COGNITIVE NEUROPSYCHOLOGY, 1999, 16 (3/4/5), 459–488

MEMORY CONJUNCTION ERRORS IN YOUNGER AND OLDER ADULTS: EVENT-RELATED POTENTIAL AND NEUROPSYCHOLOGICAL DATA

Susan R. Rubin, Cyma Van Petten, Elizabeth L. Glisky, and Wendy M. Newberg
University of Arizona, Tucson, USA

In a study/recognition paradigm, new words at test were recombinations of studied syllables (e.g. BARLEY from BARTER and VALLEY), shared one syllable with studied words, or were completely new. False alarm rates followed the gradient of similarity with studied items. Event-related potentials to the three classes of false alarms were indistinguishable. False alarms elicited different brain activity than did hits, arguing against the idea that conjunction errors occur during encoding and are later retrieved liked genuine memories. In Experiment 2, with healthy older adults, neuropsychological tests sensitive to frontal lobe function predicted false alarm rate, but not hit rate. Performance on standardised memory scales sensitive to medial temporal/diencephalic function influenced the pattern of false alarm rates across the three classes of new words. The experiments suggest that false alarms to conjunction lures are not similar to true recollections, but are products of faulty monitoring at retrieval.

INTRODUCTION

Illusions of memory have provided a perennial source of fascination for psychologists. From Bartlett's famous 1932 *War of the Ghosts* study, through the groundbreaking work on eyewitness testimony done by Loftus and colleagues in the 1970s (e.g. Loftus, Miller, & Burns, 1978), and continuing today in the renewed interest in the converging associates paradigm (Deese, 1959), psychologists never cease to be amazed at how often, and with what confidence, people will claim to remember events that never occurred (Roediger & McDermott, 1995). It has been proposed that at least some false memories are caused by binding failures, namely that components of presented information are inappropriately recombined into episodes that never occurred, producing *memory conjunction errors* (Kroll, Knight, Metcalfe, Wolf, & Tulving, 1996; Reinitz, Verfaellie, & Milberg, 1996). For instance, after meeting several new colleagues, including a Mr. Waters and a Mr. Hanson, a person might later recognise "Watson" incorrectly. A laboratory paradigm for investigating such errors was introduced by Underwood and Zimmerman (1973). In this paradigm, people study two-syllable words that are recombined at test into words that are new as wholes but whose individual features (in this case syllables) have been studied. A person who studied SPANIEL and VARNISH might incorrectly judge the conjunction lure SPANISH to be an old word during

---

Requests for reprints should be addressed to Susan Rubin, Department of Psychology, University of Arizona, Tucson, AZ 85721, USA (Fax: (520) 621-9306; E-mail: rubin@u.arizona.edu).

The experiments were the Master's thesis of the first author. Financial support was provided by the National Institute of Aging (AG14792), National Institute of Neural Disorders and Stroke (NS30825), and the Flinn Foundation. We are grateful to Ava Senkfor, Roh Ohst, and Mark Bakarich for technical support, and to Cynthia P. May, Timothy Curran, and Mark Reinitz for comments on the paper.

recognition. Such conjunction errors have been found using diverse stimuli, including compound and noncompound words, nonsense words, word pairs, sentences, pictures, and faces, indicating that the phenomenon is not purely verbal (Dorfmann, 1994; Jha, Kroll, Baynes, & Gazzaniga, 1997; Kroll et al., 1996; Reinitz, Lammers, & Cochran, 1992; Reinitz et al., 1996; Underwood, Kapelak, & Malmi, 1976; Underwood & Zimmerman, 1973).

Storing the relations among diverse aspects of a single event is a core memory function. It has been suggested that a large proportion of the distortions observed in both normal and pathological memory result from incorrect amalgams of elements drawn from external events, inferences, and other mental events (Johnson & Raye, in press). The conjunction error paradigm offers a relatively direct window into the processes underlying miscombinations of features. Experiment 1 is designed to evaluate whether such miscombinations arise primarily from faulty encoding or faulty retrieval, using a combination of behavioural and brain activity measures. Experiment 2 investigates the neuropsychological correlates of memory conjunction errors in healthy older adults.

One interpretation of conjunction errors is that they reflect a binding failure during encoding, which may consist of a complete failure to encode relational information, weak or inefficient binding such that appropriate elements are not strongly linked, or excessive binding so that fragments that belong to separate events are indiscriminately linked (Kroll et al., 1996). Another interpretation is that these errors arise from separate encoding and storage of global structure and stimulus features, which may lead to conjunction errors at retrieval, when the parts must be recombined properly (Reinitz et al., 1996).

Encoding manipulations can be used to evaluate the idea that conjunction errors are due to faulty initial learning, but the evidence to date has been mixed. Reinitz and colleagues varied the task performed during the study phase, but found equivalent false alarm rates after semantic and orthographic study tasks (Reinitz et al., 1996). Dividing attention during encoding increases false alarms to conjunctions, but also increases other sorts of errors, and so may not indicate a specific role for encoding in the binding failure (Reinitz, Morrissey, & Demb, 1994; Reinitz et al., 1996).

The effects of dividing attention during retrieval have not been examined, but other evidence indicates that conjunction errors are sensitive to testing conditions. To date, the data on memory conjunction errors in free recall is mixed. Reinitz et al. (1992) reported finding memory conjunction errors in a recall task, using stimuli of the form "The X saw the Y". However Roder, Faneuf, Burch, and Connors (1996) found no such effect using noun pairs as stimuli. It is unclear if the ambiguity of these results is due to different types of stimuli, small numbers of trials, or other methodological issues. If additional experiments support the absence of conjunction errors in explicit recall, it will argue against the idea that a false memory is formed during study and later retrieved like a real memory.

Conjunction errors have been observed in some implicit memory tests but not others. Word fragments corresponding to conjunctions of studied items were solved more successfully than fragments with unstudied completions, but conjunction words fared no better than totally new words when they had to be identified in visual noise (Reinitz & Demb, 1994). It is something of a puzzle that conjunction errors have been found in recognition, but not consistently in either implicit or recall tasks. The data concerning the testing conditions that foster or prevent conjunction errors are still sparse, and present an intriguing area for future research.

One means of evaluating the hypothesis that conjunction errors arise from faulty binding at encoding is to measure the presence or absence of an elevated false alarm rate for test items sharing only one syllable with studied words (e.g. GARMENT when only BASEMENT was studied). An account based only on faulty binding would be likely to predict no difference between such *syllable lures* and new words because neither can be generated by miscombining studied syllables. The evidence to date has been inconclusive. Two studies report no significant difference for normal subjects (Kroll et al., 1996;

Reinitz et al., 1994), a third reports significantly more false alarms to feature lures using faces as stimuli (Reinitz et al., 1994), and a fourth includes no statistical comparison although the raw numbers indicate more false alarms to syllable lures than to completely new words (Reinitz et al., 1996). One likely reason for this murky picture is low statistical power; these studies included only four to six items per condition so that differences in false alarm rates may have been difficult to detect.

The question of when false conjunctions are formed thus remains open. A likely retrieval-based explanation involves *processing fluency*, in which familiarity at retrieval is responsible for automatic "yes" responses (Jacoby, 1991). Subjects may mistake the ease in processing studied syllables with familiarity of the whole word. If they have no strong recollection to oppose this impression, they may call the recombined word "old". In contrast to the pure binding hypothesis, the processing fluency account predicts an elevated false alarm rate for syllable lures relative to wholly new words because at least one part of the word is familiar.

Event-related potentials (ERPs) are particularly well-suited to discriminating between encoding and retrieval explanations of memory conjunction errors because brain activity can be recorded during both study and test phases. ERPs are summed synaptic potentials recorded from electrodes on the scalp; the activity elicited by a particular class of stimuli are separated from the background electroencephalogram by averaging responses timelocked to the stimuli of interest (Regan, 1989).

Previous research has shown that a late positive component (LPC) of the ERP is sensitive to mnemonic processes, but only to a subset of those processes that may influence behaviour. Studied words that elicit accurate recognition or cued-recall judgements also elicit larger LPCs than new items or unrecognised old items (Allan & Rugg, 1997; Neville, Kutas, Chesney, & Schmidt, 1986; Senkfor & Van Petten, 1998). Late positivities are also triggered by studied line-drawings, novel geometric shapes, and environmental sounds in recognition tasks, and by incidentally repeated words presented during other tasks that do not call for explicit memory judgements (Bentin & Peled, 1990; Berman, Friedman, & Cramer, 1991; Besson, Kutas, & Van Petten, 1992; Chao, Nielsen-Bohlman, & Knight, 1995; Swick & Knight, 1997; Van Petten & Senkfor, 1996). This old/new effect is enhanced by the same manipulations that improve accuracy in explicit memory tasks, such as "deep" or semantic study as compared to "shallow" or orthographic study tasks (Paller & Kutas, 1992). In studies using "remember/know" judgements, the LPC is largest for "remember" items as compared to "know" items or new words (Duzel, Yonelinas, Mangun, Heinze, & Tulving, 1997; Smith, 1993). In contrast, perceptual manipulations that increase or decrease priming effects in implicit memory tasks have no influence on the LPC (Paller & Gross, 1998; Paller, Kutas, & McIsaac, 1998; Rugg & Doyle, 1994). The available evidence supports the view that the memory-related LPC indexes conscious recollection rather than nondeclarative memory processes. Experiments using difficult-to-remember stimuli like scrambled faces or geometric shapes have thus yielded faster reaction times for repeated items in the absence of any LPC repetition effect, presumably because perceptual processes influenced by familiarity can drive behavioural output in the absence of conscious recollection (Bentin & McCarthy, 1994; Van Petten & Senkfor, 1996).

We have also observed that LPCs recorded during explicit recognition tasks distinguish correct old items (hits) from incorrect new items (false alarms), although both attract the same "yes" judgement (Van Petten & Senkfor, 1996). Our prior study evaluated only the garden-variety false alarms that occur in all recognition tests, rather than false alarms that are specially motivated by events occurring during the study phase. If conjunction errors are formed by inappropriate stimulus combinations during encoding, or during a post-encoding consolidation period, then these unstudied items may be distinct from garden-variety false alarms, and instead be retrieved much like genuine memories. The hypothesis of an initial binding failure thus predicts that ERPs to conjunction lure false alarms will look much like hits. In contrast, the processing fluency (retrieval) explanation predicts the standard result, in which hits look different from all other conditions that do not involve veridical memory traces for the whole words. Previous work suggests that mere familiarity or increased processing fluency leaves no LPC record, so that conjunction words may not differ from other new words in terms of the LPC if fluency alone is the relevant factor. Alternatively, if explicit recognition of individual syllables drives "yes" responses, we may observe a graded effect of similarity to studied items in the LPC in the following pattern: hits might have the largest LPC, followed by conjunction false alarms, then syllable lure false alarms, and finally false alarms to completely new items.

An additional factor investigated in both of the present experiments is recognition confidence for false alarms as compared to hits. As a general rule, incorrect responses (both false alarms and misses) are made with lower confidence than correct responses in recognition tests (Busey, Tunnicliff, Loftus, & Loftus, 1998; Clark, 1997; Gardiner & Java, 1990; Murdock & Dufty, 1972). In experiments designed to elicit false alarms, new items that are related to studied words yield higher confidence "yes" responses than unrelated new items, but these confidence ratings are still lower than those associated with correct "yes" responses (Roediger & McDermott, 1995). Decision confidence is almost certain to have some correlate in brain activity during recognition tests. Prior ERP studies have investigated decision confidence in non-memory tasks requiring difficult perceptual discriminations. As compared to low-confidence trials, high confidence decisions elicited larger P300s, a late positive potential in much the same latency range as the memory-sensitive late positive component (Hillyard, Squires, Bauer, & Lindsay, 1971; Paul & Sutton, 1972). This result suggests that it is critical to distinguish brain activity related to accurate memory performance from that related to confidence per se.

Two prior ERP studies have offered participants an opportunity to indicate that they were uncertain about a recognition judgement. Rugg and Doyle (1992) requested a second button-press response, which could either confirm a participant's old/new judgement or indicate that it was a guess. Wilding and Rugg (1996) offered three response alternatives of "old," "new," or "don't know." In both studies, the ERPs revealed a fairly standard old/new effect when the disconfirmed or "don't know" trials were excluded from the analyses. These studies suggest that the typical old/new difference observed in recognition ERPs may not be a reflection of differential confidence for correct old (hit) and correct new (correct rejection) judgements. However, neither study included enough trials to examine the ERPs elicited by trials with low confidence ratings, or to evaluate the impact of confidence during an erroneous recognition judgement. In Experiment 1 here, participants are offered a graded confidence scale for both "old" and "new" responses so that the interactions between confidence and accuracy can be examined.

EXPERIMENT 1

Method

Participants

Participants were 27 fluent English speakers with no history of neurological disorder, who gave informed consent (15 men, 12 women, age range 21–39 years). Three participants were excluded from the study because of unacceptably high artefact rejection rates. Of the remaining 24 participants, 21 were right-handed, 2 were left-handed, and 1 was ambidextrous. Eleven of the right-handers reported having a left-handed parent or sibling.

Materials

Materials included 440 triads of 2-syllable words. Each triad was structured so that the first syllable of the first word and the second syllable of the second word could be combined to make the third or conjunction word (e.g. BARTER/VALLEY/BARLEY). The syllables in the conjunction words were both spelled and pronounced the same as in the first two words of the triad, although we allowed variation in the stress pattern across words. As shown in Table 1, the recognition tests for each

Table 1. *Sample Stimuli in Study and Recognition Phases*

	Study		*Test*	
Words 1 and 2 (N = 220/110)	barter	valley	barley	*conjunction lure*
Word 2 (N = 110)	(spaniel)	varnish	spanish	*syllable lure*
Word 3 (N = 110)	(lobby gangster)	lobster	lobster	*old*
No study (N = 110)	(vulgar creature)	(vulture)	vulture	*new*
Single syllable studied (N = 140)	plaid		plaid	*old single*
Single syllable unstudied (P = 20)	(catch)		catch	*new single*

Parentheses denote words not presented during the study phase. A total of 580 words were presented for study, 600 for test (both split among 4 study/test cycles).

participant included all 440 conjunction words, but equal numbers of these served as *old* (studied) items, *new* (unstudied) items, *conjunction lures* (first two words of the triad studied), and *syllable lures* (second word of the triad studied)[1]. Assignment of conjunction words to one of the four conditions was rotated across participants, so that each word appeared equally often in each test condition.

