Research Report

Indexing strategic retrieval of colour information with event-related potentials

E.L. Wilding*, C.S. Fraser, J.E. Herron

School of Psychology, Cardiff University, Cardiff, CF10 3AT Wales, UK

Accepted 9 April 2005
Available online 31 May 2005

Abstract

Event-related potentials (ERPs) were acquired during two experiments in order to determine boundary conditions for when recollection of colour information can be controlled strategically. In initial encoding phases, participants saw an equal number of words presented in red or green. In subsequent retrieval phases, all words were shown in white. Participants were asked to endorse old words that had been shown at encoding in one colour (targets), and to reject new test words as well as old words shown in the alternate colour (non-targets). Study and test lists were longer in Experiment 1, and as a result, the accuracy of memory judgments was superior in Experiment 2. The left-parietal ERP old/new effect—the electrophysiological signature of recollection—was reliable for targets in both experiments, and reliable for non-targets in Experiment 1 only. These findings are consistent with the view that participants were able to restrict recollection to targets in Experiment 2, while recollecting information about targets as well as non-targets in Experiment 1. The fact that this selective strategy was implemented in Experiment 2 despite the close correspondence between the kinds of information associated with targets and non-targets indicates that participants were able to exert considerable control over the conditions under which recollection of task-relevant information occurred.

© 2005 Elsevier B.V. All rights reserved.

Theme: Neural basis of behaviour
Topic: Learning and memory systems and functions
Keywords: Event-related potential; Colour information; Episodic memory; Recollection; Exclusion task

1. Introduction

Studies of retrieval from episodic memory in which event-related potentials (ERPs) were recorded have led to the identification of several modulations of scalp-recorded neural activity that index distinct classes of retrieval process (for reviews, see [16,43,53]). The most common contrast in ERP studies of memory retrieval is between the electrical activity that is evoked by old (previously studied) and new items to which correct memory judgments have been made. The differences revealed by these contrasts will be referred to here as ERP old/new effects.

There is a family of old/new effects, and the individual effects are distinguishable on the basis of their time courses, scalp distributions, and sensitivity to experimental variables [11]. The old/new effect to which probably the most attention has been paid is largest at parietal electrode locations, typically larger over the left than the right hemisphere, and comprises a relatively greater positivity for old than for new test items to which correct judgments have been made [37]. This left-parietal ERP old/new effect onsets approximately 500 ms post-stimulus, and commonly lasts between 200 and 700 ms, the duration being influenced by task demands [10].

It is now widely accepted that the left-parietal ERP old/new effect indexes recollection (for recent discussion, see [5,15]). Recollection is defined here as successful recovery of contextual information about a prior event [56], which might include one or more of information sources such as where or when the event occurred, the modality through
which information was perceived, and/or what actions were completed during the event [28].

The data that support the link between the parietal effect and recollection include the following: (1) patients with deficits in recollection (as revealed by behavioural measures) exhibit attenuated or absent left-parietal ERP old/new effects relative to appropriately matched controls [13,39,46,47]; (2) the magnitude of the effect is correlated positively with the amount of contextual information that is retrieved, as inferred from the numbers of correct source judgments that are made on retrieval tasks [50,51]; and (3) the effect is larger for correct old judgments associated with Remember (R) rather than with Know (K) judgments [12,45]. R judgments are held to be based upon recollection while K judgments are not [13].

While this brief summary provides arguably a compelling case for a link between the parietal old/new effect and recollection, some recent findings, at least at a first pass, jar with this account. These findings have come from exclusion tasks [23,24], in which ERPs were acquired time-locked to test stimuli. In the study phases of exclusion tasks, items are typically associated with one of two contexts. In subsequent retrieval phases, items associated with one context are designated as targets, while items from the other are designated as non-targets. Participants are asked to respond on one key to targets, and on another to non-targets as well as to genuinely new test items. The exclusion task was introduced as one component of the process-dissociation procedure (PDP; [23–25]), and while the details of that procedure are not important here, the critical point for present purposes is that accurate target/non-target judgments have been assumed to be based upon recollection of information associated with targets and with non-targets [23].

If this assumption is correct, then correct target as well as non-target judgments will be accompanied by reliable left-parietal ERP old/new effects. In one recent study, however, Herron and Rugg [20] reported a somewhat different pattern of effects. They acquired ERPs during the retrieval phases of two exclusion tasks. For both experiments, the first encoding phase involved generating sentences incorporating single words that were presented at study, while the second encoding phase differed across the two experiments. In the first experiment, participants rated words for pleasantness, while in the second, participants were asked simply to read words aloud. Targets were designated as words that were rated for pleasantness in Experiment 1, and words that were read aloud in Experiment 2. Thus, the non-targets were associated with the same encoding task in both experiments, and while reliable parietal ERP old/new effects were obtained for targets in both experiments, for non-targets, a reliable left-parietal effect was revealed in Experiment 2 only.

This finding is striking, because in both experiments, the information associated with non-targets is likely to be highly similar, since they were encoded using the same task (sentence generation), and in keeping with this observation, the accuracy of non-target judgments was equivalent in the two experiments. To the extent that participants were equally likely to have recollected non-targets in both experiments, therefore, the absence of a left-parietal ERP old/new effect in Experiment 1 might be seen to call into question the link between the parietal ERP old/new effect and recollection (see also [21]).

Whether these results challenge current views of the functional significance of the left-parietal ERP old/new effect, however, depends upon the assumption that participants did in fact recollect information about non-targets equally often in the two experiments. Herron and Rugg [20] questioned this assumption, noting that recollection of non-targets is not always necessary for successful completion of the exclusion task. The reason for this is that the failure to recollect information that would identify a test item as a target is one basis upon which a correct response to non-targets and new words can be made. Herron and Rugg [20] suggested that implementation of this strategy was the reason for the absence of the non-target old/new effects in Experiment 1: participants were successful in restricting recollection to information diagnostic of the target status of an item and did not recollect information about non-targets.

