The neural substrate and temporal dynamics of interference effects in working memory as revealed by event-related functional MRI

Mark D'Esposito*†, Bradley R. Postle*, John Jonides‡, and Edward E. Smith‡

*Department of Neurology, University of Pennsylvania, Philadelphia, PA 19104; and ‡Department of Psychology, University of Michigan, Ann Arbor, MI 48109

ABSTRACT Research on the prefrontal cortex (PFC) of monkeys and humans indicates that this region supports a heterogeneous repertoire of mental processes that contribute to many complex behaviors, such as working memory. Anatomical evidence for some of these processes derives from functional neuroimaging experiments using blocked experimental designs, which average signal across all components of many trials and therefore cannot dissociate distinct processes with different time courses. Using event-related functional MRI, we were able to isolate temporally the neural correlates of processes contributing to the target presentation, delay, and probe portions of an item-recognition task. Two types of trials were of greatest interest: those with Recent Negative probes that matched an item from the target set of the previous, but not the present, two trials, and those with Nonrecent Negative probes that did not match a target item from either the present or the two previous trials. There was no difference between the two trial types in target presentation (i.e., encoding) or delay-period (i.e., active maintenance) PFC activation, but there was significantly greater activation for Recent Negatives than Nonrecent Negative activation associated with the probe period within left ventrolateral PFC. These findings characterize spatially and temporally a proactive interference effect that may reflect the operation of a PFC-mediated response-inhibition mechanism that contributes to working memory performance.

Substantial evidence has accumulated that the prefrontal cortex (PFC) supports a heterogeneous set of cognitive functions. Fuster, for example, has described populations of single neurons within PFC of monkeys showing activity associated with information maintenance, inhibition, and motor-set, all within the context of working memory task performance (1). Recent functional neuroimaging studies in humans also have revealed a heterogeneity of working memory-related functions within PFC. Short-term maintenance of information, for example, may be associated with activity in several different regions of PFC whereas activity associated with the manipulation of information held in working memory is restricted to dorsolateral PFC (2, 3). Moreover, the differential dependence on PFC of the mnemonic (e.g., short-term storage, rehearsal) and nonmnemonic executive control (e.g., coordination of multitask performance, updating, attention shifting, inhibition) processes that contribute to working memory function also is supported by results from studies of patients with focal lesions (4, 5). In addition to testing hypotheses about differences in the neuroanatomical substrates of such mnemonic and nonmnemonic mental processes, another important way to elucidate the cognitive architecture of working memory-related processes is to investigate the temporal dynamics of

a process's contribution to performance on a working memory task.

A recent positron emission tomography study of an item-recognition working memory task by Jonides *et al.* (6) highlights a candidate PFC-mediated executive control process that may be dissociable in time, as well as in space, from other working memory-related processes. They found that Brodmann's area 45 in left ventrolateral PFC demonstrated greater activation under conditions in which stimulus overlap across trials created proactive interference (PI). This imaging finding, and the corresponding behavioral finding of increased reaction times in PI conditions, was attributed to the effects of inhibitory processing or some other conflict resolution process mediated by ventrolateral PFC. Because this study used a blocked design that averaged signal across all components of trials in each condition of the delayed-response task, however, it was unsuitable for investigation of temporal dynamics of the interference resolution processes ascribed to the observed PFC activation. For example, rather than reflecting a temporally discrete response-related process that was sensitive to PI, this neuroimaging effect could have occurred during stimulus presentation, delay, or retrieval components of trials in which PI was present, or in any combination of these three. Importantly, the PI-related activity reported in (6) also could have reflected a state-dependent mental "set" that persisted steadily during blocks of trials that featured high stimulus overlap, as might be expected if subjects adopted a different behavioral strategy for blocks that featured a high proportion of stimulus overlap across trials.

