Brain (1996), **119**, 889–905

An event-related potential study of recognition memory with and without retrieval of source

Edward L. Wilding and Michael D. Rugg

Wellcome Brain Research Group, School of Psychology, University of St Andrews, Scotland

Correspondence to: M. D. Rugg, School of Psychology, University of St Andrews, St Andrews, Fife KY16 9JU, Scotland

Summary

Event-related potentials (ERPs) were recorded during the test phase of a recognition memory task in two experiments. In both experiments subjects made initial old/new judgements to visually presented words, and for words judged old, indicated in which of two voices (male/female) the words had been heard at study. In the second experiment only, subjects had the option to signal that they were uncertain about the status of a test word. Two positive-going ERP effects differentiated the ERPs evoked by correctly recognized old words from those evoked by words correctly judged new. The two effects differed in their scalp topography and time course, and were both of greater magnitude in the ERPs evoked by recognized words for which a correct voice judgement was made. The findings are consistent with the view that multiple neural systems underlie the ability to recognize an item and to recall its study context. However, the findings offer little support for the view, articulated in certain 'dual-process' models of recognition memory, that recognition judgements with and without retrieval of study context depend upon qualitatively different memory processes or systems.

Keywords: event-related potential; recognition memory; memory for source; recollection; familiarity

Abbreviations: EOG = electro-oculogram; ERP = event-related potential

Introduction

The proposal that a judgement of prior occurrence can be based on two different kinds of information is central to several accounts of recognition memory (e.g. Atkinson and Juola, 1973; Mandler, 1980; Jacoby and Dallas, 1981; Humphreys *et al.*, 1989; Gardiner and Java, 1993). Common to all of these accounts is the idea that an item in a recognition memory test triggers an attempt to retrieve a memory of a specific episode involving that item, and also engenders a general sense of familiarity which, if sufficiently strong, will lead to an old judgement irrespective of whether episodic retrieval is successful. A crucial difference between these two bases of recognition memory is that only episodic retrieval yields information about the learning context; recognition based on familiarity alone is acontextual. Following Jacoby and colleagues (e.g. Jacoby and Kelley, 1992), these two bases of recognition memory will be referred to as 'recollection' and 'familiarity'.

According to Jacoby and colleagues (Jacoby and Dallas, 1981; Jacoby, 1983; Whittlesea, 1993), recognition judgements based on familiarity are made when an item is processed relatively 'fluently', and the processes supporting fluency-based recognition are held to be related to those underlying priming and other expressions of implicit memory. Evidence for this view comes from studies demonstrating that a variety of priming manipulations can, in appropriate circumstances, increase the probability that an item will be judged old on a test of recognition memory (e.g. Jacoby and Dallas, 1981; Rajaram, 1993; Whittlesea, 1993). However, the extent of the role typically played by fluency in recognition memory is unclear. First, there is little evidence that the intact priming exhibited by amnesic patients contributes to their performance on tests of recognition memory (Haist *et al.*, 1992; Knowlton and Squire, 1995; but *see* Hirst *et al.*, 1986, 1988), contrary to what would be expected if fluent processing can influence recognition memory independently of recollection. Secondly, it has been argued that fluency may make a significant contribution to recognition memory in normal subjects only when recollection is weak (Johnston *et al.*, 1991).

An alternative view is that familiarity-based recognition relies largely on 'declarative' memory, and hence on the same medial temporal and diencephalic brain structures that are necessary for other expressions of explicit memory, including recognition based upon recollection. By this argument (Moscovitch, 1992, 1994; Squire, 1994), recollection and familiarity both depend upon the successful retrieval from declarative memory of information about an item's prior occurrence. For recollection to succeed, the retrieved information must be integrated with additional information about the context in which the item was encountered, a process critically dependent upon the prefrontal cortex. Familiarity-based recognition occurs when retrieval is successful but integration fails; in this case, an item can be judged old, but information about where and when it was encountered is unavailable. This proposal accounts well for the apparent failure of amnesic patients to show preserved familiarity-based recognition memory, since it assumes that recollection and familiarity are both dependent on the memory system that is impaired in these patients. The proposal is also consistent with evidence of disproportionately poor memory for context (source memory) in patients with prefrontal damage (Schacter *et al.*, 1984; Shimamura and Squire, 1987; Janowsky *et al.*, 1989).

As this brief review indicates, the question of the functional and neurological independence of the two bases of recognition memory is unsettled. In particular, it is unclear whether familiarity-based recognition, i.e. recognition in the absence of contextual retrieval, reflects the operation of memory processes that are separate from those underlying recollection, or whether instead it depends on a subset of those same processes. One means of deciding between these opposing views is to determine whether neural activity associated with recognition judgements based on familiarity and recollection dissociate in a manner more compatible with one or other account. This is the goal of the two studies presented here, in which neural activity is monitored with ERPs.

A number of groups have recorded ERPs during the test phase of recognition memory tasks. Almost invariably, the results of these studies have shown that ERPs elicited by words correctly judged old are more positive than ERPs to words correctly judged new (for reviews, *see* Rugg, 1994; Johnson, 1995). The difference between these two classes of ERP, i.e. the ERP old/new effect, onsets some 300–400 ms post-stimulus, has a duration of ~500 ms, and is larger over the left hemisphere, most markedly at parietal sites (Neville *et al.*, 1986; Rugg and Doyle, 1992). Conflicting claims have been made for the functional significance of this effect, some proposing that it is an ERP reflection of recollection (Smith and Halgren, 1989; Van Petten *et al.*, 1991; Paller and Kutas, 1992; Smith, 1993; Rugg *et al.*, 1995), and others that it indexes familiarity (Johnson *et al.*, 1985; Friedman, 1990; Potter *et al.*, 1992; Rugg and Doyle, 1992). However, there is only one published study (Wilding *et al.*, 1995) in which investigators have attempted to compare ERPs according to whether recognition was or was not associated with retrieval of study context, a distinction central to recent attempts to dissociate the two bases of recognition memory using behavioural measures (e.g. Jacoby and Kelley, 1992; Yonelinas, 1994).

Wilding *et al.* (1995) presented half of their study words in the visual modality and the remainder auditorally. In the subsequent test task, subjects made old/new judgements to a mixed list of old and new words, presented visually in Experiment 1 and auditorally in Experiment 2. For each word judged old, subjects were required to make a further judgement about the modality in which the word had been studied. Wilding *et al.* (1995) argued that recognized words which were correctly assigned to their study modality were more likely to have been recollected than were words which attracted an incorrect modality judgement. The question of whether ERPs dissociate recognition memory decisions based on recollection and familiarity was therefore addressed by comparing the ERP old/new effects elicited by recognized words that attracted either correct or incorrect modality judgements.

In both experiments, Wilding *et al.* (1995) found that the ERPs elicited by words attracting correct modality judgements were more positive than those for words correctly judged new. In the first experiment, no such effect was found for ERPs evoked by words associated with incorrect modality judgements. In the second experiment, an effect was found for these items, but over a more restricted latency range than that seen for the ERPs evoked by the putatively recollected items.

Wilding *et al.* (1995) argued that their findings did not support the idea that recognition based on recollection and familiarity were neurologically dissociable, since there was no evidence for qualitative differences between the ERPs evoked by recollected and unrecollected items. Rather, the data were consistent with the view that recognized words that attracted an incorrect context judgement had engendered weak or partial recollection: sufficient to judge a word old but insufficient to recover its study context (Johnson *et al.*, 1993).

These conclusions are, however, tempered by the fact that the procedure adopted in the experiments necessitated that 50% of test words were presented in the same modality as at study, whereas the remainder were presented in the alternative modality. Because within-modality repetition engenders greater priming, and thus greater fluency, than does repetition across-modality (Richardson-Klavehn and Bjork, 1988), the relative fluency with which the test items were processed could have served as a basis for modality judgements in the absence of recollection (Kelley *et al.*, 1989). To the extent that this was so, the 'recollected' words in the study of Wilding *et al.* (1995) would have included an unknown proportion that had attracted correct modality judgements on the basis of fluency alone.

The two experiments reported here employed the same general logic as those of Wilding *et al.* (1995), but adopted a procedure that ensured that veridical context judgements could be made only on the basis of recollection. At study, subjects heard a mixed list of words, half spoken in a male voice and half spoken in a female voice. In a subsequent visual recognition memory test, subjects first made an old/new judgement (Decision 1) and, for each word judged old, a second judgement about the voice in which it was heard at study (Decision 2). By comparing the ERPs evoked by recognized words that were correctly or incorrectly assigned to their study context, the question of whether there are distinct ERP correlates of recognition memory with and without contextual retrieval, can be addressed.

Experiment 1

Method

Subjects

A total of 18 subjects participated in the experiment. The data from two subjects were discarded because of excessive electro-oculographic artifact. Of the remaining 16 subjects, seven were female. All 16 subjects were right-handed, as defined by writing hand. Each subject gave informed consent prior to participation in the study, which was given approval by the local Ethical Committee.

Experimental material

Stimuli consisted of 360 low frequency words (frequency range, 1–7 per million), and 90 pronounceable non-words. The words were selected from the Kučera and Francis corpus (1967). All were open-class and ranged between four and nine letters in length. They were divided into four lists, each containing 90 words. [NB A narrow range of low frequency words were employed to maximize the probability that test words would be correctly recognized (accuracy of recognition memory is negatively correlated with word frequency; for review, *see* Rugg *et al.*, 1995), and to minimize variability across items in recognition performance.]

Four study lists were produced. Each study list was formed by combining two of the four word lists and the non-word list. The word lists were rotated across study lists so that each word list (and therefore each word) appeared on two different study lists. The study lists were divided into three blocks of 90 items, with each block consisting of 60 words and 30 non-words. The order of presentation of items within each block was randomly determined. Each block began with one filler item, and thus each study list consisted of a total of 273 items.

Within each study block half of the items were spoken by a female voice, and half were spoken by a male voice. The voice in which each item was presented was rotated over study lists to control for item effects. Each non-word was spoken by the male voice on two of the study lists and the female voice on the other two lists.

