Neural Correlates of Individual Differences in Strategic Retrieval Processing

Emma K. Bridger, Jane E. Herron, Rachael L. Elward, and Edward L. Wilding
Cardiff University

Processes engaged when information is encoded into memory are an important determinant of whether that information will be recovered subsequently. Also influential, however, are processes engaged at the time of retrieval, and these were investigated here by using event-related potentials (ERPs) to measure a specific class of retrieval operations. These operations were revealed by contrasts between ERPs elicited by new (unstudied) test items in distinct tasks, the assumption being that these contrasts index operations that are engaged in service of retrieval and that vary according to the demands of different retrieval tasks. Specific functional accounts of this class of retrieval processing operations assume that they influence the accuracy of memory judgments, and this experiment was designed to test for the first time whether this is in fact the case. Toward this end, participants completed 2 retrieval tasks while ERPs were acquired, and the extent to which processes were engaged differentially across tasks in service of retrieval was operationalized as the magnitude of the differences between the new-item ERPs that were elicited. This measure correlated positively with response accuracy on the tasks, which provides strong evidence that this class of retrieval processing operations benefits the accuracy of memory judgments.

Keywords: event-related potentials (ERPs), individual differences, strategic retrieval processing

According to the source-monitoring framework (Johnson, Hashtrudi, & Lindsay, 1993), making judgments about the origin of a memory involves an assessment of various qualitative characteristics. This is typically viewed as a strategic endeavor because the characteristics that are evaluated will depend on those that are maximally diagnostic for the relevant judgment. For example, one means of distinguishing between real and imagined events would be to evaluate the perceptual information that accompanies the memory, assuming that memories for real events will carry more perceptual content than will memories for imagined events (Johnson & Raye, 1981).

This selective focus on certain kinds of task-relevant information is an example of strategic retrieval processing, with this term denoting processes that are engaged during memory retrieval according to the particular demands of the memory judgment that is required. Evidence for the deployment of the kinds of strategies described above comes from participant reports that the kinds of information they would rely on when making source judgments are in fact the kinds that differentiate between the possible sources (Johnson, Foley, Suengas, & Raye, 1988). In addition, a second line of evidence for this kind of strategic retrieval processing has come from brain imaging studies of memory retrieval, in which it has been argued that one particular contrast between measures of neural activity provides a means of indexing strategic retrieval processing operations (Johnson, Kounios, & Nolde, 1997; Rugg & Wilding, 2000; Wilding, 1999).

This contrast is between neural activity that is elicited by new (unstudied) stimuli across tasks with different retrieval demands (Rugg & Wilding, 2000; for precursors, see Johnson et al., 1997; Ranganath & Paller, 1999; Wilding, 1999). The argument is that in this contrast, any differences between the neural activities can be attributed to processes that are engaged strategically according to the specific retrieval demands of the task. This might not be the case when contrasting neural activities associated with old (studied) test items under different task demands, as for this contrast, any differences that are obtained may be a consequence of the recovery of different contents (Rugg & Wilding, 2000), a point to which we return later.

Reports of reliable differences in neural activity between classes of new test items come from a few functional magnetic resonance imaging studies of memory retrieval (e.g., Hornberger, Rugg, & Henson, 2006a; Woodruff, Uncapher, & Rugg, 2006) and, more importantly for present purposes, from several event-related potential (ERP) studies (Dzulkiﬁ, Sharpe, & Wilding, 2004; Dzulkiﬁ & Wilding, 2005; Herron & Rugg, 2003; Hornberger, Morcom, & Rugg, 2004; Hornberger, Rugg, & Henson, 2006b; Johnson et al., 1997; Ranganath & Paller, 1999, 2000; Robb & Rugg, 2002). Across these ERP studies, a number of different retrieval requirements have been used, and the ways in which ERPs elicited by new items have diverged have varied accordingly. This is important because the heterogeneity in the findings is consistent with the view that the processes that are indexed vary according to specific

Emma K. Bridger, Jane E. Herron, Rachael L. Elward, and Edward L. Wilding
Cardiff University Brain Research Imaging Centre (CUBRIC), School of Psychology, Cardiff University, Cardiff, Wales, United Kingdom.

This research was supported by the U.K. Biotechnology and Biological Sciences Research Council (BBSRC) in the form of a research studentship (to Emma K. Bridger) and research grant awards (to Jane E. Herron and Edward L. Wilding).

Correspondence concerning this article should be addressed to Emma K. Bridger, School of Psychology, Cardiff University, Cardiff, CF10 3AT, Wales, United Kingdom. E-mail: BridgerEK@cardiff.ac.uk
task demands, as would be expected if they index strategic retrieval processing operations as defined above. It is unlikely, moreover, that these differences reduce to indices of retrieval effort, as they remain after task difficulty has been equated (Robb & Rugg, 2002; for additional comments, see Dzulkiﬁ et al., 2004).

