*Memory & Cognition*
*2004, 32 (6), 932-944*

# Dissociating familiarity from recollection using rote rehearsal

IAN G. DOBBINS
*Duke University, Durham, North Carolina*

and

NEAL E. A. KROLL and ANDREW P. YONELINAS
*University of California, Davis, California*

Recollection-based recognition memory judgments benefit greatly from effortful elaborative encoding, whereas familiarity-based judgments are much less sensitive to such manipulations. In this study, we have examined whether rote rehearsal under divided attention might produce the opposite dissociation, benefiting familiarity more than recollection. Subjects rehearsed word pairs during the "distractor" phase of a working memory span task, and were then given a surprise memory test for the distractor items at the end of the experiment. Experiment 1 demonstrated that increasing rehearsal elevated the recognition rate for intact and rearranged pairs, but neither associative recognition accuracy nor implicit fragment completion benefited from rehearsal. The results suggest that rote rehearsal leads to a greater increase in familiarity than in recollection, and that the increase in observed familiarity cannot be attributed to effects of repetition priming. In Experiment 2, we tested item recognition with the remember/know procedure, and the results supported the conclusions of Experiment 1. Moreover, a signal detection model of remember/know performance systematically overpredicted rehearsal increases in remember rates, and this worsened when high-rehearsal items were assumed to be more variable in strength. The results suggest that rote rehearsal can dissociate familiarity from recollection at the time of encoding and that item recognition cannot be fully accommodated within a one-dimensional signal detection model.

The nature and number of processes underlying episodic recognition continues be widely debated. Advocates of a single-process approach contend that item recognition judgments are made on the basis of a single dimension of item familiarity or trace strength (Banks, 2000; Donaldson, 1996; Hirshman, 1995; Inoue & Bellezza, 1998). In contrast, dual-process theorists have argued that explaining recognition performance requires the postulation of two isolable and fundamentally different memory processes (Jacoby, 1991; Mandler, 1980; Tulving, 1985; Yonelinas, 1994), one registering the strength or familiarity of the item itself, and the other enabling the recall or recollection of contextual information spatiotemporally associated with the item. The latter recollection process is also typically assumed to play a more dominant role in paradigms such as recovery of source information and associative recognition judgments.

One body of research supportive of the dual-process approach to item recognition concerns introspective judgments that subjects are asked to make in order to qualify their mnemonic experiences (for a review, see Gardiner & Richardson-Klavehn, 2000). This remember/know procedure was developed by Tulving (1985) under the premise that recognition performance relies upon separate semantic and episodic memory systems. If, during testing, an item triggers recollection of contextual information consistent with first-person interaction, subjects are to respond "remember." In contrast, if they fail to recover such information, yet have a sense that an item is familiar, they are instructed to respond "know." Despite the fact that the distinction is purely subjective and sometimes difficult to convey to subjects, considerable research demonstrates that the report types can be dissociated in a principled manner, which suggests that the processes underlying the reports indeed represent qualitatively different retrieval phenomena.

On the basis of dual-process procedures such as the remember/know technique, we have postulated that the nature of encoding underlying recollection and familiarity might be differentially sensitive to depletion of attentional resources. For example, a common finding in dual-process studies is that increasing the "depth" of processing differentially increases recollective- versus familiarity-based responses (e.g., Gardiner, Java, & Richardson-Klavehn, 1996), suggesting that when attention is diverted from conceptual analysis of the materials, recollection suffers more than familiarity (or familiarity is unaffected). Conversely, the registration of information underlying judgments of item exposure frequency appears to be considerably less affected by attentional manipulation (Hasher & Zacks, 1984;

We thank Teresa Jones for help in collecting the data. Correspondence concerning this article should be addressed to I. G. Dobbins, Psychological & Brain Sciences, Duke University, Box 90086, 9 Flowers Drive, Durham, NC 27708 (e-mail: ian@psych.duke.edu).

Copyright 2004 Psychonomic Society, Inc.

Hasher, Zacks, Rose, & Sanft, 1987). To the extent that one assumes such frequency judgments to be predominantly based on item familiarity, this also suggests a differential dependence on attentional resources for recollection and familiarity during encoding. Further evidence supporting this view was provided by Gardiner, Gawlick, and Richardson-Klavehn (1994), who used a directed-forgetting paradigm in which the study word was followed by either a long or short delay prior to the cue to either "forget" or "learn" the items. They reasoned that subjects would engage only in "maintenance" rehearsal prior to the cue onset because elaborative encoding of the items would be premature until it could be determined whether the items should be committed to memory or forgotten. Thus the primary prediction was that subjects would engage in a greater degree of maintenance rehearsal in the long- versus the short-cue onset condition, which would be reflected by a greater subsequent *know* rate. Additionally, since the subjects were assumed to engage in elaborative encoding only following the memory cue, Gardiner et al. (1994) predicted that remember rates would not be influenced by the cue delay. These predictions were largely borne out. However, in order to assert that familiarity and recollection are dissociable at the point of encoding, it would be preferable to more directly manipulate the degree to which subjects attend to the materials, to examine retrieval paradigms other than the *remember/know* paradigm that are also assumed to segregate recollection and familiarity, and to potentially isolate trials on the basis of whether or not subjects diverted attention toward encoding the more elaborative aspects of study events.

To do this, we have based two experiments on a procedure used by Nairne (1983) in which attention was divided during encoding with a modified Brown–Peterson distractor paradigm (Glenberg & Adams, 1978; Rundus, 1970). For both experiments, subjects were engaged in a working memory span task for digit strings. After studying a string of digits, but before recalling the digit string, they were presented with a distractor task in which they were required to rehearse a pair of words for short or long durations. At the end of the experiments, there was a surprise recognition memory test for the words encountered during the distractor task.

In his original study, Nairne (1983) found that increased rehearsal under distraction led to similar increases in the subjects' recognition endorsements for both studied word pairs and rearranged pairs, demonstrating that the ability to recollect specific word pairings did not increase with rote repetition. In contrast, the fact that overall recognition increased indicates that some other form of memory did benefit from increasing rote repetition. One obvious possibility is that rote repetition increased the familiarity of the pairs, and thus the results may reflect a case in which familiarity is dissociated from recollection due to an attentional manipulation at encoding. However, another possible interpretation is that the increase in performance may have reflected some other form of memory such as repetition priming. We will explore this possibility below.

In Experiment 1, we have examined the ability of subjects to encode a specific form of contextual information assumed to require recollection, namely, forming an association between word pairs. The primary predictions follow Nairne (1983) in that we anticipated that subjects would be unable to benefit from rote rehearsal because their ability to discriminate intact and rearranged pairs would not improve with increasing exposure. However, we predicted that rehearsal performed under divided attention would increase the familiarity of individual items, resulting in the tendency to endorse item pairs with highly rehearsed constituents as "intact," regardless of whether or not they were in fact intact or rearranged (viz., a familiarity-induced response tendency). We extended Nairne (1983) by doubling the number of trials per condition, to increase power, and by examining the effects of rote rehearsal on an implicit measure of memory—word fragment completion—to determine whether increases in recognition memory may have reflected the contribution of repetition priming.

In Experiment 2, we have further examined the effects of rehearsal under divided attention by (1) expanding the scope of information that could contribute to our measure of contextual recollection, (2) more closely monitoring the efficacy of the divided attention manipulation, and (3) formally testing whether the pattern of responses could be accommodated within a signal detection model of the remember/know distinction. During Experiment 2, subjects again encoded word pairs under rote rehearsal conditions for either long or short durations. During subsequent testing, they were required to make an item recognition judgment accompanied by a remember/familiar report, and the prevalence of these reports was examined as a function of the number of previous rehearsals and the efficacy of the Brown–Peterson task in diverting attention during encoding. If recollection is heavily dependent on attentional resources necessary for elaborative encoding, subsequent remember reports should be more adversely affected by the attentional manipulation than familiarity-based reports.