Test lists were formed by combining all four of the initial word lists. The resulting 360 words in each test list consisted of 180 words that had been heard at study, and 180 items that were presented at test for the first time. A different test list was formed to go with each of the four study lists. Of the 180 old words in each test list, an equal number had been spoken in the male or female voice at study.

Each test list was divided into six blocks of 60 items, each block consisting of 30 new words and 30 old words. Of the old words an equal number had been spoken in the male or female voice at study. Two different random sequences were applied to the words within each block, yielding eight test lists. A filler item began each block of 60 test items, and thus each test list consisted of a total of 366 items.

Visual stimuli were presented in central vision on a TV monitor screen. Stimuli were exposed for 300 ms in white letters against a black background. The stimuli subtended a maximum horizontal visual angle of 1.5°, and a vertical angle of 0.4°. Auditory stimuli were presented binaurally at a comfortable hearing level. The auditory stimuli were digitized at 16 kHz with 8-bit resolution, and stored on the hard disk of a Commodore Amiga B2000 computer. They were edited so that the beginning of the stored sound segment corresponded to the onset of the spoken word. The mean duration of these stimuli was 650 ms, and did not differ according to the gender of the voice.

Procedure

Each subject was exposed to one of the four study lists. After a 5-min delay, subjects performed the test phase. At test, subjects were exposed to the test list appropriate for the list they had encountered at study.

The study phase consisted of a modified lexical decision task. Following electrode placement (*see* below), subjects were seated in front of the stimulus presentation monitor with the index finger and the middle finger of each hand resting on microswitches. They wore a set of headphones through which the auditory stimuli were presented. A fixation point (an asterisk) started each trial, and was removed from the screen 100 ms prior to stimulus presentation. Subjects were instructed to respond to each item by pressing one of the four keys in front of them, depending upon whether the item was a non-word or a word, and whether it had been spoken by the male or the female voice. For each subject, the lexical decisions to items spoken in one of the two voices were always made with the two response keys on the same hand, whilst the lexical decisions to words spoken by the other voice were made with the alternate hand. The hands used for lexical decisions and voice judgements were counterbalanced across subjects. Subjects were informed that accuracy and speed were of equal importance, and the fact that a recognition memory test would follow the study task was not mentioned. Subjects were asked to relax, to remain still during the task, and to minimize eye movements and eye blinks, with the exception of when the fixation point was present on the screen. A practice session consisting of 12 items preceded the study phase proper. The total interstimulus interval was 3.21 s. Responses quicker than 400 ms, or slower than 1900 ms, were treated as errors.

At test, subjects judged whether a word was old or new, and in the case of words judged old, denoted via a second key press, the voice in which the word had been spoken at study. All 366 test words were presented visually. An asterisk preceded presentation of each word, and was removed 100 ms prior to stimulus onset. Subjects made an initial old/new judgement for each word as quickly and as accurately as possible. One second after this first response a row of four question marks appeared on the screen for a duration of 2 s. For words judged old, the question marks served as the cue for the subject to report on the voice in which the word had been heard at study. All test judgements were made on the two keys on which the index fingers of the subjects rested. The hands required for the first and second decision were counterbalanced across subjects such that there was no correlation between the old/new and male/female judgements. The hand used for the male/female voice judgement at test corresponded to the same hand used for the voice judgement at study. As in the study phase, subjects were asked to restrict their eye blinks to the period when the fixation point was on the screen. Initial old/new judgements faster than 400 ms, or slower than 2100 ms, were treated as errors.

Event-related potential recording

Scalp EEG was recorded from 13 tin electrodes embedded in an elasticated cap. Recording locations were based on the International 10–20 system (Jasper, 1958). The montage comprised three midline sites (Fz, Cz, Pz), and left and right frontal (75% of the distance from Fz to F7/F8), anterior temporal (75% of the distance from Cz to T3/T4), parietal (75% of the distance from Pz to T5/T6), posterior temporal (T5, T6) and occipital sites (O1, O2). All EEG electrodes were referred to a linked pair of electrodes situated on each mastoid process. The electro-oculogram (EOG) was recorded bipolarly from electrodes placed on the outer canthus of the left eye, and above the supraorbital ridge of the right eye. All channels were amplified with a bandpass of 30–0.03 Hz (3 dB points). On-line sampling was at 6 ms per point for a duration of 1536 ms, commencing 102 ms prior to stimulus presentation.

The EOG was averaged separately for each response category to assess the influence of electro-ocular activity on the EEG. Trials on which EOG activity exceeded a pre-set criterion were rejected prior to averaging, as were trials on which baseline drift (difference between first and last data point) exceeded 80 μV at any scalp site. In order to maintain an acceptable signal/noise ratio a lower limit of 16 artifact free trials per subject per response category was set.

Results

Behavioural data

Study phase. Ninety percent of the words, and 84% of the non-words were correctly identified at study, with no difference in accuracy for items spoken in the male or the female voice. Mean reaction times for correct word and non-word decisions were 1020 ms and 1165 ms, respectively, again with no differences between items spoken in the two voices.

Table 1 *Probabilities of correct old/new judgements and correct source judgements in Experiment 1*

	Female voice	Male voice	New
Accuracy			
Old/new judgement			
P (correct)	0.71	0.70	0.72
Source judgement			
P (correct)	0.64	0.65	
Reaction times (ms)			
Old/new judgement			
Correct	1103	1139	1201
SD	332	356	373
Incorrect	1268	1263	1261
SD	426	399	386
Source judgement			
Correct	1079	1140	
SD	316	357	
Incorrect	1145	1129	
SD	336	341	

Old words are separated according to study voice. Also displayed are the mean reaction times and standard deviations (milliseconds) for correct and incorrect judgements on Decision 1, and the reaction times for old words separated according to the accuracy of the subsequent source judgement.

Test phase. Table 1 displays the probability of a correct response on Decision 1 of the test task for old and new test words, separated by study voice. For words spoken in either voice, the discrimination measure '$P_{\text{hit}} - P_{\text{false alarm}}$' (Snodgrass and Corwin, 1988) was above chance [male $t(15) = 7.59$, $P < 0.001$; female $t(15) = 12.31$, $P < 0.001$]. The two measures did not differ from one another.

The ANOVA of the reaction times for old/new judgements to test words (Table 1) employed the factors of response accuracy (correct versus incorrect) and word type (male versus female versus new). (NB For all behavioural analyses, the standard deviations of reaction time distributions were also compared, using the same factors as those employed for the mean reaction times. No differences between standard deviations were found in either experiment.) Note that in this and in all subsequent reports of analyses of variance, the analyses incorporate the Geisser–Greenhouse correction for inhomogeneity of covariance (Keselman and Rogan, 1980), and that F ratios are reported with corrected degrees of freedom. The analysis revealed a main effect of both factors [accuracy, $F(1,15) = 31.76$, $P < 0.001$; word type, $F(1.9,29.1) = 3.46$, $P < 0.05$]. The main effect of accuracy reflected the fact that correct responses were faster than incorrect responses. *Post hoc* analyses (Newman–Keuls) revealed no reliable differences between the means for old female, old male and new words. The mean reaction times to these response categories were 1186, 1201 and 1231 ms, respectively.

Table 1 also displays the probability of a correct judgement on Decision 2 of the test task for words correctly judged old on Decision 1. The probability of a correct voice judgement was 0.65, a value significantly above the chance probability of 0.5 [$t(15) = 5.56$, $P < 0.01$]. The probability of a male voice judgement to a new word incorrectly judged old was 0.49. This value did not differ significantly from 0.50, indicating the absence of a response bias in favour of one gender.

Reaction times for correct old judgements on Decision 1 are shown in Table 1, presented according to the accuracy of the subsequent voice judgement. The analysis of these reaction times [factors of response accuracy (correct versus incorrect), and word type (male versus female)] revealed no significant effects. Reaction times for the second (source) decision were not analysed, as subjects were required to delay this decision until the presentation of the response cue.

Event-related potential analyses

In accord with our previous study (Wilding *et al.*, 1995), trials on which words were correctly judged old and correctly assigned to their study context will be referred to as belonging to the 'hit/hit' response category. Trials on which words were correctly judged old and incorrectly assigned to study context will be referred to as belonging to the 'hit/miss' response category. Trials on which words were correctly judged new will be referred to as 'correct rejections'.

Preliminary analyses comparing the hit/hit ERPs separated according to study voice revealed no reliable differences. On the basis of these findings, together with the absence of any behavioural differences between items spoken in the two voices, the hit/hit and hit/miss ERPs were collapsed across study voice. The collapsed ERPs were formed by computing a weighted average of the ERPs to words spoken in the male or the female voice at study.

Grand averages of the ERPs evoked by items in the collapsed hit/hit and hit/miss categories are illustrated in Fig. 1, along with the ERPs to correct rejections. The mean numbers of trials entering into each subject's waveforms are 60, 35 and 98, for the hit/hit, hit/miss and correct rejection categories, respectively. As can be seen from the figure, from ~400 ms post-stimulus, the ERPs to hit/hit items are more positive-going than those to correct rejections. This difference is larger over the left than the right hemisphere at the parietal sites, but shows an opposite asymmetry, and a more extended time course, at the frontal electrodes. The ERPs elicited by hit/miss items are also more positive than those to correct rejections at parietal electrodes, but there is much less separation between these conditions at frontal sites. Accordingly, the most prominent differences between the ERPs elicited by hit/hit and hit/miss items are over the frontal electrodes.

The ERPs were quantified by measuring, with respect to the mean of the pre-stimulus baseline, the mean amplitudes of three consecutive latency regions, 500–800, 800–1100 and 1100–1400 ms post-stimulus. For each region, the data from the lateral electrodes were subjected to ANOVA, employing the factors of response category (hit/hit, hit/miss, correct rejection), hemisphere, and electrode site. (The outcomes of analogous ANOVAs performed on the data from the midline electrode sites are not described unless they conflict with or add to the conclusions arising from the analyses of the lateral data.) When one or more of the effects involving the factor of category was significant, subsidiary ANOVAs, contrasting the categories on a pairwise basis, were performed to elucidate these effects. As they are of no relevance to the aims of the experiment, effects solely involving the factors of electrode site or hemisphere are not reported.

