

Behavioral and neurophysiological correlates of episodic coding, proactive interference, and list length effects in a running span verbal working memory task

BRADLEY R. POSTLE

University of Wisconsin, Madison, Wisconsin

JEFFREY S. BERGER and JEREMY H. GOLDSTEIN

University of Pennsylvania Medical Center, Philadelphia, Pennsylvania

and

CLAYTON E. CURTIS and MARK D'ESPOSITO

University of California, Berkeley, California

Updating refers to (1) discarding items from, (2) repositioning items in, and (3) adding items to a running working memory span. Our behavioral and fMRI experiments varied three factors: trial length, proactive interference (PI), and group integrity. Group integrity reflected whether the grouping of items at the encoding stage was violated at discarding. Behavioral results were consistent with the idea that updating processes have a relatively short refractory period and may not fatigue, and they revealed that episodic information about group context is encoded automatically in working memory stimulus representations. The fMRI results did not show evidence that updating requirements in a task recruit executive control processes other than those supporting performance on nonupdating trials. They did reveal an item-accumulation effect, in which signal increased monotonically with the number of items presented during the trial, despite the insensitivity of behavioral measures to this factor. Behavioral and fMRI correlates of PI extended previous results and rejected an alternative explanation of PI effects in working memory.

The concept of “updating” in working memory research refers to the operations needed to modify the contents of working memory in response to the appearance of newer, superceding information in the environment. In the laboratory, updating has often been operationalized with variants of the running memory span task, in which participants are presented a serial presentation of items and must steadily update the contents of working memory in order to retain the n most recently presented items (e.g., “remember the four most recent items you have seen”; Kiss, Pisis, Francois, & Schopflocher, 1998; Morris & Jones, 1990; Salmon et al., 1996). Such a task requires at least three discrete processing steps: (1) discarding items from, (2) repositioning items in, and (3) adding items to

working memory. (By “repositioning” we mean changing the ordinal position of each item; for example, changing the item previously in the 2nd position of a 4-item list to the 1st position, the item previously in the 3rd position to the 2nd position, and so on.) These steps are also required in more complex working memory tasks, such as the n -back task, the processing components of which are considered by Jonides et al. (1997) and Postle, Stern, Rosen, & Corkin. (2000). Discarding and repositioning operations engage executive control processes not required within individual trials of simple tests of short-term memory span or duration, such as digit span and delayed-recognition tasks, respectively (D’Esposito & Postle, 1999, 2000). Indeed, tasks requiring updating of items held in working memory have often been used to assess the contributions of executive functioning to working memory (Kiss et al., 1998; Lehto, 1996; Morris & Jones, 1990; Salmon et al., 1996). This report presents the results of experiments investigating the behavioral and neural correlates of manipulating factors related to these three processes, as well as to proactive interference (PI), in a running span task.

In Experiment 1, we performed two behavioral studies in which three factors were varied. The factor of trial length controlled the number of items presented (and thus, the

This research was supported by the American Federation for Aging Research and by NIH Grants NS01762 and AG13483. B.R.P. received support from NIH Grant AG00255 to Virginia M.-Y. Lee (University of Pennsylvania). The authors thank Geoff Aguirre for consultation on fMRI analyses, Hugh Garavan, Rik Henson, Dylan Jones, Ivan Kiss, Sharon Thompson-Schill, and Eric Zarahn for helpful discussions of this work, and three anonymous reviewers for their careful critiques. Correspondence should be addressed to B. R. Postle, Department of Psychology, University of Wisconsin, Madison, WI 53706-1696 (e-mail: postle@facstaff.wise.edu).

- a. bf jc *R*
- b. c l py st fb jv mn *J*
- c. bnv qxl c r *G*
- d. m j cg fzp x *C*

Figure 1. Schematic illustration of four examples of trials in the running span task. Each row represents a trial. Grouping together of lowercase letters represents groups of memoranda presented simultaneously in a stimulus-presentation epoch. The uppercase letter bounded by asterisks represents the probe. (a) A nonupdating trial. (b) A trial requiring updating in which group integrity is preserved throughout the trial; the probe is valid. (c) A trial requiring updating in which group integrity is violated by two stimulus presentation epochs. This is a no-PI trial because the invalid probe does not match any memoranda presented during the trial. (d) A trial requiring updating in which the integrity of one group (*cg*) is violated by a subsequent stimulus presentation epoch (*fzp*) but in which group integrity is preserved in the two groups making up the final memory set (*fzp* and *x*). This is a PI trial, because a memorandum appearing during the trial matches the probe (the letter *c*) but is not a part of the final memory set.

amount of updating required) in each trial. The factor of PI reflected whether or not an invalid probe stimulus (i.e., a probe that did not match any item in the final memory set) matched an item that had been presented in that trial, but prior to the final four items of the trial. The factor of group integrity reflected the fact that our experimental design varied the grouping of items by presenting memoranda in groups of one, two, or three items per stimulus presentation epoch. For example, the letter string *bcdf* might be grouped as *b* and *cdf* or as *bc* and *df*, depending on how the items were presented. Results from a pilot study suggested that breaking up such groups at the discarding stage (Figure 1) had a deleterious effect on performance. We believed that the results of this manipulation might have bearing on our understanding of the retention in working memory of contextual information specific to the encoding episode. Guided by knowledge obtained from these behavioral studies, in Experiment 2 we employed functional magnetic resonance imaging (fMRI) to investigate the neural correlates of these experimental factors.

EXPERIMENT 1

Behavioral

Trial Length

The results of previous studies suggested that updating has an all-or-none effect on running span performance. That is, although immediate serial recall performance is significantly lower on trials that require updating than on trials that do not, performance is not further sensitive to the number of updates required in a trial (Morris & Jones, 1990). Specifically, in Morris and Jones, performance on

6-item, 8-item, and 10-item trials was equivalent. From these results, it was suggested that the control processes governing updating operations had a rapid recovery rate. Seemingly at odds with this behavioral result was the report that the magnitude of event-related potentials (ERPs) attributable to updating-related processes increased monotonically as a function of the number of updates required in a single trial (Kiss et al., 1998; behavioral data from this experiment were not suitable for assessment of possible performance correlates of this ERP effect [I. Kiss, personal communication]). The principal methodological difference between these two studies was that Morris and Jones employed an immediate serial recall procedure and Kiss et al. a yes/no recognition procedure. In the present experiment, therefore, we sought to assess the generality of the step function relating performance to updating demands (at least across a small range of trial lengths) on a task that employed a recognition procedure. We varied trial length (4, 8, or 12 items) in a running span task that required memory for the 4 most recently presented items. We considered only the latter two to require updating, because no repositioning or discarding operations were required on 4-item trials. Failure to replicate the all-or-none effect of updating on behavioral performance would call into question a fundamental assumption about the refractory rate of updating-related processes. A replication of this effect, in contrast, would highlight the divergence between behavioral and ERP correlates of updating.

PI

Previous studies of delayed item recognition have characterized behavioral and neurophysiological correlates of PI that manifests itself in “high-overlap” conditions when an item from a previous trial matches the probe on the subsequent trial (D’Esposito, Postle, Jonides, & Smith, 1999; Jonides, Marshuetz, Smith, Reuter-Lorenz, & Koeppel, 2000; Jonides, Smith, Marshuetz, Koeppel, & Reuter-Lorenz, 1998). This effect is believed to index the operation of a mechanism that resolves the ambiguity about the validity of the probe on high-overlap trials (Jonides et al., 2000). The behavioral studies presented here, in contrast to the earlier delayed item recognition studies, examined the effects of PI among items from the same trial. Our design also permitted us to investigate interactions between two putatively discrete (sets of) control functions that can contribute to working memory performance: updating and the resolution of PI.

