Electroencephalography and clinical Neurophysiology, 1987, 67: 360–371
Elsevier Scientific Publishers Ireland, Ltd.

EEG 03280

Neural correlates of encoding in an incidental learning paradigm [1]

Ken A. Paller, Marta Kutas and Andrew R. Mayes

Department of Neurosciences, University of California, San Diego, La Jolla, CA 92093 (U.S.A.)

(Accepted for publication: 22 February, 1987)

Summary Event-related brain potentials (ERPs) were recorded during an incidental learning paradigm. Recall and recognition were better for words initially presented in tasks requiring semantic decisions (i.e., 'is it living?' or 'is it edible?') than for words in tasks requiring non-semantic decisions. ERPs elicited during performance of these tasks were predictive of subsequent memory performance. A late positive ERP elicited by words later recalled or recognized was larger than that elicited by words later forgotten. This enhanced positivity for to-be-remembered words could be accounted for, in part, by the fact that words in semantic tasks were remembered better and elicited larger ERPs than did words in non-semantic tasks. Similarly, words followed by affirmative rather than negative decisions were associated both with better recognition and with larger ERPs. However, ERPs were sensitive to processes that influenced later memory performance even within an individual semantic task and within the affirmative decision condition. In addition, results showed that the ERP differences based on later memory performance did not necessarily arise from amplitude variation in P3 waves that occurred at the same time.

Key words: Memory; ERPs; P3; Humans

Memory performance can be viewed as an outcome of encoding, consolidation, and retrieval processes. The levels-of-processing framework explained memory performance with an emphasis on encoding, the initial processing of stimulus information (e.g., Craik and Lockhart 1972; Craik and Tulving 1975; Lockhart et al. 1976). Experimental manipulations of encoding can have profound effects. For example, memory is generally better when the meaning of the material is more fully processed (e.g., Craik 1979). Delineating the precise nature of these encoding processes continues to be a major research goal in experimental psychology (Tulving 1983). Towards this end, analyses of behavioral performance can be supplemented by recording electrical activity produced by the brain during perceptual and cognitive events (e.g., Hillyard and Kutas 1983).

Recent applications of such electrophysiological techniques to the study of human memory have shown that event-related brain potentials (ERPs) elicited by words can be predictive of subsequent memory performance (Sanquist et al. 1980; Karis et al. 1984; Fabiani et al. 1985, 1986; Neville et al. 1986; Kutas 1987). In these studies, the brain response to each word was sorted according to whether or not the word was remembered on subsequent memory tests. ERPs elicited during the encoding of words that were later remembered were more positive than ERPs to words that were not remembered. This difference between responses to remembered and forgotten

[1] Supported by NSF Grant BNS83 05525, NIMH Predoctoral Fellowship MH09128 to K.A.P., and Research Scientist Development Award USPHS 1KO2 MH00322 to M.K. A.R.M. was on leave from the Department of Psychology, University of Manchester. A preliminary report was presented at the 25th annual meeting of the Society for Psychophysiological Research (Paller et al. 1985). We thank Steven Hillyard for helpful comments.

Correspondence to: Ken Paller, Neuropsychology Laboratory (116B1), VA Medical Center, West Haven, CT 06516 (U.S.A.).

words will be referred to as 'Dm' (operationally defined as any ERP Difference based on later memory performance).

If Dm is a direct measure of differential processing that determines how well a word is encoded and/or consolidated in memory, then understanding this measure would have profound implications for the future study and understanding of memory. However, it is presently unclear which aspects of information processing Dm reflects. For example, Dm may reflect perceptual encoding, semantic encoding, or processes related to arousal or effort (which have an indirect influence on memory).

A number of factors operative at the time of encoding can influence later memory performance. The influence of semantic processing on memory and on ERPs was examined in the present experiment, using an incidental learning paradigm in which processing carried out during encoding was systematically controlled. Perhaps Dm has appeared solely as a consequence of differential amounts of semantic processing for to-be-remembered and to-be-forgotten words. This outcome could arise if (a) ERPs were specifically sensitive to the presence of semantic processing and (b) words processed semantically were remembered better.

In addition, this study was designed to examine possible relationships between Dm and ERP components that occur in the same time frame. The P3 (or P300), in particular, is an ERP elicited upon detection and evaluation of a relevant stimulus that is unpredictable in some manner (Pritchard 1981; Fabiani et al. 1987). P3-like responses elicited in a wide variety of situations may reflect a common cognitive process, but the precise nature of this process is still under debate. Notably, Donchin (1981) proposed that P3 amplitude indexes the incorporation of new information into existing representations of the environment. One prediction from this hypothesis would be that the larger the P3 to an item, the more likely that it will be remembered. The results of Karis et al. (1984) were interpreted as support for this hypothesis, since the amplitude of a late positive wave designated P3 was predictive of memory performance. However, not all factors that alter P3 amplitude necessarily impact memory, and P3 and Dm may be partially independent.

Method

Subjects

Sixteen adults (8 males and 8 females) were paid to participate in the experiment. They were all right-handed college students between the ages of 18 and 28 years old (mean = 21 years).

Stimuli

A total of 600 words were selected such that half contained 2 vowels and half contained more than 2 vowels, half referred to living things and half referred to non-living things, and half referred to edible things and half referred to non-edible things. Of the 600 words, 300 words were randomly chosen to be presented during the input stage on a video screen under the control of a microcomputer. Words contained 3–11 letters, each of which subtended a vertical visual angle of 1°. Stimulus duration was 200 msec. The presentation rate was one word every 2 sec.