In the present study, we used a behavioral paradigm similar to (6) but adapted it to an event-related functional MRI (fMRI) design (7) that would allow us to isolate temporally the neural correlates of different components of the task. The event-related design also permitted randomization of trial types, thereby avoiding possible contamination of our data with context-dependent effects that can manifest themselves when trials representing a particular experimental treatment are blocked together (8). Each trial of the experiment presented a target set of consonant letters, followed by a delay period, followed by a probe letter (Fig. 1); subjects had to decide as quickly as possible whether the probe letter matched a letter in the target set. Results from two types of trials are detailed in this report: (*i*) Recent Negative trials, in which the probe did not match any items in the target set of the present trial but did match an item from the target set of the two previous trials, and (*ii*) Nonrecent Negative trials, in which the probe matched items from neither the current nor the two previous target sets. We hypothesized that a correct response to a Recent Negative probe would engage interference reso-
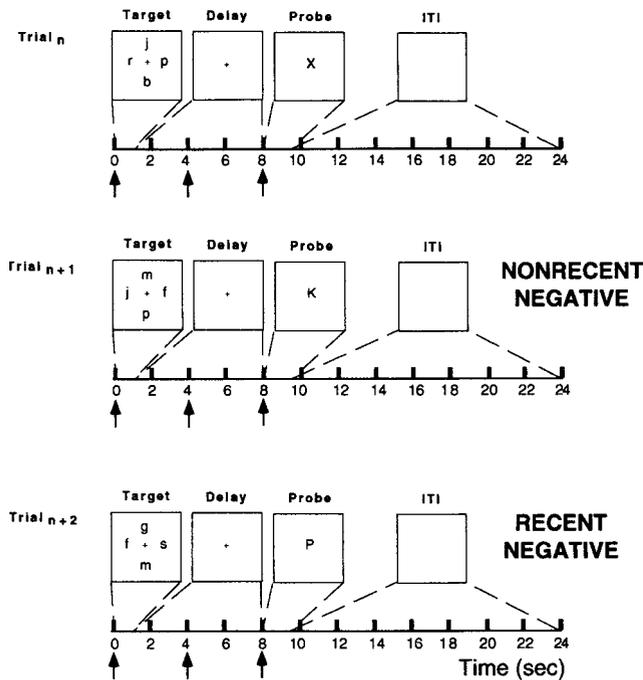

FIG. 1. Schematic illustration of the item-recognition task. Each trial lasted 9.5 sec followed by a 14.4-sec intertrial interval. Target items were presented for 950 msec followed by a 7,050-msec delay period. This interval was followed by a probe letter for 1,500 msec that either was recently presented (Recent trial) or was not recently presented (Nonrecent trial). Note that, in the Recent Negative trial, the probe letter "P" is not in the target set of that trial but that it is in the target set of the two previous trials. The arrows along the time line represent the placement of independent variables positioned to detect variance in the fMRI signal associated with target presentation, delay period, and probe portions of the trial, respectively.

lution processes, possibly inhibition of the tendency to respond "yes" based on the sheer familiarity of the probe (which had been rehearsed extensively on the two previous trials). A correct response to a Nonrecent Negative probe, by contrast, would minimize demands on interference resolution, reflecting the absence of PI from the previous trials. We predicted that interference resolution processes associated with Recent Negative trails would manifest themselves in brain activation and behavioral data.

METHODS

Subjects. Seven right-handed subjects (age 19–24) were recruited from the medical and undergraduate campuses of the University of Pennsylvania. Subjects were excluded if they had any medical, neurological, or psychiatric illness or if they were taking any type of prescription medication. All subjects gave informed consent.

Behavioral Task. Subjects performed an item-recognition task during which they judged whether the probe was from the target set of four consonant letters (Fig. 1). There were four trial types, each of which was defined by the probe: (*i*) A Recent Negative probe did not match any items from the current target set (and thus required a "no" response), but did match an item from the two previous target sets; (*ii*) a Nonrecent Negative probe did not match any items from the current target set (and thus required a "no" response) or the previous two target sets; (*iii*) a Recent Positive probe matched an item that was presented in the current target set (and thus required a "yes" response) and in the previous target set; and (*iv*) a Nonrecent Positive probe matched an item that was presented in the current target set (and thus required a "yes"

