Dissociation of the neural correlates of explicit and implicit memory

Michael D. Rugg, Ruth E. Mark, Peter Walla,
Astrid M. Schloerscheidt, Claire S. Birch & Kevin Allan

Wellcome Brain Research Group, School of Psychology, University of St Andrews, St Andrews, Fife KY16 9JU, UK
* Department of Neurology, University of Vienna, A-1180 Vienna, Austria

One presentation of a word to a subject is enough to change the way in which the word is processed subsequently, even when there is no conscious (explicit) memory of the original presentation. This phenomenon is known as implicit memory1–3. The neural correlates of implicit memory have been studied previously4–11 but they have never been compared with the correlates of explicit memory while holding task conditions constant or while using a procedure that ensured that the neural correlates were not ‘contaminated’ by explicit memory. Here we use scalp-recorded event-related brain potentials to identify neural activity associated with implicit and explicit memory during the performance of a recognition memory task. Relative to new words, recently studied words produced activity in three neuroanatomically and functionally dissociable neural populations. One of these populations was activated whether or not the word was consciously recognized, and its activity therefore represents a neural correlate of implicit memory. Then, when task and memory contamination effects are eliminated, the neural correlates of explicit and implicit memory differ qualitatively.

Memory encoding was manipulated by cueing subjects to perform either a ‘shallow’ or a ‘deep’ study task. Depth of processing affects the ability of a subject to recollect a study episode consciously, but has little influence on measures of ‘data-driven’ implicit memory, such as repetition priming2. In the first two experiments, studied (old) and unstudied (new) words were presented in a recognition memory test, and the event-related brain potentials (ERPs) produced by the different classes of test word were recorded. Crucially, we compared the ERPs produced by new words with those produced by old words that were misclassified by the subjects as new words, and thus differences between these two classes of ERP would reflect memory in the absence of awareness.

Very similar results were obtained in each experiment and we report the data collapsed across the two studies (n = 30). Of the

Figure 1 ERPs waveforms, averaged across subjects, from left and right frontal (F3, F4 of the 10–20 system12) and parietal (P3, P4) electrodes. a, ERPs elicited by correctly classified new words, and by unrecognized (unrec) and recognized (rec) shallowly studied words. ANOVA (factors of frontal versus parietal location, hemisphere and word type) of mean amplitudes between 300 and 500 ms revealed a location × word-type interaction is a standard statistical term, no explanation needed (F2, 53 = 5.97, P < 0.005). Contrasts between unrecognized and new words restricted to the parietal electrodes revealed effects that were reliable for each experiment (1 and 2) separately (F1, 14 = 5.16 and 6.52, both P values <0.005) as well as for the combined data set (F1,28 = 11.45, P < 0.005). b, ERPs elicited by correctly classified new words, and by recognized shallowly and deeply studied words. ANOVA of 300–500 ms amplitudes revealed an effect of word type (F2, 56 = 18.19, P < 0.001). There was no such effect for analysis of the data from the two types of old word. ANOVA of the 500–800 ms region revealed a significant word-type × hemisphere interaction (F2, 46 = 10.82, P < 0.001). This interaction remained when the ANOVA was restricted to data from the two types of old word (F1, 28 = 28.84, P < 0.001).
deeply studied words, 94% were recognized, as compared with 49% of the shallowly studied items ($t_{29} = 26.27, P < 0.001$). 86% of the new words were correctly classified as new.

ERPs produced by shallowly studied words are shown in Fig. 1a, and are separated according to the accuracy of recognition judgement. The scalp topographies of the respective memory effects (differences between ERPs produced in response to studied and new words) are illustrated in Fig. 2a, b. From roughly 300 to 500 ms after the onset of the stimulus, ERPs from frontal electrode sites were more positive for recognized items than they were either for new words or for old words misclassified as new. During the same latency range (300–500 ms post-stimulus), ERPs from parietal electrodes showed a different pattern: regardless of the accuracy of the recognition judgement, old words produced more positive-going waves than did new words. This effect, which was equivalent in size for recognized and unrecognized items, is a neural correlate of memory in the absence of conscious recognition. As shown in Fig. 3, the effect was insensitive not only to accuracy of recognition judgement, but also to depth of processing. These functional properties identify the effect as a correlate of implicit memory.