## EXPERIMENT 1
## Associative Recognition

### Subjects

Thirty-six subjects, undergraduate students at the University of California at Davis, received extra credit for their participation. Two groups of subjects were given identical study conditions followed by either an associative recognition test ($n = 20$) or a fragment completion test ($n = 16$). One subject was removed from the fragment group for completing only two fragments of the entire test set.

### Materials and Procedure

The stimulus pool consisted of 150 words, 105 of which were drawn from the Toronto Word Pool (mean frequency = 11.7). The average length was 7.3 letters, and each item had on average 3.3 letters missing when presented as a fragment. Previous pilot work suggested a mean baseline completion rate of 16%. Computers were used for presentation of items and collection of responses, and subjects were tested individually. One hundred twenty of the words were evenly divided and randomly assigned to each of four conditions at the beginning of each experiment; the remaining 30 were used as baseline items for word fragment testing. The order of rehearsal trials (long vs. short rehearsal) and test item trials (intact vs. rearranged item pairs) was randomized by the computer.

Prior to study, all the subjects were told that they were participating in an experiment to test their short-term memory for digit strings similar to zip codes. They were instructed that after being shown a digit string, they would be required to engage in a distracting activity—word pair repetition—after which they would be immediately tested on their memory for the digits. They were not told that they would later be tested for associative recognition of the word pairs, and that the true purpose of the Brown–Peterson manipulation was to prevent elaborative encoding while repeating the word pairs (Nairne, 1983).

There were 60 Brown–Peterson study trials, half of which were short rehearsals (three repetitions), and the other half long (six repetitions). Each of these two list conditions contained 30 pairs of items, half of which were subsequently rearranged to serve as associative recognition lures. This resulted in four critical test conditions with 15 word pairs in each (short–intact, short–rearranged, long–intact, and long–rearranged). Each member of a rearranged test pair was always rehearsed the same number of times (short or long). Each study trial began with the appearance of five capital Xs above a windows "button" with the caption "digit span" appearing directly below. All items were presented in large 36-point font. When ready, the subject pressed the "enter" key, and each X was sequentially replaced by a digit every second. Following the fifth and final digit, there was a 1.8-sec pause followed by the first appearance of the to-be-rehearsed word pair, which replaced the digits, and was separated by two asterisks (e.g., DOG**STAPLER). The subjects were instructed to read the word pair aloud every time it appeared on the screen. Each word pair remained on the screen for 1.2 sec with a 1.2-sec interval between appearances. Pairs were shown three or six times according to study condition. Following the final appearance, a highlighted box appeared with the command underneath the message "please report," at which time the subjects used the keyboard to attempt to enter the correct digit string. Following this, a new trial began. The initiation of trials and entry of digits was self-paced.

During study, the computer recorded overall span accuracy.[1] Following the study of 60 pairs, each group of subjects was given further instructions and either an associative recognition task or a fragment completion task. For the associative task, either an intact or rearranged pair was presented with two statements: "Yes—Presented Together" or "No—Not Presented Together." Subjects could either click on the appropriate response or use the "Y" and "N" keys on the keyboard. Responses were self-paced. For the priming task, the fragment was shown above a highlighted box, and subjects were instructed to type in the first word that came to mind that would complete the fragment. No mention was made of the relation between tested and studied items, and subjects were instructed to click the "Next" button after entering a completion candidate or if they could not provide an answer. The tested, "old" fragments were drawn from the first items of all the pairs previously rehearsed. The "new" items were the remaining 30 of the 150-word pool used for the study.

## Results and Discussion

Data for the associative recognition and fragment completion groups were analyzed separately, whereas digit span accuracy during study was examined jointly. All tests were conducted at the .05 alpha level unless otherwise noted.

**Digit span**. Subjects did not differ in digit span accuracy as a function of number of rehearsals (three 16.31 of 30 vs. six 16.34 of 30) [$t(34) < 1$, n.s.]. Overall, it appears that the span task was likely sufficiently difficult to impair elaborative encoding for the read word pairs because the average span accuracy was 54%.

**Associative recognition**. The results are summarized in Figure 1A. A two-way repeated measures analysis of variance (ANOVA) conducted on the proportion of "intact" responses revealed a main effect of test item type (intact vs. rearranged) [$F(1,15) = 24.83$, $MS_e = .010$] showing that the subjects were more likely to correctly respond intact (.61) than to err (.42), a main effect of rehearsals (3 vs. 6) [$F(1,15) = 20.68$, $MS_e = .028$] demonstrating that the subjects were more likely to respond *intact* to the frequently rehearsed pairs (.58) than to the less frequently rehearsed pairs (.45), and most importantly, no evidence of an interaction between the two [$F(1,15) = 0.23$, $MS_e = .024$]. This essentially replicates the main findings of Nairne (1983). Given the very small $F$ value for the interaction, and the increased power over Nairne's study (more than twice the items per subject per condition), this result suggests that over the range considered, increasing rote rehearsal under divided attention does not appreciably affect associative encoding. However, the subjects registered the prior occurrences of the individual items, leading to the systematic increase in likelihood of positive responses as a function of number of rehearsals (irrespective of actual item arrangement of the test pair).

**Fragment completion**. A paired $t$ test revealed that fragment completion rates did not significantly differ as a function of the number of rehearsals; the subjects were as likely to complete less frequently rehearsed items (.27) as the more frequently rehearsed items (.29) [$t(18) = 0.89$]. For both rehearsal conditions, studied items were more likely to be completed than new ones (.14), and $t$ values were 4.56 and 5.73, respectively (Figure 1A). These findings are critical in showing that the increase in recognition memory with repetition could not have reflected a form of repetition priming, at least as measured by fragment completion.

Overall, these data suggest three general conclusions. First, consistent with previous research, the data demonstrate that item and associative information are separable (e.g., Hockley, 1991; Nairne, 1983). Second, the data suggest that they are differentially sensitive to diverting attention during study, with the formation of associative information being potentially limited by attentional manipulation. Finally, the data indicate that the increased item recognition, that is, the general tendency to respond *intact* to more frequently seen items, did not result from perceptual priming mechanisms that support fragment completion; this indicates that the sense of familiarity for highly rehearsed items was not driven by perceptual fluency (Wagner & Gabrieli, 1998). The results suggest that rote repetition under divided attention did not lead to an appreciable increase in the ability to recollect which items were paired with which, but it did increase the familiarity of the individual items.

However, while these data suggest that item and associative information are distinct, they do not in themselves rule out a signal detection account of recognition. For ex-