In addition to the ANOVAs described above, ANOVAs on data from all electrode sites were conducted to search for topographical differences between different experimental effects. These analyses were performed directly upon the old/new effects (hit/hit minus correct rejection; hit/miss minus correct rejection) computed from the mean amplitudes of the latency regions described above, and rescaled to remove global differences in amplitude (McCarthy and Wood, 1985).

Analyses of mean amplitudes

The data from each lateral electrode site are shown in Table 2 for the three latency regions. The ANOVA of the 500–800 ms region gave rise to a main effect of response category [$F(1.2,18.7) = 7.32$, $P < 0.01$], and to an interaction between category and electrode site [$F(2.9,43.4) = 6.65$, $P < 0.001$]. A subsidiary ANOVA contrasting the hit/hit and correct rejection categories gave rise to the same two effects [category: $F(1,15) = 11.53$, $P < 0.005$; category×site: $F(1.5,22.5) = 8.15$, $P < 0.005$], along with a category× hemisphere×site interaction that approached significance [$F(2.2,32.6) = 3.15$, $P = 0.053$). These category and category×site effects arose because the differences between the hit/hit and correct rejection ERPs are greater at frontal, anterior temporal and parietal electrodes than at the posterior temporal and occipital sites. In light of the marginally significant three-way interaction, a further ANOVA was conducted on the parietal sites alone, the sites at which hemisphere asymmetries in the hit/hit old/new effects appear to be at their strongest. This revealed a significant response category×hemisphere interaction [$F(1,15) = 8.27$, $P < 0.025$], indicating that the asymmetry in the size of these old/new effects (2.60 μV and 1.23 μV for left versus right parietal sites, respectively) is reliable.

An ANOVA contrasting the hit/miss and correct rejection waveforms gave rise to a significant effect of response category [$F(1,15) = 21.69$, $P < 0.001$], indicating that the hit/miss waveforms are reliably more positive-going. A final ANOVA compared the hit/hit and hit/miss ERPs, and revealed a significant interaction between category and site

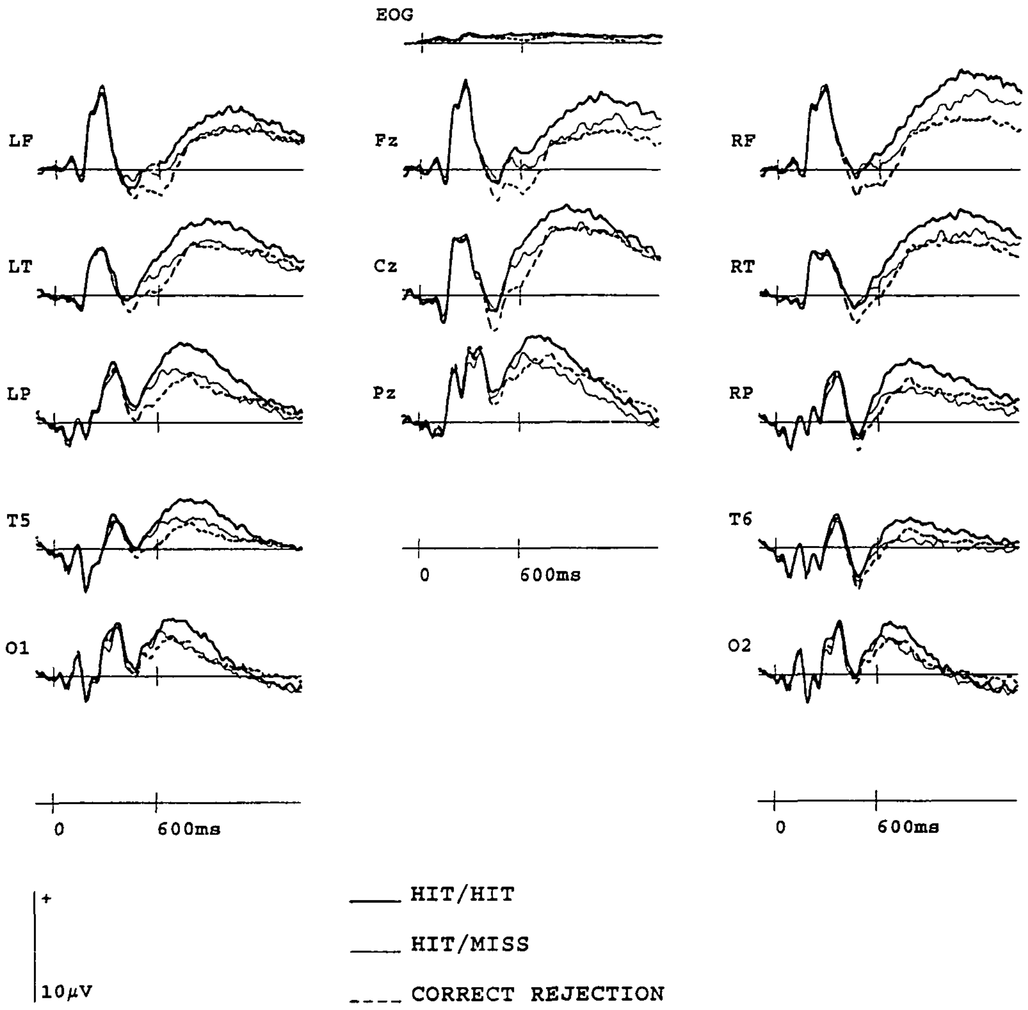

Fig. 1 Grand average ERPs associated with the hit/hit, hit/miss, and correct rejection response categories in Experiment 1. Fz, Cz and Pz signify midline frontal, central and parietal sites. LF, RF, LT, RT, LP, RP, T5, T6, O1 and O2 signify left and right frontal, anterior temporal, parietal, posterior temporal and occipital sites. The EOG activity for the three response categories is attenuated by a factor of 4 compared with the activity shown for all other scalp sites.

[$F(1.6,24.0) = 9.70$, $P < 0.001$), reflecting the frontal distribution of the differences between these two classes of waveform (1.99 µV and 0.50 µV at frontal and parietal locations, respectively).

The overall ANOVA of the 800–1100 ms region revealed a reliable category by site interaction [$F(3.4,50.8) = 3.92$, $P < 0.025$]. A subsidiary ANOVA comparing the ERPs from the hit/hit and correct rejection categories gave rise to a main effect of category [$F(1,15) = 6.19$, $P < 0.025$], and to an interaction between category and site [$F(1.6,23.9) = 4.31$, $P < 0.05$]. These effects reflect the strongly frontal distribution of the differences between these two classes of ERP. By contrast, the ANOVA comparing the hit/miss and correct rejection categories gave rise to no reliable effects. The comparison between the hit/hit and hit/miss ERPs revealed a single effect, for the interaction between category and site [$F(2.0,29.8) = 5.65$, $P < 0.01$). This effect reflects the frontal distribution of the differences between these categories (1.94 µV and 0.44 µV at frontal and parietal locations, respectively).

The overall ANOVA of the last latency region to be analysed, 1100–1400 ms, once again gave rise to a significant interaction between category and site [$F(2.9,43.2) = 6.25$, $P < 0.001$), accompanied by a significant three-way interaction between these two factors and hemisphere [$F(3.5,53.2) = 2.82$, $P < 0.05$]. The subsidiary ANOVA of the data from the hit/hit and correct rejection categories gave rise to the same two effects [category×site: $F(1.4,20.5) = 11.94$, $P < 0.001$; category×site×hemisphere: $F(2.0,31.0) = 5.46$, $P < 0.01$]. These effects reflect the frontal distribution of the differences between these waveforms, an effect that is markedly greater over the right hemisphere (1.08 µV and 0.37 µV for left frontal and parietal electrodes, respectively; 4.12 µV and –0.08 µV for the right frontal and parietal electrodes, respectively). In keeping with this pattern, an ANOVA of the data from the frontal sites

Table 2 *Mean amplitude (µV) of the 500–800, 800–1100,and 1100–1400 ms latency regions of ERPs evoked by the hit/hit, hit/miss and correct rejection response categories in Experiment 1*

	LF	LT	LP	T5	O1	RF	RT	RP	T6	O2
500–800 ms										
Hit/hit	3.80	6.07	7.61	3.99	2.93	4.06	3.32	4.70	2.16	1.77
Hit/miss	1.93	5.00	6.84	3.52	3.28	1.85	2.29	4.44	1.89	2.43
Correct rejection	1.18	3.60	4.95	2.45	2.16	1.08	1.55	3.44	1.35	1.56
800–1100 ms										
Hit/hit	5.20	6.82	7.07	4.26	2.62	7.19	6.52	5.88	3.49	1.75
Hit/miss	3.63	5.69	6.04	3.38	2.50	4.89	5.06	4.87	2.67	2.05
Correct rejection	3.85	5.46	4.98	2.83	2.51	5.54	5.54	4.87	2.50	1.97
1100–1400 ms										
Hit/hit	3.41	2.97	1.60	0.37	–0.76	8.21	5.45	2.50	2.01	–0.04
Hit/miss	2.45	2.76	2.06	0.74	–0.08	6.21	4.71	2.73	1.91	0.33
Correct rejection	2.30	2.43	1.22	0.17	0.23	4.97	4.07	2.60	1.68	0.39

LF, RF, LT, RT, LP, RP, T5, T6, O1, O2 signify left and right, frontal, anterior temporal, parietal, posterior temporal and occipital sites.

alone gave rise to a significant interaction between response category and hemisphere [$F(1,15) = 13.93$, $P < 0.001$], reflecting the fact that the old/new effects were larger on the right. The equivalent ANOVA for the data from the parietal electrodes gave rise to no significant effects.

As was the case for the 800–1100 ms latency region, ANOVAs revealed no differences between the ERPs belonging to the hit/miss and correct rejection categories. The ANOVA contrasting the hit/hit and hit/miss waveforms revealed a significant interaction between category and site [$F(1.9,27.9) = 7.69$, $P < 0.005$], reflecting the frontal distribution of the differences between these ERPs (1.45 µV and –0.34 µV at frontal and parietal locations, respectively).

