

On the flexibility and the fallibility of associative memory

KENNETH J. MALMBERG

University of South Florida, Tampa, Florida

AND

JING XU

Iowa State University, Ames, Iowa

We report the results of four experiments in which we explored the flexibility and fallibility of associative recognition memory. In each experiment, pairs were studied one or more times, and the task was to discriminate intact from rearranged pairs. The critical findings are that the pattern of false alarm rates was dependent on the nature of the recognition procedure (e.g., ratings vs. *yes-no*) and the situation in which the task was performed. The specific pattern of findings suggest that subjects adopt different recognition strategies in order to achieve a desired level of performance in the most efficient manner possible by varying the degree to which they base their decisions on familiarity versus recollected information. Implications for theories of recognition memory are discussed.

The limitations of memory are widely known. Laypersons usually condemn their “terrible memories.” In doing so, they fail to appreciate the remarkable flexibility of human memory, which represents and retrieves varieties of information obtained from different perceptual modalities in the service of countless daily tasks, each performed in countless situations. Flexibility and fallibility are hallmarks of human memory. The subject of the present investigation is what the fallibility of recognition memory can tell us about its flexibility.

Recognition tasks require the discrimination of studied items (or *targets*) from unstudied items (or *foils*). Recognition is successful when targets are endorsed (i.e., *hits*) and foils are rejected. Likewise recognition is unsuccessful when targets are rejected and foils are endorsed (i.e., *false alarms*). Thus, recognition memory improves as the difference between hit rates (HRs) and false alarm rates (FARs) increase.

Over the past dozen or so years, a dominant question in episodic memory has been how to characterize retrieval from recognition memory. One theory holds that recognition can be characterized by a single retrieval process (e.g., Donaldson, 1996; Dunn, 2004; Wixted & Stretch, 2004), and another theory holds that at least two retrieval processes are required (e.g., Joordens & Hockley, 2000; Reder et al., 2000; see Yonelinas, 2002, for a review). The former class of models is referred to as *single-process models*, and the latter is referred to as *dual-process models*. The debate is often framed so that only one model would be needed to characterize all recognition tasks, and accordingly, the same task has always been performed in the same way.

These assumptions are parsimonious, but they might overlook the inherent flexibility of human memory. For instance, there are a wide variety of recognition tasks (including single-item recognition, associative recognition, source recognition, eye witness identification, and memory scanning) and procedures (yes-no, ratings, forced choice), and they are performed in countless situations. Thus, one might assume that these tasks at least have the potential to be performed in different ways. In the present research, the flexible nature of recognition memory is explored in the context of the single- versus the dual-process debate. One hypothesis is that different recognition tasks can be performed in different ways; perhaps, even the same recognition task can be performed in different ways (e.g., Rotello, Macmillan, & Van Tassel, 2000; Yonelinas, 1997).

MODELS OF ASSOCIATIVE RECOGNITION

A major theoretical advance in human memory theory has been in its formalization, and here, we will describe three classes of models that recently have been applied to associative recognition. We will pay particular attention to the predictions that they make concerning the effect of pair repetitions on performance, because this is one way that they can be distinguished (Cleary, Curran, & Greene, 2001; Kelley & Wixted, 2001; Xu & Malmberg, 2007).

Single-Process Recall-Only Models

According to recall-only models, associative recognition is performed in a manner very similar to cued recall (but see Anderson & Watts, 1971; Postman & Stark, 1969).

K. J. Malmberg, malmberg@cas.usf.edu

Memory is probed with a retrieval cue that represents one or both of the items making up the test stimulus. As a result, episodic details may or may not be retrieved from memory. If details are retrieved, they will either match the test pair or mismatch the test pair. If they match, the test pair is positively endorsed. If they do not, the test pair is rejected. This is the simplest form of the *recall-to-reject* model, and it is usually assumed to belong to the threshold class of models because of the match versus mismatch basis for the recognition decision. Although they differ in their details, the recall-only assumption is incorporated into several modern models, including those described by Yonelinas (1997) and Diller, Nobel, and Shiffrin (2001).

A problem for the recall-only models is recent findings that show that increasing the number of target presentations has little or no effect on FARs (Cleary et al., 2001; Kelley & Wixted, 2001; Xu & Malmberg, 2007). It is simplest to assume that increasing the number of times that pairs are studied should steadily improve the recall-to-reject process. If so, the single-process recall-only model predicts that false alarms should steadily decrease with target presentations, and they apparently do not.

Single-Process Familiarity-Only Models

Single-process familiarity-based models assume that the basis for the recognition decision is a continuous random variable (e.g., Green & Swets, 1966), which is often referred to as *familiarity*. The simplest of these models is the *compound cue model* (e.g., Gronlund & Ratcliff, 1989), which assumes that the items that make up a pair are stored jointly in a common memory trace. When memory is probed with a retrieval cue consisting of both test items, the output of memory is a level of familiarity, and the response is positive if it exceeds a subjective criterion. Since familiarity is assumed to be a nonlinear positive function of the similarity between the retrieval cue and the contents of memory, compound cue models predict that intact pairs will have a greater average familiarity than will rearranged pairs. Strengthening targets increases the average familiarity of rearranged pairs, and thus the simplest version of the compound cue model predicts an increase in FARs with an increase in target repetitions, which is inconsistent with extant findings (Cleary et al., 2001; Kelley & Wixted, 2001; Xu & Malmberg, 2007).

A different type of familiarity-only model is the *independent cue model* (e.g., Doshier, 1984; Murdock, 1997). Like compound cue models, they assume that familiarity is the basis for recognition. Unlike compound cue models, independent cue models assume that the pairwise association of two items is represented separately from the items themselves or is otherwise stochastically independent. When a test pair is presented, the pairwise association is created anew and used to probe memory.

The strongest versions of the independent cue models assume that the probe involves only comparing the independent cue with the contents of memory (Murdock, 1982). The items making up a pair play no systematic role in the recognition decision. Since an intact pair is represented in memory but a rearranged pair is not (they match the contents of memory only randomly), intact pairs tend

to be more familiar than rearranged pairs. In addition, strengthening target pairs should have no effect on the familiarity of rearranged pairs, and hence, the independent cue predictions are supported by a null effect of repetitions on FARs (Cleary et al., 2001; Kelley & Wixted, 2001; Xu & Malmberg, 2007).

A weaker version of the independent cue model assumes that the probe of memory consists of both item and associative information (e.g., Criss & Shiffrin, 2004; Hockley, 1992; Kelley & Wixted, 2001). Thus, the familiarity of a test pair is a function of the joint familiarity of the items and the association of the items. What these models predict when targets are repeated depends on whether repetitions increase, decrease, or have no effect on the familiarity of the rearranged pairs. Hockley suggested that item and associative information might be combined for intact pairs but that only item familiarity contributes to the recognition decision involving rearranged pairs. Accordingly, false alarm rates should increase, and hence, this version of the independent cue model has been disconfirmed. In contrast, Kelley and Wixted assumed that increasing repetitions increases item familiarity but decreases the associative strength of rearranged pairs. When these two sources of evidence are of equal magnitude, the prediction is a null effect of repetitions on false alarm rates.