In extending this account to explain why non-targets were associated with a reliable left-parietal ERP old/new effect in Experiment 2, Herron and Rugg noted that the reliability of the strategy described above decreases along with the likelihood of recollecting information about targets. They noted that target judgments in Experiment 2 were less accurate than in Experiment 1 and proposed that because of this, participants attempted to recollect information about both targets and non-targets in order to ensure successful completion of the task.

There are several implications of the findings of Herron and Rugg, one of which is the way in which these data impact on assumptions concerning the use of the exclusion task and the process-dissociation procedure [20,24,57,58]. A second implication, and the principal focus here, concerns the boundary conditions for when a strategy of recollecting information associated with targets only can be implemented. The experiments that are presented here are concerned with these boundary conditions, and are motivated by findings suggesting that under some circumstances, even when the accuracy of target judgments is relatively high, recollection of information associated with non-targets occurs. The critical data come from studies in which the target/non-target distinction was based upon study voice (male or female; [52,54]) or the colour in which pictures were presented at study (red or green; [7–9]). In the case of both of these manipulations, and in contrast to the findings of Herron and Rugg described previously, the researchers have reported reliable non-target ERP old/new effects (also see [19]).

One way to reconcile these findings is to assume that the precision with which recollection can be controlled depends upon the relationship between the information that either is or is not to be recollected. That is, in the case of the voice
and colour studies described above, one explanation for the presence of the non-target ERP old/new effects is that the high degree of similarity between the target and non-target information precluded the restriction of recollection to information associated with targets only. In Experiment 1 of Herron and Rugg [20], meanwhile, the cognitive operations underlying sentence generation and ratings for pleasantness were sufficiently distinct to permit recollection to be restricted to the latter only.

While this explanation for the colour and voice data is reasonable, the majority of the foregoing inferences about the conditions under which recollection of non-targets occurs in the colour and voice studies are based only on informal observations of the correspondences between electrophysiological as well as behavioural data across studies. In the case of the colour and voice data, moreover, the published non-target old/new effects have, at least on the basis of visual inspection, been of smaller magnitude than the target old/new effects. The reason that this is important is that this pattern of findings raises the possibility that larger non-target old/new effects (relative to target old/new effects) might be obtained as target accuracy—hence, the likelihood of target recollection—decreases. This pattern of findings would be consistent with the view that at least partial control of recollection of non-target material can be exerted even when the information associated with targets and with non-targets corresponds closely. There are, however, no studies to date in which the outcomes of the appropriate contrasts—those between target and non-target old/new effects for the same kinds of perceptual information and separated according to levels of target accuracy—have been reported.

The two experiments presented here were designed in order to address this disparity, and in both experiments, the colour of visually presented words was the contextual information that was designated as being important for the target/non-target distinction in subsequent retrieval phases. The two experiments differed only in the number of study—test cycles and the number of words per cycle. Experiment 1 contained fewer study—test blocks and more study and test words per block than Experiment 2. This difference was introduced in order to encourage more accurate performance in Experiment 2, thereby permitting an analysis of the left-parietal ERP old/new effects for the same forms of information under conditions in which target accuracy differed.

In addition, the lengths of study and test lists in the two experiments were set at a level that ensured the accuracy of target judgments was higher (Experiment 2) and lower (Experiment 1) than in the majority of previous ERP studies in which voice or colour has been the source of information on which target/non-target judgments were to be made. Non-target left parietal ERP old/new effects of smaller amplitude in Experiment 2 than Experiment 1 (relative to the relevant target effects in each case) would suggest that participants can exert some degree of control over recollection of colour information even under conditions where target recollection also depends on accessing colour information. Equivalent non-target ERP old/new effects across the two experiments would provide important information concerning limits on the resolution with which control over recollection can be exerted.

2. Methods

Due to the similarity of the two experiments, they are described jointly.

2.1. Participants

Twenty participants (12 female, mean age = 21 years) took part in Experiment 1. The data from two participants (one male) were discarded due to insufficient trials in at least one response category of interest because of excessive EOG artefacts. Eighteen participants (eight female, mean age = 22 years) took part in Experiment 2. No participant took part in both experiments. All participants had normal or corrected to normal vision and were right-handed, as assessed by writing-hand. No participants reported suffering from red/green colour blindness, all were native English speakers, and each gave informed consent before completing the task. The data for Experiment 1 were collected prior to the data for Experiment 2.

2.2. Materials

The stimuli comprised low-frequency words (MRC psycholinguistic database: frequency 1–9/million); 240 words were used in Experiment 1 and 180 words were used in Experiment 2. All words were open-class and ranged between four and nine letters in length. Words were presented visually on a monitor 1.2 m from the participant, in red or green text in the study phases and white text in the test phases. A black background was used during all study and test phases. The stimuli subtended maximum visual angles of 0.5° (vertical) and 2.2° (horizontal).

2.3. Design and procedure

Participants were fitted with an electrode cap before the experiment (see below) and were seated in a sound-attenuated room facing a monitor with their thumbs resting on two response keys. There was a short practice session before the experiment and short breaks between each study—test cycle. There were 8 study—test cycles in Experiment 1, and 12 cycles in Experiment 2. In Experiment 1, the 240 words were divided randomly into 8 equal blocks, creating 8 study—test cycles each containing 20 study words and 30 test words (an equal number of old red, old green and new words). In Experiment 2, the 180 words were divided randomly into 12 equal blocks creating 12 study—
test cycles, each containing 10 study words and 15 test words (again, an equal number of old red, old green, and new words). In both experiments, the words in each block were rotated so that, across participants, each word was encountered as an old red word, an old green word, and a new word. This manipulation resulted in the creation of three complete task-lists in each experiment, and an equal number of participants completed each list. The order of word presentation at both study and test within each study–test cycle of each list was determined randomly for each participant.

