Journal of Experimental Psychology:
Learning, Memory, and Cognition
1994, Vol. 20, No. 6, 1341–1354

0278-7393/94/$3.00

Receiver-Operating Characteristics in Recognition Memory: Evidence for a Dual-Process Model

Andrew P. Yonelinas

Evidence is presented that recognition judgments are based on an assessment of familiarity, as is described by signal detection theory, but that a separate recollection process also contributes to performance. In 3 receiver-operating characteristics (ROC) experiments, the process dissociation procedure was used to examine the contribution of these processes to recognition memory. In Experiments 1 and 2, reducing the length of the study list increased the intercept (d') but decreased the slope of the ROC and increased the probability of recollection but left familiarity relatively unaffected. In Experiment 3, increasing study time increased the intercept but left the slope of the ROC unaffected and increased both recollection and familiarity. In all 3 experiments, judgments based on familiarity produced a symmetrical ROC (slope = 1), but recollection introduced a skew such that the slope of the ROC decreased.

Since the early 1960s, signal detection theory (Green & Swets, 1966; Norman & Wickelgren, 1969) has played a crucial role in memory theory. One of its most important contributions is the idea that recognition judgments can be based on an assessment of strength or familiarity—a notion that still plays a dominant role in current theorizing (i.e., in connectionist models as well as global memory models). Presumably, studying an item temporarily increases the item's familiarity, such that old items will on average be more familiar than new items. Thus, an assessment of familiarity provides a good basis for recognition memory judgments. However, subjects may not be limited to assessments of familiarity. If some aspect of the study event can be recollected (e.g., "I remember seeing that word. . . . It was the first one in the list"), this could also serve as a basis for recognition judgments. The aim of this article is to show that the use of familiarity is well described by signal detection theory, and that a weakness of earlier applications of that theory was the failure to separate the effects of recollection. A review of a number of recent experiments examining receiver-operating characteristics (ROCs) shows that a simple signal detection theory cannot account for the data without the introduction of an additional factor or process. A dual-process model is proposed, which accounts for the previous data. Finally, three ROC experiments are conducted to assess the dual-process model further by testing a number of its predictions.

Signal Detection Theory

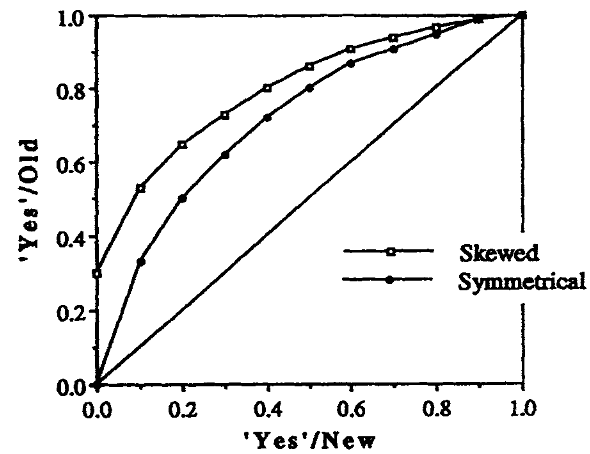
Probably the strongest support for the use of signal detection theory in recognition memory came from the analysis of ROCs. An ROC is the function that relates the proportion of correct recognitions (hit rate) to the proportion of incorrect recognitions (false-alarm rate). Typically, performance is examined across levels of confidence. For example, after studying a list of words, subjects are presented with a mixture of old and new items and are required to make recognition judgments on a confidence scale ranging from *sure it was old* to *sure it was new*. The number of different response categories on the scale typically range from *6* to *10*. Points on the ROC are plotted as a function of confidence, such that the first point includes only the most confidently remembered items (i.e., items eliciting a response of 1). The second point includes all of the most confident responses as well as the next most confident responses (i.e., items eliciting a response of 1 or 2). In this way a 6-point response scale provides 5 points on the ROC.

Figure 1 presents hypothetical ROCs derived from a 10-point confidence scale plotted on probability coordinates as well as z coordinates. Plotting the data on z coordinates provides two important measures of performance. The intercept of the transformed ROC provides a convenient measure of discriminability (d'), and the slope of the transformed ROC provides a measure of the symmetry of the ROC. An ROC that is symmetrical along the diagonal produces a slope on the transformed ROC of 1.0. However, asymmetrical or skewed ROCs are also possible (see Figure 1). Asymmetrical ROCs are still curvilinear but are pushed up and are no longer symmetrical along the diagonal. This skew is reflected as a slope of less than unity when the z scores are plotted. So, as the ROC becomes more asymmetrical, the slope of the transformed curve will fall away from 1.0.