A Dual-Process Model

Generally speaking, dual-process models assume that familiarity and recollective details may be used as evidence on which to base a recognition judgment (Atkinson & Juola, 1974). The dual-process model that we are considering is the one Xu and Malmberg (2007) recently developed within the framework of the retrieving effectively from memory (REM) theory (Malmberg, Holden, & Shiffrin, 2004; Malmberg & Shiffrin, 2005; Malmberg, Zeelenberg, & Shiffrin, 2004; Shiffrin & Steyvers, 1997). The formal details of that model are presented in the Appendix, but at least for the moment, we need consider only five basic assumptions.

1. There are two types of information on which to potentially base a recognition decision (Atkinson & Juola, 1974). A sense of familiarity is produced in response to a probe of memory by a global-matching process (e.g., Clark & Gronlund, 1996; Shiffrin & Steyvers, 1997), and episodic details are produced by a sampling and recovering process (e.g., Malmberg & Shiffrin, 2005; Raaijmakers & Shiffrin, 1980).

2. The familiarity process provides information more quickly than does the sampling and recovery process (e.g., Diller et al., 2001; Gronlund & Ratcliff, 1989).

3. Performance of explicit memory tasks involves the interaction of permanent structural processes and strategic control processes (Atkinson & Shiffrin, 1968).

4. Recognition should be performed in the most efficient manner, and this involves a subjective weighting of the speed and accuracy with which recognition is performed (Shiffrin & Steyvers, 1997; Xu & Malmberg, 2007).

5. The decision-making steps of this model are shown in Figure 1. We assume that memory is probed with a compound cue and that the familiarity of the test pair is

compared with a subjective criterion. If it does not exceed the criterion, the response is negative. If the familiarity of the test pair does exceed the criterion and the sampling and recovery processes are able to retrieve episodic details, those details are compared with the stimulus. If they do not match the stimulus, the response is negative. If the sampling and recovery processes fail, the subject guesses, with a far greater tendency to guess positively (Malmberg, Holden, & Shiffrin, 2004).

In addition to considering how recognition *is* performed, one can consider how a recognition task *should* be performed within this framework. For instance, Malmberg, Zeelenberg, and Shiffrin (2004) concluded that when single-item recognition requires the discrimination of targets from randomly similar foils, a single-process continuous-state model suffices. Subjects adopt a single-process familiarity-based strategy when targets and foils are randomly similar, because targets are generally more familiar than foils, and it is faster than the dual-process strategy because it does not require the outcome of the sampling and recovery process. On the other hand, Malmberg, Holden, and Shiffrin (2004) argued that a dual-process strategy is used when single-item recognition requires the discrimination of targets from similar foils, because both seem relatively familiar: The episodic details provided by the sampling and recovery process are utilized to reject the otherwise familiar foils.

Xu and Malmberg (2007) proposed that associative recognition is performed in a manner close to the way that single-item recognition is performed when targets and foils are similar. Importantly, the tendency to rely on the recall-to-reject strategy is assumed to be under subjective control (perhaps, by altering the initial criterion). This tendency is captured by the a parameter of the model. Figure 2 shows the predicted effect of repetitions on HRs and FARs as a function of a . Regardless of the level of a , HRs increase with increases in repetitions. In fact, the level of a has only a negligible effect on HRs, because subjects are biased to guess *old* to otherwise familiar items even when recollection fails. The level of a has a large effect on

the pattern of FARs. When the recall-to-reject strategy is often used, FARs increase and then decrease. At intermediate levels of a , the FAR function is fairly steady after the initial increase. Hence, the dual-process model is not challenged by the null effect of repetitions on FARs (Kelley & Wixted, 2001; Malmberg, Holden, & Shiffrin, 2004), because it is possible to pick two similar points on a nonlinear function. For instance, the 1- and 12-presentation FARs are similar in Figure 2 when $a = 1.0$.

Thus, it is not possible to discriminate between independent cue models on the basis of a null pattern of FARs. On the other hand, Xu and Malmberg (2007) observed a variety of FAR functions. For unfamiliar items (e.g., pseudowords and Chinese characters), the FAR function *increased*. For familiar items (e.g., words and faces), the FAR function *decreased*. Dual-process models predict the variable false alarm functions in an a priori manner (see Figure 2). Other models might be able to do so, but they need to adopt new assumptions in order to do so. Before doing so, however, it seems prudent to further explore the effect of repetitions on associative recognition. In the following experiments, we explored the form of the FAR function for different recognition tasks and different situations. Because the methods used were very similar, we will describe them in general terms first.

GENERAL METHOD

Subjects

A total of 176 undergraduate students enrolled in introductory psychology courses at Iowa State University participated in exchange for course credit. All the subjects were sampled from the Iowa State University subject pool, which consists of approximately equal numbers of males and females, primarily between 18 and 21 years of age. Experiments 1 and 2 were conducted between November 27 and December 8, 2004, and Experiments 3 and 4 were conducted between January 24 and February 18, 2005.

Design and Materials

These associative recognition experiments required the discrimination of intact pairs from rearranged pairs. In each experiment, 160 pairs of nouns were randomly created anew for each subject. These pairs were randomly assigned to one of four 40-pair study lists and to one of the five number-of-presentation conditions. The number of pair presentations was varied on five levels within subjects. Each pair was presented for 1.2 sec of study 1, 2, 3, 6, or 12 times. There was at least 1 intervening pair between each presentation of a given pair. Following each study list, single digits were mentally added for 20 sec. Four of the pairs in each condition were randomly determined to serve as intact pairs, and the remaining 4 pairs served as rearranged pairs. The rearranged pairs were formed by combining single items from the same number-of-presentation condition.

Procedure

The methods used in the four experiments varied only in the task instructions given to the subjects. The tasks used in Experiments 1–4 were, respectively, confidence ratings, *yes–no* recognition, *yes–no* recognition with a 2-sec delay, and confidence ratings when the test list contained pairs constructed from unstudied items (i.e., *XY* pairs, in addition to rearranged pairs). We will refer to these as the *ratings*, *yes–no*, *yes–no delay*, and *XY* experiments, respectively. In all the experiments, pairs of words were presented one word above the other both at study and at test. The subjects were instructed that their task was to discriminate intact from rearranged pairs. Test pairs were presented

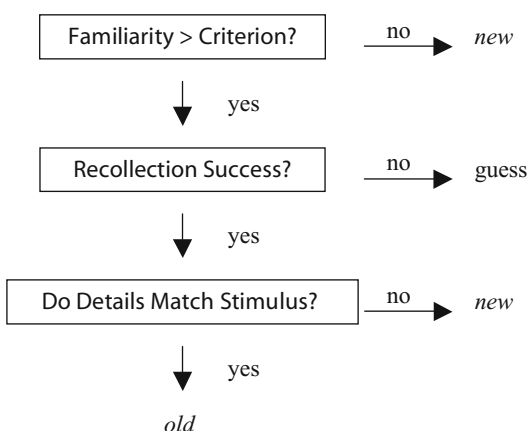


Figure 1. Flowchart for the REM dual-process model.