Design

The experiment consisted of 3 stages: input, recall, and recognition. Subjects were given instructions prior to each stage. In addition, subjects were periodically reminded to restrict body and eye movements and to stay relaxed but alert.

Input stage. A question displayed below the screen indicated which of 4 processing tasks was to be performed. The 4 tasks were as follows: (E) Is it edible?, (L) Is it living?, (V) Are there exactly 2 vowels?, (A) Are first and last letters in alphabetical order? The distinction between the semantic tasks (E and L) and the non-semantic tasks (V and A) refers to the fact that processing of a word's meaning is necessary for correct performance only in the case of semantic tasks. Subjects indicated an affirmative response or a negative response after each word by pressing 1 of 2 buttons. For half of the subjects, affirmative responses were made with the right hand and negative responses were made with the left hand; the opposite combination was used for the remaining subjects. Subjects were advised to respond as quickly as possible without making errors.

Prior to EEG recording, subjects were given 5 practice words with each task. Thereafter, the processing task was changed after every 25 words such that a total of 75 words were processed with each of the 4 tasks. The specific sets of words used with each task were counterbalanced across subjects.

Subjects were not told that memory tests were forthcoming, and a subsequent query showed that none suspected the tests. Immediately after the final input word, subjects were instructed to count backwards by threes. This procedure for preventing rehearsal continued for 1 min, at which point subjects were given a blank sheet of paper and a pen.

Free recall test. Approximately 5–10 min were allotted for subjects to write all input words that they could recall. A 20 min delay then ensued during which subjects listened to tone sequences for a different experiment (the results of which were reported by Paller 1986).

Recognition test. The video screen was used to present the entire set of 600 words. After each word was presented, subjects responded via key presses to 3 questions in the following order: (1) had the word also been presented during the input stage?, (2) how confident was this recognition decision?, and (3) was the word part of their present vocabulary?. Results from the third question will not be discussed further, since it was almost always answered in the affirmative. The recognition test was self-paced, but usually lasted about 60 min. Because the order of the words was randomized, the delay between a word's appearance during the input stage and during the recognition test varied across a large range (30–120 min). Chance recognition performance was 50% correct.

ERP recording

At the beginning of the experiment, Ag-AgCl electrodes were affixed to the subject's head with collodion. The following electrode sites of the 10-20 system (Jasper 1958) were used: Fz, Cz, Pz, Oz, F3, F4, T3, T4, P3, P4, left mastoid, and right mastoid. During the input and recognition stages, EEG was amplified with an 8 sec time constant and 60 Hz half-amplitude upper cutoff. In addition, horizontal eye movements were recorded from electrodes placed lateral to each eye and vertical eye movements were recorded from an electrode below the right eye referred to the left mastoid electrode. Electrophysiological and behavioral data were digitized on-line at a rate of 170 samples/sec.

Here we report only on ERPs elicited during the input stage. Average ERPs were computed for each scalp electrode referenced to the mean of left and right mastoid recordings. (These data were derived from original recordings with a left mastoid reference by subtracting one-half the amplitude of the right mastoid to left mastoid recording.) Averaging epochs (trials) were 1500 msec in duration, including 200 msec prior to stimulus onset. Trials were rejected if a computer algorithm detected electro-ocular or other artifacts. The mean percentage of trials rejected was 16%. Trials were sorted and ERPs computed for each level of the following factors.

Independent variables: (1) task (*edible, living, vowel,* or *alphabet*); (2) word frequency (high or low); (3) typicality (normal or atypical decisions). For example, with task E the word 'burrito' involves a normal decision, whereas the word 'mammal' involves some ambiguity and would be an atypical decision. This applied to semantic tasks only, since decisions in non-semantic tasks were unambiguously correct or incorrect. Values for these last two variables, as well as for decision accuracy (below), were based on ratings made by 10 people who did not serve as subjects for the ERP recordings.

Performance variables: (1) decision (affirmative or negative task decision); (2) reaction time (above or below the median); (3) decision accuracy (correct or incorrect task decision); (4) recall (recalled or not recalled); (5) recognition (recognized or not recognized); (6) recognition confidence (high or low).

ERPs also were computed for relevant 2-factor and higher-order interactions, but many interactions were not analyzed due to the small number of trials in some conditions. In order to assess experimental effects, 2 ERP measurements were used: (1) mean amplitude in the 250–400 msec latency range and (2) mean amplitude in the 400–800 msec latency range. These measurements were made relative to a 200 msec prestimulus baseline. The measurement intervals were selected on the basis of effects apparent in ERPs averaged across subjects. ERP measurements and behavioral results were submitted to repeated measures analyses of variance (ANOVAs) using the 5% significance level. Recognition scores and reaction times were analyzed by 2-way ANOVAs (4 processing tasks × 2 levels of one of the following factors: word frequency, typicality, decision, reaction time, or decision accuracy). ERP measurements were analyzed by 4-way ANOVAs (4 processing tasks × 2 levels of one of the above factors × 2 levels of recall or recognition × 6 or 10 electrode locations). Because some higher-order conditions included few or no trials, lower-order ANOVAs were used whenever necessary. Sources of significant effects involving electrode as a factor were determined by separate tests on recordings from each electrode (using the Bonferroni correction for inflated probability of a type I error). Significant effects involving processing task were further analyzed by pairwise comparisons using the Tukey test (Keppel 1982).