Prior to commencing the experiments, participants were informed that, in each study cycle, they would see some words presented in red and some in green, and that they were to make key presses with either their right or left hand based upon the colour in which the words were presented. They were instructed that, in each test phase, words presented in one colour at study (targets) would require a key press with one hand, while a key press with the other hand would be required for words presented in the other colour (non-targets) as well as for words that were new to the experiment. Participants were also informed that target designation would only be revealed at the start of each test phase, and that no words would appear in more than one study–test cycle.

In both experiments, each study trial began with an asterisk (*), which was displayed for 500 ms and followed by a blank screen (200 ms), after which the study word was presented for 300 ms. Half of the study words were presented in red/green in a random order. Participants indicated the study–test cycle at test. An interval of approximately 1 min intervened between each study and test phase, and that no words would appear in more than one study–test cycle.

Each test trial also began with an asterisk (500 ms duration), and 200 ms then intervened before the presentation of a test word. Each word was visible for 300 ms and then the screen remained blank until 500 ms after the response of the participant, at which time the asterisk signalling the start of the next trial appeared. Participants were asked to make a target or non-target/new judgement via key press when each test word appeared. For each participant, the first half of the study–test cycles (4 in Experiment 1 and 6 in Experiment 2) had the same target designation (red or green) and the remainder had the alternate designation. Blocking of designation was employed in order to reduce the number of times in which the response requirements changed for participants during the experiments. The order of target designation was balanced across participants, as were the hands required for the binary red/green judgment at study and the binary target/non-target/new judgment at test. Participants were encouraged to respond while balancing speed and accuracy equally. Test responses faster than 650 ms were treated as errors.

2.4. Electrophysiological recording procedure

EEG was recorded from 25 silver/silver chloride electrodes housed in an elastic cap. They were located at midline (Fz, Cz, Pz) and left/right hemisphere locations (FP1/FP2, F7/F8, F5/F6, F3/F4, T3/T4, C5/C6, C3/C4, T5/T6, P5/P6, P3/P4, O1/O2). Additional electrodes were placed on the mastoid processes, and in Experiment 2, a further electrode was placed on the nose-tip. EEG was recorded continuously at 166 Hz (6 ms per point) with Fz as the reference electrode, and was re-referenced computationally off-line to a linked mastoid reference into epochs of 1536 ms (256 data points), each including a 102-ms pre-stimulus baseline relative to which all post-stimulus amplitudes were measured. EOG was recorded from above and below the right eye (VEOG) and from the outer canthi (HEOG). Trials containing large EOG artefact were rejected, as were trials containing A/D saturation or baseline drift exceeding ±80 μV. Other EOG blink artefacts were corrected using a linear regression estimate [44].

Averaged ERPs were formed for correct judgments at test to target, non-target, and new words for each participant in each experiment. Participants were excluded if they did not contribute at least 16 artefact-free trials to each response category [51,55]. The ERPs were collapsed across colour, on the basis of null results for this factor in preliminary analyses of both the behavioural and the electrophysiological data (in keeping with the findings of Cycowicz et al. [8]). On average, less than 20% of trials per participant per response category were rejected due to artefact. In Experiment 1, the mean numbers of trials per category were 39, 47, and 65 for correct responses to target, non-target, and new words, respectively. The equivalent values in Experiment 2 were 39, 40, and 50. The averaged ERPs for each participant and for each category of interest in each experiment were subjected to a 7-point binomially weighted smoothing filter prior to analysis.

3. Results

3.1. Behavioural data

The probabilities of correct responses to red and green words during the study phases of both experiments were high and statistically equivalent, as assessed by ANOVA\(^1\) with between and within factors of experiment and colour, respectively (Experiment 1: red = 0.98, green = 0.99; Experiment 2: red = 0.99, green = 0.99). In two further analyses, the reaction times for the colour judgments at study were analysed separated according to whether they

---

\(^1\) These and all subsequent analyses of variance incorporated the Geisser–Greenhouse correction for non-sphericity where necessary [18]. Corrected degrees of freedom are shown where appropriate.
were designated as targets or non-targets in the subsequent test phase(s). These analyses were restricted to the blocks where participants had completed three previous blocks with the same target designation (data from blocks 4 and 8 in Experiment 1 and data averaged across blocks 4–6 and 10–12 in Experiment 2). Separate analyses were conducted on the data from each experiment, including the factors of subsequent target designation and block (4 vs. 8 in Exp. 1, 4–6 vs. 10–12 in Exp. 2). No reliable effects were revealed in either analysis. The mean study RTs for targets and non-targets in Experiment 1 were 875 and 850 ms, respectively, for block 4, and 865 and 840 ms for block 8. In Experiment 2, the values were 1070 and 1101 ms (blocks 4–6) and 1143 and 1113 ms (blocks 10–12)\(^2\). The reason for these restricted block analyses was to determine whether there was any evidence to support the view that, as the task unfolded, participants treated study items differentially, which might impact upon the patterns of ERP data that were acquired in subsequent retrieval phases. We return to this issue in the Discussion.

Table 1 shows the mean probabilities of, and reaction times (RTs) for, correct responses to target, non-target, and new words in Experiments 1 and 2. Correct responses to targets and non-targets will be referred to as target hits and non-target hits, respectively. The term old words will refer to targets and non-targets jointly. In each experiment, the probabilities of a target hit were reliably greater than the probabilities of a target response to a non-target [Experiment 1: \(t(17) = 6.85, P < 0.001\); Experiment 2: \(t(17) = 14.89, P < 0.001\)] or a new word [Experiment 1: \(t(17) = 13.11, P < 0.001\); Experiment 2: \(t(17) = 31.10, P < 0.001\)].