This design maximises the number of trials in the critical conditions, but also results in only 25% of the critical items being truly old at recognition. Unequal probabilities of old and new items are undesirable, as they may encourage indiscriminate "yes" responses at test, or introduce probability effects in the ERP measures. Single syllable filler words were thus included to provide better balance between the proportions of old and new items; 140 were presented during the study and test phases (old items at test), and 20 only during the recognition phase (new items)[2].

The stimuli were split among four study-test cycles. Each study list thus included 110 two-syllable words, equally split between conjunction words, first and second members of a triad (presented in canonical order), and second members of a triad, together with 35 single-syllable filler words. Mean frequency of usage for all words was 45 (SD = 6; summed frequency of all regularly inflected forms, Francis & Kuçera, 1982); word length averaged 6 (SD = 0.1) letters. Care was taken to avoid inadvertent syllable overlap in either the study or test phases. For instance, if GARTER and SUDDEN appeared on a given study list (to suggest the conjunction GARDEN), that list would not include DENIM, nor would the syllable –DEN– appear in the other list presented in the same session, thereby avoiding ambiguity as to when it may have been acquired.

Procedure

The experiment was conducted in two sessions scheduled about a week apart, each consisting of a short practice list and two study-test cycles. Words were presented in the middle of the screen for 200 msec, at a rate of one every 3.2 sec at study and every 5 sec at test. Participants were informed of the upcoming memory test, but also assigned a study task of classifying each word as noun, verb, or modifier. At recognition, participants used three response keys with one hand to indicate Definite/Probable/Maybe Old, and three keys with the other hand for Definite/Probable/Maybe New.

[1] Experiment 2 indicated little difference in the recognition of unstudied words sharing the first versus second syllable of a studied word.

[2] None of the syllables in the single-syllable words were shared with the two-syllable words.

Left and right hands were counterbalanced across the two sessions.

The electroencephalogram (EEG) was recorded with tin electrodes mounted in a commercially available elastic cap. Midline central (Cz) and parietal (Pz) recording sites were used, along with lateral pairs of electrodes over the prefrontal (Fp 1, Fp2) and occipital (O1, O2) scalp as defined by the 10–20 system (Jasper, 1958). Three additional pairs of lateral electrodes were used: (1) a frontal pair placed midway between F7–8 and T3–4 (approximately over Broca's area and its right homologue, Bl and Br); (2) a temporoparietal pair placed 30% of the interaural distance lateral and 12.5% of the inion–nasion distance posterior to Cz (approximately over Wernicke's area and its right hemisphere homologue, Wl and Wr); and (3) a temporal pair 33% lateral to Cz (Tl and Tr). Each scalp site was referred to an off-line average of the left and right mastoids (see Van Petten & Kutas, 1988). Vertical eye movements and blinks were monitored via an electrode placed below the right eye referred to the left mastoid. Horizontal eye movements were monitored via a right to left bipolar montage at the external canthi.

The EEG was amplified by a Grass Model 12 polygraph with half-amplitude cutoffs of 0.01 and 100 Hz, digitised at an on-line sampling rate of 170 Hz. Trials with horizontal eye movement, excessive muscle activity, or amplifier saturation effects were rejected prior to averaging the single EEG trials into ERPs. Trials with blink artefacts were corrected using an algorithm developed by A. Dale.

Results

Behavioural Performance

The proportion of words called "old" in the four stimulus classes are shown in Fig. 1. Single-syllable filler words are not included in any of the figures or analyses as they were not of interest in this experiment. Conjunction lures elicited the highest false alarm rate, followed by syllable lures, and then new words. The false alarm rate for unrelated words was 21%, close to that of a previous study using study lists of similar length, but recognition lists in which all of the new words were unrelated to studied words (Van Petten & Senkfor, 1996). Syllable lures elicited a 6% higher false alarm rate than new words, and conjunction lures elicited a 6% higher false alarm rate than syllable lures (standard error of these differences was 2%, with a 95% confidence interval of 2–10%). Conjunction lures elicited 12% more false alarms than unrelated new words (standard error of this difference was 2%, with a 95% confidence interval of 9–15%). Pairwise comparisons confirmed that each of the false alarm rates was significantly different from the others: conjunction lure vs. syllable lure, $F(1,23) = 19.2$, $P < .005$; syllable lure vs. new, $F(1,23) = 11.3$, $P < .005$.

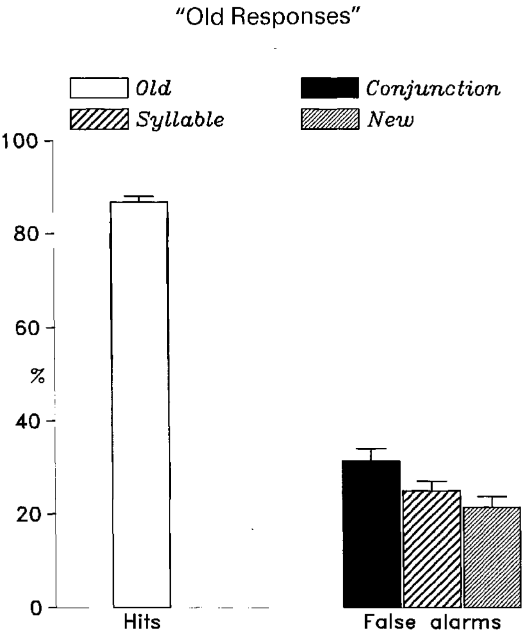

Fig. 1. *Mean percentage of items judged "old" during the recognition test (maximum = 100% in each category) of Experiment 1. Error bars show standard error of the mean. Single-syllable items are not included.*

Confidence levels were higher for hits (mean = 2.73, where "definite" = 3.0) than false alarms (mean = 2.05), correct rejections (mean = 2.31), and misses (mean = 2.06), all Fs $(1,23) > 49.15$, all Ps $< .0001$. Confidence levels for the two broad error categories—false alarms and misses—were equivalent ($F < 1$). Closer examination of the false alarms revealed no confidence differences among the three categories of conjunction lures, syllable lures, and new words (Fs < 2.61).

Reaction times are shown in Table 2, and were analysed by an initial 2 x 2 ANOVA with factors of accuracy and response type ("old" vs. "new"). One subject had no misses and was excluded from the above analysis. Accurate responses (hits and correct rejections) were faster than inaccurate (misses and false alarms), $F(1,22) = 9.96$, $P < .005$. "Old" responses were faster than "new", $F(1,22) = 18.2$, $P < .0001$, and there was a significant interaction of accuracy by "old/new", $F(1,22) = 5.03$, $P < .05$, reflecting the reaction time pattern of hits faster than correct rejections faster than false alarms that were equivalent to misses. A separate analysis revealed no reaction time differences among the three false alarm types. Not surprisingly, reaction times were correlated with confidence level. Collapsed across accuracy, reaction times ranged from a mean of 1354 msec for "definite" responses to 1772 msec for "probable" responses, to 2014 msec for "maybe" responses, $F(2,44) = 50.0$, $P < .0001$.

Event-related Potentials

Old versus New Items. Figure 2 shows that correctly recognised words (hits) elicited a substantially more positive ERP than correctly rejected new words, beginning at about 350 msec post-stimulus onset. Figure 3 shows a similar difference between hits and false alarms (all types collapsed) that was of somewhat longer duration. ERPs elicited at test were analysed in an early time window (300–600 m after stimulus onset) and a late time window (600–900 msec after stimulus onset), using 100msec of pre-stimulus activity as a baseline.[3] In the early time window, the mean amplitude of ERPs elicited by hits was significantly more positive than that of all correct rejections together, $F(1,23) = 42.2$, $P < .0001$, but there was no difference in the late time window. Hits were more positive than false alarms in both the early and late time windows, $F(1,23) = 24.4$ and 17.6, Ps < .0001 and .0005.

Figure 4 shows that the ERPs elicited by false alarms to conjunction lures, syllable lures, and new words were indistinguishable from one another at any latency. The ERPs elicited by correct rejections were also unaffected by word type, $Fs(2,46) < 2$, as shown in Fig. 5.

Because confidence was shown to vary systematically with accuracy and response type, we restricted our analyses from this point on to the "definite" items to ensure that the observed effects were not due to this potentially confounding factor. This restriction decreased the trial numbers in each category since some individuals made very few high confidence errors. Subjects with less than 10 artefact-free trials in each category of interest were excluded from the analyses in order to preserve an adequate signal-to-noise ratio in the individual subject averages, which form the basis for all statistical comparisons. A minimum of 16 subjects were included in each analysis of high confidence responses.

After controlling for confidence, most of the ERP effects remained the same. "Definite" hits were still more positive than "definite" correct rejections in the early time window, $F(1,23) = 17.6$, $P < .001$. Hits were also more positive than "definite" false alarms in the late time window, $F(1,19) = 14.25$, $P < .001$, but the previously observed difference in the early time window faded into marginal significance, $F(1,19) = 3.94$, $P < .06$.

Table 2. *Means and standard errors of reaction times at test*

	All	*"Definites" only*
	Mean (SE)	*Mean (SE)*
Hits	1502 (67)	1053 (43)
Correct rejections	1687 (75)	1393 (57)
False alarms	1758 (81)	1262 (81)
Misses	1760 (106)	—

Not all subjects made misses with high confidence, so the mean for "definite" misses was not calculated.

[3] Electrode site was used as a factor (13 levels) in this and all subsequent ANOVAs. We report only main effects and interactions involving the condition factors in the text, as these serve as the most conservative tests of the experimental effects. As evident in Figs. 2 and 3, the ERP memory effects were widespread across the scalp and not confined to restricted scalp regions.

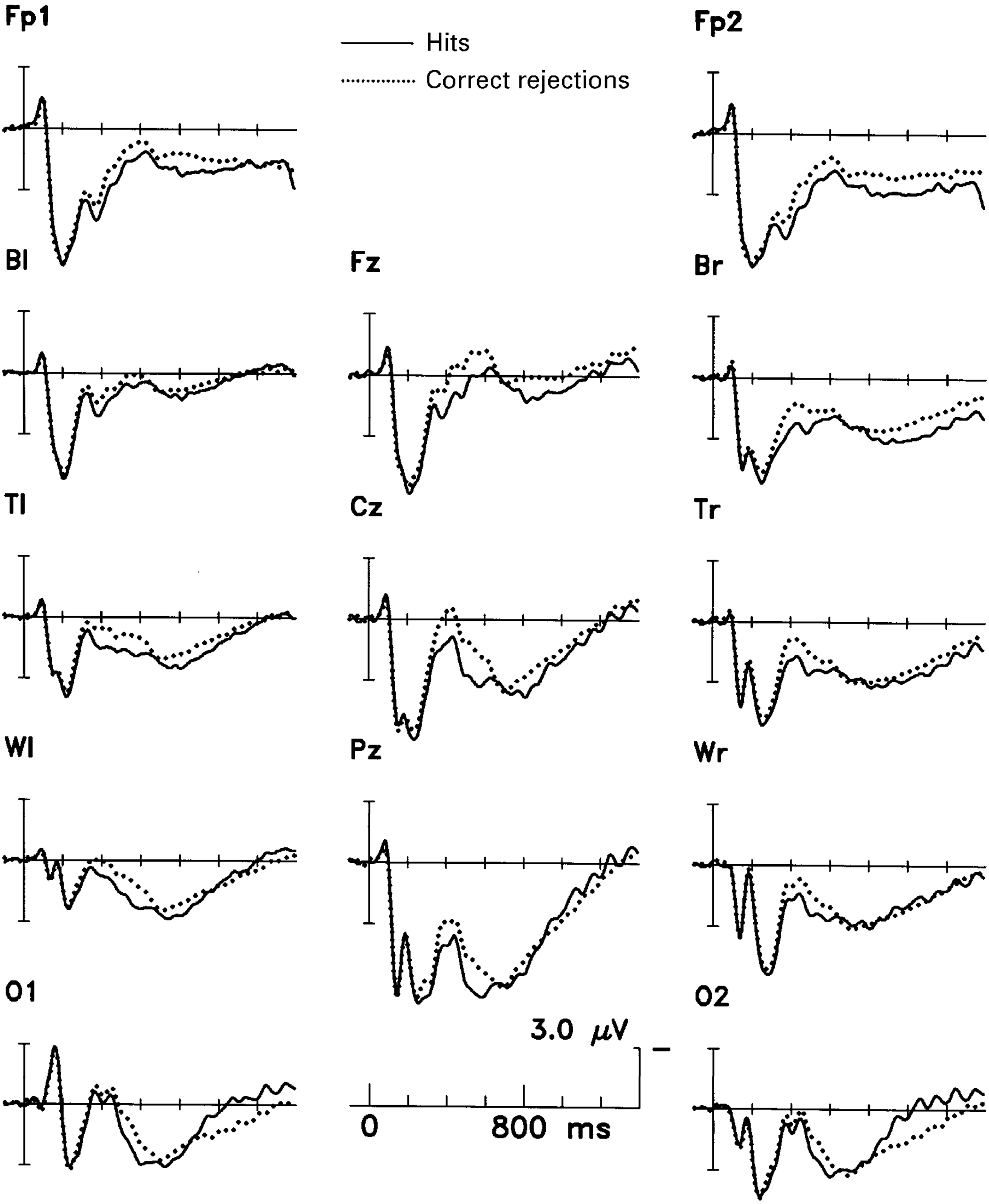

Fig. 2. *Grand average ERPs elicited by words correctly recognised (hits) and correctly rejected (correct rejections) at test. Single-syllable items are not included. Negative voltage is plotted up. Electrode sites over the left scalp are displayed in the left column, midline sites in the centre column, and sites over the right scalp in the right column. The most anterior scalp sites appear at the top and the most posterior sites at the bottom of the figure.*