Symmetrical ROCs are perfectly described by a simple signal detection theory. By such a theory, recognition judgments are based on the assessment of item familiarity. All items have some level of preexperimental familiarity and there is some variability from one item to the next, such that the familiarity of new items is described by a normal distribution, as in Figure 2. Studying a list of items temporarily increases the familiarity of those items, which has the effect of shifting the distribution to the right. The subject selects some level of familiarity so that only the items exceeding this level are

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. I thank Larry Jacoby, Lee Brooks, and Ian Begg for their critical comments and helpful suggestions, as well as Scott Gronlund, Janet Metcalfe, and an anonymous reviewer.

Correspondence concerning this article should be sent to Andrew P. Yonelinas, Department of Psychology, McMaster University, Hamilton, Ontario, Canada L8S 4K1. Electronic mail may be sent to yonelinas@mcmail.cis.mcmaster.ca.



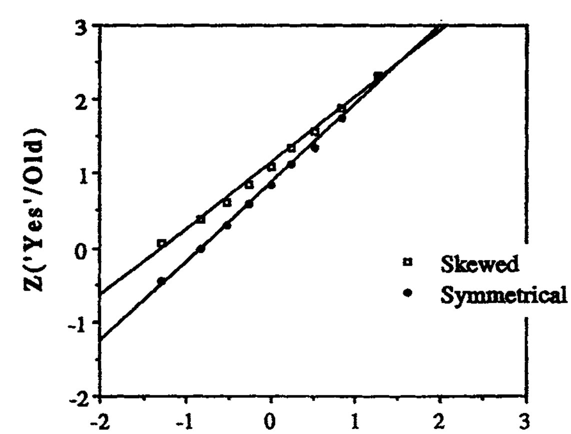


Figure 1. Symmetrical and skewed receiver-operating characteristics plotted on probability coordinates as well as *z*-coordinates.

judged as old. In a confidence judgment task, subjects select a number of different criteria along the familiarity scale (see Figure 2). The most familiar items lead to the most confident yes response, the second most familiar items lead to the second most confident yes response, and so on. In this way, the model produces an ROC that has a slope of 1.0. Some manipulations will produce greater increases in familiarity than others, and thus the intercept will increase, but the shape of the old and new item distributions will stay the same and the slope of the transformed ROC will remain constant at 1.0. This model is referred to as a Gaussian equal-variance signal detection model. Slopes other than 1.0 can only be generated by complicating the model and introducing another parameter or factor (e.g., the ratio of the new item variance to old item variance).

Although symmetrical ROCs are sometimes found in psychophysical detection experiments (see Green & Swets, 1966), they are rarely observed in recognition memory studies. Ratcliff, Sheu, and Gronlund (1992) examined ROC slopes as a function of item strength and found that although increasing the study duration or the number of study presentations dramatically increased the intercept of the ROC, the slope remained constant at approximately 0.8. Similarly, Egan (1958) reported that performance was greater for items presented twice than those presented once but that the slopes for both curves were close to 0.7. Ratcliff et al. (1992) examined data from two other studies (Mandler & Boeck, 1974; Murdock & Dufty, 1972) and found the slopes to be close to 0.8.

However, there are a number of studies in which an increase in recognition performance was accompanied by a decrease in ROC slope. For example, Donaldson and Murdock (1968) used a continuous recognition paradigm and found that items that were tested immediately after being studied produced the greatest performance and the shallowest slope. As study-test lag was varied from 0 to 9 items, d' decreased from 3.6 to 1.2, and the ROC slope increased from 0.3 to slightly less than 1. Furthermore, they found the same pattern of results across the test sessions, where d' dropped from 1.6 to 1.4 but the ROC slope increased from less than 0.7 to about 0.9. In another study, Gehring, Toglia, and Kimble (1976) examined ROC curves for pictures and words at delays of 15 min, 1 month, and 2 months. They found that picture recognition was superior to word recognition and that performance for both was decreased across delay. An examination of their ROC curves (see Gehring et al., 1976, Figure 3) shows that the ROC slope for words was less than that for pictures and that ROC slope decreased across delay. Finally, Glanzer and Adams (1990) examined ROC curves as a function of word frequency, word concreteness, and the type of decoding required to read the word (they compared performance on words that were presented normally with words that were presented in reverse order at study and test, e.g., "emoh"). They found that across all the manipulations, as performance increased the ROC slope decreased.