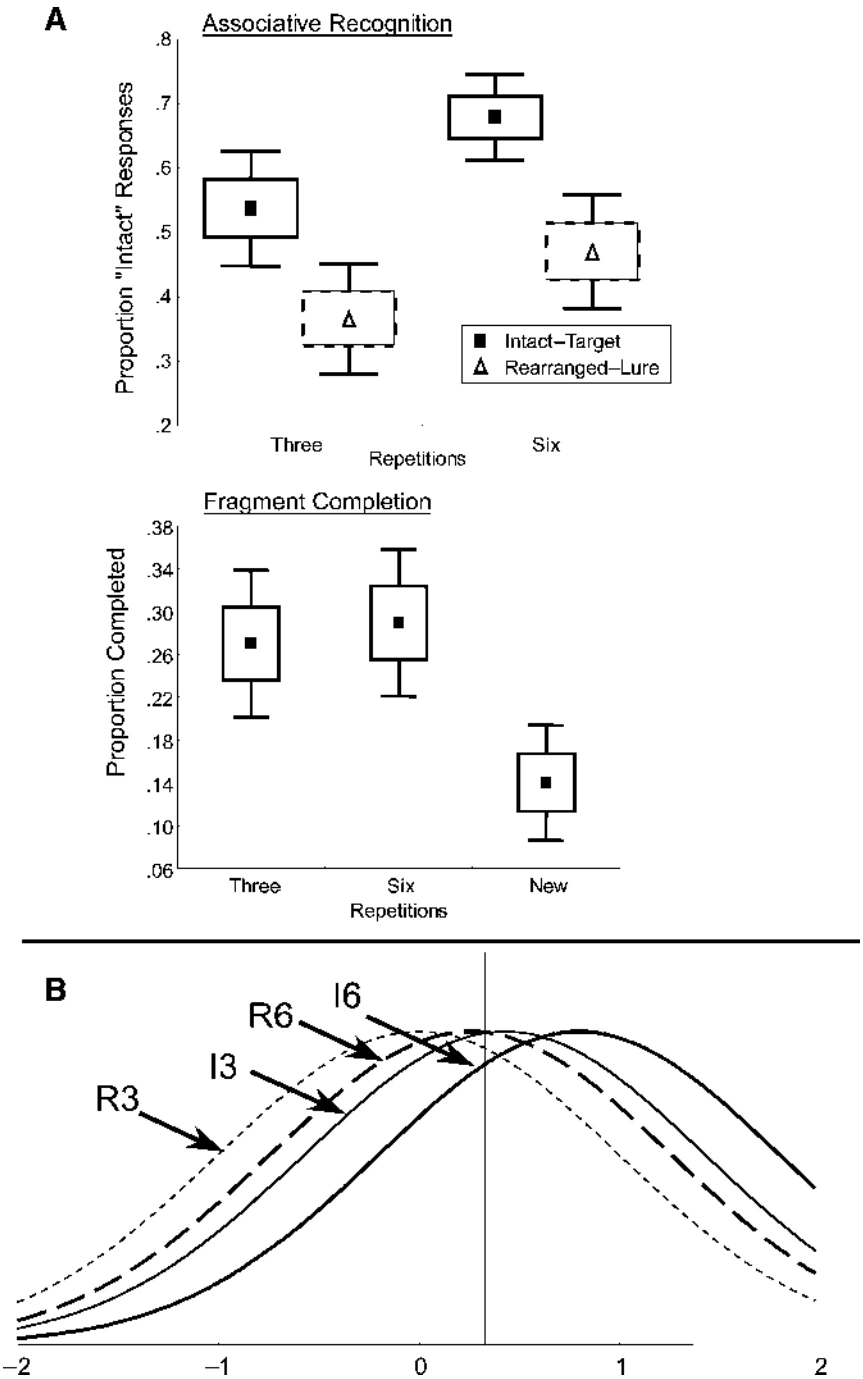


**Figure 1. Panel A illustrates mean associative recognition and fragment completion performance. Top section shows proportion of intact responses as a function of rehearsals and actual test pairing (intact or rearranged). Bottom section shows fragment completion rates for items rehearsed three times or six times, or new. Each box represents one standard error of the mean; boxes plus whiskers represent two standard errors of the mean. Panel B illustrates one of many potential signal detection models that mirror the dissociation between associative recognition accuracy and item recognition.**

ample, the data could be described entirely within a signal detection framework by assuming that item and associative information jointly contribute to a global familiarity signal. Figure 1B demonstrates one such potential model in which the variance of all four distributions is assumed to be equal. If, in addition to the means, one allows the relative variances of the distributions to differ, a multitude of other arrangements would also yield the observed proportions of intact and rearranged responses. Thus a single dissociation between item and associative memory does not necessarily render the signal detection model invalid.

However, there are at least two reasons to favor the dual-process interpretation. First, the shapes of associative recognition receiver operating characteristics (ROCs) have been shown to be inconsistent with signal detection theory in isolation (Kelley & Wixted, 2001; Yonelinas, 1997). Second, the signal detection model does not predict a priori that associative encoding should be more severely disrupted by diverting attention than should item encoding. That is, although the model is able to describe the proportions post hoc, nothing within the model would have led to the initial prediction that the two should dissociate. In contrast, such predictions more naturally arise from dual-process theory because recollection is assumed to differentially benefit from increased elaborative or conceptual processing, and such processing is likely severely hampered by diverting attentional resources during encoding. Nevertheless, a more compelling case for assuming qualitatively different processes could be made if the signal detection model failed to accommodate performance when retrieval was restricted to single-item recognition. This point will be addressed in the next experiment.

## EXPERIMENT 2
## Old/New Recognition

Despite the high prevalence of dissociations of *remember* and *know* reports in the literature, many forms of these dissociations can be accommodated under a single-process signal detection model by simply assuming two different response cutoffs, one used to determine that an item is old and a second, more stringent, criterion reserved for making remember responses (Donaldson, 1996; see also Hirshman, 1998; Hirshman & Henzler, 1998; Inoue & Bellezza, 1998). Figure 2 demonstrates the model and shows how increasing the old item strength across conditions can easily result in modulation of remember rates with know rates left largely unaffected, increased, or decreased, depending upon the starting position of the old/new and remember response criteria (left and right side of the rectangles in Figure 2, respectively). This happens because the know rate arises from the proportion of the target distribution trapped between the two criteria. This proportion can rise, remain unchanged, or fall as the distribution moves rightward past the fixed criteria as a result of the strengthening.

In contrast, any manipulation that strengthens overall performance necessarily increases the observed remember rate if the criteria remain fixed. It should be noted that these constraints only apply under conditions in which the criteria are assumed to remain fixed across the strong and weak conditions. However, this assumption must be adopted if items of different strengths are mixed within a single test list because under such conditions, subjects have no additional information, other than strength itself, that would enable systematic adjustment of the response criteria.

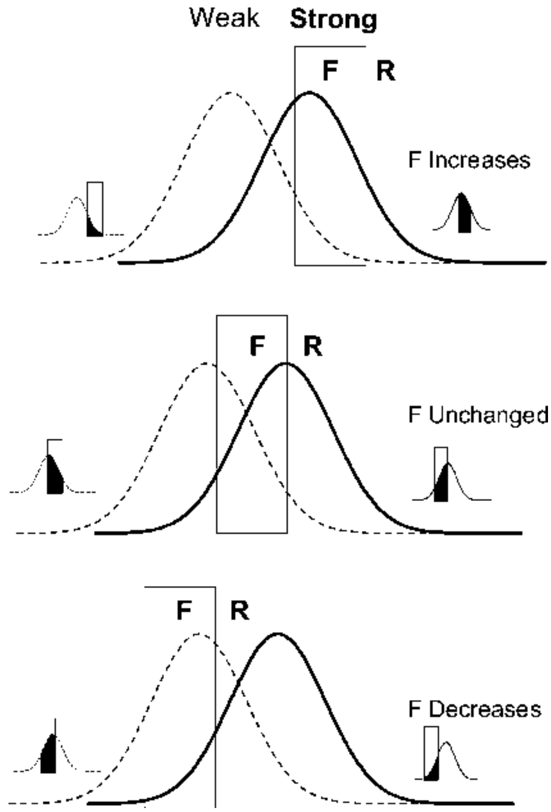


**Figure 2. Graphic representation of two-criterion signal detection for *remember* (R) and *familiar* (F) responses during old/new recognition. Left and right sides of the rectangle denote the old/new and remember response criteria, respectively. The proportion of *familiar* responses is constrained to fall in this region. That is, the item is sufficiently familiar to be called "old" but not familiar enough to exceed the remember criterion. Distributions on the left represent the "weak" or low-rehearsal condition, and those on the right the "strong" or high-rehearsal condition. Regardless of initial position of criteria, strengthening results in an increase in the proportion of the distribution falling above the remember criterion. Starting position of criteria affects whether strengthening increases (top panel), leaves unchanged (middle panel), or decreases (bottom panel) the observed proportion of familiarity-based responses. Small insets illustrate the constrained areas yielding *familiar* responses.**

If our interpretations of the results from Experiment 1 were correct, and rote repetition increases familiarity more than recollection, this manipulation when used in conjunction with a remember/know test should prove problematic for the signal detection model because it would produce the pattern of results just described: an increase in overall performance but a relatively minor effect or no effect on the proportion of *remember* responses. One of our aims in Experiment 2 was to test this potential weakness of signal detection theory. Another was to verify the results of Experiment 1 with a different method of operationalizing recollection and familiarity.

