

# Associative recognition: A case of recall-to-reject processing

CAREN M. ROTELLO

*University of Massachusetts, Amherst, Massachusetts*

and

EVAN HEIT

*University of Warwick, Coventry, England*

Two-process accounts of recognition memory assume that memory judgments are based on both a rapidly available familiarity-based process and a slower, more accurate, recall-based mechanism. Past experiments on the time course of item recognition have not supported the recall-to-reject account of the second process, in which the retrieval of an old item is used to reject a similar foil (Rotello & Heit, 1999). In three new experiments, using analyses similar to those of Rotello and Heit, we found robust evidence for recall-to-reject processing in associative recognition, for word pairs, and for list-discrimination judgments. Put together, these results have implications for two-process accounts of recognition.

One of the key issues in recognition memory research is the nature of underlying processing, such as whether there is more than one process involved and what each process accomplishes. According to single-process accounts of recognition (e.g., Estes, 1994; Gillund & Shiffrin, 1984; Heit, 1993; Hintzman, 1988; Nosofsky, 1988), judgments are based on a unitary assessment of strength, corresponding to the summed similarity of the test item to all the relevant items in memory. Single-process memory models have been shown to account for a wide variety of data (see Clark & Gronlund, 1996, and Raaijmakers & Shiffrin, 1992, for reviews).

A growing body of research, however, suggests that recognition judgments may involve a second, recall-like process (e.g., Atkinson & Juola, 1973, 1974; Doshier, 1984; Gronlund & Ratcliff, 1989; Hintzman & Curran, 1994; Humphreys, 1978; Jacoby & Dallas, 1981; Jones & Heit, 1993; Mandler, 1980; Ratcliff & McKoon, 1982, 1989; Yonelinas, 1994; see Mandler, 1991, for a brief review). For example, one dual-process account assumes that recognition decisions are based on the familiarity of the target item if that familiarity is either sufficiently strong or suf-

ficiently weak that a response (*old* or *new*, respectively) may be made with confidence. However, if the familiarity of the item cannot support a confident decision, a secondary process is used (Atkinson & Juola, 1973, 1974).

Most relevant to our work, there is research suggesting that faster memory processing responds to global aspects of the test stimulus, whereas slower processing is affected by the finer details of the stimulus. This result has been demonstrated in a number of experiments using a response signal procedure. In this paradigm, a signal to respond occurs a variable amount of time (typically, 100–2,000 msec) after the onset of a test stimulus, and the subject is asked to make a judgment quickly after the signal. Using this methodology, Doshier (1984) has shown that when subjects study word pairs that are unrelated to one another (e.g., *king-pear*; *house-queen*), they tend to false alarm to rearranged distractor pairs that are semantically related (e.g., *king-queen*) for faster response signals—say, within about 600 msec of stimulus presentation. At longer response lags, however, subjects correctly reject these semantically related pairs (for related work, see Ratcliff & McKoon, 1982, 1989). In contrast, new items that are not similar to presented stimuli and are not semantically related to each other are correctly rejected even in very rapid responses (see Gronlund & Ratcliff, 1989). Similarly, Hintzman and Curran (1994) reported that subjects who studied lists of stimuli of varying plurality (e.g., *frog*, *trucks*) tend to false alarm to highly similar foils (*frogs*, *truck*) early in the time course of judgment but to correctly reject those test stimuli later in processing. These studies suggest that the familiarity of a stimulus is used as the first basis of response and that a recall-like process may be used for slower or more accurate recognition decisions.

This evidence points to two mechanisms that underlie recognition, but little research has directly addressed the specific nature of the presumed recall process. We think

---

This research was supported by Grant SBR-9753136 from the NSF and by a Faculty Research Grant from the University of Massachusetts to C.M.R. under the name Caren M. Jones. This research was also supported by a grant from the Biotechnology and Biological Sciences Research Council to E.H. and Koen Lamberts. We thank Scott Gronlund, William Hockley, Michael Humphrey, and Sandy Pollatsek for insightful comments on this work, Neil Macmillan and Arnie Well for statistical advice, and Sonya Dougal and Christine Higgins for assistance with stimulus materials and data collection. Correspondence concerning this article should be sent to either C. M. Rotello, Department of Psychology, Box 37710, University of Massachusetts, Amherst, MA 01003-7710 (e-mail: [caren@psych.umass.edu](mailto:caren@psych.umass.edu)) or E. Heit, Department of Psychology, University of Warwick, Coventry CV4 7AL, England (e-mail: [e.heit@warwick.ac.uk](mailto:e.heit@warwick.ac.uk)).

it is crucial to address specific versions of two-process accounts, rather than considering this issue at a level that is too general to make sharp predictions. Rotello and Heit (1999) described several possible ways in which a recall process could operate in recognition memory. For example, an *exhaustive search* recall process would work by searching through memory for a memory trace that matches the test probe. Successful location of a match would lead to its acceptance as a studied item; failure to find a match would lead to the rejection of the probe, regardless of how familiar it seemed. A *recall-to-accept* recall process would supplement a familiarity-based process by increasing the probability of an *old* response when a memory trace matching the probe is found. Such a process would not influence the responses to test foils, which would presumably be based on their familiarity alone (see Rotello & Heit, 1999, for details). However, the dominant conception of the recall-like process is that it would provide mismatching information from memory so that test foils that are similar to studied items may be correctly rejected and called *new* (e.g., Doshier, 1984; Yonelinas, 1997). This type of process has been called a *recall-to-reject* process (Clark, 1992, p. 241; Clark & Burchett, 1994, p. 57; Gronlund & Ratcliff, 1989, p. 857; Hintzman & Curran, 1994, p. 14; Rotello & Heit, 1999; see Clark & Gronlund, 1996, pp. 56-57, for a review). That is, when a subject is tested on a similar foil, the corresponding old item is recalled, leading the subject to reject the foil.

The recall-to-reject account makes a distinctive prediction, unique among these possible recall processes in recognition, when a similar item, such as *frogs*, is tested but the word *frog* was studied (Rotello & Heit, 1999). When *frogs* is tested, the old item, *frog*, is recalled, but the similar item, *frogs*, cannot be recalled. In the face of this mismatching evidence, *frogs* is rejected and is given a judgment of *new* regardless of the output of the similarity-based recognition process. This recall-to-reject process would not affect a dissimilar item, such as *chair*, which is not particularly similar to any presented item. Indeed, such a completely new item would likely be rejected on the basis of the rapid similarity-based process. Thus, the difference in false alarms to *frogs* and to *chair* should decrease late in processing (say, after about 600–700 msec, based on past research), as the second process helps people to reject *frogs*, relative to *chair*. These differences in false alarms can be computed as  $d'$ 's or, for convenience, as  $d_L$ 's, which indicate the standardized distance between the means of the two putative normal ( $d'$ ) or logistic ( $d_L$ ) distributions of memory strength.<sup>1</sup> Doshier (1984) used such an analysis for recognition judgments, referring to the measure as *pseudo- $d'$* ; more recently, this measure has been used for similar experiments by Doshier and Rosedale (1991), Gronlund and Ratcliff (1989), Hintzman and Curran (1994), Ratcliff and McKoon (1989), and Rotello and Heit (1999). Figure 1 presents the recall-to-reject prediction, in terms of  $d_L$ , in idealized graphical form.

Other performance measures, such as the false alarm rate to similar foils as a function of processing time, could

also be considered when evaluating the presence of a recall process. Say that early in processing, false alarms to similar foils tend to increase, whereas later in processing they tend to decrease. Such a pattern of false alarms has sometimes been taken as general evidence for a recall-like process (e.g., Hintzman & Curran, 1994). However, it is normally accepted that hit rates and false alarm rates should be interpreted with caution, because hits and false alarms depend not only on signal strength, but also on response bias. In a typical response signal experiment, in which a subject might sometimes have to respond after 100 msec and, other times, have as much as 2,000-msec to respond, a subject may well have different response biases at different deadlines. Hardly any information would be available early in processing, so a lenient criterion for responding *old* might apply. On the other hand, when further information is available, a stricter criterion might be in effect. Indeed, Rotello and Heit (1999) reported significant changes in bias measures as a function of response signal lag. However, we note that statistical assessment of whether bias is changing, or not changing, is complex when sensitivity or  $d'$  is also changing, as in the case of response signal experiments (Macmillan & Creelman, 1991).

In sum, using a difference score such as  $d_L$  has two main advantages over uncorrected measures, such as false alarm rate. First and more important, it is a targeted measure of recall-to-reject processing, rather than a general measure of a secondary process. Second, because it is, in essence, the difference in response to two items at a given response signal deadline, any response bias itself should be canceled out. We emphasize that the issue of changing response bias is secondary, however. This difference

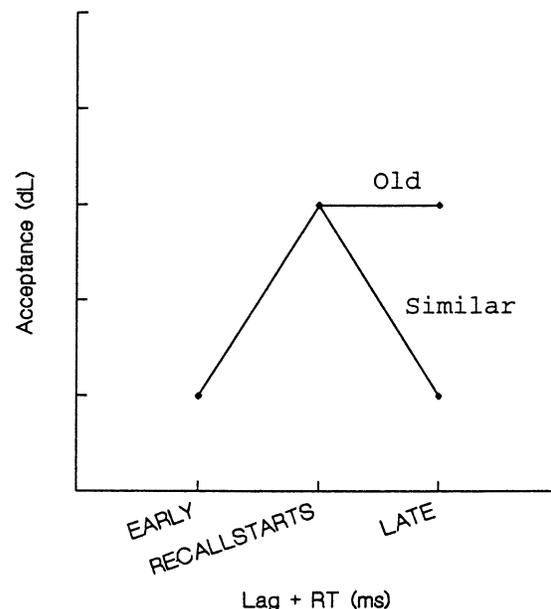


Figure 1. Idealized prediction of the recall-to-reject account of recognition judgments, in terms of  $d_L$ .

score measure is targeted at recall-to-reject processing regardless of whether response bias changes at different response signals.