The analysis of the accuracy data [factors of experiment (between) and response category (within)] revealed a main effect of experiment \([F(1,34) = 9.24, P < 0.01]\) and an interaction between this factor and category \([F(1,6,55.4) = 4.31, P < 0.05]\). Post hoc analyses (Newman–Keuls) revealed that only target accuracy was superior in Experiment 2 to Experiment 1. RT varied according to response category only \([F(1,8,62.4) = 38.40, P < 0.001]\), and post hoc analyses revealed that RTs for correct judgments to new words were reliably faster than those for targets, which were in turn reliably faster than those for non-targets.

### 3.2. Results: ERPs

The ERPs from a \(3 \times 3\) array of 9 electrode sites for Experiments 1 and 2 are shown in Fig. 1. The figure shows the ERPs associated with correct judgments to target, non-target, and new words. The fact that this selection of electrode locations permits inspection of the principal divergences between response categories is attested to by the data shown in Fig. 2, which comprises the scalp distributions of the ERP old/new effects for targets and for non-targets in the two experiments over 4 time windows: 300–500, 500–700, 700–900, and 900–1400 ms. These scalp maps are derived from the difference scores obtained by subtracting the ERPs evoked by new words from those evoked by targets and non-targets, respectively. Fig. 3 shows the same data as for the lower half of Fig. 1, with the exception that the data is referenced to nose-tip. We return to the importance of this figure in the Discussion.

#### 3.2.1. Analysis of left-parietal ERP old/new effects

The initial directed analysis of the ERPs was conducted in order to determine the relationships between the left-parietal ERP old/new effects that were obtained in the two experiments. This analysis was restricted to the P5 electrode location, and included the factors of experiment and response category (correct responses to target, non-target, and new test words). The analysis revealed an interaction between these two factors \([F(1,8,60.9) = 5.02, P < 0.025]\) and was followed up by separate analyses of the target and non-target ERP old/new effects in each experiment. These revealed that the ERPs evoked by old words were reliably more positive than those evoked by new words with the exception of the non-target vs. new contrast in Experiment 2 [Exp. 1: targets \(F(1,17) = 48.08, P < 0.001\), non-targets \(F(1,17) = 14.42, P < 0.01\); Exp. 2: targets \(F(1,17) = 24.40, P < 0.001\), non-targets \(F < 1\)]. Target old/new effects were also reliably larger than non-target effects in both experiments [Experiment 1: \(F(1,17) = 17.84, P < 0.01\); Experiment 2: \(F(1,17) = 32.39, P < 0.001\)].

#### 3.2.2. Global analyses

The directed analyses were followed by a series of analyses conducted using data from the 9 electrodes shown in Fig. 1 (F5, Fz, F6, C5, Cz, C6, P5, Pz, P6). The analyses incorporated the factors of location in the anterior/posterior (AP: anterior, central, posterior) and left–right (LR: left-hemisphere, midline, right-hemisphere) planes. They were conducted separately for each experiment in order to determine the correspondence between the ERP old/new effects in these experiments and those reported in previous studies, as well as to license subsequent between-experiment analyses [42,54]. They were conducted employing

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Target</th>
<th>Non-target</th>
<th>New</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RT</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experiment 1</td>
<td>1070 (279)</td>
<td>1101 (338)</td>
<td>886 (246)</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>977 (258)</td>
<td>1118 (338)</td>
<td>871 (279)</td>
</tr>
</tbody>
</table>

\(^2\) We are grateful to an anonymous referee for suggesting this specific analysis.
data from the 300–500, 500–700, 700–900, and 900–1400 ms time windows, selected in order to encompass the epochs in which ERP old/new effects have been reported previously (e.g., [54]), as well as to capture the differences between experiments that are evident in Figs. 1 and 2.

For each experiment, an initial analysis was conducted incorporating data from all three response categories as well as the factor of epoch, in addition to the site factors (AP and LR) described above. Both analyses revealed interactions between epoch (EP) and response category (RC), including in each case a four-way interaction between these two factors, AP and LR [Exp. 1: $F(7.4,126.0) = 5.08, P < 0.001$; Exp. 2: $F(6.0,101.9) = 2.20, P < 0.05$]. For each experiment, these analyses were followed up by paired contrasts for the data from successive epochs. All other factors were as described above, and the outcomes of these analyses (restricted to interactions involving epoch and response category) are shown in Table 2, where it can be seen that there are either reliable or marginal interactions between the relevant factors for each of the paired epoch contrasts.

The purpose of the analyses that included epoch as a factor (see Table 2) was to establish that the epochs employed in this analysis section are the ones that likely separate the ERP data into periods during which reliably different patterns of ERP old/new effects occurred. In order to establish the profile of the old/new effects within each epoch, these initial analyses were followed by separate analyses within each epoch. The analyses comprised all possible paired contrasts of the ERPs associated with correct judgments to targets, non-targets, and new test words (see Tables 3 and 4: main effects or interaction terms that do not include the factors of response category are not shown as they are not the outcomes of principal interest here). The following description of the outcomes of these analyses is restricted to the highest-order interactions that were obtained in each case.
3.2.3. Experiment 1 — see Table 3

For all epochs, the contrasts involving new words revealed interactions between response category, AP, and LR. For the 300–500 and 500–700 ms epochs, these interactions reflect the vertex maximum of the positive-going ERP old/new effects, and the left-lateralisation of the effects at posterior electrodes, as Fig. 2 illustrates. The same posterior lateralisation is evident in the 700–900 ms epoch, while at frontal locations over this time window, the small positive-going old/new effect is distributed relatively more evenly across the two hemispheres. In the 900–1400 ms epoch, the right-lateralised ERP old/new effects at anterior
locations are accompanied by a negative-going old/new effect with a posterior midline (Pz) maximum.