Encoding during Experiment 2 differed from that in Experiment 1 in two ways. First, the range of rehearsal was increased by contrasting two versus eight rehearsals.

Second, the computer program kept track of the trials on which the subjects erred or successfully completed the Brown–Peterson task, so that subsequent memory performance could be conditionalized by the efficacy of the distractor task (Naveh-Benjamin & Jonides, 1984). It is important to note that from a signal detection perspective, separating the encoding trials on the basis of the outcome of the Brown–Peterson task should not have any particularly deleterious effect on the fit of the two-criteria signal detection model. Such a split essentially just separates strong from weak encoding trials, both of which are presumably subject to the same principles. Our initial assumption, which was supported by a subsequent difference in overall recognition accuracy between correct and incorrect encoding trials, was that error trials would represent cases in which attention was inappropriately diverted toward the rehearsed words. This could occur for several reasons. For example, initial difficulty in reading the words may lead subjects to shift greater attentional resources to the words, thereby failing to sufficiently maintain the digits in memory. Similarly, some words may be more salient as a function of characteristics such as low word frequency, emotional, or personally relevant content. Despite these considerations, from a strength-based perspective, the words are either more strongly, more weakly, or similarly encoded across correct and incorrect trials. Whether this arises from differences in the mechanism(s) underlying attention or a systematic difference based on some property of the items is largely irrelevant because these factors can only affect the strength and perhaps variance of encoding of the words from a signal detection perspective.

### Subjects

Two separate groups of 16 undergraduate students at the University of California at Davis participated as subjects for extra credit. An additional 11 undergraduates were discarded for reasons described below. Because the accuracy of the groups did not differ, they were combined into one analysis. All subjects were native speakers of English.

### Materials and Procedure

The materials and procedure were as in Experiment 1, with the following exceptions. Two hundred sixty-four low-frequency English words were selected from the Oxford Dictionary Electronic Database with a mean Kučera–Francis frequency of 11.9 (Kučera & Francis, 1967). The length of the words ranged from 6 to 8 letters with an average length of 7.11 letters. These words were randomly assigned to four lists, 66 words per list. For each subject, the words from one of these lists were used in the two-rehearsal condition, the words from another of these lists were used in the eight-rehearsal condition, and the words from the other two lists were used as foils. Across subjects, each list was used equally often in all conditions. In addition, 72 five-digit numbers were constructed such that none of the numbers contained an obvious pattern.

Each trial began with a "+" sign in the middle of the computer monitor. When ready, the subject pressed the space bar, and the five-digit number appeared sequentially on the screen. One second after the appearance of the fifth digit, the entire number disappeared from the screen and was replaced by two words (e.g., SCORPION ** TRACTOR). The subjects had been instructed to say the words out loud each time they appeared on the screen. For each repetition, the duration of the word pair was 1.2 sec with a .8-sec interstimulus interval. The word pairs were presented twice before asking the subject to recall the digits on half of the trials and presented eight times before asking the subject to recall the digits on the other half of the trials. At the bottom of the screen, the subjects were given a running total of how many numbers they had remembered correctly (i.e., all five digits in the correct order) in each of the two- and eight-repetition conditions. The purpose of this running feedback was to emphasize the importance of performing well on the number-recall task.

In the recognition test phase, one word appeared on the screen during each trial, and subjects were instructed that this was a test of their ability to remember which words had been used as distractors during the number-memory phase of the experiment. They were told,

> On each trial, choose one of three keys to indicate your judgment about the word on the screen: R = You remember the word—i.e., you actually remember some detail about having seen or said this word during the number-memory phase of the experiment. F = You do not remember the word having been in the number-memory phase of the experiment, but the word seems so familiar that you feel it must have been in the previous phase of the experiment. N = You do not remember the word, nor do you find it very familiar, but, rather, believe it to be a new word—i.e., a word not in the previous phase of the experiment.

In addition, subjects were told, "You should not use the 'R' key unless you truly remember the word. That is, while fully half of the words will be old, you should use the 'F' key to indicate 'old' when you just have the feeling of familiarity."

The words from the first six trials (three with two repetitions and three with eight repetitions) and the last six trials (three with two repetitions and three with eight repetitions) of the number-memory task were not scored, but were instead combined with 12 lures taken from the distractor list, and presented as test trials and used as practice trials to ensure that the subjects understood the instructions. For example, the first time the subjects used any of the three response keys, they were asked to explain their choice of response. In addition, if subjects responded "R" to a lure during these first 48 trials, they were reminded that they should use the "R" response only when they truly remembered the word and, that while false memories were possible, they were unlikely. At the end of these 48 trials, they were told how many times they had judged a word to be old (i.e., via the "R" and "F" keys) and how many times they had judged a word to be new ("N"), and they were reminded that half of the words were, in fact, old words. One subject was replaced because of never using the familiarity response during the critical memory trials, and 10 others were replaced because they had fewer than six 8-repetition trials with correct digit span responses.

## Results and Discussion

Accuracy scores from the Brown–Peterson trials indicated that the digit span task was an effective attentional distractor. Subjects successfully completed on average 62% (37 of 60) of the two-rehearsal trials, whereas for the eight-rehearsal condition they completed 40% (24 of 60). On this basis, the analysis was segregated into successful versus unsuccessful Brown–Peterson trials, under the hypothesis that the rehearsal manipulation would have different effects on remember and familiar performance, depending on the degree to which subjects were distracted from elaboratively encoding the words. Collapsed across rehearsal levels when subjects were accurate on the Brown–Peterson task, accuracy as indexed by $d'$ was lower (1.08) than when they were inaccurate on the Brown–

**Table 1**
**Familiar and Remember Response Proportions as a Function of Number of Rehearsals and Accuracy on Dual-Task Performance**

| | Rehearsals | | | | |
|---|---|---|---|---|---|
| | Two | | Eight | | Rehearsal |
| | Familiar | Remember | Familiar | Remember | Increase |
| Correct BP | .26 [.33] | .24 | .35 [.49] | .30 | .15 |
| Incorrect BP | .26 [.39] | .34 | .25 [.48] | .49 | .14 |
| False alarms | .18 | .03 | | | |

Note—BP, Brown–Peterson task. Incorrect BP data was computed only for subjects with 14 or more incorrect trials in the low-rehearsal condition ($n$ = 22). Bracketed proportions are dual-process estimates of familiarity [familiar/(1 − remember)]. Rehearsal increase is the net gain in hit rate proportion as a function of increased rehearsal (eight vs. two).

Peterson task (1.27) [$t(31) = 2.07, p < .05$]. Thus the items were more effectively encoded during the unsuccessful Brown–Peterson trials.

The next section analyzes the raw response proportions and determines whether these proportions can be accommodated within the two-criteria model of Donaldson (1996). In the following section, the familiarity proportions are corrected assuming a dual-process model, and the data are reanalyzed to determine whether recollection appears more disrupted than familiarity by attentional diversion, as predicted under dual-process theory.

## Raw Proportions

The raw remember and raw and dual-process corrected familiar proportions for test items are shown in Table 1 as a function of accuracy on the Brown–Peterson task and the number of encoding rehearsals.