### Evidence Regarding Recall-to-Reject Processing

Rotello and Heit (1999) assessed the evidence for recall-to-reject processing in item recognition, using data from Hintzman and Curran (1994, Experiments 2 and 3), as well as data from two new experiments. In each experiment, subjects studied lists of stimuli and then were asked to make recognition judgments in a response signal paradigm. There were three types of test items: (1) *old* items that had appeared on the study list; (2) *new* items that had not been studied and were unrelated to any studied items; and (3) *similar* items that had not been studied but were similar to studied items. Across the experiments, similar items were defined either as nouns that differed in plurality from the studied items (e.g., *frogs*, when *frog* had been presented) or as pseudowords that were orthographically similar to studied pseudowords (e.g., PRUMIR, when PRAMAD had been studied). None of the analyses in any of the four experiments showed evidence for recall-to-reject processing. More specifically, no decrease in the rate of acceptance of similar items was found later in processing, as would be predicted by the recall-to-reject account. Similar foils were not being increasingly rejected, relative to completely new foils, as subjects were allowed more time to make their judgments. Although there may be some results that are generally suggestive of multiple processes, our analyses specifically targeted the distinctive time course prediction for recall-to-reject processing. Now that these strict tests have been developed, we intend to apply them to other tasks that may well lead to recall-to-reject processing.

One task in which recall-to-reject processing is perhaps more plausible is associative recognition. In associative recognition, two items or events are recognized as having occurred together at some point in the past. If pairings of events are mutually exclusive, it may be assumed that recalling that one event had occurred with the test cue precludes another event's having occurred with the cue. More concretely, if you are asked whether John is married to Mary, recalling that Susan is John's wife is sufficient evidence to recognize that John and Mary are not married. Several researchers have proposed that associative recognition includes a recall-to-reject process (e.g., Clark, 1992, p. 241; Clark & Burchett, 1994, p. 57; Gronlund & Ratcliff, 1989, p. 857; Humphreys, 1976, 1978; Mandler, 1980). However, they have not provided distinctive time course evidence for the recall-to-reject mechanism.

Indeed, there are reasons to predict that no evidence would be found for a recall-to-reject process in associative recognition. It may seem unlikely that there is a different mechanism for pairs of words than for single words. For example, Hockley (1994) has argued from experiments on the mirror effect that "similar retrieval and decision processes underlie recognition of item and asso-

ciative information" (p. 713). Several memory models treat the storage of associations between items in essentially the same way as that for individually studied items (e.g., Hintzman, 1988) and assume that both item and associative information are retrieved from memory simultaneously (e.g., Gillund & Shiffrin, 1984; Hintzman, 1988; Pike, 1984; see Clark & Gronlund, 1996, for a brief summary). Although these models can accommodate additional assumptions about recall processes, they do not at present assume that recall is part of associative recognition. We see the value of these models, in fact, as showing what can be done without building in recall-like processes when they are not actually necessary to account for results.

Nonetheless, many differences have been found between item and associative recognition. For example, low-frequency words are typically better recognized than high-frequency words in an item recognition task (see Gregg, 1976, for a review), but in an associative recognition task, either there is no frequency effect (Hockley, 1994) or high-frequency words are better recognized (Clark, 1992; Clark & Shiffrin, 1992). The associative recognition data are more consistent with frequency effects in recall (see, e.g., Sumbly, 1963) than with those in item recognition. Similarly, instructional effects on associative recognition and free recall performance also parallel one another and are different than the effects for item recognition. When subjects are told to study word pairs in preparation for an associative recognition test, rather than an item recognition test, their associative recognition improves, but their item recognition is unaffected (Hockley & Cristi, 1996). Analogously, subjects who expect a recall test of studied materials have better recall scores than those who expect a recognition test, but the two groups have equivalent recognition performance (May & Sande, 1982). Results like these led Clark and Gronlund (1996) to argue that recall was more likely to operate in associative recognition than in item recognition and, moreover, that it was likely to operate in a recall-to-reject fashion.

Several other studies have provided evidence suggestive of a recall-to-reject process in associative recognition. Perhaps the most compelling data come from an experiment by Humphreys (1978, Experiment 1). Humphreys evaluated whether recognition of associated words (i.e., intact,  $A-B$  pairs) could be predicted from the probability of recalling  $B$  given  $A$  as a cue and the probability of recognizing  $B$  in the absence of other memory cues. Subjects studied pairs of words ( $A-B$ ) and then recognized old ( $B$ ) and new ( $X$ ) words. Finally, half of the subjects took an associative recognition test in which intact pairs ( $A-B$ ) and old-new pairs ( $A-X$ ) were presented. The remaining subjects were given a cued recall test, with the  $A$  words as retrieval cues. Performance in the associative recognition condition was well predicted by the probability of a  $B$  word's being recognized and, if  $B$  was not recognized, the probability of its being recalled when cued with  $A$ . The close match between the observed and the predicted data in the associative recognition task supports the hypothesis that recall is involved in associative

recognition. However, Humphreys' data do not speak to the time course predictions for a recall-to-reject process.

More recent support for the use of recall-to-reject processing in associative recognition comes from receiver operating characteristic (ROC) data (Rotello, Macmillan, & Van Tassel, 2000; Yonelinas, 1997). In Yonelinas's (1997) experiments, subjects were asked to discriminate *intact* pairs of words (i.e., those that had been studied together) from *rearranged* test pairs in which both words had been studied, but not as part of the same pair. (Departing from standard terminology, Yonelinas, 1997, labeled those rearranged pairs *new*.) Yonelinas (1997) argued that sometimes subjects were able to recall that rearranged pairs had not been studied, because they could recall that they had been paired with other study partners. The evidence for this claim centered on the (cumulative) false alarm rate that was associated with the upper portion of the ROC. Specifically, "if subjects recollect that 20% of the new [i.e., rearranged] pairs are new, then their false alarm rate should not exceed 80%" (Yonelinas, 1997, p. 752). Indeed, under conditions in which recollection of studied pairs was relatively easy (i.e., each word was paired with only one study partner; short study lists), the cumulative false alarm rate declined, as compared with an experiment in which study lists were long and study words had multiple partners. Rotello et al. (2000) recently showed that this decline in the cumulative false alarm rate to rearranged pairs occurs only under conditions in which the recall process has time in which to operate. For recognition judgments made very quickly, in which the familiarity-based process was assumed to dominate the decision, no evidence for recall-to-reject processing emerged in the ROC data.

Finally, there is some evidence collected over the full time course of judgment that is suggestive of a recall-to-reject process. Gronlund and Ratcliff (1989) asked subjects to discriminate intact and rearranged test pairs; in addition, they presented subjects with new test pairs in which neither word had been studied. Their data suggest that people may be increasingly able to reject similar, rearranged distractors, relative to completely new foils, after about 600–700 msec of processing time. That is, in their published figure, the false alarms to the rearranged pairs appeared to decrease more rapidly over time than did the false alarms to completely new foils, perhaps indicating the presence of a recall-to-reject process in an associative recognition task.

### Overview

We see experiments on the time course of recognition judgments, using methods such as the response signal technique, as being crucial for assessing any multiprocess account of recognition. The recall-to-reject account makes a distinctive prediction for response signal experiments, in terms of a difference score comparing responses on similar foils and completely new items. To the extent that this prediction is not supported by particular results (as

in Hintzman & Curran, 1994, and Rotello & Heit, 1999), we see a failure of the recall-to-reject account. Again, we are focusing on the most appropriate measure of recall-to-reject processing, rather than assessing other measures that are not targeted at this specific account of the recall process. On the basis of past work, there is some reason to think that recall-to-reject processing may be more likely in associative recognition than in item recognition. However, this past work for the most part did not address the time course of judgments, nor did it apply the targeted, difference-score measure that is most appropriate for assessing recall-to-reject processing. Therefore, in this paper, we are applying the analyses in Rotello and Heit to associative recognition, to better assess the evidence for recall-to-reject processing.

Experiment 1 replicated the critical aspects of Gronlund and Ratcliff's (1989) Experiment 2 with a substantially larger number of subjects. We also simplified their design. Concurrent with the associative recognition task, Gronlund and Ratcliff's subjects were given an item recognition task. That is, their study lists and test lists were mixtures of individual words and word pairs. This design may have focused the subjects' attention on the individual words in the study pairs, rather than on the associations. Such a focus on individual study items might make the task somewhat similar to an item recognition task, in which recall-to-reject processes have not been found. In our first experiment, the subjects studied and were tested only on word pairs.

In our Experiment 2, the subjects again studied only pairs of words, but they were tested on word pairs as well as single words. We examined whether recall-to-reject processing depended on studying pairs of words together or whether it was a feature of associative (i.e., word pair) recognition judgments. Finally, we continued along these lines in Experiment 3, in which the subjects studied only single words, but in two different list contexts. The retrieval task was associative in that the subjects judged whether a particular word occurred in a particular list.

## EXPERIMENT 1

### Method

**Subjects.** The subjects were 14 members of the University of Massachusetts community, mainly undergraduates, who were paid for their participation. An additional 3 volunteers participated, but their data were excluded from the analyses because they failed to discriminate studied test pairs from completely new pairs even at the longest test lags.