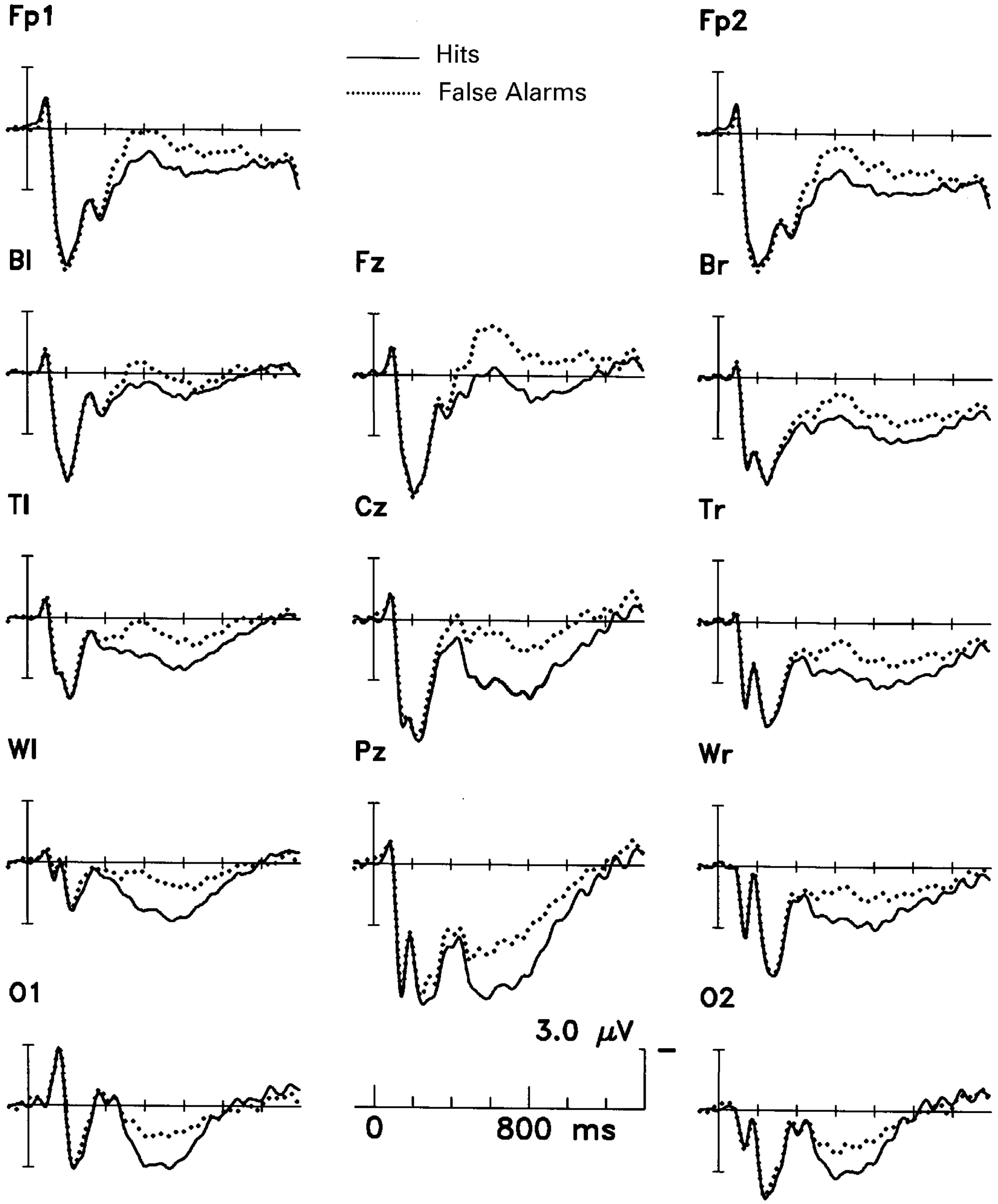

Fig. 3. *Grand average ERPs elicited by words correctly and incorrectly categorised as "old" (hits and false alarms) at test. Single-syllable items are not included.*

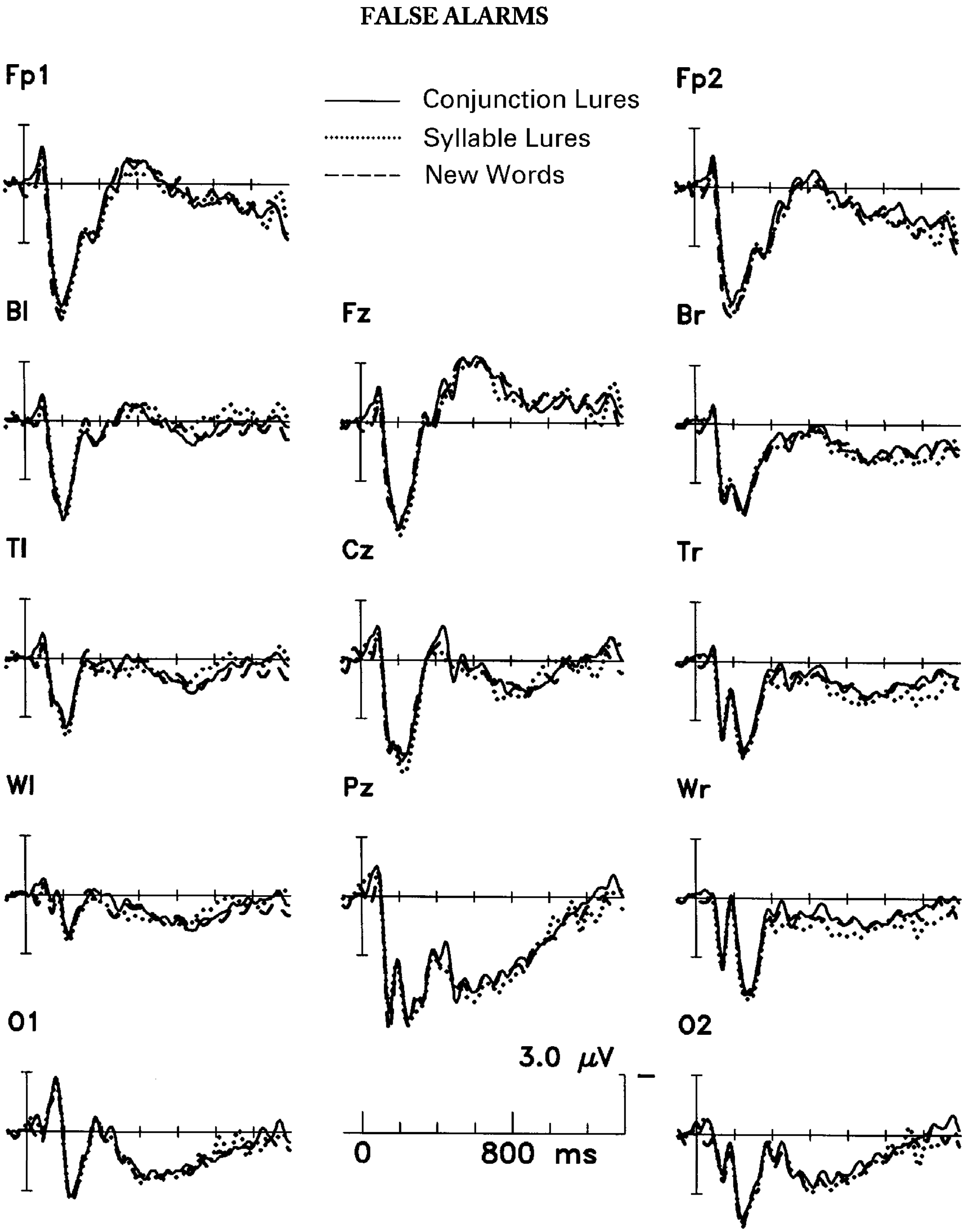

Fig. 4. *Grand average ERPs elicited by conjunction lures, syllable lures, and new words falsely categorised as "old" at test. Single-syllable items are not included.*

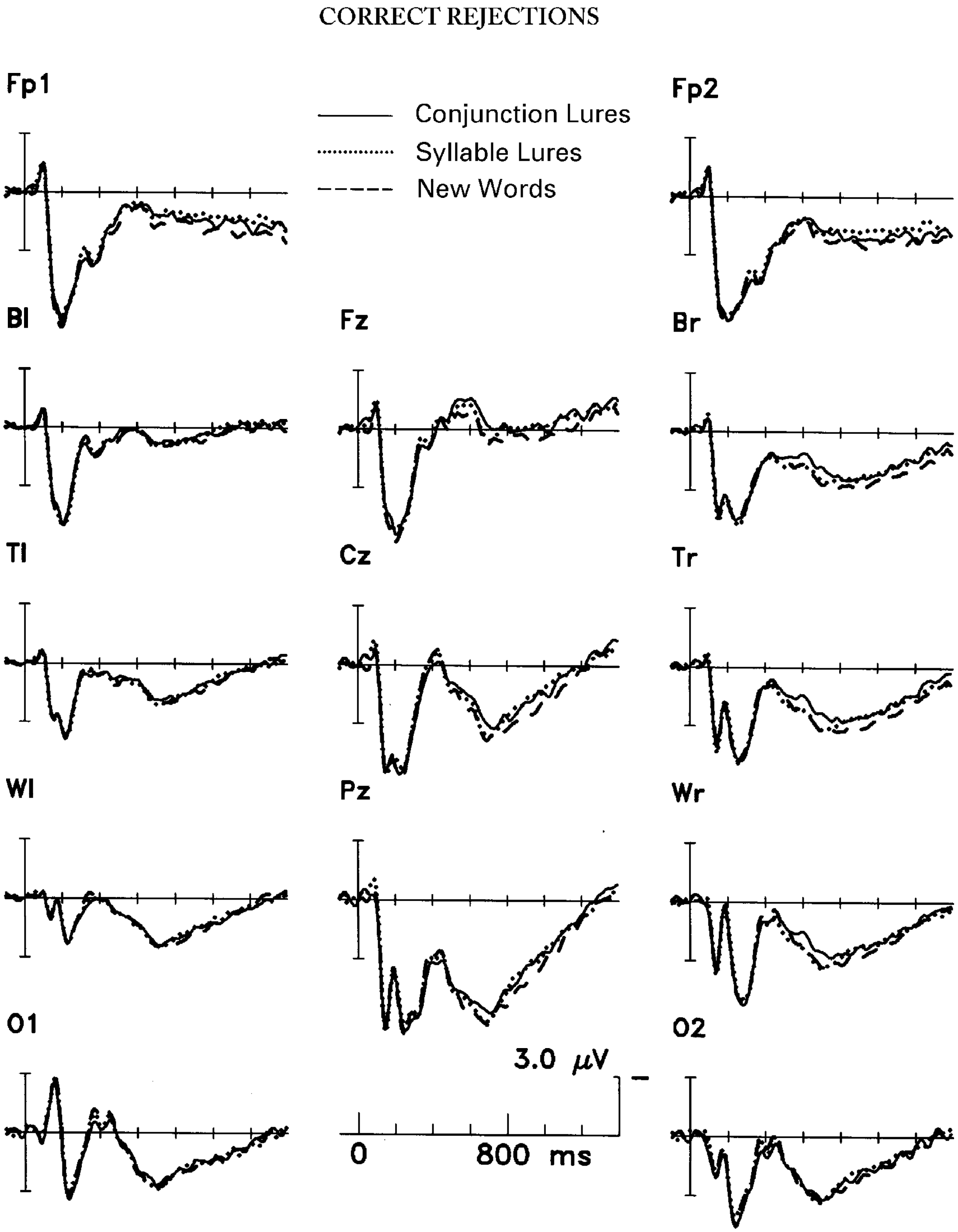

Fig. 5. *Grand average ERPS elicited by conjunction lures syllable lures, and new words correctly rejected as "new" at test. Single-syllable items are not included.*

Perhaps the most central comparison of the experiment is shown in Fig. 6, which contrasts the ERPs elicited by old words and conjunction lures when both were accompanied by "definite old" responses. The ERPs elicited by veridical recognition were significantly more positive than conjunction lure false alarms in the late time window, $F(1,15) = 7.72$, $P < .01$, and marginally more positive in the early time window, $F(1,15) = 4.20$, $P < .06$.

ERPs elicited during the study phase were examined for signs of differential encoding leading to correct or incorrect recognition judgements in the subsequent test. Too few studied words were missed at test to form reliable ERP averages for these words when they were initially studied. Three other categories of study-phase items were evaluated: (1) words subsequently presented, and recognised at test (*hits at test*) (2) words whose individual syllables were presented as parts of other words at test (i.e., the studied counterparts of the conjunction and syllable lures), and those test lure words were correctly rejected (*correct rejections at test*), and (3) words whose individual syllables were presented as parts of other words at test, and those test words were falsely categorised as old (*false alarms at test*). Figure 7 shows that there were no differences among these three categories, all *F*s < 1.

