NEUROIMAGE **7,** 163–175 (1998)
ARTICLE NO. NI980328

Functional–Anatomic Study of Episodic Retrieval

II. Selective Averaging of Event-Related fMRI Trials to Test the Retrieval Success Hypothesis

Randy L. Buckner,*,† Wilma Koutstaal,‡ Daniel L. Schacter,‡ Anders M. Dale,† Michael Rotte,‡,§ and Bruce R. Rosen†

**Department of Psychology, Washington University, St. Louis, Missouri 63130; †Massachusetts General Hospital–Nuclear Magnetic Resonance Center; Department of Radiology, Harvard University Medical School, and ‡Department of Psychology, Harvard University; Boston, Massachusetts; and §Department of Neurophysiology, Otto-von-Guericke-University, Magdeburg, Germany*

Received September 24, 1997

In a companion paper (R. L. Buckner *et al.*, 1998, *NeuroImage* 7, 151–162) we used fMRI to identify brain areas activated by episodic memory retrieval. Prefrontal areas were shown to differentiate component processes related to retrieval success and retrieval effort in block-designed paradigms. Importantly, a right anterior prefrontal area was most active during task blocks involving greatest retrieval success, consistent with an earlier PET study by M. D. Rugg *et al.* (1996, *Brain* 119, 2073–2083). However, manipulation of these variables within the context of blocked trials confounds differences related to varying levels of retrieval success with potential shifts in subjects' strategies due to changes in the probability of target events across blocks. To test more rigorously the hypothesis that certain areas are directly related to retrieval success, we adopted recently developed procedures for event-related fMRI. Fourteen subjects studied words under deep encoding and were then tested in a mixed trial paradigm where old and new words were randomly presented. This recognition testing procedure activated similar areas to the blocked trial paradigm, with all areas showing similar levels of activation across old and new items. Of critical importance, significant activation was detected in right anterior prefrontal cortex for new items when subjects correctly indicated they were new (correct rejections). These findings go against the retrieval success hypothesis as formally proposed and provide an important constraint for interpretation of this region's role in episodic retrieval. Furthermore, anterior prefrontal activation was found to occur late, relative to other brain areas, suggesting that it may be involved in retrieval verification or monitoring processes or perhaps even in anticipation of subsequent trial events (although an alternative possibility, that the late onset is mediated by a late vascular response, cannot be ruled out). These findings and their relation to the results obtained in the companion blocked-trial paradigm are discussed.

In a companion paper (Buckner *et al.*, 1998a), we identified brain areas involved in episodic memory retrieval using a blocked fMRI recognition paradigm where trials of a similar type were presented sequentially, in sets or "blocks" of items. Across recognition task blocks, we varied the manner in which the target items had been previously studied (either under deep or shallow encoding conditions). The main finding was that blocks of recognition trials composed of items that were earlier presented under deep encoding conditions, and where retrieval success was high, were correlated with greater activation in right anterior prefrontal cortex as contrasted with low retrieval success (left anterior prefrontal cortex also showed a tendency for such an effect but was not included in targeted regional hypotheses). Rugg *et al.* (1996) and Tulving *et al.* (1994) have observed a similar effect in PET experiments examining recognition of visual words and auditory sentences, respectively. Thus, our companion study and these two previous studies all point to the possibility that anterior prefrontal cortex, particularly on the right, is differentially involved in processes related to retrieval success. However, these studies are all characterized by a possible complicating factor—groups of trials of the same type were presented in blocks, thereby allowing the more general task context within which the individual trials occurred to change across task blocks of different types. Blocks with lower or higher percentages of successfully retrieved items may have served to modify how subjects engaged in the recognition task, and/or the manner in which they arrived at their recognition decisions, and thereby have indirectly contributed to modulation of anterior prefrontal cortex, activation. Thus, for these three recognition studies we do not know whether differential right anterior prefrontal activation was due directly to retrieval success at an item-specific level or to shifts in subject-initiated strategies that might ensue when

1053-8119/98 $25.00

many trials of the same type are encountered (Wagner *et al.*, 1996).

The hypothesis that anterior prefrontal cortex is involved in retrieval success makes a direct prediction at the level of individual target items: there should be greater anterior prefrontal activation for trials where subjects correctly recognize previously studied or "old" items, compared to trials where subjects correctly rejected nonstudied or "new" items. Moreover, if retrieval success is a *necessary* condition for anterior prefrontal activation, then there should not be any significant anterior prefrontal activation associated with correct rejections. Thus, evidence for or against the retrieval success hypothesis could be generated by selectively averaging individual trials based on subjects' responses, contrasting activation observed exclusively under conditions of successful or correct recognition with that observed exclusively under conditions of correct "nonrecognition" (i.e., correct rejections). Such a hypothesis was previously impossible to test with PET and fMRI, given that the only available methods and analysis tools utilized blocks of trials averaged over many seconds; in these circumstances, subjects' responses within each block might include both correct and incorrect responses. However, a flurry of recent technical and analysis innovations have made it possible to conduct "event-related" fMRI where the hemodynamic signal contributions of individual trials in mixed designs are extracted (Boynton *et al.*, 1996; Buckner *et al.*, 1996a; Kim *et al.*, 1997; Konishi *et al.*, 1996; Zarahn *et al.*, 1997; Dale and Buckner, 1997; Josephs *et al.*, 1997; for review see Rosen *et al.*, 1998).

Our goal in the present study was to use these methods to selectively average individual trials based on subject performance during yes/no recognition and directly test the hypothesis that anterior prefrontal cortex is related to retrieval success. Importantly, the designs and analysis methods were structured such that evidence supporting or refuting the retrieval success hypothesis could be obtained. A secondary goal was to begin to characterize the data produced by event-related fMRI procedures, including examination of within- and across-subject variance of the hemodynamic response.

MATERIALS AND METHODS

Subjects

Twenty-six right-handed subjects between the ages of 18 and 35 years volunteered and received $50 as payment for participation. Fourteen of these subjects (7 male) contributed data to the main experiment. These subjects are nonoverlapping with the subjects studied in the companion paper (Buckner *et al.*, 1998a). One subject from this group produced data with sufficient artifacts to preclude further analysis. In addition, a further cohort of 12 subjects contributed control data (same control subject group as in companion paper). Informed consent for all subjects was obtained prior to scanning in a manner approved by the Human Studies Committee of the Massachusetts General Hospital.

General Magnetic Resonance (MR) Procedures

Imaging parameters and acquisition procedures were identical to the methods described in Buckner *et al.* (1998a) except that a larger number of runs and time points were acquired. The functional acquisition sequence was a T2*-weighted asymmetric spin echo sequence sensitive to blood oxygenation-level-dependent (BOLD) contrast (TE, 50 ms; offset, −25 ms). Functional images were acquired within eight runs of 128 time points each, with each time point sampling data over the entire brain, including the cerebellum (16 slice, in plane resolution 3.125 mm, 7 mm skip 1 mm thickness, acquisition aligned to the plane intersecting the anterior and posterior commissures, TR, 2 s). A total of 128 time points was chosen because 16 eight-time point trials could be included within each run. Four discarded time points were acquired prior to each run to allow T1 stabilization.

Data for each individual subject were transformed into the stereotaxic atlas space of the Talairach and Tournoux atlas (1988) using procedures identical to those applied in the companion paper (Buckner *et al.*, 1998a). However, because the analysis was not based on a block-designed paradigm, a number of steps intervened prior to Talairach and Tournoux atlas (1988) transformation (see Methods for Selective Averaging of fMRI data). These procedures generated stacks of images which summarized mean responses for each event (trial) type and variance at each time point within the event type. As these were the only data needed for our analyses, it was only these processed images that were transformed to atlas space and then averaged across subjects rather than raw time courses.