Results

Memory performance

Mean percent correct across subjects was 73% on the recognition test and 8% on the recall test. Recognition scores analyzed in relation to the independent variables are shown in Table I. Processing task had a significant influence on recognition (F (3, 45) = 47.3, $P < 0.001$) as the 2 scores from semantic tasks were higher than the 2 scores from non-semantic tasks, which also differed significantly from each other. Similarly, recall performance was better in the semantic tasks (12% correct) than in the non-semantic tasks (3% correct).

Recognition memory scores were significantly higher for words associated with affirmative deci-

TABLE I

Behavioral and electrophysiological results.

Factor	Number of words	Recognition memory (% correct)	Reaction time (msec)	LPC	Dm
All words	300	73 (8)	987 (40)	3.1 (0.4)	1.0 (0.1)
Task E (edible)	75	**83** (12)	**879** (38)	**3.6** (0.4)	2.2 (0.4)
Task L (living)	75	**83** (7)	**934** (42)	**3.7** (0.4)	1.1 (0.4)
Task V (vowel)	75	**65** (10)	**986** (47)	**3.6** (0.4)	0.4 (0.2)
Task A (alphabet)	75	**57** (13)	**1161** (50)	**1.5** (0.4)	0.3 (0.2)
Affirmative decisions	149	**76** (18)	993 (39)	**3.6** (0.2)	1.5 (0.2)
Negative decisions	148	**71** (15)	979 (42)	**2.7** (0.2)	0.5 (0.2)
Fast reaction time	148	73 (16)	**779** (32)	**3.6** (0.4)	0.9 (0.2)
Slow reaction time	151	74 (16)	**1194** (48)	**2.5** (0.4)	0.8 (0.2)
High-frequency words	179	73 (17)	**970** (38)	**3.3** (0.4)	1.2 (0.2)
Low-frequency words	121	73 (17)	**1017** (43)	**2.7** (0.4)	1.1 (0.2)
Normal decisions	130	**84** (8)	**889** (39)	3.7 (0.4)	1.6 (0.3)
Atypical decisions	20	**89** (11)	**1009** (44)	3.4 (0.5)	2.5 (0.8)
Correct decisions	251	73 (16)	**976** (38)	**3.3** (0.4)	1.0 (0.1)
Incorrect decisions	42	73 (21)	**1198** (76)	**2.0** (0.4)	1.6 (0.5)

Note: Numbers in parentheses indicate standard errors of the mean. Numbers in boldface indicate significant main effects. LPC and Dm are mean amplitude measurements in μV, from lateral parietal and midline electrodes in the 400–800 msec latency range.

sions than for words associated with negative decisions (F (1, 15) = 11.9, $P < 0.004$). As shown in Table II, this effect depended upon processing task (F (3, 45) = 12.9, $P < 0.001$) and was significant only in semantic tasks. Also, words involving atypical decisions were remembered better than words involving normal decisions (F (1, 15) = 10.3, $P < 0.006$). None of the other variables (reaction time, word frequency, and decision accuracy) influenced recognition memory.

Reaction time

Mean reaction times for the various conditions are shown in Table I. Although the reaction time distributions for the 4 tasks overlapped to a great extent, there was a significant main effect of processing task (F (3, 45) = 34.2, $P < 0.001$). The mean reaction time (RT) during the *alphabet* task was significantly slower than the mean RT in any of the 3 remaining tasks. The only other significant difference between pairs of means was between those from the *vowel* task and the *edible* task. RT was also influenced by 3 other factors: (1) decisions on high-frequency words were quicker than decisions on low-frequency words (F (1, 15) = 48.6, $P < 0.001$), (2) normal decisions were quicker than atypical decisions (F (1, 15) = 34.0, $P < 0.001$), and (3) correct decisions were quicker than incorrect decisions (F (1, 15) = 12.1, $P <$ 0.003). Decision type (affirmative versus negative) did not reliably influence reaction time (F (1, 15) = 1.0).

ERPs and processing task

In general, ERPs elicited during the input stage did not differ across conditions before a latency of about 250 msec. The mean amplitude measured in the 400–800 msec latency range was highly positive at all posterior electrodes, and will henceforth be designated the late positive complex or LPC [2]. As shown in Fig. 1 and in Table I, LPC differed