Topographic analyses

Two topographical analyses were conducted. The first compared the distribution of the hit/hit and hit/miss old/new effects in the 500–800 ms latency region, the region in which the effects were reliable for both response categories. The interaction between category and electrode site was non-significant, indicating that the scalp distributions of these effects are equivalent. The second topographic analysis contrasted old/new effects as a function of time, comparing the distribution of the 500–800 ms latency region with that of the 1100–1400 ms region. As only the hit/hit category was associated with reliable effects in the latter region, the analysis was restricted to this category. The ANOVA revealed a reliable interaction between latency region and electrode site [$F(3.4,50.3) = 3.99$, $P < 0.01$], indicating that the scalp distribution of the hit/hit old/new effects differs in the two latency regions. A subsidiary ANOVA conducted on the frontal sites alone gave rise to a significant hemisphere×region interaction [$F(1,15) = 5.01$, $P < 0.05$], reflecting a larger asymmetry in favour of the right hemisphere in the 1100–1400 ms region. The equivalent ANOVA conducted on the data from the parietal electrodes also revealed a reliable interaction between hemisphere and region [$F(1,15) = 8.03$, $P < 0.025$]. In this case, the interaction reflected the weakening with time of the strong left>right asymmetry occurring in the 500–800 ms region.

Misses and false alarms

Figure 2 illustrates the ERPs evoked by old words incorrectly categorized as new (misses), and new words wrongly categorized as old (false alarms; these ERPs were obtained from only 13 subjects, three subjects making too few false alarm responses to permit the formation of reliable ERPs). The ANOVAs contrasting each of these classes of ERP with those evoked by correct rejections revealed no evidence of positive-going effects analogous to the old/new effects found for the ERPs evoked by words correctly judged old.

Discussion

As in our previous study of source memory (Wilding *et al.*, 1995), words correctly judged old evoked ERPs that were more positive-going than those elicited by correctly identified new items. Unlike in that study, however, this old/new effect could clearly be dissociated into two topographically and temporally distinct components: while one had a left parietal maximum and a time-course of a few hundred milliseconds, the other was maximal at the right frontal electrode and showed no sign of abating by the end of the recording epoch. The characteristics of the first of these components indicate that it corresponds to the late positive wave ('P600'; Rugg and Doyle, 1992) whose modulation underlies the old/new effects described in previous studies (e.g. Neville *et al.*, 1986; Rugg *et al.*, 1995). The second component has, to our knowledge, not been reported previously, although it may be related to the 'right frontal P300' described by Johnson (1995). As is discussed later, this component may reflect processes engaged by the requirement to retrieve and make use of information about study context.

Echoing the findings of Wilding *et al.* (1995), the ERPs

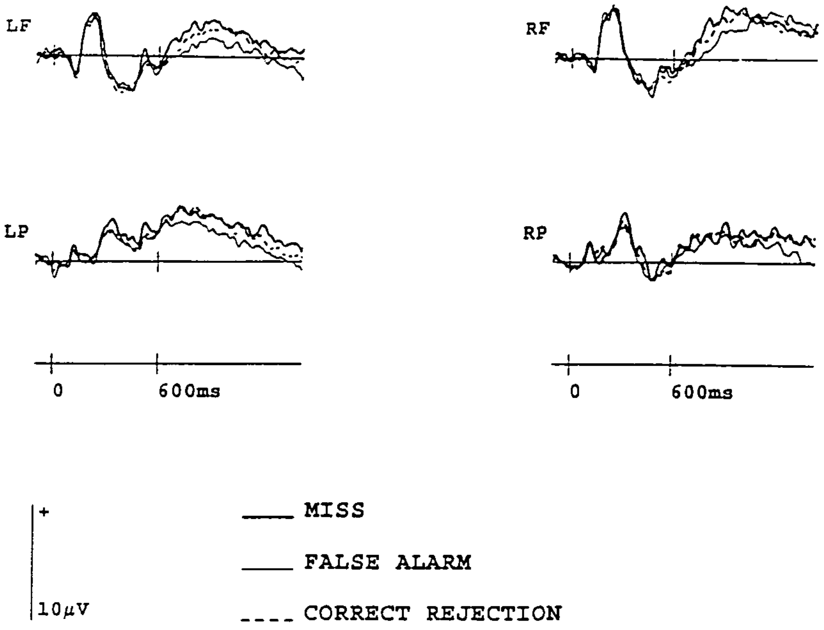

Fig. 2 Grand average ERPs to misses, false alarms, and correct rejections in Experiment 1. Data are displayed for the lateral sites LF, RF, LP and RP. The misses are collapsed across study voice, and the false alarms are collapsed across the Decision 2 voice judgement.

evoked by old words differed according to whether the words were associated with successful (hit/hit) or unsuccessful (hit/miss) retrieval of the study context. The difference took the form of greater positivity in the hit/hit ERPs, especially from around 800 ms post-stimulus. (NB Here and elsewhere terms such as 'more positivity' are intended to be descriptive only. There is no means of determining whether a positive-going shift in an ERP waveform represents increased ouput from a generator producing a scalp field with positive polarity relative to some reference site, as opposed to decreased activity in a generator whose scalp field has a negative polarity.) There was, however, little evidence to suggest that these old/new effects differed qualitatively, since the scalp distributions of the two effects in the 500–800 ms latency region, the only region in which both effects were reliable, did not differ significantly. Thus, as in the study of Wilding *et al.* (1995), the findings are consistent with the idea that ERP differences between 'recollected' and 'unrecollected' words are more quantitative than qualitative.

One rather mundane interpretation of these differences is that they reflect the differential influence of guessing on the ERPs from the hit/hit and hit/miss categories. This possibility arises because anything up to 28% of studied words (the equivalent of the false alarm rate) might have attracted a correct recognition judgement on the basis of a lucky guess. Presumably, these words would have failed to elicit ERP correlates of recognition memory (whether based on recollection or familiarity). Thus, averaging the ERPs elicited by these words together with those evoked by genuinely recognized words would attenuate the memory related effects carried by the latter class of ERP, the degree of attenuation depending on the ratio of the two trial types. This raises a problem in interpreting the differences between the hit/hit and hit/miss ERPs, because correct source judgements for study words that were correctly guessed old on Decision 1 must also have been the outcome of a lucky guess, whereas correct voice judgements on genuinely recognized words would comprise a mixture of guesses and veridical judgements. Thus, in comparison with the hit/hit response category, the hit/miss category must have contained a higher proportion of study words that were correctly guessed old. It follows that the differences between the ERPs evoked by hit/hit and hit/miss words may merely reflect the differing extents to which otherwise equivalent old/new effects were attenuated by trials associated with words that were guessed old on Decision 1. If this is so, attempts to account for these differences in terms of the different kinds of information that are retrieved on hit/hit and hit/miss trials are superfluous. Further discussion of the functional significance of these ERP differences is therefore deferred until this issue has been addressed.

Experiment 2

Introduction

Experiment 2 was designed primarily to address the question of whether the ERP differences found in Experiment 1 between the hit/hit and hit/miss response categories were caused by the differential influence of guessing on the initial old/new judgement. To this end, the binary forced choice procedure employed in Experiment 1 was modified to provide three response alternatives at test. The third alternative allowed subjects to signal that they were unable to decide about the status of a test word. Thus, a significant proportion of words that would otherwise have attracted a guess about their old/new status should receive instead the 'don't know' alternative. To the extent that these items are responsible for the differences observed between the hit/hit and hit/miss ERPs in Experiment 1, the differences should be reduced in the present experiment.

A don't know response alternative was also provided for the subsequent context judgement. The aim here was to obtain a cleaner separation between trials on which study voice was successfully recollected and those on which it was not, thereby reducing the proportion of trials contributing to the hit/hit ERPs which were lucky guesses on Decision 2. In Experiment 1, the mixing of these guesses with responses representing genuine retrieval of the study context would presumably have acted to reduce the magnitude of the differences between the ERPs evoked by words from the hit/hit and hit/miss categories.

Method

Subjects

A total of 19 subjects participated in the experiment. The data from two subjects were discarded due to a technical error. The data from a further subject was discarded because too few correct old judgements were made to permit formation of reliable averaged waveforms. Of the remaining 16 subjects, 12 were female. All subjects were right-handed, as defined by writing hand. Each subject gave informed consent prior to participation in the study.

Experimental material

The items differed from those used in Experiment 1, but were selected according to the same criteria. The procedures for selecting items and forming study and test lists were the same as in the first experiment.

Visual stimuli subtended a maximum horizontal visual angle of 2.0°, and a maximum vertical angle of 0.6°. Auditory stimuli were presented binaurally at a comfortable hearing level. They were digitized at 22 kHz with 16 bit resolution, and stored on the hard disk of an IBM-compatible PC. Mean duration for auditorally presented stimuli was 660 ms for words spoken by the male voice, and 630 ms for words spoken by the female voice.

Procedure

The only difference in the study task from the procedure adopted in Experiment 1 was that the inter-trial interval was lengthened to 4.1 s. The additional delay preceded presentation of the fixation asterix on each trial.

For the test phase, all aspects of the procedure were the same as in Experiment 1, with the exception that subjects had the option to make a don't know response for both judgements. As in Experiment 1, the old/new and voice judgements were made on the response keys on which the index fingers of the subjects rested. Subjects made a don't know response by pressing one of the response keys on which their middle fingers rested; the finger used for this response was counterbalanced across subjects. The hands required for the first and second judgement were also counterbalanced across subjects such that there was no correlation between the old/new and male/female judgements. As in the study phase, subjects were asked to restrict their eye blinks to the period when the fixation point was on the screen.

Event-related potential recording

With the following exceptions, EEG and EOG recording procedures and criteria were the same as in Experiment 1. First, EEG was recorded from a further four electrode sites. These additional sites were left and right prefrontal (FP1 and FP2), and the centroparietal sites, P3 and P4. Secondly, all channels were amplified with a bandpass of 35 to 0.03 Hz (3 dB points).