Group Integrity

Presenting verbalizable memoranda in groups, as opposed to one at a time, has salutary effects on immediate serial recall (Ryan, 1969). Such grouping effects are believed to be represented at the level of the phonological loop (i.e., they are represented by verbal rehearsal; Hitch, Burgess, Towse, & Culpin, 1996) and may be encoded by internal timing signals that represent temporal order (Burgess & Hitch, 1999). Grouping effects in working memory, when induced externally, can be viewed as

Table 1
Itemization by Type of the 96 Trials Administered to
Each Participant in Experiments 1A and 1B

	Trial Length					
	4		8		12	
	Number of Stimulus Presentation Epochs		Number of Stimulus Presentation Epochs		Number of Stimulus Presentation Epochs	
Group Integrity	2	3	4	5	6	7
Valid-Probe Trials						
Preserved	8	8	4	4	4	4
Violated	n.a.	n.a.	4	4	4	4
Invalid-Probe Trials						
Preserved						
PI	n.a.	n.a.	2	2	2	2
No-PI	8	8	2	2	2	2
Violated						
PI	n.a.	n.a.	2	2	2	2
No-PI	n.a.	n.a.	2	2	2	2

Note—Trial length was determined by the number of items presented as memoranda.
n.a. = not applicable.

an influence of episode-specific information about context at encoding. We investigated the nature of this episodic code in verbal working memory by exploring what happens when the integrity of a group defined by temporal grouping of items at encoding is subsequently violated at the discarding component of the updating task.

Method

Participants. We tested a total of 25 healthy undergraduate volunteers in two experiments (Experiments 1A and 1B). Twelve, recruited and tested at the University of Pennsylvania (5 males, 7 females; mean age = 21.0 years), participated in Experiment 1A, and 15 (4 males, 11 females; mean age = 19.0 years), recruited and tested at the University of Wisconsin—Madison, participated in Experiment 1B. Informed consent was obtained from each participant prior to testing.

Procedure. Stimuli were drawn randomly for each trial from the set of 21 consonant letters. No items repeated as memoranda on a trial. Each stimulus presentation event (hereafter referred to as an *epoch*) presented one, two, or three lowercase letters for 500 msec. Items in two- and three-letter groups were presented simultaneously, in a horizontal row. Group size within a trial varied pseudo-randomly and was constrained by the group integrity factor (see below). Epoch onset asynchrony (EOA) was 4 sec. Trials presented a total of 4, 8, or 12 stimuli, followed by a probe stimulus. (Within each of these levels of the factor of trial length—defined by number of items—were two temporal lengths. See Table 1.) The participants were instructed to maintain a memory of the 4 most recently presented items and to press the “yes” (right) or the “no” (left) button to indicate whether or not the probe matched an item in the memory set (Figure 1). They were also instructed to respond as quickly and accurately as possible, but not to sacrifice accuracy for the sake of speed. Valid and invalid probes were equiprobable. Each participant was tested on 32 trials of each length, which occurred in a randomly determined order.

PI was manipulated on 8- and 12-item trials featuring invalid probes (32 trials per participant) by the presence or absence of a lure item that matched the probe but appeared among the first 4 items on 8-item trials or among the first 8 items on 12-item trials (Figure 1). Item-specific PI across adjacent trials was not studied.

Group integrity was manipulated on all 8- and 12-item trials (i.e., on all trials that required updating). On one half of all such trials,

the final stimulus presentation epoch required the breakup of a group (Figure 1). Such violations of group integrity could also occur unpredictably in portions of the trial preceding the final stimulus presentation epoch, so that these events would not be predictive of the end of the trial.¹

A comprehensive accounting of the exact numbers of each trial type is provided in Table 1.

Results: Experiment 1A

The data revealed an effect of length consistent with the hypothesized all-or-none effect of updating on performance and revealed a robust PI effect. There was no evidence of a reliable effect of group integrity, assessed across invalid probes. The planned analysis of group integrity collapsed across all 8- and 12-item trials was not performed, due to an interaction of PI with group integrity in the reaction time (RT) data, which signaled the possibility of a speed–accuracy tradeoff in performance.

Trial length. An analysis of variance (ANOVA) assessing accuracy data from the 12 participants yielded a main effect of length (Length 4, $M = 97.7\%$, $SE = 0.6$; Length 8, $M = 90.0\%$, $SE = 1.8$; Length 12, $M = 88.5\%$, $SE = 2.2$) [$F(2,22) = 13.9$, $p < .0001$]. The planned comparison confirmed that this effect was carried by the predicted all-or-none effect of updating on running span performance: Performance at Length 4 was significantly better than performance at Lengths 8 and 12 [$t(11) = 4.3$, $p < .001$] (Figure 2). Although mean RT values increased monotonically with trial length (Length 4, $M = 1,538.5$ msec, $SE = 92.9$; Length 8, $M = 1,599.3$ msec, $SE = 97.0$; Length 12, $M = 1,642.3$ msec, $SE = 161.5$), this effect was not reliable [$F(2,22) = 0.76$, n.s.] (Figure 2).

PI and group integrity. These data assessed accuracy and RT data from updating trials with invalid probes [accuracy: PI/violated, $M = 79.2\%$, $SE = 5.8$; no-PI/violated, $M = 96.9\%$, $SE = 1.6$; PI/preserved, $M = 86.5\%$, $SE = 2.9$; no-PI/preserved, $M = 99.0$, $SE = 1.0$; RT: PI/violated, $M = 1,898.0$ msec, $SE = 128.1$; no-PI/violated,

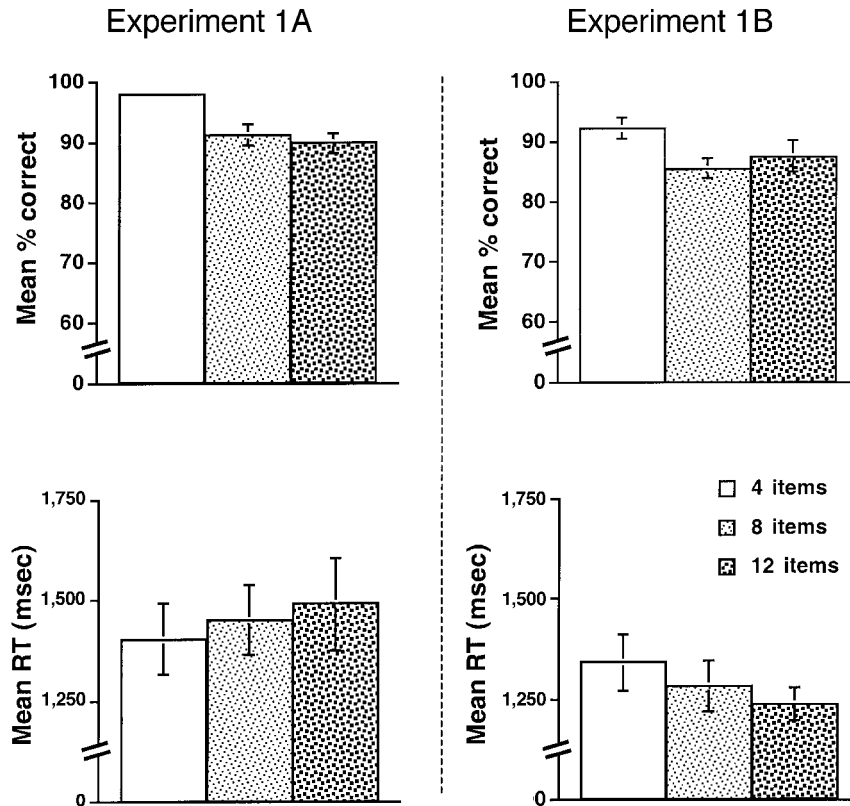


Figure 2. Trial-length data of Experiments 1A and 1B. Top: Accuracy data illustrating the all-or-none effect of updating on performance (on this and on all subsequent graphs, bars represent *SEs*). Bottom: RT data show a numerical, but nonsignificant, increasing trend in Experiment 1A and show the opposite trend in Experiment 1B. This difference may be explained in part by the absence of within-trial PI in Experiment 1B.