Confidence Effects. The preceding analyses confirmed that the ERP differences between hits and new words (both correct rejections and false alarms) could not be attributed to differential confidence, because the differences were still evident when the analyses were restricted to "definite" responses. However, the relationship between decision confidence and accuracy, as well as the impact of confidence per se on the ERPs are also of interest. A priori, we can imagine two very different scenarios about the role of confidence in recognition ERPs. One possibility is that confidence exerts an independent influence on ERPs in all decision tasks, including recognition tests, so that this influence is additive with the old/new memory effects. This account predicts equivalent confidence effects (differences between high and low confidence responses) for hits, correct rejections, false alarms, and misses. An alternate possibility is that the influence of decision confidence on the ERPs elicited in a memory task is largely mediated by accurate retrieval; strong memory traces result in both high ratings of confidence, and a larger LPC in the ERP waveform. This latter account predicts a confidence effect for hits, and no confidence effect for misses or false alarms. The prediction of this account for correct rejections is less clear. If participants consult their memories for studied items in order to rule out these new items, correct rejections may also be subject to a confidence effect; but if making a correct rejection does not involve explicit retrieval, this process may leave no record in the memory-sensitive LPC.

We examined the relationship between decision confidence and accurate retrieval by comparing the ERPs elicited by hits, false alarms, and correct rejections at different confidence levels. There were too few misses to form ERP averages with an adequate signal-to-noise ratio. Similarly, very few correct responses (hits or correct rejections) were accompanied by "maybe" responses, so that it was not possible to form separate ERP averages for this lowest confidence ranking. Instead, "probable" and "maybe" responses were summed for comparison to "definite" responses. For the six categories of interest—hits, correct rejections, and all varieties of false alarm accompanied by "definite" and "probable/maybe" ratings—15 of the 24 subjects had at least 10 artefact free trials per category and the analyses here are based on those subjects.

Figure 8 shows that confident and correct judgements about old words elicited more positive ERPs than hits accompanied by lower confidence, beginning about 500msec post-stimulus onset. Confidence had a smaller impact in correct judgements about new words, and essentially no effect on incorrect judgements about new words. Analyses of mean amplitudes in the 600–900msec epoch thus yielded a significant confidence effect for hits [$F(1,14) = 8.78$, $P < .02$], a nonsignificant trend for correct rejections [$F(1,14) = 3.24$, $P < 10$], and no difference for false alarms [$F(1,14) = 1.40$]. Comparisons between hits and both categories of new items confirmed that the confidence effect was largest for correct old items

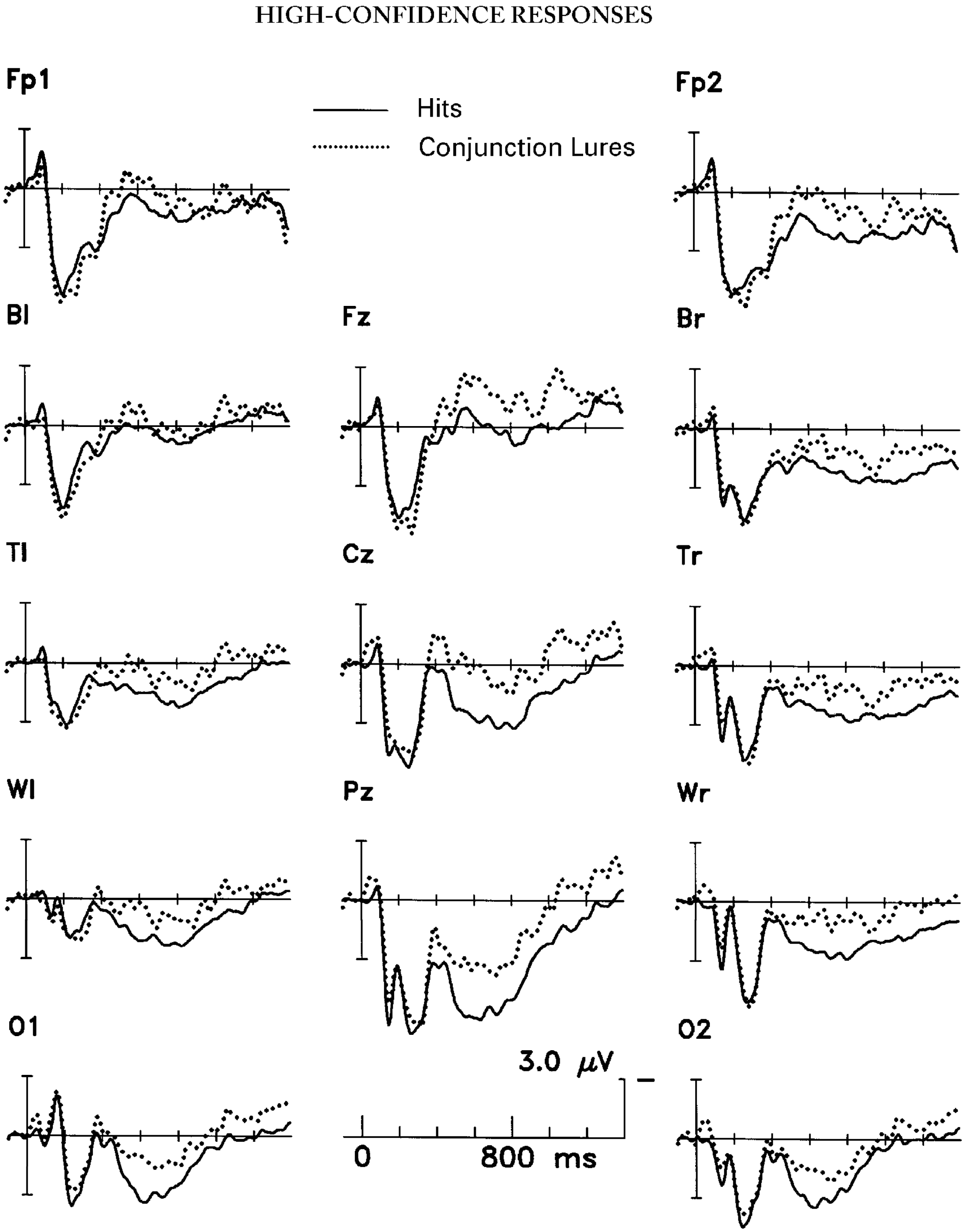

Fig. 6. *Grand average ERPs elicited by truly old words and conjunction lures caterorised as "definitely old" at test. Single-syllable items are not included.*

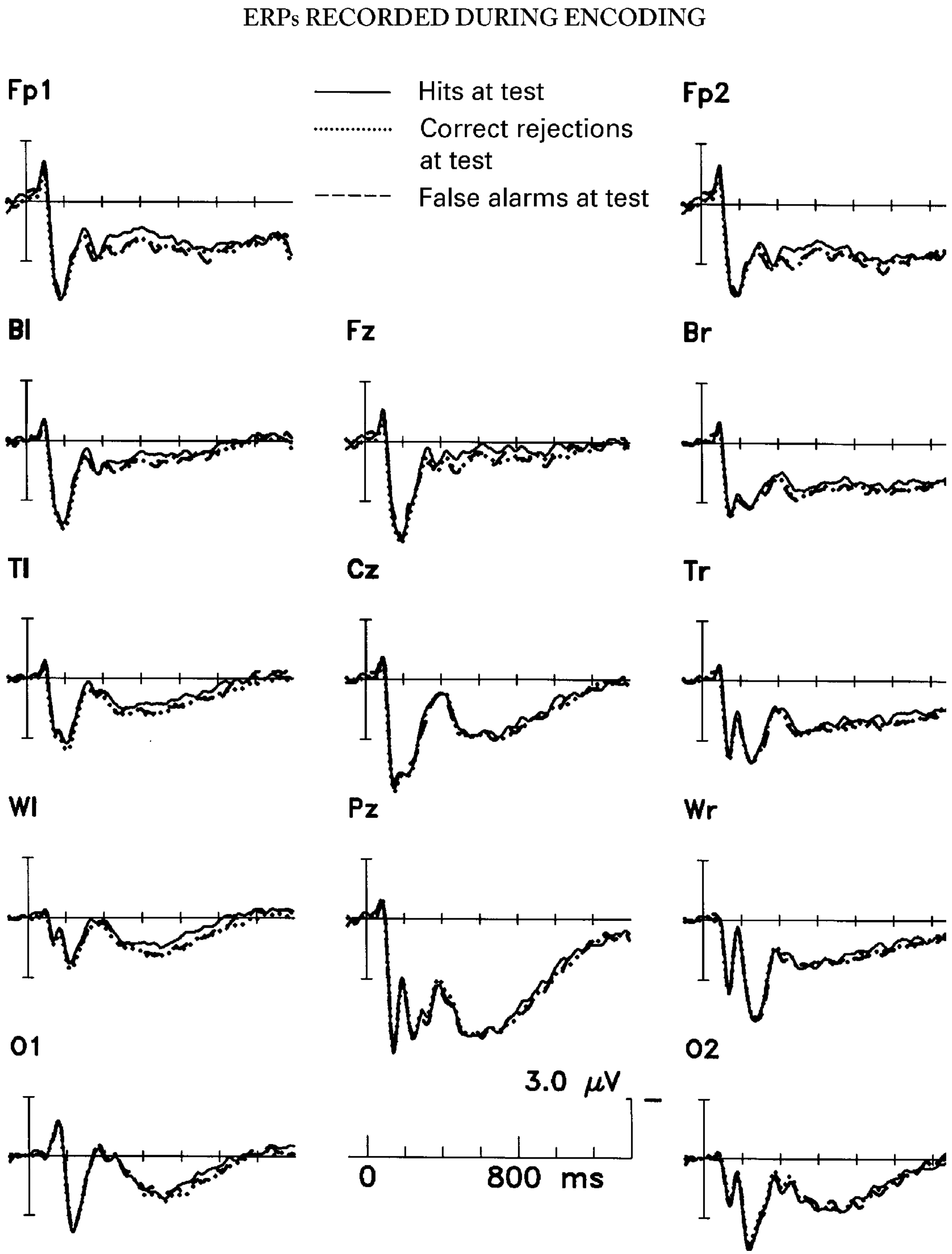

Fig. 7. *Grand average ERPs elicited during study by words later correctly categorised as "old" (hits) correctly categorised as "new" (correct rejections), and incorrectly categorised as "old" (false alarms) at test. Single-syllable items are not included.*

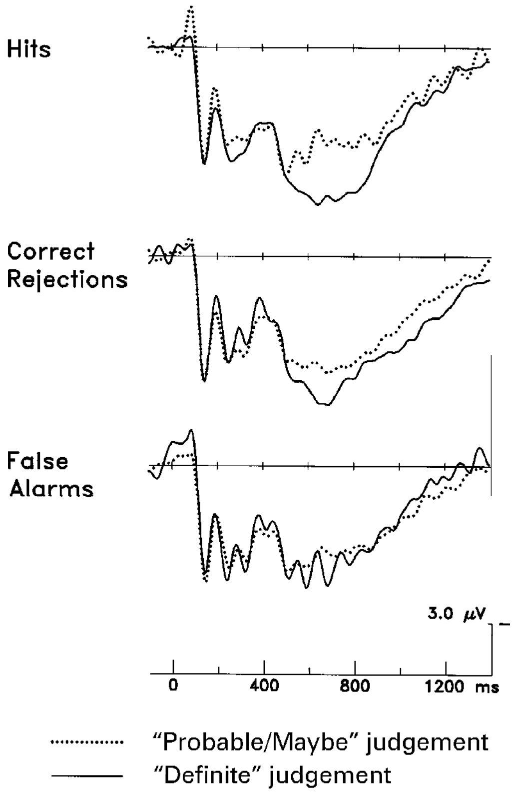

Fig. 8. *Effect of confidence on grand average ERPs elicited by different-response types (hits, correct rejections, and false alarms) at electrode site Pz. Single-syllable items are not included.*

[interaction of Old/New by Confidence: hits vs. correct rejections, $F(1,14) = 4.59$, $P < .05$; hits vs. false alarms, $F(1,14) = 8.43$, $P < .02$]. There were no significant confidence effects in analyses of the 300-600msec epoch.

The outcome of these confidence analyses is most compatible with the second hypothesis outlined earlier. Although it will be important to examine recognition misses in some future experiment that includes more highly confident misses, the present results suggest that the amplitude of the LPC is primarily driven by successful retrieval so that a clear recollection of a studied item will yield both high decision confidence and larger late positivities.

Discussion

Encoding versus Retrieval

Our results are most compatible with a processing fluency explanation of false alarms to conjunction lures. The critical piece of behavioural evidence is that subjects had consistently higher false alarm rates to syllable lures than to new words. If the binding hypothesis were correct, we would have expected no difference between these two word types. False alarms to syllable lures follow presentation of only one of the relevant syllables during encoding, so this type of error is unlikely to be attributable to inappropriate binding during the study phase. Two previous studies have reported no differences between syllable lures and new words, but these suffered from low statistical power due to small numbers of trials, which is not the case in the current study (Kroll et al., 1996; Reinitz & Demb, 1994). The graded pattern of false alarms (conjunction lures < syllable lures < new words) observed here is predicted by a processing fluency account since conjunction lures have two familiar components, syllable lures have one, and new words have none.

An even stronger argument against the binding hypothesis comes from the ERP data, which indicates that conjunction errors are formed at retrieval rather than encoding. The key finding here was that ERPs elicited by false alarms to conjunction lures, syllable lures, and new words at test were indistinguishable from one another, but all were significantly different from the ERPs elicited by hits. The largest LPC, previously thought to be sensitive to conscious recollection, was observed for hits. Moreover, high confidence hits elicited larger LPCs than low-confidence hits, providing further support for the idea that this component reflects the recollective experience of previously studied items. No differences were observed among the three types of false alarms, suggesting that these responses were not based on explicit recollection of words created by mis-binding during the study phase, or even of individual syllables viewed earlier, but on a feeling of familiarity, which led to a "yes" response. However, it is still unclear what this feeling of familiarity is based on. It could plausibly be due to orthographic, phonological, or visual similarity, and the current study cannot distinguish between these possibilities.