It has also been shown that ERP indices of strategic retrieval processing are evident primarily in tasks in which retrieval is required explicitly, rather than when old and new test items are encountered, but task instructions make no reference to a prior learning episode (Hornberger et al., 2006b). This finding is consistent with the assumption that the engagement of strategic retrieval processing is contingent upon goal-directed retrieval requirements, and in keeping with this assumption, functional accounts of differences in neural activity across classes of new items have been cast in terms of the benefits that accrue from using strategic processing operations. For example, it has been proposed that the processes indexed by ERPs reﬂect operations that maximize the overlap between processes engaged at encoding and retrieval (Rugg, Herron, & Morcom, 2002) or that they index processes that increase the likelihood that representations of cues will interact with some memory traces rather than others (Dzulkiﬁ, Herron, & Wilding, 2006; Herron & Rugg, 2003).

These accounts predict that response accuracy in memory tasks will beneﬁt when strategic retrieval processing of the kind described above is engaged, and the work described in this article is a direct test of that proposal. An alternative proposal, however, is that the processes indexed in neural contrasts of this kind reﬂect operations that assist simply the efﬁciency with which information is retrieved or analyzed in service of task demands. If this account is correct, then the strategic retrieval processing operations indexed in the critical ERP contrasts might relate to processes such as the speed with which memory judgments were made but would not extend to the accuracy of those judgments.

To adjudicate between these competing accounts, we recorded ERPs from the retrieval phases of two verbal memory tasks with different demands. The relationship between ERP indices of strategic retrieval processing and response accuracy on the tasks was then assessed. For both tasks, the encoding requirements were for participants to generate either a function for or judge how easy it would be to draw the objects denoted by study words. A binary response was required in both retrieval tasks, in which unstudied words were presented intermixed with words that had been encoded under either function or drawing instructions. In one task, participants were asked to respond on one key to words encoded in the function task, and on a second key to respond to new words as well as to words encoded in the drawing task. In the other retrieval task, the designation for which encoding task shared the same response key as new test words was reversed. Tasks with response requirements such as these have been termed exclusion tasks (see Jacoby, 1991).

ERPs were recorded for all words at test, and the critical ERPs were those elicited by new words in each task. The key experiment prediction was that ERP evidence for the degree to which strategic retrieval processing was engaged would be correlated positively with response accuracy, in support of the view that these processes beneﬁt the accuracy of memory judgments. Evidence for the degree to which strategic retrieval processing was engaged was operationalized as the magnitude of the voltage differences between the ERPs elicited by new items in the two retrieval tasks for each participant. The magnitude of these voltage differences is thus assumed to index to what extent strategic retrieval processing operations differed across the two tasks.

The experiment prediction was tested via staged analyses. First, by using a median split to allocate participants to one of two groups depending upon response accuracy. Second, by investigating whether the magnitude of the differences between the critical classes of ERPs was larger for the group who responded more accurately on the tasks. To anticipate the results, we obtained this outcome, and in a second analysis stage, we selected the scalp sites at which the ERP differences were largest for the group with superior response accuracy as those for which, on an individual participant basis, we plotted response accuracy against the ERP index of the degree to which strategic retrieval processing was engaged. A positive relationship between these two measures would provide, for the first time, strong support for the view that strategic retrieval processing operations indexed by ERPs beneﬁt the accuracy of memory judgments. It would also point to the utility of real-time measures of retrieval processing in determining the factors that inﬂuence accurate source attributions.

Method

Participants

Forty-seven right-handed participants (7 men) gave informed consent and were paid the sterling equivalent of $12 per hour. All spoke English as a ﬁrst language, did not have a diagnosis of dyslexia, and were free from neurotropic medication. Data from 11 participants (1 man) were discarded because of experimenter error (1), failure to follow instructions (3), and excessive electro-oculogram (EOG) artifacts resulting in insufficient trials per condition (7). Mean age of the remaining participants was 21 years (range = 18–27 years).

Design

Three hundred and sixty words were taken from Kucera and Francis’s (1967) corpus and presented in white letters on a black background on a computer monitor placed approximately 1 m from participants. Words were concrete nouns with a frequency of 1–15 per million and a letter range of 4–9. Maximum horizontal and vertical visual angles were 4.6° and 0.6°, respectively. Six groups of 60 words were selected at random for a full experiment list. Each experiment list comprised two study–test cycles. Each study phase comprised two word groups (120 words). These were repeated at test together with a third word group to give 180 test words per cycle. No words were repeated across cycles. Word groups were rotated fully across experiment lists, resulting in the formation of six complete lists.

Procedure

Participants were ﬁrst ﬁtted with an electrode cap (see below). They completed two study–test cycles after a short practice phase in which they were familiarized with task response requirements. The requirements in the test phase of the practice session corresponded to the ﬁrst set of retrieval instructions that each participant was given. The researcher read aloud the task instructions, and participants were also given written descriptions.
In each study phase, participants completed two tasks. Cues preceding each word signaled which task to complete: “FUNCTION?” for the function task and “DRAW?” for the drawing task. In the function task, they were asked to say aloud a function for the object denoted by the word. In the drawing task, they were asked to rate verbally how difficult it would be to draw the object denoted by each word. The four response options were as follows: very easy, fairly easy, fairly difficult, and very difficult. Cues remained on the screen for 1,000 ms, followed by a blank screen for 500 ms. Order of encoding task cues was pseudorandomized; no more than three consecutive words were preceded by the same cue. Each study word was presented for 300 ms before the screen was made blank. Participants initiated the next trial by pressing a key on a response pad, and the trial started 2,000 ms after this response.