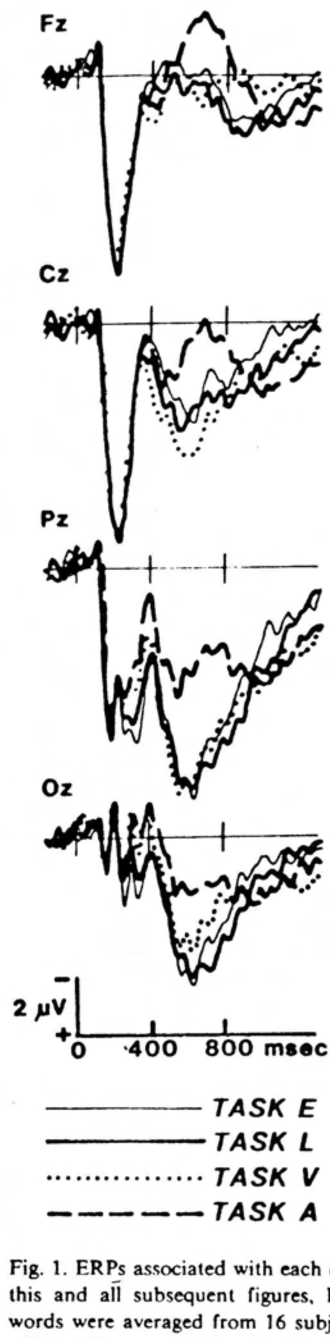

Fig. 1. ERPs associated with each of the 4 processing tasks. In this and all subsequent figures, ERPs elicited by the input words were averaged from 16 subjects for a 1500 msec epoch with a 200 msec prestimulus baseline. Positive activity at the active site is plotted downwards. A large positive deflection (peaking at about 600 msec at the Pz electrode) resembled the P3 wave. At all electrodes, the task requiring the alphabetical judgment was associated with less positive activity (or equivalently, greater negative activity).

[2] The term LPC is used as a descriptive label for the mean amplitude measurement, which bears no necessary relationship to measurements termed LPC in previous reports. In many different paradigms, late positive ERPs with parietal maxima have been identified as P3 waves. Although ERPs analogous to these P3 waves may be measured by LPC, this measurement may be sensitive to other processes as well, as discussed below.

significantly as a function of processing task (F (3, 45) = 8.8, $P < 0.001$). A very large LPC was observed in recordings from parietal electrodes in all tasks except the one requiring alphabetical judgments. This tendency for LPC in task A to be less positive than LPC in the other tasks was significant in all but the frontal recordings.

Processing task had a significant influence on the earlier measurement (250–400 msec) only at posterior electrodes. That is, the task by electrode interaction was significant (F (27, 405) = 2.9, $P <$ 0.001) and, at all 3 parietal electrodes, ERPs from task A were significantly less positive than those from the 2 semantic tasks. Also, accuracy in task performance differed with processing task (F (3, 45) = 3.7, $P < 0.018$), as there were significantly more errors for alphabetical decisions (19%) than for counting vowels (10%).

ERPs and subsequent memory performance

ERPs averaged over all tasks on the basis of subsequent recognition memory performance (Fig. 2A) revealed greater late positive activity following subsequently recognized words than subsequently unrecognized words (F (1, 15) = 8.6, $P <$ 0.01, as measured in the 400–800 msec latency range). In the 250–400 msec latency range, the amplitude difference between ERPs elicited by recognized words and ERPs elicited by unrecognized words was not significant (F (1, 15) = 0.6).

The ERP difference based on recognition memory performance (Dm) can be seen in the difference wave forms computed by subtracting ERPs to unrecognized words from ERPs to recognized words (Fig. 2B). The effect of recognition was clear at midline electrodes and lateral parietal electrodes. Dm was virtually absent at lateral frontal and lateral temporal electrodes, so data from these electrodes were excluded from further analyses. Dm, which will henceforth be measured over the 400–800 msec latency range across the remaining 6 electrodes, averaged 1.0 μV for recognition across all tasks.

The relationship between input ERPs and successful recall was basically the same as that for recognition (Fig. 3). Subsequently recalled words elicited approximately 1.2 μV more than did words that were not recalled (F (1, 15), = 7.1, $P < 0.017$).

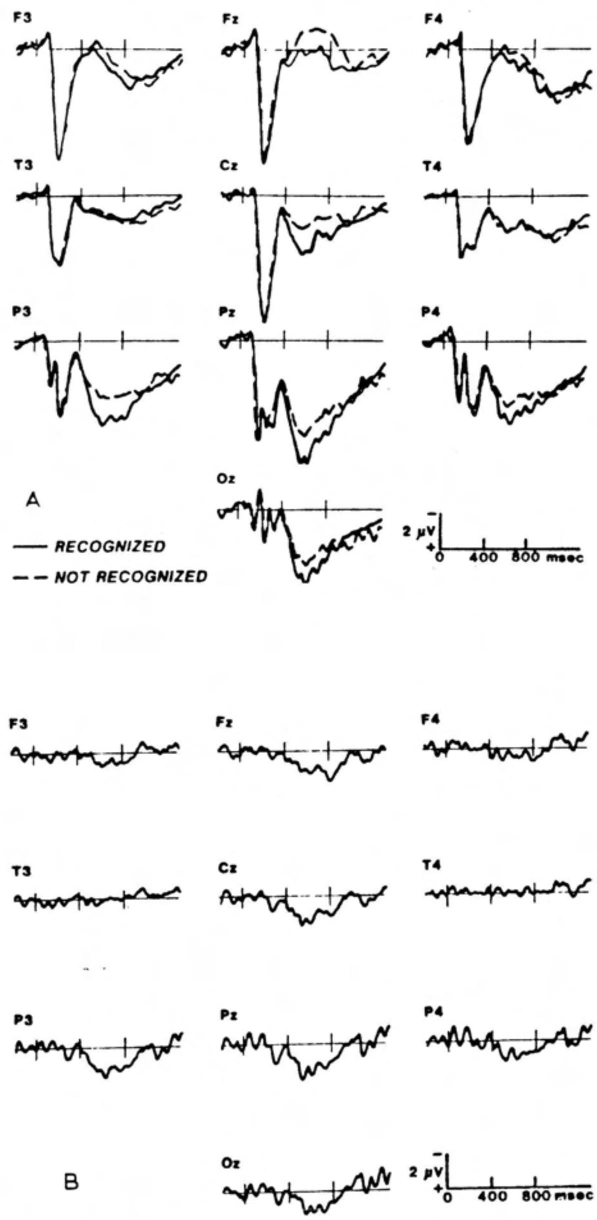

Fig. 2. A: ERPs averaged based on later recognition performance. Words subsequently recognized elicited greater late positive activity than did words subsequently not recognized. B: the ERP difference based on later recognition memory performance (or Dm, the difference between the two wave forms in A).