$M = 1,454.0$ msec, $SE = 85.2$; PI/preserved, $M = 1,689.2$ msec, $SE = 105.3$; no-PI/preserved, $M = 1,435.7$, $SE = 99.0$) (Figure 3). The two-factor ANOVA assessing accuracy revealed a main effect of PI [$F(1,11) = 19.18$, $p < .001$], no effect of group integrity [$F(1,11) = 1.9$, $p = .19$], and no interaction [$F(1,11) = 0.53$, n.s.] (Figure 3, top).² The two-factor ANOVA assessing RT data from these same trials in the same participants revealed a main effect of PI [$F(1,11) = 12.87$, $p < .0005$], no effect of group integrity [$F(1,11) = 2.9$, $p = .11$], and an interaction [$F(1,11) = 7.12$, $p < .05$] (Figure 3, bottom).

Because the two-factor ANOVAs described in the previous paragraph were restricted to updating trials with invalid probes, we had planned to increase our power to detect effects of group integrity by performing paired t tests on data from all 8- and 12-item updating trials (i.e., including trials with valid and invalid probes). The appropriateness of collapsing across levels of PI was questionable, however, due to a possible speed-accuracy tradeoff in performance: Although there was no PI \times group integrity interaction in the accuracy data, these two factors did interact in the RT data.

Thus, in order to investigate the group integrity effect in an unequivocal way, we performed a second experiment (Experiment 1B), in which no trials featured within-trial PI. Experiment 1B was identical in design to Experiment 1A in all other respects.

Results: Experiment 1B

The accuracy results replicated the trial-length effect from Experiment 1A, whereas the effect of trial length on RT was, unexpectedly, in the opposite direction from Experiment 1A. A reliable effect of group integrity in Experiment 1B was evident in both the accuracy data and the RT data.

Trial length. The accuracy results from these 15 participants (Length 4, $M = 92.1\%$, $SE = 1.7$; Length 8, $M = 85.4\%$, $SE = 1.6$; Length 12, $M = 87.5\%$, $SE = 2.7$) replicated those from Experiment 1A, with a main effect of length [$F(2,28) = 4.83$, $p < .05$] and performance at Length 4 significantly better than performance at Lengths 8 and 12 [$t(14) = 3.2$, $p < .01$] (Figure 2). The RT results (Length 4, $M = 1,342.1$ msec, $SE = 70.0$; Length 8, $M = 1,282.9$ msec, $SE = 64.2$; Length 12, $M = 1,237.9$ msec,

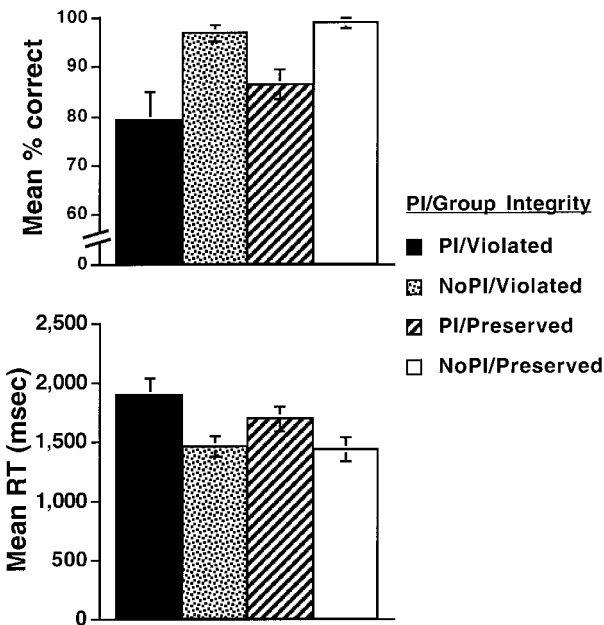


Figure 3. PI \times group integrity data of Experiment 1A. **Top:** Accuracy data from 8- and 12-item invalid-probe trials, illustrating a main effect of PI, no effect of group integrity, and no interaction. **Bottom:** RT data from the same trials, illustrating a main effect of PI, no effect of group integrity, and an interaction.

$SE = 41.7$), in contrast, differed markedly from those of Experiment 1A, decreasing monotonically with trial length [$F(2,28) = 5.14, p = .01$] (Figure 2).

Group integrity. The group integrity data revealed significantly lower accuracy on trials violating group integrity (violated, $M = 84.8\%$, $SE = 2.3$; preserved, $M = 89.4\%$, $SE = 2.0$) [$t(14) = 2.2, p < .05$] (Figure 4, top) and longer RTs on these trials (violated, $M = 1,263.2$ msec, $SE = 49.7$; preserved, $M = 1,200.5$ msec, $SE = 51.8$) [$t(14) = 3.0, p < .01$] (Figure 4, bottom).

Discussion

Two experiments failed to demonstrate a monotonic effect of list length on accuracy, a result that is consistent with the suggestion that, across a limited number of trial lengths, the effect on performance of updating is all or none, not cumulative (Morris & Jones, 1990). These results indicate that this property generalizes across testing procedures. This is consistent with the idea that updating-related processes have a refractory period that is shorter than 4,000 msec (Morris & Jones, 1990)³ and that they do not decline in efficiency with repeated use. Of course, what appears to be a step function when plotted across only three data points may be only a nonlinear beginning segment of a function that may contain a monotonically decreasing segment. RTs associated with another putative executive process—attention switching among representations in working memory (Garavan, 1998)—increased monotonically as list length increased from 2 to

24 items (H. Garavan, personal communication, July 2000). Thus, additional studies of the running span task that feature longer trials will be required to more completely characterize the effect of trial length on updating-related processes. Nonetheless, the fMRI study presented as Experiment 2 was designed to assess whether the hemodynamic correlate of list length would echo the step function suggested by our accuracy data, whether it would increase monotonically with list length (as does the ERP correlate of list length; Kiss et al., 1998) or whether it would demonstrate a novel relationship with trial length.

Interpretation of the RT data is complicated by the reversal in sign, between Experiments 1A and 1B, of the slope of the line characterizing RT as a function of list length (Figure 2). This may be explained in part by the absence of intratrial PI in Experiment 1B. Perhaps the unpredictable presence of intratrial PI in Experiment 1A, which could occur only at longer length trials, induced a strategy of greater caution on longer trials. It also bears restating that, although the participants were instructed to respond as quickly as possible, they were also instructed not to sacrifice accuracy in doing so. Thus, the RT data might provide a less reliable index of updating-related processes than the accuracy data.

The robust effect of (intratrial) PI on accuracy and RTs, mirroring those from previous studies of tasks that generated PI across trials (D'Esposito, Postle, Jonides, & Smith, 1999; Jonides et al., 2000; Jonides et al., 1998), indicated that we could reasonably predict an fMRI correlate of PI in the left inferior prefrontal cortex (PFC), mirroring neuroimaging results from those previous studies.