Behavioral Methods

The experiment was designed to explicitly test the hypothesis that anterior prefrontal activation was due to successful recognition, as suggested by the blocked design study of Rugg *et al.* (1996) and also by our companion study (Buckner *et al.*, 1998a). Prior to imaging, items were studied under deep encoding conditions. During imaging, items were tested in a single-trial task design consisting of randomly intermixed "old" (studied) items and "new" (nonstudied) items. Event-related fMRI procedures were used to sort the items *post hoc* based on subject performance. The logic of the test was that if anterior prefrontal cortex activation is driven by successful recognition, then greater levels of activation would be observed when

subjects correctly recognized items than when they correctly rejected items. Thus, the hypothesis would be supported if there is a significant difference between the two trial types, with significantly greater activation associated with correct recognition than with correct rejections. The stronger hypothesis that retrieval success is a necessary condition would be supported if significant activity were found only in the correctly recognized condition. If, however, anterior prefrontal cortex activation is driven by retrieval attempt or factors associated with retrieval attempt, regardless of trial outcome, then activation of this region would be significant during both kinds of trials, and the two trial types would not be significantly different from each other.

For study, words were presented, one at a time, approximately 40 min prior to recognition testing. The study words (abstract and concrete nouns) were presented visually in a single long series of items (approx 5 min, 136 items), using a presentation format identical to that of the companion paper (Buckner *et al.*, 1998a) except that only deep encoding was used to orient subjects (one word every 2 s, stimulus duration, 1 s; words presented in a 36-point Geneva font, white on black; fixation cross-hair displayed between words). For the deep encoding condition subjects again decided whether the words were abstract or concrete. Responses were indicated by a left-hand keypress. Four filler items were included at the beginning and end of the study block to reduce primacy and recency effects.

Recognition testing occurred across eight separate event-related fMRI runs (Fig. 1). Each run contained two trial types randomly intermixed and separated by visual fixation, including (i) New items and (ii) Old items. Item order was counterbalanced such that old and new items followed each other equally often and the position (across runs) was mirrored for item type (e.g., the position of a new item in one run would be the position of an old item in another run). Sixteen trials (each 1-s stimulus duration) were tested per run, separated by 16 s to allow near complete decay of the hemodynamic response (Buckner *et al.*, 1996a). It should be noted, however, that methods for rapid presentation of trials within event-related fMRI, which correct for or accommodate overlapping hemodynamic responses, have since become available and applied to memory

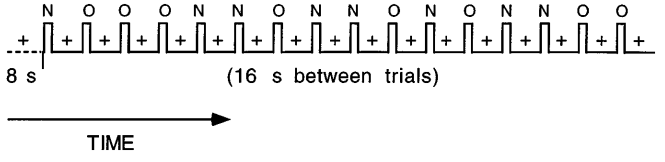


FIG. 1. A schematic illustration of the temporal organization of the task paradigm. One trial is presented every 16 s, with fixation between trials. Trials are randomly intermixed and present either Old or New items.

experiments (Clark *et al.*, in press; Buckner *et al.*, 1998b; Dale and Buckner, 1997). Subjects were instructed to decide, by pressing a key with their left hand, whether each item was old or new. Five seconds were allowed for each response although no feedback was given in the rare instance when a response was made outside that window. Subjects were instructed to respond quickly and accurately and to return to focusing on the fixation point as soon as they completed the trial.

In addition to the 14 subjects who performed the above recognition task, 12 additional control subjects underwent identical scanning procedures except there were no experimental task events. A fixation cross-hair appeared in the center of the screen and subjects were instructed to fixate for the duration of the run. The logic of this manipulation was that false positive rates could be determined from this data set by *post hoc* treating the data as if there were sham event trials (similar to Hunton *et al.*, 1996; Zarahn *et al.*, 1997b). Because no real event trials were present, all else being equal, any "significant" findings from the sham trials would be false positives and would thus allow empirical estimation of the frequency of such errors.

Event-Related fMRI Analysis Methods

For each fMRI run, the data were subjected to a series of normalization procedures. First, voxels falling outside the brain were masked using a threshold value found to consistently identify the brain/skull versus external space border (areas of signal loss due to susceptibility artifact are also masked using this procedure). Mean voxel intensity was calculated for the voxels within the brain. Images were then rescaled to a set whole-brain value of 1000 to eliminate variance across runs and subjects. Second, the linear slope was removed from each run on a voxel-by-voxel basis to counteract first order effects of drift (Bandettini *et al.*, 1993). Third, a one-voxel wide spatial smooth with a Hanning filter was applied to increase the signal-to-noise ratio (Xiong *et al.*, 1996). Finally, the mean signal intensity of the run was subtracted from every image within the run on a voxel-by-voxel basis. This last step has the effect of setting the mean value to zero so that signal intensity changes due to anatomic variance do not contribute to the retained signal. Without removing the mean signal intensity, anatomic variance across subjects can overwhelm the comparatively small contributions of BOLD contrast.

Normalized fMRI runs were then selectively averaged within each subject such that eight mean images (16 s at TR = 2 s) were retained for each trial type, as well as the variance for each of the eight images per trial type (see Dale and Buckner, 1997 for details of method). Five trial types were included: Hits, Misses, Correct Rejections, False Alarms, and Errors. An Error

was coded when the subject responded in less than 300 ms, failed to respond, or RT data were lost due to equipment malfunction. Once all trials were selectively averaged for each subject, the mean and variance images were transformed into stereotaxic atlas space (see General MR Procedures). Data were then averaged across subjects, weighting the variance and means by the number of trials, to yield a final set of event types that contained the composite data across all subjects for all runs.

Data from over 1500 individual trials contributed to this composite data set. By studying a relatively large cohort of subjects and trials, central tendencies in the hemodynamic response for each event type could be identified with a considerable amount of certainty and hence power for statistical analyses. However, because there is some debate in the field about how much variance exists in the hemodynamic response we also explicitly characterized variance across subjects and subject groups (see Examination of Between-Subject Variance of the Hemodynamic Response).

Activation maps were initially constructed in an exploratory manner using a *t* statistic as described by Dale and Buckner (1997; see also Schacter *et al.*, 1997). For this analysis, predicted hemodynamic curves were generated based on a gamma function (Boynton *et al.*, 1996). These functions used fixed parameters $d = 2.5$ s and $t = 1.25$ s (see Dale and Buckner, 1997) and a latency delay parameter that was varied between 2 s (early), 4 s (mid), and 6 s (late). Multiple delays were employed because considerable latency variation has been previously observed across brain regions (Schacter *et al.*, 1997). Main effects of trial type were explored by considering the mean hemodynamic response without explicit reference to any other condition. Because subjects were fixating between trials, this procedure has the effect of contrasting each trial type (or composite of trial types) with fixation. Direct contrasts between trial types were explored by covarying the predicted hemodynamic response functions against the difference timecourse generated by the subtraction of the two trial types being considered. A second type of *t* statistic map was created when more targetted temporal hypotheses were possible (e.g., after activation within a region was initially identified and the shape of the hemodynamic curve known). Main effects were explored using a paired *t* test across time points of the maximal difference (e.g., time 0 s versus time 6 s of all trial types). Effects across trial type were explored by contrasting the same time points across relevant trial types (e.g., time 6 s of Hits versus time 6 s of Correct Rejections).

The exact same set of normalization and statistical test procedures was performed on the control data set where subjects simply fixated a cross-hair for the entire run. For this analysis, two sham event types were used. These sham events were then explored for main effects and effects across trial type for both of the analysis procedures described above.