## Correct Brown–Peterson Trials

Response type (familiar vs. remember) and rehearsal (two vs. eight) were analyzed with a two-way ANOVA on the raw proportions. Because these data share a common false alarm rate, increases in the proportions are monotonically related to accuracy. There was a main effect of rehearsal [$F(1,31) = 49.12, MS_e = .004, p < .001$], demonstrating that correct responses increased with rehearsal, but no main effect of response type or interaction of response type and rehearsal. Although the interaction was not significant, we further analyzed the changes in familiar and remember rates as a function of rehearsal because the two-criteria signal detection model specifically predicts a selective increase in remember and not familiar rates under these conditions. Post hoc pairwise comparisons (Tukey's HSD) indicated that the familiar rate significantly increased with rehearsal (.26 vs. .35, $p < .01$), whereas the remember rate did not (.24 vs. .30, $p = .08$). Overall, the analysis suggests a clear increase in familiar rates, with a borderline increase in remember rates.

To determine whether the observed increases were consistent with the two-criteria signal detection model, we contrasted the observed increases with the predicted increases under the model in both the aggregate and individual subject data. Figure 3A graphically presents the predictions of the model for the aggregate data in which we used the hit and remember rates under two rehearsals to estimate the location of the old/new and remember re-

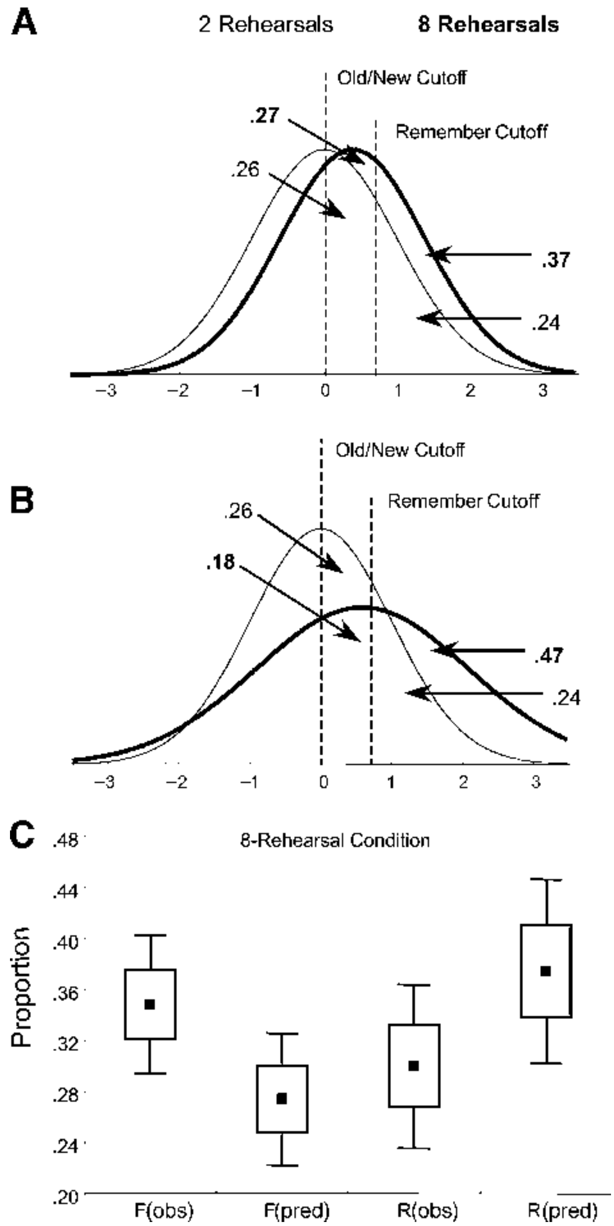


**Figure 3. Predictions of the two-criterion signal detection model for correct Brown–Peterson data. Panel A illustrates the predicted proportion of *familiar* and *remember* responses for the eight-rehearsal condition assuming equivalent standard deviations of the two- and eight-rehearsal item distributions. Panel B demonstrates the predictions assuming the eight-rehearsal distribution is 1.5 times more variable than the two-rehearsal condition. Heavy curve is for eight rehearsals; light, for two. Panel C shows the mean difference between predicted (pred) and observed (obs) data for individual subjects in the eight-rehearsal condition. F is familiarity, and R is *remember* rate.**

sponse criteria. Following this, we calculated the increase in overall strength necessary to accommodate the increased hit rate for the items rehearsed eight times (viz., the difference in $d'$ values across the low- and high-rehearsal conditions). This allowed us to calculate the predicted familiar and remember rates and to compare these with what was actually observed in the eight-rehearsal condition. As can be seen, the two-criteria model appears to predict little, if any, increase in the familiarity rate but a large increase in *remember* responses, a pattern potentially opposite to what was observed in the ANOVA above. More specifically, the model predicts that familiar rates (the area between the two criteria) should be essentially unchanged by rehearsal (.26 to .27) when in fact they reliably increased by .09. In contrast, the model predicts a notable increase in the remember rates (.24 to .37), whereas the increase in the actual data was more slight (.24 to .30). One method of improving the fit of the signal detection model to ROCs has been to allow the studied item distribution to become more variable than the new item distribution. The underlying logic is that the variance increase results from the learning process, and this assumption is an essential part of the learning algorithm for several prominent computational memory models (Gillund & Shiffrin, 1984; Glanzer, Adams, Iverson, & Kim, 1993; Hintzman, 1984). Applied to the current data, the natural extension of this approach is to predict that studying the items eight times leads to more variability than studying them only twice. However, as Figure 3B demonstrates, this assumption increases, not reduces, the discrepancy between the signal detection model and the actual data. For illustration, in Figure 3B we assumed that the standard deviation of the eight-rehearsal condition was one and one half times that of the two-rehearsal condition and again placed this distribution in the appropriate location in order to match the overall hit rate of the eight-rehearsal condition. As can be seen, the remember rate is now overpredicted by an even greater margin, and the anticipated familiar rate is lessened as a result of the distribution "flattening," the exact opposite of what is needed to accommodate the actual data.

In order to fit individuals, we performed the same procedure that was conducted on aggregate data on each subject's data. Namely, the old/new and remember criterion locations were determined with the two-rehearsal proportions, and these criteria were applied to the eight-rehearsal distribution implied by the increase in the hit rate in order to calculate the anticipated familiar and remember rates. The pattern in the aggregate data was borne out in the individual data, the model systematically underpredicted the increase in the familiar rate and hence overpredicted the increases in remember rates [$t(31) = 4.24, p < .001$] (Figure 3C). Only one $t$ value is reported because an underprediction in one response type necessarily translates to an equal overprediction in the other (i.e., the proportions must add up to the observed overall hit rate).

We conducted a final test of the two-criteria model by comparing the accuracy of remember versus overall responding for each rehearsal condition. As shown in Figure 2, remember and overall responses are assumed to be governed by the same underlying distributions. If so, accuracy estimated from the overall hit and false alarm rates should not systematically differ from that based on the hit and false alarm rates for remember responses because the two accuracies arise from a common signal detection process (i.e., $d'$ is unaffected by the placement of the response criteria) (Donaldson, 1996). With $d'$ as a measure of accuracy, the data demonstrated significantly higher accuracy for remember compared with overall responding in the two-rehearsal condition [$t(26) = 4.85$, $p < .001$], whereas accuracy did not statistically differ in the eight-rehearsal condition [$t(26) = 1.29$, n.s.]. The data from 5 subjects were not analyzed because they did not commit any remember false alarms.

