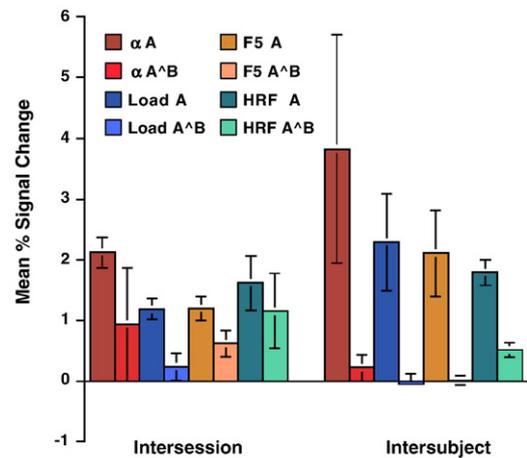


Fig. 4. Effect sizes (as mean % signal change) for each contrast-of-interest, extracted from suprathreshold voxels Session A, and from these same voxels from Session B (i.e., A^B). Error bars represent standard errors.

This follows from the fact that the time series data extracted from load-sensitive regions in the majority of SS data sets clearly reflected task-related activity. Another way to frame this is with the following thought experiment: Imagine one was performing a repetitive transcranial magnetic stimulation (rTMS) experiment with Subject 3 from Fig. 1 (or Subject 5, or Subject 9...), the goal of which was to disrupt delayed-recognition performance with rTMS delivered to a region of Subject 3's brain during the delay period. In such a situation we would surely choose to stimulate the load-sensitive region of superior temporal gyrus identified in the SS analysis of this individual's data rather than the SNGA analysis-identified region in frontal cortex. This is particularly true because the results of Experiment 2 suggest that at least a portion of the anatomical network engaged by this subject during the fMRI scan will also be engaged during the repeated performance of the task in the subsequent rTMS study. (A caveat about this reasoning, however, is that the predicted outcome of SS-targeted rTMS being more effective than SNGA-targeted rTMS would not necessarily tell us about storage. This is because, strictly, it is more correct to say that SS analyses tell us about how the brain of the subject in question supports the *behavior* in which the subject is engaging. In the present experiment, we do not know how much of the intersubject topographical variability is due to differences in strategy, rather than to differences in how individual brains accomplish storage in working memory.)

The goal of SNGA analyses, in contrast, is to extract from a sample of subjects a measure of structure–function relations that generalize to the population. One of its virtues is that it will “clean up” a data set by filtering out subject-specific effects that are due to e.g. different strategies, thereby isolating only those processes shared in common across subjects. To return to the thought experiment from the previous paragraph, what if the subject for this hypothetical rTMS experiment was someone new who we have not previously scanned? In this case one might choose to target the SNGA-defined region, because the logic of the SNGA analysis permits inference from the sample to others from the same population.

Because this paper makes the case that the SS approach provides useful information about the neural bases of working memory function, we will briefly consider some of the practical problems associated with it. Primary among these is how to deal with intersubject topographical variability. The results from Experiment 1 make clear that having access to load-sensitive maps from one or more subjects does not allow one to predict where in the brain one will observe load sensitivity in the next subject to be scanned, at least not with greater specificity than *somewhere in the left or right hemisphere bounded rostrally by the precentral sulcus and caudally by the intraparietal sulcus*. With such poor predictive ability, it is understandable that interpretation of SS analyses of working memory data can give the impression of a *post hoc* exercise. One solution is to adopt a criterion of “task relatedness”, as we did in Experiment 1. That is, we did not simply accept every suprathreshold “activation” in the load sensitivity map as being storage related, but, rather, we carefully inspected the trial-averaged time series of each, only accepting those demonstrating a positive delay effect for *Forward 5* trials, a positive effect for at least one epoch of *Forward 2* trials, and the absence of a large negative delay effect in *Forward 2* trials. A second, which could be combined with the first, could be to perform an initial localizer scan. This would be analogous to the way that localizer scans can be used to predefine category-sensitive ROIs in visual cortex, for example, and then using these ROIs in the analysis of

visual working memory data (e.g., Ranganath et al., 2004). For some particularly subtle effects, such as load sensitivity, an entire “localizer session” might be required before the subsequent hypothesis-testing session could be performed. It may be that one or more of the relatively recently developed multivariate analysis techniques (e.g., Lin et al., 2003; Polyn et al., 2005; Rissman et al., 2004; Woodward et al., 2006) will be particularly useful in this regard. Such an approach might be expected to provide more sensitive, and more complete, measures of the topography of, for example, the brain systems that are sensitive to manipulations of load.