A novel finding from Experiment 1 was the group integrity effect. The sensitivity of updating performance to violations of group integrity suggests that episodic information about group context at encoding, although irrelevant to task demands (and, indeed, deleterious to performance), is encoded automatically as a feature of stimulus representations in working memory. We will consider further the implications of this result in the General Discussion.

EXPERIMENT 2 fMRI

Our fMRI method featured an event-related technique that permitted us to assess fMRI signal associated with each stimulus presentation epoch, uncontaminated by variance in the signal arising from epochs preceding or following the epoch in question (Postle, Zarahn, & D'Esposito, 2000; Zarahn, Aguirre, & D'Esposito, 1997b). We could thus examine directly the evolution of the fMRI signal across the seven or eight stimulus presentation epochs associated with a 12-item trial and the probe-evoked signal from PI versus no-PI trials. In Experiment 2, we addressed several basic questions about updating that could not be addressed with behavioral studies. Among these, we investigated whether the processes supporting updating can be dissociated neurally from nonupdating processes

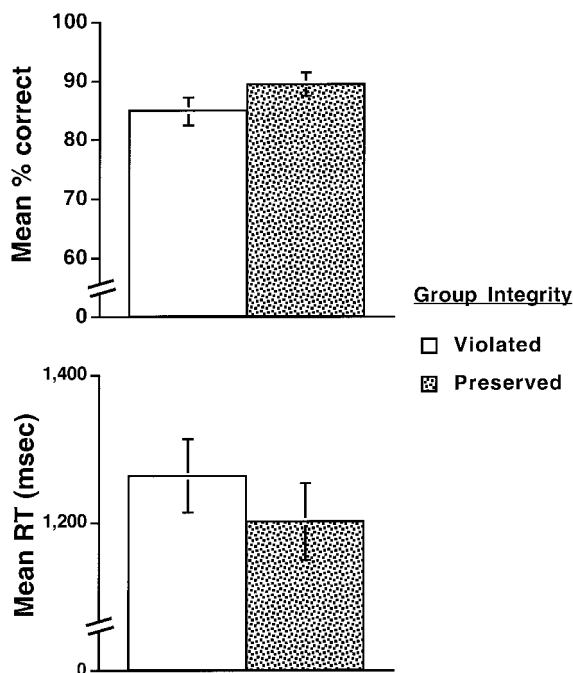


Figure 4. Group integrity data in Experiment 1B (an experiment that featured no within-trial PI). Top: Accuracy data from all 8- and 12-item trials, revealing significantly lower scores on trials violating group integrity. Bottom: RT data from the same trials, illustrating a significantly longer RTs on trials violating group integrity.

(i.e., stimulus encoding and maintenance) also engaged by our task. An affirmative result would provide evidence consistent with models positing fundamental dissociability of updating-specific from nonupdating processes. Neural dissociability is a sufficient, although not necessary, feature of the relationship between two independent mental processes.⁴ Other examples of hypothesized extramnemonic executive control processes are those required to manipulate items in working memory. For example, the processes recruited to alphabetize randomly ordered letter stimuli in working memory are supported by PFC voxels that do not support maintenance of those same items (Postle, Berger, & D'Esposito, 1999). And if updating and nonupdating processes are neurally dissociable, are updating-related processes supported by a relatively circumscribed region of PFC, as are alphabetization-related processes (D'Esposito, Postle, Ballard, & Lease, 1999; Postle et al., 1999)? The fMRI experiment was also expected to provide additional information about two of the factors that we investigated in depth in Experiment 1: trial length and PI.

Trial Length

Experiment 1 demonstrated the generalizability across testing procedures of the all-or-none effect of updating on performance and highlighted the divergent results of studies of performance (Morris & Jones, 1990) and studies of ERP measures of updating (Kiss et al., 1998), the latter

being positively correlated with the number of updates in a trial. Would fMRI data, like ERPs, reveal a process that corresponds to the accumulation of items within a trial and that is not detectable by behavioral methods? To address this question, we obtained the appropriate behavioral and physiological measures in the same experimental session.

PI

Recently, it has been proposed that the neuroimaging effects obtained in high-overlap conditions of the delayed item recognition task (D'Esposito, Postle, Jonides, & Smith, 1999; Jonides et al., 2000; Jonides et al., 1998) may not reflect the operation of a mechanism that resolves PI (Jonides et al., 2000) but rather the fact that high-overlap conditions effectively increase memory load (Bunge, Matsumoto, Desmond, Glover, & Gabrieli, 2000). That is, a span of four items maintained under high-overlap conditions may be treated by the nervous system as a span of five or more items because of intrusion into the memory set by proactively interfering items. Although Experiment 2 did not vary memory load, it did vary trial length, permitting us to assess these two alternatives by examining whether the probe-specific fMRI sensitivity to trial length differed qualitatively (as would be predicted by Jonides et al., 2000) or quantitatively (as would be predicted by Bunge et al., 2000) from probe-specific fMRI sensitivity to trials featuring PI. Specifically, the memory-load hypothesis (Bunge et al., 2000) would predict that the distribution and activity patterns of voxels sensitive to trial length would resemble those of voxels sensitive to high-overlap conditions (which we are referring to as the PI condition). The resolution-of-PI hypothesis (Jonides et al., 2000), in contrast, would predict that PI effects would be focused within left Brodmann's area (BA) 45, whereas trial-length effects would be more broadly distributed, particularly in dorsolateral PFC (Rypma & D'Esposito, 1999).

Group Integrity

Although the results of Experiment 1A indicated that a speed-accuracy tradeoff may have masked a possible interaction of the factors of PI and group integrity, we were interested in testing the two mutually exclusive explanations of the PI effect (see previous paragraph). Thus, we chose to keep the PI manipulation in our fMRI study (Experiment 2) at the expense of generating interpretable group integrity data. Thus, although the behavioral procedure for Experiment 2 was closely modeled after that used in Experiment 1A, we do not report the fMRI correlates of group integrity.

Power Considerations

One way in which fMRI and behavioral studies differ from one another is the markedly higher costs (financial, time, human resources) of the former. This unavoidable intrusion of an extrascientific factor into the experimental design process typically induces investigators to include fewer participants in an fMRI study than in a behavioral study. Our approach on this project was to first perform a

series of behavioral studies whose statistical power would permit adequate characterization of the theoretical phenomena of interest, and then to investigate the neurophysiological correlates of these phenomena in a smaller group of participants. This approach made the assumption that the psychological processes engaged by the tasks featured in the behavioral studies would be similarly engaged during the fMRI study, because the latter featured the same behavioral tasks as the former.⁵ This assumption was critical, because the smaller N in the fMRI experiment was expected to preclude detection, at a statistically reliable level, of the behavioral effects corresponding to these psychological processes. Also critical to this approach was the fact that the fMRI data from each participant were expected to feature sufficient power to detect, within individual participants, two of the neurophysiological effects of interest (updating and list length). Previous event-related fMRI studies in our laboratory that employed similar designs have featured upward of 1,100–1,200 effective degrees of freedom (Postle, Zarahn, & D'Esposito, 2000). We planned, therefore, to treat the fMRI results from each participant as a case study. For data such as these, an N of between 5 and 10 participants can suffice to assess the reliability of an effect across participants. Our previous experience with investigations of PI, however, indicated that the fMRI correlate of this effect was too subtle to investigate with the case-study approach (D'Esposito, Postle, Jonides, & Smith, 1999). Nonetheless, the N of 7 in that study sufficed to demonstrate a reliable, topographically specific effect of PI in the fMRI data.