Results

Behavioural data

Study phase. Identification accuracy was almost identical to that in Experiment 1. In the only disparity between the two experiments, responses to items spoken in the female voice were significantly faster than responses to items spoken in the male voice [1124 versus 1165 ms, $F(1,15) = 19.91$, $P < 0.001$]. This difference in all likelihood reflects the fact that the average length of the female speech samples was 30 ms less than that of the male samples.

Test phase. Table 3 displays the probability of correct, incorrect and don't know judgements to old and new test words on Decision 1. Old words are separated according to study voice. A discrimination measure of '$P_{\text{hit}} - P_{\text{false alarm}} + P_{\text{(don't know/new')}}$ was computed for words spoken in either voice. This measure represents the lower bound on discrimination estimates commonly obtained for tests of recognition memory by the index 'phit – pfalse alarm'.

Discrimination was above chance for words spoken in either voice [male: $t(15) = 7.01$, $P < 0.001$; female: $t(15) = 8.10$, $P < 0.001$] and these two indices did not differ significantly. The ANOVA comparing the probabilities of incorrect responses to old female, old male and new words also revealed no significant differences. However, ANOVA comparing the probabilities of a don't know response to these three word types revealed a main effect [$F(1.3,19.5) = 6.56$, $P < 0.05$). *Post hoc* analyses (Newman–Keuls) revealed that, whilst the probabilities of don't know responses to old female and old male words were not reliably different, both

Table 3 *Probabilities of correct, incorrect and don't know old/new judgements and source judgements in Experiment 2*

	Female voice	Male voice	New
Accuracy			
Old/new judgement			
P (correct)	0.70	0.67	0.62
P (incorrect)	0.21	0.22	0.23
P (don't know)	0.09	0.11	0.15
Source judgement			
P (correct)	0.50	0.50	
P (incorrect)	0.27	0.24	0.23/0.26
P (don't know)	0.23	0.26	0.51
Reaction times (ms)			
Old/new judgement			
Correct	1206	1203	1382
SD	322	313	358
Incorrect	1477	1514	1425
SD	341	337	369
Don't know	1867	1810	1907
SD	229	305	276
Source judgement			
Correct	1079	1140	
SD	316	357	
Incorrect	1145	1129	
SD	336	341	
Don't know	1291	1348	
SD	298	314	

Old words are separated according to study voice. Also displayed are the mean reaction times and standard deviations (milliseconds) for correct, incorrect and don't know judgements on Decision 1, and the reaction times for old words separated according to the subsequent source judgement.

were significantly lower than the probability of a don't know response to a new word.

Table 3 shows the reaction times for correct, incorrect and don't know judgements to old and new words on Decision 1 of the test task. Given the low number of don't know responses, and the fact that two subjects made no don't know responses for the first decision, analysis of the reaction times was restricted to correct and incorrect judgements. The analysis employed the factors of accuracy (correct versus incorrect) and word type (new versus old male versus old female). Reaction times for correct judgements were reliably faster than reaction times for incorrect judgements [$F(1,15) = 32.29$, $P < 0.001$]. In addition, the analysis revealed an interaction between accuracy and word type [$F(1.4,20.3) = 8.35$, $P < 0.01$]. *Post hoc* analyses (Newman–Keuls) revealed that incorrect judgements were slower than correct judgements for old words, but that reaction times to new words did not differ as a function of response accuracy. Further, correct responses to old words were faster than correct responses to new words, but the reaction times for incorrect responses did not differ according to word type.

The probabilities of male, female and don't know judgements on Decision 2 of the test task are also shown in Table 3. For words spoken in either voice, the probability of a correct voice judgement was reliably higher than the probability of an incorrect judgement [male: $t(15) = 4.17$, $P < 0.001$; female: $t(15) = 5.77$, $P < 0.001$]. Comparison of the probabilities of male and female voice judgements to new words incorrectly judged old (false alarms) revealed no evidence for a voice response bias. The ANOVA comparing the probability of a don't know response to correctly identified old words and to false alarms revealed a main effect of category [$F(1.4,21.7) = 36.53$, $P < 0.001$]. *Post hoc* analyses (Newman–Keuls) revealed no difference between the probabilities of a don't know judgement to old words separated according to study voice, but both were significantly lower than the probability of making a don't know response to false alarms.

Reaction times to words correctly judged old on Decision 1 are displayed in Table 3, presented according to the response they attracted on Decision 2. The ANOVA of these reaction times employed the factors of response type (correct versus incorrect versus don't know) and word type (male versus female). The analysis revealed a main effect of response type [$F(1.8,27.2) = 10.82$, $P < 0.001$). *Post hoc* tests (Newman–Keuls) revealed that the reaction times for correct and incorrect judgements did not differ (1161 ms and 1214 ms, respectively), but that both were significantly faster than the reaction times for don't know responses (1320 ms).

Event-related potential analyses

Trials associated with don't know responses on Decision 1 of the test task were discarded. Hit/hit ERPs were formed as for Experiment 1. When the hit/miss response category was defined as in Experiment 1 (incorrect voice judgements to words correctly judged old), only 12 subjects contributed sufficient trials to permit the formation of reliable ERPs. A preliminary analysis was conducted on the data from these 12 subjects to ascertain whether ERPs formed by pooling trials associated with recognized items attracting incorrect and don't know voice judgements differed from those associated with incorrect voice judgements alone. The ANOVA comparing these two classes of ERPs revealed no differences over the duration of the recording epoch. On the basis of this finding, the ERPs for all 16 subjects were collapsed across the incorrect and don't know response categories, yielding waveforms for recognized words for which the subsequent source judgement was either incorrect or uncertain. This class of ERPs will be referred to below as the 'hit/miss' response category, although it should be noted that this category is not strictly equivalent to the hit/miss category as defined in Experiment 1. Finally, the ERPs associated with the hit/hit and hit/miss response categories were collapsed across study voice, since a preliminary comparison of the hit/hit ERPs separated according to study voice revealed no reliable differences.

The grand average ERPs evoked by items in the hit/hit, hit/miss and correct rejection response categories are shown in Fig. 3. The mean numbers of trials entering into each

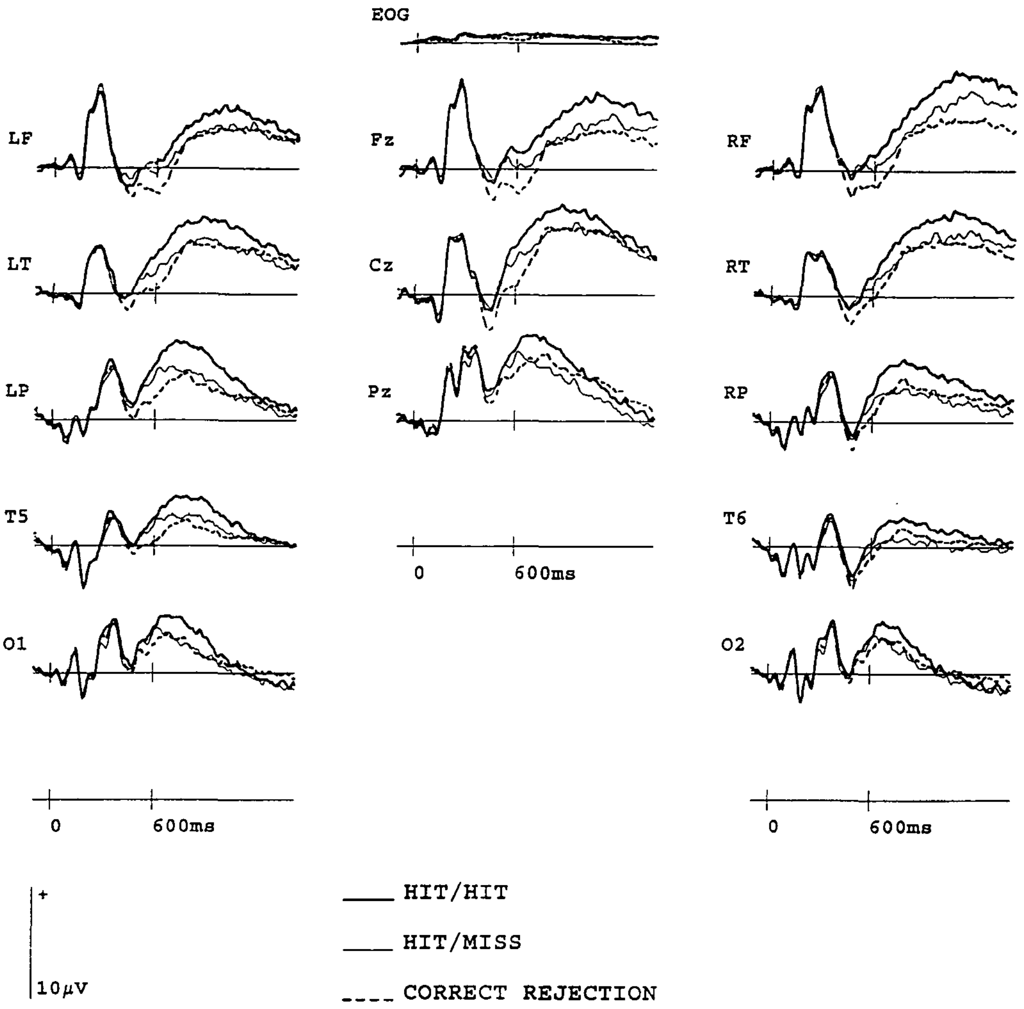

Fig. 3 Grand average ERPs associated with the hit/hit, hit/miss, and correct rejection response categories in Experiment 2. Electrode sites as for Fig. 1.

subject's waveforms were 51, 52 and 91, for the hit/hit, hit/miss, and correct rejection categories, respectively. (For ease of comparability with Experiment 1, the waveforms from the additional prefrontal and parietal electrode sites are not illustrated or analysed here. Analyses including these sites did not alter any of the conclusions drawn below from the analyses of the more restricted data set.)