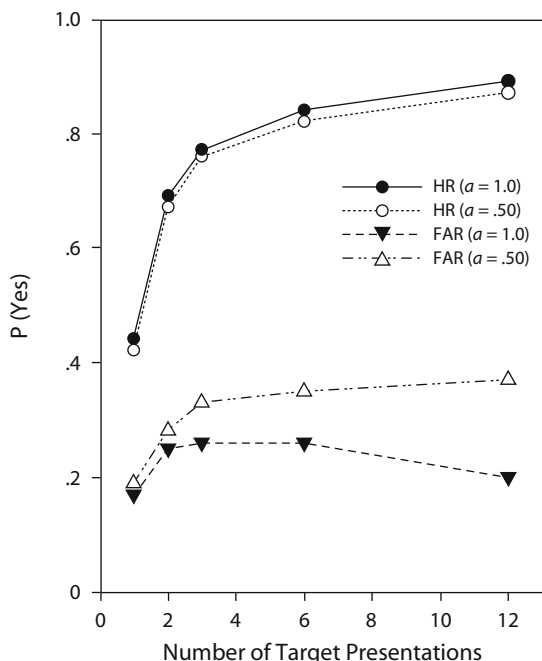


Figure 2. The effect of varying the use of a recall-to-reject strategy (a) on associative recognition. In this simulation, the contribution of recollection (a) is varied from .2 to 1.0. The parameters used in these simulations are $w = 20$, $t_1 = 4$, $u^* = .04$, $c = .7$, $g = .4$, *old-new criterion* = 2.2, $\gamma = 0.8$. HR, hit rate; FAR, false alarm rate.

until a response was made. We will note the differences between the tasks in subsequent sections that describe each experiment.

EXPERIMENT 1 Ratings

As was noted above, Kelley and Wixted (2001) found that FARs were not unaffected by increasing pair presentations. However, presentations were varied on only two levels (1 vs. 6 presentations) in those experiments; hence, it is unknown whether presentations have no effect on FARs or whether the function relating presentations to FARs is nonlinear. To obtain a better understanding of this function, we parametrically varied the number of pair presentations over a wider range and over five levels (1, 2, 3, 6, or 12 presentations). Once the form of the FAR function has been established for the ratings task, we will use it as a standard of comparison for the subsequent experiments.

Ratings Procedure

Sixty-eight subjects were instructed to judge on a 4-point Likert scale how confident they were that the pair of items had been studied together (1 = *high confidence studied*, 2 = *moderate confidence studied*, 3 = *moderate confidence not studied*, 4 = *high confidence not studied*). The subjects were free to distribute their responses over the ratings as they saw fit. To make their response, the subjects were to enter the appropriate number into the computer, using the keyboard.

Results and Discussion

All the contrasts that are reported are from repeated measures ANOVAs. The t tests are two tailed, and the standard of reliability is .05 for planned comparisons. The HRs and FARs for all the experiments are plotted as a function of the number of pair presentations in Figure 3. Table 1 reports these in a numerical fashion and shows the results of a regression analysis of the relationship between the number of target presentations and FARs.

Figure 3A shows that HRs increased [$F(1,67) = 502.28$, $p < .0005$] and FARs remained fairly steady as the number of target presentations increased [linear, $F(1,67) = 0.03$, $p = .87$; quadratic, $F(1,67) = 2.76$, $p = .10$]. These results replicate the findings of Kelley and Wixted (2001); there was no reliable difference between the FARs in the one-presentation condition and those in the six-presentation condition [$t(67) = 1.55$, $p = .13$]. Because the ANOVA contrasts assume equal intervals between the levels of the number of presentations, Table 1 also reports the results of a regression analysis of the relationship between FARs and the number of presentations. Neither the linear nor the quadratic relationship accounted for much of the variance in the FARs, although more was accounted for by the quadratic trend.¹

EXPERIMENT 2 Yes-No Recognition

The goal of the present experiment was to generalize the findings from the ratings task, where we found a flat, almost linear FAR function, to the *yes-no* task. Many single-process models of associative recognition assume that for the ratings and the *yes-no* tasks, the same information is utilized to make a decision; hence, they predict equivalent levels of performance and the same patterns of HRs and FARs for each task. Likewise, Xu and Malmberg's (2007) dual-process model predicts that the form of the functions relating FARs to repetitions will be the same for the ratings and the *yes-no* tasks if subjects wait to make their recognition decision until all the available evidence is obtained from memory.

The *yes-no* task is, however, a simpler task than the ratings task. According to many single-process familiarity-based models, ratings are based on a comparison of a familiarity value with a set of criteria, whereas *yes-no* recognition is based on a comparison with a single criterion (Green & Swets, 1966), and additional comparisons should slow the performance in the ratings task, relative to that in the *yes-no* task. Alternatively, Baranski and Petrusic (1998) speculated that confidence is determined after an initial *yes-no* decision has been made, especially when subjects emphasize speed over accuracy. In several visual discrimination experiments, they observed that initial discrimination decisions are often reversed when a subsequent confidence rating is given. More recently, similar observations were made using a recognition memory task. Van Zandt and Moldonado-Molina (2004) had subjects perform a speeded *yes-no* task, followed by an unsped

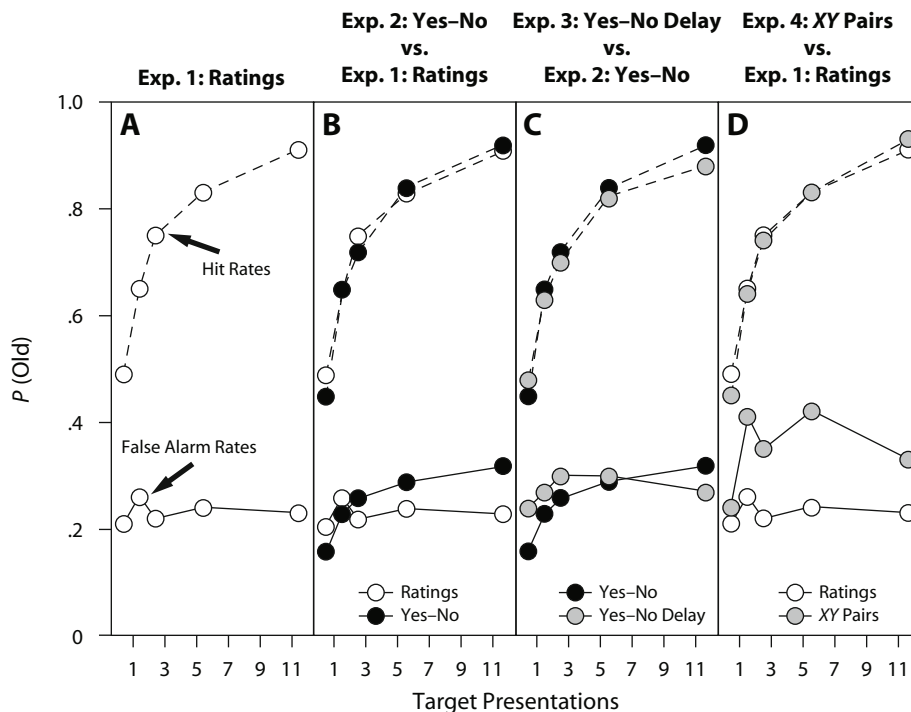


Figure 3. Results of Experiments 1-4.

ratings task. Again, the subjects often reversed their initial *yes-no* decision.