Method

Participants. Five healthy adults (3 males, 2 females; mean age = 19.8 years) recruited from the undergraduate and medical campuses of the University of Pennsylvania, none of whom participated in Experiment 1, participated in Experiment 2. Additionally, 2 healthy adults recruited from the undergraduate campus of the University of California–Berkeley (1 male and 1 female; 23 and 21 years of age, respectively) also participated. Informed consent was obtained from each participant prior to testing.

Behavioral procedure. Stimulus presentation parameters and instructions were the same as those in Experiment 1, with the exception that stimulus exposure duration was increased to 1 sec, to ensure adequate encoding. Scanning was conducted during eight behavioral blocks, each 6 min 48 sec long and containing 12 trials. Trial type was balanced for length, PI status, and number of integrity-violating versus integrity-preserving epochs within each block. Trial order was pseudorandomized. Each trial was separated by an intertrial interval of 12 sec (Berkeley) or 14 sec (Pennsylvania).

Data acquisition. The fMRI scanning was conducted at the University of Pennsylvania with a GE Signa 1.5 T scanner equipped with a fast gradient system for echoplanar imaging. High-resolution sagittal and axial T1-weighted images were obtained in every participant, and a gradient echo, echoplanar sequence (TR = 2,000 msec, TE = 50 msec) was used to acquire data sensitive to the blood oxygen level dependent (BOLD) signal. Whole-brain images were acquired in twenty-one 5-mm-thick axial slices. The fMRI scanning was conducted at the University of California–Berkeley with a Picker 1.5 T scanner equipped with a fast gradient system for echoplanar imaging and employing all the same image acquisition parameters with the exception of slice number (20) and TE (40 msec). For all participants, scans of the working memory task were preceded by a scan in which we derived the hemodynamic response function (HRF) for each par-

ticipant (Aguirre, Zarahn, & D'Esposito, 1998). The HRF, which characterizes the fMRI response resulting from a brief impulse of neural activity (Boynton, Engel, Glover, & Heeger, 1996), was used to convolve independent variables entered into the modified general linear model (GLM) (Worsley & Friston, 1995) that we used to analyze the results of the scans of our working memory task.

Data processing. The fMRI time series data were filtered to remove high (>0.244 Hz) and low (<0.05 Hz) frequencies and were adjusted to remove the effects of nuisance covariates (Friston, Holmes, Poline, Heather, & Frackowiak, 1995). The principle of the fMRI time series analysis was to model the fMRI signal changes evoked by each stimulus presentation epoch with a covariate shaped like the HRF (Postle, Zarahn, & D'Esposito, 2000; Zarahn et al., 1997b). The smoothness of the fMRI response to neural activity allows evoked responses that arise from temporally dependent events to be resolved on the order of 4 sec (Zarahn et al., 1997b). Thus, because different event types in our experiment occurred in a temporally dependent manner, we selected the 4-sec EOA so that we could estimate the fMRI response evoked by each stimulus presentation epoch. This approach permitted us, for example, to contrast updating- and nonupdating-related activity (see section on "Updating versus nonupdating activity," further along in the Method section) or to contrast probe-related activity on trials featuring PI and probe-related activity on trials that did not feature PI. The least-squares solution of the GLM of the fMRI time series data yielded parameter estimates that were associated with each covariate of interest. Differences in fMRI signal (either between conditions or vs. baseline) were tested by computing t statistics resulting from linear combinations of the covariates in question.

Controlling Type I error. Our procedure for fMRI analysis is a "massively univariate" one in which the GLM is solved, independently and in parallel, in each voxel in the hypothesis-testing space. To control the omnibus false-positive rate of each statistical test, we applied Bonferroni correction, which has been demonstrated to provide an almost exact correction (to $p \leq .05$) for small voxel-wise false-positive rates in spatially unsmoothed data (Zarahn, Aguirre, & D'Esposito, 1997a). With these methods, the sensitivity of any hypothesis-testing contrast was inversely related to the number of voxels in the hypothesis-testing space. Thus, we maximized the sensitivity of each contrast by restricting each to regions of interest (ROIs) that were selected a priori (Postle, Zarahn, & D'Esposito, 2000).

ROIs. Our analyses were performed with anatomically defined ROIs. ROIs for dorsolateral and ventrolateral PFC were created by first defining them on the "canonical" representation of a brain in Talairach space that is provided in SPM96b, using the atlas of Talairach and Tournoux (1988) to confirm our identification of anatomical landmarks. (The dorsolateral PFC ROI corresponded to BAs 9 and 46 of the middle and inferior frontal gyri, and the ventrolateral PFC ROI to BAs 44, 45, and 47 of inferior frontal gyrus; we also created a BA 45 ROI.) Next, we transformed these ROIs from Talairach space into the native space in which each participant's data had been acquired by applying the 12-parameter affine transformation (Friston, Ashburner, et al., 1995) with nonlinear deformations (Ashburner & Friston, 1996), routine in SPM96b (effectively, a "reverse normalization"). Because some individual anatomical variability is not accounted for in the reverse normalization process, we adjusted the ROIs after transformation to better correspond to the anatomical images of each participant so that they would cover perfectly the intended brain regions. Our analyses focused on PFC because this region is widely accepted as a critical neural substrate of extramnemonic executive control processes that are recruited to support working memory performance. PI analyses focused on BA 45 (D'Esposito, Postle, Jonides, & Smith, 1999). All ROIs were also broken down by hemisphere for some analyses.

Updating versus nonupdating activity. We determined whether the processes supporting updating can be dissociated neurally from nonupdating processes with the contrast [updating – nonupdating],

where nonupdating epochs were defined as the initial epochs of the trial that did not require a discarding operation. Any voxels demonstrating significantly greater updating activity than nonupdating activity could be classified, within the context of this experiment, as specialized for updating. Once voxels had been so classified, we planned to assess quantitatively the level of nonupdating activity in updating-specialized voxels (with the contrast [nonupdating – baseline]), and vice versa. These two contrasts were not mathematically independent, because of the way in which the interrogated voxels had been defined. We felt that they would nevertheless provide interpretable initial indices of the degree of topographical overlap of updating- and nonupdating-related activity.

A second type of quantitative analyses, normalized voxel counts, were employed to assess the spatial extent of nonupdating- and updating-specialized activity (e.g., Is more cortex in region A than in region B recruited by updating-specialized operations?). In particular, these would permit us to assess the relative extent of particular types of activity in dorsolateral versus ventrolateral PFC, the former region being associated with a privileged role in many non-maintenance processes (D'Esposito, Postle, Ballard, & Lease, 1999; Petrides, 1989; Rowe, Toni, Josephs, Frackowiak, & Passingham, 2000). To perform these analyses, we first normalized the number of updating-specialized voxels within each ROI of each participant by dividing the number of suprathreshold voxels by the number of voxels in the ROI. Then, we assessed mean differences by ROI with paired *t* tests. Critical to the validity of this analysis (which is treated in more detail in Postle & D'Esposito, 2000) was our ability to define objectively the threshold for determining the significance of updating-specific activity on a voxel-by-voxel basis in our unsmoothed data.