The direct contrast between targets and non-targets revealed that the ERPs evoked by targets are more positive-going from 500 to 900 ms, with this positivity being most pronounced at the midline for the 500–700 ms epoch, reflected in a RC × LR interaction. The RC × AP × LR interaction over the 900–1400 ms epoch reflects the fact that the greater relative positivity for targets is larger at midline and right than at left anterior locations, while at Pz, the ERPs evoked by non-targets are relatively more positive-going.

3.2.4. Experiment 2 — see Table 4

For the target vs. new contrast, the RC × LR interaction over the 900–1400 ms epoch reflects the fact that the greater positivity for targets is smaller at anterior than at central and posterior locations, and larger at the midline than at left and right hemisphere locations. For the target vs. non-target contrast, the RC × AP × LR interaction for the 300–500 ms epoch, the same three-way interaction reflects the vertex maximum of the old/new effect. For the non-target versus new contrast, there were no reliable ERP old/new effects between 500 and 900 ms. The analyses in the early (300–500 ms) epoch revealed a main effect of response category only—despite the apparent anterior maximum of the ERP old/new effect—and reflects the relatively greater positivity for non-targets. For the 900–1400 ms epoch, the RC × AP × LR interaction revealed in the non-target vs. new contrast reflects primarily the fact that the differences between the ERPs are restricted to midline central and posterior scalp locations, where the ERPs evoked by new test words are more positive-going than those evoked by non-targets.

3.2.4. Experiment 2 — see Table 4

For the target vs. new contrast, the interactions between RC, AP, and LR from 500 to 900 ms and the marginal interaction in the 900–1400 ms epoch come about for similar reasons to those described above for Experiment 1 for the contrasts involving new words. In the 300–500 ms

Table 2

Results of the paired epoch contrasts between the ERPs evoked by correct judgments to target, non-target, and new words in Experiments 1 and 2

<table>
<thead>
<tr>
<th>Effect</th>
<th>df</th>
<th>300–500 vs. 500–700</th>
<th>500–700 vs. 700–900</th>
<th>700–900 vs. 900–1400</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>ɛ</td>
<td>F</td>
</tr>
<tr>
<td>Experiment 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EP × RC</td>
<td>2,34</td>
<td>7.79**</td>
<td>0.89</td>
<td>13.75***</td>
</tr>
<tr>
<td>EP × RC × LR</td>
<td>4,68</td>
<td>3.24*</td>
<td>0.72</td>
<td>12.27***</td>
</tr>
<tr>
<td>EP × RC × AP</td>
<td>4,68</td>
<td>3.53*</td>
<td>0.54</td>
<td>3.56*</td>
</tr>
<tr>
<td>EP × RC × AP × LR</td>
<td>8,136</td>
<td>5.42***</td>
<td>0.45</td>
<td>7.65***</td>
</tr>
<tr>
<td>Experiment 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EP × RC</td>
<td>2,34</td>
<td>ns</td>
<td></td>
<td>11.11***</td>
</tr>
<tr>
<td>EP × RC × LR</td>
<td>4,68</td>
<td>ns</td>
<td></td>
<td>11.68***</td>
</tr>
<tr>
<td>EP × RC × AP</td>
<td>4,68</td>
<td>ns</td>
<td></td>
<td>ns</td>
</tr>
<tr>
<td>EP × RC × AP × LR</td>
<td>8,136</td>
<td>2.58* (P = 0.05)</td>
<td>0.45</td>
<td>ns</td>
</tr>
</tbody>
</table>

EP = epoch, RC = response category, AP = anterior/central/posterior, LR = left/midline/right. *P < 0.1, **P < 0.05, ***P < 0.01, ****P < 0.001, ɛ = epsilon value.

Table 3

Results of the paired contrasts between ERPs evoked by correct judgments to target, non-target, and new words in Experiment 1 over the 300–500, 500–700, 700–900, and 900–1400 ms epochs

<table>
<thead>
<tr>
<th>Effect</th>
<th>df</th>
<th>300–500</th>
<th>500–700</th>
<th>700–900</th>
<th>900–1400</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>F</td>
<td>ɛ</td>
<td>F</td>
<td>ɛ</td>
</tr>
<tr>
<td>Target vs. new</td>
<td>1,17</td>
<td>35.93***</td>
<td></td>
<td>42.26***</td>
<td></td>
</tr>
<tr>
<td>RC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC × AP</td>
<td>2,34</td>
<td>ns</td>
<td>35.93***</td>
<td></td>
<td>42.26***</td>
</tr>
<tr>
<td>RC × LR</td>
<td>2,34</td>
<td>4.78*</td>
<td>0.88</td>
<td>9.92**</td>
<td>0.89</td>
</tr>
<tr>
<td>RC × AP × LR</td>
<td>4,68</td>
<td>4.26*</td>
<td>0.69</td>
<td>9.48***</td>
<td>0.72</td>
</tr>
<tr>
<td>Non-target vs.</td>
<td>1,17</td>
<td>31.15***</td>
<td></td>
<td>10.49**</td>
<td></td>
</tr>
<tr>
<td>new</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC × AP</td>
<td>2,34</td>
<td>ns</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC × LR</td>
<td>2,34</td>
<td>8.85**</td>
<td>0.86</td>
<td>5.23*</td>
<td>0.97</td>
</tr>
<tr>
<td>RC × AP × LR</td>
<td>4,68</td>
<td>4.36*</td>
<td>0.57</td>
<td>10.64***</td>
<td>0.75</td>
</tr>
<tr>
<td>Target vs. non-target</td>
<td>1,17</td>
<td>ns</td>
<td></td>
<td></td>
<td>23.68***</td>
</tr>
<tr>
<td>RC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC × AP</td>
<td>2,34</td>
<td>ns</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC × LR</td>
<td>2,34</td>
<td>3.86*</td>
<td>0.92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RC × AP × LR</td>
<td>4,68</td>
<td>ns</td>
<td></td>
<td></td>
<td>2.47*</td>
</tr>
</tbody>
</table>