In conclusion, although we identified why different groups obtain different results with respect to the role of the PFC in working memory, we did not make a strong case that one or the other sets of results is “correct.” What we can do, however, is summarize some questions that need to be resolved. For the SS analyses, would an experimental procedure that more tightly controlled the mental processes that subjects engage have an effect on the degree intersubject variability that we observed in this study? Orthogonal to this, would a study that included independent measures of individual differences be able to account for some of this variability? For the SNGA analysis, what processes are reflected by the left PFC region that it identified? What is one to make of the lack of obvious “task relatedness” in the time series data pulled from SNGA-identified region of the data of individual subjects? Can confounding factors (e.g., difficulty) be ruled out? For both analyses, how do the regions that they identify as being load sensitive respond to other experimental manipulations that target storage, such as duration of the delay period, stimulus domain, or secondary task interference? Finally, what will be the effect of rTMS applied to the regions identified by each method?

Acknowledgments

The authors would like to thank Charan Ranganath, Rik Henson and Tom Johnstone for helpful discussions of these ideas, and Andrew Nick and Jarrod Lewis-Peacock for experimental assistance. Supported by NIH MH064498 (B.R.P.).

References

- Aguirre, G.K., Zarahn, E., D'Esposito, M., 1998a. An area within human ventral cortex sensitive to “building” stimuli: evidence and implications. *Neuron* 21, 373–383.
- Aguirre, G.K., Zarahn, E., D'Esposito, M., 1998b. The variability of human, BOLD hemodynamic responses. *NeuroImage* 8, 360–369.
- Baddeley, A.D., 1986. *Working Memory*. Oxford Univ. Press, London.
- Baddeley, A.D., 2000. The episodic buffer: a new component of working memory? *Trends Cogn. Sci.* 4, 417–423.
- Baddeley, A.D., Hitch, G.J., 1974. Working memory. In: Bower, G.H. (Ed.), *The Psychology of Learning and Motivation*. Academic Press, New York, pp. 47–89.
- Baddeley, A.D., Logie, R.H., 1999. Working memory: the multiple-component model. In: Miyake, A., Shah, P. (Eds.), *Models of Working Memory*. Cambridge Univ. Press, Cambridge, UK, pp. 28–61.
- Bor, D., Duncan, J., Wiseman, R.J., Owen, A.M., 2003. Encoding strategies dissociate prefrontal activity from working memory demand. *Neuron* 37, 361–367.
- 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.
- Buckner, R.L., Petersen, S.E., Ojemann, J.G., Miezen, F.M., Squire, L.R.,