To explain these findings, Baranski and Petrusic (1998) proposed and Van Zandt and Molodono-Molina (2004) formally described a single-item recognition model that assumes, for a binary-choice task, that evidence accumulates in two counters in response to a stimulus. As soon as one counter reaches threshold, a *yes-no* decision is made. Importantly, evidence continues to accumulate after the initial *yes-no* decision has been made, and this additional evidence can influence the subsequent rating decision. Sometimes, the additional information leads to a decision reversal, particularly when the initial *yes-no* decision is incorrect.

Van Zandt and Molodono-Molina's (2004) findings also suggest, within the dual-process framework, that familiarity and recollective details might accumulate at different rates during retrieval. The time course of retrieval

during item and associative recognition was compared by Gronlund and Ratcliff (1989), using a signal-to-respond method. In these experiments, pairs were studied, and single-item targets, single-item foils, intact pairs, and rearranged pairs were tested. The subjects were instructed to respond positively to single-item targets and intact pairs and negatively to single-item foils and rearranged pairs when a signal was presented. The main finding was that the rate with which performance increased was much greater for single-item than for associative recognition, primarily because of a nonmonotonic relationship between response lag and the FARs for rearranged pairs: FARs initially increased and then decreased after about 700 msec, asymptoting at about 1,400 msec subsequent to the probe.

Gronlund and Ratcliff (1989) concluded that item familiarity accumulated more quickly than associative

Table 1
Effects of the Number of Target Presentations (With Standard Deviations) on False Alarm Rates in Experiments 1-4

Task	Number of Target Presentations						Regression Coefficients
	0	1	2	3	6	12	
Exp. 1: Ratings	–	.21 (.16)	.26 (.18)	.22 (.17)	.24 (.18)	.22 (.20)	Linear: $R^2 = .01$ Quadratic: $R^2 = .19$
Exp. 2: Yes-No	–	.15 (.13)	.23 (.20)	.26 (.20)	.29 (.23)	.32 (.27)	Linear: $R^2 = .75$ Quadratic: $R^2 = .92$
Exp. 3: Yes-No delay	–	.24 (.16)	.27 (.16)	.30 (.20)	.30 (.20)	.27 (.21)	Linear: $R^2 = .05$ Quadratic: $R^2 = .82$
Exp. 4: XY pairs	.10 (.14)	.23 (.25)	.41 (.21)	.35 (.25)	.42 (.26)	.33 (.26)	Linear: $R^2 = .03$ Quadratic: $R^2 = .58$

familiarity, within the framework of the compound cue model. However, there is a different interpretation of these findings within the framework of a dual-process model. Specifically, it might be the case that the recollected information becomes available later than familiarity.

According to the dual-process model, responding before the sampled and recovered evidence has been obtained will primarily diminish the tendency to effectively use a recall-to-reject strategy. If so, Figure 2 shows that the effect will be observed almost entirely in the pattern of FARs. HRs should be little affected, because subjects tend to guess *old* to targets regardless of whether the recollection is successful or not. Responding before the recollection process is complete, however, should result in a strictly positive relationship between repetitions and FARs. If subjects overemphasize speed and underemphasize accuracy for the *yes-no* task, lower HRs and greater FARs are predicted by the single-process models as the result of having incomplete evidence on the basis of which to make a decision.

Yes-No Procedure

The design of this experiment was exactly like the design of Experiment 1, except that 36 subjects were instructed to respond either *yes* or *no* to indicate whether the items making up a pair had been studied together. To respond *yes* and *no*, the subjects pressed the “1” and “2” keys, respectively.

Results and Discussion

Figure 3B and Table 1 show that repetitions increased HRs [$F(1,35) = 337.41, p < .0005$] and FARs (see Table 1). The linear trend in FARs was reliable [$F(1,35) = 25.78, p < .0005$], and there was a smaller quadratic trend [$F(1,35) = 3.93, p = .06$]. Again, however, the ANOVA contrast assumes equal intervals; hence, the nature of the trend is better informed by an additional regression analysis that shows that the quadratic trend accounts for more of the variance than does the linear trend.

This suggests that the pattern of FARs is different from that obtained in Experiment 1, in which a ratings task was used. A between-subjects analysis comparing the patterns of HRs and FARs for the *yes-no* and the ratings (Experiment 1) tasks indicates that task did not affect them (both $F_s < 1$); nor was there an interaction between presentation and task for HRs [$F(1,102) = 1.35$]. There was, however, a reliable interaction between the task and presentation for FARs [$F(1,102) = 16.77, p < .0005$].

These analyses confirm what is apparent in Figure 3B: The pattern of FARs, but not the pattern of HRs, depends on the nature of the recognition task. For ratings, FARs are little affected by the number of target presentations. For *yes-no* recognition, FARs increase with presentations. This disconfirms models that must predict a null effect of presentations on FARs—most notably, independent cue models that assume that item and associative information are statistically independent and that only associative information is used to probe memory (Murdock, 1997). Within the framework of these models, the present results

suggest that both item and associative information contribute to *yes-no* associative recognition, although it is unclear why, within the present independent cue frameworks, the nature of the retrieval cue depends on the distinction between ratings and *yes-no* recognition.

The combined results of Experiments 1 and 2 also disconfirm independent cue models that assume that the ratings and the *yes-no* tasks differ only in the number of criteria used to make decisions (e.g., Kelley & Wixted, 2001), since these models predict equivalent levels of performance regardless of whether a *yes-no* rating procedure is used. We note that the median latencies of correct responses were approximately twice as fast for the *yes-no* task as for the ratings task (1.8 vs. 3.4 sec; $p < .0005$). Thus, it is possible that the slower rating decisions are due to the additional decisions that are required.

It is possible that different subjects use different recognition strategies. If so, we might expect that those who rely more heavily on a recall-to-reject strategy will respond more slowly than those who rely more heavily on a familiarity-based strategy. On the other hand, the *yes-no* task might have encouraged all (or most) of the subjects to rely on the familiarity-based strategy. If so, there should be little or no relationship between response latencies and false alarms. To investigate these possibilities, we analyzed the correlation between FARs and mean latencies across subjects, and we divided the subjects evenly by assigning them to groups on the basis of the mean false alarm latencies, in order to see whether FARs would be affected by whether the subjects were assigned to the fast or the slow group. The results of both analyses showed no reliable relationship between false alarms and response latencies, which suggests that the subjects were adopting similar recognition strategies.

EXPERIMENT 3

Yes-No Delay

The variable patterns of false alarms found in Experiments 1 and 2 disconfirm the simplest forms of all the single-process models that we have considered. However, the independent cue models can be salvaged if associative information is available later than item information and if subjects adopt different strategies in response to task demands and situational factors. It makes sense within some independent cue frameworks (Criss & Shiffrin, 2004; Kelley and Wixted, 2001; Murdock, 1997) that associative information becomes available later than item information, because test items need to be processed prior to creating a compound retrieval cue. If subjects adopt a strategy that emphasizes speed over accuracy, at least some of the response might be based on item familiarity, and not on associative familiarity.

These results are also consistent with the hypothesis that the ratings task requires two different decisions (Baranski & Petrusic, 1998; Van Zandt & Moldonado-Molina, 2004). Within the framework of the dual-process model, one might suppose that initially, a familiarity value is compared with a criterion, followed by a recall-to-reject

strategy if the criterion is exceeded (as is shown in Figure 1). If the subjects in Experiment 2 emphasized speed to the detriment of accuracy, *yes* responses might sometimes have been based on familiarity when the use of recollected information would have led to a *no* decision.