### Incorrect Brown–Peterson Trials

Overall, subsequent accuracy was greater for trials on which subjects committed an error during the earlier Brown–Peterson task versus those on which they were accurate (Table 1), and this suggests that subjects diverted attention toward the words during the retention interval. On those trials in which subjects did commit digit span errors, the average number of correctly recalled digits was 2.79 for the low-rehearsal condition and 2.34 for the high-rehearsal condition. Response type (familiar vs. remember) and rehearsal (two vs. eight) were analyzed with a two-way ANOVA. There were main effects of rehearsal [$F(1,21) = 23.06$, $MS_e = .004$, $p < .001$] and response type [$F(1,21) = 7.44$, $MS_e = .08$, $p < .05$] and a rehearsal by response type interaction [$F(1,21) = 15.07$, $MS_e = .009$, $p < .001$]. Post hoc pairwise comparisons suggested that the interaction occurred because rehearsal significantly increased the remember rate (.34 vs. .49, $p < .001$) but did not influence the familiarity rate (.26 vs. .25, $p = .96$). Figure 4A graphically presents the predictions of the two-criteria model for the aggregate data and demonstrates that the predictions match the results of the ANOVA, unlike the case of correct Brown–Peterson trials. More specifically, the aggregate model predicted little change in the familiarity rate (.26 to .25) and a notable increase in the *remember* rate (.34 to .49). The accuracy of the predictions extended to the 22 subjects with sufficient error trials, with the predicted familiar and remember rates not significantly differing from what was observed [$t(21) < 1$] (Figure 4B). Although the familiarity and remember proportions of the high-rehearsal data are in line with the two-criteria model predictions, the accuracy of *remember* performance and that of overall performance were not similar, as expected under the model. *Remember* accuracy was higher than overall accuracy in the two-rehearsal condition [$d'$ of 1.39 vs. 1.17; $t(19) = 2.53, p < .05$] and the eight-rehearsal condition [$d'$ of 1.84 vs. 1.59; $t(19) = 2.73, p < .05$]. The low-

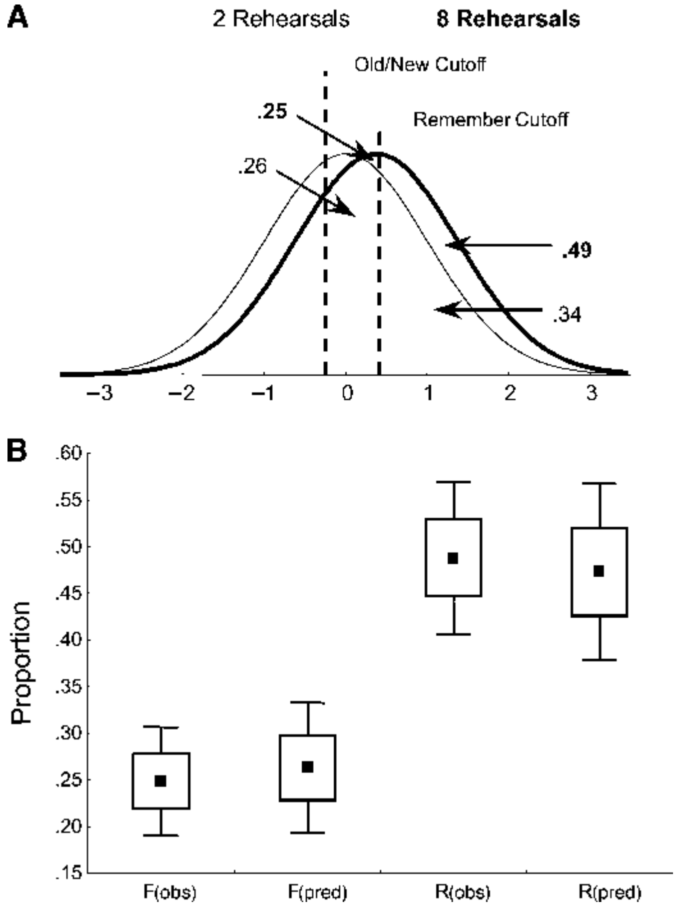


**Figure 4. Predictions of the two-criterion signal detection model for incorrect Brown–Peterson data. Panel A illustrates the predicted proportion of *familiar* and *remember* responses for the eight-rehearsal condition assuming equivalent standard deviations of the two- and eight-rehearsal item distributions. Heavy curve is for eight rehearsals; light, for two. Panel B shows the mean difference between predicted (pred) and observed (obs) data for individual subjects. F is familiarity, and R is *remember* rate.**

ered degrees of freedom in the latter statistical tests resulted from removing subjects without sufficient error trials and those who did not commit false alarms for remember responses.

## Estimating Recollection and Familiarity

Under the dual-process model, the observed proportions of *remember* responses are typically considered direct estimates of recollection, and therefore recollection is assumed to be captured by the observed remember rates listed in Table 1. In contrast, because subjects were instructed to respond *familiar* whenever the item was sufficiently familiar in the absence of recollection [$F(1-R)$], familiarity was estimated as the probability of a *familiar* response independent of whether or not the item was remembered [viz., $F/(1-\mathrm{R})$]. This correction is important under dual-process theory because conditions that promote remembering will necessarily constrain the potential for familiarity-based responding (Yonelinas & Jacoby, 1995). Without such corrections, one would be led to the nonsensical conclusion that operations that greatly facilitate remembering necessarily reduce the familiarity of the items.

In order to analyze the effects of rehearsal and Brown–Peterson outcome on the process estimates, a three-way ANOVA was conducted with factors of rehearsal (two vs. eight), outcome (Brown–Peterson correct or incorrect trial), and process (remember or familiar). The ANOVA revealed main effects of outcome with correct Brown–Peterson trials leading to lower subsequent hit rates than did incorrect trials [.35 vs. .42; $F(1,21) = 16.97$, $MS_e = .013$, $p < .001$] and of rehearsals [.33 vs. .45; $F(1,21) = 41.26$, $MS_e = .015$, $p < .001$] with higher hit rates for more frequent rehearsal. The data also displayed a two-way outcome × process interaction [$F(1,21) = 7.77$, $MS_e = .015$,

$p < .05$]. This interaction occurred because whereas mean familiarity estimates were similar across correct and incorrect Brown–Peterson trials, recollection significantly increased following incorrect trials (Tukey's HSD) (Figure 5A). Finally, there was a significant three-way rehearsal × outcome × process interaction [$F(1,21) = 8.49$, $MS_e = .010$, $p < .01$]. Post hoc interaction contrasts suggested that this occurred because, although familiarity and recollection did not interact following incorrect Brown–Peterson trials [$F(1,21) = 1.5$, $MS_e = .014$, n.s.], following correct Brown–Peterson trials process and rehearsal interacted [$F(1,21) = 7.13$, $MS_e = .014$, $p < .05$], such that only familiarity increased with rehearsal; recollection did not significantly increase with greater rehearsal (Tukey's HSD) (Figure 5B).

## DISCUSSION

The present data support our initial hypothesis that the encoding of information supportive of later recollection is more easily disrupted by diverting attention than by the encoding of that supporting future familiarity-based judgment. However, it is useful to consider whether or not the data are also consistent with a single-process, signal detection account. In the case of remember/know (familiar) paradigms, the most widely cited model is the two-criteria signal detection model (Donaldson, 1996; Hirshman & Henzler, 1998; Hirshman, Lanning, Master, & Henzler, 2002; Inoue & Bellezza, 1998). Indeed a recent search of the Social Science Citation Index indicated that the Donaldson (1996) article had been cited 119 times. Critically, the two-criteria model likely presents the only way to characterize the remember versus familiar distinction while remaining within the basic unidimensional signal detection framework, and therefore determining its viability is important.