Although the ERPs elicited by false recognition of new items were clearly distinct from those elicited by accurate retrieval of old items, judgements of "old" and "new" for new items did not lead to identical ERPs. Comparison of Fig. 2 and 3 (as well as the statistical analyses) indicates that the difference between hits and false alarms is of longer duration than the analogous difference between hits and correct rejections. We have suggested that both false alarms and correct rejections differ from hits because they share one core feature: the absence of a clear recollective experience. But false alarms and correct rejections differ along other dimensions, some of which may be general to tasks that involve an element of target detection—accepting or rejecting an item as "old" in the case of a recognition task. Reaction times for high-confidence false alarm judgements were shorter than for correct rejections, although slower than for hits, which may reflect the absence of a clear conscious recollective experience on false alarm trials. There was also a large difference in the variability of reaction times across the three trial types: Table 2 shows that the standard error of the mean of the reaction time was substantially higher for false alarms than for either hits or correct rejections. This table shows the variability across subjects; examination of reaction times for individual subjects revealed a more striking difference. Even for high-confidence responses, the average within subject standard errors for hits and correct rejections were 31 and 40 msec respectively, whereas the average standard error for false alarms was 104 msec. Variability in the timing of cognitive processes also influences the appearance of ERP waveforms. Increased latency jitter across the individual trials contributing to an averaged ERP produces broader, flatter waveforms than averages with a narrower range of latency variability. The degree of latency variability within and between conditions can thus determine the duration of the difference between two ERP waveforms. The reaction times for hits and correct rejections showed relatively small and equivalent variability; Fig. 2 correspondingly shows a relatively short-duration difference between the ERPs in these two conditions. In contrast, the false alarm category showed substantially greater reaction time variability than the hit category; Fig. 3 correspondingly shows a long duration ERP difference between these two conditions.

In contrast to the robust difference between old and new words observed during the recognition tests, the ERPs recorded during the study phase did not reveal any encoding differences among items. In previous research, the typical analysis of study-phase ERPs has compared items that are subsequently recognised (hits at test) to those subsequently unrecognised (misses at test). A convenient term for this conventional analysis is Dm (difference according to memory). Under some study conditions, particularly those that include a "yes/no" study task (e.g. living vs. nonliving judgements), the Dm comparison reveals a larger late positivity for successfully encoded items (Fabiani & Donchin, 1995; Paller, Kutas, & Mayes, 1987; Van Petten & Senkfor, 1996). The present experiment did not allow this typical analysis because of the relatively small number of studied words that were unrecognised at test. Instead, we conducted a related analysis of study-phase words whose individual syllables were presented at test; in this analysis "good encoding" might lead to correct rejection of recognition lures containing only one studied syllable, whereas "poor encoding" might lead to incorrect acceptance of these lures. However, this prediction was not verified; no difference was observed between study-phase items whose related lures were rejected or accepted at test (Fig. 7). The null result of our study-phase analysis lends no support to the hypothesis that false alarms at test are due to a specific defect in prior encoding processes. However, the null result should also be interpreted with some caution for two reasons. First, the present analysis is a novel one, so that there are no prior studies in which "good encoding" is associated with both a correct rejection at test, and some demonstrable ERP effect during the study phase. Second, although the more conventional Dm analysis has yielded a distinction between good versus poor encoding in some dozen studies, two recent studies have also suggested that the presence or absence of a typical Dm effect is contingent on the exact form of the study-phase task (Senkfor & Van Petten, 1998; Van Petten & Senkfor, 1996). Because the non-binary “noun/verb/modifier” study task used in the present experiment is also novel, we do not yet know if it is well suited to detecting encoding differences.

Relationship to Prior Studies of Brain Activity and False Recognition

As noted in the Introduction, previous ERP studies have compared correct and incorrect “old” responses during recognition tests in which the new items bear no special relationship to studied items. These studies yielded the same results as those observed here: false alarms are accompanied by much smaller late positive ERP components than hits (Neville et al., 1986; Sanquist, Rohrbaugh, Syndulko, & Lindsley, 1980; Van Petten & Senlefor, 1996). We are aware of four recent studies examining brain activity when new items are closely related to studied items and thus elicit higher than typical false alarm rates. Each of these used the converging semantic associate paradigm in which the critical lures are semantically associated with studied words; all report great similarity between hits and false alarms (Duzel et al., 1997; Johnson et al., 1997; Schacter, Buckner, Koutstaal, Dale, & Rosen, 1997; Schacter, Reiman, et al., 1996).

Johnson and colleagues (1997) observed essentially identical ERPs for hits and false alarms to related lures, when these related lures were randomly intermixed with old words and unrelated new words. However, Johnson et al., observed larger late positivities for hits than false alarms in a separate experiment which presented old, new, and related new words in distinct blocks. Johnson and colleagues interpret the different pattern of results in the blocked and random versions as reflective of different response criteria. Although the contrast between hits and correct rejections was not analysed in this study, it appears that this basic old/new effect also differed between the participants in the two experiments, so that careful examination of response criteria, confidence levels, and/or individual differences in the converging semantic associates paradigm will be important avenues of future investigation.

Using a randomised design very similar to Johnson et al. (1997), Duzel and colleagues (1997) observed no late positive difference between hits and false alarms to related lures, but instead a larger N400 component for hits. A direct relationship between the N400 component and episodic memory has not been established, but N400 amplitude is exquisitely sensitive to semantic context (see Kutas & Van Petten, 1994, for review). The N400 result may reflect differential semantic context effects for old items and related lures in the specific implementation of the converging semantic associate paradigm used by Duzel and other investigators. Related lures are, by definition, related to a large number of studied items, some of which re-occur during the recognition test. Old items are subject to less semantic context at both study and test, because these words are selected for their semantic relationships to the critical related lures, but not for their relationships to each other (e.g. TOOTH is associated with the critical lure SWEET, but is not as strongly related to HONEY, SOUR, CANDY, etc). At test, old words may thus provide semantic context for the lures, but not for each other. Similarly, the semantic associations between old words and lures may not be directionally symmetric (e.g. SWEET may elicit TOOTH in a production norm, but TOOTH may not elicit SWEET). At test, the truly old items may thus serve as semantic context for the related lures, but the reverse is less likely to occur. Differential semantic context effects during recognition may account for the N400 difference between hits and false alarms observed by Duzel et al. (1997); the strength of this factor across studies using the converging semantic associate paradigm will depend on the exact composition and ordering of recognition word lists.

With positron emission tomographic (PET) measures, Schacter and colleagues (Schacter, Reiman et al., 1996) observed greater blood flow in left temporoparietal cortex during blocks of old words than blocks of related lures. Prefrontal cortex showed a statistical trend for greater blood flow during blocks of related lures. Relative to a fixation baseline, other brain regions such as the medial temporal lobe showed similar increases in blood flow for old words and related lures. The left temporoparietal result was attributed to the retrieval of auditory/phonological information for old words heard during the study phase, information that would be absent in the case of related lures. However, the proximity of this region to those implicated in semantic processing and the possibility of differential semantic context effects in the converging semantic associate paradigm suggests that this result should be interpreted with caution. With functional magnetic resonance (fMRI) measures, Schacter and colleagues (1997) have recently compared old items to related lures with both blocked and randomly intermixed presentation formats. Blood oxygenation levels in left temporoparietal cortex did not discriminate the two word types in either format. The mixed format yielded no prefrontal differences between hits and false alarms, while the possibility of motion artefacts made it difficult to determine if there was differential prefrontal activity in the blocked format. As in the PET study, numerous brain regions showed equivalent haemodynamic changes for both old words and related lures relative to a fixation baseline.

Overall, studies of brain activity using the converging semantic associate paradigm have shown a striking similarity between hits and false alarms, particularly when old and new items are intermixed during the recognition test. These results are in dramatic contrast to the large and robust ERP differences observed here with a random presentation format. Although further investigation of the conjunction lure paradigm is warranted, the discrepancy between the converging semantic associate and conjunction paradigms strongly suggests that not all false alarms are the same. The conjunction lure, syllable lure, and new items used in the present study are not semantically related to studied items, and the results suggest that false alarms to these items are generated by retrieval errors. Other authors have suggested that the related lures in the converging semantic associate paradigm are activated during the study phase (Seamon, Luo, & Gallo, 1998), so that false alarms in this paradigm may reflect encoding errors; at test, semantic lures may indeed be retrieved much like veridical memory traces and elicit similar brain activity. Although our study task was a semantic one, subjects apparently did not use that information very effectively at test and so became entrapped by the orthographic and phonological similarities of the lures. It is also interesting to note that false alarms to related lures are made with higher confidence than false alarms to unrelated words in the converging semantic associates paradigm, but we observed no confidence differences between false alarms to conjunction lures and new words. This suggests that semantic context may affect confidence in a way that familiarity per se does not.

The data from Experiment 1 all converge on the conclusion that familiarity or processing fluency for studied syllables led to a memory illusion at retrieval, but this process was not accompanied by the same LPC that accompanies retrieval of veridical memories.

EXPERIMENT 2

Experiment 1 has provided some insight about when and how memory fails in the conjunction lure paradigm, but additional information can be gleaned from examining the relations between false recognition and other cognitive processes. Experiment 2 is designed to provide such information by using neuropsychological methods in a group of older adults.

Adult age differences in recognition are weaker than those observed in recall (Craik & Jennings, 1992), but our primary interest is in utilising the variability within the elderly population to shed light on the nature of the processes involved in memory conjunction errors. There is general agreement that various cognitive functions do not decrease at the same rate with age, and that variability within a given task is typically greater within elderly than young samples (Albert, 1988; Albert, Duffy, & Naeser, 1987; Welford, 1993).

Neuropathological and neuroimaging studies indicate that prefrontal and medial temporal cortex are particularly sensitive to the normal ageing process, so that both regions are potential candidates for explanations of age-related changes in memory performance (Coffey et al., 1992; Double et al., 1996; Haug, 1985; Martin, Friston, Colebatch, & Frackowiak, 1991; Murphy et al., 1996; Raz, Torres, Spencer, & Acker, 1993; Terry, DeTheresa, & Hansen, 1987; Troncoso, Martin, Dalforrio, & Kawas, 1996). By relating older adults' performance on standardised neuropsychological tests to their experimental performance, we hope to gain insight into either the gross neuroanatomical locus and/or the basic cognitive processes underlying false recognition (for a review of the cognitive neuroscience of false memories, see Schacter, Norman, & Koutstaal, 1998). Later, we describe our approach to characterising individual differences among older adults, but first we briefly review the general background for supposing that one or both of these broadly defined brain regions may be critical for determining false alarm rates.

In a previous study using the conjunction lure paradigm, Kroll and colleagues (1996) observed similar hit rates for normal controls and patients with damage to the hippocampal system. However, patients with left hippocampal damage had false alarm rates of nearly 40% for verbal conjunction lures, as compared to about 14% for older control subjects. Patients with right hippocampal damage showed the same pattern of results for miscombined pictorial stimuli. This result supports the idea that the hippocampal formation plays a vital role in the binding of individual stimulus components into an integrated whole (Kroll et al., 1996). A handful of studies suggest that reduced recognition accuracy within normal elderly samples is due more to inflated false alarm rates than to depressed hit rates (Chao & Knight, 1997; Flicker, Ferris, Crook, & Bartus, 1989; Fulton & Bartlett, 1991; Isingrini, Fontaine, Taconnat, & Duportal, 1995; Norman & Schacter, 1997). Kroll et al.'s results might then suggest that high false alarm rates among the elderly should be attributed to age-related dysfunction of the medial temporal lobe. In contrast, other studies of patients with organic brain damage suggest that prefrontal cortex plays a critical role in determining false alarm rates. Patients with prefrontal damage show only a weak impairment in recognition memory tasks (Wheeler, Stuss, & Tulving, 1995), but this deficit is primarily due to high false alarm rates (Swick & Knight, in press). High false alarm rates in recognition tasks, and confabulation in recall tasks have both been noted in numerous case studies of patients with frontal lesions (Benson et al., 1996; Curran, Schacter, Norman, & Galluccio, 1997; Delbecq-Derousné, Beauvois, & Shallice, 1990; Fischer, Alexander, D'Esposito, & Otto, 1995; N. Kapur & Coughlan, 1980; Moscovitch, 1995; Parkin, Bindschaedler, Harsent, & Metzler, 1996; Rapscak, Polster, Corner, & Rubens, 1994; Schacter, Curran, Galluccio, Milberg, & Bates, 1996). These data then suggest that high false alarm rates among normally ageing individuals might be attributed to mild prefrontal dysfunction.

Overall, the neuropsychological literature indicates that intact medial temporal and prefrontal cortices are both necessary for accurate memory judgements. Moscovitch (1994) has proposed a general theory about the relationship between a *memory system* and a *working-with-memory* system. He argues that the medial temporal lobe is a primary component of the memory system, which is responsible for the actual encoding, storage, and retrieval of memory traces. The working-with-memory system consists primarily of prefrontal cortex, whose roles are to select input, and to monitor or interpret the output of the memory system. High false alarm rates can thus arise from generally weak memories, or a failure to assess the strength of these memories accurately. The working-with-memory system is likely to be particularly taxed in situations that involve a high degree of similarity between old and new items, as in the conjunction lure paradigm.

The results of our first experiment are most compatible with viewing false alarms as arising from a failure to assess memory output adequately. We observed an increasing false alarm rate as the test stimuli became more similar to studied items, indicating that false alarms were indeed based on some form of memory for the studied items.