response) but not in the previous target set. Note that Recent Negative and Recent Positive probes each matched a stimulus that had appeared in the target set of two successive trials. Recent Negative probes appearing as the second trial of a block only matched an item from the previous target set. Each target set contained two items that had appeared in the previous target set and two items that had not appeared in either of the two previous target sets. An individual item never appeared in more than two consecutive target sets. On average, 4.8 trials separated a Nonrecent Negative probe from the previous appearance of this same item in the target set of a previous trial. Each subject performed a total of 128 trials (16 trials per run) in the proportion of 5 Recent Negative:4 Nonrecent Negative:3 Recent Positive:4 Nonrecent Positive. A roughly equal balance between positive and negative trials ensured attentiveness and obviated expectancy effects in our subjects, but data from the positive trials were not germane to the questions addressed in this report. Subjects were not informed about the manipulation of probe recency.

MRI Technique. Imaging was carried out on a 1.5T SIGNA scanner (GE Medical Systems, Milwaukee, WI) equipped with a prototype fast gradient system for echo-planar imaging. A standard radiofrequency head coil was used with foam padding to restrict comfortably head motion. High-resolution sagittal and axial T1-weighted images were obtained in every subject. A gradient echo, echoplanar sequence (TR = 2,000 ms, TE = 50 ms) was used to acquire data sensitive to the blood oxygen level-dependent signal. Resolution was 3.75×3.75 mm in plane and 5 mm between planes (21 axial slices were acquired). Subjects viewed a backlit projection screen from within the magnet bore through a mirror mounted on the head coil.

Data Analysis. Off-line data processing was performed on Ultra workstations (Sun Microsystems, Palo Alto, CA) by using programs written in Interactive Data Language (Research Systems, Boulder, CO). Initial data preparation proceeded in the following steps: image reconstruction; sinc interpolation in time (to correct for the fMRI slice acquisition sequence (9); motion correction [six-parameter, rigid-body, least squares realignment (10)]; and slice-wise motion compensation [to remove spatially coherent signal changes via the application of a partial correlation method to each slice in time (11)]. Note that the data were not smoothed spatially. Because fMRI data are temporally autocorrelated under the null-hypothesis (11), the data analysis was conducted within the framework of the modified general linear model for serially correlated error terms (12). Low frequency sinusoids and trial-mean covariates were included as nuisance covariates in our model (13). In the convolution matrix (12) was placed a time-domain representation of the expected 1/f power structure (11) and a notch filter that removed frequencies >0.244 Hz (i.e., at and around the Nyquist frequency) and <0.01 Hz (i.e., the portions of highest power in the noise spectrum). The data were not smoothed temporally with a low pass filter as advocated by Worsley and Friston (12) because including a 1/f model appears to control adequately the false-positive rate in relatively high frequency paradigms (11). Additionally, temporal smoothing would be undesirable with our technique because it would reduce the ability to detect signal changes during the delay (because of increasing colinearity of the delay-targeted covariates with those modeling the stimulus presentation and motor response).

The event-related fMRI analysis used in this study consisted of modeling fMRI signal changes occurring during the three procedurally and temporally discrete periods of each behavioral trial (i.e., target presentation, delay period, probe) with covariates composed of shifted blood oxygen level-dependent impulse response functions (7). An impulse response function estimate was obtained from each subject immediately before each fMRI experiment (9). We tested our hypotheses in a region of interest (ROI) within left ventrolateral PFC that was

Table 1. Behavioral performance [mean (SD)] in fMRI experiment

Subject	Accuracy, %				Reaction time, msec			
	Nonrecent Negative	Recent Negative	Nonrecent Positive	Recent Positive	Nonrecent Negative	Recent Negative	Nonrecent Positive	Recent Positive
MA	95	93	78	93	774	837	910	870
WB	100	98	97	89	719	758	726	688
MC	89	98	92	93	953	952	910	870
JH	97	98	97	88	910	995	1100	1034
BM	100	100	100	89	914	915	815	813
HR	100	100	97	100	798	835	798	835
EW	100	96	92	96	655	670	604	571
Mean	97 (4)	97 (2)	93 (8)	92 (5)	818 (111)	850 (113)	833 (155)	814 (148)