This Dm for recall was significantly greater at the left parietal electrode than at the right parietal electrode (2.1 μV versus 0.9 μV, F (1, 15) = 7.2,

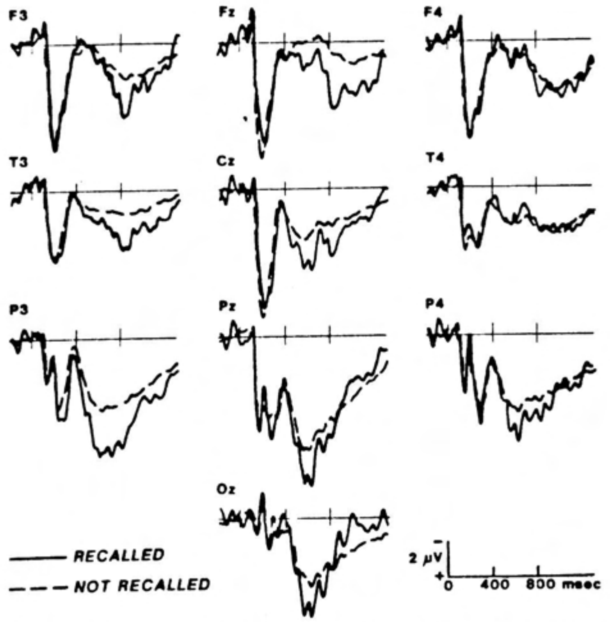

Fig. 3. ERPs averaged based on later free recall performance. Words subsequently recalled elicited greater late positive activity than did words subsequently not recalled.

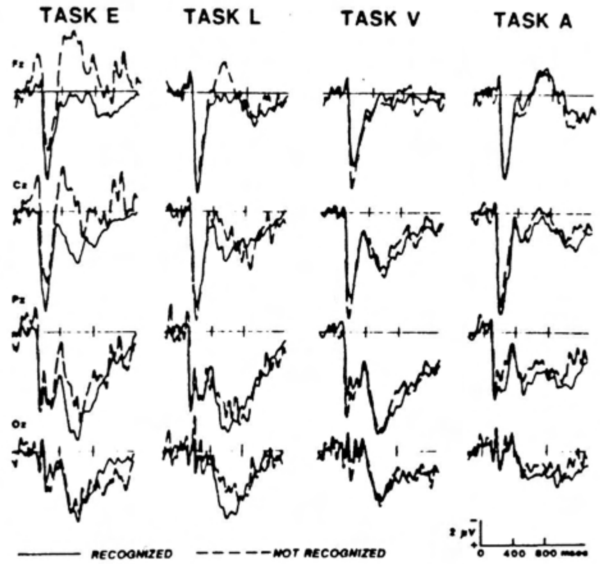

Fig. 4. ERPs associated with each processing task, averaged based on later recognition performance. The ERP difference between words later recognized and words not recognized was greater in the 2 semantic tasks (E and L).

$P < 0.016$). Because so few words were recalled, all further analyses focus on the recognition data.

ERPs elicited by recognized and unrecognized words differed depending on the processing task (Fig. 4). Dm was much greater in semantic tasks than in non-semantic tasks (Table I). However, the recognition × task interaction was statistically significant only at anterior electrodes (recognition × task × electrode interaction, $F\ (15,\ 225) = 1.9$, $P < 0.031$).

ERPs and recognition confidence

LPC to recognized words was significantly larger for words associated with high recognition confidence (3.6 μV) than for words associated with low recognition confidence (2.6 μV). Recognition confidence also influenced LPC to unrecognized words; when an input word was mistakenly identified as a new word, but with low confidence, LPC was larger (2.7 μV) than when subjects were more sure about their incorrect response (1.1 μV). This effect fell short of statistical significance in the

TABLE II

Processing task and decision interaction.

Processing task	Memory		LPC		Dm	
	+	−	+	−	+	−
E (edible)	90	83	4.4	3.0	3.7	1.1
L (living)	92	78	4.7	3.0	1.1	0.6
V (vowel)	67	63	3.4	3.9	0.8	0.3
A (alphabet)	55	59	1.7	1.1	−0.3	1.3

Note. '+' denotes an affirmative processing task decision; '−' denotes a negative processing task decision. Memory is percent correct recognition. LPC and Dm are mean amplitude measurements in μV, from lateral parietal and midline electrodes in the 400–800 msec latency range.

case of unrecognized words (F (1, 15) = 3.1) but was significant for recognized words (F (1, 15) = 6.5, $P < 0.022$). Recognition confidence effects were small and non-significant within individual processing tasks, especially non-semantic tasks. The overall recognition confidence effect may have stemmed in part from the relatively low confidence associated with words in non-semantic tasks. In semantic tasks, 77% of the recognized words were recognized with high confidence, whereas only 49% of the recognized words in non-semantic tasks were recognized with high confidence.