The present data indicate that the two-criteria model is deficient for several reasons. First, in the case of the correct Brown–Peterson trials, the two-criteria model systematically underpredicted the rehearsal gains in famil-

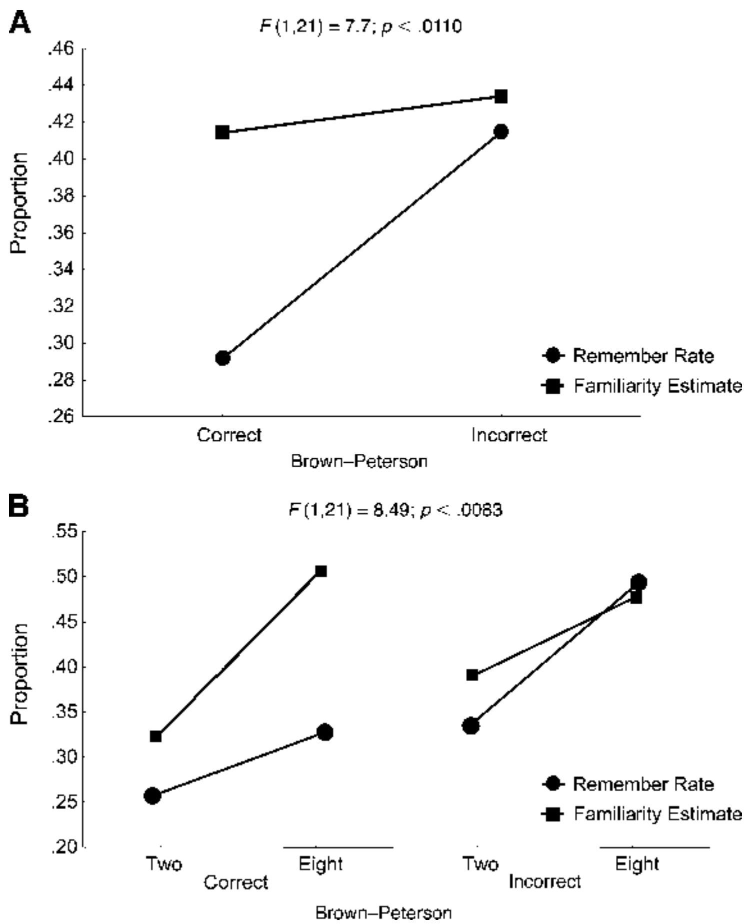


**Figure 5. Graph of the two- and three-way interactions involving dual-process estimate of familiarity and the observed remember rate. Panel A shows the interaction of the Brown–Peterson outcome and the estimates (remember vs. familiar). Panel B shows the interaction of the Brown–Peterson outcome, process estimate, and level of rehearsal (two vs. eight).**

iarity while overpredicting those in *remember* responses, both at the aggregate and individual subject levels. This misfit was in fact worsened by assuming that the variance in strength increased with rehearsal, which interestingly represents one of the few cases in which the unequal variance assumption does not improve the fit of the signal detection model. Indeed, in order to improve the model fit we would have had to assume that the eight-rehearsal condition was less variable than the two-rehearsal condition, the opposite of what has been typically been called for under signal detection accounts. In the case of incorrect Brown–Peterson performance, the two-criteria model made accurate predictions regarding familiar and remember rates but a critical assumption of the model—the equivalence of overall recognition and remember rate accuracy estimates—was violated for both levels of rehearsal. This assumption was also violated for the low-rehearsal condition during correct Brown–-Peterson performance where again accuracy was higher for remember than for overall responding. It is important to note that the segregation of items dependent on the outcome of the Brown–Peterson task should have no major significance from a signal detection perspective. That is, successful or unsuccessful performance on the maintenance task can only affect strength. Thus, conditionalizing performance based on outcome is in principle no different from any other strength manipulation. In contrast, under dual-process theory, we anticipated that the more effectively attention was diverted from the materials, the more unlikely conceptual analysis would be, and this would limit gains in subsequent remembering as a function of rehearsal opportunity.

This predicted dissociation was supported on two levels. First, there was an overall increase in remembering during incorrect versus correct Brown–Peterson trials that did not occur with familiarity estimates (Figure 5A). Second, within the correct Brown–Peterson trials, familiarity significantly increased with rehearsal, whereas gains in remembering were reduced, presumably because subjects were devoting considerable attention to the maintenance task (Figure 5B). Although the observed dissociations support a dual-process interpretation, it should be noted that the predicted null effect of rehearsal on *remember* rates during correct Brown–Peterson trials was only marginally supported. Although post hoc testing did not demonstrate a significant increase in remembering as a function of rehearsal in this condition, the *remember* rate was numerically higher during high versus low rehearsals, and the ANOVA did not reveal a significant interaction of the response types as a function of rehearsal. One explanation for this may be that *remember* and *know* rates are not process-pure estimates of recollection and familiarity. For example, the dual-process signal detection model of Rotello and colleagues (Rotello, Macmillan, & Reeder, 2004) avoids the process-pure assumption, yet unlike the two-criteria signal detection model examined here, it assumes that recognition is composed of separate and orthogonal global and specific memory mechanisms (viz., dual processes) that are either summed or differenced in order to arrive at recognition and remember judgments, respectively. Alternatively, one could adhere to a process-pure interpretation but acknowledge that the Brown–Peterson design may not completely eliminate the potential for elaborative encoding. This does not seem entirely unreasonable, given that subjects rehearsed the items eight times and that individuals with exceptional working memory spans or easily chunked digit strings may have enabled a small degree of elaborative encoding to occur on some trials. Regardless of whether one does or does not make the process-pure assumption, however, the present dissociations clearly suggest that multiple processes are at play and that the pattern in the correct Brown–Peterson performance challenges the most popular single-process interpretation of remember/know data.

In the case of Experiment 1, a one-dimensional signal detection model was able to provide a post hoc fit of the data potentially accommodating the dissociation between associative and item recognition. However, it is not clear how one would integrate both the associative recognition and remember/familiar data sets within the same signal detection framework, and perhaps more importantly, whether the outcome of one would lead to the prediction of the other. In contrast, in addition to the remember/familiar prediction, the dual-process model would suggest that associative encoding of word pairs should also be severely hampered by diverting attention. Notably, trials during the associative recognition study were not conditionalized on the basis of the outcome of the Brown–Peterson task. Given that Experiment 2 suggested that during inattention (and hence errors on the Brown–Peterson task) there should be large increases in associative encoding, it may be surprising to see that associative recognition accuracy did not benefit from rehearsal since we were unable to remove the incorrect Brown–Peterson trials. We can think of two potential post hoc explanations why this might be the case. First, the encoding required to form a successful association between two words is highly specific and likely more demanding than that required to form an association between a single item and any aspect of the experimental context. Thus two items that both evoke recollection may nevertheless be poorly associated and given this, it is perhaps reasonable to assume that associative encoding will accrue at a slower rate than that which supports more general recollections (i.e., remember reports). Similarly, associative encoding may be dependent on not only attending to conceptual aspects of each item, but on deliberately contrasting or relating each with the other (viz., relational encoding). Since subjects had no reason to anticipate an associative recognition test, it is possible that even when they diverted attention toward the pairs, they still failed to engage in such relational encoding.

Overall, these results complement a large body of literature demonstrating that many variables including levels of processing and generate/read manipulations have large effects on recollection while leaving familiarity relatively unaffected (for a review, see Yonelinas, 2002). They are also consistent with research comparing rote and elaborative rehearsal across different types of retrieval tasks (for a review, see Greene, 1987). The failure to fully accommodate basic item recognition within the signal detection framework is noteworthy because several current approaches assume that old/new recognition can be fully accommodated within a basic signal detection process (e.g., Banks, 2000; Donaldson, 1996; Hilford, Glanzer, Kim, & DeCarlo, 2002; Hirshman, 1998; Hirshman & Henzler, 1998; Inoue & Bellezza, 1998). For example, following a brief review of the findings from experiments using ROCs, Hilford et al. concluded, "These findings support a normal signal detection model as fully adequate for explaining item recognition memory" (p. 496). The current data demonstrate that this is not the case and make it likely that an effective model of old/new recognition either needs to recruit an additional process (e.g., dual-process theories) or perhaps abandon the assumption of normal distributions in favor of more complicated distributional assumptions (e.g., DeCarlo, 2002).