Each test trial began with a fixation asterisk that remained on the screen for 500 ms, followed by a 500 ms blank screen and then the test word for 300 ms. The screen was then made blank until the participant responded. The next trial began 1,500 ms later. Participants were instructed to respond using one thumb to words from one of the two encoding tasks (hereafter targets) and with the other thumb to new test words as well as those from the other task (hereafter nontargets). Target designation (function/drawing) changed across study–test cycles. Participants were informed of target designation immediately prior to each test phase, and they were not informed during the practice session that a change would occur. The thumbs used for responses were balanced across participants, and 50% of participants completed the function task designation first. There was a break of 2 min between each study–test cycle and between study and test phases.

**Electroencephalogram (EEG) Recording**

EEG was recorded from 32 locations on the basis of the International 10–20 system (Jasper, 1958), including midline (Fz, Cz, Pz, Oz) and left/right hemisphere sites (FP1/FP2, F7/F8, F5/F6, F3/F4, F1/F2, T7/T8, C5/C6, C3/C4, C1/C2, P7/P8, P5/P6, P3/P4, P1/P2, O1/O2). Additional electrodes were placed on the mastoid processes. Electro-ocular activity (EOG) was recorded from above and below the left eye (vertical EOG) and from the outer canthi (horizontal EOG). EEG (range = DC-419 Hz; sampling rate = 2048 Hz) was acquired referenced to linked electrodes located midway between POz and PO3/PO4, and re-referenced off-line to the average signal at the mastoids. Trials containing large EOG artifact were rejected, as were trials containing A/D saturation or baseline drift exceeding ±80 μV. Other EOG blink artifacts were corrected using a linear regression estimate (Semlitsch, Anderer, Schuster, & Presslich, 1986). The data were band-pass filtered off-line (0.03–60 Hz), epoched, and down-sampled to 167 Hz. Epochs were 1,536 ms in length, including a 102 ms baseline relative to which all mean amplitudes were computed. A 7-point (22 Hz) binomially weighted smoothing filter was applied to the averaged ERPs before analysis.

**Results**

**Behavioral Data**

Table 1 shows the proportions of correct judgments and associated reaction times (RTs) to old and new words separated according to whether words encoded under function or drawing instructions were designated as targets. For both of these target designations, target responses to targets were significantly more likely than target responses to nontargets and new words—for all comparisons, t(35) ≥ 25.00, p < .001.

Further analyses were guided by the intention to compare ERP indices of strategic retrieval processing according to how well people completed the tasks (see the introduction section). In a first step, participants were separated according to an averaged measure of target/nontarget discrimination \( p(\text{target hit}) - p(\text{nontarget false alarm}) \) collapsed across whether words from the function or drawing encoding condition were designated as targets. Table 2 shows (for both groups) mean proportions of correct responses and their corresponding RTs. Mean target/nontarget discrimination scores for the high-accuracy group were 0.83 (function) and 0.79 (drawing). The corresponding mean values for the low-accuracy group were 0.63 and 0.58. These discrimination scores were assessed via a two-way analysis of variance (ANOVA) with factors of group (high/low) and target designation (function/drawing). The Geisser–Greenhouse correction for nonsphericity was used where appropriate here and in the subsequent ERP analyses (Greenhouse & Geisser, 1959).

There was superior target/nontarget discrimination in the high group, \( F(1, 34) = 35.33, p < .001 \), and the decision to use discrimination scores collapsed across target designation was supported by the absence of a reliable effect for this factor, although it did approach significance, \( F(1, 34) = 8.00, p = .06 \), reflecting a trend for superior discrimination in the function target designation. An independent t-test showed that the likelihood of a correct response to a new test word did not vary with group.\(^1\) This is important because it suggests that any differences between the ERPs elicited by new tests items and separated according to target

---

\(^1\) The exclusion task response requirements mean that some responses labeled as correct rejections may be a response to a new item on the basis of the belief that the item was a nontarget, or a guess that the item was a nontarget in the presence of incorrect information indicating that the item was old rather than new. Given the relatively high levels of response accuracy in the tasks, trials of this type are unlikely to be present in significant proportions; in addition, the pattern of behavioral data gives little incentive to believe that any such trials will be distributed in meaningfully unequal proportions across the two tasks.
Table 2
Mean Proportions of Correct Responses to Targets, Nontargets, and New Words Separated According to Group (High/Low) and Collapsed Across Target Designation

<table>
<thead>
<tr>
<th>Group</th>
<th>Word status</th>
<th>p(correct)</th>
<th>RT (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Target</td>
<td>.88 (.07)</td>
<td>1,288 (218)</td>
</tr>
<tr>
<td></td>
<td>Nontarget</td>
<td>.92 (.04)</td>
<td>1,328 (230)</td>
</tr>
<tr>
<td></td>
<td>New</td>
<td>.98 (.03)</td>
<td>1,079 (201)</td>
</tr>
<tr>
<td>Low</td>
<td>Target</td>
<td>.74 (.11)</td>
<td>1,211 (294)</td>
</tr>
<tr>
<td></td>
<td>Nontarget</td>
<td>.86 (.10)</td>
<td>1,269 (344)</td>
</tr>
<tr>
<td></td>
<td>New</td>
<td>.96 (.05)</td>
<td>1,014 (233)</td>
</tr>
</tbody>
</table>

Note. Standard deviations are in parentheses. RT = reaction time.

designation are unlikely to reflect task difficulty differences (cf. Robb & Rugg, 2002).