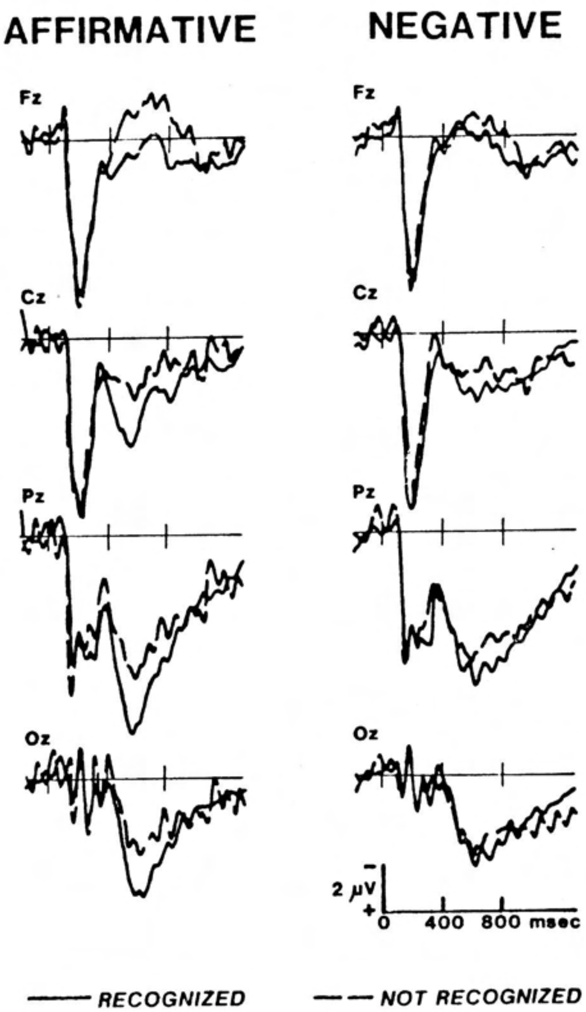

Fig. 5. ERPs associated with each type of decision, averaged based on later recognition performance. Greater late positive activity was elicited by words followed by affirmative responses than by words followed by negative responses. Further, Dm was greater in the former condition.

ERPs and processing task decision

In general, words given affirmative decisions elicited greater late positivity than did words given negative decisions (F (1, 15) = 9.3, $P < 0.008$). However, this difference between affirmative and negative decisions depended upon processing task (F (3, 45) = 4.5, $P < 0.008$), in that the LPC difference was greater in semantic tasks (Table II). A large ERP difference based on recognition performance was evident when only words associated with affirmative decisions were considered (Fig. 5), and there was a non-significant tendency for this Dm to be larger than Dm associated with negative decisions (F (1, 15) = 2.5).

ERPs and other variables

Reaction time. The influence of reaction time was assessed via a median split of the RT distributions for each task and subject. LPC associated with the faster RTs was significantly larger than LPC associated with the slower RTs (F (1, 15) = 8.8, $P < 0.009$). This effect was larger at posterior electrodes (F (5, 75) = 2.4, $P < 0.049$) but was not significantly influenced by processing task. Differences in the amplitude of Dm in the 2 RT conditions were non-significant, although Dm appeared more prolonged in the slow RT condition.

Word frequency. LPC to high-frequency words was significantly more positive than LPC to low-frequency words (F (1, 15) = 4.7, $P < 0.047$). At approximately 700 msec, however, the effect reversed. ERPs to low-frequency words were significantly more positive between 700 and 1000 msec after stimulus onset (F (1, 15) = 4.9, $P < 0.043$). Dm measured within the 400–800 msec epoch was uninfluenced by word frequency.

Typicality. Neither LPC nor Dm were significantly influenced by whether the decision was normal or atypical.

Decision accuracy. LPC to words followed by correct decisions was significantly larger than LPC to words followed by incorrect decisions (F (1, 15) = 15.5, $P < 0.001$). Dm, on the other hand, was not significantly influenced by whether the subject's decision at input was right or wrong.

Discussion

An incidental learning paradigm provided the means to manipulate degree of semantic analysis and other independent variables that influenced (or were correlated with) memory performance and/or ERPs. The fundamental finding was that ERPs recorded during encoding were sensitive to processes that correlated with subsequent memory performance. This type of result was first reported by Sanquist et al. (1980) using a recognition test. Donchin and colleagues found a similar ERP difference based on free recall performance, but only in subjects who did not use 'elaborative' encoding strategies, either by choice (Karis et al. 1984), by instructions (Fabiani et al. 1985), or by participation in an incidental learning paradigm (Fabiani et al. 1986). Neville et al. (1986) also reported an ERP difference based on subsequent recognition performance. In their study, words were either congruous or incongruous with a preceding phrase; Dm was significant in both conditions but began about 200 msec later for incongruous words. Kutas (1987) likewise examined Dm for congruous and incongruous words, using a cued-recall test that provided the sentence context in which the input words had been presented. Johnson et al. (1985) reported that ERPs to subsequently recognized words were slightly (but non-significantly) enhanced in amplitude and significantly shortened in latency relative to ERPs to unrecognized words, using an intentional learning paradigm. Overall, the finding of an ERP difference based on subsequent memory performance — Dm — has been demonstrated several times, although it remains an empirical question whether Dm measured the same type of processing in the different experimental paradigms.