In Fig. 1b we contrast the ERPs produced by new items with those elicited by recognized words. Between approximately 300 and 500 ms after the onset of the stimulus, memory effects for the two classes of studied word were equivalent in magnitude (Fig. 3) and scalp topography (Fig. 2a, c). From ~500 ms onwards, the ERPs to deeply studied words differed from those to both new and shallowly studied items. The scalp topography of this effect differed reliably from the memory effect elicited by the same words in the earlier, 300–500 ms, latency region (Fig. 2c, d), indicating that the two effects reflect the activity of distinct neural populations.

Our findings show that old words in a recognition memory test can produce three different patterns of memory-related activity. The functional characteristics of two of the patterns correspond closely to the characteristics of two kinds of memory—implicit memory and conscious recollection—that have been distinguished on the basis of behavioural and neuropsychological evidence.

Indeed, the ERP correlate of recollection identified here closely resembles that identified in previous studies. The third pattern of activity, maximal over the frontal scalp from 300–500 ms after the onset of the stimulus, is less easily characterized. This was found only for recognized old items, but was insensitive to depth of study processing. It may reflect item 'familiarity', a form of explicit memory held to be dissociable from recollection.

Our results go beyond previous findings by showing directly that neural activity elicited by recently experienced words that are not consciously recognized differs from activity elicited by genuinely new words. An important question is whether this effect merely represents weak explicit memory that is sufficient to be manifest in ERPs but too weak to lead to a positive recognition judgement. This possibility can be rejected on two grounds. First, the magnitude of the proposed neural correlate of implicit memory did not vary between recognized and unrecognized words, whereas a correlate of explicit memory should vary with recognition accuracy. Second, unrecognized and recognized words produced qualitatively different patterns of neural activity, indicating that explicit memory may involve the engagement of neural populations that are separate from those supporting memory without awareness.

If the ERP differences produced in response to new and unrecognized words are a correlate of implicit memory, similar effects should be seen in the kind of task standardly used to study implicit memory. In two other experiments, we investigated the ERPs produced by old and new words in a semantic judgement task, where it was irrelevant whether the words were old or new. A reliable repetition priming effect was observed on reaction time (old words, 891 ms; new words, 907 ms; $F_{1,30} = 7.19, P < 0.02$). This effect did not vary with depth of study processing. A positive-going memory effect was seen in ERPs at ~300–600 ms after stimulus onset (Figs 3 and 4).

**Figure 2** Scalp distributions of the differences between the ERPs to new words and different classes of studied word. 

(a) Recognized shallowly studied minus new, 300–600 ms post-stimulus. 
(b) Unrecognized shallowly studied minus new, 300–600 ms post-stimulus. 
(c) Recognized deeply studied minus new, 300–500 ms post-stimulus. 
(d) Recognized deeply studied minus new, 500–800 ms post-stimulus.

ANOVA revealed that the contrasts between a and b, and between c and d, were significant (a vs b: site × word type interaction $F_{4,115} = 11.5, P < 0.005$; c vs d: site × latency region interaction $F_{4,115} = 4.14, P < 0.005$). The contrast between a and c was not significant.
course, and insensitivity to depth of processing, the effect resembles that elicited by unrecognized old words in the first two experiments. Thus, the neural correlate of implicit memory identified in those experiments is also present when test items are not intentionally used as retrieval cues.

Our findings are the first demonstration that the neural correlates of implicit and explicit memory can be dissociated within a single task. They provide strong support for the view\textsuperscript{21,22} that these two forms of memory reflect the operation of qualitatively distinct neural systems.

**Methods**

**Recognition task.** Subjects (n = 16 and 18 in experiments 1 and 2, respectively) were healthy, young, right-handed adults, naive to the purpose of the experiment. In each experiment, 15 subjects provided enough data (>16 artefact-free trials) to form ERPs for both correctly classified items and shallowly studied items misclassified as new.

Two (experiment 1) or three (experiment 2) study test blocks were used. In each study phase, 68 critical and 4 filler words were presented one at a time on a TV monitor. Each study word was preceded by a cue which indicated the nature of the task to be performed on that word. In response to one cue, subjects determined whether the first and last letters of the word were in alphabetic order (shallow task). In response to the other cue, they incorporated the word into a short sentence (deep task). After ~5 min, the test task was administered. The 68 critical words from the study phase were presented, randomly intermixed with 34 critical unstudied words and 10 unstudied fillers. The words (presentation duration 300 ms, maximum visual angle 1.5 × 0.4 degrees (experiment 1); or 500 ms, 2.8 × 0.5 degrees (experiment 2)) were presented every 5.3 s (experiment 1) or 5.5 s (experiment 2). Each word was preceded for 2.1 s by a fixation character, which was erased 102 ms before the word's onset. Subjects discriminated between old and new items as quickly and as accurately as possible by pressing one of two response keys. The study and test lists were constructed so that, across subjects, each critical item served equally often as a word that was new, deeply studied, and shallowly studied. The words ranged