Abbreviations and meanings of symbols are as for Table 2.
the 500–700 ms epoch, the RC × AP × LR interaction for the contrast between the two classes of old word is due to reasons similar to those responsible for the same interaction term that was revealed in the target vs. new contrast: the fact that the greater relative positivity for targets is most pronounced at left hemisphere central and posterior scalp locations.

3.2.5. Between-experiment analyses

These comprised direct contrasts of the ERP old/new effects from the two experiments, and were completed on mean amplitudes computed from subtraction waveforms obtained by subtracting the ERPs evoked by correct responses to new words from those evoked by target and non-target hits. The initial analyses were completed separately for targets and for non-targets, and were restricted to targets for the 500–700 and 700–900 ms epochs as the within-experiment analyses described above revealed no reliable old/new effects for non-targets over these epochs in Experiment 2. The analyses each included factors of experiment, AP and LR, using the same epochs and electrode locations described above. Where these analyses revealed interactions between experiment and scalp locations, they were followed up by analyses of rescaled data in order to determine whether the interactions came about because of differences between the scalp distributions of the ERP old/new effects, or simply because of amplitude differences [33]. The data were rescaled using the vector length method [33], and the rescaling was computed over all 25 sites from which data were collected, although the analyses included the same factors and electrode locations that were employed for the data that were analysed prior to rescaling.

For the 300–500 ms epoch, the only reliable effect involving experiment was an interaction with the LR dimension for targets [F(1.6,54.2) = 3.83, P < 0.05], reflecting the fact that the old/new effects in Experiment 1 are larger than those from Experiment 2 at left hemisphere locations but smaller at the midline. This interaction was not reliable for the analysis of rescaled data. For the 500–700 and 700–900 ms epochs, the target old/new effects in Experiment 1 were larger than those from Experiment 2 [500–700: F(1,34) = 5.76, P < 0.05; 700–900: F(1,34) = 8.66, P < 0.01]. For the 900–1400 ms epoch, the contrast for non-targets revealed a main effect of experiment [F(1,34) = 4.31, P < 0.05], reflecting the fact that the positive-going old/new effects at anterior locations were larger in Experiment 1, while the negative-going effects at posterior locations were larger in Experiment 2. The contrast for targets over this epoch revealed an interaction between experiment and the anterior/posterior dimension [F(1,47.1) = 4.17, P < 0.05], reflecting the fact that while the negative-going ERP old/new effects at posterior electrodes differ minimally across experiments, at anterior electrode locations the positive-going old/new effects are markedly larger in Experiment 1 than in Experiment 2. This interaction was not, however, reliable for the analysis of rescaled data.

4. Discussion

Participants were able to discriminate between targets and non-targets in both experiments, but the proportion of correct target judgments was higher in Experiment 2 than in Experiment 1. This pattern of behavioural data indicates that participants were able to use recollection of the colour of study presentation as a basis for distinguishing between targets and non-targets, and that the amount or quality of information available for this discrimination was greater in the second experiment.
In and of themselves, however, the behavioural findings in these experiments do not support unequivocally the assumption that participants recollected information about both targets and non-targets in order to make judgments at test. Succeeding or failing in the recollection only of colour information for targets can also serve as the basis for distinguishing targets from non-targets and new words [19,20]. As noted in the Introduction, this strategy becomes increasingly effective as the likelihood of recollecting information associated with targets increases, and the analysis of ERPs evoked by test stimuli in these two experiments provides an opportunity to assess the contribution of recollection during the completion of exclusion tasks that differed primarily in terms of the accuracy of target judgments.

Reliable left-parietal ERP old/new effects were obtained for targets in both experiments, but for non-targets in Experiment 1 only, suggesting that participants depended upon recollection of information associated with non-targets to a greater degree in Experiment 1 than in Experiment 2. Since the accuracy of task judgments was superior in the second experiment, this finding is consistent with Herron and Rugg’s proposal that a strategy of depending primarily on recollection of target information for task judgments is employed when the likelihood of recollecting information about targets is high [14,19,20].

It is also of critical importance here that this evidence for selective recollection was obtained in tasks where the principal feature distinguishing targets from non-targets was colour. Thus, participants were able to restrict recollection of information associated with non-targets, while in both experiments, recollecting colour information about targets. These results suggest that a considerable degree of control can be exerted over what is and what is not recollected, even under circumstances when there is a close correspondence between these two classes of information.

A second important aspect of the findings in Experiment 1 is the fact that the left-parietal ERP old/new effect for non-targets, while reliable, was smaller than the effect for targets. This finding raises the possibility that, while relying on recollection of non-target colour information to a greater extent in Experiment 1 than Experiment 2, participants in Experiment 1 also relied upon recollection of target colour information to a greater degree than upon recollection of non-target colour information. This could have come about because of variability within participants with respect to the retrieval strategy that was adopted during the course of the retrieval phases, or because of variability across participants, whereby some but not all participants in Experiment 1 depended solely upon the success or failure of recollection of colour information about targets in order to make task judgments [52,54].