The two principal factors of interest. Trial-length effects were assessed by inspection of trial-averaged time series data from 12-item 6-epoch trials and from 8-item 4-epoch trials (these were the two trial types for which the epoch on which updating processes were first engaged could be determined unambiguously). PI effects were assessed by identifying voxels in BA 45 of the left inferior PFC that demonstrated invalid probe-evoked activity, as assessed with the contrast [$\text{probe}_{\text{invalid, PI}} + \text{probe}_{\text{invalid, no-PI}}$], extracting the spatially averaged time series from these voxels and interrogating this time series with the orthogonal two-tailed contrast [$\text{probe}_{\text{invalid, PI}} - \text{probe}_{\text{invalid, no-PI}}$]. The result of this contrast was a *t* value, a noise-normalized index of the PI effect (Postle, Zarahn, & D'Esposito, 2000). A positive value would indicate greater invalid probe-evoked activity on trials featuring PI, and a negative value would indicate greater invalid probe-evoked activity on trials that did not feature PI. In the event that we found reliably a PI effect in left BA 45 (i.e., across participants), the same procedure would be applied to other PFC areas (right BA 45, and left and right dorsolateral PFC) to assess the anatomical specificity of this effect (D'Esposito, Postle, Jonides, & Smith, 1999).

Additionally, we planned to assess the memory-load alternative explanation of the PI effect (Bunge et al., 2000) by also assessing trial-length effects in each of these four ROIs. This would be implemented by identifying probe-sensitive voxels with the contrast [$\text{probe}_{\text{all}} - \text{baseline}$] and interrogating them with the two-tailed contrast [$\text{probe}_{12\text{-item}} - \text{probe}_{4\text{-item}}$]. Note that these analyses measured an effect that has been attributed to memory scanning (Rypma & D'Esposito, 1999), an effect that is unrelated to the item-accumulation effect despite the fact that both relate to the independent variable trial length.

Results

Behavior. We did not perform inferential statistical tests on these data because of the small *N* of this experiment. Overall mean accuracy was 91.5% ($SD = 5.8$) correct. Accuracy on trials that did not require updating

($M = 95.6\%$, $SD = 3.6$) was higher than on trials requiring updating ($M = 89.4\%$, $SD = 7.4$). Mean accuracy by trial length was 96.7%, 95.6%, and 91.9%, for 4-, 8-, and 12-item trials, respectively. Accuracy on invalid trials that did not feature PI ($M = 94.7\%$, $SD = 9.1$) was superior to that of invalid trials featuring PI ($M = 88.0\%$, $SD = 13.0$). Mean RTs increased with trial length (4 items, $M = 1,286.9$ msec, $SD = 139.4$; 8 items, $M = 1,338.3$ msec, $SD = 241.4$; 12 items, $M = 1,399.7$ msec, $SD = 214.5$). Mean RTs were shorter on invalid trials that did not feature PI ($M = 1,420.2$ msec, $SD = 266.5$) than on those that did feature PI ($M = 1,629.6$ msec, $SD = 342.5$), yielding a group mean PI effect of 209.3 msec ($SD = 118.2$). One participant (W) did not show an appreciable PI effect (1.7 msec).

fMRI. Voxels displaying sensitivity to nonupdating portions of the task (i.e., encoding and maintenance), as well as voxels demonstrating sensitivity to portions of the task requiring updating, were identified in many brain regions, including PFC, superior frontal areas 8 and 6, anterior cingulate cortex, posterior parietal cortex, inferior temporal cortex, and occipital cortex, a result broadly consistent with a previous PET study (Salmon et al., 1996). In all these brain regions, updating-related activity was more extensive than nonupdating-related activity.

Normalized voxel count analyses indicated that nonupdating-sensitive voxels (identified with the contrast [nonupdating – baseline]) were identified to a comparable extent in dorsolateral and ventrolateral PFC, a result that was consistent with several previous fMRI studies of working memory for letters (Postle & D'Esposito, 2000; Rypma & D'Esposito, 1999). The majority of these voxels also demonstrated sensitivity to updating portions of the trial, and only a very few voxels demonstrated greater nonupdating-related activity than updating-related activity. Updating-sensitive voxels (identified with the contrast [updating – baseline]) were also identified in dorsolateral and ventrolateral PFC. A group mean of 43.3% of all updating-sensitive voxels in PFC demonstrated significantly greater updating activity than nonupdating activity and thus could be classified as specialized for updating-related processes (see Method). These voxels, too, were found in both dorsolateral and ventrolateral PFC, and the normalized voxel count analysis revealed no difference in the extent of updating-specialized voxels in these two ROIs (dorsolateral PFC, $M = 14.5\%$, $SE = 2.2$; ventrolateral PFC, $M = 16.5\%$, $SE = 3.8$) [$t(6) = 0.8$; n.s.]. Quantitative assessment of comparative intensity of updating- versus nonupdating-related activity in PFC revealed that nonupdating-sensitive voxels also demonstrated robust updating-related activity (mean $t = 24.5$, $SE = 1.6$), and updating-sensitive voxels also demonstrated robust nonupdating-related activity (mean $t = 10.2$, $SE = 1.6$). This result confirmed our impression that the difference between updating- and nonupdating-related activity was quantitative, not qualitative, thereby suggesting that updating and nonupdating processes are not highly segregated at the neural level.

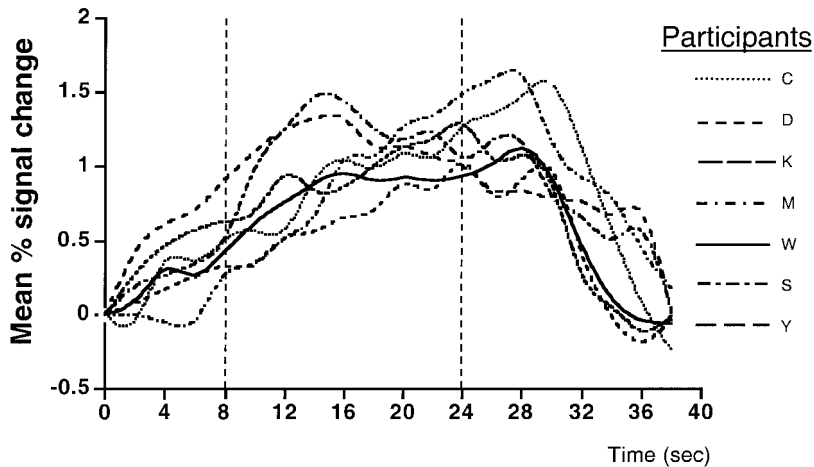


Figure 5. The fMRI time series data from each participant in Experiment 2, illustrating the item-accumulation effect. Data are spatially averaged across PFC updating-sensitive voxels and are trial-averaged across 12-item 6-epoch trials. Vertical dotted lines bound the portion of the trial during which updating operations were required. The probe stimulus appeared at time 28 sec.

Trial length. Examination of the time courses of updating-sensitive voxels revealed an item-accumulation effect in each participant: an increase of fMRI signal with each increase in the number of items presented on a trial (Figure 5). The slope of the linear regression line fit to the group data was .037, and the y intercept was .342.

PI. Each of the 6 participants whose RT data showed a PI effect also showed a positive fMRI PI effect in left BA 45, a reliable effect [$t(5) = 2.6, p < .05$] (Table 2). The 7th participant (W), whose behavioral PI effect was negligible (1.7 msec), also showed only a negligible positive fMRI PI effect in left BA 45 [$t(1409) = 0.08$]. PI effects did not approach significance in right BA 45, left dorsolateral PFC, or right dorsolateral PFC of these same 6 participants (Table 2).

The trial-length effect, in contrast with the PI effect, was stronger in dorsolateral ROIs than in ventrolateral ROIs and was significant in left dorsolateral PFC [$t(5) = 4.4, p < .01$] (Table 2). Thus, the PI effect and the trial-length effect differed qualitatively.

