Recognition memory for emotionally negative and neutral words: an ERP study

Elizabeth J. Maratos*, Kevin Allan, Michael D. Rugg

Institute of Cognitive Neuroscience and Department of Psychology, University College London, 17 Queen Square, London WC1N 3AR, UK

Received 12 July 1999; received in revised form 27 March 2000; accepted 3 April 2000

Abstract

Scalp recorded event-related potentials were used to investigate the neural activity elicited by emotionally negative and emotionally neutral words during the performance of a recognition memory task. Behaviourally, the principal difference between the two word classes was that the false alarm rate for negative items was approximately double that for the neutral words. Correct recognition of neutral words was associated with three topographically distinct ERP memory ‘old/new’ effects: an early, bilateral, frontal effect which is hypothesised to reflect familiarity-driven recognition memory; a subsequent left parietally distributed effect thought to reflect recollection of the prior study episode; and a late onsetting, right-frontally distributed effect held to be a reflection of post-retrieval monitoring. The old/new effects elicited by negative words were qualitatively indistinguishable from those elicited by neutral items and, in the case of the early frontal effect, of equivalent magnitude also. However, the left parietal effect for negative words was smaller in magnitude and shorter in duration than that elicited by neutral words, whereas the right frontal effect was not evident in the ERPs to negative items. These differences between neutral and negative words in the magnitude of the left parietal and right frontal effects were largely attributable to the increased positivity of the ERPs elicited by new negative items relative to the new neutral items. Together, the behavioural and ERP findings add weight to the view that emotionally valenced words influence recognition memory primarily by virtue of their high levels of ‘semantic cohesion’, which leads to a tendency for ‘false recollection’ of unstudied items. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Emotional memory; Familiarity; Recollection; Episodic memory; Semantic cohesion

1. Introduction

Numerous studies have demonstrated that performance on episodic memory tasks is influenced by the emotional nature of the test items. Notably, recall of emotionally negative material (and, in some studies, positive material also) is enhanced relative to the recall of neutral items [5,23,26]. It has been suggested that differences in memory performance for emotional and non-emotional material arise through multiple mechanisms, two of the most important of which are ‘arousal’, and ‘semantic cohesiveness’ [22]. Items that are arousing (i.e. items that elicit significant autonomic responses as indexed, for example, by skin conductance) gain a mnemonic advantage over non-arousing items through the enhancement of their encoding and consolidation in memory. This advantage seems likely to be mediated at the neural level by the modulatory influence of the amygdala on hippocampal and cortical components of the network supporting episodic memory [10,22]. By contrast, non-arousing emotional items (e.g. words with negative emotional valence) influence memory largely because, unlike emotionally neutral items, they tend to belong to categories that are semantically ‘cohesive’, that is, categories in which the constituent items share strong inter-item associations.

The foregoing distinction provides a useful framework in which to view findings of studies comparing
memory for emotionally valenced and emotionally neutral words. With the exception of certain classes of item such as ‘taboo’ words, emotional and non-emotional words differ little in their arousing properties [19] raising the possibility that differences in memory for these two word classes are due mainly to differences in semantic cohesiveness. Consistent with this possibility, Phelps and LaBar (cited in LaBar and Phelps, [19]) found that when inter-item associations were controlled, the recall advantage normally found for emotionally valenced words over neutral items was eliminated. Also consistent are findings showing that the above-mentioned recall advantage does not extend to tests of recognition memory (e.g. see [5,20]). In one of these studies [5], discrimination (d') between studied (old) and unstudied (new) words was lower for emotional than for neutral items (hit and false alarm rates were not reported). In the second study [20], emotional words were associated with a higher hit rate but, because this effect was offset by an equivalent elevation of the false alarm rate, recognition accuracy did not differ for the two word types.

The disparate pattern of findings for recall and recognition fit well with the idea that emotional valence exerts much of its effect on memory for words through the mechanism of semantic cohesion. The fact that emotional words generally belong to a relatively ‘closed’ semantic category means that recall for these items benefits for the same reasons that underlie the recall advantage for words from any semantically categorised set relative to uncategorised words [2]. In the case of recognition, the findings of Leiphart et al. [20] suggest that emotionally valenced words may act like the ‘related lures’ that elicit high false alarm rates in studies of ‘false recollection’. In these studies (e.g. [24,25,35] see [15] for review) subjects typically study lists of semantically related items and subsequently attempt to discriminate between these items and two kinds of new word — semantic associates of studied words (related lures), and words semantically unrelated to items that had been shown at study (unrelated lures). Relative to unrelated lures, related lures generate high false alarm rates, with subjects reporting many of these responses to be based on a recollection of the lure as a member of the study list (‘Remember’ responses, e.g. [25]). Neuroimaging and electrophysiological evidence suggest that the illusory recollection of related lures relies on much of the same neural circuitry that underlies veridical recollection [7,14,34,35]. Since sets of emotionally valenced words tend to be semantically and associatively related, it is easy to see how these items might also engage processes supporting false recollection.