While the present data do not permit a clear separation of these alternatives, the accuracy of either does not challenge two central claims. First, that the degree of dependence upon recollection of colour information associated with non-targets is correlated inversely with the accuracy of target judgments. Second, that this strategic control of retrieval can occur despite an apparently close correspondence between the sources of information associated with targets and with non-targets.

Whether the same conclusions hold for auditory source information remains to be determined. As noted in the Introduction, reliable non-target ERP old/new effects were reported by Wilding and Sharpe [54] in two experiments where target designation was based on the voice (male/female) in which words were spoken at study and target accuracy was ≥0.70. Whether these non-target effects would be attenuated as accuracy increased to the levels achieved by participants in Experiment 1 is a question for future research, as is the question of whether there are other forms of information for which selective recollection does not or cannot occur during completion of the exclusion task [21].

It is also important to note that the analyses of the study-phase RT data provide no support for the possibility that some of the differences between the ERPs at test in Experiments 1 and 2 came about because of differential encoding of targets and of non-targets. In both experiments, the colour of words designated as targets was changed halfway through the experiment—after four study–test blocks in Experiment 1 and 6 study–test blocks in Experiment 2. Participants were not informed of target designation until the start of each test phase, they were not informed of the total number of study–test blocks in each experiment, and they were told that target designation might change for each retrieval phase. Despite these aspects of the design, however, the possibility remains that participants made assumptions about likely target designation and encoded information at study differentially on the basis of these assumptions.

In order to assess this possibility, the RTs for words presented during the study phases in the blocks where the same target designation had been required in the three preceding blocks were analysed after being separated according to whether they were designated subsequently as targets or non-targets. The analyses revealed no effects involving block position or target designation, a finding consistent with the view that the differences between the ERPs at test can be ascribed to factors operating at retrieval rather than at encoding5. Support for the claim that the test data is a function of strategies adopted at test also stems from the fact that qualitatively similar patterns of target and non-target ERP old/new effects have been obtained in studies where there was only one study phase [20,21].

A second challenge for the conclusions drawn above stems from the consideration that there may be variations across the two experiments in the levels of confidence associated with task judgments. Johnson et al. [29] reported that ERPs elicited during a recognition memory task may be sensitive to variations in response confidence, and in their study, these variations manifested themselves as an increased positivity with increasing confidence, the effects being largest at left-central scalp locations (see also [15]). Due to...
the fact that these modulations linked with decision confidence were identified at scalp locations and in time periods (~500–900 ms post-stimulus) that overlap with those in which the left-parietal ERP old/new effect is typically observed, it is possible that variations in decision confidence may be responsible for the patterns of ERP old/new effects that were obtained in the experiments reported here.

For the following reason, however, this account seems unlikely. Response confidence may on average have been greater in Experiment 2 than in Experiment 1 because of superior response accuracy. In Experiment 2, however, left-parietal ERP old/new effects were non-significant for non-targets, while the effects for the same category were reliable in Experiment 1. Given that the data indicating ERP sensitivity to response confidence predict relatively greater positivity with increasing confidence [15,29], and the reasonable inference that the confidence in judgments increases with response accuracy, this pattern of data suggests that variations in confidence do not explain satisfactorily the pattern of data across the two experiments: for non-targets, the left-parietal ERP old/new effects are larger in the experiment where response confidence is if anything likely to be lower. Rugg et al. [40], moreover, show data illustrating that ERP old/new effects at left-parietal scalp locations differ little according to response confidence (see also [36,38]).

What mechanisms might operate during retrieval tasks in order to permit control over episodic retrieval? Common to several conceptual frameworks is the assumption that processes engaged before and/or after an interaction between a retrieval cue and a memory trace can influence what information is available and/or employed for memory judgments [3,35,48]. Anderson and Bjork [2] identified two classes of process that operate prior to this interaction and which might support selective retrieval (also see [1,30]). The first, target bias, describes processes that operate directly on memory representations themselves. This class of process enables selective retrieval by making some representations more likely to interact with an appropriate retrieval cue than others. A second class of process described by Anderson and Bjork is cue bias, which describes processes that operate directly on retrieval cues and a memory trace can influence the nature of that interaction. A third class of process discussed by Anderson and Bjork—attention bias—has a different locus, operating downstream of any interactions between a retrieval cue and a memory trace. This class of process enables selective retrieval by ensuring that processing resources are allocated only to task-relevant material that becomes available as a result of the cue/trace interaction.

There is to date no electrophysiological data that distinguish unequivocally between these classes of bias mechanism, although it has been argued that target bias mechanisms are likely to be maintained throughout a retrieval task and thus will not be revealed in time-locked and baseline-corrected event-related potential measures of neural activity [22]. Herron and Rugg [19] favoured a cue-bias account of the ERP data they acquired in a task where the encoding phase comprised presentation of words and pictures. Test stimuli were words, and the old words were either re-presentations of words, or of words corresponding to the objects shown in the pictures. In separate retrieval phases, targets were designated as old words encountered either as words or pictures at encoding.

There were reliable parietal old/new effects for targets in both target designations, but reliable parietal effects for non-targets only when pictures were designated as targets. Herron and Rugg [19] noted that these data could not be accommodated straightforwardly within an attention bias account, but could be within a cue-bias account if it was assumed that: (1) in the word target condition, participants were able to process cues sufficiently selectively to restrict recollection to information associated with studied words, perhaps because of the perceptual match for the target stimuli, while (2) in the picture target condition, the absence of the perceptual match for target stimuli meant that processes engaged to recover information about studied pictures involved processing operations that also resulted in recollection of information associated with studied words.