Compared with the ERPs evoked by correct rejections, the hit/hit waveforms show a widespread positive-going modulation. This is larger over the left than the right hemisphere at posterior sites, but shows an opposite asymmetry, and a more extended time course, at frontal electrodes. The ERPs from the hit/miss category show a similar, but attenuated pattern of effects.

These ERPs were subjected to the same set of analyses as those from Experiment 1, and the mean amplitudes of the 500–500, 800–1100 and 1100–1400 ms latency regions are given in Table 4.

The overall ANOVA of the data from the 500–800 ms region revealed a single significant effect of category [$F(1.9,28.3) = 7.86$, $P < 0.005$]. The follow-up analysis contrasting the hit/hit and correct rejection categories gave rise solely to the same effect [$F(1,15) = 12.52$, $P < 0.005$], suggesting that the differences between the two categories are distributed homogeneously over the scalp. An analysis restricted to the parietal electrodes did, however, reveal a significant interaction between category and hemisphere [$F(1,15) = 4.73$, $P < 0.05$], indicating that, as in Experiment 1, the differences in this latency range are distributed asymmetrically between these sites. The contrast between the hit/miss and correct rejection categories revealed significant interactions between category and site [$F(2.1,30.9) = 3.55$, $P < 0.05$], and between these two factors and hemisphere [$F(2.3,35.1) = 3.89$, $P < 0.025$]. These effects reflect the fact that the differences between these categories are distributed symmetrically at frontal sites (1.50 μV and 1.63 μV for left and right sites, respectively), but are asymmetric more posteriorly (1.52 μV and 0.46 μV). A final subsidiary ANOVA, comparing the hit/hit and hit/miss categories directly, gave rise solely to a category effect [$F(1,15) = 6.13$, $P < 0.05$], indicating that the differences between these ERPs are statistically equivalent at all scalp sites.

Table 4 *Mean amplitude (μV) of the 500–800, 800–1100 and 1100–1400 ms latency regions of ERPs evoked by the hit/hit, hit/miss and correct rejection response categories in Experiment 2*

	LF	LT	LP	T5	O1	RF	RT	RP	T6	O2
500–800 ms										
Hit/hit	1.85	5.21	6.73	3.71	5.07	2.96	3.57	4.17	1.23	4.17
Hit/miss	0.75	3.34	4.85	2.49	3.73	1.53	2.26	2.29	–0.18	2.97
Correct rejection	–0.74	2.16	3.34	1.23	3.23	–0.10	1.33	1.83	–0.35	2.75
800–1100 ms										
Hit/hit	6.02	7.71	6.35	3.93	2.91	8.75	8.30	5.53	2.26	1.66
Hit/miss	4.26	5.39	3.98	2.31	1.47	6.25	6.04	2.88	0.34	0.49
Correct rejection	3.93	5.17	3.40	1.55	1.62	4.92	5.65	3.39	1.04	1.01
1100–1400 ms										
Hit/hit	5.02	5.29	2.51	0.95	–0.65	9.82	8.01	3.43	1.03	–1.32
Hit/miss	3.81	3.72	1.40	0.49	–0.98	7.69	6.25	1.59	–0.28	–1.62
Correct rejection	3.77	4.19	1.84	0.65	0.22	5.15	5.23	2.38	0.43	–0.38

For the 800–1100 ms region the overall ANOVA gave rise to reliable effects for category [$F(1.4,21.2) = 9.83$, $P < 0.005$] and for the interaction between category, site and hemisphere [$F(3.3,49.4) = 4.40$, $P < 0.01$]. The ANOVA contrasting the hit/hit and correct rejection categories gave rise to the same two effects [category: $F(1,15) = 9.69$, $P < 0.01$; category×site×hemisphere: $F(2.1,31.7) = 5.77$, $P < 0.01$]. These effects reflect the fact that the differences between these categories are greater over the right hemisphere anteriorly (2.09 and 3.84 μV for left and right frontal electrodes, respectively), but show a slightly reversed asymmetry at posterior sites (2.96 and 2.15 μV at the left and right parietal electrodes, respectively). The equivalent ANOVA for the hit/miss category gave rise solely to a reliable interaction between category, site and hemisphere [$F(2.1,30.9) = 4.78$, $P < 0.02$], reflecting the fact that the differences between these ERPs are qualitatively similar to those between the hit/hit and correct rejection ERPs (0.32, 1.33, 0.57 and –0.51 μV for left and right frontal and parietal sites, respectively). The final comparison, between the hit/hit and hit/miss categories, gave rise only to an effect of category [$F(1,15) = 20.39$, $P < 0.001$], indicating that the hit/hit waveforms are more positive at all electrode sites.

The overall ANOVA of the 1100–1400 ms region gave rise to a reliable category by site interaction, and to a three-way interaction between these factors and hemisphere (category×site: $F(2.7,40.4) = 6.63$, $P < 0.001$; category×site×hemisphere: $F(3.8,57.5) = 6.09$, $P < 0.001$]. These same two interactions were significant when the hit/hit and correct rejection data were compared [$F(1.6,23.6) = 9.49$, $P < 0.005$, and $F(2.5,37.5) = 9.41$, $P < 0.001$, respectively]. These effects arose because of the strongly asymmetric and frontal distribution of the differences between these waveforms (2.45, 4.68, 0.67 and 1.04 μV for the left and right frontal and parietal sites, respectively). As would be expected from this pattern of results, a subsidiary ANOVA on the data from the frontal sites alone gave rise to a reliable interaction between response category and hemisphere [$F(1,15) = 21.07$, $P < 0.001$], whereas the equivalent ANOVA for the parietal sites revealed no significant effects.

The ANOVA contrasting the 1100–1400 ms data from the hit/miss and correct rejection categories gave rise to the same two interactions [category×site: $F(1.7,25.7) = 8.56$, $P < 0.005$; category×site×hemisphere: $F(2.3,34.6) = 7.02$, $P < 0.005$], which reflects the same pattern of differences, albeit of a smaller magnitude. The direct comparison of the hit/hit and hit/miss data gave rise solely to an effect of category [$F(1,15) = 4.68$, $P < 0.05$], reflecting the greater overall positivity of the hit/hit waveforms.

Topographic analyses

A single ANOVA was employed to compare the scalp distributions of the hit/hit and hit/miss old/new effects over the 500–800 and 1100–1400 ms latency regions. This analysis revealed no effects involving the factor of response category, but did give rise to a significant interaction between latency region and electrode site [$F(5.3,78.9) = 5.08$, $P < 0.001$], indicating a change in scalp distribution with time. An ANOVA restricted to the data from the frontal sites gave rise to a significant response category×hemisphere interaction [$F(1,15) = 5.75$, $P < 0.05$], reflecting more asymmetric (right>left) old/new effects in the latter of the two regions. The equivalent ANOVA for the parietal sites also gave rise to a category×hemisphere interaction. In this case, the interaction [$F(1,15) = 16.10$, $P < 0.001$]. arose because an initially strong left>right asymmetry became significantly weaker with time.

Misses and false alarms

The ERPs to misclassified old and new words are shown in Fig. 4, compared with the ERPs evoked by words correctly judged new. In comparison with this latter category, neither class of ERP showed any evidence of the positive-going effects observed for the ERPs evoked by words that were correctly judged old.

Discussion

The findings from this experiment resemble those from Experiment 1 in two respects. First, there was once again

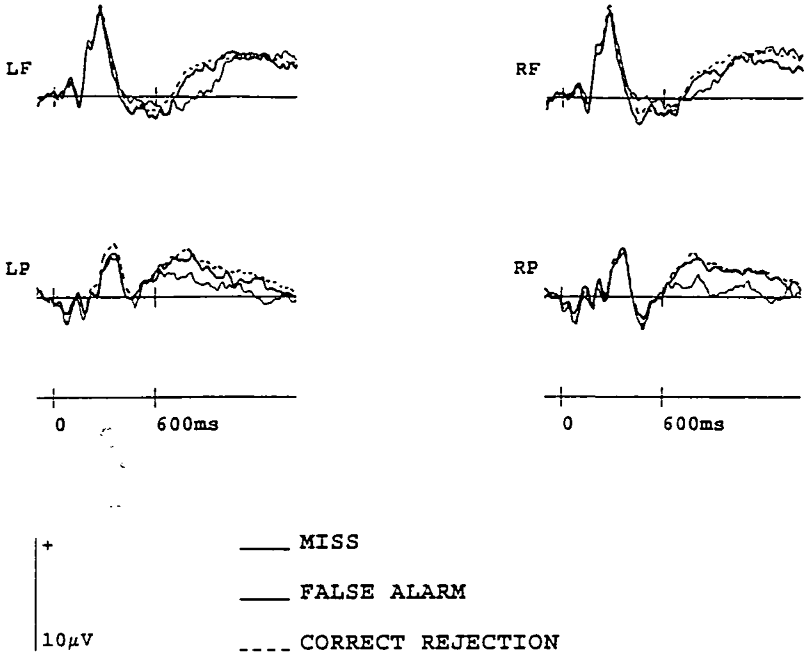

Fig. 4 Grand average ERPs to misses, false alarms, and correct rejections in Experiment 1. Electrode sites as for Fig. 2. The misses are collapsed across study voice, and the false alarms are collapsed across the Decision 2 voice judgement.

clear evidence that the differences between ERPs elicited by correctly classified old and new words reflected the modulation of two components, one maximal at the left parietal electrode, the other at the right frontal site. Secondly, the ERPs elicited by words belonging to the hit/hit category were more positive-going than those evoked by words from the hit/miss trials, although these latter waveforms also differed from the ERPs evoked by words correctly judged new.