According to the foregoing analysis, therefore, emotionally valenced words do not exert their effect on memory by virtue of their valence per se unless they are arousing. Rather, they do so because emotionally valenced words have, on average, stronger inter-item associations than do sets of unselected neutral words. Thus, memory for emotional words is mediated not by systems or processes to which these items have privileged access, but by the same cognitive operations that support memory for non-emotional material.

In the present experiment, we assess the foregoing proposal by investigating the neural correlates of recognition memory for emotionally valenced and neutral words using event-related brain potentials (ERPs). In general terms, the proposal that recognition memory for the two classes of word engages equivalent cognitive operations can be assessed by determining whether the memory-related ERP effects they elicit differ with respect to scalp distribution. If the effects for emotional and neutral words do differ in this respect, it would indicate that memory for the two kinds of material is neurally (and, therefore, most likely functionally) dissociable [29]. Such a finding would be inconsistent with the proposal that memory for emotional and neutral words relies on functionally equivalent memory mechanisms.

The results of previous research investigating the ERP correlates of recognition memory allow predictions arising from the ‘semantic cohesiveness’ hypothesis to be addressed more specifically. It has long been known that, relative to new words, ERPs elicited by correctly classified old words in tests of recognition memory are more positive-going — the so-called ERP ‘old/new’ effect (see [33] for review). Recent work (for review see [28]; see also [31]) has led to the identification of three old/new effects which appear to index functionally distinct aspects of recognition. For present purposes, the most important effect is one which onsets around 400–500 ms post-stimulus, and is maximal over the left parietal scalp (the ‘left parietal’ old/new effect). The effect appears to be a neural correlate of the episodic retrieval (recollection) of study items [39] and, in studies of false recollection [7,14] is elicited by both genuinely old words and related lures. Another old/new effect involves an earlier (ca. 300–500 ms), bilateral shift with a frontal distribution. Rugg et al. [31] proposed that this effect was a neural correlate of familiarity-based recognition. This is a form of recognition memory held by some authors to be independent of the ‘semantic cohesiveness’ hypothesis and to underlie recognition judgements associated with ‘Know’ rather than ‘Remember’ responses [9,38]. The third old/new effect onsets quite late (ca. 500–700 ms), persists for a second or more, and is distributed over the right frontal scalp [6,39]. The ‘right frontal’ old/new effect is held to reflect processes that operate on the products of memory retrieval [28].

If, as suggested above, unstudied emotionally valenced words in tests of recognition memory act like
associative lures in studies of false recollection, these items should be difficult to discriminate from studied emotional items because of their tendency to elicit ‘recollection’ of their study presentation. As a consequence, unstudied emotional items should be associated with an elevated false alarm rate relative to unstudied neutral words. Furthermore, even on trials on which unstudied emotional items are correctly classified as new, the tendency of these items to elicit illusory recollection might be expected to both impede the decision to respond ‘new’ and, critically, to be manifest in a pattern of neural activity — the left parietal ERP effect — signifying the engagement of neural systems supporting episodic retrieval. Thus, relative to the left parietal old/new effect elicited by neutral words, the effect elicited by emotionally valenced words should be smaller in magnitude. Furthermore, the difference in the magnitude of the left parietal effects elicited by the two word types should be carried mainly by the ERPs elicited by new items, those elicited by emotional words exhibiting the greater positivity.

We tested these predictions by recording ERPs while subjects discriminated between studied and unstudied emotionally negative and neutral words. For the reasons already noted, we expected negative words to give rise to an excess of false alarms, a smaller left parietal old/new effect, but no evidence for the recruitment of memory mechanisms additional to those engaged by neutral items. Further, to the extent that emotional valence influences recognition memory exclusively through its effects on recollection, the ERP old/new effect held by Rugg et al. [31] to index familiarity-driven recognition should be unaffected by this variable.

2. Method

2.1. Subjects

Twenty young male and female right-handed subjects were employed in the study. Each subject gave informed consent prior to participation in the study and all were remunerated at the rate of £5 per hour. Four subjects’ data were discarded prior to data analysis due to excessive electro-oculographic artefact leading to a failure to provide 16 or more artefact free trials for one or more of the ERP categories. The 16 subjects whose data were analysed consisted of 11 females and five males.