Having constructed statistical activation maps, three-dimensional regions were automatically defined around the peak activations and the time course observed. For this analysis, an algorithm identified all contiguous voxels within 12 mm of the peak that reached a significance level of $P < 0.0001$, similar to the methods applied to block data in the companion paper.

Examination of Between-Subject Variance of the Hemodynamic Response

An important question for event-related fMRI is how much variance there is in the hemodynamic response across individual subjects and subject groups. In particular, our analyses are based on a range of predicted hemodynamic response functions, assuming that individuals within a group (and hence the averaged subject group) converge toward a modal hemodynamic response when many events are considered. However, it has been argued that this assumption is not justified and that there exists too much variance to average across subjects—and perhaps even over trials (Kim *et al.*, 1997). To justify this assumption, we characterized across-subject variance in the hemodynamic response for two regions: supplementary motor area (SMA) and extrastriate cortex. These two regions represent the largest perceptual and motor response activations that we observed in the present paradigm and could thus be identified most reliably within subjects. A caveat regarding this analysis procedure is that, in this approach, only hemodynamic responses of a certain class are examined—responses with a single peak occurring several seconds after the stimulus onset. This shape of response has been observed across a wide range of studies and thus comprises a reasonable starting point for analysis (e.g., Buckner *et al.*, 1996a; Konishi *et al.*, 1996; McCarthy *et al.*, 1997; Zarahn *et al.*, 1997a). If responses exist that are outside our proposed hemodynamic shape, we would not identify those voxels. Thus, our analysis should be considered an exploration of variance across a certain class of hemodynamic responses but not exhaustive.

Peaks of activation were identified from the statistical maps in the Talairaich atlas space for each individual subject without averaging across subjects. Automated three-dimensional regions (see Generation of Statistical Activation Maps for Event-Related fMRI Data, above) were then defined, seeded by these peak locations. Both regions could be identified in all 13 subjects. To characterize the variance in our sample, the mean responses for Hits and Correct Rejections were obtained for each region, for each subject and

plotted (26 separate hemodynamic curves). Standard error estimates of the 26 curves for all time points during the hemodynamic response were also derived. As a second analysis, to characterize the stability of the hemodynamic response across independent subject groups, the pool of 13 subjects was divided into two subgroups containing 7 and 6 subjects each. The mean hemodynamic response for each subgroup was then calculated, separately for Hits and Correct Rejections, and contrasted. As a final analysis, we correlated the hemodynamic response across subjects and conditions by deriving a complete correlation matrix across all 13 subjects for both Hits and Correct Rejections. This procedure quantitatively estimates how much variance in the hemodynamic response can be predicted between subjects and how much variance can be predicted within subject.

Behavioral Results

Subjects correctly recognized 78% of the old items (Hits) and correctly rejected 76% of the new items. Mean probability of a Hit minus probability of a False Alarm ($P_{HIT} - P_{FA}$) was 0.54, suggesting good discrimination between old and new items. Mean reaction time significantly varied across the trial types (ANOVA $F[3, 11] = 8.93$, $P < 0.001$), with correct rejections (1773 ms) showing significantly slower reaction times than hits (1616 ms) as indicated by a *post hoc t* test ($t[12] = 3.21$, $P < 0.01$). These reaction times were considerably slower than those observed in the companion study where a 2-s intertrial interval was imposed (Buckner et al., 1998).

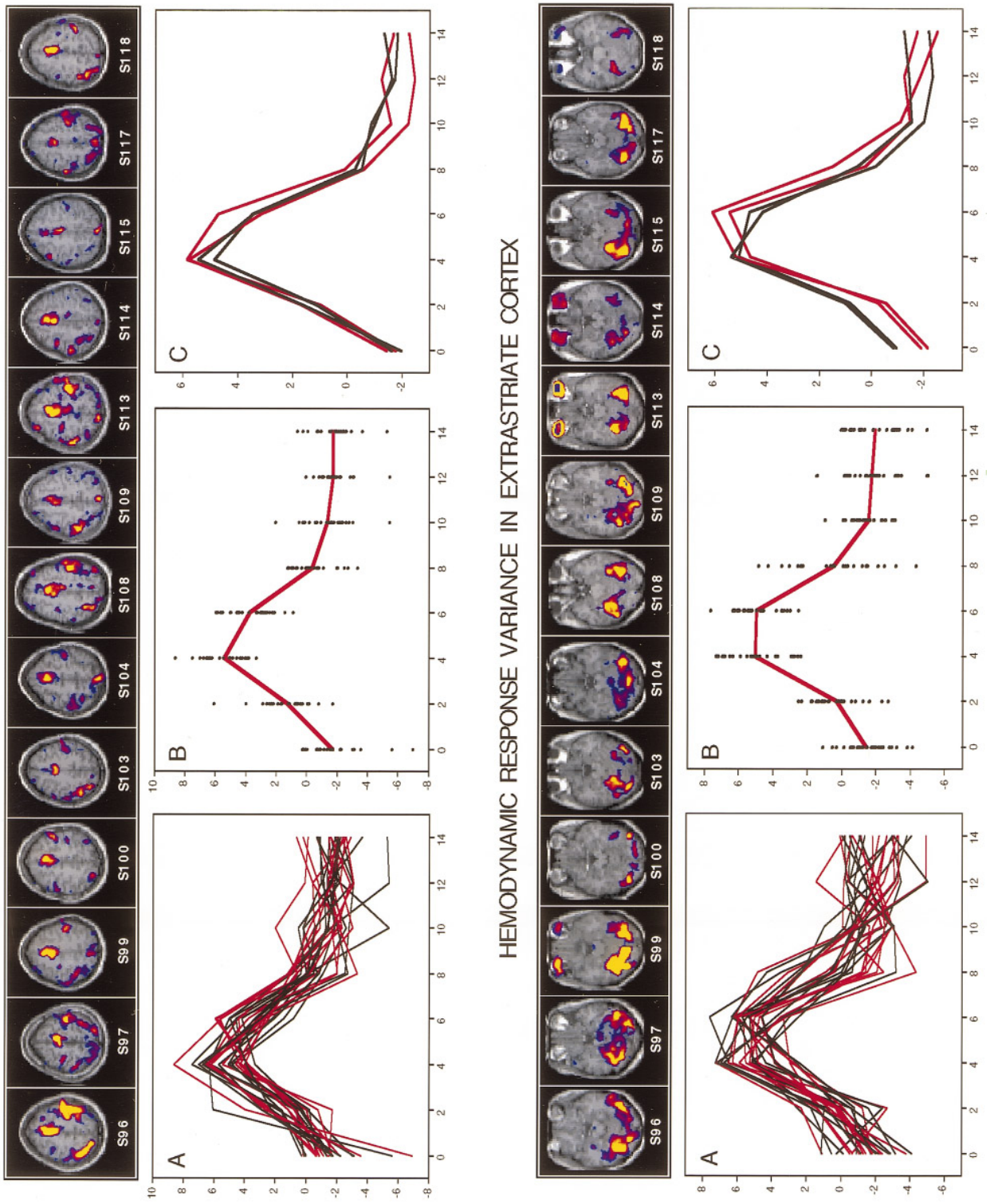
Because our procedure involved presentation of widely spaced trials—a procedure unconventional for typical behavioral and ERP studies—we explored the effects of trial position within runs and across blocks to determine if subject performance was stable. No effect of trial position within run was found ($r^2 = 0.00$), suggesting that subjects did not fatigue or change strategy as trials progressed within runs. The effect of run, however, was significant ($r^2 = 0.11$, $t[12] = 3.94$, $P < 0.001$), with subjects showing a decrease in reaction time across each run, perhaps reflecting a learning effect or a greater reliance on familiarity based retrieval across runs.