Discussion

Updating engaged broad extents of dorsolateral and ventrolateral PFC, as well as of other brain regions. This

is consistent with previous PET results with a similar task (Salmon et al., 1996) and with fMRI results from the *n*-back task (Braver et al., 1997; D'Esposito, Aguirre, Zarahn, & Ballard, 1998; Nystrom et al., 2000; Postle, Stern, Rosen, & Corkin, 2000; Schumacher et al., 1996), which is believed to engage updating processes. Updating processes were associated with greater signal intensity than initial (i.e., nonupdating) encoding and maintenance processes, in both dorsolateral and ventrolateral PFC. But updating voxels were identified in equivalent proportions in dorsolateral and ventrolateral PFC. Additionally, updating voxels, on average, also demonstrated robust sensitivity to nonupdating portions of the task, and the converse was also true for voxels identified with the nonupdating contrast. Thus, we did not find evidence that the neural substrates of updating-related processes are anatomically discrete from the strictly mnemonic portions of the running span task. Although fMRI dissociability is not a necessary feature of two independent cognitive processes, its demonstration would be strong evidence for independence. For example, the present results differ from those of at least two previous neuroimaging studies that sought explicitly to dissociate neurally executive control functions contributing to working memory performance from maintenance functions. In

Table 2
Mean Group PI and Trial-Length Effects ($n = 6$),
Indexed as t Values, by PFC ROI, from Experiment 2

Analysis	ROI			
	BA 45		PFC	
	Left	Right	Left Dorsolateral	Right Dorsolateral
PI	1.92 (0.7)*	1.29 (1.1)	0.96 (0.5)	0.58 (0.8)
Trial length	1.17 (0.89)	1.0 (1.0)	1.6 (0.4)*	1.24 (0.7)

Note—Standard errors of the mean are presented in parentheses. * $p < .0$.

one study, alphabetization-specific activity (an operationalization of the concept of manipulation of items in working memory) was observed primarily in dorsolateral PFC and was doubly dissociated from activity that was sensitive to manipulations of the mnemonic components of the task (Postle et al., 1999). This result was interpreted as evidence that alphabetization-related processes are fundamentally different from the processes supporting delayed item recognition when no manipulation is required. In a second study, attention switching dissociated from encoding and maintenance processes in a running count task, a dissociation interpreted as evidence for central executive functioning (Garavan, Ross, Li, & Stein, 2000). The results from the present study, however, do not rule out the possibility that updating requirements in a task may simply recruit, to a greater degree, the same processes that are recruited to support performance on simpler working memory tasks.

Trial length. The fMRI data from updating epochs revealed an item-accumulation effect: Signal increased linearly with the number of items presented during the trial. This result is at variance with behavioral data from this and other experiments but is consistent with ERP data (Kiss et al., 1998). We characterize this as an item-accumulation effect, rather than a PI effect, because there is no evidence that the accumulation of items interferes with behavior.⁶ Thus, this effect differs importantly from “load-sensitivity” effects in which increased working memory load affects both performance and neuroimaging signal (e.g., Braver et al., 1997; Jonides et al., 1997; Postle et al., 1999).

Because the longest trials in which we observed this item-accumulation effect were relatively short (presenting a total of 12 items and just six updating operations), we do not know whether, at longer list lengths, it may saturate in an item-dependent (or a time-dependent) manner and whether it may begin interfering with task performance. The answers to these questions may help determine whether the item-accumulation effect reflects the operation of a PI-related process (Kiss et al., 1998), the operation of previously undiscovered updating-related processes, or simply the epiphenomenal buildup of a by-product of updating activity, analogous to the buildup of lactic acid with extended anaerobic operation of a muscle.

PI. The fMRI correlates of the probe-related intratrial PI effect were found reliably only in BA 45—and thus were anatomically specific—and were consistent with the results obtained with intertrial PI in delayed item recognition tasks (D’Esposito, Postle, Jonides, & Smith, 1999; Jonides et al., 2000; Jonides et al., 1998). This suggests that this physiological correlate of the PI effect may generalize to tasks in which PI is controlled within trials, as well as across trials. Importantly, the analyses of the effect of trial length on probe-related activity were inconsistent with the memory-load alternative explanation of the PI effect (Bunge et al., 2000), because PI and

trial-length effects differed qualitatively from each other, not just quantitatively. Specifically, only left BA 45 showed a reliable PI effect in the fMRI data, whereas probe-related activity that was sensitive to trial length was numerically stronger in the dorsolateral PFC ROIs and was reliable only in left dorsolateral PFC. This neurophysiological double dissociation of PI and trial-length effects is inconsistent with the idea that these two effects reflect the operation of the same underlying mechanisms. The locus of the trial-length effect in the present study—which may index memory-scanning-related processes—is consistent with the results of a previous study that examined the fMRI correlates of memory scanning directly (Rypma & D’Esposito, 1999).

GENERAL DISCUSSION

The experiments described in this report contain many novel features. They are the first to explore the role of group integrity in working memory, the first to collect suitable high temporal resolution behavioral and physiological measures of updating in the same experiment, and the first to effect detailed quantitative comparisons of neuroimaging correlates of updating versus nonupdating processes in PFC. Thus, it is perhaps not surprising that they raise more questions than they resolve. The results of these experiments reveal that working memory encoding of letter strings automatically incorporates an episodic contextual code, but the breadth of contextual information that can be incorporated in this code and its accessibility to volitional control remain undetermined. Future research is required to address at least three important questions about this episodic code. First, do participants assess the probe against more items on trials that violate group integrity than on those that do not? That is, is the effective memory set larger on group integrity-violating trials than on group integrity-preserving trials (a model similar to the trial-length explanation of PI; Bunge et al., 2000)? Second, what, if any, other kinds of contextual information can be contained in this code? Third, is this episodic code generated and maintained obligatorily, or can either its encoding or its maintenance be suppressed with appropriate training?

These experiments have also confirmed that the physiological item-accumulation effect has no correlate in accuracy or RT data, but the processes revealed by this effect remain obscure. These experiments have failed to reveal evidence that updating-related processes can be dissociated neurally from nonupdating-related processes, a result that calls into question the assumption that the discarding and repositioning operations assumed to be required by the updating task are fundamentally different from the encoding- and maintenance-related processes that are engaged by all, even the simplest, working memory tasks. Thus, further work is required to resolve whether or not performance on the updating task reveals the op-

eration of executive control functions in working memory (Kiss et al., 1998; Lehto, 1996; Morris & Jones, 1990; Salmon et al., 1996). Finally, by rejecting an alternative explanation, these experiments have strengthened the case that the neural effect associated with high-overlap conditions of delayed-recognition tasks is indeed related to the resolution of PI.