2.2. Experimental material

The critical stimuli consisted of 224 words selected from The Balanced Affective Word Project (a corpus of words normed for emotional valence; [36]). Half of these words were emotionally negative (e.g. Fright), whereas the remainder were emotionally neutral (e.g. Mention). Five neutral buffer words were also used. The words varied in length between three and nine letters. All words were semantically distinct from each other in the sense that derivatives of words from a common root were not used.

Two study lists (A and B) were created, each containing a set of 56 negative and 56 neutral words. The words in the study lists were randomly ordered and a neutral buffer word was added to the beginning and end of each list. Two test lists (1 and 2) were created using all 224 negative and neutral critical words in two randomised orders, with a buffer word added to the beginning of both lists.

2.3. Procedure

The experiment took the form of a single study-test cycle. The combination of study and test list used was counterbalanced across subjects, thus ensuring there was no correlation between word type and old/new status.

The stimuli were presented in central vision on a computer monitor (white upper case words on a dark screen). Each word subtended a maximum vertical visual angle of approximately 0.4° and a maximum horizontal angle of approximately 1.2°. Study words were presented with their first letter replacing the fixation character. Test words were presented with the third letter replacing the fixation character. Subjects were tested in a sound attenuated recording booth. Prior to the experiment, subjects were fitted with an electrode cap. Subjects were informed that the experiment consisted of two parts but they were not informed that the second part would involve a memory test. An interval of around 5 min separated the study and test phases, and during this time subjects were instructed to relax.

2.4. Study phase

Each trial consisted of a presentation of the word for 300 ms, followed immediately by the restoration of the fixation character. This character remained in view until the next trial was initiated (see below). Subjects were required to give an affective rating to each study word using a scale that ranged from −3 (negative) through 0 (neutral) to +3 (positive). They were instructed to make positive ratings to words they associated with feelings of happiness, content or satisfaction, negative ratings to words they associated with unpleasant feelings of sadness, anger or anxiety, and neutral ratings for words which were neither negative
nor positive. The experimenter initiated each trial after recording the response to the preceding item.

2.5. Test phase

Subjects were presented with one of the two test lists. Each trial consisted of the presentation a fixation character for 2100 ms, followed by a 124 ms period during which the screen was blank. This period was followed by the display of a test item for 300 ms, following which the monitor was blanked for 2700 ms. Subjects were required to judge whether or not each word had appeared previously in the study phase. Responses were made by button press using one or other index finger. The mapping of keys to response type was counterbalanced across subjects. Subjects were instructed to relax, to keep body movements to a minimum, and to blink only when the fixation character was in view. The test phase was computer controlled. Responses faster than 300 ms or slower than 2500 ms were treated as errors.

2.6. Event-related potential recording

EEG was recorded from 25 tin electrodes embedded in an elasticated cap and from an electrode placed on the right mastoid process. All channels were referenced to a left mastoid electrode, and re-referenced off-line to represent recordings with respect to linked mastoids. EOG was recorded bipolarly from electrodes placed above the supraorbital ridge of the right eye and on the outer canthus of the left eye. EEG electrodes were located according to the International 10-20 system [13]. These were located over the midline (Fz, Cz, Pz), and at the following additional locations: lateral frontal (FP1/FP2, F3/F4, LF/RF (75% of the distance from Fz to F7/8), and F7/F8), central/anterior temporal (C3/C4, LT/RT (75% of the distance from Cz to T3/4), and T3/T4), parietal/posterior temporal (P3/P4, LP/RP (75% of the distance from Pz to T5/6), T5/T6), and occipital (O1, O2). Data were sampled at a rate of 8 ms per point and digitised with 12 bit resolution. The duration of the recording epoch was 2048 ms with a 104 ms pre-stimulus baseline period.

All channels were amplified with a bandpass of 0.032–35 Hz. Trials on which base to peak EOG activity exceeded 98 µV were rejected, as were trials on which A/D saturation occurred, or on which baseline drift across the recording epoch (i.e. amplitude of the first point minus the amplitude of the last point of the epoch) exceeded ±55 µV in any EEG channel.

### Table 1

<table>
<thead>
<tr>
<th>Word type</th>
<th>Hit</th>
<th>CR</th>
<th>Pr</th>
<th>Br</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>0.83 (0.10)</td>
<td>0.86 (0.12)</td>
<td>0.73</td>
<td>0.41</td>
</tr>
<tr>
<td>Negative</td>
<td>0.89 (0.07)</td>
<td>0.66 (0.14)</td>
<td>0.57</td>
<td>0.72</td>
</tr>
</tbody>
</table>

ERPs were formed for four response conditions: correctly recognised old neutral words (neutral hits); correctly rejected new neutral words (neutral correct rejections); correctly recognised old negative words (negative hits); and correctly rejected new negative words (negative correct rejections). The waveforms were smoothed with a five point binomially weighted filter. There were insufficient trials to form ERPs to items associated with error trials (i.e. misses and false alarms).