defined by creating a sphere (1 cm^3) around the local maximum [coordinates: −48, 21, 9 (14)] reported in Jonides *et al.* (6). Activity during each component of the task was assessed with contrasts (yielding *t* statistics with ≈1,300 df) of the parameter estimates associated with each task component covariate against baseline activity. The false-positive rates for all contrasts were controlled at the region-wise level at $\alpha = 0.05$ by Bonferroni correction for the number of voxels per ROI. Hypothesis tests were performed by identifying voxels within an ROI that were active during a specific component of the task [e.g., Recent Negative probe + Nonrecent Negative probe] and then performing the orthogonal contrast of [e.g., Recent Negative probe − Nonrecent Negative probe] on the spatially averaged time series that was extracted from those voxels. The resultant *t* value was a normalized index of the PI effect within each subject. Assessments of the reliability of trends in an effect across subjects were performed with paired *t* tests. The results from these random effects group analyses could generalize from our sample of subjects to the population represented by this sample, an inferential step that cannot be made with the fixed effects group analyses have been used by the majority of fMRI experimentalists (15, 16). These analyses were repeated in a second ROI comprising left middle frontal gyrus (Brodmann's area 9/46) that was created by outlining this ROI on a "canonical" representation of a brain in Talairach space that is provided in SPM96b, using the atlas of Talairach and Tournoux (14) to confirm our identification of anatomical landmarks. Next, we transformed this ROI from Talairach space into the native space in which each subject's data had been acquired by applying the 12-parameter affine transformation (10) with nonlinear deformations (17) routine. This ROI served as a control to verify the anatomical specificity of results obtained in the left ventrolateral PFC ROI.

RESULTS

Behavioral Data. All analyses were restricted to the Recent Negative and Nonrecent Negative conditions. The behavioral prediction was borne out in the data of each of the subjects and was confirmed in a group comparison: Recent Negative probes resulted in responses that were slower than Nonrecent Negative probes [818 vs. 850 msec; $t(6) = 2.82$; $P = 0.03$] but the accuracy between trial types did not differ [$t(6) = -0.16$] (see Table 1). These data are consistent with the hypothesis that a memory trace of items from the previous trials interfered with comparison of the probe to items from the current trial in the Recent Negative condition, resulting in longer latencies in making a decision about a negative probe item than in the Nonrecent Negative condition.

Imaging Data. Evoked signal within the left PFC ROI (inferior frontal gyrus; Brodmann's area 45) was determined for each subject during the three task components: target presentation, delay period, and probe. This ROI was chosen *a priori*, based on Jonides *et al.* (6). Each subject showed significant activity in the ROI during each of the three components of the task, when data were collapsed across all trial types. Across subjects, tests of the Recent Negative vs. Nonrecent Negative hypothesis revealed no differences in the evoked fMRI signal in the target presentation [$t(6) = 0.09$] or delay period [$t(6) = 0.39$] components of the task. At the probe, however, there was significantly greater evoked fMRI signal in the Recent trials as compared with the Nonrecent trials [$t(6) = 2.99$; $P = 0.03$]. Activation during each of the components of the task during these two trial types in ventrolateral PFC is illustrated in a representative subject in Fig. 2. Greater evoked fMRI signal in Recent Negative as compared with Nonrecent Negative trials, during the probe period, was evident in each individual subject (see Fig. 3). Inspection of the