Among the many encoding processes influencing memory, the present experiment focused on the possible role of semantic processing in the Dm phenomenon. Indeed, processing task influenced both memory and LPC. The non-semantic task requiring a decision on the alphabetical order of the initial and final letters of each word (task A) was associated both with poorer memory and with a smaller LPC relative to that in the 3 other tasks. Thus, Dm can be attributed in part to the dual effects of task A on memory and on LPC, in that the ERP to forgotten words included a greater proportion of words from task A (associated with smaller LPC measurements) than did the ERP to remembered words. In support of this suggested role of task A, it is interesting to note that although Dm was negligible in task A alone, the largest Dm for any 2 tasks combined was for the combination of task E with task A.

But why was task A associated with a smaller LPC than were the other tasks? One cause could be greater trial-to-trial variability in the processes underlying LPC in task A. This may be unlikely given that the standard errors of the mean reaction times were similar across tasks. Another hypothesis would be that task A prevented semantic processing because of its greater difficulty and demands compared to the other non-semantic task. By this account, counting the number of vowels in each word (task V) was sufficiently less demanding so as to allow some semantic processing in addition to the required orthographic processing. This is consistent with the findings that error rate was highest and reaction time was longest in task A relative to the other tasks, in addition to the anecdotal finding that several subjects expressed concern over the difficulty of making the alphabetical judgments.

Differential semantic processing between task A and the other 3 tasks, however, cannot account for the entire ERP difference based on recognition since Dm was present within tasks as well as across tasks. Task E, for example, required that the meaning of each word be analyzed. Dm in this case cannot be simply an ERP difference between words processed semantically and words not processed semantically. Furthermore, affirmative decisions yielded better memory and a greater LPC than did negative decisions. These dual effects of task decision might account for some of Dm, since remembered items included a greater proportion of affirmative decisions and, accordingly, greater positivity. Dm was still significant, however, when the ERP comparison for recognized and unrecognized words was limited to trials associated with affirmative decisions.

These results are consistent with the possibility that uncontrolled variation in semantic analysis or correlated processes could have made a partial contribution to ERP differences between remembered and forgotten words in previous reports. The experiment of Sanquist et al. (1980) also involved systematically varied degrees of semantic processing. A derived measure of ERP late positivity did not reliably differentiate between 3 conditions requiring either orthographic, phonemic, or semantic judgments in response to a pair of words. But both this study and the present study found the typical memory advantage for congruent words (i.e., the affirmative decision condition in the present study) in addition to a parallel enhancement of late positive ERP activity for congruent words. Whereas Dm was analyzed separately for congruent and incongruent words above, Sanquist et al. were unable to compute Dm in the absence of this confound. It has been suggested that words that fit with the preceding context are remembered better because of greater associative processing [3] (Craik and Tulving 1975). Thus, words with affirmative responses in the present experiment might have elicited a greater LPC and been remembered better because of associative processing contingent on their membership in the category of living or edible things. The fact that Dm was largest in task E provokes the speculation that the edible/non-edible distinction provided greater variability in associative processing compared to the living/non-living distinction. It is further possible that the same associative processing factor varies within the affirmative condition to give rise to Dm.

On the other hand, perhaps a crucial requirement for Dm is sufficient variability in memory strength between words. Negligible Dm waves might have been elicited in non-semantic tasks because memory strength was uniformly low. The lack of an effect of confidence on LPC in non-semantic tasks casts some doubt on this idea of a recognition floor effect, but this possibility cannot be eliminated. Further, it remains possible that associative processing was not the variable mediating Dm in semantic tasks and that a more non-specific process, such as arousal, was involved instead.

An arousal explanation for Dm would predict that some words were delivered when the subject was in a low state of arousal, and that these words were associated with poorer memory and a smaller LPC. Related explanations could substitute a construct of cognitive or attentional effort for arousal. If we consider reaction time as an index of effort, then more effort would be associated with slower reaction times, which might then correlate with poorer memory performance, but this was not the case. Nevertheless, the arousal explanation cannot be rejected on this basis because reaction time may be a poor index of effort (Eysenck and Eysenck 1979) [4]. Karis et al. (1984) attempted to dismiss the arousal explanation by virtue of their finding that subjects using elaborative encoding strategies did not show ERP differences based on subsequent recall tests. However, recall performance in subjects processing words using an elaborative strategy should be primarily determined by this highly effective processing, whereas rote strategies might leave recall performance more sensitive to arousal effects.

Understanding Dm would be easier if Dm solely reflected variations in a well-characterized ERP component [5]. We chose to combine all ERP components elicited in the 400–800 msec latency range into one measure termed LPC. This objectively defined complex probably includes a late positive component akin to measures studied by others and labeled P3 or P300. We believe that, without additional manipulations, a P3 component cannot be rigorously isolated from other ERP components occurring at the same time. Whereas principal component analysis has been used to sep-

[3] This *associative* processing must differ from the *elaborative* processing mentioned above (i.e., mnemonic strategies used in intentional learning paradigms by some subjects, such as linking adjacent words with an absurd storyline), in that only elaborative processing detracts from the Dm phenomenon.

[4] The dual-task procedure adapted from Eysenck and Eysenck (1979) has demonstrated that task L and task V are equivalent in their interference with the secondary task and so are perhaps equivalent in terms of effort (Kirkham, Meudell and Mayes, personal communication).