Analyses of the RT data were conducted for correct responses only via ANOVA with factors of response category (target/nontarget/new), target designation (function/drawing), and group (high/low). The analysis revealed a main effect of category only, \( F(2, 68) = 66.71, p < .001 \), reflecting the faster RTs for responses to new test words. A planned comparison of the RTs for new words separated according to designation did not reveal a reliable difference (\( t < 1 \)).

ERP Analyses

The critical analyses for the experiment predictions were those for the ERPs elicited by unstudied test words, and these are described first below. These are followed by descriptions of the outcomes of comparable analyses for the ERPs elicited by targets and nontargets, again separated according to target designation. The outcomes of the analyses for targets and nontargets offer the opportunity to constrain possible interpretations of the findings for the ERPs elicited by new words (see the Discussion section, and for a conceptually similar approach, see Ranganath & Paller, 1999, 2000). The mean numbers of trials contributing to the averaged ERPs for each participant and response category can be seen in the Appendix.

The initial analysis strategy was the same for all of the paired contrasts (new words, targets, and nontargets). The first analyses were conducted separately for four epochs (300–500 ms, 500–800 ms, 800–1,100 ms, and 1,100–1,400 ms), including four regions of interest each comprising five sites: left-frontal (FP1/F1/F3/F5/F7), right-frontal (FP2/F2/F4/F6/F8), left-posterior (O1/P1/P3/P5/P7), and right-posterior (O2/P2/P4/P6/P8). For each epoch and paired contrast, a global analysis was conducted with factors of group (high/low), target designation (function/drawing), hemisphere (left/right), anterior/posterior dimension (anterior/posterior), and site (inferior/midlateral/superior/midline/presuprior). When interactions involving group and target designation were obtained in these analyses, subsequent analyses within group were conducted to determine the reasons for the interaction terms.

As noted in the introduction section, the outcomes of these analyses were also used to guide the selection of scalp sites that were used when calculating correlations between response accuracy and differences between ERPs separated according to target designation. The key analysis outcomes for the test of the experiment prediction are as follows: (a) indications that the magnitude of the differences between new-item ERPs is larger in the high-accuracy group than in the low-accuracy group, and (b) at the individual participant level, there will be a positive correlation between target/nontarget discrimination and the size of the differences between the new-item ERPs.

**ERPs Elicited by New Items (Correct Rejections)**

Figure 1 (upper panel) depicts the grand averaged waveforms elicited by correct rejections in the two tasks for the high-accuracy group. The lower panel depicts the ERPs for the same response categories for the low-accuracy group. The upper panel shows that at midline-posterior sites from 700 ms onward, ERPs associated with correct rejections in the function target designation are relatively more positive going than those associated with the drawing target designation. There are not comparable differences between the ERPs in the low-accuracy group, although there is some differentiation between these classes of ERPs, primarily at electrodes P3 and O1.

The initial analyses of these ERPs revealed reliable interactions involving group and target designation in each epoch (see Table 3). Follow-up analyses for each group revealed reliable effects of target designation only in the high group. A three-way interaction between designation, the anterior/posterior dimension, and hemisphere was reliable from 500 to 1,100 ms: 500–800 ms, \( F(1, 17) = 5.35, p < .05 \); 800–1,100 ms, \( F(1, 17) = 4.86, p < .05 \). Furthermore, the designation by anterior/posterior term was reliable for the 1,100–1,400 ms epoch, \( F(1, 17) = 5.89, p < .05 \). These reliable effects came about because of the posterior and generally left-sided maxima of the differences between the ERPs that were elicited by correct rejections in the high-accuracy group (see Figure 1).

The presence of reliable differences between the ERPs elicited by correct rejections in the high-accuracy group only is consistent with the view that strategic retrieval processing influences positively the accuracy of memory judgments. The next analysis stage was implemented to establish whether evidence consistent with this account could be obtained at the level of individual participants. Toward this end, the outcomes of the preceding analyses guided the selection of left posterior/occipital sites (O1, P1, P3, P5, and P7) as those from which an ERP measure of the degree to which strategic retrieval processing differed across tasks was plotted against target/nontarget discrimination for each participant. In keeping with the operational definition given in the introduction section, this ERP measure for each site and epoch was a difference score obtained by subtracting mean amplitudes associated with correct rejections in the drawing designation from those associated with the function designation. Critically, significant positive correlations between these measures were obtained at several electrodes in all three time windows (see upper section of Table 4). The lower section in Table 4 shows that the correlations remain significant from 800 to 1,400 ms when 2 participants with poor discrimination scores are removed, and these findings converge with those that were obtained in the group-level analyses already described.
ERPs Elicited by Old Test Items Attracting Correct Judgments

Figure 2 depicts the grand averaged waveforms elicited by correct responses to nontargets in the two tasks for both groups. Although there is little evidence for marked differences between these classes of ERPs for the high-accuracy group (upper panel), the lower panel demonstrates some differentiation between these classes of ERPs at posterior (and primarily right hemisphere) sites for the low-accuracy group. These ERPs were subjected to the same analysis strategy as for the ERPs elicited by correct rejections, and Table 3 shows that a reliable four-way interaction was revealed in the 300–500 ms epoch. However, separate follow-up analyses within each group revealed no reliable effects.