**Stimuli.** There were 3,600 words (3,100 nouns, 500 verbs) selected from the Francis and Kučera (1982) word frequency norms. The words were 3–11 letters in length, with initial letters covering the entire alphabet, and were selected without regard to linguistic frequency. Other than length, the two constraints on selection were that there could be no obvious synonym pairs represented on the list and that an average University of Massachusetts undergraduate would be likely to recognize each word.

The 3,600 stimulus words were randomly organized into pairs for each subject. The word pairs were then assigned to one of 50 study–

test blocks. Each block consisted of a study list of 27 pairs of words presented one at a time for 4 sec and a test list of 27 pairs. On each test list, there were 9 *intact* pairs from the just-studied list, 9 *rearranged* pairs from that list, and 9 pairs made up of completely new words that were never studied in any list. The rearranged pairs were generated by combining the left-hand word from one study pair with the right-hand word from another study pair, maintaining the left–right ordering. The study pairs from which the rearranged test pairs were created appeared in random positions on the study list.

Six test lags were used (100, 300, 500, 750, 1,200, and 2,000 msec), so there was one test trial of each type (intact, rearranged, new) at each test lag. In even-numbered blocks, the 3 remaining trials occurred at the 100-, 500-, and 1,200-msec lags; in odd-numbered blocks, they occurred at the 300-, 750-, and 2,000-msec lags. So, within each experimental session (i.e., 10 blocks) there were 15 trials of each type at each lag. Across blocks and sessions (excluding Session 1 as practice), there were 60 trials of each type at each lag for each subject.

The order of presentation of both study and test pairs was randomly determined for each study–test block for each subject. To minimize list interference, each stimulus word appeared in only one study pair and/or one test pair in the experiment.

**Procedure.** Each subject participated in 5-h-long experimental sessions over a period of several days. At the beginning of each session, there was a block of 36 training trials in which the subjects practiced responding in the response signal procedure. For each of the 36 training trials, they were shown an orienting cue ( $\rightarrow+\leftarrow$ ) centered on the computer monitor for 1 sec. The cue was replaced by the word *yes* or *no*, which appeared on the screen for a variable amount of time (100, 300, 500, 750, 1,200, or 2,000 msec) before being replaced by a mask (????). At the same time as the mask appeared, a brief signal tone sounded, signaling that the subject should press 1 if the word *yes* had appeared and 0 if *no* had been presented. They were instructed to respond as quickly as possible after the signal, even if they had to make a guess. However, accuracy feedback was provided on each trial, and latency feedback was provided only when the response occurred too quickly after the signal (i.e., less than 50 msec after the signal) or when it occurred too slowly (i.e., more than 350 msec after the signal). There were 6 training trials at each of the six test lags; half were *yes*, and half were *no*.

After the practice trials, each session was made up of 10 study–test blocks. The 27 pairs of study words were presented one at a time, centered on the screen, at a 4-sec rate. Immediately following the study list, the subjects made 27 speeded recognition judgments about pairs of words from that study list. As in the practice trials, each test trial began with an orienting cue, which was replaced by a test pair that remained on the screen for a variable amount of time (i.e., one of the six test lags) before being replaced by the mask. At the same time as the mask appeared, a signal tone sounded, indicating to the subject that an immediate response was required. The subjects were told to respond *yes* to test pairs containing words that had appeared *together* on the study list (i.e., to intact pairs) and to respond *no* to all other pairs. As in the practice trials, the subjects were urged to respond immediately after the signal; they were given accuracy feedback on each test trial and latency feedback after responses that were too slow or too fast. The subjects were instructed to take self-paced breaks between blocks.

## Results

Data from the first session for each subject were discarded as practice trials. All analyses were based on only those responses that occurred between 100 and 350 msec after the response signal. Responses outside that window, 13.3% overall, were considered to reflect anticipa-

**Table 1**  
Mean Hit and False Alarm (FA) Rates  
for the Stimuli in Experiment 1

Type	Response Signal Lag (msec)					
	100	300	500	750	1,200	2,000
Intact (Hit)	.12	.31	.53	.70	.75	.78
Rearranged (FA)	.19	.21	.34	.29	.19	.19
New (FA)	.15	.09	.07	.05	.03	.04

tory responses or additional uncontrolled processing of the stimulus.<sup>2</sup>

For completeness, we provide the hit rates and false alarm rates for the intact, rearranged, and new pairs in Table 1. These values were adjusted by adding 0.5 to the number of *yes* responses and dividing by the total number of responses + 1.0, so that  $d_L$  could be calculated later (Hintzman & Curran, 1994; Rotello & Heit, 1999; Snodgrass & Corwin, 1988). There are two notable aspects of the data. The false alarms to rearranged pairs increased early in processing (i.e., until the 500-msec response signal) and then declined later in processing. Although this peaked pattern of false alarms to similar or rearranged test stimuli could be taken as evidence of the contribution of recall to recognition, the raw scores must be interpreted with caution, because the false alarms to completely new pairs declined continuously across processing time. That is, the subjects' decision criteria appear to become more conservative as they are allowed more time to make recognition decisions, although we again note that this issue is difficult to assess statistically when sensitivity is changing as well (Macmillan & Creelman, 1991).

Next, we calculated acceptance rates ( $d_L$ ) for intact and rearranged pairs by comparing the *old* response rates on these items to the *old* response rates to completely new pairs. The results of these analyses are shown as the data points in Figure 2. The subjects were well able to discriminate intact test pairs from completely new pairs: The acceptance of intact pairs, relative to completely new pairs, rises continuously to a high asymptote. However, the subjects showed a nonmonotonic function of acceptance for rearranged pairs. That is, acceptance of rearranged pairs increased early in processing and then declined later in processing. These data appear to be consistent with the predictions of the recall-to-reject account (compare Figures 1 and 2).

There are several ways in which these  $d_L$  data could be evaluated statistically. Most obviously, an analysis of variance could be conducted on the data, and post hoc tests or trend analyses could be used to evaluate whether the acceptance of rearranged pairs (relative to completely new pairs) shows a peak early in processing (see Rotello & Heit, 1999). A better focused analysis would use nonlinear estimation procedures that capture the pattern or shape of the data. That is, the data can be fitted with mathematical functions that either do or do not have a peaked shape. If the fit is statistically better with a peaked (i.e.,

nonmonotonic) function than with a nonpeaked (i.e., monotonically increasing) function, there would be distinctive evidence for a rise-then-fall pattern in the data. We adopted this curve-fitting approach for our present analyses.

To determine whether the drop in acceptance of the rearranged pairs at the later response signal lags was statistically significant, the data in Figure 2 were fitted with two separate functions. The first function represents monotonic growth to a limit (Ratcliff, 1978):

$$d_L(lag) = \frac{\lambda_1}{\sqrt{1 + v/(lag - \delta)}}, \tag{1}$$

where *lag* is the response signal lag plus the latency to respond at that lag (i.e., total processing time),  $\lambda_1$  is the asymptotic level of discrimination, *v* is the rate of approach to that asymptote, and  $\delta$  is the point at which discrimination rises above chance. The second function we fit to the data is a nonmonotonic function of a similar format. It assumes that the data conform to Equation 1 until a particular point in processing time, *lag\**, at which point discrimination begins to decline to a second asymptote,  $\lambda_2$ . (Because we were interested specifically in a rise-then-fall pattern in the acceptance data, the parameters were constrained so that the second asymptote was no higher than the first.) The nonmonotonic function is given by

$$d_L(lag) = \frac{\lambda_2 + (\lambda_1 - \lambda_2)(lag^* - \delta)/(lag - \delta)}{\sqrt{1 + v/(lag - \delta)}} \tag{2}$$

for *lag* ≥ *lag\**.

Equations 1 and 2 (and variants) have been widely used (e.g., Doshier, 1984; Gronlund & Ratcliff, 1989; Ratcliff, 1980; Rotello & Heit, 1999; Wickelgren & Corbett, 1977).

For each type of stimulus, distinct rates and asymptotes were estimated, but a single intercept ( $\delta$ ) was used for all the stimuli to reduce the number of parameters to be fit.<sup>3</sup> Model fitting was performed using SYSTAT's Simplex estimation method to iteratively adjust parameters so as to minimize the residual sum of squares (*RSS*). The fits to these data confirmed the visually observable conclusion of a nonmonotonic acceptance rate for rearranged pairs. Although the monotonic fit was quite good ( $R^2_{adjusted} = .986$ , *RSS* = .981),<sup>4</sup> the nonmonotonic fit was significantly better [ $R^2_{adjusted} = .990$ , *RSS* = .515;  $\chi^2(2) = 7.733$ , *p* < .05].<sup>5</sup> The best-fitting parameter values for both functions are shown in Table 2, and the nonmonotonic func-

tion's fitted values are shown in Figure 2 as the curves superimposed on the data.

These model-based analyses focused on average group responses, rather than on individual subjects' responses. It is useful to look at group responses, particularly when not much data has been collected for each individual, but there is some risk that the averaged group-level responses may not capture all the characteristics of individual subjects' data. To evaluate whether the group responses were indeed representative of the individuals', the data from individual subjects were also fitted with both functions. Of course, these data were necessarily noisier, possibly making it more difficult to find a reliably better fit for the nonmonotonic function. Nonetheless, for 4 of the 14 subjects, the nonmonotonic function fit significantly better than the monotonic function. With an  $\alpha = .05$  criterion, the probability of 4 or more significant results out of 14, by chance alone, is less than 1%. Also, it is possible to sum up the evidence for the nonmonotonic model obtained for each individual subject. The additional extent of data fitting provided by the nonmonotonic model can be represented as a chi-squared statistic with 2 degrees of freedom (see note 5). The summation of these statistics would also have a chi-squared distribution, with 28 degrees of freedom. The obtained value for the summation across subjects was  $\chi^2(28) = 66.88$ , *p* < .001, indicating that the total improvement of fit that was due to the nonmonotonic model was significantly greater than chance. Hence, when applying the models to individual subjects' data, the results supported the analyses based on group averages. (See Rotello & Heit, 1999, for further examples of these analyses.)