The ratings task used in Experiment 1 is necessarily more time consuming than the *yes-no* task because there are two decisions. The first is a *yes-no* judgment performed on the basis of familiarity. The second is a confidence judgment, which might be based on a combination of familiarity and recollected evidence. The additional confidence decision might have allowed time for the recollected information to become available without a time cost to the subject. That is, on some test trials, an initial *yes* decision based on the pair's familiarity (see above) might be overridden by the occurrence of recollected information while the confidence judgment is being made, and this ultimately results in a *no* decision.

If the subjects emphasized speed over accuracy in Experiment 2, imposing a delay before the *yes-no* decision could be made would have allowed the subjects to utilize recollected information without a time cost (cf. Van Zandt & Moldonado, 2004). The predictions of the dual-process model are shown in Figure 2. An increase in the *a* parameter corresponds to an increase in the tendency to utilize information that has been sampled and recovered. Because both familiarity and recollected information indicate that a target pair has been studied, increasing *a* does not substantially affect HRs. However, increasing *a* will increase the tendency to reject otherwise familiar rearranged pairs on the basis of sampled and recovered information, and thus FARs will generally be lower and the function relating repetitions to FARs will be flat or inverted-U shaped.

Yes-No Delay Procedure

The *yes-no* delay condition was the same as the *yes-no* condition, except that each test pair was presented on the computer monitor for 2 sec before the subjects were prompted for their *yes-no* decision. Thirty-eight subjects participated in this experiment.

Results and Discussion

Figure 3C shows that repetitions increased HRs [$F(1,37) = 211.69, p < .0005$]. A visual inspection of Figure 3C indicates that the function relating presentations to FARs has a decidedly inverted-U-shaped nonlinearity. Table 1 shows that the linear trend in the FARs was not reliable [$F(1,37) = 1.61, p = .21$] but that the quadratic trend was reliable [$F(1,37) = 7.22, p = .011$]. The regression analysis leads to a similar conclusion, since the variance accounted for by the quadratic trend is much greater than the variance accounted for by the linear trend.

A planned between-subjects analysis of the data from Experiments 2 and 3 indicates no main effect of task (*yes-no* vs. *yes-no* delay) on HRs or FARs [both $F(1,72) < 1$]. There was also no interaction between task and presentation on HRs [$F(1,72) = 1.45, p = .23$]. However, there was a significant interaction between task and presentation on FARs [$F(1,72) = 10.01, p < .002$]. When the sub-

jects responded freely, FARs increased steadily. The delay imposed on making a *yes-no* response produced an initial increase, followed by a decrease in FARs. Moreover, the delay imposed on the recognition decision had no effect on HRs.

These results are consistent with the hypothesis that when subjects perform a *yes-no* recognition task, they sometimes tend to overemphasize speed over accuracy. The result is an overemphasis on the familiarity component of retrieval and a neglect of a slower but more accurate recollective component. When a delay in responding is imposed and the outcome of the recollective component of retrieval is made available at no time cost, subjects will use this information to reject rearranged pairs, because there is no time cost in doing so.

EXPERIMENT 4 **XY Pairs**

The present assumption is that recognition tasks can be performed in a variety of ways and that the strategy adopted seeks to balance speed with accuracy to meet a subjective goal of efficient performance. For the *yes-no* task in Experiment 2, the subjects appeared to sacrifice accuracy in favor of quicker responses, and FARs rose steadily. When, in Experiment 3, there was no time cost associated with slower responses, FARs did not rise steadily and, in fact, decreased a small amount after six pair presentations.

We note that the subjects also apparently adopted different associative recognition strategies in the experiments conducted by Postman and Stark (1969) and Anderson and Watts (1971). Postman and Stark had subjects study one or two lists of pairs. The group that studied one list served as the control group (*A-B*); a second group studied two lists that shared no common items (*A-B, C-D*), and the last group studied two lists that did share cues (*A-B, A-D*). Memory was tested for the pairs only on the first list by either cued recall or associative recognition. Importantly, the rearranged foils used in the associative recognition condition were from pairs that came only from the most recent list. Postman and Stark observed interference in the *A-B, A-D* group for cued recall, but not for associative recognition. In contrast, Anderson and Watts used foils that were constructed from items from both lists, and interference was observed. These findings demonstrate a certain amount of flexibility in associative recognition. The same task—namely, associative recognition—can be performed in different ways, depending on the testing conditions.

The difference between Postman and Stark's (1969) and Anderson and Watts's (1971) experiments was the compositions of the test lists. In Postman and Stark's experiment, the foils were always rearranged pairs formed from items studied on the last study list, whereas some of the foils were formed from pairs from the first study list in Anderson and Watts's experiment. In Postman and Stark, the foils consisted of a target item paired with an item from the most recent list. Thus, the subject needed only

to determine whether the foil item was from the most recent list. If so, the pair could be rejected. In Anderson and Watts's experiment, some of the foils were drawn from pairs from the first list. Thus, the subjects could not rely on a strategy by which items could be rejected solely on the basis of list membership and had to determine whether the two items had been studied together on the first list.

Here, we adopted a similar but different method to explore whether subjects would adapt to task conditions, by presenting at test some pairs consisting of items that were not studied (i.e., *XY* pairs).² The *XY* pairs were only randomly similar to the studied pairs, and hence, they would, on average, be far less familiar than the intact pairs. We hypothesized that under these conditions, the subjects would be less inclined to use a recall-to-reject strategy because, on the majority of the test trials, they could respond as accurately but more quickly if they used a strategy that emphasized familiarity (cf. Xu & Malmberg, 2007). If so, we would expect to see relatively high FARs, as compared with those obtained in Experiment 1, where only intact and rearranged pairs were tested. In addition, our model predicted that hit rates should not be substantially affected by the change in the strategy.

XY Procedure

Thirty-four subjects performed the confidence ratings task as in Experiment 1, but they were also told that some of the test pairs were made up of items that were not studied, and they were instructed to respond negatively to these pairs. Here, 4 of the pairs from each condition were randomly selected to be intact pairs, and 2 of the pairs were selected to be rearranged pairs. In addition, each test list contained 10 *XY* pairs, which were randomly constructed from words that were not studied. Thus, each test list contained equivalent numbers of targets (20) and foils (20).

Results and Discussion

Figure 3D shows that repetitions increased HRs [$F(1,33) = 187.20, p < .0005$]. Table 1 reports an ANOVA conducted on the FARs for the rearranged pairs (i.e., the *XY* pair FARs were not included). The linear trend in FARs was not reliable [$F(1,33) = 2.31$], but the quadratic trend was [$F(1,33) = 17.02, p < .0005$], suggesting that form of the function relating FAR to presentation is nonlinear. The FARs appear to initially increase and then decrease as the number of target presentations increase. This conclusion is supported by the regression analysis, which shows that the quadratic coefficient is much greater than the linear coefficient.