Figure 3 shows the grand averaged waveforms for the two groups for the ERPs elicited by correct responses to targets. The ERPs associated with targets differ little according to target designation for the low-accuracy group, but they do vary with designation for the high group. From approximately 400 ms onward, there is a sustained relative positivity for targets associated with the drawing rather than the function condition, primarily at right-frontal sites. In keeping with this description of differences according to group, Table 3 shows that interactions including the factors of group and target designation occurred from 500 ms onwards, and subsequent within-group analyses revealed interactions with target designation only in the high group. For this group, the interaction term including target designation, anterior/posterior, and site was significant in all epochs after 500 ms: 500–800 ms, $F(3.3, 56.5) = 5.81, p < .01$; 800–1,100 ms, $F(3.4, 57.3) = 3.08, p < .05$; 1,100–1,400 ms, $F(3.3, 55.4) = 3.46, p < .05$. In addition, the factors of designation and hemisphere interacted from 800 to 1,400 ms—$F(1, 17) = 6.05, p < .05$, and $F(1, 17) = 15.49, p < .01$—moderating a main effect of target designation in the 800–1,100 ms window, $F(1, 17) = 7.38, p < .05$. These interactions confirm that targets were more positive
going in the drawing designation over the right hemisphere, in particular at frontal scalp sites and extending to some degree to posterior midline sites, as depicted in Figure 4 (lower panel).

As for the correct rejection analysis, the sites where target differences were greatest in the high-accuracy group were taken as those from which amplitude differences were derived and correlated with target/nontarget discrimination. At right-prefrontal and parietal sites (FP2, F2, F4, F6, F8, P1, P3), mean amplitudes associated with target hits in the drawing designation were subtracted from those associated with the function designation, mirroring the subtraction completed for the new-item ERPs. Moderate correlations between these measures and discrimination were observed at a number of sites in all three time windows (see Table 5). Although these outcomes are somewhat sensitive to the inclusion of two outliers, some effects remain after they have been removed. In addition, note that because the ERPs associated with the drawing condition for targets are more positive going than those from the alternate condition, the subtraction used here (function minus drawing, as was done for correct rejections) results generally in negative values for the magnitude of the differences between conditions for each participant; however, the correlation that was obtained indicates that the size of the differences between conditions increases as accuracy increases, as was the case for the new-item ERP contrasts.

### Analyses of Scalp Distributions

To determine whether the scalp distributions of the ERP effects shown in Figure 4 change with time, we conducted separate analyses of the ERP differences between those elicited by correct rejections and those elicited by targets. These analyses were restricted to data from the high-response-accuracy group, because only in this group were there reliable differences when the ERPs were separated according to target designation. The absence of robust within-group differences also explains why analyses of scalp distributions were not conducted for nontargets.

For new words and for targets, the ANOVAs included the factors of site (20) and epoch (3). The data submitted to analysis were the difference scores from which the Figure 4 maps were generated, and these data points were first rescaled using the maximum–minimum method to avoid confounding effects that result from variable magnitudes over conditions of interest from effects that are due to differences between the shapes of distributions (McCarthy & Wood, 1985). The analysis of the rescaled difference scores associated with the correct rejections subtraction revealed no significant effects involving epoch and site. This was also true for the analysis based on the target ERP data. These findings are consistent with the view that the same neural generators (hence the same cognitive processes) were engaged in each epoch.

### Table 3

**Outcomes of Global ANOVAs for Comparisons Between ERPs Associated With Each Item Type in Each Time Window**

<table>
<thead>
<tr>
<th>Item type</th>
<th>Epoch</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>300–500 ms</td>
</tr>
<tr>
<td>Correct rejections</td>
<td></td>
</tr>
<tr>
<td>GP × TD, F(1, 34)</td>
<td>4.23*</td>
</tr>
<tr>
<td>GP × TD × AP, F(1, 34)</td>
<td></td>
</tr>
<tr>
<td>GP × TD × HM, F(1, 34)</td>
<td>4.21*</td>
</tr>
<tr>
<td>GP × TD × AP × HM, F(1, 34)</td>
<td>9.48**</td>
</tr>
<tr>
<td>Nontarget hits</td>
<td></td>
</tr>
<tr>
<td>GP × TD × AP × HM, F(1, 34)</td>
<td>4.50*</td>
</tr>
<tr>
<td>Target hits</td>
<td></td>
</tr>
<tr>
<td>TD × HM, F(1, 34)</td>
<td>4.18*</td>
</tr>
<tr>
<td>GP × TD × HM, F(1, 34)</td>
<td></td>
</tr>
<tr>
<td>GP × TD × AP × ST, F(4, 136)</td>
<td>5.50**</td>
</tr>
</tbody>
</table>

*Note. Outcomes are shown only for those terms that included a reliable interaction involving group (GP) and target designation (TD) in at least one time window. Epsilon values computed as part of the Geisser–Greenhouse correction for deviations from sphericity are shown as subscripts. ANOVAs = analyses of variance; ERPs = event-related potentials; AP = anterior/posterior; HM = hemisphere; ST = site. *p < .05. **p < .01.