One aspect not addressed by these data is why the division of attention via a working memory maintenance task so effectively prevents increases in the type of encoding necessary for future recollective experience. A possible answer to this question can be seen in recent functional brain imaging research. For example, Wagner and others have demonstrated that activity in regions of the prefrontal cortex is predictive of whether or not subjects will later remember or forget studied items (e.g., Brewer, Zhao, Desmond, Glover, & Gabrieli, 1998; Davachi, Maril, & Wagner, 2001; Nyberg, 2002; Wagner, Maril, & Schacter, 2000; Wagner et al., 1998). In addition, this activity is particularly predictive of whether or not items will later be accompanied by remember reports and often does not distinguish between items identified in the absence of recollection or endorsements based on low confidence. Importantly, many of these same cortical regions are often implicated in working memory studies where subjects must maintain or rehearse a set of items over a delay period (e.g., Clark & Wagner, 2003; Schumacher et al., 1996; Smith, Jonides, Marshuetz, & Koeppe, 1998), in addition to studies that require the controlled retrieval or selection of semantic information during study (Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997; Wagner, Paré-Blagoev, Clark, & Poldrack, 2001). Thus the functional brain imaging literature indicates that requiring the maintenance of verbal information during the distractor task of the present study potentially precluded the participation of prefrontal cortex regions critical for controlled semantic evaluation, which are also predictive of later episodic memory and apparently critical in furthering the type of mnemonic traces necessary for future recollective experience. In short, the brain imaging and current data suggest that the Brown–Peterson task limits the ability to meaningfully analyze the rehearsal words and that this poses a greater impediment to encoding that supports later recollection versus that which is sufficient to support familiarity-based responding.

## REFERENCES


Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science*, **11**, 267-273.

Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1998). Making memories: Brain activity that predicts how well visual experience will be remembered. *Science*, **281**, 1185-1187.

Clark, D., & Wagner, A. D. (2003). Assembling and encoding word representations: fMRI subsequent memory effects implicate a role for phonological control. *Neuropsychologia*, **41**, 304-317.

Davachi, L., Maril, A., & Wagner, A. D. (2001). When keeping in mind supports later bringing to mind: Neural markers of phonological rehearsal predict subsequent remembering. *Journal of Cognitive Neuroscience*, **13**, 1059-1070.

DeCarlo, L. T. (2002). Signal detection theory with finite mixture distributions: Theoretical developments with applications to recognition memory. *Psychological Review*, **109**, 710-721.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition*, **24**, 523-533.

Gardiner, J. M., Gawlick, B., & Richardson-Klavehn, A. (1994). Maintenance rehearsal affects knowing, not remembering; elaborative rehearsal affects remembering, not knowing. *Psychonomic Bulletin & Review*, **1**, 107-110.

Gardiner, J. M., Java, R. I., & Richardson-Klavehn, A. (1996). How level of processing really influences awareness in recognition memory. *Canadian Journal of Experimental Psychology*, **50**, 114-122.

Gardiner, J. M., & Richardson-Klavehn, A. (2000). Remembering and knowing. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 229-244). New York: Oxford University Press.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1-67.

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review*, **100**, 546-567.

Glenberg, A., & Adams, F. (1978). Type I rehearsal and recognition. *Journal of Verbal Learning & Verbal Behavior*, **17**, 455-463.

Greene, R. L. (1987). Effects of maintenance rehearsal on human memory. *Psychological Bulletin*, **102**, 403-413.

Hasher, L., & Zacks, R. T. (1984). Automatic processing of fundamental information: The case of frequency of occurrence. *American Psychologist*, **39**, 1372-1388.

Hasher, L., Zacks, R. T., Rose, K. C., & Sanft, H. (1987). Truly incidental encoding of frequency information. *American Journal of Psychology*, **100**, 69-91.

Hilford, A., Glanzer, M., Kim, K., & DeCarlo, L. T. (2002). Regularities of source recognition: ROC analysis. *Journal of Experimental Psychology: General*, **131**, 494-510.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, **16**, 96-101.

Hirshman, E. (1995). Decision processes in recognition memory: Criterion shifts and the list-strength paradigm. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 302-313.

Hirshman, E. (1998). On the utility of the signal detection model of the remember–know paradigm. *Consciousness & Cognition: An International Journal*, **7**, 103-107.

Hirshman, E., & Henzler, A. (1998). The role of decision processes in conscious recollection. *Psychological Science*, **9**, 61-65.

Hirshman, E., Lanning, K., Master, S., & Henzler, A. (2002). Signal-detection models as tools for interpreting judgements of recollections. *Applied Cognitive Psychology*, **16**, 151-156.

Hockley, W. E. (1991). Recognition memory for item and associative information: A comparison of forgetting rates. In W. E. Hockley &

S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 227-248). Hillsdale, NJ: Erlbaum.

Inoue, C., & Bellezza, F. S. (1998). The detection model of recognition using know and remember judgments. *Memory & Cognition*, **26**, 299-308.

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory & Language*, **30**, 513-541.

Kelley, R., & Wixted, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 701-722.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence: Brown University Press.

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, **87**, 252-271.

Nairne, J. S. (1983). Associative processing during rote rehearsal. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **9**, 3-20.

Naveh-Benjamin, M., & Jonides, J. (1984). Cognitive load and maintenance rehearsal. *Journal of Verbal Learning & Verbal Behavior*, **23**, 494-507.

Nyberg, L. (2002). Levels of processing: A view from functional brain imaging. *Memory*, **10**, 345-348.

Rotello, C. M., Macmillan, N. A., & Reeder, J. A. (2004). Sum-difference theory of remembering and knowing: A two-dimensional signal detection model. *Psychological Review*, **111**, 588-616.

Rundus, D. (1970). Paired-associate recall following a distractor task: Initial and second choice performance. *Journal of Mathematical Psychology*, **7**, 362-370.

Schumacher, E. H., Lauber, E., Awh, E., Jonides, J., Smith, E. E., & Koeppe, R. A. (1996). PET evidence for an amodal verbal working memory system. *NeuroImage*, **3**, 79-88.

Smith, E. E., Jonides, J., Marshuetz, C., & Koeppe, R. A. (1998). Components of verbal working memory: Evidence from neuroimaging. *Proceedings of the National Academy of Sciences*, **95**, 876-882.

Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, **94**, 14792-14797.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology*, **26**, 1-12.

Wagner, A. D., & Gabrieli, J. D. (1998). On the relationship between recognition familiarity and perceptual fluency: Evidence for distinct mnemonic processes. *Acta Psychologica*, **98**, 211-230.

Wagner, A. D., Maril, A., & Schacter, D. L. (2000). Interactions between forms of memory: When priming hinders new episodic learning. *Journal of Cognitive Neuroscience*, **12**(Suppl. 2), 52-60.

Wagner, A. D., Paré-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. *Neuron*, **31**, 329-338.

Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., Rosen, B. R., & Buckner, R. L. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity [see comments]. *Science*, **281**, 1188-1191.

Yonelinas, A. P. (1994). Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 1341-1354.

Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition*, **25**, 747-763.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory & Language*, **46**, 441-517.

Yonelinas, A. P., & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory & Language*, **34**, 622-643.

## NOTES

1. In Experiment 2, we wanted to have greater certainty of the degree to which subjects were engaged in the distractor task, and this, combined with a clear difference in span accuracy across short and long conditions, led us to design the program with the ability to isolate the correct and incorrect Brown–Peterson encoding trials.

2. All errors were combined into the same condition regardless of the number of missed digits during the span task. We did not attempt further analyses conditional on the size of the error (e.g., 2 vs. 4 missed digits) because the bin sizes were so small as to likely be unreliable.

(Manuscript received February 24, 2003;
revision accepted for publication January 8, 2004.)