3. Results

3.1. Behavioural data

3.1.1. Affective ratings

The mean rating assigned to the negative words was \(-1.73\), (across subject SD = 0.35; range \(-2.38\) to \(-1.07\)) whereas the mean rating for the neutral words was 0.52, (across subject SD = 0.40; range 0.04 to 1.30). These means differed reliably (\(t_{15}=15.28, p < 0.001\)). For every subject the mean rating given for the negative words was significantly lower than that given for the neutral items (minimum \(t_{110}=11.66, p < 0.001\)).

Hit and correct rejection rates for the neutral and negative items, along with discrimination (Pr) and bias (Br) indices [37], are shown in Table 1. Pr for neutral items was significantly greater than that for negative items (\(t_{15}=5.01, p < 0.001\)). Bias also differed significantly between word types (\(t_{15}=7.45, p < 0.001\)), such that subjects were more willing to respond ‘yes’ to negative items. Additional \(t\)-tests revealed that hit rates and false alarm rates were both significantly greater for negative items (\(t_{15}=2.93, p = 0.01\), and \(t_{15}=6.28, p < 0.001\), respectively).

### Table 2

<table>
<thead>
<tr>
<th>Word type</th>
<th>Hit</th>
<th>CR</th>
<th>False alarm</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>999 (242)</td>
<td>1097 (267)</td>
<td>1313 (272)</td>
</tr>
<tr>
<td>Negative</td>
<td>975 (249)</td>
<td>1184 (268)</td>
<td>1188 (292)</td>
</tr>
</tbody>
</table>
Reaction time (RT) data are shown in Table 2. ANOVA of the RTs for hits and correct rejections revealed main effects of response type ($F(1,15) = 10.32, p < 0.01$), valence ($F(1,15) < 17.90, p = 0.001$), and an interaction between these factors ($F(1,15) = 17.46, p < 0.001$). Tukey HSD tests revealed that RTs were faster for correct rejections to the neutral rather than the negative items, and that RTs for hits were faster than those for correction rejections in the case of the negative items only. A further test revealed that RTs to false alarms were faster for negative words ($t_{15}=2.81, p < 0.01$).

Fig. 1. Top: Grand average ERP waveforms elicited by correctly classified old and new neutral words. Bottom: Grand average ERP waveforms elicited by correctly classified old and new negative words. Electrode locations are described in the text. Gain for the EOG channel is $\times 5$ lower than for the EEG.
3.1.2. ERPs

The mean numbers of trials (range in brackets) contributing to the ERPs for neutral hits, neutral correct rejections, negative hits and negative correct rejections were 38 (22–49), 39 (23–50), 38 (20–50) and 29 (18–38), respectively. The proportion of trials lost because of artefact was 18, 19, 22 and 20%, respectively. These proportions did not differ significantly. Grand average waveforms from selected lateral and midline sites are shown overlaid according to response category and valence in Figs. 1 and 2 respectively. ERPs begin to diverge as a function of response category and word type from around 300 ms post-stimulus onset. ERPs to old words become more positive-going than those to new words.

**NEW**

**OLD**

Fig. 2. Top: Grand average ERP waveforms elicited by correctly classified neutral and negative new words. Bottom: Grand average ERP waveforms elicited by correctly classified neutral and negative old words.
new words, and ERPs to negative items become more positive-going than those to the neutral items. From approximately 500 ms post-stimulus, old/new effects demonstrate a left posterior maximum. From around 800 ms post-stimulus, the old/new effects for negative words are no longer evident, whereas those for neutral words begin to exhibit a right frontal maximum which extends for approximately a further 500 ms.

ERPs were quantified by measuring (with respect to the mean of the pre-stimulus baseline) the mean amplitudes of four consecutive latency regions (300–500 ms, 500–800 ms, 800–1100 ms, 1100–1400 ms). These intervals correspond closely with the regions chosen for analysis in previous studies (e.g. [6,39]).

Two sets of analyses were performed on these data. The first set investigated differences in the amplitude of ERP effects associated with the factors of valence and category. The second set tested for differences in the scalp distribution of these effects, and whether the distribution of these effects changed over time. In all ANOVAs, degrees of freedom were corrected for non-sphericity by application of the Greenhouse–Geisser procedure [16] and F ratios are reported with corrected degrees of freedom. Data employed in the topographic analyses were rescaled to eliminate the confounding effects of between-condition and between-epoch differences in amplitude [21].