- Raichle, M.E., 1995. Functional anatomical studies of explicit and implicit memory retrieval tasks. *J. Neurosci.* 15, 12–29.
- Christoff, K., Prabhakaran, V., Dorfman, J., Zhao, Z., Kroger, J.K., Holyoak, K.J., Gabrieli, J.D.E., 2001. Rostrolateral prefrontal cortex involvement in relational integration during reasoning. *NeuroImage* 14, 1136–1149.
- Corbetta, M., Miezin, F.M., Dobmeyer, S., Shulman, G.L., Petersen, S.E., 1990. Attentional modulation of neural processing of shape, color, and velocity in humans. *Science* 248, 1556–1559.
- Courtney, S.M., 2004. Attention and cognitive control as emergent properties of information representation in working memory. *Cogn. Affect. Behav. Neurosci.* 4, 501–516.
- Courtney, S.M., Ungerleider, L.G., Keil, K., Haxby, J.V., 1996. Object and spatial visual working memory activate separate neural systems in human cortex. *Cereb. Cortex* 6, 39–49.
- Daneman, M., Carpenter, P.A., 1980. Individual differences in working memory and reading. *J. Verbal Learn. Verbal Behav.* 19, 450–466.
- Davachi, L., Romanski, L.M., Chafee, M.V., Goldman-Rakic, P.S., 2004. Domain specificity in cognitive systems. In: Gazzaniga, M.S. (Ed.), *The Cognitive Neurosciences III*. The MIT Press, Cambridge, MA, pp. 665–678.
- Duncan, J., Owen, A.M., 2000. Dissociative methods in the study of frontal lobe function. In: Monsell, S., Driver, J. (Eds.), *Control of Cognitive Processes: Attention and Performance XVIII*. MIT Press, Cambridge, MA, pp. 567–576.
- Engel, S.A., Rumelhart, D.E., Wandell, B.A., Lee, A.T., Glover, G.H., Chichilnisky, E.J., Shadlen, M.N., 1994. fMRI of human visual cortex. *Nature* 369, 525.
- Engle, R.W., Kane, M.J., Tuholski, S.W., 1999. 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. (Eds.), *Models of Working Memory*. Cambridge Univ. Press, Cambridge, UK, pp. 102–134.
- Evans, A.C., Beil, C., Marrett, S., Thompson, C.J., Hakim, A., 1988. Anatomical–functional correlation using an adjustable MRI-based region of interest atlas with positron emission tomography. *J. Cereb. Blood Flow Metab.* 8, 513–530.
- Fox, P.T., Mintun, M.A., Reiman, E.M., Raichle, M.E., 1988. Enhanced detection of focal brain responses using intersubject averaging and change-distribution analysis of subtracted PET images. *J. Cereb. Blood Flow Metab.* 8, 642–653.
- Friston, K.J., Henson, R.N., 2006. Commentary on: divide and conquer; a defence of functional localisers. *NeuroImage* 30, 1097–1099.
- Friston, K.J., Frith, C.D., Liddle, P.F., Dolan, R.J., Lammertsma, A.A., Frackowiak, R.S., 1990. The relationship between global and local changes in PET scans. *J. Cereb. Blood Flow Metab.* 10, 458–466.
- Friston, K.J., Holmes, A.P., Poline, J.-B., Heather, J.D., Frackowiak, R.S.J., 1995. Analysis of fMRI time-series revisited. *NeuroImage* 2, 45–53.
- Friston, K.J., Rotshtein, P., Geng, J.J., Henson, R.N., 2006. A critique of functional localisers. *NeuroImage* 30, 1077–1087.
- 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.
- Fuster, J.M., 2002. Physiology of executive functions: the perception-action cycle. In: Stuss, D.T., Knight, R.T. (Eds.), *Principles of Frontal Lobe Function*. Oxford Univ. Press, Oxford, pp. 96–108.
- Fuster, J.M., Alexander, G.E., 1971. Neuron activity related to short-term memory. *Science* 173, 652–654.
- Gathercole, S.E., Pickering, S.J., 2000. Working memory deficits in children with low achievements in the national curriculum at 7 years of age. *Br. J. Educ. Psychol.* 70, 177–194.
- Glahn, D.C., Kim, J., Cohen, M.S., Poutanen, V.P., Therman, S., Bava, S., Van Erp, T.G.M., Manninen, M., Huttunen, M., Lonnqvist, J., Standertskjold-Nordemstam, C.G., Cannon, T.D., 2002. Maintenance and manipulation in spatial working memory: dissociations in the prefrontal cortex. *NeuroImage* 17.
- Goldman-Rakic, P.S., 2000. Localization of function all over again. *NeuroImage* 11, 451–457.