REFERENCES

- AGUIRRE, G. K., ZARAHN, E., & D'ESPOSITO, M. (1998). The variability of human, BOLD hemodynamic responses. *NeuroImage*, **8**, 360-369.
- ASHBURNER, J., & FRISTON, K. (1996). Fully three-dimensional nonlinear spatial normalization: A new approach. *NeuroImage*, **3**, S111.
- BOYNTON, G. M., ENGEL, S. A., GLOVER, G. H., & HEEGER, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *Journal of Neuroscience*, **16**, 4207-4221.
- BRAVER, T. S., COHEN, J. D., NYSTROM, L. E., JONIDES, J., SMITH, E. E., & NOLL, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *NeuroImage*, **5**, 49-62.
- BUNGE, S. A., MATSUMOTO, A., DESMOND, J. E., GLOVER, G. H., & GABRIELI, J. D. E. (2000, March). *The neural basis of interference resolution: Manipulations of interference and working memory load in the item recognition paradigm*. Poster presented at the Rotman Research Institute 10th Annual Conference: The frontal lobes.
- BURGESS, N., & HITCH, G. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, **106**, 551-581.
- D'ESPOSITO, M., AGUIRRE, G. K., ZARAHN, E., & BALLARD, D. (1998). Functional MRI studies of spatial and non-spatial working memory. *Cognitive Brain Research*, **7**, 1-13.
- D'ESPOSITO, M., & POSTLE, B. R. (1999). The dependence of span and delayed-response performance on prefrontal cortex. *Neuropsychologia*, **37**, 1303-1315.
- D'ESPOSITO, M., & POSTLE, B. R. (2000). Neural correlates of component processes of working memory: Evidence from neuropsychological and pharmacological studies. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 579-602). Cambridge, MA: MIT Press.
- D'ESPOSITO, M., POSTLE, B. R., BALLARD, D., & LEASE, J. (1999). Maintenance versus manipulation of information held in working memory: An event-related fMRI study. *Brain & Cognition*, **41**, 66-86.
- D'ESPOSITO, M., POSTLE, B. R., JONIDES, J., & SMITH, E. E. (1999). The neural substrate and temporal dynamics of interference effects in working memory as revealed by event-related functional MRI. *Proceedings of the National Academy of Sciences*, **96**, 7514-7519.
- FRISTON, K. J., ASHBURNER, J., FRITH, C. D., POLINE, J.-B., HEATHER, J. D., & FRACKOWIAK, R. S. J. (1995). Spatial registration and normalization of images. *Human Brain Mapping*, **2**, 165-189.
- 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.
- GARAVAN, H. (1998). Serial attention within working memory. *Memory & Cognition*, **26**, 263-276.
- GARAVAN, H., ROSS, T. J., LI, S.-J., & STEIN, E. A. (2000). A parametric manipulation of central executive functioning. *Cerebral Cortex*, **10**, 585-592.
- HITCH, G. J., BURGESS, N., TOWSE, J. N., & CULPIN, V. (1996). Temporal grouping effects in immediate recall: A working memory analysis. *Quarterly Journal of Experimental Psychology*, **49A**, 116-139.
- JONIDES, J., MARSHUETZ, C., SMITH, E. E., REUTER-LORENZ, P. A., & KOEPPE, R. A. (2000). Age differences in behavior and PET activation reveal differences in interference resolution in verbal working memory. *Journal of Cognitive Neuroscience*, **12**, 188-196.
- JONIDES, J., SCHUMACHER, E. H., SMITH, E. E., LAUBER, E. J., AWH, E., MINOSHIMA, S., & KOEPPE, R. A. (1997). Verbal working memory load affects regional brain activation as measured by PET. *Journal of Cognitive Neuroscience*, **9**, 462-475.
- JONIDES, J., SMITH, E. E., MARSHUETZ, C., KOEPPE, R. A., & REUTER-LORENZ, P. A. (1998). Inhibition in verbal working memory revealed by brain activation. *Proceedings of the National Academy of Sciences*, **95**, 8410-8413.
- KISS, I., PISIO, C., FRANCOIS, A., & SCHOPFLOCHER, D. (1998). Central executive function in working memory: Event-related brain potential studies. *Cognitive Brain Research*, **6**, 235-247.
- LEHTO, J. (1996). Are executive function tests dependent on working memory capacity? *Quarterly Journal of Experimental Psychology*, **49A**, 29-50.
- MORRIS, N., & JONES, D. M. (1990). Memory updating in working memory: The role of the central executive. *British Journal of Psychology*, **81**, 111-121.
- NYSTROM, L. E., BRAVER, T. S., SABB, F. W., DELGADO, M. R., NOLL, D. C., & COHEN, J. D. (2000). Working memory for letters, shapes and locations: fMRI evidence against stimulus-based regional organization of human prefrontal cortex. *NeuroImage*, **11**, 424-446.
- PETRIDES, M. (1989). Frontal lobes and memory. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (3rd ed., Vol. 3, pp. 75-90). New York: Elsevier.
- POSTLE, B. R., BERGER, J. S., & D'ESPOSITO, M. (1999). Functional neuroanatomical double dissociation of mnemonic and executive control processes contributing to working memory performance. *Proceedings of the National Academy of Sciences*, **96**, 12959-12964.
- 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., STERN, C. E., ROSEN, B. R., & CORKIN, S. (2000). An fMRI investigation of cortical contributions to spatial and nonspatial visual working memory. *NeuroImage*, **11**, 409-423.
- 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 Research Protocols*, **5**, 57-66.
- ROWE, J. B., TONI, I., JOSEPHS, O., FRACKOWIAK, R. S. J., & PASSINGHAM, R. E. (2000). The prefrontal cortex: Response selection or maintenance within working memory? *Science*, **288**, 1656-1660.
- RYAN, J. (1969). Grouping and short-term memory: Different means and patterns of grouping. *Quarterly Journal of Experimental Psychology*, **21**, 137-147.
- RYPMA, B., & D'ESPOSITO, M. (1999). The roles of prefrontal brain regions in components of working memory: Effects of memory load and individual differences. *Proceedings of the National Academy of Sciences*, **96**, 6558-6563.
- SALMON, E., VAN DER LINDEN, M., COLLETTE, F., DELFIORE, G., MAQUET, P., DEGUELDRE, C., LUXEN, A., & FRANCK, G. (1996). Regional brain activity during working memory tasks. *Brain*, **119**, 1617-1625.
- SARTER, M., BERNTSON, G. G., & CACIOPPO, J. T. (1996). Brain imaging and cognitive neuroscience: Toward strong inference in attributing function to structure. *American Psychologist*, **51**, 13-21.
- SCHUMACHER, E. H., LAUBER, E., AWH, E., JONIDES, J., SMITH, E. E., & KOEPPE, R. A. (1996). PET evidence for an amodal verbal working memory system. *NeuroImage*, **3**, 79-88.
- TALAIRACH, J., & TOURNOUX, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.
- TEUBER, H.-L. (1955). Physiological psychology. *Annual Review of Psychology*, **6**, 267-296.
- 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). Empirical analyses of BOLD fMRI statistics: I. Spatially unsmoothed data collected under null-hypothesis conditions. *NeuroImage*, **5**, 179-197.
- ZARAHN, E., AGUIRRE, G. K., & D'ESPOSITO, M. (1997b). A trial-based experimental design for fMRI. *NeuroImage*, **6**, 122-138.

NOTES

1. In pilot testing, one half of all updating trials featured two or more discards that required violation of group integrity, and one half featured zero or one such discard. Inspection of the data, however, indicated that

performance was sensitive only to violations of the integrity of groups containing one or two items that would be in the final span of items tested by the probe.

2. Note that the apparent ceiling effect could not be avoided, because pilot testing indicated that maintaining a running span of 5 items was too difficult.

3. Morris and Jones (1990) did not report EOA.

4. It bears restating that neural dissociability cannot be conclusively demonstrated by physiological measures (such as fMRI) (Sarter, Berntson, & Cacioppo, 1996). Doubly dissociable effects of functional

disruption (e.g., a lesion) in two discrete loci are required to draw this inference conclusively (Teuber, 1955).

5. Note that the design of the behavioral experiments was also influenced by fMRI-related constraints, in that the 4-sec EOA represented the closest together we could position epochs whose evoked fMRI signal would be mutually resolvable (Zarahn et al., 1997b).

6. We thank Sharon Thompson-Schill for this observation.

(Manuscript received August 8, 2000;
revision accepted for publication January 2, 2001.)