**Discussion**

Most critically, the subjects showed a peaked pattern of acceptance for rearranged test pairs, as a function of response deadline, providing evidence for recall-to-reject processing. Apparently, retrieving old word pairs did help the subjects to reject rearranged word pairs. In contrast to these results supporting recall-to-reject processing, Rotello and Heit (1999) found no evidence consistent with recall-to-reject processing in several item recognition experiments. This difference may be surprising in light of past accounts that recognition of items and associations is dependent on the same underlying mechanisms. However, as we argued earlier, a number of differences between item and associative recognition have been reported

**Table 2**  
**Best-Fitting Parameter Values for the Monotonic and Nonmonotonic Functions, Fit to Data from Experiment 1**

Type	Monotonic Model			Nonmonotonic Model				
	Intercept	Asymptote	Rate	Intercept	Lag*	Asy. 1	Asy. 2	Rate
Intact	497.6	5.78	341.5	493.6	1,105.3	5.83	—	363.2
Rearranged	497.6	2.14	31.0	493.6	1,105.3	2.76	1.39	106.9

Note—Asy. 1, the earlier asymptote; Asy. 2, the later (lower) asymptote.

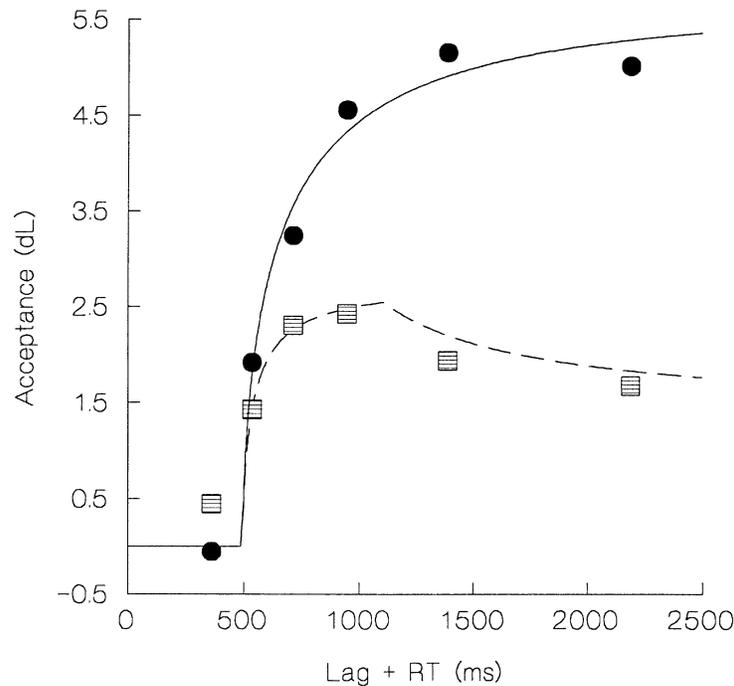


Figure 2. Mean acceptance rates, in terms of  $d_L$ , for the intact (circles) and rearranged (squares) test pairs in Experiment 1. The curves show the best-fitting nonmonotonic functions.

(e.g., different effects of linguistic frequency), many of which are consistent with a recall process's playing a larger role in associative recognition than in item recognition.

Our next two experiments were designed, in the first instance, as attempts to extend and replicate the results supporting recall-to-reject processing. Experiment 1 is perhaps the most distinctive time course evidence for recall-to-reject processing to date, and it did seem useful to try to replicate this singular result. Experiment 2 was intended to explore further the differences between item recognition (no recall-to-reject processing) and pair recognition (recall-to-reject processing). What is it about associative recognition tasks that would lead to recall-like processes? Does studying items in pairs facilitate later recall-like processes? Past research has indicated that the manner in which information is encoded does affect retrieval in several ways. For example, experiments in which subjects study pairs of words and are instructed to form either a visual image of each word or a single image of the words together have shown that recognition of the individual words (i.e., item recognition) tends to be better under separate imagery instructions, whereas associative recognition is best under interactive imagery instructions (e.g., Begg, 1979; McGee, 1980). In addition, studying pairs of words differs from studying single words in terms of memory load (on the surface, there is twice as much information being encoded for word pairs, as compared with single words). Perhaps the changes in the study task in Experiment 1 could have been responsible for the different results from those of Rotello and

Heit (1999). On the other hand, our own account is that the nature of associative retrieval itself would facilitate recall-to-reject processing. That is, when tested on rearranged pairs, subjects will be able to readily recall old pairs of words and use this information to reject the rearranged pair.

In Experiment 2, the subjects studied pairs of words; then they were tested both on word pairs and on single words. Thus, the subjects had an associative encoding task, followed by a test of associative recognition (as in Experiment 1) and of item recognition. Our prediction was that recognition would be suggestive of recall-to-reject processing for word pairs, replicating Experiment 1, but that recognition would not indicate a second recall-like process for single items, replicating Rotello and Heit (1999). Such results would indicate that recall-to-reject does not depend on studying pairs of words but, rather, on being tested on pairs of words.

Table 3  
Stimulus Structure for Experiment 2

Study Pairs	Test Stimuli	Type
A-B	A-B	Intact pair
C-D	C-F	Rearranged pair
E-F	X-Y	New pair
G-H	G	Old item
I-J	I'	Similar item
	Z	New item

Note—This structure represents one seventh of a study-test block; ' indicates semantic relatedness (synonymy) between two stimuli with the same letter name.

## EXPERIMENT 2

## Method

**Subjects.** There were 13 volunteers from the same population as that in Experiment 1. They were paid \$8 for each of five experimental sessions. One additional subject's data were excluded because fewer than a third of the responses were made within 350 msec of the response signal and the remaining trials revealed chance-level recognition at the longest test lags.

**Stimuli.** The basic stimulus structure is shown in Table 3. Essentially, the subjects studied lists of 35 word pairs and were tested on intact, rearranged, and new word pairs, as well as on individual studied words, synonyms of studied words, and completely new words.

The stimuli were 3,298 words selected from the larger pool of words used in Experiment 1 and 245 synonym pairs (e.g., *attorney-lawyer*; *rabbit-bunny*) that were selected from the Whitten, Suter, and Frank (1979) norms (the set of synonyms did not overlap with the larger pool of words). From the larger pool, 2,808 words were selected and randomly arranged into pairs, some of which served as study pairs and some of which were reserved for use as completely new test pairs. An additional 245 words from the larger pool were selected at random and served as the completely new foils on item recognition trials. From each synonym pair, one word was selected at random and assigned to be studied with one of the 245 remaining words from the larger pool (it was randomly assigned to the left or right position within the pair); the other word in each synonym pair served as a similar distractor on the item memory tests.

For each subject, the word pairs were divided evenly into 35 study lists of 35 pairs each. The presentation order of the pairs on each study list was randomized for each subject. For each study list, a corresponding test list was created so that there were seven intact test pairs, seven rearranged test pairs, seven new test pairs, seven studied test words, seven unstudied synonyms, and seven completely new words. One test item of each type occurred at each of the seven signal lags (i.e., 100, 300, 500, 750, 1,000, 1,250, and 2,000 msec). The test order was randomized for each list and for each subject. For the test pairs, words always appeared in their original left-right ordering. For the test items, the test word of a given studied pair was selected at random from the left and right words. Across lists and blocks (excluding those in the first session), there were 28 test trials of each type at each response signal lag.

**Procedure.** The procedure was similar to that of Experiment 1, with only minor changes. First, the subjects were given 42 practice trials at the beginning of each experimental session, 6 at each of the seven possible signal lags (100, 300, 500, 750, 1,000, 1,250, and 2,000 msec). Second, the subjects were told to respond positively to test *pairs* only if they had studied the words together on the list, whereas they were told to respond positively to test *words* if they had studied the words in any pair. For all other test items, they were to respond negatively.

**Table 4**  
Mean Hit and False Alarm (FA) Rates  
for the Stimuli in Experiment 2

Type	Response Signal Lag (msec)						
	100	300	500	750	1,000	1,250	2,000
Test pairs							
Intact (hit)	.30	.32	.48	.59	.57	.59	.60
Rearranged (FA)	.28	.26	.38	.41	.40	.38	.25
New (FA)	.26	.26	.15	.14	.15	.15	.14
Test items							
Old (hit)	.25	.36	.50	.61	.58	.60	.58
Similar (FA)	.29	.26	.28	.29	.25	.26	.21
New (FA)	.26	.27	.23	.22	.24	.21	.16

## Results

The data from each subject's initial session were excluded as training trials. In addition, we excluded responses that were less than 100 msec or more than 350 msec after the response signal, because those data include anticipatory responding or additional uncontrolled processing. The trimming process excluded 20.4% of the data, which suggests that this experimental task, with intermixed item and pair recognition, was more difficult than the straightforward pair recognition task in Experiment 1 (for which only 13.3% of the data were excluded).<sup>6</sup>

The mean hit and false alarm rates to both pairs and items are shown in Table 4. As in Experiment 1, one notable aspect of the data is the systematic decrease in the false alarm rate to new pairs and to new items as the response signal lag increases. Also noteworthy is the apparent increase in false alarms to rearranged word pairs early in processing, followed by the substantial decrease in those false alarms later in processing. This pattern of responding is suggestive of a late recall-like process that allows rejection of similar foils.