[5] We have adopted the distinction between ERP *components*, which each arise from a distinct neural process, and ERP *deflections*, which are local maxima and minima that can include multiple overlapping components.

arate overlapping components, this practice is controversial and can lead to erroneous conclusions (Wood and McCarthy 1984). Although P3 waves may contribute a great deal to the LPC measurement, it seems likely that other ERP components occurring within the same epoch are being measured as well.

The findings of the present experiment nevertheless suggested a dissociation between LPC and Dm (both defined as composite measures). First, the scalp distributions of Dm and LPC differed. Dm was large at all midline electrodes and was enhanced at left compared to right parietal electrodes. In contrast, LPC was highly localized to parietal electrodes and symmetric across left and right sides. The task manipulation also dissociated the two measures; LPC was small in task A and large in the 3 other tasks, whereas Dm was small in task A but just as small in task V. Several variables correlated with LPC but not with Dm (but such comparisons must be viewed with caution since non-significant effects on Dm might reflect greater variability in conditions with very few events). On the other hand, reaction time, word frequency, and decision accuracy all influenced LPC but not memory, whereas decision typicality influenced memory but not LPC.

Although some subset of the potentials constituting LPC necessarily gave rise to Dm, dissociations between measures cast doubt on the idea that the primary factors that governed LPC were identical to those that governed Dm. Rather, only part of LPC varied across words to give rise to Dm. This subset of LPC may have included ERP reflections of encoding processes engaged when a word's meaning is accessed and associated with a semantic category, but further evidence will be required to substantiate these speculations. In any event, the ERP appears to provide a moment-to-moment measure of processes that occur at the time of encoding and a means with which to study mechanisms of human memory.

References

Craik, F.I.M. Human memory. Ann. Rev. Psychol., 1979, 30: 63–102.

Craik, F.I.M. and Lockhart, R.S. Levels of processing: a framework for memory research. J. verb. Learn. verb. Behav., 1972, 11: 671–684.

Craik, F.I.M. and Tulving, E. Depth of processing and the retention of words in episodic memory. J. exp. Psychol., 1975, 104: 268–294.

Donchin, E. Surprise! ... surprise?. Psychophysiology, 1981, 18: 493–513.

Eysenck, M.W. and Eysenck, M.C. Processing depth, elaboration of encoding, memory stores, and extended processing capacity. J. exp. Psychol.: Hum. Learn. Mem., 1979, 5: 472–484.

Fabiani, M., Karis, D. and Donchin, E. Effects of strategy manipulation on P300 amplitude in a Von Restorff paradigm. Psychophysiology, 1985, 22: 588.

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

Fabiani, M., Gratton, G., Karis, D. and Donchin, E. P300: methodological and theoretical issues. In: P.K. Ackles, J.R. Jennings and M.G.H. Coles (Eds.), Advances in Psychophysiology, Vol. 2. JAI Press, Greenwich, CT, 1987: in press.

Hillyard, S.A. and Kutas, M. Electrophysiology of cognitive processing. Ann. Rev. Psychol., 1983, 34: 33–61.

Jasper, H.H. Report to the committee on methods of clinical examination in electroencephalography. Appendix: the ten-twenty system of the International Federation. Electroenceph. clin. Neurophysiol., 1958, 10: 371–375.

Johnson, R., Pfefferbaum, A. and Kopell, B. P300 and long-term memory: latency predicts recognition performance. Psychophysiology, 1985, 22: 497–507.

Karis, D., Fabiani, M. and Donchin, E. 'P300' and memory: individual differences in the Von Restorff effect. Cogn. Psychol., 1984, 16: 177–216.

Keppel, G. Design and Analysis: a Researcher's Handbook (2nd edn). Prentice-Hall, Englewood Cliffs, NJ, 1982.

Kutas, M. An event-related potential analysis of priming, congruity, and recall. 1987, submitted for publication.

Lockhart, R.S., Craik, F.I.M. and Jacoby, L.L. Depth of processing, recognition and recall; some aspects of a general memory system. In: J. Brown (Ed.), Recall and Recognition. Wiley, London, 1976.

Neville, H.J., Kutas, M., Chesney, G. and Schmidt, A.L. Event-related brain potentials during initial encoding and recognition memory of congruous and incongruous words. Mem. Lang., 1986, 25: 75–92.

Paller, K.A. Effects of Medial Temporal Lobectomy in Monkeys on Brain Potentials Related to Memory. Doctoral Dissertation, University of California, San Diego, CA, 1986.

Paller, K.A., Kutas, M. and Mayes, A. An investigation of neural correlates of memory encoding in humans. Psychophysiology, 1985, 22: 607.

Pritchard, W.S. Psychophysiology of P300: a review. Psychol. Bull., 1981, 89: 506–540.

Sanquist, T.F., Rohrbaugh, J.W., Syndulko, K. and Lindsley, D.B. Electrocortical signs of levels of processing: perceptual analysis and recognition memory. Psychophysiology, 1980, 17: 568–576.

Tulving, E. Précis of elements of episodic memory. Behav. Brain Sci., 1983, 7: 223–268.

Wood, C.C. and McCarthy, G. Principal component analysis of event-related potentials: simulation studies demonstrate misallocation of variance across components. Electroenceph. clin. Neurophysiol., 1984, 59: 249–260.