Unlike in Experiment 1, the present experiment offered subjects the opportunity to signal that they were uncertain about the status of a test word, and trials on which a don't know response was made on the first decision were discarded. Thus, compared with the first experiment, a smaller proportion of correct old and new judgements should have been based on a guess, providing less opportunity for ERPs to hit/hit and hit/miss items to be differentially contaminated by trials associated with guesses. Therefore, if the differences between the ERPs evoked by these items in Experiment 1 reflected such differential contamination, they should have been less evident in the waveforms from the present experiment. Because subjects made rather little use of the don't know response option for the initial recognition decision, employing it on average on 10% of the trials containing a new word and 15% of the new word trials, it is arguable that the present experiment does not provide a very strong test of this hypothesis. Nonetheless, the findings reveal no sign of a reduction in the magnitude of the old/new effects that were obtained, compared with those found in Experiment 1; on the contrary, the effects are somewhat larger than they were in the previous experiment. Thus, differences between the two classes of ERP may indeed reflect differences in the information retrieved on the two types of trial: information sufficient in both cases for a correct old judgement, but in only one case for a correct context judgement.

As already noted, the data from the present experiment resemble those from Experiment 1 in a number of respects. Importantly, these resemblances include the failure to find any evidence for a qualitative difference in the old/new effects for the hit/hit and hit/miss ERPs. Despite the differences in their magnitudes, the scalp distributions of the two effects were equivalent, strongly suggesting that they reflect the modulation of the same population of generators. Thus, as in Experiment 1 and the study of Wilding *et al.* (1995), the findings do not offer support for the view that recognition judgements with and without contextual retrieval are based upon functionally or neurologically distinct processes.

The present findings differ from those of Experiment 1 in two main respects. First, the magnitude of both components of the old/new effect evoked by hit/hit items was greater in the present experiment. These differences most likely reflect the influence of the don't know response alternative available for each of the two decisions, the combined effects of which would have been to increase the proportion of trials on which both the recognition and source judgements were based on veridical information, and hence the proportion of single trials in the hit/hit averages carrying an old/new effect.

The second difference between the two experiments concerns the hit/miss waveforms, which in the present experiment differed at the right frontal electrode from the ERPs elicited by correct rejections. One possibility is that this effect reflects the contribution to the hit/miss ERPs of

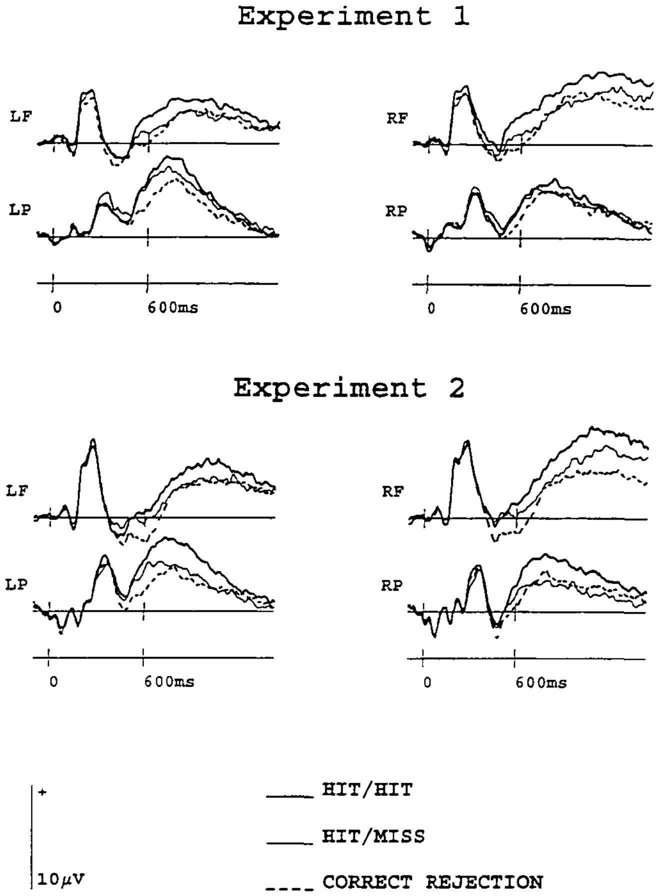

Fig. 5 Grand average ERPs for the hit/hit, hit/miss, and correct rejection response categories. Electrode sites as for Fig. 2.

trials on which, in the absence of a don't know response alternative for the second judgement, a correct, albeit unconfident voice judgement would have been made. In other words, the provision of the second don't know decision meant that some proportion of weakly recollected words were transferred from the hit/hit to the hit/miss response categories. In addition, these effects may also reflect the reduction in the frequency with which a source judgement was made to a word that had been guessed old.

General discussion

The principal findings from the two experiments are summarized in Fig. 5. In both experiments, the ERP old/new effects took the form of the enhancement of two positive-going ERP features, which could be dissociated on a combination of topographic and temporal characteristics. One effect onset at ~400 ms post-stimulus, showed a left parietal maximum and persisted for <1 s. The other effect onset at about the same latency, displayed a right frontal maximum and was still near to maximum amplitude at the end of the 1434 ms recording epoch.

Both of these effects were greater for ERPs to words for which a correct old judgement and a correct source judgement had been made (hit/hit items) than they were for ERPs elicited by words that, although correctly judged old, attracted either incorrect or uncertain source judgements (hit/miss items). Topographic analyses suggested that the differences between the old/new effects associated with these two classes of ERP were largely quantitative, and thus that the hit/hit and hit/miss old/new effects reflected the activity of the same combination of intracerebral generators.

As already noted, the differences found in these experiments between the hit/hit and hit/miss ERPs are unlikely to reflect the differential contribution of guessing to the two classes of ERP. It is also unlikely that these differences reflect gross differences in the confidence with which the initial recognition decision was made. First, the differences were undiminished when, in Experiment 2, subjects were offered the option of signalling a don't know response, an option which would presumably have attracted some of the unconfident old responses which might otherwise have contributed disproportionately to the hit/miss response category. Secondly, in neither experiment did the reaction times for the initial recognition decision differ according to whether the subsequent source judgement was correct or incorrect. To the extent that reaction time for this decision was influenced by confidence, this finding suggests that the initial recognition decisions for these two classes of item were made equally confidently.

As has been reported previously (Smith, 1993; Wilding *et al.*, 1995), neither old words misclassified as new (misses), nor new words misclassified as old (false alarms) elicited old/new effects. Taken together, these findings indicate that these effects are a consequence neither of item repetition, nor of the execution of an 'old' response, unless these events are associated with some form of successful memory retrieval.

What light do the findings from these experiments shed on the different models of recognition memory outlined in the Introduction? Clearly, they do not offer support for the idea that the recognition judgements for hit/hit and hit/miss items were based on functionally or neurally distinct processes. In particular, the absence of an ERP effect that was of equal or greater magnitude in the ERPs elicited by hit/miss items, suggests that no reliable ERP correlate of familiarity-based recognition exists in these data. Thus the findings do not provide support for the kind of dual process model of recognition memory proposed by Jacoby and colleagues (Jacoby and Kelley, 1992). Instead, they appear consistent with proposals (Moscovitch, 1992; Squire *et al.*, 1993) that the difference between recognition with and recognition without contextual retrieval is more one of degree than of kind.

Although the present findings do not provide direct evidence for the idea that the recognition of hit/miss items was more likely to be based on a fluency-based familiarity process than was the recognition of hit/hit items, they are not wholly inconsistent with this idea. It is possible that fluency-based recognition did play a role in the initial recognition decision, and indeed that this role was greater for hit/miss items, but that ERPs are insensitive to the processing underlying this form of recognition (Smith and Halgren, 1989). That said, the findings show that on hit/miss trials the test items evoked brain activity qualitatively similar to, though of smaller amplitude than that associated with 'full' recollection (as indexed by accurate source memory). On the assumption that this attenuated activity reflects the retrieval of information sufficient for an accurate recognition judgement, but lacking the detail required for the more demanding source judgement, there seems little need to postulate any additional basis for the initial recognition memory judgement.

As in previous studies of recognition memory (Neville *et al.*, 1986; Rugg and Doyle, 1992), the ERP old/new effects in the present experiments involved a positive component, the modulation of which was greater over the left than over the right hemisphere. Several lines of evidence converge to suggest that this effect is evoked when items engender recognition accompanied by recollection (Smith and Halgren, 1989; Paller and Kutas, 1992; Smith, 1993; Paller *et al.*, 1995; Rugg *et al.*, 1995; Wilding *et al.*, 1995), and the present findings are consistent with this suggestion. They are also consistent with the proposal of Rugg *et al.* (1995) that the magnitude of this old/new effect is correlated with the quality or amount of information retrieved in response to the test item.

The old/new effects in the present experiments also involved the modulation of a second ERP component. This took the form of the enhancement of a slow, frontally distributed positive wave, which was both larger and more sustained in time over the right than the left hemisphere. The prominence of this old/new effect here, as opposed to its lack of prominence in previous ERP studies of recognition memory, may be a consequence of the requirement explicitly to retrieve information about the study context of each item. The finding that its magnitude was greater for words attracting correct rather than incorrect context judgements suggests that the processing reflected by this component may play a functional role in such judgements. The results of Experiment 2, in which an attenuated version of the effect was evident in the ERPs to hit/miss items, suggest that, like its more posteriorly distributed counterpart, it too may represent a graded rather than an all-or-none process.

Taken together, the present and previous findings suggest that the two ERP old/new effects described above may index dissociable functions. These functions map fairly directly onto those that have been suggested previously to underlie recognition with and without contextual retrieval (Moscovitch, 1992; Squire, 1994). One function, indexed by the parietal old/new effect, involves the retrieval of information from declarative/episodic memory, and supports simple judgements of prior occurrence. The second, indexed by the right frontal effect, operates on the products of this retrieval process, and is necessary for the recovery of contextual information.

In view of the similarity between the proposed functional correlates of these two ERP components, and the functional organization of recognition memory proposed by, among others, Squire (1994) and Moscovitch (1992), the question arises as to whether there is evidence for a correspondence at the anatomical level. According to the neuropsychological models, the explicit retrieval of recently acquired information is dependent upon the hippocampal formation and its associated medial temporal and diencephalic structures, whereas the integration of this information into a coherent representation of the learning episode requires, in addition, the participation of one or more regions of the prefrontal cortex.