- Goldman-Rakic, P.S., Leung, H.-C., 2002. Functional architecture of the dorsolateral prefrontal cortex in monkeys and humans. In: Stuss, D.T., Knight, R.T. (Eds.), *Principles of Frontal Lobe Function*. Oxford Univ. Press, Oxford, UK, pp. 85–95.
- Grill-Spector, K., Malach, R., 2001. fMR-adaptation: a tool for studying the functional properties of human cortical neurons. *Acta Psychol.* 107, 293–321.
- Handwerker, D.A., Ollinger, J.M., D'Esposito, M., 2004. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage* 21, 1639–1651.
- Haxby, J.V., Ungerleider, L.G., Horwitz, B., Rapoport, S.I., Grady, C.L., 1995. Hemispheric differences in neural systems for face working memory: a PET-rCBF study. *Hum. Brain Mapp.* 3, 68–82.
- Haxby, J.V., Petit, L., Ungerleider, L.G., Courtney, S.M., 2000. Distinguishing the functional roles of multiple regions in distributed neural systems for visual working memory. *NeuroImage* 11, 380–391.
- Haynes, J.-D., Rees, G., 2005. Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat. Neurosci.* 8, 686–691.
- Jonides, J., 1995. Working memory and thinking. In: Smith, E.E., Osherson, D.N. (Eds.), *An Invitation to Cognitive Science*. MIT Press, Cambridge, MA, pp. 215–265.
- Josephs, O., Turner, R., Friston, K., 1997. Event-related fMRI. *Hum. Brain Mapp.* 5, 243–248.
- Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* 8, 679–685.
- Kanwisher, N., McDermott, J., Chun, M.M., 1997. The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J. Neurosci.* 17, 4302–4311.
- 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., Cheng, H.M., Brady, T.J., Rosen, B.R., 1992. Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proc. Natl. Acad. Sci. U. S. A.* 89, 5675–5679.
- Leung, H.-C., Gore, J.C., Goldman-Rakic, P.S., 2002. Sustained mnemonic response in the human middle frontal gyrus during on-line storage of spatial memoranda. *J. Cogn. Neurosci.* 14, 659–671.
- Leung, H.-C., Seelig, D., Gore, J.C., 2004. The effect of memory load on cortical activity in the spatial working memory circuit. *Cogn. Affect. Behav. Neurosci.* 4, 553–563.
- Levy, I., Hasson, U., Avidan, G., Hendler, T., Malach, R., 2001. Center-periphery organization of human object areas. *Nat. Neurosci.* 4, 533–539.
- Lin, F.H., McIntosh, A.R., Agnew, J.A., Eden, G.F., Ziffo, T.A., Belliveau, J.W., 2003. Multivariate analysis of neuronal interactions in the generalized partial least squares framework: simulations and empirical studies. *NeuroImage* 20, 625–642.
- Logie, R.H., Della Salla, S., 2003. Working memory as a mental workspace: why activated long-term memory is not enough. *Behav. Brain Sci.* 26, 745–746.
- Manoach, D.S., Schlag, G., Siewert, B., Darby, D.G., Bly, B.M., Benfield, A., Edelman, R.R., Warach, S., 1997. Prefrontal fMRI signal changes are correlated with working memory load. *NeuroReport* 8 (2), 545–549.
- Miller, E.K., 2000. The prefrontal cortex: no simple matter. *NeuroImage* 11, 447–450.
- Miller, E.K., Cohen, J., 2001. An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* 24, 167–202.
- Miller, G.A., Galanter, E., Pribram, K.H., 1960. *Plans and the Structure of Behavior*. Henry Holt and Company, New York.
- Miller, M.B., Van Horn, J.D., Wolford, G.L., Handy, T.C., Valsangkar-Smyth, M., Inati, S., Grafton, S., Gazzaniga, M.S., 2002. Extensive individual differences in brain activations associated with episodic retrieval are reliable over time. *J. Cogn. Neurosci.* 14, 1200–1214.
- Mottaghy, F.M., Gangitano, M., Sparing, R., Krause, B.J., Pascual-Leone, A., 2002. Segregation of areas related to visual working memory in prefrontal cortex revealed by rTMS. *Cereb. Cortex* 12, 369–375.
- Muller, N.G., Machado, L., Knight, R.T., 2002. Contributions of subregions of the prefrontal cortex to working memory: evidence from brain lesions in humans. *J. Cogn. Neurosci.* 14, 673–686.