However, despite the fact that brain activity clearly differentiated these similar items from those actually studied, the participants' behavioural responses did not. We interpret the ERP difference between hits and false alarms as reflecting a distinction between recollection and familiarity, but it appears that this distinction can be uncoupled from the decision or criterion-setting processes that lead to a behavioural response. In Moscovitch's (1994) framework, the failure on a false alarm trial is most likely to reside in the working-with-memory system. To the extent that this functional system is contained in prefrontal cortex, we might expect that the propensity to make false alarms will be associated with other functions ascribed to this general brain region. This prediction is tested by charactersing healthy elderly adults via their performance on neuropsychological tests that are sensitive to the presence of prefrontal lesions in patients with frank organic damage.

The proposed relationship between memory monitoring and prefrontal cortex leads to a secondary prediction about decision confidence during recognition judgements. Experiment 1 yielded the typical result of higher confidence in correct than incorrect decisions; this result indicates an intact relationship between memory strength and memory monitoring. At least one observation indicates that the subjective confidence ratings of patients with frontal lobe damage show little relationship to their actual response accuracy (Rapscak et al., 1998). If the relationship between memory strength and monitoring is weakened with advancing age, or, more specifically, a decline in prefrontal function, the correlation between accuracy and confidence might also be associated with other indices of prefrontal function.

Assessing Individual Differences in the Elderly

The present study uses a methodology developed by Glisky, Polster, and Routhieaux (1995) to categorise cognitive strengths and weaknesses among older adults. One hundred healthy elderly participants were given a battery of standardised neuropsychological tests and the scores subjected to a factor analysis. These tests are listed in Table 3. Five tests sensitive to prefrontal function in brain-damaged patients loaded on a single factor and four tests sensitive to medial temporal and/or diencephalic function in brain-damaged patients loaded on a second factor uncorrelated with the first one. Because we were interested in the differential contributions of the frontal lobes and medial temporal lobes/diencephalon that were independent of age, variance attributable to age was removed from each of the nine test scores through regression analyses (see Glisky et al., for further details). On the basis of the factor analysis, two z-scores were calculated for each subject, representing composite measures of what we will refer to as *Frontal* (F) and *Medial temporal/diencephalic* (M) functions. It is important to note that none of the tests in the frontal battery tap long-term memory; instead they test such abilities as inhibition, set switching, working memory, and strategy formation. Paired associate and other cued recall tests contributed to the M factor score. The two factor scores were then used to categorise participants as having "high Frontal" and "high Medial temporal/diencephalic" function (F+M+), "high Frontal" and "low Medial temporal/diencephalic" function (F+M-), "low Frontal" and "high Medial temporal/diencephalic" function (F-M+), or "low Frontal" and "low Medial temporal/diencephalic" function (F-M-), depending on

Table 3. *Neuropsychological Tests Loading on Frontal and Medial temporal diencephalic*

Factors
Frontal
Modified Wisconsin Card Sorting Task (Hart et al., 1988)
FAS (Benton & Hamsher, 1976)
Mental Arithmetic from Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981)
Backward Digit Span from Wechsler Memory Scale-Revised (WMS-R; Wechsler, 1987)
Mental Control from WMS-R (Wechsler, 1987)
Medial temporal diencephalic
Logical Memory I from WMS-R (Wechsler, 1987)
Verbal Paired Associates I from WMS-R (Wechsler, 1987)
Visual Paired Associates II from WMS-R (Wechsler, 1987)
California Verbal Learning Test-Long Delay Cued Recall (Delis et al., 1987)

whether the z-scores fell above or below zero. We used a subset of these subjects (six from each category) in the present study.

Method

Participants

Participants included 18 young adults (21–39 years old, 8 women, 10 men) and 24 healthy, community-dwelling elderly subjects (66–88 years old; 16 women, 8 men), all of whom gave informed consent. See Table 4 for demographic variables of the older participants broken down by factor score. Three additional young adults participated, but their data were excluded from the analyses due to memory performance (hits – false alarms to new words) more than two standard deviations below the mean.

Materials and Procedure

Materials included 240 of the 440 triads and 70 of the single-syllable filler words described in Experiment 1. As before, assignment of conjunction words to one of the four conditions was rotated across participants, so that each word appeared equally often in each test condition. Unlike Experiment 1, the syllable lures of Experiment 2 consisted of test items sharing either first or second syllables with study items (e.g. study PANIC, test PANDER, or study BLENDER, test PANDER). The stimuli were divided into two lists and split between two study-test cycles.

The experiment consisted of a short practice list and two study-test cycles. Participants were assigned the same study task, and words were presented the same way as in Experiment 1, except that the stimulus duration was increased from 200 to 300 msec for the older adults (stimulus onset asynchrony remained the same).

Results

Effects of Age on Performance

The proportions of two-syllable words called "old" by young and elderly participants are shown in Fig. 9. Older adults correctly classified 71% of the two-syllable words during the recognition tests, as compared to 78% for the young participants, $F(1,40) = 6.25$, $P < .02$. This small impairment in overall accuracy arises from both a slightly lower hit rate and a slightly higher false alarm rate in the elderly group, but neither the hit nor the false alarm rate showed significant age effects when analysed separately, Fs$(1,40) < 2.98$, Ps $< .09$.

Mean confidence ratings for hits, correct rejections, false alarms, and misses were analysed in an ANOVA using age, accuracy, and Old/New as variables. Not surprisingly, both old and young subjects were more confident in their correct than incorrect responses, yielding a main effect of accuracy, $F(1,40) = 111.46$, $P < .0001$. There was also an interaction between accuracy and Old/New due to the fact that old correct items (hits, mean = 2.76 out of 3) had the highest confidence ratings, followed by new correct items (correct rejections, mean = 2.45), old incorrect items (misses, mean = 2.33), and new incorrect items (false alarms, mean = 2.22). Finally, the accuracy by age interaction was also significant, $F(1,40) = 7.86$, $P < .008$. Further analyses revealed that older adults were slightly superior in their metamemory judgement, in that younger adults were a bit more confident that their false recognitions were actually correct, $F(1,40) = 3.04$, $P < .089$. There were no

Table 4. *Mean Demographic Variables of Older Adults by Factor Score*

Category	*Age*	*Education*	*Gender*	*F score*	*M score*
F+M+	73	17 years	3 males, 3 females	.68	.67
F+M−	72	15 years	1 male, 5 females	.63	−.46
F-M+	76	16 years	1 male, 5 females	−.43	.37
F-M-	75	16 years	3 males, 3 females	−.71	−.40

None of the demographic variables were significantly correlated with factor score.

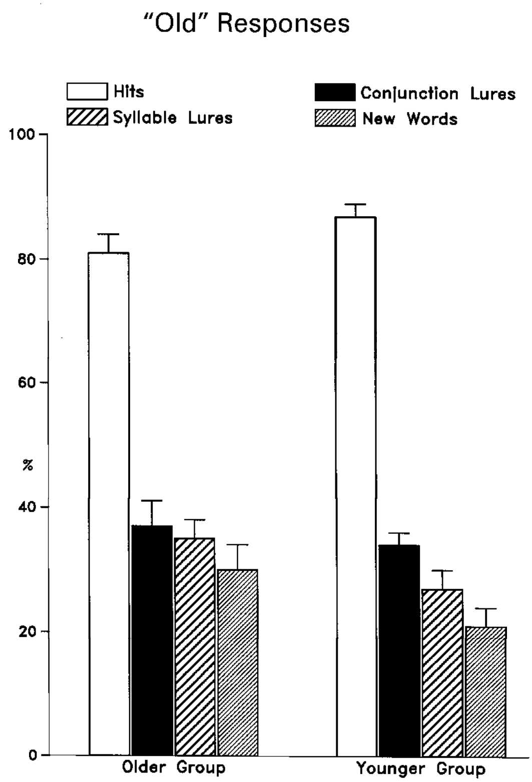

Fig. 9. *Mean percentage of items judged "old" during the recognition test (maximum = 100% in each category) of Experiment 2. Single-syllable items are not included. Error bars show standard error of the mean.*

age differences on confidence ratings for hits, correct rejections, or misses, all *F*s < 1.

Analyses of the false alarm categories in Fig. 9 revealed a significant effect of word type, $F(2,80) = 20.7$, $P < .0001$. There were no significant differences between false alarm rates for first- and second-syllable lures, *F*s < 1.0, so these were collapsed into a category of "syllable lures" in all analyses. There was no interaction between word type and age, indicating that both groups showed the same general pattern of false alarm rates (conjunction lures > syllable lures > new words). Older adults thus had no special propensity to make conjunction lure false alarms.

Effects of Neuropsychological Factor Score on Performance

A breakdown of older adults' recognition accuracy according to the neuropsychological factors is shown in Fig.10. The initial analysis evaluated overall accuracy for the two-syllable words (both hits and correct rejections as a proportion of the total trials), using the F and M factors as independent variables in a 2 × 2 ANOVA. F-participants had lower overall accuracy rates, $F(1,23) = 7.99$, $P < .01$, whereas the M score did not influence overall accuracy. An independent analysis of hit rates indicated that neither factor score yielded a significant main effect, but a somewhat lower hit rate in the F+M- group led to an interaction between the two factors, $F(1,23) = 5.73$, $P < .02$.

False alarms were examined in an ANOVA using the two neuropsychological factors and word type (conjunction lure, syllable lure, or new) as variables. This showed that the reduced overall accuracy in the low frontal groups was due to an elevated false alarm rate, $F(1,20) = 5.74$, $P < .02$. The overall false alarm rate was not influenced by the M factor. However, Fig. 10 shows that the M factor did influence the pattern of false alarms. Unlike either the young adults or the M+ older adults, both M− groups failed to show the typical gradient of false alarms. Instead, these participants displayed nearly identical false alarm rates for conjunction lures, syllable lures, and new words, leading to an interaction between the M factor score and Word Type, $F(2,40) = 3.56$, $P < .03$.

Mean confidence ratings for hits, correct rejections, false alarms, and misses were analysed in an ANOVA using F score, M score, accuracy, and Old/New as variables. Like the young adults in Experiment 1, the older adults were more confident in their correct responses than their incorrect responses, $F(1,20) = 39.06$, $P < .0001$. Our initial hypothesis was that individuals who performed well in the "frontal battery" of neuropsychological tests would also show stronger relationships between accuracy and confidence. This prediction was not supported. The F+M+ participants showed the largest differences between confidence ratings for their correct and incorrect responses, leading to a significant three-way interaction (F score ×

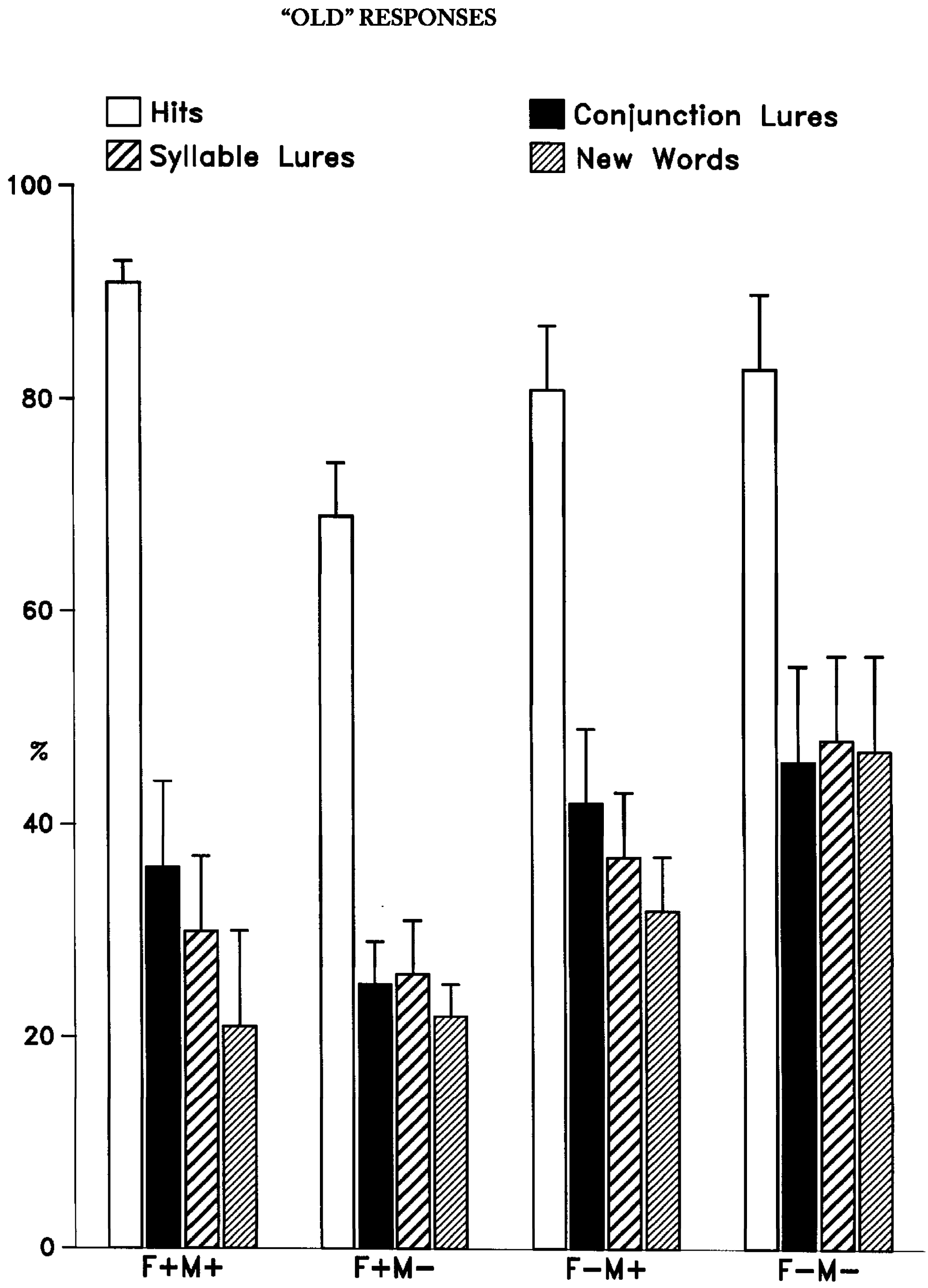

Fig. 10. *Mean percentage of items judged "old" during the recognition test (maximum 100% in each category) of Experiment 2 by neuropsychological factor score. Single-syllable items are not included. Error bars show standard error of the mean.*

M score × Accuracy), $F(1,20) = 6.04$, $P < .05$. Support for our original hypothesis would require that the F+M− group also show a stronger relationship between confidence and accuracy than the two F-frontal groups. However, an analysis comparing these three groups included only a main effect of accuracy $F(1,15) = 16.73$, $P < .001$, but no group interactions. The results thus showed that the individuals who performed the best on the entire neuropsychological test battery also had the strongest confidence/accuracy relationships.