### Table 4

**Values for Pearson’s R Relating ERP Correct Rejection Difference Score Amplitudes (Function − Drawing) at Left Posterior Sites (O1, P1, P3, P5, P7) With Target/Nontarget Discrimination**

<table>
<thead>
<tr>
<th>Group</th>
<th>Site</th>
<th>500–800 ms</th>
<th>800–1,100 ms</th>
<th>1,100–1,400 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>All participants (N = 36)</td>
<td>O1</td>
<td>.38*</td>
<td>.47**</td>
<td>.46**</td>
</tr>
<tr>
<td></td>
<td>P1</td>
<td></td>
<td>.39*</td>
<td>.38*</td>
</tr>
<tr>
<td></td>
<td>P3</td>
<td>.33*</td>
<td>.44**</td>
<td>.49**</td>
</tr>
<tr>
<td></td>
<td>P5</td>
<td>.43*</td>
<td>.45**</td>
<td>.44**</td>
</tr>
<tr>
<td></td>
<td>P7</td>
<td>.47**</td>
<td>.47**</td>
<td>.42**</td>
</tr>
<tr>
<td>Outliers removed (N = 34)</td>
<td>O1</td>
<td>.49**</td>
<td>.52**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P1</td>
<td>.43*</td>
<td>.45**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P3</td>
<td>.47**</td>
<td>.47**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P5</td>
<td>.47**</td>
<td>.52**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P7</td>
<td>.38*</td>
<td>.37**</td>
<td></td>
</tr>
</tbody>
</table>

*Note. All tests were two-tailed. ERP = event-related potential. *p < .05. **p < .01.
Discussion

ERPs elicited by new test items attracting correct responses varied according to retrieval task but only when how well the participants were able to discriminate between targets and nontargets was taken into account: The magnitude of the differences between the two classes of new-item ERPs was correlated positively with target/nontarget discrimination. If the size of the differences between these ERPs is a marker of the degree to which strategic retrieval processing differed across the tasks, then the data support the view that strategic retrieval processing influences the accuracy of memory judgments.

Robust differences between ERPs elicited by classes of new test items have been reported in several studies (Dzulkifli & Wilding, 2005; Herron & Rugg, 2003; Hornberger et al., 2004, 2006b; Robb & Rugg, 2002), but to our knowledge there are no published reports of the correspondence between these differences and the accuracy of memory judgments. In the absence of such a correspondence, one plausible explanation for the ERP differences that have been reported is that they index operations influencing the speed or efficiency with which retrieved information is processed.

This account is challenged strongly by the fact that the magnitude of ERP differences at the scalp locations where these differences were largest correlated positively with target/nontarget discrimination. This can be seen clearly in Figure 5, and in combination, these data points support the claim that the degree to which strategic retrieval operations are engaged is beneficial to task performance.

Functional Significance of the Differences Between ERPs Elicited by New Items

The findings in this experiment also offer several insights into the form that the strategic retrieval processing operations identified here might take, and one of these follows from the time course of the differences between the ERPs elicited by new test items. A common finding in other studies is that differences in contrasts
between new-item ERPs have an extended time course, with little
evidence for changes in the distribution of these ERP effects over
time (e.g., Dzulkifli et al., 2004; Dzulkifli & Wilding, 2005;
Hornberger et al., 2004; Ranganath & Paller, 1999). This is also
the case in this experiment, as the scalp maps in the upper panel of
Figure 4 show clearly. The extended time course of these effects is
important because it continues beyond the period in which suc-
cessful episodic retrieval is typically considered to occur (Yoneli-
nas, 2002). Data from response-deadline memory paradigms show
that processes that support retrieval of contextual information
become available within 900 ms (Yonelinas & Jacoby, 1994).
Likewise, the ERP correlate of successful episodic retrieval is
pronounced between 500 and 800 ms after the presentation of a
retrieval cue (Allan, Wilding, & Rugg, 1998; Curran, 2000).

In light of this information about the time course of retrieval,
Hornberger et al. (2004) proposed that temporally extended ERP
differences between new test items reflect the maintenance of
internal representations of retrieval cues in service of recovery of
task-relevant information for old items (for related comments, see
Dzulkifli et al., 2004). This account provides a good interpretation
for the data reported here, which in turn leads to the proposal that
the differences between the new-item ERPs in this experiment
reflect the generation of internal representations that will bias
retrieval toward content that is specific to the drawing and the
function target designation tasks, respectively. This might reflect a
bias toward recovery of imagery-related information in the draw-
ing target designation and an emphasis on the cognitive operations
involved in accessing the semantic information necessary to make
function judgments in the alternate target designation.