fMRI Results

Variance of the hemodynamic response. The onset and resolution of the hemodynamic response varied across the 13 subjects with a range of about two seconds for both the SMA and extrastriate regions tested (Fig. 2A). The mean hemodynamic response started at about 2–4 s for both regions, peaked between 4 and 6 s (note that time point 4 s represents the start of image acquisition which lasts 2 s), and was resolved by 10–12 s post-stimulus onset (Figs. 2A and 2B). This general pattern was consistent across all 13 individual subjects—although it should be noted that our analysis techniques would not identify voxels with marked deviations from this pattern. The relatively fast onset was statistically significant, with the response at 2–4 s being larger than the response at 0 s for both regions ($t[25] = 5.25$, $P < 0.001$, and $t[25] = 5.33$, $P < 0.001$). When the mean hemodynamic responses for independent groups of subjects and conditions were considered, they were extremely similar in timing and shape (Fig. 2C). This stability carried to the level of the individual subject, as the variance of the peak response across subjects was nonoverlapping with the variance at the onset of the response (time 0 s; see Fig. 2B). Quantitatively, in our sample, knowing the shape of the curve for one subject could predict—on average—72% of the variance (i.e., r^2) in the shape of another subject's response. That value went up to 76% if we used the shape of the curve for individual subjects as a means of predicting their own responses across the two conditions. These observations suggest that across-subject averaging of the hemodynamic response is appropriate and benefits from strong central tendencies in the shape of the response. Nonetheless, it is also evident that variance is clearly present, and methods accounting for this additional quarter of the variance may further increase statistical power.

Main effects of recognition. Direct examination of the Recognition trials (independent of trial subtype) revealed activation of a large number of brain areas. These areas overlapped considerably with those found to be activated in the companion blocked-trial paper (Fig. 3; Buckner *et al.*, 1998a) and included visual striate and extrastriate cortex, right-lateralized motor cortex, premotor cortex, SMA, and multiple cerebellar

FIG. 2. Results from the analysis of hemodynamic response variance are displayed for two separate regions [supplementary motor area (SMA) and extrastriate cortex] across the 13 subjects. For each region, the set of 13 images on top are the statistical activation maps for each individual subject in Talairach and Tournoux (1988) atlas space (threshold for all maps, $P < 0.001$). The slice plane is selected to cut through the center of the region. As can be seen, all subjects showed significant activation in approximately the same regions. (A) The BOLD hemodynamic response for each subject, for each of the two conditions of interest (Hits, red; Correct Rejections, black) is plotted. The onset and time of peak response varied over a range of about 2 s across subjects. (B) The mean hemodynamic response is displayed (red) along with the range of responses from each individual hemodynamic curve (black filled circles). (C) The mean response from two separate averaged cohorts of subjects are shown (subject group one, red; subject group two, black) for each of the two conditions. The two groups, although differing slightly, show highly similar modal hemodynamic responses for both regions.


HEMODYNAMIC RESPONSE VARIANCE IN SMA
S96
S97
S99
S100
S103
S104
S108
S109
S113
S114
S115
S117
S118
A
B
C
HEMODYNAMIC RESPONSE VARIANCE IN EXTRASTRIATE CORTEX
S96
S97
S99
S100
S103
S104
S108
S109
S113
S114
S115
S117
S118
A
B
C

2

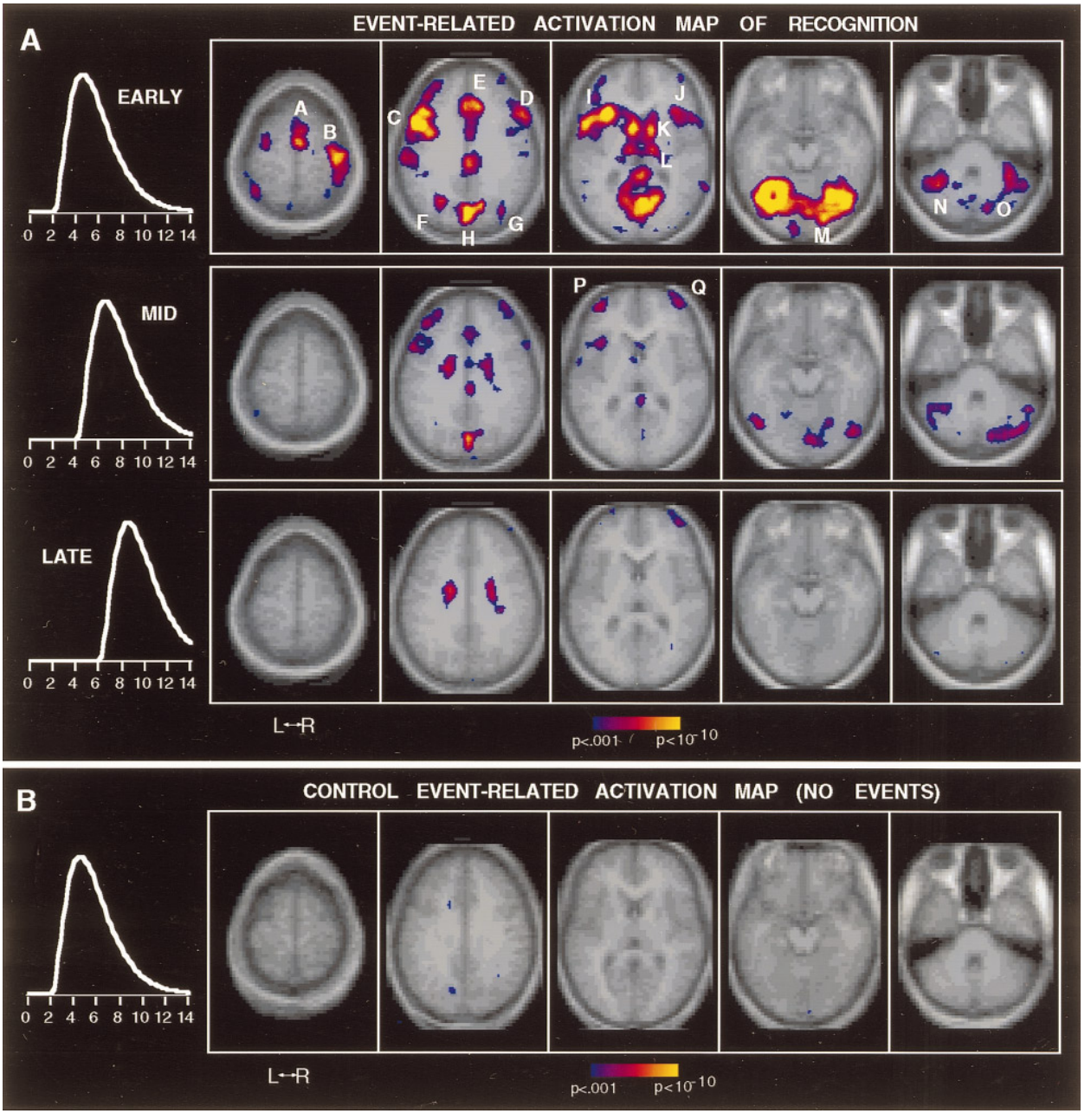


FIG. 3. (A) BOLD signal increases are shown for the combined recognition events modeled separately against three predicted hemodynamic response functions differing in onset time (Early, Mid, and Late). Statistical maps (colored scale) overlay the averaged SPGR anatomic image. Many areas are activated including (A) SMA, (B) right motor/somatosensory cortex, (C) left dorsal prefrontal/motor cortex, (D) right dorsal prefrontal/motor cortex, (E) SMA extending into anterior cingulate, (F and G) lateral parietal cortex, (H) posterior medial parietal cortex near precuneus, (I and J) bilateral anterior insular cortex near the frontal-operculam, (K) basal ganglia, (L) medial thalamus, (M) striate and extrastriate cortex, (N and O) lateral cerebellum, and (P and Q) anterior prefrontal cortex. Certain regions show significant activation only when predicted by the hemodynamic responses with late onsets (e.g., anterior prefrontal cortex is active in the Mid and Late maps). (B) BOLD signal increases for the Early hemodynamic response function are shown for the sham control data set where no real events existed. False positives were not detected at the applied significance level.

regions. In addition, bilateral anterior insular cortex near frontal-opercular cortex, left dorsolateral prefrontal cortex, anterior prefrontal cortex near Brodmann area 10, and posterior medial parietal cortex were also all activated and correspond to the collective set of areas that have been consistently identified during episodic retrieval tasks (see companion paper for discussion).