A planned comparison of the performance of the subjects in the ratings experiment (i.e., Experiment 1) with the performance of the present subjects shows that task did not affect HRs [$F(1,100) < 1$] but did affect FARs [$F(1,100) = 12.58, p < .001$]. An inspection of Figure 3D and Table 1 indicates that the FARs were substantially greater when *XY* pairs were included on the test list. In addition, there was a task \times presentation interaction on FARs [$F(1,100) = 13.07, p < .0005$], suggesting that the forms of the functions relating FAR to presentation vary

depending on whether *XY* pairs are included in the test list. When this is so, the function is much more nonlinear than when only intact and rearranged pairs are tested.³

MODEL FITTING

The different patterns of FARs suggest, within the current framework, that the contribution of recollection to associative recognition depends on the nature of the task. Specifically, the contribution of recollection to performance—*a*, in the model—appears greater in Experiments 1 (ratings) and 3 (*yes-no* delay) than in Experiments 2 (*yes-no*) and 4 (*XY* pairs).

To verify this conclusion, we conducted a simulation to determine whether the *a* parameter varies in a manner consistent with it. In this Monte Carlo simulation, all parameter values were held fixed, with the exception of *a*, which was varied from .30 to 1.0 in increments of .025. The remaining parameter values are listed in Figure 4. For each value of *a* for each experiment, 1,500 subjects were simulated. The *best-fitting* value of *a* was that which minimized the mean squared difference between the model and the data. As was expected, the contribution of recollection to associative recognition was greater for the ratings (.975) and the *yes-no* delay (.775) tasks and lesser for the *yes-no* and *XY* pairs (.700 and .500) tasks. Given that only one parameter varied between each simulation, the model fits are reasonable. The best fit occurred for the *yes-no* task (Experiment 2), with a mean squared difference of .0008, which indicates that the average deviation of the model from the data was less than .03. The worst fit occurred for the *XY* pairs task, which had a mean squared difference of .0031, which indicates that average deviation of the model from the data was about .055.

To further explore the fits of the model to the data, we plotted the best fits of the model to the data in Figure 4. The predicted rate of increase in hit rates was more rapid than that observed. The other significant departure of the model from the data was for the FARs for the *XY* pairs task. The model tended to underestimate those FARs by about .06. The problem for the model was the FAR in the 12-presentation condition. When minimizing the squared deviation, the model attempted to capture this point by lowering all the other FARs. Thus, the model underestimated the remaining FARs. To model the performance for *XY* pairs, we assumed that recollection always failed because one cannot recall what was not studied (i.e., u^* for *XY* pairs was zero). An underestimate of the FAR for *XY* pairs occurred, which was .05 versus the .10 observed in the data.

The departure of the model from the hit rates is potentially the more serious of the two shortcomings, for two reasons. First, it occurs in varying degrees in all the experiments. Second, according to the model, the contribution of recollection to performance, *q*, is limited by how well items are encoded: $\hat{q} = ac [1 - (1 - u^*)^r]^t$, where *a* is the effectiveness of the sampling and recovery processes and u^* and *c* determine how many features are encoded and how accurately they are encoded, given that a pair was studied *r* times and, each time, *t* attempts at storage were

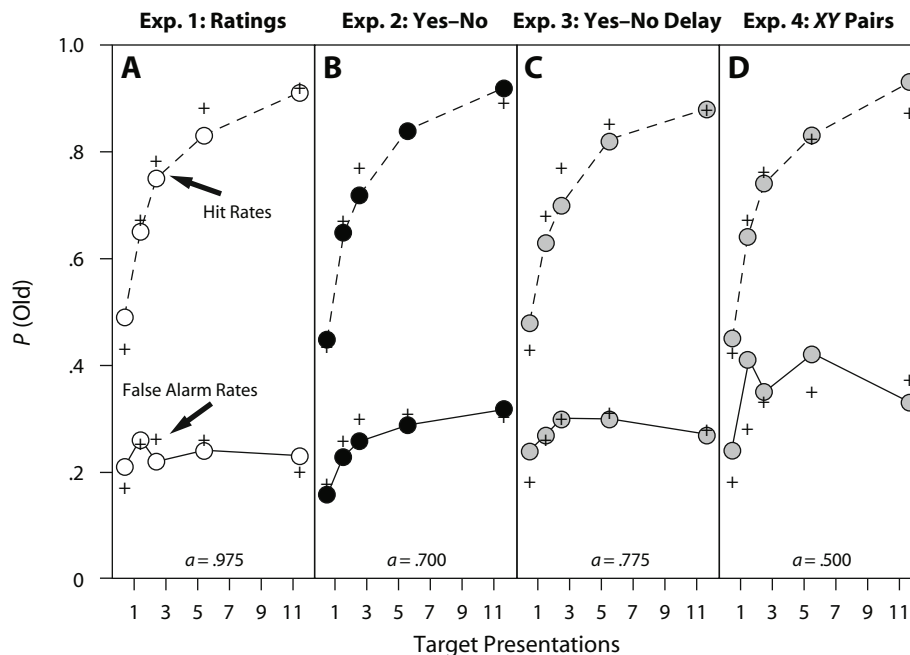


Figure 4. Dual-process model fit to Experiments 1–4. The plus signs represent the best fit of the model to the data when only the degree to which a recall-to-reject strategy is utilized (a) varies between fits. The other parameters used in these simulations are $w = 20$, $t_1 = 4$, $u^* = .04$, $c = .7$, $g = .4$, *old–new criterion* = 2.2, $\gamma = 0.8$.

made. The assumption is that one cannot recall what has not been encoded. If, however, encoding (i.e., u^*) occurs at too rapid a rate with respect to repetitions, the contribution of recollection to performance (i.e., q) also increases too rapidly. Perhaps the simplest way to amend the model is to assume that the amount of encoding decreases with each subsequent target presentation. That is, u^* is negatively related to repetitions. The challenge, which we defer, is to incorporate this assumption both during encoding and during retrieval.

GENERAL DISCUSSION

We sought to observe the flexibility of associative memory, using an associative recognition task. The results of Experiment 1 replicate those in Kelley and Wixted (2001), insofar as the number of target presentations had little or no effect on FARs when the ratings method was used. In contrast, the *yes–no* procedure used in Experiment 2 produced steadily increasing FARs, suggesting that the evidence used to perform the ratings and the *yes–no* tasks is not always the same. We assumed that recollective evidence becomes available later during retrieval than does familiarity, and the subjects performing the *yes–no* task responded, at least sometimes, before the contribution of recollective evidence was available. We tested this hypothesis in Experiment 3 by imposing a 2-sec delay on the *yes–no* responses. The finding of a concave down nonlinear function relating repetitions to FARs supported this hypothesis, suggesting that the evidence used to perform a given task can be influenced by situational factors. In

this case, the 2-sec delay imposed on the subjects allowed them to utilize the slower but more accurate recollective evidence without a time cost.

Given this apparent flexibility in associative memory, we took a page from the verbal-learning literature and further speculated that when the familiarity process usually provides accurate information, subjects will rely less on the slower, more effortful recollective information. Thus, we introduced a significant number of randomly similar *XY* pairs to the test list in Experiment 4, on the assumption that when targets and foils are randomly similar, familiarity is a relatively efficient means for discriminating targets from foils. We observed a substantial increase in FARs for the rearranged pairs, and the function relating repetitions to FARs was more nonlinear, relative to that observed in Experiment 1. This provides additional support for the hypothesis that subjects adapt their recognition strategy to the task situation and suggests that familiarity was more likely to be the basis of the recognition task when *XY* pairs were tested than when they were not tested.