---

**Figure 3** Magnitudes of ERP memory effects. a, b, Mean (+s.e.m.) differences in the amplitude of the 300–500 (a) and 500–800 ms (b) latency regions of ERPs to new words and ERPs to recognized deeply studied words and recognized and unrecognized shallowly studied words, shown for left (F3, P3) and right (F4, P4) frontal and parietal electrodes. c, Mean (+s.e.m.) of the differences in the semantic judgement task between the amplitude of the 300–500-ms latency region of ERPs to new words and ERPs to deeply and shallowly studied old words. Double asterisks indicate that these results differ from zero with $P < 0.005$; single asterisks indicate that these results differ from zero with $P < 0.05$.

---

**Figure 4** The semantic judgement task. a, ERPs from frontal (F3, F4) and parietal (P3, P4) electrodes in the semantic judgement task. ANOVA of the 300–500-ms latency region revealed a location × word-type interaction ($F_{2,58} = 5.51, P < 0.01$). Contrasts between word types restricted to the parietal electrodes revealed effects that were reliable for both experiments (3 and 4) separately ($F_{2,28} = 4.17, F_{2,24} = 3.59$, values $P < 0.05$) and for the combined data set ($F_{2,60} = 7.73, P < 0.001$). b, Scalp topography of the ERP differences between old (collapsed over study conditions) and new words.

---

Two (experiment 1) or three (experiment 2) study test blocks were used. In each study phase, 68 critical and 4 filler words were presented one at a time on a TV monitor. Each study word was preceded by a cue which indicated the nature of the task to be performed on that word. In response to one cue, subjects determined whether the first and last letters of the word were in alphabetic order (shallow task). In response to the other cue, they incorporated the word into a short sentence (deep task). After ~5 min, the test task was administered. The 68 critical words from the study phase were presented, randomly intermixed with 34 critical unstudied words and 10 unstudied fillers. The words (presentation duration 300 ms, maximum visual angle 1.5 × 0.4 degrees (experiment 1); or 500 ms, 2.8 × 0.5 degrees (experiment 2)) were presented every 5.3 s (experiment 1) or 5.5 s (experiment 2). Each word was preceded for 2.1 s by a fixation character, which was erased 102 ms before the word's onset. Subjects discriminated between old and new items as quickly and as accurately as possible by pressing one of two response keys. The study and test lists were constructed so that, across subjects, each critical item served equally often as a word that was new, deeply studied, and shallowly studied. The words ranged
from low to medium frequencies of occurrence, and from four to nine letters in length.

ERPs (sampling rate 6 ms per point, epoch length 1,536 ms, prestimulus baseline 102 ms) produced in response to different classes of test word were obtained from 25 scalp sites as described. ERPs were quantified by measuring the mean amplitude of two latency regions, 300–500 ms and 500–800 ms after the onset of stimulus. Amplitude differences were assessed by analysis of variance (ANOVA) (degrees of freedom (d.f.) corrected for non-sphericity). Differences in scalp topography were assessed by d.f.-corrected ANOVA of the data from all 25 electrodes after rescaling.

Semantic judgement task. Two groups of young adults (n = 16) were used in each experiment. For experiment 3, the subjects were the same individuals as those used in experiment 1. A new sample was used in experiment 4.

Experimental items were drawn from the same word pool used to construct the lists for experiments 1 and 2. For experiment 3, two study test blocks identical in structure to those used in experiment 1, although using different items, were administered. The only difference in procedure from experiment 1 was that subjects were required during the test to classify each test word as animate or inanimate. The procedure for experiment 4 was identical to that for experiment 3 except that the depth of processing manipulation at study was blocked. Stimulus display parameters and electroencephalogram recording for these experiments were as for experiment 1. As the differences between old and new words in the two experiments were very similar, and there were no block effects in experiment 4, the data are reported collapsed across experiments.

Received 1 October 1997; accepted 25 February 1998.


Acknowledgements. This research was supported by the Wellcome Trust. P.W. was supported by the Austrian Research Foundation, and K.A. by the Biotechnology and Biological Sciences Research Council, UK.

Correspondence and requests for materials should be addressed to M.D.R. (e-mail: mdr@st-andrews.ac.uk).