Also of interest was the finding that activation within certain brain areas was best fit when predicted by hemodynamic curves of varied lags (Fig. 3). For ex-

ample, bilateral anterior insular cortex was most significant when analyzed with a predicted response with a 2-s lag to onset, whereas anterior prefrontal cortex showed greatest significance at lags of 4 and 6 s. As suggested by previous studies, these findings again demonstrate that certain brain areas reveal delayed hemodynamic lags relative to other brain areas, with the magnitude of that delay falling within the order of seconds (Buckner *et al.*, 1996a; Schacter *et al.*, 1997). To better characterize this phenomenon, a *single* slice was selected that included activation within visual-extrastriate, bilateral anterior insular cortex, and anterior prefrontal cortex. The raw time course was calculated across the three activations by defining *two-dimensional* regions confined to the slice. By confining regions to a single slice, acquisition timing (which is varied across slices) does not artifactually contribute to timing offsets across regions. Furthermore, in order to appreciate the entire timecourse of the response, the time window was extended to include the period after each 16-s trial into the next trial. The results are shown in Fig. 4 and are consistent with the notion that anterior prefrontal areas can show both delayed, and prolonged, responses relative to more posterior visual and extrastriate regions. However, it should be noted that this analysis does not afford an unbiased method of estimating the time course of regions: regions are defined within the same data set used to derive the absolute time course and not on, for example, an independent data set (e.g., Schacter *et al.*, 1997).

Hypothesis testing. Tests of the retrieval success hypothesis were performed by examining both the activation patterns for the Hits versus the Correct Rejections separately and by examining the two conditions against each other. Three critical activations were explored that were nearly identical in location to the regions explored in our companion paper (Buckner *et al.*, 1998). These regions included right anterior prefrontal cortex, left dorsolateral prefrontal cortex, and bilateral anterior insular cortex. The location of the peak coordinate at the center of these regions are listed in Fig. 5 and can be directly contrasted to the location in the companion manuscript (Fig. 4 or Buckner *et al.*, 1998). The main findings are shown in Figures 5 and 6. Both Hits and Correct Rejections demonstrated similar activation levels across the three regions, with similarly shaped hemodynamic responses.

The two separate conditions were further explored to determine whether each condition, on its own, activated these three regions. All regions were significantly activated by both conditions. Of central importance, selectively averaged Correct Rejections showed significant activation of right anterior prefrontal cortex, with some activation in the homologous region on the left. The peak coordinate of the right lateralized activation was located at $x = 37$, $y = 56$, $z = 9$ in the Hit condition and $x = 34$, $y = 59$, $z = 9$ in the Correct Rejection condition. The distance between these present coordinates and the location in the companion blocked-trial paper is within 3 mm in each dimension. Furthermore,

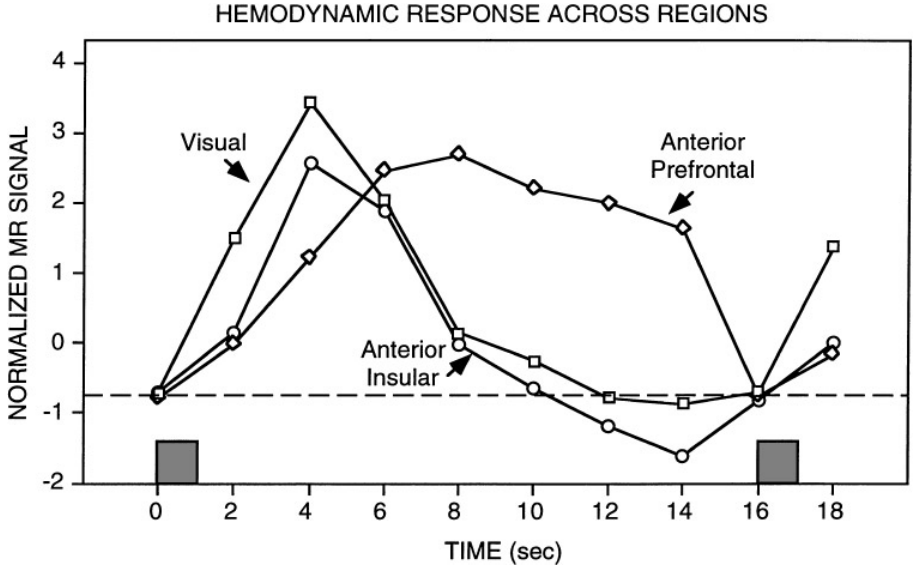


FIG. 4. The time course for the BOLD hemodynamic response within three regions (visual, anterior insular, and anterior prefrontal) are shown for the grand-averaged trial across all trial types. The time period extends into the next trial to appreciate the full evolution of the hemodynamic response. The position of the trial being averaged and the relative position of the following trial are shown by shaded rectangles. The visual and anterior insular regions demonstrate quick to develop hemodynamic responses peaking between 4 and 6 s, while the anterior prefrontal region does not show a peak until 8 to 10 s. Moreover, the anterior prefrontal hemodynamic response is temporally extended and does not dissipate until the onset of the next trial.