- Munk, M.H.J., Linden, D.E.J., Muckli, L., Lanfermann, H., Zanella, F.E., Singer, W., Goebel, R., 2002. Distributed cortical systems in visual short-term memory revealed by event-related functional magnetic resonance imaging. *Cereb. Cortex* 12, 866–876.
- Murray, S.O., Olman, C.A., Kersten, D., 2006. Spatially specific fMRI repetition effects in human visual cortex. *J. Neurophysiol.* 95, 2439–2445.
- Narayanan, N., Prabhakaran, V., Bunge, S.A., Christoff, K., Fine, E.M., Gabrieli, J.D., 2005. The role of prefrontal cortex in the maintenance of verbal working memory information: an event-related fMRI analysis. *Neuropsychology* 19, 223–232.
- Niki, H., Watanabe, M., 1976. Prefrontal unit activity and delayed response: relation to cue location versus direction of response. *Brain Res.* 105, 79–88.
- Ogawa, S., Tank, D.W., Menon, R., Ellermann, J.M., Kim, S.G., Merkle, H., Ugurbil, K., 1992. Intrinsic signal changes accompanying sensory stimulation: functional brain mapping using MRI. *Proc. Natl. Acad. Sci. U. S. A.* 89, 5951–5955.
- Owen, A.M., Herrod, N.J., Menon, D.K., Clark, J.C., Downey, S.P.M.J., Carpenter, T.A., Minhas, P.S., Turkheimer, F.E., Williams, E.J., Robbins, T.W., Sahakian, B.J., Petrides, M., Pickard, J.D., 1999. Redefining the functional organization of working memory processes within human lateral prefrontal cortex. *Eur. J. Neurosci.* 11, 567–574.
- Passingham, R.E., Rowe, J.B., 2002. Dorsal prefrontal cortex: maintenance in memory or attentional selection? In: Stuss, D.T., Knight, R.T. (Eds.), *Principles of Frontal Lobe Function*. Oxford Univ. Press, Oxford, pp. 221–232.
- Peelen, M.V., Downing, P.E., 2005. Within-subject reproducibility of category-specific visual activation with functional MRI. *Hum. Brain Mapp.* 25, 402–408.
- Pessoa, L., Gutierrez, E., Bandettini, P.A., Ungerleider, L.G., 2002. Neural correlates of visual working memory: fMRI amplitude predicts task performance. *Neuron* 35, 975–987.
- Petersen, S.E., Fox, P.T., Posner, M.I., Mintun, M., Raichle, M.E., 1988. Positron emission tomographic studies of the cortical anatomy of single-word processing. *Nature* 331, 585–589.
- Petrides, M., 2000. Frontal lobes and memory, In: Boller, F., Grafman, J. (Eds.), *Handbook of Neuropsychology*, 2nd edition. Elsevier, Amsterdam, pp. 67–84.
- Polyn, S.M., Natu, V.S., Cohen, J.D., Norman, K.A., 2005. Category-specific cortical activity precedes retrieval during memory search. *Science* 310, 1963–1966.
- Postle, B.R., 2005. Analysis of fMRI data from tasks containing temporal dependencies: an evaluation of Slotnick (2005). *Cogn. Neuropsychol.* 22, 921–924.
- Postle, B.R., 2006. Working memory as an emergent property of the mind and brain. *Neuroscience* 139, 23–38.
- Postle, B.R., D'Esposito, M., 2000. Evaluating models of the topographical organization of working memory function in frontal cortex with event-related fMRI. *Psychobiology* 28, 132–145.
- Postle, B.R., Berger, J.S., D'Esposito, M., 1999. Functional neuro-anatomical double dissociation of mnemonic and executive control processes contributing to working memory performance. *Proc. Natl. Acad. Sci. U. S. A.* 96, 12959–12964.
- Postle, B.R., Zarahn, E., D'Esposito, M., 2000. Using event-related fMRI to assess delay-period activity during performance of spatial and nonspatial working memory tasks. *Brain Res. Protoc.* 5, 57–66.
- Postle, B.R., Druzgal, T.J., D'Esposito, M., 2003. Seeking the neural substrates of working memory storage. *Cortex* 39, 927–946.
- Postle, B.R., Ferrarelli, F., Hamidi, M., Feredoes, E., Massimini, M., Peterson, M.J., Alexander, A., Tononi, G., 2006. Repetitive transcranial magnetic stimulation dissociates working memory manipulation from retention functions in prefrontal, but not posterior parietal, cortex. *J. Cogn. Neurosci.* 18 (1712), 1722.