The functional correlates proposed for the parietal old/new effect imply that it should reflect processes that are dependent upon the medial temporal lobes, and reports that the effect is abolished or attenuated in patients who have sustained damage to these regions are consistent with this proposal (Smith and Halgren, 1989; Rugg *et al.*, 1991; Johnson, 1995). It is, however, very unlikely that the generators of the effect are localized to this region, as scalp electrodes appear to be largely insensitive to ERP activity generated locally within the hippocampus and adjacent structures (Rugg, 1995). Thus, while the parietal old/new effect may serve as an index of memory processes subserved by such structures as the hippocampal formation, the effect is most likely generated elsewhere, perhaps in cortical regions responsive to input from the medial temporal memory system (Teyler and DiScenna, 1986; McClelland *et al.*, 1995).

A somewhat stronger hypothesis can be advanced about the second of the old/new effects identified in the present study. While considerable caution is necessary in making inferences about generator location on the basis of scalp distribution alone, the distribution of these effects is none the less suggestive of a locus in prefrontal cortex, with a greater contribution coming from the right than the left hemisphere. Support for this proposal comes from two sources. First, given the putative functional correlates of this component, the idea that it is generated in prefrontal cortex fits well with the finding that prefrontal lesions are associated with disproportionately poor source memory (Schacter *et al.*, 1984; Shimamura and Squire, 1987; Janowsky *et al.*, 1989). Secondly, findings from recent functional neuroimaging studies have shown that the right dorsolateral prefrontal cortex is selectively activated during tasks requiring episodic memory retrieval (Buckner and Tulving, 1995), although there have been no reports to date of investigations of source memory *per se*. These findings converge with those from the present study to suggest that neural activity in the right prefrontal cortex supports the recovery of contextual information about the item's study episode. If this proposal is correct, lesions of the right prefrontal cortex should result in a greater reduction in the magnitude of the frontal old/new ERP effect, and a more profound impairment of source memory, than otherwise equivalent left-sided lesions.

In summary, the findings from these experiments offer little support for the view that recognition memory judgements that are accompanied or unaccompanied by contextual retrieval engage qualitatively different memory processes. The findings suggest instead that these two forms of recognition engage functionally equivalent processes, but to differing extents. They further suggest that these processes are neurologically dissociable, and provide additional evidence for the involvement of the frontal lobes in tasks requiring memory for study context.

Acknowledgements

This research was funded by a programme grant from the Wellcome Trust. E. L. Wilding was supported by a Wellcome Prize Studentship.

References

Atkinson RC, Juola JF. Factors influencing speed and accuracy of word recognition. In: Kornblum S, editor. Attention and performance IV. New York: Academic Press, 1973: 583–612.

Buckner AL, Tulving E. Neuroimaging studies of memory: theory and recent PET results. In: Baron JC, Grafman J, editors. Handbook of neuropsychology, Vol. 10. Amsterdam: Elsevier, 1995: 439–66.

Friedman D. ERPs during continuous recognition memory for words. Biol Psychol 1990; 30: 61–87.

Gardiner JM, Java RI. Recognising and remembering. In: Collins AF, Gathercole SE, Conway MA, Morris PE, editors. Theories of memory. Hove (UK): Lawrence Erlbaum, 1993: 163–88.

Haist F, Shimamura AP, Squire LR. On the relationship between recall and recognition memory. J Exp Psychol Learn Mem Cogn 1992; 18: 691–702.

Hirst W, Johnson MK, Kim JK, Phelps EA, Risse G, Volpe BT. Recognition and recall in amnesics. J Exp Psychol Learn Mem Cogn 1986; 12: 445–51.

Hirst W, Johnson MK, Phelps EA, Volpe BT. More on recognition and recall in amnesics. J Exp Psychol Learn Mem Cogn 1988; 14: 758–62.

Humphreys MS, Bain JD, Pike R. Different ways to cue a coherent memory system: a theory for episodic, semantic, and procedural tasks. Psychol Rev 1989; 96: 208–33.

Jacoby LL. Perceptual enhancement: persistent effects of an experience. J Exp Psychol Learn Mem Cogn 1983; 9: 21–38.

Jacoby LL, Dallas M. On the relationship between autobiographical memory and perceptual learning. J Exp Psychol Gen 1981; 110: 306–40.

Jacoby LL, Kelley C. Unconscious influences of memory: dissociations and automaticity. In: Milner AD, Rugg MD, editors. The neuropsychology of consciousness. London: Academic Press, 1992: 201–33.

Janowsky JS, Shimamura AP, Squire LR. Source memory impairment in patients with frontal lobe lesions. Neuropsychologia 1989; 27: 1043–56.

Jasper HH. The ten twenty electrode system of the international federation. Electroencephalogr Clin Neurophysiol 1958; 10: 371–5.

Johnson R. Event-related potential insights into the neurobiology of memory systems. In: Baron JC, Grafman J, editors. Handbook of neuropsychology, Vol. 10. Amsterdam: Elsevier, 1995: 135–64.

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

Johnston WA, Hawley KJ, Elliott JMG. Contribution of perceptual fluency to recognition judgments. J Exp Psychol Learn Mem Cogn 1991; 17: 210–23.

Johnson MK, Hashtroudi S, Lindsay DS. Source monitoring. [Review]. Psychol Bull 1993; 114: 3–28.

Kelley CM, Jacoby LL, Hollingshead A. Direct versus indirect tests of memory for source: judgments of modality. J Exp Psychol Learn Mem Cogn 1989; 15: 1101–8.

Keselman HJ, Rogan JC. Repeated measures F tests and psychophysiological research: controlling the number of false positives. Psychophysiology 1980; 17: 499–503.

Knowlton BJ, Squire LR. Remembering and knowing: two different expressions of declarative memory. J Exp Psychol Learn Mem Cogn 1995; 21: 699–710.

Kučera H, Francis WN. Computational analysis of present-day American English. Providence (RI): Brown University Press, 1967.

Mandler G. Recognizing: the judgment of previous occurrence. Psychol Rev 1980; 87: 252–71.

McCarthy G, Wood CC. Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. Electroencephalogr Clin Neurophysiol 1985; 62: 203–8.

McClelland JL, McNaughton BL, O'Reilly RC. Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. [Review]. Psychol Rev 1995; 102: 419–57.

Moscovitch M. Memory and working-with-memory: a component process model based on modules and central systems. J Cognit Neurosci 1992; 4: 257–67.

Moscovitch M. Models of consciousness and memory. In: Gazzaniga MS, editor. The cognitive neurosciences. Cambridge (MA): MIT Press, 1994: 1341–56.

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

Paller KA, Kutas M. Brain potentials during memory retrieval provide neurophysiological support for the distinction between conscious recollection and priming. J Cognit Neurosci 1992; 4: 375–91.

Paller KA, Kutas M, McIsaac HK. Monitoring conscious recollection via the electrical activity of the brain. Psychol Sci 1995; 6: 107–11.

Potter DD, Pickles CD, Roberts RC, Rugg MD. The effects of scopolamine on event-related potentials in a continuous recognition memory task. Psychophysiology 1992; 29: 29–37.

Rajaram S. Remembering and knowing: two means of access to the personal past. Mem Cognit 1993; 21: 89–102.

Richardson-Klavehn A, Bjork RA. Measures of memory. Annu Rev Psychol 1988; 39: 475–543.

Rugg MD. ERP studies of memory. In: Rugg MD, Coles MGH, editors. Electrophysiology of mind: event-related brain potentials and cognition. London: Oxford University Press, 1994: 113–27.

Rugg MD. Cognitive event-related potentials: Intracranial and lesion studies. In: Baron JC, Grafman J, editors. Handbook of Neuropsychology, Vol. 10. Amsterdam: Elsevier, 1995: 165–86.

Rugg MD, Doyle MC. Event-related potentials and recognition memory for low- and high-frequency words. J Cognit Neurosci 1992; 4: 69–79.

Rugg MD, Roberts RC, Potter DD, Pickles CD, Nagy ME. Event-related potentials related to recognition memory. Brain 1991; 114: 2313–32.

Rugg MD, Cox CJC, Doyle MC, Wells T. Event-related potentials and the recollection of low and high frequency words. Neuropsychologia 1995; 33: 471–84.

Schacter DL, Harbluk JL, McLachlan DR. Retrieval without recollection: an experimental analysis of source amnesia. J Verb Learn Verb Behav 1984; 23: 593–611.

Shimamura AP, Squire LR. A neuropsychological study of fact memory and source amnesia. J Exp Psychol Learn Mem Cogn 1987; 13: 464–73.

Smith ME. Neurophysiological manifestations of recollective experience during recognition memory judgements. J Cognit Neurosci 1993; 5: 1–13.

Smith ME, Halgren E. Dissociation of recognition memory components following temporal lobe lesions. J Exp Psychol Learn Mem Cogn 1989; 15: 50–60.

Snodgrass JG, Corwin J. Pragmatics of measuring recognition memory: applications to dementia and amnesia. J Exp Psychol Gen 1988; 117: 34–50.

Squire LR. Declarative and nondeclarative memory: multiple brain systems supporting learning and memory. In: Schacter DL, Tulving E, editors. Memory systems 1994. Cambridge (MA): MIT Press, 1994: 203–31.

Squire LR, Knowlton B, Musen G. The structure and organization of memory. [Review]. Annu Rev Psychol 1993; 44: 453–95.

Teyler TJ, DiScenna P. The hippocampal memory indexing theory. Behav Neurosci 1986; 100: 147–54.

Van Petten C, Kutas M, Kluender R, Mitchiner M, McIsaac H. Fractionating the word repetition effect with event-related potentials. J Cognit Neurosci 1991; 3: 131–50.

Whittlesea BWA. Illusions of familiarity. J Exp Psychol Learn Mem Cogn 1993; 19: 1235–53.

Wilding EL, Doyle MC, Rugg MD. Recognition memory with and without retrieval of context: an event-related potential study. Neuropsychologia 1995; 33: 743–67.

Yonelinas AP. Receiver-operating characteristics in recognition memory: evidence for a dual-process model. J Exp Psychol Learn Mem Cogn 1994; 20: 1341–54.

Received September 14, 1995. Revised December 15, 1995. Accepted January 29, 1996