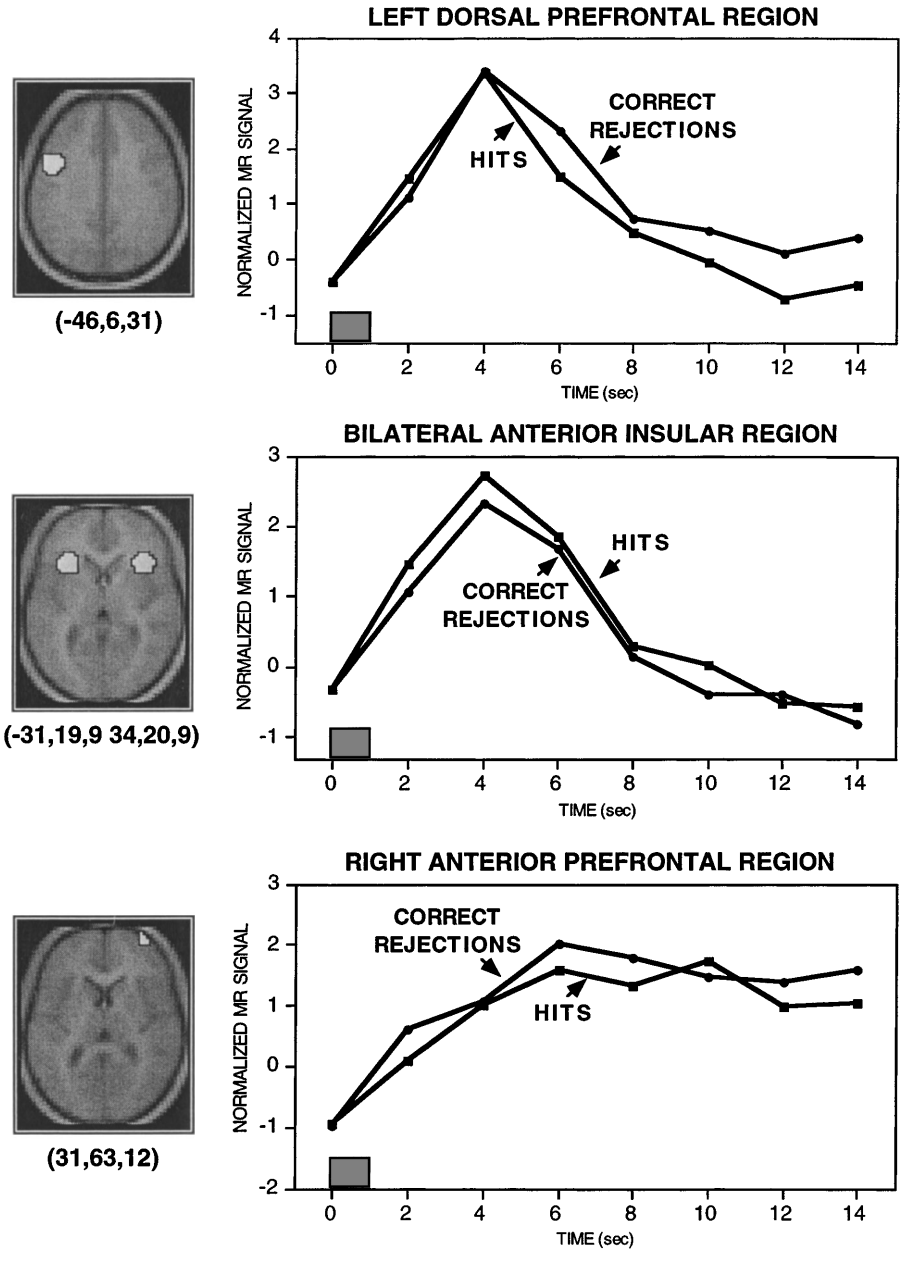


FIG. 5. The BOLD signal time course is displayed for each of the three regions found to be significantly modulated by retrieval effort or retrieval success in our companion paper (Buckner et al., 1998). For each region, one slice from the region is shown superimposed on the averaged anatomic image (leftmost panels) with the peak coordinates of the region listed below [*x*, *y*, *z*, Talairach and Tournoux (1988) atlas]. These regions, which are defined based on the present data set, are highly similar to the regions defined in the companion paper. BOLD signal, in all instances, was not found to significantly differ between the Hits and Correct Rejections.

the location is quite close to that reported by Rugg *et al.* (1996) in their examination of brain areas related to retrieval success (40,50,8 and 38,48,8 across two contrasts in their study). Moreover, this region did not show significant modulation when the Hits were directly compared to the Correct Rejections, for any of the analysis procedures—consistent with the overlapping timecourse of response observed in Fig. 5. Results thus indicate that, under these event-related procedures, successful recognition and correct rejections *both* activate anterior prefrontal cortex and differential activation is minimal, if present at all.

General exploration of differences across hits and correct rejections. Exploratory maps were generated by directly contrasting Hits and Correct Rejections. Results were marginal. No regions showed a clear change across the two trial types for any of the three hemodynamic delays.

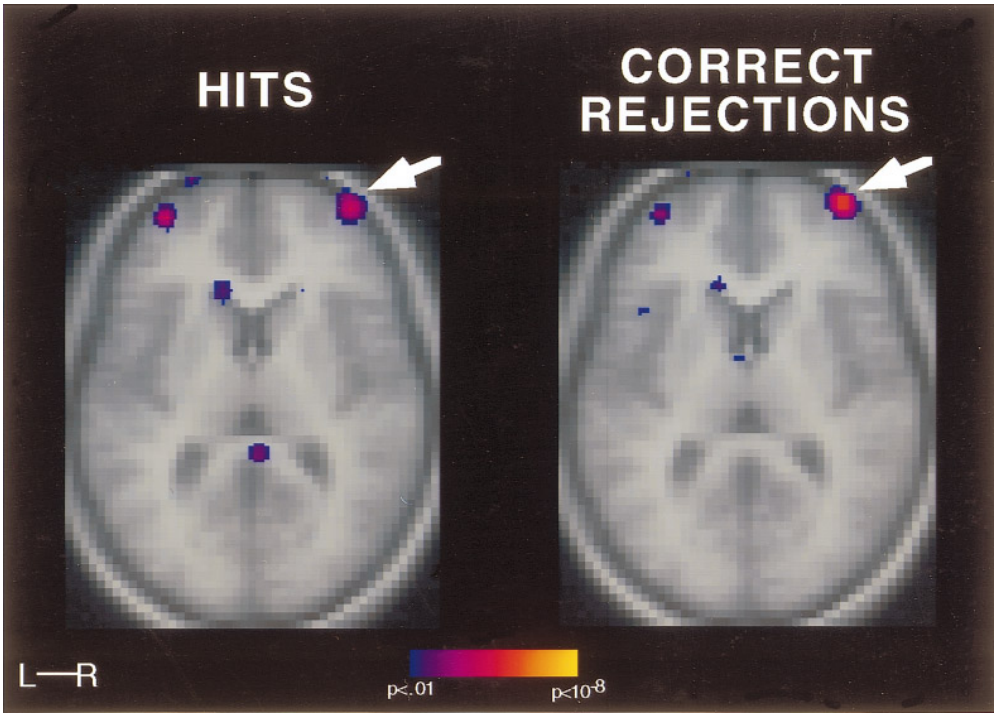


FIG. 6. Statistical maps show BOLD signal increases separately for the selectively averaged trials of Hits and of Correct Rejections. The activation maps show the contrast between signal level at the start of the trial (time 0–2 s) to late in the trial (time 8–10 s). Significant activation is observed in anterior prefrontal cortex for both trial types. Posterior regions, which also show significant signal change, are not highlighted in the present activation maps which target signal changes showing late and prolonged hemodynamic responses that are characteristic of anterior prefrontal cortex.

DISCUSSION

The main findings of theoretical interest are that (1) right anterior prefrontal cortex was significantly activated not only during trials in which subjects correctly recognized items, but also during trials in which subjects correctly rejected new items, and (2) no differential activation was detected between these two types of trials (i.e., Correct Rejections, and trials in which subjects successfully identified old items, or Hits). The failure to find greater anterior prefrontal activation for hits than correct rejections goes directly against a retrieval success hypothesis as outlined by Rugg *et al.* (1996) and, on first order, apparently supported by our companion paper (Buckner *et al.*, 1998a). The fact that anterior prefrontal regions were activated significantly for correct rejections provides evidence against the even stronger hypothesis that successful retrieval is a necessary condition for activation within this area.

Nonetheless, there need not be conflict among any of the various results concerning the relation between retrieval success and anterior prefrontal regions. If both kinds of trials (Hits and Correct Rejections) can equally elicit right anterior prefrontal activation during episodic retrieval, as suggested by the present results, then the explanation of the different pattern of findings across several studies likely lies in differences between the testing procedures that were used or—more specifically—the manner in which those procedures alter the *context* in which individual trials occur. Blocks of trials that mostly contain items of a similar type, such as were used by Rugg *et al.* (1996) and in our companion study, might allow subjects to consciously or unconsciously detect the likelihood that a trial of a certain type will follow. Subjects may come to anticipate items, adjust how they respond to them, or even modulate the degree to which they evaluate their responses. By contrast, in mixed trial paradigms, such as that used here, context effects arising from subjects' perception of, and response to, varying between-trial contingencies are held to a minimum: subjects cannot predict an item or their response in advance as the average stimulus probabilities for different types of items remain constant both within and across runs.