The acceptance rate (in terms of  $d_L$ ) for test pairs is shown in Figure 3, as a function of response lag plus reaction time; corresponding results for single items are shown in Figure 4. Consistent with past research (e.g., Humphreys, 1976, 1978), recognition of individual words that were studied in pairs is somewhat lower than recognition of the pairs. Perhaps the most striking aspect of these two figures, however, is the difference in the shape of the  $d_L$  curves. Whereas the pair data, shown in Figure 3, look quite similar to the data from Experiment 1, the item recognition data look quite different. As was predicted, the pair recognition results show the distinctive peaked pattern of recall-to-reject processing, but the item recognition data do not. (The results in Figure 4 do appear to be comparable with item recognition results reported by Rotello & Heit, 1999.)

Next, we conducted model-based analyses of the data from the associative recognition task. That is, we fitted the monotonic and nonmonotonic functions (Equations 1 and 2) to the data in Figure 3. Although the monotonic fit was good ( $R^2_{\text{adjusted}} = .976$ ,  $RSS = .796$ ), the fit of the nonmonotonic function was significantly better [ $R^2_{\text{adjusted}} = .988$ ,  $RSS = .265$ ;  $\chi^2(2) = 15.398$ ,  $p < .001$ ]. The best-fitting parameter values for both functions are shown in Table 5, and the predictions of the nonmonotonic function are shown as the curves in Figure 3.

We also conducted model-based analyses of the data from the item recognition task. Although the nonmonotonic fit was good ( $R^2_{\text{adjusted}} = .962$ ,  $RSS = .425$ ), it was not reliably better than the monotonic fit for the data in Figure 4 [ $R^2_{\text{adjusted}} = .968$ ,  $RSS = .455$ ;  $\chi^2(2) = .955$ , n.s.]. The best-fitting parameter values for both the monotonic and the nonmonotonic functions are shown in Table 5, and the predictions of the monotonic function are shown as the curves in Figure 4.

Finally, we fit the monotonic and nonmonotonic functions to individual subjects' data. Those data were neces-

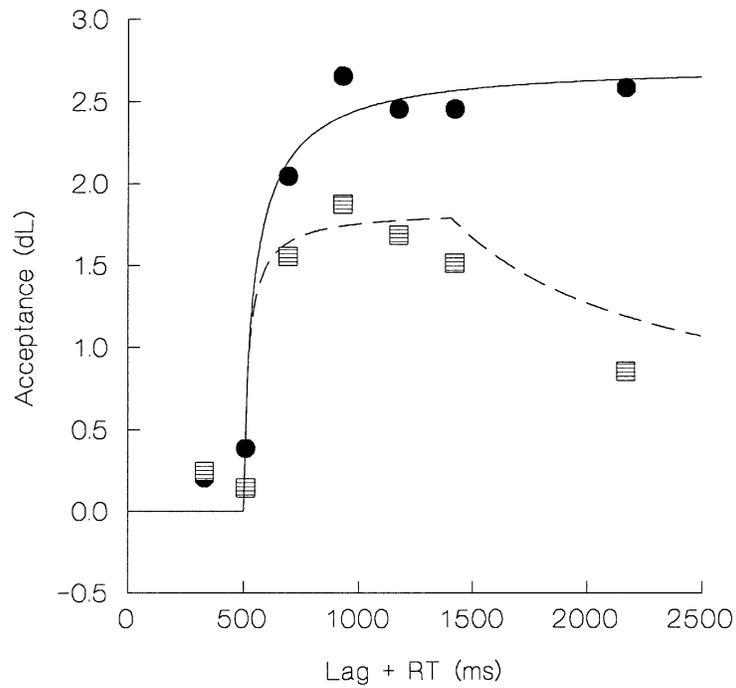


Figure 3. Mean acceptance rates, in terms of  $d_L$ , for the intact (circles) and rearranged (squares) test pairs in Experiment 2. The curves show the best-fitting nonmonotonic functions.

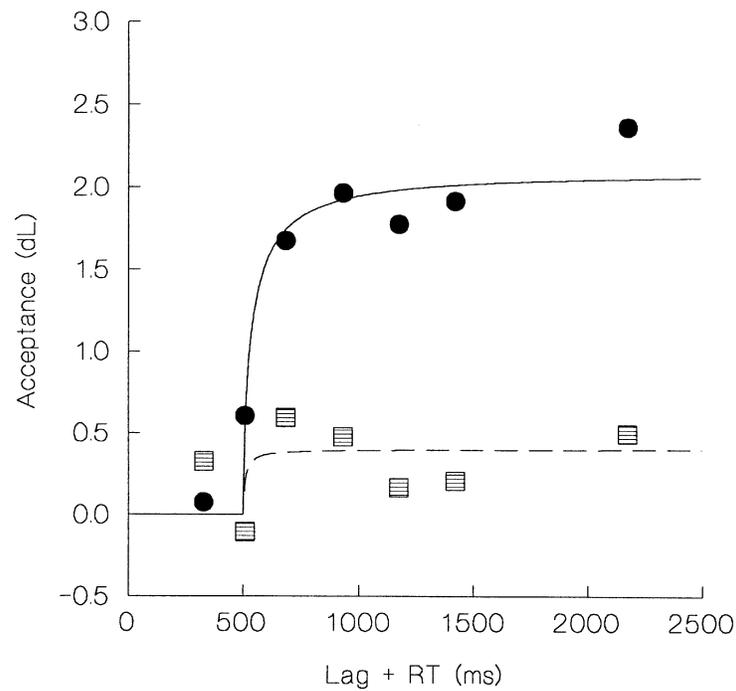


Figure 4. Mean acceptance rates, in terms of  $d_L$ , for the old (circles) and similar (squares) test items in Experiment 2. The curves show the best-fitting monotonic functions.

**Table 5**  
**Best Fitting Parameter Values for the Monotonic and Nonmonotonic Functions,**  
**Fit to Data from Experiment 2**

Type	Monotonic Model			Nonmonotonic Model				
	Intercept	Asymptote	Rate	Intercept	Lag*	Asy. 1	Asy. 2	Rate
Associative recognition								
Intact	506.0	2.72	113.7	505.7	1,406.0	2.73	—	119.1
Rearranged	506.0	1.51	0	505.7	1,406.0	1.84	.46	48.7
Item recognition								
Old	504.0	2.10	82.0	475.2	654.4*	2.59	—	466.0
Similar	504.0	.40	18.9	475.2	654.4*	.22	.22	521.5

Note—Asy. 1, the earlier asymptote; Asy. 2, the later (lower) asymptote. \*Multiple values of this parameter make the same predictions. A sample value is shown.

sarily noisier than the overall means (they were also noisier than the individual data from Experiment 1, because only half as many associative recognition judgments were collected). Nonetheless, of the 13 subjects, 3 had associative recognition data that were significantly better fit by the nonmonotonic model. There would only be a 1.1% chance of three of these outcomes resulting by chance alone. For the item recognition data, only 1 subject had data that were significantly better fitted by the nonmonotonic model ( $p > .35$ ). In addition, the incremental fit of the nonmonotonic model, summed over all the subjects' individual fits, was highly significant for the recognition of word pairs [ $\chi^2(26) = 65.749, p < .001$ ] but was not significant for item recognition [ $\chi^2(26) = 38.160, n.s.$ ]. These analyses of individual subjects' data fit well with the analyses for group averages.

### Discussion

Most importantly, this experiment replicated the findings from Experiment 1 suggesting recall-to-reject processing for word pairs and likewise replicated the findings of Rotello and Heit (1999) that did not find evidence for recall-to-reject processing for single words. Notably, Rotello and Heit's main analyses were for experiments that used pseudowords and words that differed in plurality. Hence, these findings for a lack of recall-to-reject processing in the case of item recognition of synonyms are an extension of Rotello and Heit's reported results. It appears that the difference between item recognition and pair recognition is not some consequence of simply studying words in pairs; rather, recall-to-reject processing seems to be particularly facilitated by testing on paired items.

In Experiment 3, we again attempted to extend our results, here by using a different kind of associative retrieval task. Instead of testing subjects on word pairs, however, we asked the subjects to study individual items on two separate study lists (List 1 and List 2) and gave them list-specific recognition tests. Hence, we created an associative recognition task by asking the subjects to respond positively to words that were associated with a target list and to reject both new words and studied words that were associated with the other (nontarget) list. These latter items were comparable with the rearranged pairs in Experiments 1 and 2, which likewise put together two fa-

miliar cues that had not been encoded together. We predicted that the subjects in this experiment would show evidence for recall-to-reject processing, because of the associative nature of the retrieval task, even though, strictly speaking, the subjects only studied single words. We are not the first to argue that a list discrimination task is a type of associative recognition: Indeed, classic work on recognition memory has emphasized the parallels between list-specific recognition judgments and word pair memory judgments, in terms of the associative nature of both tasks (Anderson & Bower, 1974; Hintzman, Block, & Summers, 1973).

Two similar studies have been conducted recently. First, Yonelinas and Jacoby (1994, Experiment 3) had subjects study sets of word lists and then make list-specific recognition judgments. This experiment used a response signal procedure, but with only two test lags (600 and 1,600 msec). Although the false alarms to completely new items did not differ with response signal lag, false alarms to words from the nontarget list decreased slightly from the 600- to the 1,600-msec lag. This pattern of data is consistent with a decrease in similar-new discrimination that would be indicative of a recall-to-reject process. However, Yonelinas and Jacoby did not report  $d_L$  or  $d'$ , so the predictions of the recall-to-reject account could not be evaluated.