3.2. Mean amplitude analyses

To assess amplitude differences between the conditions, ANOVAs were conducted for each latency region on the data from lateral frontal (F7/8, LF/RF, F3/4), temporal/central (T3/4, LT/RT, C3/4), and parietal (T5/6, LP/RP, P3/4) electrode sites, employing the factors of valence (negative vs neutral), response category (hit vs correct rejection), hemisphere, location (frontal, temporal/central, parietal) and site (inferior, middle, superior). The results of these ANOVAs with respect to the valence and response category factors are shown in Table 3. The ANOVAs were followed up by additional subsidiary ANOVAs as reported below.

As can be seen in Table 3, analysis of the 300–500 ms latency range revealed effects of valence and response category, reflecting the greater positivity of ERPs elicited by negative words and hits respectively. Crucially, while the factor of valence interacted with the factors of location, hemisphere, and site (see Table 3), it did not interact with response category (maximum $F = 1.57$).

A further ANOVA was conducted on the data from the 300–500 ms latency region. This focused on the three sites — F3, Fz, F4 — at which the putative ERP correlate of familiarity described by Rugg et al. [30,31] was of maximum amplitude in that study. The waveforms from these sites are shown in Fig. 3. The ANOVA revealed significant effects of valence ($F(1,15)=26.53, p < 0.001$) and response category ($F(1,15)=7.67, p < 0.025$), but no interaction between these factors (max $F(1.6, 24.4)=2.27, p > 0.1$). Separate ANOVAs showed that the effect of valence was reliable for both old and new items ($F(1,15)=18.21, p < 0.001$ and $F(1,15)=18.68, p = 0.001$, respectively), and that the effect of response category was reliable for both neutral and negative items ($F(1,15)=5.95, p < 0.05$, and $F(1,15)=5.14, p < 0.05$, respectively).

Table 3 shows that the ANOVA of data from the 300–500 ms region revealed main effects of valence and response category, along with three interactions (one of marginal significance) involving these factors. As can be seen in Fig. 4, old/new effects in this latency range showed a left posterior maximum for both

<table>
<thead>
<tr>
<th></th>
<th>300–500 ms</th>
<th>500–800 ms</th>
<th>800–1100 ms</th>
<th>1100–1400 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>NE</strong></td>
<td>$F(1,15)=31.6, p &lt; 0.001$</td>
<td>$F(1,15)=8.4, p &lt; 0.025$</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><strong>ON</strong></td>
<td>$F(1,15)=6.9, p &lt; 0.025$</td>
<td>$F(1,15)=11.1, p = 0.005$</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><strong>NE × ON</strong></td>
<td>$F(1,15)=4.2, p = 0.058$</td>
<td>$F(1,15)=14.3, p &lt; 0.0025$</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><strong>NE × CH</strong></td>
<td>$F(1.6,24)=19.6, p &lt; 0.001$</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><strong>ON × ST</strong></td>
<td>–</td>
<td>$F(1,15)=22.1, p &lt; 0.001$</td>
<td>$F(1,15)=7.4, p = 0.01$</td>
<td>–</td>
</tr>
<tr>
<td><strong>ON × HM</strong></td>
<td>–</td>
<td>$F(1.5,4)=6.5, p &lt; 0.025$</td>
<td>$F(1,1,16.1)=4.7, p &lt; 0.05$</td>
<td>–</td>
</tr>
<tr>
<td><strong>NE × ON × ST</strong></td>
<td>–</td>
<td>$F(1.1,16.2)=9.9, p = 0.005$</td>
<td>$F(1,15.7)=8.5, p = 0.01$</td>
<td>–</td>
</tr>
<tr>
<td><strong>NE × CH × HM</strong></td>
<td>–</td>
<td>$F(1.9,28.5)=7.3, p &lt; 0.005$</td>
<td>$F(1,5.22.6)=4.3, p &lt; 0.05$</td>
<td>–</td>
</tr>
<tr>
<td><strong>NE × CH × ST</strong></td>
<td>$F(2.6,38.4)=6.6, p &lt; 0.005$</td>
<td>–</td>
<td>–</td>
<td>$F(2.5,37.2)=3.1, p = 0.05$</td>
</tr>
<tr>
<td><strong>NE × HM × ST</strong></td>
<td>$F(1.8,26.9)=6.6, p &lt; 0.01$</td>
<td>$F(1.5,22.9)=5.9, p = 0.01$</td>
<td>$F(1,4.20.4)=5.1, p &lt; 0.05$</td>
<td>–</td>
</tr>
<tr>
<td><strong>ON × CH × HM</strong></td>
<td>–</td>
<td>$F(1,7.25.3)=6.7, p &lt; 0.01$</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td><strong>NE × ON × CH × HM</strong></td>
<td>–</td>
<td>$F(1,4,21.5)=6.0, p = 0.01$</td>
<td>$F(1,7.25.5)=9.5, p = 0.001$</td>
<td>$F(1,3,19.9)=4.5, p &lt; 0.05$</td>
</tr>
</tbody>
</table>