Rugg *et al.* (1996) first suggested the possibility that context effects might be influencing anterior prefrontal results, but viewed the possibility as unlikely because their subjects did not appear to change performance following blocks of trials with different probabilities of retrieval success. However, in light of the present fMRI results, it would seem that a context account is likely.

Further support of the context account comes from Wagner and colleagues (submitted for publication) who have recently shown that test instructions that provide information about target probabilities can alter anterior prefrontal activation levels in blocked-trial recognition paradigms. In such instances, subjects might anticipate the nature of the requisite responses for upcoming trials and adjust their strategies even before the actual block begins. Importantly, the strong hypothesis that right anterior prefrontal cortex is obligatorily related to retrieval success is falsified by our result of significant activation during trials that are explicitly selected to contain no retrieval success.

Although it is always possible that a certain level of differential modulation due to trial type was present but below our sensitivity, both the reproducible and the robust main effects that were observed in the present study, and our ability to modulate anterior prefrontal cortex at a detectable level in the companion paper, argue against this possibility and suggest that lack of statistical power is an unlikely explanation. Moreover, the time course of the hemodynamic response for the Correct Rejections overlapped with that for the Hits (Fig. 5).

We must therefore reevaluate what the role of anterior prefrontal cortex might be in episodic retrieval. The present findings and those already in the literature place a number of constraints on the potential role. (1) Right anterior prefrontal cortex at or near Brodmann area 10 is activated by many episodic retrieval tasks compared to a number of different reference control tasks (both when studied with PET and based on blood flow and when studied with fMRI and based on BOLD contrast). (2) Participation is amodal and generalizes across many kinds of episodic retrieval task: Auditory and visual stimuli, recall and recognition, and veridical and illusory recognition have all activated anterior prefrontal cortex (see Buckner, 1996, for review; Schacter *et al.*, 1997). (3) Activation can be elicited whether or not the task involves successful recognition of studied items or the absence of recognition (correct rejections) as indicated by the present results. (4) Activation can be modulated in blocked-trial designs, presumably due to context effects (Rugg *et al.*, 1996; Buckner *et al.*, 1998; Wagner *et al.*, 1996) and (5) Activation can occur during certain working memory and semantic retrieval tasks—although less consistently so than during episodic retrieval (MacLeod *et al.*, 1998). These multiple constraints do not point toward a selective role of anterior prefrontal cortex in episodic memory in terms of initially proposed ideas about basic processes involved in episodic retrieval (e.g., retrieval effort and/or retrieval success). We suspect that the role is more general and perhaps involves higher level monitoring or evaluation.

Right anterior prefrontal cortex was also observed to have a delayed and prolonged hemodynamic response relative to other regions. We have observed this pattern previously (Schacter *et al.*, 1997). One possible account of this delay is that the regions are sampling different vascular distributions with intrinsic variation in the observed shape of the hemodynamic response (Lee *et al.*, 1995; Rosen *et al.*, 1998). This possibility cannot currently be ruled out as an account of the present data. However, assuming that the effect is not vascular in origin, the observation of the delayed response further helps to suggest a processing role for this region. The present data are consistent with (1) a role in postretrieval verification and/or monitoring or (2) a decreasing response in anticipation of the following trial. The latter possibility, while currently speculative, can account for the observation that the response builds up as the subsequent trial approaches and (assuming a decreasing response) resolves following the onset of the trial.

These possible roles might also help to explain the results observed in the companion paper (Buckner *et al.*, 1998). In the blocked-trial study, anterior prefrontal cortex activation varied inversely with the amount of time required to make a response on the trial: High Recognition, which showed the fastest reaction times, demonstrated the highest levels of activation in anterior prefrontal cortex. Perhaps the relevant processes are engaged after the trial has been completed. Thus, allowing more time between the completion of the trial and the onset of the next trial would actually increase time allotted for postretrieval processes and thereby increase activation. Further exploration is needed to sort amongst these alternatives.

One puzzling aspect of the delayed anterior prefrontal region activation is that the region is activated in the blocked task paradigm where trials are presented and responded to every 2 s (Buckner *et al.*, 1998a). Thus, a multisecond hemodynamic delay as observed in the present study would show overlap across trials in the blocked-trial study. We cannot currently resolve this issue but several possibilities exist. If the delayed hemodynamic response is related to differential vascular sampling then the answer may simply lie in the fact that the delay does not reflect delayed neuronal activity. Alternatively, the time course in the present situation, where 16 s elapses between successive trials, may not generalize to the blocked-trial paradigm where only 2 s elapse between trials. Systematically varying the time between trials, or adding a secondary task that is performed between the experimental trials, may add insight into this possibility as well as the possibility that the response is in anticipation of the subsequent trial. A final alternative is counterintuitive—that there exist a class of neuronal responses that are sustained well after a trial, perhaps into subsequent trials, but are nonetheless directly related to the earlier trial. We

do not know of a precedent for such a neuronal response, but the possibility is consistent with our data.

It is somewhat disconcerting that differences in activation pattern were not clearly identified across trial type. The data revealed highly significant activation maps when the main effect of recognition was considered, suggesting that the data were stable and central tendencies across averaged subjects could be extracted. Nonetheless, comparisons between Hits and Correction Rejections yielded negative results. For the critical regions tested, the evoked hemodynamic responses across trial types were largely similar, suggesting that the negative result was not a failure attributable to weak statistical procedures but is rather attributable to the fact that the data were indeed similar across conditions.

We have previously failed to find differences across seemingly distinct retrieval trial types in both PET (Buckner *et al.*, 1996b) and fMRI (Schacter *et al.*, 1997) in studies where considerable power was present for detecting main effects. While it is too early to draw strong conclusions, these failed attempts may be providing boundary conditions for what we can and cannot detect with current functional neuroimaging procedures. In the present study, encoding of words consisted of a relatively long study list with comparatively little accompanying contextual information or variation in perceptual or conceptual details that might later provide additional "recollective" material that could be remembered: Each of the words was presented in isolation, in a similar-appearing font, for a comparatively short time, and subjects performed the same abstract versus concrete judgment for all of the items. Under such context-barren circumstances, the difference between old and new items may be subtle, and the recognition decision is driven primarily by familiarity (Mandler, 1980; Tulving, 1985; Gardiner, 1988). These factors may diminish the magnitude of the old/new recognition effect. We are currently exploring event-related yes/no recognition under conditions that maximize study context to further explore differential activation across trial types.

A number of additional directions for further exploration are also suggested by our results. Most importantly, while we have conducted a direct test of the retrieval success hypothesis, we have provided little insight into the nature of the processing changes that might lead to strategy shifts in blocked task paradigms. It would be worthwhile to show—in a *within-subjects* design—that modulation for two kinds of trial types can be observed in blocked task paradigms but not during mixed event-related task paradigms. Such an effect is suggested by the results across our two studies, by a previous fMRI study by Schacter *et al.* (1997) and also by an ERP study of episodic retrieval (Johnson *et al.*, 1997). However, no definitive fMRI data showing a significant interaction across paradigm type has yet been produced. Similarly, identifying task variables that modulate anterior prefrontal cortex activation during blocked task paradigms might shed insight into the role of these regions as they pertain to episodic retrieval and to cognitive processes more generally (e.g., Wagner *et al.*, submitted for publication).