More recently, Hintzman, Caulton, and Levitin (1998) reported list-discrimination experiments that also used a response signal paradigm. In their Experiment 1, subjects studied lists of words and were given a list-discrimination test after each study list, so that List 2 on test trial  $n$  became List 1 on trial  $n+1$ . Some of the test words occurred only in List 1; others occurred only in List 2; the rest were either completely new or had been studied on both Lists 1 and 2. They did not report  $d_L$ s for the nontarget list items, as compared with completely new foils, but the pattern of data shown in their Figure 2 suggests that the false alarms to words that had been studied on the nontarget list (i.e., familiar words that were to be rejected) decreased more rapidly over time than did the false alarms to completely new items. That is, their data provide some support for our prediction that an item recognition task that requires list discrimination would involve a recall-to-reject process. However, their task was rather complex;

subjects were required to keep track of the list(s) on which a word had been studied, and also to keep track of what the appropriate label for each list was at the moment of testing. The complexity of the task may have led the subjects to adopt unusual response strategies not typically found in recognition judgments. Hintzman et al. (1998) simplified the task for the subjects in their Experiment 2, in that study words occurred on only one list and testing occurred after pairs of study lists. However, the subjects who made list discriminations were only tested on studied words from the two lists (there were no truly new test items), so  $d_L$  could not be computed or estimated for those data. The design of our own Experiment 3 is similar to that of Experiment 2 of Hintzman et al. (1998), but with the addition of completely new test words.

### EXPERIMENT 3

#### Method

**Subjects.** There were 15 volunteers from the same population as that in the previous experiments. They were paid \$8 for each of five experimental sessions.

**Stimuli.** The stimuli were 1,680 words selected at random from the larger pool of words used for Experiment 2. Stimulus selection was different for each subject. Of these words, 1,120 were randomly assigned to 40 pairs of study lists; each study list consisted of 14 words. Eight pairs of study lists were used in each of the five experimental sessions. Each pair of study lists had one accompanying recognition test, which consisted of the 28 studied words (14 from each list), as well as 14 completely new, unrelated words. The order of occurrence of the test words was randomized for each test and for each subject. Across test blocks and experimental sessions, there were 64 trials of each type at each response signal lag (100, 300, 500, 750, 1,000, 1,250, and 2,000 msec).

**Procedure.** The procedure was similar to that of Experiment 2, with only a few minor changes. In this experiment, the subjects studied two word lists prior to each of the eight recognition tests in each experimental session. Each study list contained 14 words shown for 3 sec each; the beginning of List 2 was clearly indicated (i.e., the subjects were told to "Press Enter for List 2"). For half of the recognition tests in each experimental session, the subjects were told to respond positively (by pressing the Z key) to words studied on List 1 and to respond negatively (by pressing the / key) to new words and to words from List 2. For the remaining tests, they were told to respond positively to words studied on List 2 and negatively to new words and words studied on List 1. Lists 1 and 2 occurred as the critical, target list in a different random order for each experimental session and for each subject. The subjects were informed which list was the target list for that recognition test only after the study of both lists in that block.

#### Results

The data from each subject's initial session were excluded as training trials. In addition, the data were lost

from Subject 4's third session owing to experimenter error, and a power outage caused the loss of the data from the final three blocks of Subject 1's second session. Of the remaining data, 18.5% were excluded from the statistical analyses because the responses were more than 350 msec or less than 100 msec after the response signal.<sup>7</sup>

For completeness, the hit and false alarm data are reported in Table 6. The subjects made more positive responses overall when List 1 was the target list, but target list (1 or 2) did not interact with any other variables, so we have collapsed over target lists in Table 6 and in all statistical analyses of these data. Unlike many past experiments, the false alarm rate to completely new items did not decrease systematically at longer response signal lags. Also, the pattern of false alarms for nontarget-list items was not obviously suggestive of multiple processes. However, again our main focus is the  $d_L$  analysis, rather than the raw scores, because the  $d_L$  measures are focused specifically on the recall-to-reject prediction.

The acceptance rates (in terms of  $d_L$ ) for the words from the target list and for the words from the nontarget list are shown in Figure 5 as a function of response signal lag. Clearly, the subjects were able to discriminate words from the target list and completely new foils: Their  $d_L$ s were well above zero. It would also appear that a recall-to-reject mechanism was operating, in that the acceptance of words from the nontarget list initially increased somewhat and then decreased later in processing, as compared with the acceptance of completely new foils.

To evaluate these observations, we conducted model-based analyses of the data in Figure 5. Specifically, we applied the monotonic and nonmonotonic functions (Equations 1 and 2) to the  $d_L$  data from this experiment. Although the monotonic fit was quite good ( $R^2_{\text{adjusted}} = .980$ ,  $RSS = .418$ ), the nonmonotonic fit was significantly better [ $R^2_{\text{adjusted}} = .984$ ,  $RSS = .251$ ;  $\chi^2(2) = 7.140$ ,  $p < .05$ ]. The best-fitting parameter values for both functions are shown in Table 7, and the predicted values of the nonmonotonic functions are shown as the curves in Figure 5.

As in Experiments 1 and 2, we also fit both functions to the data provided by each subject individually. Of the 15 subjects, 3 provided data that were significantly better fit by the nonmonotonic than by the monotonic function. With  $\alpha = .05$ , there is less than a 1% chance of 3 significant results out of 15 by chance alone. In addition, the incremental fit of the nonmonotonic model, summed over all the subject's individual fits, was highly significant [ $\chi^2(30) = 61.025$ ,  $p < .001$ ].

#### Discussion

Despite subjects' having studied individual words, success on the list-discrimination task required the retrieval of specific associative information (i.e., in which list was this word studied). The associative nature of the task led again to clear evidence for recall-to-reject processing in recognition. Of course, it could be argued that any recognition task would require some kind of associative or context-specific information to be retrieved. For example, in an ordinary recognition experiment with a single

**Table 6**  
Mean Hit and False Alarm (FA) Rates  
for the Stimuli in Experiment 3

Item	Response Signal Lag (msec)						
	100	300	500	750	1,000	1,250	2,000
Target list (hit)	.16	.32	.42	.49	.53	.52	.54
Nontarget list (FA)	.16	.19	.25	.28	.26	.27	.26
New (FA)	.14	.14	.14	.13	.13	.16	.16

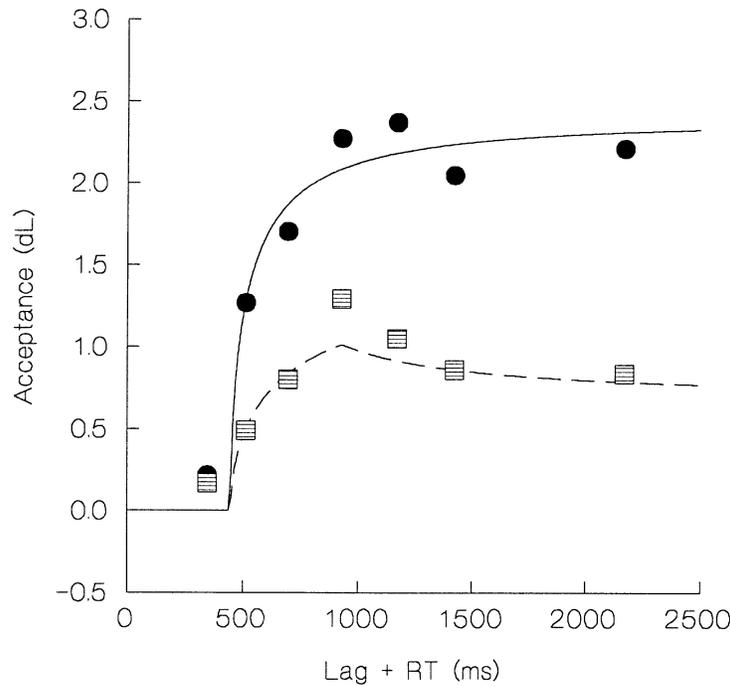


Figure 5. Mean acceptance rates, in terms of  $d_L$ , for the target list (circles) and nontarget list (squares) test items in Experiment 3. The curves show the best-fitting nonmonotonic functions.

list of words, subjects still must distinguish list-specific memories from observations of the words outside the experimental context (see, e.g., Estes, 1994). Apparently, the associative nature of the retrieval task has to be of some degree of difficulty. Discriminating between two highly similar list contexts within an experiment would seem to encourage or necessitate recall-like processes more than would simply discriminating an experimental context from the outside world.

One possible interpretation of the difference between distinguishing list membership (List 1 or List 2) of a test word and deciding whether it occurred in the experimental context is a *mutual exclusivity* assumption. Specifically, in Experiment 3, the subjects studied each word on only one of the two lists in a study block. Presumably, they could use this information when asked to respond positively to words from one of the lists: If they could recall the word's occurrence on the alternative, nontarget list, they could be certain that it had not occurred on the

target list. In contrast, consider the simpler recognition paradigm in which subjects merely decide whether a word occurred on a single study list. There, recalling the test word from a preexperimental context would not be informative about its presence or absence in the experimental study list. These differences might suggest that a recall-to-reject process would be an effective strategy in Experiment 3, but not in simple item recognition tasks (e.g., Rotello & Heit, 1999).