*ON = old/new (response category), NE = neutral/negative (valence), CH = frontal/anterior/temporal location, HM = hemisphere (left, right), ST = electrode site (Inferior, Mid-lateral, Superior). Significant F values of interactions involving the factors of valence and response category are shown in bold.
classes of word. Furthermore, while old/new effects were larger for the neutral than for the negative words at all scalp regions, the disparity between the effects varied across regions.

In order to assess the reliability of the old/new effects elicited by each class of word, separate ANOVAs were conducted on the data for the neutral and negative items. The ANOVA for the neutral items revealed a main effect of response category ($F(1,15)=20.64, p < 0.001$), an interaction between response category and hemisphere ($F(1,15)=15.10, p = 0.001$), and a three-way interaction between response category, location and hemisphere ($F(1.5, 22.2)=12.33, p = 0.001$). The effects reflect the robust nature of the old/new effects elicited by the neutral items, and their left posterior maximum. ANOVA of the data for the negative items revealed interactions between response category and hemisphere ($F(1,15)=8.36, p = 0.01$), and between response category and location ($F(1.3,19.7)=4.35, p < 0.05$). These effects reflect the tendency for the old/new effects elicited by these items to be left lateralised, and in addition to show an anterior-posterior gradient (see Fig. 4).

A final set of analyses were conducted on the data from the 500–800 ms latency in order to focus specifically on the left parietal old/new effect, the putative ERP signature of episodic retrieval (see introduction). An ANOVA was conducted on the data from the left parietal electrode alone. This revealed effects of valence ($F(1,15)=5.40, p < 0.05$) and response category ($F(1,15)=33.90, p = 0.001$), as well as a marginally significant interaction between these factors ($F(1,15)=4.32, p < 0.06$). Separate ANOVAs revealed that the effects of response category were reliable for both the neutral and negative items ($F(1,15)=50.05, p < 0.001$, and $F(1,15)=12.89, p < 0.005$, respectively).

Two further ANOVAs revealed that the waveforms elicited by the new negative words were more positive than those for the new neutral items ($F(1,15)=7.09, p = 0.025$), whereas no such effect was evident for the ERPs to the two classes of old item ($F(1,15) < 1$).

As shown in Table 3, ANOVA of data from the 800–1100 ms region revealed three interactions between the factors of valence and response category, including the four-way interaction between these factors, location and hemisphere. As shown in Fig. 4, these effects reflect the left posterior maximum of the old/new effects in this latency range and the apparent absence of such effects in the ERPs to negative words.

As for the previous latency region, old/new effects in the 800–1100 ms region were assessed separately for neutral and negative words. ANOVA of the data for the neutral items revealed a main effect of response category ($F(1,15)=8.62, p = 0.010$), and interactions between response category and hemisphere ($F(1,15)=7.04, p < 0.025$), and between response category, location and hemisphere ($F(1.7,25.8)=8.56, p < 0.005$). The effects reflect the left posterior distribution of the old/new effects for these items. ANOVA of the data for the negative items revealed no effects involving the factor of response category.

Once again, a further set of analyses was directed toward the left parietal electrode alone. An initial ANOVA revealed a significant effect of response category ($F(1,15)=19.12, p = 0.001$), and a response category $\times$ valency interaction ($F(1,15)=14.62, p < 0.01$). Further ANOVAs revealed a reliable old/new effect for the neutral items ($F(1,15)=38.88, p < 0.001$) but not for the negative items. A final pair of ANOVAs indicated that the ERPs for new negative words were more positive than those for new neutral items ($F(1,15)=8.63, p = 0.01$), whereas the reverse pattern...
was found for the ERPs to old words \((F(1,15)=9.02, p < 0.01)\); (see Fig. 4).

ANOVA of the 1100–1400 ms data revealed yet another four way interaction between valence, response category, anterior/posterior locality and hemisphere (Table 3). Fig. 4 shows that old/new effects in this latency region were confined largely to ERPs elicited over the right anterior scalp by neutral words. In keeping with the impression given by the figure, a further ANOVA confined to data from the frontal sites alone gave rise to an interaction between valence, response category, and hemisphere \((F(1,15)=8.46, p < 0.025)\), as well as to an interaction between valence, response category and site \((F_{1,3.19.8}=4.62, p < 0.05)\).