Discussion

The behavioural results of Experiment 2 replicate those of the first experiment; both young and elderly adults overall showed a gradient of false alarm rates determined by the similarity of the recognition lures to studied items. Older adults showed slightly reduced accuracy overall, but no qualitative difference in the balance of hits to false alarms. Although there have been reports of inflated false alarm rates in elderly samples as compared to young, this finding is by no means uniform. Numerous studies also show no greater change in false alarm rate than hit rate with age (Gunter, Jackson, & Mulder, 1992; Mantyla & Backman, 1992; Mark & Rugg, 1998; Parkin & Lawrence, 1994; Swick & Knight, 1997, Trott, Friedman, Ritter, & Fabiani, 1997). Kroll et al., (1996) included control groups of older and younger adults in their study of memory conjunction errors, and they found decreased accuracy overall rather than increased false alarm rates without accompanying hit rate decreases in all but one of their conditions, and this condition appeared to have a ceiling effect insofar as the hit rates were between 95.6 and 100% even for the patient groups. Older adults did not show a disproportionate false alarm rates for conjunction lures in the current study either. This is in contrast to some studies using the converging semantic associates paradigm, which found inflated false alarm rates for older adults in the absence of depressed hit rates (Norman & Schacter, 1997; Tun, Wingfield, Rosen, & Blanchard, 1998), and points to the possibility that making a false alarm in this paradigm may involve different processes than in other paradigms.

Analyses of relationship of confidence to accuracy in the current study also revealed few differences between younger and older adults. In fact, the only age difference indicated that the older adults had somewhat better metamemory than the young when indicating their confidence level when they falsely recognised a word. Although this may seem counter-intuitive, there is quite a bit of evidence supporting the idea that the elderly are not differentially impaired on metamemory tasks (see Light, 1991, for a review).

Despite the overall similarities between older and younger adults, the neuropsychological approach employed here distinguished different patterns of memory performance among the elderly sample. The Frontal factor score, which reflects a composite of cognitive abilities excluding long-term memory, predicted overall false alarm rate. This finding meshes well with reports that patients with prefrontal lesions also show inflated false alarm rates in recognition tasks, while not exhibiting an overall memory impairment (for a comparison between frontal and hippocampal patients, see Swick & Knight, in press). This finding also lends support to Moscovitch's (1994) proposal that prefrontal cortex is critical for monitoring the output of the memory system, and making decisions about the veridicality of that output. This monitoring function is likely to be especially critical in recognition paradigms that include a high degree of similarity between new and studied items, or perhaps even long lists of studied items that increase the possibility of proactive and retroactive interference, two characteristics of the paradigm used here.

Our factor analytic approach to characterising prefrontal function in the elderly population draws on standardised neuropsychological tests generally thought to be sensitive to frontal damage. It has previously been reported that individual differences in the Frontal factor score are related to performance in another memory task known to be sensitive to prefrontal damage—source memory judgements (Glisky et al., 1995; Henkel, Johnson, & De Leonardis, 1998). More recently, we have noted that ERPs recorded at prefrontal scalp sites during a source memory paradigm differentiate healthy older adults characterised as "high Frontal" or "low Frontal" by the factor score approach (Senkfor & Van Petten, unpublished observation). These relationships between the factor analytic method and other methods in cognitive neuroscience are reassuring, but it is important to note that this statistical approach does not offer the direct anatomical link, or the anatomical precision possible in the study of patients with circumscribed lesions. It is possible that none of the tests in the frontal battery are functionally related to recognition memory, or share any anatomical substrates with recognition memory. However, the appeal of using a composite neuropsychological measure is that the tests may collectively offer a broad and stable measure of the integrity of this large brain region. To the extent that mnemonic and non-mnemonic functions of prefrontal cortex are similarly influenced by normal ageing, the observed relationships between the factor score and experimental measures of memory offer at least coarse information about the localisation of these mnemonic processes.

The Medial temporal/diencephalic factor also influenced recognition performance in a distinctive way. Participants with M scores below the mean were just as likely to judge a new word as old when it shared two, one, or no syllables with a studied item. The flat false alarm pattern in the M− group suggests that they simply did not have a good enough memory for the stimulus elements to differentiate between words composed of old and new syllables.

Although it is tempting to consider our M− group as analogous to the amnesic patient groups in other memory conjunction error studies, we feel it is inappropriate to make direct comparisons between our results and the data of Kroll et al. (1996) and Reinitz et al. (1996), because our participants' performance was in the normal range and was only slightly inferior to young participants, whereas their participants were severely impaired. Thus, differences between the current study and prior studies with varied patient groups might be accounted for in two ways. First, it is possible that the relatively small number of trials used in prior studies obscured some effects while exaggerating others. Second, the heterogeneity and severity of deficit of the groups in the various studies prohibits generalising from these groups to healthy elderly individuals. The current study takes the heterogeneity of older adults into account, and provides sufficient numbers of trials to support the conclusion that errors induced by similarity between studied and new items are most likely to be due to monitoring failures of the frontal lobes under circumstances when processing fluency creates confusion between truly old words and recombined syllables.

SUMMARY

In Experiment 1, both behavioural and ERP data indicated that so-called conjunction errors can best be explained by processing fluency during retrieval. The probability of false alarms followed the gradient of the similarity with studied words (conjunction lures > syllable lures > new words), indicating that the elevated false alarm rate to syllable lures cannot easily be attributed to faulty binding. Different classes of false alarms were accompanied by much the same brain activity, regardless of whether the item was a recombination of studied material or not. Hits and false alarms elicited different brain activity, arguing against the notion that conjunction errors are formed at encoding (or during a brief consolidation period) and later retrieved like genuine memories. Thus, although it may be possible under some circumstances to affect conjunction lure false alarm rates through an encoding manipulation, our current data point to retrieval as the time when these errors are actually generated. Experiment 1 further showed that strong memory traces lead to high confidence ratings, and that the late positive component of the ERP reflected confidence in veridical retrieval, but not false confidence when making a recognition error.

Experiment 2 indicated that although there are age differences in recognition accuracy, older adults are not disproportionately likely to recognise conjunction lures falsely. It also indicated that the neuropsychological measures tapping prefrontal cortex are most important in predicting false alarm rates. Taken together, the two experiments suggest that false alarms to conjunction lures are not similar to true recollections, but are products of faulty monitoring processes at retrieval: Items that seem familiar based on their features may generate automatic "yes" responses when there are no accurate recollections to oppose these errors.

REFERENCES

Albert, M.S. (1988). General issues in geriatric neuropsychology. In M.S. Albert & M.B. Moss (Eds.), *Geriatric neuropsychology* (pp. 7–10). New York: Guilford Press.

Albert, M.S., Duffy, F.H., & Naeser, M. (1987). Nonlinear changes in cognition with age and their neuropsychological correlates. *Canadian Journal of Psychology, 41,* 141–157.

Allan, K., & Rugg, M.D. (1997). An event-related potential study of explicit memory on tests of cued recall and recognition. *Neuropsychologia, 35,*(4), 387–397.

Bartlett, F.C. (1932). *Remembering*. Cambridge, MA: Cambridge University Press.

Benson, D.F., Djenderedjian, A., Miller, B.L., Pachana, N.A., Chang, L., Itti, L., & Mena, I. (1996). Neural basis of confabulation. *Neurology, 46,* 1238–1243.

Bentin, S., & McCarthy, G. (1994). The effects of immediate stimulus repetition on reaction time and event-related potentials in tasks of different complexity. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20* (1), 130–149.

Bentin, S., & Peled, B.S. (1990). The contribution of task-related factors to ERP repetition effects at short and long lags. *Memory and Cognition, 18,* 359–366.

Benton, A.L., & Hamsher, K. de S. (1976). *Multilingual Aphasia Examination manual.* Iowa City: University of Iowa.

Berman, S., Friedman, D., & Cramer, M. (1991). ERPs during continuous recognition memory for words and pictures. *Bulletin of the Psychonomic Society, 29,* 113–116.

Besson, M., & Kutas, M. (1993). The many facets of repetition: A cued recall and event related potential analysis of repeating words in same versus different sentence contexts. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19,* 1115–1133.

Besson, M., Kutas, M., & Van Petten, C. (1992). An event-related potential analysis of semantic congruity and repetition effects in sentences. *Journal of Cognitive Neuroscience, 4,* 132–149.

Busey, T.A., Tunnicliff, L., Loftus, G.R., & Loftus, E.F. (1998). *Accounts of the confidence-accuracy relation in recognition memory*. Manuscript submitted for publication.

Chao, L.L., & Knight, R.T. (1997). Age-related prefrontal alterations during auditory memory. *Neurobiology of Aging, 18,* (1), 87–95.

Chao, L.L., Nielsen-Bohlman, L., & Knight, R.T. (1995). Auditory event-related potentials dissociate early and late memory processes. *Electroencephalograpy and Clinical Neurophysiology, 96,* 157–168.

Clark, S.E. (1997). A familiarity-based account of confidence-accuracy inversions in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 232–238.

Coffey, C.E., Wilkinson, W.E., Parashos, I.A., Soady, S.A., Sullivan, R.J., Patterson, L.J., Figiel, G.S., Webb, M.C., Spritzer, C.E., & Djang, W.T. (1992). Quantitative cerebral anatomy of the aging human brain. *Neurology, 42,* 527–536.

Crak F.I.M., & Jennings, L.M. (1992). Human memory. In F.I.M. Craik & T.M. Salthouse (Eds.), *Handbook of aging and cognition* (pp. 51–1100). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Craile, F.I.M., Morris, L.W., Morris, R.G., & Loewen, E.R. (1990). Relations between source amnesia and frontal lobe functioning in older adults. *Psychology and Aging, 5,* 148–151.

Curran, T., Schacter, D.L., Norman, K.A., & Galluccio, L. (1997). False recognition after a right frontal lobe infarction: Memory for general and specific information. *Neuropsychologia, 35,* (7), 1035–1049.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58,* 17–22.

Delbecq-Derousné, L, Beauvois, M.F., & Shallice, T. (1990). Preserved recall vs. impaired recognition: A case study. *Brain, 113,* 1045–1074.

Delis, D.C., Kramer, L., Kaplan, E., & Ober, B.A. (1987). *The California Verbal Learning Test.* San Antonio, TX: Psychological Corporation.

Dorfinan, J. (1994). Sublexical components in implicit memory for novel words. *Journal of Experimental Psychology, 20* (5), 1108–1125.

Double, K.L., Halliday, G.M., Krill, J.J., Harasty, J.A., Cullen, K., Brooks, W.S., Creasey, H., & Broe, G.A. (1996).Topography of brain atrophy during normal aging and Alzheimer's disease. *Neurobiology of Aging, 17,* 513–521.

Duzel, E., Yonelinas, A.P., Mangun, G.R., Heinze, H.-J., & Tulving, E. (1997). Event related brain potential correlates of two states of conscious awareness in memory. *Proceedings of the National Academy of Sciences, 94,* 5973–5978.

Eichenbaum, K. (1994). The hippocampal system and declarative memory in humans and animals: Experimental analysis and historical origins. In D.L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 147–201). Cambridge, MA: MIT Press.

Eichenbaum, H., & Bunsey, M. (1995). On the binding of associations in memory: Clues from studies on the role of the hippocampal region in paired-associate learning. *Current Directions in Psychological Science, 4,* 19–23.

Fabiani, M., & Donchin, E. (1995). Encoding processes and memory organization: A model of the von Restorff effect. *Journal of Experimental Psychology: Learning, Memory and Cognition, 21,* 224–240.

Fabiani, M., Karis, D., & Donchin, E. (1986). P300 and recall in an incidental memory paradigm. *Psychophysiology, 23,* 298–308.

Fischer, R.S., Alexander, M.P., D'Esposito, M.D., & Otto, R. (1995). Neuropsychological and neuroanatomical correlates of confabulation. *Journal of Clinical and Experimental Neuropsychology, 17,* 20–28.

Flicker, C., Ferris, S.H., Crook, T., & Bartus, R.T. (1989). Age differences in the vulnerability of facial recognition memory to proactive interference. *Experimental Aging Research, 15* (3–4), 189–194.

Francis, W.N., & Kuçera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar.* Boston, MA: Houghton Mifflin.

Fulton, A., & Bartlett, J.C. (1991). Young and old faces in young and old heads: The factor of age in face recognition. *Psychology and Aging, 6* (4), 623–630.

Gardiner, J.M., & Java, R.I. (1990). Recollective experience in word and nonword recognition. *Memory and Cognition, 18,* 23–20.

Glisky, E.L., Polster, M.R., & Routhicaux, B.C. (1995). Double dissociation between item and source memory. *Neuropsychology, 9* (2), 229–235.

Gunter, T.C., Jackson, J.L., & Mulder, G. (1992). An electrophysiological study of semantic processing in young and middle-aged academics. *Psychophysiology, 29,* 38–54.

Hart, R.P., Kwentus, J.A., Wade, J.B., & Taylor, R. (1988). Modified Wisconsin Card Sorting test in elderly normal, depressed and demented patients. *Clinical Neuropsychologist, 2,* 49–56.