FIG. 2. Pattern of activation with each of the task components. Shown are axial slices encompassing the ventral PFC ROI (white circles) in a representative subject (WB). Significantly greater activation (relative to the intertrial interval) is seen during the Recent Negative trials as compared with the Nonrecent Negative trials during the Probe period and not during the presentation of the Target or during the Delay period.
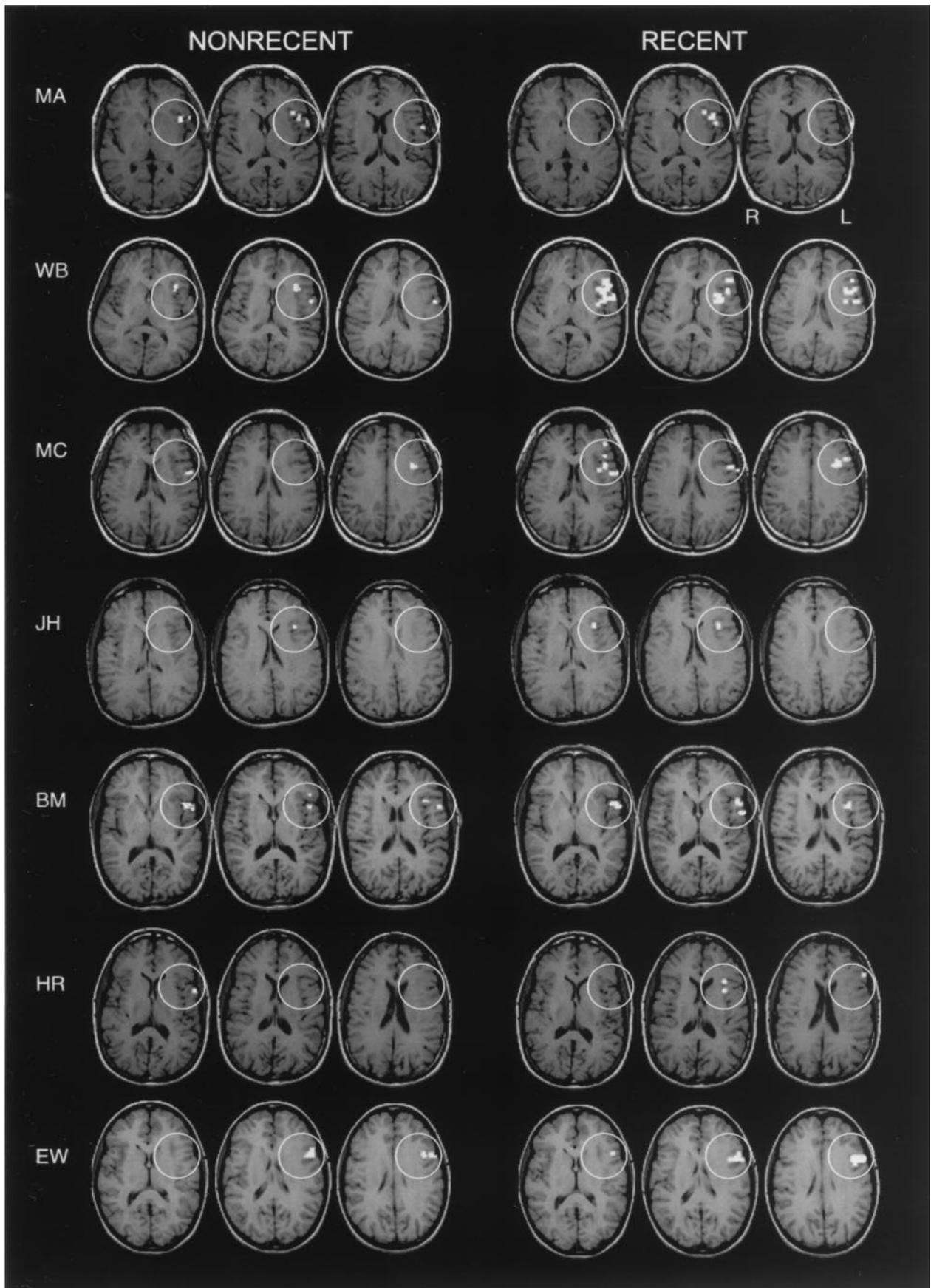
FIG. 3. Pattern of activation within the *probe* period of the task. Shown are axial slices encompassing the ventral PFC ROI (white circles) in each subject showing activation in the Recent Negative and Nonrecent Negative trials.
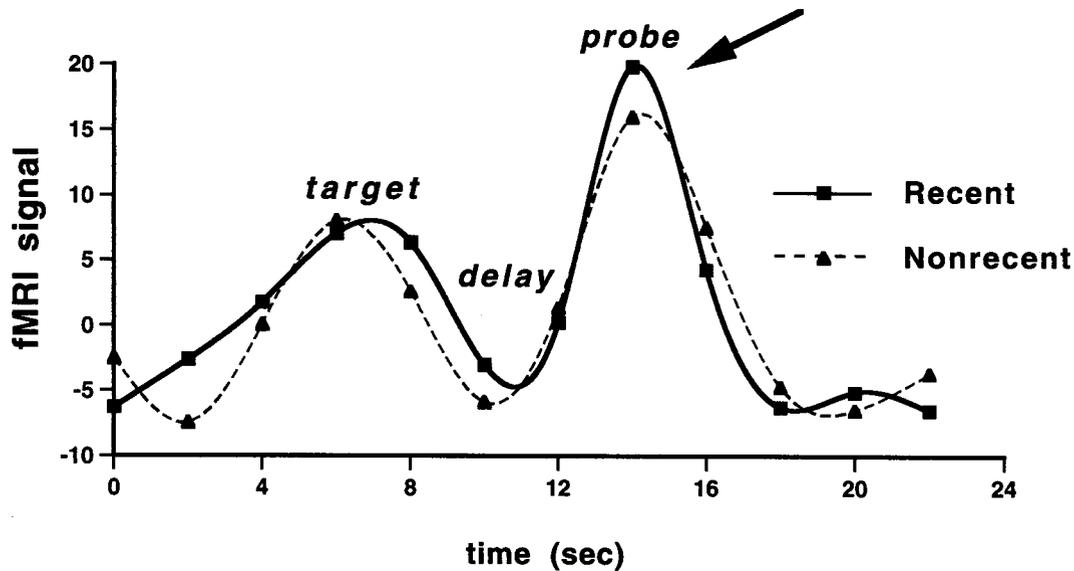
FIG. 4.   Trial averaged time series. Shown are the trial-averaged time series data for a representative subject extracted from voxels demonstrating a main effect for the Probe period across both Recent Negative and Nonrecent Negative trials. Activity in the two conditions differed statistically only during the Probe portion of the task (see arrow).