ANOVA performed on these data for the two word classes separately revealed, for the neutral items, interactions between response category and hemisphere, and response category and electrode site \((F_{1,15}=8.56, p < 0.01, \text{ and } F_{1,3.18.9}=4.18, p < 0.025, \text{ respectively})\). These effects reflect the tendency for the old/new effects elicited by these words to be greater over the right than the left hemisphere, and at superior rather than inferior sites. For the negative items, the

---

**Fig. 4.** Differences in mean amplitude for old negative minus new negative, and old neutral minus new neutral, for the 500–800, 800–1100, and 1100–1400 ms, latency regions. Amplitude measures are averaged over the electrode site indicated and the sites immediately adjacent to it.
Fig. 5. Topographic maps illustrating the distribution of the differences between ERPs to correctly classified old and new neutral words (upper row), and between ERPs to correctly classified old and new negative words (lower row), in successive latency regions.
ANOVA revealed no effects involving the factor of response category.

### 3.3. Topographic analyses

Four sets of topographical analyses were performed. The first set compared the scalp distributions of the old/new effects (i.e. the differences in voltage between the ERPs elicited by correctly classified old and new words) for negative and neutral words between 300 and 500 ms. These distributions are illustrated in Fig. 5. ANOVA (factors of electrode site and valence) of these data revealed no evidence of a valence × site interaction ($F = 1$), indicating that the two scalp distributions are statistically equivalent.

A second analysis contrasted the topographies of the old/new effects with the effects of valence in the same latency range, that is, over the time interval in which the analyses of the mean amplitudes between 300 and 500 ms suggested that old/new and valence effects were additive. The old/new effects consisted of difference scores (old–new) collapsed over the factor of valence. Valence effects consisted of the difference scores (negative — neutral) collapsed over the factor of response category. The topographies of the two effects are illustrated in Fig. 6, where it can seen that each displays a mid-frontal maximum. Unsurprisingly, given their similarity, an ANOVA contrasting these topographies gave rise to a site effect ($F(3.6, 53.7) = 7.38, p < 0.001$), but not to a site × condition interaction ($F < 1$).

A third analysis compared the topographies of the old/new effects elicited by neutral and negative items in the 500–800 ms latency region (i.e. the only region post-500 ms in which effects were reliable for the negative words). The scalp distributions of these effects are illustrated in Fig. 5, where it can be seen that both show a strong left posterior focus. ANOVA of these data revealed a reliable site effect ($F(3.9, 58.7) = 4.40, p < 0.005$, but no evidence of a site × valence interaction ($F = 1.3$).

The final set of analyses compared the scalp distributions of the old/new effects elicited by the neutral words as a function of latency region, in order to confirm that the effects do indeed reflect the contributions of generator configurations that differ over time. The comparison was conducted across the three latency regions — 300–500, 500–800, and 1100–1400 ms — in which old/new effects appeared to show the most disparate distributions (see Fig. 5). ANOVA revealed a significant site by latency region interaction ($F(4.9, 73.3) = 4.06, p < 0.005$), indicating that the old/new effects for the neutral words do indeed differ with time. As shown in Fig. 5, these distributions initially evolve from a fronto-central to a left parietal maximum, and culminate in a right frontal focus.

### 3.4. Summary of ERP results

Reliable old/new effects were elicited by negative and neutral words in both the 300–500 and 500–800 ms latency regions. In the first of these regions, the effects had a fronto-central scalp distribution, whereas in the later region their topography showed a left parietal maximum. Neither the magnitude nor the scalp topography of the old/new effects differed according to

![Fig. 6. Topographic maps illustrating the distribution of the differences between correctly classified old and new items (left), and between neutral and negative items (right), in the 300-500 ms epoch.](image-url)
valence in the 300–500 ms region. In the 500–800 ms region, the scalp topography of the old/new effects elicited by the two classes of word was again equivalent, but the magnitude of the effects was greater for the neutral items. This difference was largely due to the enhanced positivity of the ERPs to new negative items, rather than an attenuation of the waveforms elicited by old negative words. From 800 ms onwards, old/new effects were reliable for the neutral items only, and evolved with time from a left posterior to a right frontal focus.

4. Discussion

4.1. Performance data

Recognition memory was poorer for negative than for neutral items. Crucially, this difference in recognition accuracy was entirely due to a marked difference in the rate of false alarms elicited by the two classes of word (34 vs 14% for negative and neutral items, respectively). These findings are broadly consistent with previous research investigating recognition memory for emotionally valenced words [5,20] and are well accounted for by the ‘semantic cohesiveness’ hypothesis outlined in the introduction. Specifically, it is proposed that the excessive false alarm rate for negative words reflects the same mechanisms that give rise to false recollection effects in studies investigating recognition memory for semantic associates of emotionally neutral study items [25].