However, there is some evidence that shows that the use of mutually exclusive stimuli is not sufficient to promote the use of recall-to-reject processing. Rotello and Heit (1999, Experiment 2) directly evaluated the effect of a mutual exclusivity assumption in item recognition. The subjects studied pseudowords (such as PRUMIR) and were tested on studied items (PRUMIR), unrepresented pseudowords that were similar to a studied item (e.g., PRAMIR or PRAMAD), and completely new foils (e.g., DECHOP). In the *paired* condition, the subjects studied both members

Table 7  
Best-Fitting Parameter Values for the Monotonic and Nonmonotonic Fits to the  $d_L$  Data in Experiment 3

Item	Monotonic Model			Nonmonotonic Model				
	Intercept	Asymptote	Rate	Intercept	Lag*	Asy. 1	Asy. 2	Rate
Target list	472.4	2.35	109.0	448.7	930.0	2.42	—	165.7
Nontarget list	472.4	1.06	97.1	448.7	930.0	1.42	.68	465.7

Note—Asy. 1, the earlier asymptote; Asy. 2, the later (lower) asymptote.

of a similar pair (i.e., both  $X$  and  $X'$ ) about a third of the time. For these subjects, intuition suggested that recall-to-reject processing would be an unlikely strategy, because recalling  $X$  did not rule out the presence of  $X'$  on the study list. In the *unpaired* condition, the stimuli were mutually exclusive: The subjects never studied both members of a similar pair (i.e., if  $X$  was presented for study,  $X'$  was not). Intuition suggested that recall-to-reject processing would be very likely in the unpaired condition. Contrary to intuition, no evidence for recall-to-reject processing emerged in either condition: The mutually exclusive nature of the stimulus pairs in the unpaired condition did not facilitate use of a recall-to-reject mechanism. Furthermore, in Hintzman and Curran (1994), the word pairs were organized in a mutually exclusive fashion, but our analyses of their data (Rotello & Heit, 1999) did not show evidence for recall-to-reject processing.

A remaining question, of course, is whether mutually exclusive stimuli are *necessary* for the use of recall-to-reject processing. One way to address this issue is to compare experimental conditions likely to generate recall-to-reject processing that involve mutually exclusive stimuli, such as the associative recognition experiments reported here, with similar experimental conditions that did not use mutually exclusive pairs, such as those in Dyne, Humphreys, Bain, and Pike (1990). They had subjects study some pairs of words that were mutually exclusive (i.e., the words occurred in one study pair only;  $A-B$ ,  $C-D$ ) and others that were not (i.e., the words occurred in more than one study pair;  $A-B$ ,  $A-C$ ). If mutual exclusivity of stimuli were a necessary condition for recall-to-reject processing, it would appear in our Experiment 1 (which it did), but not in Dyne et al.'s experiment. Unfortunately, the subjects in Dyne et al.'s experiments were specifically instructed not to use a recall-to-reject strategy, and instructions about the use of a recall-to-reject strategy have recently been found to be effective (Rotello et al., 2000). Moreover, the Dyne et al. study did not measure recognition over the time course of judgment. Thus, whereas we can rule out the sufficiency of mutually exclusive stimuli for recall-to-reject processing, we cannot draw any strong conclusions on the necessity of mutually exclusive stimuli.

## GENERAL DISCUSSION

These experiments have tested the dominant account of recall in recognition, the recall-to-reject account, which makes a distinctive prediction about the ability of subjects to discriminate completely new test stimuli from stimuli that are new but similar to studied materials. Specifically, the similar test foils should be increasingly rejected, relative to a baseline of completely new items, late in processing (i.e., after information from the recall component of recognition becomes available). The data from Experiments 1 and 2 are consistent with that prediction: We did find that rearranged pairs were increasingly rejected, relative to new pairs, after about 1,000 msec

of processing. Many researchers have referred to the recall-to-reject account of recognition, particularly in the context of associative recognition (e.g., Clark, 1992, p. 241; Clark & Burchett, 1994, p. 57; Gronlund & Ratcliff, 1989, p. 857; Humphreys, 1976, 1978; Mandler, 1980), but these data are the most distinctive evidence in support of such a hypothesis, appropriately addressing the time course of processing. Moreover, the data from Experiments 2 and 3 indicate that the use of a recall-to-reject process is a consequence of associative information's being required to recognize the test probe, not a consequence of merely having to study words in pairs.

The data from these three experiments are consistent with recall-to-reject processing, but we would not rule out the possibility that other recall-like processes have some role in recognition. Rotello and Heit (1999) reviewed other possible two-process descriptions of recognition and described their predictions. For example, according to the exhaustive search account, a familiarity-based process is supplemented with a complete search through a set of memory traces aimed at finding an exact match to the test item. The exhaustive search would facilitate the rejection of both similar foils and completely new items and thus does not make a distinctive prediction about the difference in acceptance rate for these two kinds of items. Therefore our  $d_L$  results for associative recognition do not especially support the exhaustive search account. Another possible two-process account would be the recall-to-accept account, in which a second, recall-like process helps people to accept studied items as old. The recall-to-accept account does not make predictions about the form of responses for similar foils, and it certainly does not predict the nonmonotonic pattern for similar foils that we found in all three experiments. These experiments do not provide distinctive evidence for an additional recall-to-accept process, although we admit that our data do not rule it out either (these experiments were not designed to assess the evidence for this account).

Also, we note that our results for associative recognition do not rule out all possible single-process, non-recall-based accounts of recognition. For example, Rotello and Heit (1999) argued that a single process may apply over the entire time course of judgment but that it makes use of different information at different points in time (e.g., Brockdorff & Lamberts, 2000; Lamberts, 1995). That is, a test pair might initially seem familiar on the basis of its salient features, but later on, when more detailed information is available from the test stimulus, the pair might be a worse match to what is in memory. In this case, the rearranged pairs might initially seem familiar; indeed, they would be familiar because both words in each pair had been studied. However, those same rearranged pairs would seem less familiar as additional information about the more specific details about the test pair is compared with what is in memory. Ratcliff and his colleagues have made the very similar suggestion that individual feature information would be available early in processing but that information about feature relations or associations

would be available later in the time course of judgment (Gronlund & Ratcliff, 1989; Ratcliff & McKoon, 1989), leading to early false alarms on rearranged test pairs. Analogously, these early and late types of information might come from different sources, such as visual and auditory information (see McElree, Dolan, & Jacoby, 1999).

Still, there is some converging evidence that generally points to the two-process account, in the absence of a more completely developed one-process account. For example, in the introduction, we described data reported by Humphreys (1978), who showed that the probability of recognizing an intact ( $A-B$ ) word pair could be predicted from the probabilities of recognizing  $B$  alone and of recalling  $B$  given  $A$  as a cue (if  $B$  was not recognized). Humphreys' fits to his data were strongly indicative of a recall-based process's contributing to associative recognition.

Indirect evidence for the use of recall has been reported by Clark and his colleagues in forced-choice associative recognition experiments (Clark & Hori, 1995; Clark, Hori, & Callan, 1993). In those experiments, subjects studied pairs of words, such as  $A-B$ ,  $C-D$ , and  $E-F$ . At test, they were presented with triads of word pairs and were asked to identify the pair that had been studied. There were two kinds of test triads: OLAP trials, in which one of the test words appeared in all three test pairs (e.g.,  $A-B$ ,  $A-D$ ,  $A-F$ ), and NOLAP trials, in which all of the test words were unique (e.g.,  $A-B$ ,  $C-F$ ,  $G-I$ ). After short study lists (i.e., 29 word pairs), recognition was best in the NOLAP triads (Clark et al., 1993). Clark et al. argued that the NOLAP advantage was a consequence of the larger number of unique test words in the NOLAP condition than in the OLAP condition (i.e., six rather than four), each of which could be used as a cue in a mental cued recall task to evaluate the test pair. After longer study lists (100 word pairs), when such a cued recall strategy would be less effective, there was no difference in the recognition performance on NOLAP and OLAP triads (Clark & Hori, 1995).

In comparison, the evidence against the use of recall in associative recognition is relatively weak. Gronlund and Ratcliff (1989, Experiment 4) presented 2 subjects with pairs of words to study, then tested them on intact, rearranged, and new pairs in a response signal paradigm. The critical aspect of their experiment was the manner in which test pairs were presented: The left-hand word of each test pair was presented 200 msec before the right-hand word. Gronlund and Ratcliff's intention was to give the recall process a *head start* by allowing processing on the first word (including cued recall of the second word) to begin before processing on the right-hand word. They found no recognition advantage in this experiment, as compared with an earlier experiment in which no head start was provided. Thus, they concluded that recall was unlikely to operate in associative recognition. Given that it takes approximately 200 msec to read a word (Rayner & Pollatsek, 1989) and that subjects read from left to right, however, their 200-msec head start actually provided

little to no additional processing time for the left-hand word. In other words, the amount of time it would take subjects to read the left-hand word in the test pair is nearly exactly the same amount of time as occurs before the onset of the right-hand word in their delayed task. Hence, there is no real advantage provided by the head start, and the evidence against the recall process in associative recognition is thin.

We conclude, therefore, that the bulk of the extant data are consistent with our interpretation of our time course results as indicating a recall-to-reject process in associative recognition. Together with the lack of evidence for recall-to-reject processing in item recognition (Rotello & Heit, 1999), these results present a challenge for accounts of recognition memory. Memory models need to be flexible enough to predict a second, recall-like process for pair recognition and list-specific recognition, but not for item recognition.