Critically, however, this interpretation is based upon only a
paired contrast between the new-item ERPs, which means that
these ERP data in and of themselves cannot adjudicate between
accounts that postulate the engagement of distinct strategic re-
trieval processes in each target designation, as described above,
from accounts that assume a difference solely in the degree to
which the same set of processes are engaged. For example, the

Figure 3. Upper panel: Grand average event-related potentials (ERPs) elicited by target hits in the two target
designation conditions for the high-accuracy group. Lower panel: Grand average ERPs elicited by target hits in
the two target designation conditions for the low-accuracy group. In both panels, data are shown for 12 electrode
locations at left and right hemisphere sites over prefrontal (FP1, FP2), anterior (F7, F3, F4, F8), posterior (P7,
P3, P4, P8), and occipital (O1, O2) scalp sites.
differences between the new-item ERPs might equally well be interpreted as reflecting a greater relative emphasis on one kind of information in one task than in the other (cf. Rugg, Allan, & Birch, 2000). Data that encourage favoring the former account over the latter, however, come from considering the functional implications of the outcomes of the contrasts between the ERPs that were elicited by the targets and the nontargets, which are discussed below.

Functional Significance of the Differences Between ERPs Elicited by Old Items

The magnitude of the differences between the ERPs elicited by targets also predicted target/nontarget discrimination, although the sites at which these effects were evident were not the same as those for the new-item ERPs. There was not robust evidence, however, for comparable differences between the ERPs elicited by nontargets. These findings suggest that participants were prioritizing source information associated with targets over information associated with nontargets, and therefore suggest that the differences between the new-item ERPs index operations that were engaged specifically to encourage the retrieval of different contents in the tasks in which either words from the function or the drawing target designation from those in the function designation, and they are shown for the 500–800 ms, 800–1,100 ms, and 1,100–1,400 ms time windows. Each map is proportionately scaled between the maxima (red) and minima (blue) of the depicted effect, and the maximum and minimum values (microvolts) are shown below each map.

Figure 4. Upper panel: Topographic maps showing the scalp distributions of the differences between neural activity elicited by new test words for the high-accuracy group. Lower panel: Topographic maps showing the scalp distributions of the differences between neural activity elicited by targets for the same group. Voltage maps are computed on the basis of difference scores obtained by subtracting mean amplitudes for the event-related potentials elicited by words in the drawing target designation from those in the function designation, and they are shown for the 500–800 ms, 800–1,100 ms, and 1,100–1,400 ms time windows. Each map is proportionately scaled between the maxima (red) and minima (blue) of the depicted effect, and the maximum and minimum values (microvolts) are shown below each map.

Note. All tests were two-tailed. ERP = event-related potential.

Table 5
Values for Pearson’s R Relating ERP Target Difference Score Amplitudes (Function – Drawing) at Right Frontal and Parietal Sites (FP2, F2, F4, F6, F8, P1, P3) With Target/Nontarget Discrimination

<table>
<thead>
<tr>
<th>Group</th>
<th>Site</th>
<th>500–800 ms</th>
<th>800–1,100 ms</th>
<th>1,100–1,400 ms</th>
</tr>
</thead>
<tbody>
<tr>
<td>All participants (N = 36)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FP2</td>
<td></td>
<td>-.40*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F2</td>
<td></td>
<td>-.35*</td>
<td>-.42*</td>
<td>-.33*</td>
</tr>
<tr>
<td>F4</td>
<td></td>
<td>-.36*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F6</td>
<td></td>
<td>-.37*</td>
<td>-.35*</td>
<td></td>
</tr>
<tr>
<td>F8</td>
<td></td>
<td>-.39*</td>
<td>-.54*</td>
<td>-.35*</td>
</tr>
<tr>
<td>P1</td>
<td></td>
<td>-.39*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td></td>
<td></td>
<td>-.45*</td>
<td></td>
</tr>
<tr>
<td>Outliers removed (N = 34)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FP2</td>
<td></td>
<td>-.44**</td>
<td>-.44**</td>
<td>-.35*</td>
</tr>
<tr>
<td>F2</td>
<td></td>
<td>-.39*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* p < .05. ** p < .01.
Strategies for Completing Exclusion Tasks

Irrespective of the accuracy of this account, however, a critical point for present purposes is that the ERPs elicited by targets do not diverge from each other in the same way as they do for nontargets, thereby providing evidence in support of the view that participants prioritized retrieval processing relevant to source information associated with targets. In combination with the findings for the new-item ERPs, these data points provide insights into how participants complete exclusion tasks.

The principal difference across the two retrieval tasks was whether items encoded in the function or the drawing conditions were designated as targets. One way to complete this task is to rely equally on the recovery of information associated with targets as well as nontargets, and this might be regarded as an optimal strategy in so far as it involves taking both potentially relevant forms of source information into account to make target/nontarget discriminations (Wilding & Herron, 2006). The ERP data reported here suggest, however, that this strategy was not adopted by all participants in this experiment.

Relying on the recovery of information about targets as well as nontargets equates to using the same strategic retrieval operations irrespective of target designation; hence, ERPs elicited by correct rejections would not vary with target designation if this account was correct. In addition, when combined with the fact that the ERPs elicited by targets but not nontargets differed according to task demands, these findings are consistent with the view that participants who performed relatively better were prioritizing source information associated with targets over information associated with nontargets to complete the exclusion tasks. These data therefore converge with previous observations that the binary test distinction in an exclusion task can be made by prioritizing information associated with targets, and in the limiting case reducing the task judgment for old items to a decision about whether there is sufficient target-relevant source information available to make a target response (Dzulkifli & Wilding, 2005; Herron & Rugg, 2003).