- Raichle, M.E., Martin, W.R.W., Herscovitch, P., Mintun, M.A., Markham, J., 1983. Brain blood flow measured with intravenous H_2O^{15} : II. Implementation and validation. *J. Nucl. Med.* 24, 790–798.
- Ranganath, C., DeGutis, J., D'Esposito, M., 2004. Category-specific modulation of inferior temporal activity during working memory encoding and maintenance. *Cogn. Brain Res.* 20, 37–45.
- Rissman, J., Gazzaley, A., D'Esposito, M., 2004. Measuring functional connectivity during distinct stages of a cognitive task. *NeuroImage* 23, 752–763.
- Rypma, B., 2006. Factors controlling neural activity during delayed-response task performance: testing a memory organization hypothesis of prefrontal function. *Neuroscience* 139, 223–235.
- 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., 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.
- Sala, J., Rama, P., Courtney, S.M., 2003. Functional topography of a distributed neural system for spatial and nonspatial information maintenance in working memory. *Neuropsychologia* 41, 341–356.
- Saxe, R., Brett, M., Kanwisher, N., 2006. Divide and conquer: a defense of functional localizers. *NeuroImage* 30, 1088–1096.
- Sereno, M.I., Dale, A.M., Reppas, J.B., Kwong, K.K., Belliveau, J.W., Brady, T.J., Rosen, B.R., Tootell, R.B., 1995. Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268, 889–893.
- Slotnick, S.D., 2005. Spatial working memory specific activity in dorsal prefrontal cortex? Disparate answers from fMRI beta-weight and timecourse analysis. *Cogn. Neuropsychol.* 22, 905–920.
- Swallow, K.M., Braver, T.S., Snyder, A.Z., Speer, N.K., Zacks, J.M., 2003. Reliability of functional localization using fMRI. *NeuroImage* 20, 1561–1577.
- Talairach, J., Tournoux, P., 1988. *Co-Planer Stereotaxic Atlas of the Human Brain*. Thieme Medical Publishers, New York.
- Tek, C., Gold, J., Blaxton, T., Wilk, C., McMahon, R.P., Buchanan, R.W., 2002. Visual perceptual and working memory impairments in schizophrenia. *Arch. Gen. Psychiatry* 59, 146–153.
- Tootell, R.B.H., personal communication.
- Tootell, R.B.H., Kwong, K.K., Belliveau, J.W., Baker, J.R., Stern, C.E., Hockfield, S.J., Breiter, H.C., Born, R., Benson, R., Brady, T.J., Rosen, B.R., 1993. Mapping human visual cortex: evidence from functional MRI and histology. *Invest. Ophthalmol. Visual Sci.* 34, 813.
- Tootell, R.B., Reppas, J.B., Kwong, K.K., Malach, R., Born, R.T., Brady, T.J., Rosen, B.R., Belliveau, J.W., 1995. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *J. Neurosci.* 15, 3215–3230.
- Tsao, D.Y., Freiwald, W.A., Tootell, R.H., Livingstone, M.S., 2006. A cortical region consisting entirely of face-selective cells. *Science* 311, 670–674.
- Ungerleider, L.G., Courtney, S.M., Haxby, J.V., 1998. A neural system for visual working memory. *Proc. Natl. Acad. Sci. U. S. A.* 95, 883–890.
- Veltman, D.J., Rombouts, S.A., Dolan, R.J., 2003. Maintenance versus manipulation in verbal working memory revisited: an fMRI study. *NeuroImage* 18, 247–256.
- Woodward, T.S., Cairo, T.A., Ruff, C.C., Takane, Y., Hunter, M.A., Ngan, E.T.C., 2006. Functional connectivity reveals load dependent neural systems underlying encoding and maintenance in verbal working memory. *Neuroscience* 139 (1), 317–325.
- Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited—Again. *NeuroImage* 2, 173–182.
- Zarahn, E., Aguirre, G.K., D'Esposito, M., 1997a. A trial-based experimental design for fMRI. *NeuroImage* 6, 122–138.
- Zarahn, E., Aguirre, G.K., D'Esposito, M., 1997b. Empirical analyses of BOLD fMRI statistics: I. Spatially unsmoothed data collected under null-hypothesis conditions. *NeuroImage* 5, 179–197.
- Zarahn, E., Raskin, 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.