This account does not preclude the possibility that a combination of these classes of bias mechanism is typically responsible for selective episodic retrieval, nor whether the extent to which these processes act in concert varies according to factors such as the content that is to be retrieved and the structure of retrieval tasks. Identification of boundary conditions for when selective retrieval can occur—a goal of the study described here—is a precursor to attempts to delineate further the mechanisms that enable selective retrieval of episodic information to be accomplished.

Over and above the implications of the pattern of left-parietal ERP old/new effects in these two experiments for the resolution with which recollection can be constrained, the contrasts between the target and non-target ERP old/new effects revealed other modulations of the electrical record that have been linked to memory processes. The analyses of the ERP old/new effects provided no evidence for qualitative differences between the ERP old/new effects in the two experiments, but there were amplitude differences of note. In the 300–500 ms post-stimulus time window, a fronto-central modulation comprising a greater positivity for correctly identified old than for new items has been linked to familiarity [4,6,34,41,59], which according to dual-process accounts of recognition memory is one basis for making old/new recognition memory judgments [23,32,57]. Although the scalp maps in Fig. 2 show that there are some differences according to experiment and target/non-target status in this epoch, these were not borne out by the analyses, which revealed only quantitative differences between the ERP old/new effects for targets.
A disparity between the ERP old/new effects in the two experiments that does receive statistical support is evident primarily at frontal electrode locations, where the right-lateralised ERP old/new effects were markedly more prominent in Experiment 1 than in Experiment 2, and larger for targets than for non-targets. According to one account of the right-frontal ERP old/new effect, it indexes processes responsible for monitoring retrieval processing in service of task goals [42,43]. The current data are broadly compatible with this account, in so far as the relatively lower level of memory accuracy in Experiment 1 might well have encouraged a greater degree of content and/or response monitoring than in Experiment 2, where the superior level of memory accuracy may have precluded requirements for extensive monitoring.

A further notable aspect of the ERP old/new effects is the fact that common to both experiments is the relatively focal negativity for old compared to new words that is evident primarily during the 900–1400 ms epoch, and which has a posterior midline maximum. While Fig. 1 shows that there are small differences between these posterior negativities for targets and for non-targets across experiments, the principal divergence is between the ERPs evoked by new words and those evoked by both classes of old word. This late posterior negativity is evident in a number of recent ERP studies of source memory (for examples, see [27,31]), and the functional significance of the modulation is a matter of continuing debate. In a series of studies, Cycowicz, Friedman, and colleagues [7–9] have reported pronounced posterior negativities in tasks where the information to be retrieved was the colour (red/green) in which pictures were presented in a prior study phase. In their experiments, the effect did not differ reliably according to the accuracy of source judgments, leading to the proposal that it reflects processes linked closely to the search for and/or retrieval of colour information [7], or perhaps processes that are linked with retrieval and/or evaluation of source information but not necessarily restricted to colour [17].

The data in this experiment contribute in two ways to questions about the functional significance of this posteriorly distributed old/new effect. First, the fact that the effect is of comparable magnitude across the two experiments while the right-frontal ERP old/new effect is markedly larger in Experiment 1 is consistent with the view that the two effects are functionally dissociable (see also [49]). The second point stems from the fact that the ERP waveforms in the studies of Cycowicz and colleagues have been displayed with respect to a nose-tip reference only, with the exception of one recent publication [17]. In the majority of ERP memory studies, a linked mastoid reference has been employed. The data from Experiment 2 are relevant in this regard, as EEG was also acquired from the nose-tip, permitting the data to be re-referenced to that site. The outcome of this procedure is shown in Fig. 3, where it can be seen that the consequence of this change of reference is a considerable increase in the amplitude of the late posterior negativity, and a small reversal of the positive-going ERP old/new effects at anterior electrodes.

The data referenced to linked mastoids that is shown in Figs. 1 and 2 bear strong similarities to that which has been reported for different kinds of source information such as voice [51] or study task [49]. One interpretation of this correspondence is that the posterior negativity is not linked strongly to colour-specific processing, given the apparent match between the data reported here and in other studies in which a linked mastoid reference has been employed and in which retrieval of different kinds of source information has been encouraged. The data in Fig. 3 support this account in so far as when a nose-tip reference is employed, the pattern of data resembles closely that reported previously by Cycowicz and colleagues [7–9]. In combination, therefore, the data provide little support for a colour-specific account of the late posterior negativity.

From this perspective, Johannsen and Mecklinger’s recent proposal that one aspect of the posterior negativity reflects processes related to maintaining and/or forming representations of attribute conjunctions is parsimonious, since it makes no claims about the content of those conjunctions [17,26]. It is important to observe, however, that these suggestions are based on informal comparisons across studies, and in determining the functional significance of this late posterior modulation, it will be important in future studies to contrast ERPs associated with retrieval of different contents while equating all other factors [17].

To summarise, the principal findings in these two experiments concern the pattern of left-parietal ERP old/new effects—the electrophysiological signature of recollection. In the first experiment, overall task accuracy was inferior to that in the second experiment, and targets as well as non-targets were associated with reliable left-parietal ERP old/new effects. In the second experiment, only targets were associated with the effect. In combination, these findings are consistent with Herron and Rugg’s proposal that relying upon the success or failure of recollection of information associated with targets only is a strategy that is adopted when memory for targets is high [19,20]. In addition, the findings in these two experiments demonstrate that control over what is and what is not recollected can be exercised with a high degree of precision, as attested to by the fact that the data in Experiment 2 suggest that participants were able to recollect some colour information while not recollecting other highly similar information.

Acknowledgments

This research was supported by the UK Biotechnology and Biological Sciences Research Council (BBSRC). We would like to thank Mariam Dzulkifli and Nazanin Azimian-Faridani for comments on a draft of the manuscript.
References


[45] M.E. Smith, Neurophysiological manifestations of recollective expe-