Implications for Models of Memory Retrieval

Another perspective on the pattern of ERP findings in this experiment is that the greater the degree of similarity between the ways in which new items were processed, the lower the accuracy of target/nontarget judgments. An important theoretical implication that follows from this finding is that attempting to recover or assess multiple elements of information associated with a retrieval cue may not confer benefits on binary discriminations that outweigh those that are conferred by prioritizing some elements over others. From the perspective of the source monitoring framework (Johnson et al., 1993), this suggests for the first time that relying on some forms of information for source attributions is beneficial not only because they may be diagnostic for a particular source decision but also because there are costs associated with attempting to recover and/or assess multiple contents.

An additional insight for models of how retrieval occurs stems from consideration of the ways in which selective retrieval of task-relevant content might be implemented. Anderson and Bjork (1994) discussed this possibility in terms of cue-bias and target-bias. The term cue-bias refers to the generation of internal representations that will bias retrieval cues to interact with certain classes of targets (memory representations). The term target-bias refers to operations that will influence memory representations such that they are more likely to interact with some classes of retrieval cue than with others. The outcome of the engagement of either of these classes of process is an increase in the likelihood of recovery of task-relevant information, and distinguishing between these alternative accounts is not straightforward.

The findings for the ERPs elicited by new items in this experiment, however, support strongly the view that benefits in the recovery of task-relevant information do not reduce to target-bias operations. In tasks in which the same retrieval requirements are maintained throughout blocks of trials, it seems reasonable to assume that the cognitive processes supporting target-bias operations would be sustained, rather than being reimplemented trial-by-trial (Wilding, Fraser, & Herron, 2005). Consequently, the
differences between the new-item ERPs, and critically their correlation with response accuracy, implicate cue-bias processes in memory retrieval. The findings in this experiment do not speak to whether target-bias operations are also engaged, but one way to investigate this issue would be to acquire measures of sustained neural activity in retrieval tasks alongside the kinds of item-related measures that are reported here.

Finally, complementary behavioral evidence that unstudied test items are subjected to differential processing in line with task demands comes from a study reported by Jacoby, Shimizu, Daniels, and Rhodes (2005). They compared performance on recognition memory tasks across two groups of participants who had initially processed words under deep or shallow encoding conditions. Each group then completed two recognition memory tasks, in the second of which the “old” words were those that had served as “new” words in the first task. Old/new discrimination on the second task was superior for the deep encoding group, and this benefit was interpreted as a levels-of-processing effect (Crain & Lockhart, 1972). This account was based on the assumption that new as well as old words on the first recognition memory task were subjected to different retrieval processing operations according to the processing emphasis in the initial encoding phases (see also Rugg et al., 2000) and that the discrimination benefit on the second task arose as a result of the deeper processing of all test items in one group than in the other. These data points are therefore consistent with the view that the specific retrieval processing operations to which items are subjected during a memory test have consequences for their later memorability. The new claim that the present data license is that specific retrieval processing operations confer benefits on response accuracy at the time they are engaged. Whether these benefits remain when retrieval tasks require only recognition judgments (as in the study of Jacoby et al., 2005) is still to be determined.

To summarize, participants completed two retrieval tasks with different demands. The magnitudes of the differences between ERPs elicited by new items in the tasks were taken as an index of the degree of engagement of strategic retrieval processing operations, and the size of this index correlated positively with response accuracy. These findings indicate for the first time that the engagement of strategic retrieval processing operations indexed by ERPs benefits the accuracy of memory judgments. Moreover, the nature of the task demands under which these results were obtained, in combination with the ways in which ERPs elicited by old and by new items differed, offer insights into how strategic retrieval processing is implemented and how selective recovery of task-relevant information might come about.

References


### Appendix

#### Mean Number of Trials per Participant and Response Category

<table>
<thead>
<tr>
<th>Category</th>
<th>Designation</th>
<th>Accuracy group</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>High</td>
<td>Low</td>
</tr>
<tr>
<td>Correct rejections</td>
<td>Function</td>
<td>40 (25–47)</td>
<td>43 (25–48)</td>
</tr>
<tr>
<td></td>
<td>Drawing</td>
<td>38 (25–47)</td>
<td>43 (34–52)</td>
</tr>
<tr>
<td>Nontarget hits</td>
<td>Function</td>
<td>36 (26–45)</td>
<td>37 (21–49)</td>
</tr>
<tr>
<td></td>
<td>Drawing</td>
<td>36 (17–45)</td>
<td>38 (27–49)</td>
</tr>
<tr>
<td>Target hits</td>
<td>Function</td>
<td>35 (19–47)</td>
<td>34 (25–47)</td>
</tr>
<tr>
<td></td>
<td>Drawing</td>
<td>33 (16–53)</td>
<td>31 (18–39)</td>
</tr>
</tbody>
</table>

*Note.* Ranges are in parentheses.