On the Strategic Use of Recognition Memory

We hypothesized that subjects attempt to use an efficient recognition strategy, and an interesting question concerns *how* subjects adapt their strategy to the task and situation at hand. One possibility is that they attempt to achieve a specific level of accuracy from situation to situation. If so, we would expect to see similar overall HRs and FARs from situation to situation, even though the patterns of FARs might vary, and our results are consistent with this expectation.

The pattern and magnitudes of the HRs were virtually indistinguishable in all four experiments. For FARs, we compared *yes-no* performance (Experiment 2) with ratings performance (Experiment 1), and there was no main effect of task, even though the pattern of FARs varied between the tasks. The same was true when we compared *yes-no* performance in the delay condition with that in the no-delay condition.

The comparison between the ratings (Experiment 1) and the *XY* experiments is particularly interesting (Xu & Malmberg, 2007). The FARs for rearranged pairs were substantially higher when *XY* pairs were tested than when they were not. However, the rearranged false alarms represented only half of the foils tested. The remaining half came from the *XY* pairs, which had substantially lower FARs. When a weighted average was obtained, the overall FAR when *XY* pairs were tested was .18, versus .23 when they were not tested (Experiment 1). The difference in overall FARs is reliable [$t(102) = 2.04, p = .044$]. However, since this difference is rather small and given the results from the ratings versus the *yes-no* and the *yes-no* delay comparisons, it is perhaps safest to conclude that subjects attempt to achieve similar levels of recognition performance across different tasks or situations.

The question of how subjects adjust performance need not imply that subjects are able to monitor accuracy on a trial-by-trial basis. Presumably, everyone knows that accuracy decreases as the amount of time spent on a task decreases. For example, students allocate a subset of a finite amount of time to each question on an exam. Thus, subjects may estimate, on the basis of *metamemorial knowledge*, how long it will take to achieve a given level of performance and will allocate their time accordingly (Nelson & Narens, 1990).

The *a* parameter is assumed to reflect the contribution of recollection to performance, and we have assumed that *a* may vary as a function of interference or aging and/or it may vary due to situational or strategic factors (Xu & Malmberg, 2007). Variations in the contribution of recollection to performance due to interference or aging are clearly out of control of the subject. However, in the present experiments, we have assumed that variations in the contribution of recollection to performance are strategic, but we have not modeled *how* the subject adopts different recognition strategies. The present simulation offers one possible explanation.

Note that the *yes-no* criterion used in Xu and Malmberg's (2007) simulation was lower (1.0) than the *yes-no* criterion used in the present simulation (2.2). Xu and Malmberg found that the present model had difficulty predicting the FAR in the one-presentation condition when the contribution of recollection to performance was relatively low. Here, the higher *yes-no* criterion allowed the model to predict the relatively low FARs for the *XY* pairs and the rearranged pairs in the one-presentation condition, particularly for Experiments 2 and 4. This is potentially important because the *yes-no* criterion affects the contribution of recollection to performance. That is, responses

are based on familiarity whenever an item's familiarity does not exceed the *yes-no* criterion (see Figure 1). Hence, setting a relatively high criterion produces less of a contribution of recollection to performance. This might be what happened in Experiments 2 and 4.

Thus, it is possible that subjects adopt different recognition strategies by shifting their *yes-no* criterion. In the present model, an extremely high *yes-no* criterion will produce a preponderance of rejections, but one can imagine a different decision model (e.g., Atkinson & Juola, 1974) in which there is a strict criterion and a lenient criterion. When familiarity exceeds the stricter *yes-no* criterion, a *yes* response is made, and when familiarity does not exceed the lenient criterion, a *no* response is made. Recollection is invoked when the familiarity values falls between the two *yes-no* criteria. In a model such as this, the contribution of recollection to performance is a positive function of the difference between the two criteria. Assume that these criteria converge to the point where the underlying familiarity distributions cross, which is optimal, and, hence, the model reverts to a single-process familiarity-based model. In this case, we will expect faster responses than if the dual-process strategy is invoked and will expect approximately equal *yes* and *no* responses.

This speculation concerning how different recognition strategies are implemented has implications for how one might interpret Xu and Malmberg's (2007) findings. In that experiment, the nature of the stimuli making up the pairs was varied, and this factor was crossed with study time and the number of target presentations. The results of the model simulations suggested that the contribution of recollection to performance was greater when words and faces were paired than when pseudowords and Chinese characters were paired. As in the present simulation, a single *yes-no* criterion was used in all the conditions in Xu and Malmberg's simulation. However, a statistical analysis showed that the criterion location (C_a) was stricter for pseudowords and Chinese characters. Interestingly, the model fit in Xu and Malmberg's simulation was best for words and faces and worst for pseudowords and Chinese characters. Indeed, when different criteria were assumed for words and faces and for pseudowords and Chinese characters, Xu and Malmberg found that a better fit of the model to all the data could be achieved, which is consistent with the hypothesis that the *yes-no* criterion is involved in implementing different recognition strategies.

CONCLUSIONS

The present set of experiments demonstrates the flexible nature of human recognition memory. Our conclusion is not that the availability of recollective information versus familiarity information varies from situation to situation. Rather, we conclude that the basis for the associative recognition decision varies from situation to situation and that subjects adopt an associative recognition strategy that will attain a given level of accuracy in the most efficient way possible.

AUTHOR NOTE

Portions of this research were presented at the 30th Annual Interdisciplinary Conference in Teton Village, Wyoming, and the 3rd Annual Context and Episodic Memory Symposium in Philadelphia. The authors thank Doug Hintzman, Bill Hockley, and Bill Petrusic for helpful comments on the manuscript. Correspondence concerning this article should be addressed to K. J. Malmberg, Department of Psychology, University of South Florida, 4202 East Fowler Ave., PCD 4118G, Tampa, FL 33620-7200 (e-mail: malmberg@cas.usf.edu).