trial-averaged times series illustrated that this difference occurred only at the time of the probe (see Fig. 4).

To test the anatomical specificity of this result, we performed the same analyses as described above in a different area, the region of left dorsolateral PFC corresponding to Brodmann's areas 9 and 46. These analyses revealed no significant Recent Negative vs. Nonrecent Negative differences in any component of the task, including the probe [$t(6) = 0.44$].

DISCUSSION

Our results replicate the functional neuroanatomical evidence for a role for left ventrolateral PFC during trials of a verbal working memory task with a PI component (6) and extend our understanding of the PI resolution mechanisms in question by characterizing their temporal dynamics. The predicted temporal profile for these mechanisms became evident as soon as we reorganized the task from a blocked design to a trial-based design, of course, because the identity of a trial as a Recent or a Nonrecent trial only manifested itself with the onset of the probe. The empirical fMRI evidence for this PI effect, however, permits us to posit explicitly the existence of a discrete PFC-mediated mechanism that can contribute to working memory function and to characterize its temporal profile. This mechanism is engaged only during the comparison of the probe against mnemonic representations of the memory set. Presently, we characterize this mechanism as "interference resolution" because its fMRI and behavioral profiles are consistent with a mechanism that would detect and resolve conflicting signals between the negative outcome of a comparison operation between the probe and the current target set and a positive outcome based on the sheer familiarity of the probe. At present, however, we cannot adjudicate among alternative theoretical accounts of these data, which include response inhibition (6), selection among candidate memoranda (18), and probe discriminability (19, 20). Additional theorizing and empirical study will be necessary to elucidate the computational nature of this interference resolution mechanism. We can, however, classify it as a PFC-supported executive control process that can contribute to working memory function, just as are the attention-shifting, updating, and coordination mechanisms that we referred to in the introduction.