## REFERENCES

- ANDERSON, J. R., & BOWER, G. H. (1974). Interference in memory for multiple contexts. *Memory & Cognition*, *2*, 509-514.
- ATKINSON, R. C., & JUOLA, J. F. (1973). Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV* (pp. 583-612). New York: Academic Press.
- ATKINSON, R. C., & JUOLA, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary development in mathematical psychology: Learning, memory, and thinking* (pp. 243-293). New York: Freeman.
- BEGG, I. (1979). Trace loss and the recognition failure of recalled words. *Memory & Cognition*, *7*, 113-123.
- BOROWIAK, D. S. (1989). *Model discrimination for nonlinear regression models*. New York: Dekker.
- BROCKDORFF, N., & LAMBERTS, K. (2000). A feature-sampling account of the time course of old-new recognition judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *26*, 77-102.
- CLARK, S. E. (1992). Word frequency effects in associative and item recognition. *Memory & Cognition*, *20*, 231-243.
- CLARK, S. E., & BURCHETT, R. E. R. (1994). Word frequency and list composition effects in associative recognition and recall. *Memory & Cognition*, *22*, 55-62.
- CLARK, S. E., & GRONLUND, S. D. (1996). Global matching models of memory: How the models match the data. *Psychonomic Bulletin & Review*, *3*, 37-60.
- CLARK, S. E., & HORI, A. (1995). List length and overlap effects in forced-choice associative recognition. *Memory & Cognition*, *23*, 456-461.
- CLARK, S. E., HORI, A., & CALLAN, D. E. (1993). Forced-choice associative recognition: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *19*, 871-881.
- CLARK, S. E., & SHIFFRIN, R. M. (1992). Cuing effects and associative information in recognition memory. *Memory & Cognition*, *20*, 580-598.
- DOSHER, B. A. (1984). Discriminating preexperimental (semantic) information from learned (episodic) associations: A speed-accuracy study. *Cognitive Psychology*, *16*, 519-555.
- DOSHER, B. A., & ROSEDALE, G. (1991). Judgments of semantic and episodic relatedness: Common time-course and failure of segregation. *Journal of Memory & Language*, *30*, 125-160.
- DYNE, A. M., HUMPHREYS, M. S., BAIN, J. D., & PIKE, R. (1990). Associative interference effects in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *16*, 813-824.
- ESTES, W. K. (1994). *Classification and cognition*. New York: Oxford University Press.

- FRANCIS, W. N., & KUČERA, H. (1982). *Frequency analysis of English usage*. Boston: Houghton Mifflin.
- GILLUND, G., & SHIFFRIN, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, **91**, 1-67.
- GREGG, V. H. (1976). Word frequency, recognition and recall. In J. Brown (Ed.), *Recall and recognition* (pp. 183-216). Chichester, U.K.: Wiley.
- GRONLUND, S. D., & RATCLIFF, R. (1989). Time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **15**, 846-858.
- HEIT, E. (1993). Modeling the effects of expectations on recognition memory. *Psychological Science*, **4**, 244-252.
- HEIT, E. (1998). Influences of prior knowledge on selective weighting of category members. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20**, 712-731.
- HINTZMAN, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, **95**, 528-551.
- HINTZMAN, D. L., BLOCK, R. A., & SUMMERS, J. J. (1973). Contextual associations and memory for serial position. *Journal of Experimental Psychology*, **97**, 220-229.
- HINTZMAN, D. L., CAULTON, D. A., & LEVITIN, D. J. (1998). Retrieval dynamics in recognition and list discrimination: Further evidence of separate processes of familiarity and recall. *Memory & Cognition*, **26**, 449-462.
- HINTZMAN, D. L., & CURRAN, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory & Language*, **33**, 1-18.
- HOCKLEY, W. E. (1994). Reflections of the mirror effect for item and associative recognition. *Memory & Cognition*, **22**, 713-722.
- HOCKLEY, W. E., & CRISTI, C. (1996). Tests of encoding tradeoffs between item and associative information. *Memory & Cognition*, **24**, 202-216.
- HUMPHREYS, M. S. (1976). Relational information and the context effect in recognition memory. *Memory & Cognition*, **4**, 221-232.
- HUMPHREYS, M. S. (1978). Item and relational information: A case for context independent retrieval. *Journal of Verbal Learning & Verbal Behavior*, **17**, 175-187.
- JACOBY, L. L., & DALLAS, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, **110**, 306-340.
- JONES, C. M., & HEIT, E. (1993). An evaluation of the total similarity principle: Effects of similarity on frequency judgments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 799-812.
- LAMBERTS, K. (1994). Towards a similarity-based account of compatibility effects. *Psychological Research*, **56**, 136-143.
- LAMBERTS, K. (1995). Categorization under time pressure. *Journal of Experimental Psychology: General*, **124**, 161-180.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
- MANDLER, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, **87**, 252-271.
- MANDLER, G. (1991). Your face looks familiar but I can't remember your name: A review of dual process theory. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory to data: Essays on human memory in honor of Bennet B. Murdock* (pp. 207-225). Hillsdale, NJ: Erlbaum.
- MAY, R. B., & SANDE, G. N. (1982). Encoding expectancies and word frequency in recall and recognition. *American Journal of Psychology*, **95**, 485-495.
- MCÉLREE, B., DOLAN, P. O., & JACOBY, L. L. (1999). Isolating the contributions of familiarity and source information to item recognition: A time-course analysis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **25**, 563-582.
- MCGEE, R. (1980). Imagery and recognition memory: The effects of relational organization. *Memory & Cognition*, **8**, 394-399.
- NOSOFSKY, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 700-708.
- PIKE, R. (1984). A comparison of convolution and matrix distributed memory systems. *Psychological Review*, **91**, 281-294.
- RAAIJMAKERS, J. G., & SHIFFRIN, R. M. (1992). Models for recall and recognition. *Annual Review of Psychology*, **43**, 205-234.
- RATCLIFF, R. (1978). A theory of memory retrieval. *Psychological Review*, **85**, 59-108.
- RATCLIFF, R. (1980). A note on modeling accumulation of information when the rate of accumulation changes over time. *Journal of Mathematical Psychology*, **21**, 178-184.
- RATCLIFF, R., & MCKOON, G. (1982). Speed and accuracy in the processing of false statements about semantic information. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **8**, 16-36.
- RATCLIFF, R., & MCKOON, G. (1989). Similarity information versus relational information: Differences in the time course of retrieval. *Cognitive Psychology*, **21**, 139-155.
- RAYNER, K., & POLLATSEK, A. (1989). *The psychology of reading*. Hillsdale, NJ: Erlbaum.
- ROTELLO, C. M., & HEIT, E. (1999). Two-process models of recognition memory: Evidence for recall-to-reject? *Journal of Memory & Language*, **40**, 432-453.
- ROTELLO, C. M., MACMILLAN, N. A., & VAN TASSEL, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory & Language*, **43**, 67-88.
- SNODGRASS, J. G., & CORWIN, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, **117**, 34-50.
- SUMBY, W. H. (1963). Word frequency and the serial position effect. *Journal of Verbal Learning & Verbal Behavior*, **1**, 443-450.
- WHITTEN, W. B., II, SUTER, W. N., & FRANK, M. L. (1979). Bidirectional synonym ratings of 464 noun pairs. *Journal of Verbal Learning & Verbal Behavior*, **18**, 109-127.
- WICKELGREN, W. A., & CORBETT, A. T. (1977). Associative interference and retrieval dynamics in yes-no recall and recognition. *Journal of Experimental Psychology: Human Learning & Memory*, **3**, 189-202.
- 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., & JACOBY, L. L. (1994). Dissociations of processes in recognition memory: Effects of interference and of response speed. *Canadian Journal of Experimental Psychology*, **48**, 516-534.

## NOTES

1.  $d_L$  is equal to roughly 1.67 times  $d'$  and is computed as

$$d_L = \ln \left[ \frac{HR(1-FA)}{FA(1-HR)} \right],$$

where  $HR$  is the hit rate and  $FA$  is the false alarm rate.

2. We also analyzed the data by using a 100- to 500-msec inclusion window. That analysis, from which only 5.4% of the data were excluded on average, led to the same conclusions.

3. The use of a single intercept was based on the assumption that the very earliest stages of processing are probably identical for all types of stimuli. For the sake of generality, we also fit the data from each experiment with monotonic and nonmonotonic models in which the intercepts were allowed to vary for the different stimulus types; these alternative model applications led to the same conclusions in every case and hence will not be reported.

4.  $R^2_{\text{adjusted}}$  is the proportion of variance accounted for by the model, adjusted for the number of free parameters:

$$R^2_{\text{adjusted}} = 1 - \frac{\sum_{i=1}^N (d_i - \hat{d}_i)^2}{\sum_{i=1}^N (N-k)},$$

$$\frac{\sum_{i=1}^N (d_i - \bar{d})^2}{\sum_{i=1}^N (N-1)},$$

where  $N$  is the number of data points ( $d_i$ ),  $\hat{d}_i$  is the predicted value,  $k$  is the number of free parameters, and  $\bar{d}$  is the overall mean.

5. These nested models were compared by using the technique of Borowiak (1989). In brief, when Model A is a nonlinear model with  $a$  free parameters estimated using a least-squares criterion, and B is a restricted version of this model with  $b$  free parameters, the likelihood ratio statistic is

$$\lambda = \left( \frac{RSS_A}{RSS_B} \right)^{\frac{k}{2}}$$

where  $RSS$  is the residual sum of squares of the model and  $k$  is the number of data points to be predicted (for this experiment, 12). Borowiak showed that  $-2 \ln(\lambda)$  has a  $\chi^2$  distribution with  $(a - b)$  degrees of freedom. (See Heit, 1998, Rotello & Heit, 1999, and Lamberts, 1994, for

other applications of this technique.) The difference in number of free parameters between models indicated two degrees of freedom, and we employed an  $\alpha = .05$  criterion unless otherwise stated.

6. As in Experiment 1, we also analyzed the data by using a 100- to 500-msec inclusion window. That analysis, which excluded an average of 15.3% of the data, led to the same conclusions.

7. We also analyzed the data by using a 100- to 500-msec inclusion window. That analysis, which excluded an average of 11.6% of the responses, led to the same conclusions.

(Manuscript received November 9, 1998;  
revision accepted for publication October 21, 1999.)