REFERENCES

- ANDERSON, R. C., & WATTS, G. H. (1971). Response competition in the forgetting of paired associates. *Journal of Verbal Learning & Verbal Behavior*, **10**, 29-34.
- 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 developments in mathematical psychology: Vol. 1. Learning, memory, and thinking* (pp. 243-293). San Francisco: Freeman.
- ATKINSON, R. C., & SHIFFRIN, R. M. (1968). Human memory: A proposed systems and its control processes. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2, pp. 89-195). New York: Academic Press.
- BARANSKI, J. V., & PETRUSIC, W. M. (1998). Probing the locus of confidence judgments: Experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 929-945.
- CLARK, S. E., & GRÖNLUND, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, **3**, 37-60.
- CLEARY, A. M., CURRAN, T., & GREENE, R. L. (2001). Memory for detail in item versus associative recognition. *Memory & Cognition*, **29**, 413-423.
- CRISS, A. H., & SHIFFRIN, R. M. (2004). Pairs do not suffer interference from other types of pairs or single items in associative recognition. *Memory & Cognition*, **32**, 1284-1297.
- DILLER, D. E., NOBEL, P. A., & SHIFFRIN, R. M. (2001). An ARC-REM model for accuracy and response time in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 414-435.
- DONALDSON, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition*, **24**, 523-533.
- DOSHER, B. A. (1984). Discriminating preexperimental (semantic) from learned (episodic) associations: A speed-accuracy study. *Cognitive Psychology*, **16**, 519-555.
- DUNN, J. C. (2004). Remember-know: A matter of confidence. *Psychological Review*, **111**, 524-542.
- GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- GRÖNLUND, 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.
- HOCKLEY, W. E. (1992). Item versus associative information: Further comparisons of forgetting rates. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **18**, 1321-1330.
- JOORDENS, S., & HOCKLEY, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 1534-1555.
- 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.
- MALMBERG, K. J., HOLDEN, J. E., & SHIFFRIN, R. M. (2004). Modeling the effects of repetitions, similarity, and normative word frequency on old-new recognition and judgments of frequency. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 319-331.
- MALMBERG, K. J., & SHIFFRIN, R. M. (2005). The "one-shot" hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **31**, 322-336.
- MALMBERG, K. J., ZEELLENBERG, R., & SHIFFRIN, R. M. (2004). Turning up the noise or turning down the volume? On the nature of the impairment of episodic recognition memory by midazolam. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 540-549.
- MURDOCK, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, **89**, 609-626.
- MURDOCK, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review*, **104**, 839-862.
- NELSON, T. O., & NARENS, L. (1990). Metamemory: A theoretical framework and new findings. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 26, pp. 125-169). San Diego: Academic Press.
- POSTMAN, L., & STARK, K. (1969). Role of response availability in transfer and interference. *Journal of Experimental Psychology*, **79**, 168-177.
- RAAIJMAKERS, J. G. W., & SHIFFRIN, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 14, pp. 207-262). New York: Academic Press.
- REDER, L. M., NHOUYVANISVONG, A., SCHUNN, C. D., AYERS, M. S., ANGSTADT, P., & HIRAKI, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 294-320.
- ROTTELO, C. M., MACMILLAN, N. A., & VAN TESSEL, G. (2000). Recall-to-reject in recognition: Evidence from ROC curves. *Journal of Memory & Language*, **43**, 67-88.
- SHIFFRIN, R. M., & STEYVERS, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, **4**, 145-166.
- SHIFFRIN, R. M., & STEYVERS, M. (1998). The effectiveness of retrieval from memory. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 73-95). Oxford: Oxford University Press.
- VAN ZANDT, T., & MALDONADO-MOLINA, M. M. (2004). Response reversals in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **30**, 1147-1166.
- WIXTED, J. T., & STRETCH, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review*, **11**, 616-641.
- XU, J., & MALMBERG, K. J. (2007). Modeling the effects of verbal and nonverbal pair strength on associative recognition. *Memory & Cognition*, **35**, 526-544.
- 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.

NOTES

- As in Xu and Malmberg (2007), the tendency to use the high-confidence *old* rating increased with increases in the number of target presentations when a false alarm was made [$F(1,48) = 10.64, p < .003$].
- Our *XY* pairs were different from Anderson and Watts's (1971) extra-list foils in two respects. Both members of their extra-list foils were studied, and they consisted of one item appearing on each list.
- The average response time (RT) in Experiment 4 was 2.3, which falls in between the latencies in Experiment 1 (3.4) and Experiment 2 (1.8). This is what would be expected, given that half of the foils in Experiment 4 could be discriminated from the targets only on the basis of familiarity, unlike in Experiment 1. In addition, one would not expect the RTs in Experiment 4 to be as short as those in Experiment 2, because the rating task is more complicated than the *yes-no* task.

APPENDIX

Xu and Malmberg REM Dual-Process Model

In REM, generic knowledge is stored in lexical/semantic memory traces and events are stored in episodic traces. When words are studied, their lexical/semantic traces are retrieved into a short-term memory buffer. New episodic traces are created by copying the values from the lexical/semantic traces residing in the short-term memory buffer to new incomplete and error-prone episodic vectors.

Specifically, lexical/semantic traces are assumed to be vectors of w features whose values, V , are geometrically distributed integers. (g is the geometric distribution parameter.) u^* is the probability of storing a feature in an episodic trace, and c is the probability of copying that feature correctly from a lexical/semantic trace. If an error in encoding is made, a feature value is drawn randomly from the geometric distribution. A zero value represents a case in which no feature is stored.

The number of features stored increases with study time, but the amount of extra storage diminishes over time: $t_j = t_{j-1}(1 + e^{-b})$, where t_j is the number of attempts at storing a feature for an item residing continuously in the short-term buffer for j seconds and b is a rate parameter (Malmberg & Shiffrin, 2005). When an item is repeated, an additional t_j attempts at storing features occur. The additional features are usually accumulated in an existing trace, rather than a new trace being created (Malmberg & Shiffrin, 2005; Shiffrin & Steyvers, 1997, 1998), in order to account for the effects of a variety of strengthening operations. Here, we simply assume that this probability is 1.0. Thus, the probability of storing a feature, m , given that the item it makes up was studied r times and each time t attempts were made to store the feature, is

$$P(m) = 1 - (1 - u^*)^{rt} \quad (\text{A1})$$

When a pair of items is studied, a concatenation of two episodic traces is stored.

We assume that the concatenation of the two lexical/semantic or temporary short-term vectors corresponding to stimulus pair serves as a retrieval cue (cf. Gronlund & Ratcliff, 1989). The compound cue is matched to all of the concatenated traces stored during study. For each trace j , a likelihood ratio, λ_j , is calculated as

$$\lambda_j = (1 - c)^{n_{jq}} \prod_{i=1}^{\infty} \left[\frac{c + (1 - c)g(1 - g)^{i-1}}{g(1 - g)^{i-1}} \right]^{n_{ijm}}, \quad (\text{A2})$$

where n_{jq} is the number of mismatching features in the j th concatenated trace and n_{ijm} is the number of features in the j th concatenated trace that match the features in the compound retrieval cue. λ_j represents the degree of match, or *activation*, of the concatenated trace j in response to a probe with the retrieval cue. The more similar the compound cue is to the concatenated trace j , the greater λ_j will be. The recognition decision is based on

$$\Phi = \frac{1}{n} \sum_{j=1}^n \lambda_j.$$

If Φ exceeds a subjective criterion, the response is positive.

We assume that recall conforms to the standard set of assumptions in SAM (Raaijmakers & Shiffrin, 1980) and REM (Malmberg & Shiffrin, 2005). After the probe of memory, a trace is sampled with replacement. The greater the similarity between trace i and the cue, the more likely it is that trace i will be sampled. The greater the similarity between trace k and the cue, the less likely it is that trace i will be sampled. Once a trace is sampled, an attempt is made to recover its contents. If recovery is successful and the contents match the stimulus, the response is *old*. If recovery is successful and the contents do not match the stimulus, the response is *new*. Otherwise, the subject guesses *yes* with probability γ .

The contribution of recollection to recognition performance is determined partly by how well traces are encoded and by control processes that implement different retrieval strategies. We further assume that the recall-to-reject strategy requires a degree of cognitive effort or control beyond that which familiarity-based retrieval imposes. Hence, strategic, situational, and subjective factors can affect the tendency to rely on recollection as a source of evidence on the basis which to make the recognition decision.

The relatively complex control and retrieval assumptions are expressed by a simple function (Malmberg, Holden, & Shiffrin, 2004). The probability that recollection is the basis for the *old–new* decision is

$$\hat{q} = ac \left[1 - (1 - u^*)^{rt} \right]. \quad (\text{A3})$$

The value of a varies between 0 and 1. It scales the contribution of recollection to performance independently of how well items are encoded.