The results of this experiment establish that the functional heterogeneity of PFC-mediated processes contributing to

working memory function can be characterized in terms of the temporal asynchrony of different processes, as well as of their anatomical organization. Temporally, the interference resolution processes investigated in the present study are engaged with the onset of the probe and are thus dissociable in time from the processes that contribute to (*i*) encoding of the memory set stimuli (target) and (*ii*) maintenance of information across the delay period. Anatomically, these interference resolution processes are supported by left inferior PFC, leading us to predict that they are temporally and neurally dissociable from other executive control processes that can contribute to working memory performance. The processes recruited to manipulate information during the delay period (e.g., to rearrange letter memoranda into alphabetical order), for example, have been localized to dorsolateral PFC areas 9 and 46 (3). The finding that ventrolateral, and not dorsolateral, PFC seems to subserve interference resolution processes that also might be interpreted as response inhibition processes is consistent with previous work in monkeys that demonstrated impaired go/no-go performance after inferior convexity lesions (21). A recent imaging study in humans using a go/no-go task also demonstrated that activity attributable to the no-go trials (i.e., those requiring response inhibition) was localized to ventrolateral PFC (22).

1.   Fuster, J. (1997) *The Prefrontal Cortex: Anatomy, Physiology, and Neuropsychology of the Frontal Lobes* (Raven, New York).
2.   Owen, A. M., Evans, A. C. & Petrides, M. (1996) *Cereb. Cortex* **6,** 31–38.
3.   D'Esposito, M., Postle, B. R., Ballard, D. & Lease, J. (1999) *Brain Cognit.*, in press.
4.   D'Esposito, M. & Postle, B. R. (1999) in *Attention and Performance XVIII*, eds. Monsell, S. & Driver, J. (MIT Press, Cambridge, MA), in press.
5.   D'Esposito, M. & Postle, B. R. (1999) *Neuropsychologia*, in press.
6.   Jonides, J., Smith, E. E., Marshuetz, C., Koeppe, R. A. & Reuter-Lorenz, P. A. (1998) *Proc. Natl. Acad. Sci.* **95,** 8410–8413.
7.   Zarahn, E., Aguirre, G. K. & D'Esposito, M. (1997) *NeuroImage* **6,** 122–138.

8. Johnson, M. K., Nolde, S. F., Mather, M., Kounios, J., Schacter, D. L. & Curran, T. (1997) *Psychol. Sci.* **8,** 250–257.

9. Aguirre, G. K., Zarahn, E. & D'Esposito, M. (1998) *NeuroImage* **8,** 360–369.

10. Friston, K. J., Ashburner, J., Frith, C. D., Poline, J.-B., Heather, J. D. & Frackowiak, R. S. J. (1995) *Hum. Brain Mapp.* **2,** 165–189.

11. Zarahn, E., Aguirre, G. K. & D'Esposito, M. (1997) *NeuroImage* **5,** 179–197.

12. Worsley, K. J. & Friston, K. J. (1995) *NeuroImage* **2,** 173–182.

13. Friston, K. J., Holmes, A. P., Poline, J.-B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S. J. & Turner, R. (1995) *NeuroImage* **2,** 45–53.

14. Talairach, J. & Tournoux, P. (1988) *Co-Planar Stereotaxic Atlas of the Human Brain* (Thieme, New York).

15. Woods, R. P. (1996) *NeuroImage* **4,** S84–S94.

16. Holmes, A. P. & Friston, K. J. (1998) *NeuroImage* **7,** S754.

17. Ashburner, J. & Friston, K. (1996) *NeuroImage* **3,** S111.

18. Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K. & Farah, M. J. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 14792–14797.

19. McElree, B. & Dosher, B. A. (1987) *J. Exp. Psychol.* **118,** 346–373.

20. Monsell, S. (1978) *Cognit. Psychol.* **10,** 465–501.

21. Iversen, S. D. & Mishkin, M. (1970) *Exp. Brain Res.* **11,** 376–386.

22. Konishi, S., Nakajima, K., Uchida, I., Sekihara, K. & Miyashita, Y. (1998) *Eur. J. Neurosci.* **10,** 1209–1213