Haug, H. (1985). Are neurons of the human cerebral cortex really lost during aging? A morphometric examination. In J. Traber & W.H. Gispen (Eds.), *Senile dementia of the Alzheimer type.* Berlin: Springer.

Henkel, L.A., Johnson, M.K., & De Leonardis, D.M. (1998). Aging and source monitoring: Cognitive processes and neuropsychological correlates. *Journal of Experimental Psychology, 127* (3), 251–268.

Hillyard, S.A., Squires, K.C., Bauer, J.W., & Lindsay, P.H. (1971). Evoked potential correlates of auditory signal detection. *Science, 172,* 1357–1360.

Isingrini, M., Fontaine, R., Taconnat, L., & Duportal, A. (1995). Aging and encoding in memory—False alarms and decision criteria in a word-pair recognition task. *International Journal of Aging and Human Development, 41* (1), 79–88.

Jacoby, L.L. (1991). A process dissociation framework: Separating automatic from intentional use of memory. *Journal of Memory and Language, 30,* 513–541.

Janowsky, J.S., Shimamura, A.R., & Squire, L.K. (1989). Source memory impairments in patients with frontal lobe lesions. *Neuropsychologia, 27,* 1043–1056.

Jasper, H.H. (1958). The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology, 10,* 371–375.

Jha, A.P., Kroll, N.E.A., Baynes, K., & Gazzaniga, M.S. (1997). Memory encoding following complete callosotomy. *Journal of Cognitive Neuroscience, 9* (1), 143–159.

Johnson, M.K., Hashtroudi, S., & Lindsay, S.D. (1993). Source monitoring. *Psychological Bulletin, 114,* 3–28.

Johnson, M.K., Nolde, S.F., Mather, M., Kounios, J, Schacter, D.L., & Curran, T. (1997). The similarity of brain activity associated with true and false recognition depends on test format. *Psychological Science, 8* (3), 250–257.

Johnson, M.K., & Raye, C.L. (in press). Cognitive brain mechanisms of false memories and beliefs. In D.L. Schacter & E. Scarry (Eds.), *Memory, brain, and belief.* Cambridge, MA: Harvard University Press.

Kapur, N., & Coughlan, A.K. (1980). Confabulation and frontal lobe dysfunction. *Journal of Neurology, Neurosurgery and Psychiatry, 43,* 461-463.

Kapur, S., Craik, F.I.M., Jones, C., Brown, G.M., Houle, S., & Tulving, E. (1995). Functional role of prefrontal cortex in memory retrieval: A PET study. *Neuroreport, 6,* 1880–1884.

Karis, D., Fabiani, M., & Donchin, E. (1984). "P300" and memory: Individual differences in the von Restorff effect. *Cognitive Psychology, 16,* 177–216.

Kroll, N.E.A., Knight, R.T., Metcalfe, L., Wolf, E.S., & Tulving, E. (1996). Cohesion failure as a source of memory illusions. *Journal of Memory and Language, 35,* 176–196.

Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 83–143). New York: Academic Press.

Light, L.L. (1991). Memory and ageing: Four hypotheses in search of data. *Annual Review of Psychology, 42,* 333–376.

Loftus, E.F., Miller, D.G., & Burns, H.J. (1978). Semantic integration of verbal information into visual memory. *Journal of Experimental Psychology, 4,* 19–31.

Mantyla, T., & Backman, L. (1992). Aging and memory for expected and unexpected objects in real-world settings. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1298–1309.

Mark, R.E., & Rugg, M.D. (1998). Age effects on brain activity associated with episodic memory retrieval: An electrophysiological study. *Brain, 121,* 868–873.

Martin, A.J., Friston, K.J., Colebatch, J.G., & Frackowiak, R.S.J. (1991). Decreases in regional cerebral blood flow with normal aging. *Journal of Cerebral Blood Flow and Metabolism, 11,* 684–689.

Moscovitch, M. (1994). Memory and working with memory: Evaluation of a component process model and comparisons with other models. In D.L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 269–310). Cambridge MA: MIT Press.

Moscovitch, M. (1995). Confabulation. In D.L. Schacter (Ed.), *Memory distortion: How minds, brains and societies reconstruct the past* (pp. 227–251). Cambridge, MA: Harvard University Press.

Moscovitch, M., & Winocur, G. (1995). Frontal lobes, memory and ageing. In J. Grafman, K.J. Holyoak, & F. Boller (Eds.), *Structure and functions of human prefrontal cortex* (pp. 119–150). New York: New York Academy of Sciences.

Murdock, B.B., & Dufty, P.O. (1972). theory and recognition memory. *Journal of Experimental Psychology, 94,* 284–290.

Murphy, D.G.M., DeCarli, C., McIntosh, A.R., Daly, E., Mentis, M.L., Pictrini, P., Szczepanik, J., Schapiro, M.B., Grady, C.L., Horwitz, B., & Rapoport, S.I. (1996). Sex differences in human brain morphometry and metabolism—an in vivo quantitative magnetic resonance and position emission tomography study on the effect of aging. *Archives of General Psychiatry, 53,* 585–594.

Neville, H.J., Kutas, M., Chesney, G., & Schmidt, A. (1986). Event-related brain potentials during the initial encoding and subsequent recognition memory of congruous and incongruous words. *Journal of Memory and Language, 25,* 75–92.

Norman, K.A., & Schacter, D.L. (1997). False recognition in younger and older adults - Exploring the characteristics of illusory memories. *Memory Cognition, 25* (6), 838–848.

Paller, K A., & Gross, M. (1998). Brain potentials associated with perceptual priming versus explicit remembering during the repetition of visual word-form. *Neuropsychologia, 36,* 559–571.

Paller, K.A., & Kutas, M. (1992). Brain potentials during memory retrieval provide neurophysiological support for the distinction between conscious recollection and priming. *Journal of Cognitive Neuroscience, 4,* 375–391.

Paller, K.A., Kutas, M., & Mayes, A.R. (1987). Neural correlates of encoding in an incidental learning paradigm. *Electroencephalography and Clinical Neurophysiology, 67,* 360–371.

Paller, K.A., Kutas, M., & McIsaac, H.K. (1998). An electrophysiological measure of priming of visual word-form. *Consciousness and Cognition, 7,* 54–66.

Parkin, A.L., Bindschaedler, C., Harsent, L., & Metzler, C. (1996). Pathological false alarm rates following damage to left prefrontal cortex. *Brain and Cognition, 32* (1), 14–27.

Parkin, A.L., & Lawrence, A. (1994). A dissociation in the relation between memory tasks and frontal lobe tests in the normal elderly. *Neuropsychologia, 32,* 1523–1532.

Paul, D.D., & Sutton, S. (1972). Evoked potential correlates of response criterion in auditory signal detection. *Science, 177,* 362–364.

Rapcsak, S.Z., Kaszniak, A.W., Reminger, S.L., Glisky, M.L., Glisky, E.L., & Comer, J.F. (1998). Dissociation between verbal and autonomic measures of memory following frontal lobe damage. *Neurology, 50,* 1259–1265.

Rapcsak, S.Z., Polster, M.R., Comer, LS., & Rubens, A.B. (1994). False recognition and misidentification of faces following right hemisphere damage. *Cortex, 30,* 565–583.

Raz, N., Torres, I.J., Spencer, W.D., & Acker, J.D. (1993). Pathoclysis in aging human cerebral cortex: Evidence from in vivo MR1 morphometry. *Psychobiology, 21,* 151–160.

Regan, D. (1989). *Human-brain electrophysiology.* New York: Elsevier.

Reinitz, M.T., & Demb, J. (1994). Implicit and explicit memory for compound words. *Memory and Cognition, 22* (6), 687–694.

Reinitz, M.T., Lammers, W.J., & Cochran, B.P. (1992). Memory-conjunction errors: Miscombination of stored stimulus features can produce illusions of memory. *Memory and Cognition, 20* (1), 1–11.

Reinitz, M.T., Morrissey, L., & Demb, J. (1994). Role of attention in face encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 161–168.

Reinitz, M.T., Verfaellie, M., & Milberg, W.P. (1996). Memory conjunction errors in normal and amnesic subjects. *Journal of Memory and Language, 35,* 286–299.

Roder, B.J., Faneuf, S., Burch, W., & Connors, M. (1996). *Memory for conjoined-concepts.* Poster presentation, Psychonomic Society Annual Meeting.

Roediger, H.L., III, & McDermott, K.B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803–814.

Rudy, L.W., & Sutherland, R.J. (1994). The memory-coherence problem, configural associations, and the hippocampal system. In D.L. Schacter & E. Tulving (Eds.), *Memory systems 1994* (pp. 19-146). Cambridge, MA: MIT Press.

Rugg, M.D., & Doyle, M.C. (1992). Event-related potentials and recognition memory for low- and high-frequency words. *Journal of Cognitive Neuroscience, 4,* 69–79.

Rugg, M.D., & Doyle, M.C. (1994). Event-related potentials and stimulus repetition in direct and indirect tests of memory. In H.J. Heinze, T.F. Munte, & G.R. Mangun (Eds.), *Cognitive electrophysiology* (pp. 124–138). Boston, MA: Birkhauser.

Sanquist, T.F., Rohrbaugh, J.W., Syndulko, K., & Lindsley, D.B. (1980). Electrocortical signs of levels of processing: Perceptual analysis and recognition memory. *Psychophysiology, 17,* 568-576.

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

Schacter, D.L., Curran, T., Galluccio, L., Milberg, W.P., & Bates, LF. (1996). False recognition and the right frontal lobe: A case study. *Neuropsychologia, 34* (8), 793–808.

Schacter, D.L., Harbluk, J.L., & McLachlan, D.R. (1984). Retrieval without recollection: An experimental analysis of source amnesia. *Journal of Verbal Learning and Behavior, 23,* 593–611.

Schacter, D.L., Norman, K.A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology, 49,* 289–318.

Schacter, D.L., Reiman, E.M., Curran, T., Yun, L.S., Bandy, D., McDermott, K.B., & Roediger, H.L. III (1996). Neuroanatomical correlates of veridical and illusory recognition memory: Evidence from positron emission tomography. *Neuron, 17,* 267–274.

Seamon, J.G., Luo, C.R., & Gallo, D.A. (1998). Creating false memories of words with or without recognition of list items. Evidence for nonconscious processes. *Psychological Science, 9,* 20–26.

Senkfor, A.J., & Van Petten, C. (1998). Who said what? An event-related potential investigation of source and item memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 1005–1025.

Smith, M.E. (1993). Neurophysiological manifestations of recollective experience during recognition memory judgements. *Journal of Cognitive Neuroscience, 5,* 1–13.

Smith, M.E., & Guster, K. (1993). Decomposition of recognition memory event-related potentials yields target, repetition, and retrieval effects. *Electroencephalography and Clinical Neurophysiology, 86,* 335-343.

Swick, D., & Knight, R.T. (1997). Event-related potentials differentiate the effects of aging on word and nonword repetition in explicit and implicit memory tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 123–142.

Swick, D., & Knight, R.T. (in press). Contributions of prefrontal cortex to recognition memory: Electrophysiological and behavioral evidence. *Neuropsychology.*

Terry, R.D., DeTheresa, R., & Hansen, L.A. (1987). Neocortical cell counts in normal human adult aging. *Annals of Neurology, 21,* 530–539.

Troncoso, L.C., Martin, L.J., Dalforno, G., & Kawas, C.H. (1996). Neuropathology in controls and demented subjects from the Baltimore longitudinal study of aging. *Neurobiology of Aging, 17,* 365–371.

Trott, C.T., Friedman, D., Ritter, W., & Fabiani, M. (1997). Item and source memory: Differential age effects revealed by event-related potentials. *Neuroreport, 8,* 3373–3378.

Tulving, E., Kapur, S., Markowitsch, H.J., Craik, F.I.M., Habib, R., & Houle, S. (1994). Neuroanatomical correlates of retrieval in episodic memory: Auditory sentence recognition. *Proceedings of the National Academy of Science, 91,* 2012–2015.

Tulving, E., Markowitsch, H.J., Craik, F.I.M., Habib, R., & Houle, S. (1996). Novelty and familiarity activations in PET studies of memory encoding and retrieval. *Cerebral Cortex, 6,* 71–79.

Tun, P.A., Wingfield, A., Rosen, M.J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology and Aging, 13* (2), 230–241.

Underwood, B.J., Kapelak, S.M., & Malmi, R.A. (1976). of discrete verbal units in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 293–300.

Underwood, B.J., & Zimmerman, J. (1973). The syllable as a source of error in multisyllable word recognition. *Journal of Verbal Learning and Verbal Behavior, 12,* 338–344.

Van Petten, C., & Kutas, M. (1988). The use of event-related potentials in the study of brain asymmetries. *International Journal of Neuroscience, 39,* 91–99.

Van Petten, C., & Senkfor, A.J. (1996). Memory for words and novel visual patterns: Repetition, recognition, and encoding effects in the event-related brain potential. *Psychophysiology, 33,* 491–508.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised manual.* New York: Psychological Corporation.

Wechsler, D. (1987). *Wechsler Memory Scale—Revised manual.* New York: Psychological Corporation.

Welford, A.T. (1993). The gerontological balance sheet. In J. Cerella, J. Rybash, W. Hoyer, & M.L. Commons (Eds.), *Adult information processing: Limits on loss* (pp. 3–10). New York: Academic Press.

Wheeler, M.A., Stuss, D.T., & Tulving, E. (1995). Frontal lobe damage produces episodic memory impairment. *Journal of the International Neuropsychologica Society, 1,* 525–536.

Wilding, E.L., & Rugg, M.D. (1996). An event-related potential study of recognition memory with and without retrieval of source. *Brain, 119,* 889–905.