Finally, a second, more general, goal of this study was to begin to characterize the degree of variability observed in the hemodynamic responses of individual subjects under event-related testing procedures such as those used here. Toward this end, we selected two regions—SMA and extrastriate cortex—that were consistently activated in all subjects and examined the similarity of the response profiles that were obtained in these two regions. We found that there was a high degree of similarity in both the timing and shape of the hemodynamic response curves and this similarity was observed both across different subjects (comparing curves for each of these areas separately) and within subjects (comparing curves for two conditions within subject). Although some caution is necessary in generalizing from these results because the procedure we employed considered only hemodynamic responses of a certain class (those with a single peak occurring several seconds after the stimulus onset) and so may have missed hemodynamic responses of a different shape, responses of the "delayed, single peak" form have been observed across a wide range of studies. These results thus provide evidence that, contrary to the reservations expressed by some researchers (Kim *et al.*, 1997), the assumption that there is across-subject convergence toward a modal hemodynamic response is justified—and thus averaging responses across subjects is both appropriate and meaningful. Nonetheless, further exploration of this question, particularly in paradigms employing individual trials that are more closely spaced than in the present study (cf. Buckner *et al.*, 1998b; Clark *et al.*, in press; Dale and Buckner, 1997), is also warranted, and it is possible that the development of new analysis and/or data collection procedures may allow reductions of both across-subject and within-subject sources of variance, further increasing the sensitivity and power of event-related procedures.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (NIDCD DC03245 and NIA AG08441), the Charles A. Dana Foundation, the McDonnell Center for Higher Brain Function, and the Human Frontiers Science Program.

REFERENCES

Bandettini, P. A., Jesmanowicz, A., Wong, E. C., and Hyde, J. S. 1993. Processing strategies for time-course data sets in functional MRI of the human brain. *Magn. Reson. Med.* **30:** 161–173.

Boynton, G. M., Engel, S. A., Glover, G. H., and Heeger, D. J. 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. *J. Neurosci.* **16:** 4207–4221.

Buckner, R. L. 1996. Beyond HERA: Contributions of specific prefrontal brain areas to long-term memory retrieval. *Psychonom. Bull. Rev.* **3:** 149–158.

Buckner, R. L., Bandettini, P. A., O'Craven, K. M., Savoy, R. L., Petersen, S. E., Raichle, M. E., and Rosen, B. R. 1996a. Detection of cortical activation during averaged single trials of a cognitive task using functional magnetic resonance imaging. *Proc. Natl. Acad. Sci. USA* **93:** 14878–14883.

Buckner, R. L., and Koutstaal, W. 1998. Functional neuroimaging studies of encoding, priming, and explicit memory retrieval. *Proc. Natl. Acad. Sci. USA,* **95:** 891–898.

Buckner, R. L., Koutstaal, W., Schacter, D. L., Wagner, A. D., and Rosen, B. R. 1998a. Functional-anatomic study of episodic retrieval using fMRI. I. Retrieval effort versus retrieval success. *NeuroImage* **7,**000–000.

Buckner, R. L., Goodman, J., Burock, J., Rotte, M., Koutstaal, W., Schacter, D., Rosen, B., and Dale, A. M. 1998b. Functional–anatomic correlates of object priming in humans revealed by rapid presentation event-related fMRI. *Neuron* **20:**285–296.

Buckner, R. L., Raichle, M. E., Miezin, F. M., and Petersen, S. E. 1996b. Functional anatomic studies of memory retrieval for auditory words and visual pictures. *J. Neurosci.* **16:**6219–6235.

Clark, V. P., Maisog, J. M., and Haxby, J. V. in press. FMRI studies of face memory using random stimulus sequences. *J. Neurophysiol.*

Dale, A. M., and Buckner, R. L. 1997. Selective averaging of rapidly presented individual trials using fMRI. *Hum. Brain Map.* **5:**329–340.

Gardiner, J. M. 1988. Functional aspects of recollective experience. *Memory Cog.* **16:** 309–313.

Hunton, D. L., Miezin, F. M., Buckner, R. L., van Mier, H. I., Raichle, M. E., and Petersen, S. E. 1996. An assessment of functional-anatomical variability in neuroimaging studies. *Hum. Brain Map.* **4:** 122–139.

Johnson, M. K., Nolde, S. F., Mather, M., Kounios, J., Schacter, D. L., and Curran, T. 1997. Test format can affect the similarity of brain activity associated with true and false recognition memory. *Psychol. Sci.* **8:** 250–257.

Josephs, O., Turner, R., and Friston, K. 1997. Event-related fMRI. *Hum. Brain Map.* **5:**243–248.

Kim, S. G., Richter, W., and Ugurbil, K. 1997. Limitations of temporal resolution in functional MRI. *Magn. Reson. Med.* **37:** 631–636.

Konishi, S., Yoneyama, R., Itagaki, H., Uchida, I., Nakajima, K., Kato, H., Okajima, K., Koizumi, H., and Miyashita, Y. 1996. Transient brain activity used in magnetic resonance imaging to detect functional areas. *NeuroReport* **8:**19–23.

Lee, A. T., Glover, G. H., and Meyer, C. H. 1995. Discrimination of large venous vessels in time-course spiral blood-oxygen-level-dependent magnetic-resonance functional neuroimaging. *Magn. Reson. Med.* **33:** 745–754.

MacLeod, A. K., Buckner, R. L., Miezin, F. M., Petersen, S. E., and Raichle, M. E. 1998. Right anterior prefrontal cortex activation during semantic monitoring and working memory. *NeuroImage* **7:** 41–48.

Mandler, G. 1980. Recognizing: The judgment of previous occurrence. *Psych. Rev.* **87:**252–271.

McCarthy, G., Luby, M., Gore, J., and Goldman-Rakic, P. 1997. Infrequent events transiently activate human prefrontal and parietal cortex as measured by functional MRI. *J. Neurophysiol.* **77:**1630–1634.

Richter, W., Georgopoulos, A. P., Ugurbil, K., and Kim, S.-G. 1997. Detection of brain activity during mental rotation in a single trial by fMRI. *NeuroImage* **5:** S49.

Rosen, B. R., Buckner, R. L., and Dale, A. M. 1998. Event related fMRI: Past, present, and future. *Proc. Natl. Acad. Sci. USA,* **95:**773–780.

Rugg, M. D., Fletcher, P. C., Frith, C. D., Frackowiak, R. S. J., and Dolan, R. J. 1996. Differential response of the prefrontal cortex in successful and unsuccessful memory retrieval. *Brain* **119:** 2073–2083.

Schacter, D. L., Buckner, R. L., Koutstaal, W., Dale, A. M., and Rosen, B. R. 1997. Late onset of anterior prefrontal activity during true and false recognition: An event-related fMRI study. *NeuroImage* **6:**259–269.

Talairach, J., and Tournoux, P. 1988. *Co-planar Stereotaxic Atlas of the Human Brain.* Thieme, New York.

Tulving, E. 1985. Memory and consciousness. *Can. Psychologist* **26:**1–12.

Tulving, E., Kapur, S., Markowitsch, H. J., Craik, F. I. M., Habib, R., and Houle, S. 1994. Neuroanatomical correlates of retrieval in episodic memory: Auditory sentence recognition. *Proc. Natl. Acad. Sci. USA* **91:** 2012–2015.

Wagner, A. D., Gabrieli, J. D. E., Desmond, J. E., Joaquim, S., and Glover, G. H. 1996. Prefrontal mediation of episodic memory performance. *Soc. Neurosci. Abstr.* **22:** 719.

Xiong, J., Gao, J.-H., Lancaster, J. L., and Fox, P. T. 1996. Assessment and optimization of functional MRI analyses. *Hum. Brain Map.* **3:**153–167.

Zarahn, E., Aguirre, G., and D'Esposito, M. 1997a. A trial-based experimental design for fMRI. *NeuroImage,* **6:**122–138.

Zarahn, E., Aguirre, G. K., and D'Esposito, M. 1997. Empirical analyses of BOLD fMRI statistics. I. Spatially unsmoothed data collected under null-hypothesis conditions. *NeuroImage* **5:** 179–197.