The RT data add support to the foregoing hypothesis. Subjects were some 100 ms quicker to respond correctly to new neutral words than to new negative words. By contrast, false alarm responses to negative words were 125 ms faster than those to neutral words. This pattern of results presumably reflects the propensity of new negative items to elicit information which is either difficult (in the case of correct rejections) or impossible (in the case of false alarms) to discriminate from that associated with a veridical memory of the study episode.

4.2. ERP data

Between 300 and 500 ms post-stimulus ERPs were modulated by both the valence and the study status of the test items. Crucially, however, these two factors did not interact with respect to either the magnitude of their effects, or their scalp topographies. To the extent that Rugg et al. [31] are correct in their proposal that frontally distributed old/new effects in this latency range reflect familiarity-driven recognition rather than episodic retrieval (recollection), these findings suggest that valence has little effect on familiarity. This conclusion adds weight to the proposal (see below) that valence influences recognition memory primarily through its effects on recollection.

Valence exerted a marked effect on ERPs between 300 and 500 ms. Specifically, neutral items elicited a fronto-centrally distributed negative-going deflection, peaking around 400 ms, that was approximately 3 µV greater than the deflection elicited by negative words. One possibility is that this effect reflects the modulation of one or more of the generators of the ‘N400’, an ERP component well known for its sensitivity to semantic relatedness and association [3,12,18,32]. According to this hypothesis, the high level of inter-item association between members of the negatively valenced word set meant that these items tended to prime one another semantically, leading to an attenuation of N400 in a fashion similar to that seen in studies of semantic priming (e.g. [12,32]). To the extent that this account is correct, it underscores the differences that exist with respect to semantic ‘cohesiveness’ between otherwise unselected sets of emotionally negative and neutral words.

As is evident from Fig. 6, the scalp topographies of the old/new and valence effects in the 300–500 ms latency range were remarkably similar. While this finding does not rule out the possibility that there exist brain regions, undetectable by the ERP method, which respond differentially to the two variables, it suggests that valence and study status modulate a common generator population. According to this suggestion, therefore, these generators are sensitive both to ‘semantic’ (valence) and ‘episodic’ (old vs new) factors. A discussion of possible reasons for this observation can be found in Ref. [27].

4.3. Left parietal old/new effect

Neutral words elicited a prominent left parietally distributed old/new effect which extended from approximately 500 to 1200 ms post-stimulus. In the early part of the same latency range, negative words also elicited a reliable left parietal effect. While this latter effect was topographically indistinguishable from that elicited by the neutral items, it was both smaller in magnitude and considerably more short-lived. Crucially, the smaller magnitude of the left parietal effect elicited by negative items was carried (wholly, in the 500–800 ms latency range, and partially, in the later time region) by the greater positivity of the ERPs elicited by new negative words relative to those elicited by new neutral words. For the reasons outlined in the introduction, we interpret this pattern of results as a reflection of the capacity of new negative items to elicit spurious or ‘illusory’ episodic memories. Presumably, the fact that this effect was observed in the ERPs elicited by items which were correctly (albeit relatively
Toring operations were engaged whenever a test item during the processing of associative lures in a PET for their observation of right prefrontal activation. Schacter et al. [35] advanced this framework to account for veridical episode from the study phase (a similar argument was advanced by Schacter et al. [35] to account for the observation of right prefrontal activation during the processing of associative lures in a PET study of false recollection). Thus, post-retrieval monitoring operations were engaged whenever a test item elicits episodic information from memory. The only class of items in the present experiment for which such operations were not necessary, therefore, were the new, neutral items. Hence, neutral items, but not negative ones, elicited a right frontal effect.

5. Concluding comments

To summarise, we found that ERP old/new effects differed according to the emotional valence of the words that elicited them. We further found, however, that these differences involved modulations of a common set of memory-related effects, which, along with the behavioural findings, could be understood on the assumption that negatively valenced words share higher levels of inter-item associations than do words of neutral valence (the semantic cohesion hypothesis of Phelps et al. [22]). There was no evidence that recognition memory for the two different word classes engaged qualitatively distinct neural systems. With the caveat that accompanies any conclusion based on a null result, the findings suggest that emotionally negative words engage memory retrieval operations that are neither functionally nor neurally distinct from those that support the retrieval of emotionally neutral items.

Acknowledgements

EJM is supported by the UK Medical Research Council. MDR and KA are supported by the Wellcome Trust. This study was conducted while the authors were in the School of Psychology, University of St Andrews.

References