In summary, the slope of the transformed ROCs in recognition studies is often much less than 1. Moreover, the slope changes across some manipulations but not across others. When strength was manipulated by increasing study time, d' increased but the slope was unaffected. However, interference manipulations such as lag and delay, as well as material

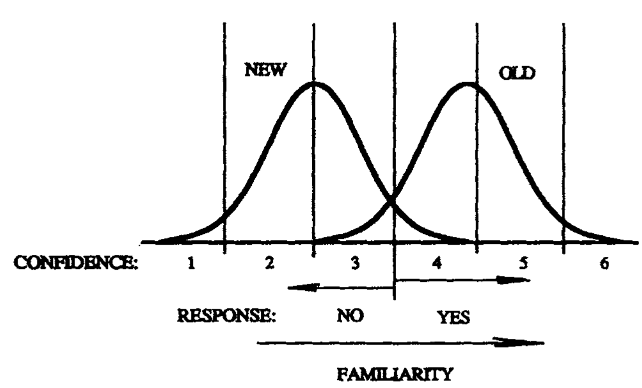


Figure 2. Familiarity distributions representing old and new items in the classical signal detection model.

manipulations such as word frequency, influenced both d' and slope: When d' increased, slope decreased.

The Dual-Process Model

The simple signal detection theory discussed previously predicts an ROC slope of 1 regardless of d'. However, the problem of accounting for slopes of less than 1.0 can be overcome by introducing another parameter or factor. For example, suppose recognition judgments were not based solely on assessments of familiarity but reflected the product of a separate recollection process. Consider the case in which a subject successfully retrieves something about a study event, like what the item was paired with or what list it was presented in. In this case, one would expect that the recognition judgment would be a highly confident one. Recollection would tend to increase the number of old items eliciting a high-confidence response without influencing the false-alarm rate. This would push the ROC up and produce the skew that is often seen.

In terms of signal detection theory, this may be described, at least roughly, as an increase in the variance of the old item distribution. However, it would be more correct to represent it as a skewing of the old item distribution such that it is pulled to the right. In principle, one can distinguish between these two alternatives by examining the shape of the *z*-transformed ROCs. If both distributions are normally distributed, then the ROC should be a straight line; if the old item distribution is pulled to the right, then the *z*-transformed ROC should exhibit a slight U shape. In practice, though, distinguishing between these two alternatives is quite difficult (see Experiment 1).

The model I propose is loosely based on a number of previous dual-process models of recognition (Atkinson & Juola, 1974; Mandler, 1980). However, it is most similar to a dual-process model proposed for cued recall by Jacoby, Toth, and Yonelinas (1993). In recognition memory, the basic idea is that judgments can be based upon an assessment of item familiarity or on the product of a conscious recollection process. Recollection is assumed to be an all-or-none retrieval process, such that for any item the subject either succeeds or fails at retrieving something about that specific study event. A successful retrieval is expected to lead to a highly confident response. Familiarity, on the other hand, is assumed to be well described by the standard equal-variance signal detection theory described earlier. The two processes are assumed to contribute independently to overall recognition performance. That is, that the probability of recognizing an old item is equal to the probability that it is recollected (R), plus the probability that its familiarity exceeds some criterion (F > cr), minus the intersect of the two:

$$P(\text{'yes'/old}) = P(R) + P(F > cr) - [P(R) \times P(F > cr)].$$

It is important to note that in this model the response criterion is applied only to the use of familiarity. Recollection is expected to be independent of false-alarm rate. So as the criterion changes, the number of items accepted on the basis of familiarity will change, but the probability of recollection should remain fixed.

If performance were based solely on the assessment of familiarity, the model would predict a symmetrical ROC curve (slope = 1), like that of the standard signal detection theory. Adding recollection will increase the high-confidence hit rate, resulting in a skewed ROC (slope < 1). To illustrate this, hypothetical ROCs were derived based on different levels of familiarity and recollection. Because familiarity is thought to reflect a signal detection process, the product of this process can be measured in terms of d'. Because recollection is assumed to be an all-or-none retrieval process, the product of this process can be measured as a simple probability. ROC curves were calculated by first selecting a level of familiarity (d'). For a range of false-alarm rates, corresponding hit rates were found using standard d' tables. In the following example, nine false-alarm rates ranging from 0.1 to 0.9 at increments of 0.1 were used. This reflects the expected probability of accepting an old item on the basis of familiarity for each false-alarm rate. To calculate the overall hit rate, a set proportion of recollected items were added (by the independence formula presented above) to each familiarity value. Figure 3 (top) presents two ROC curves derived in this way. Both curves reflect an equal contribution of familiarity (the discrimination afforded by familiarity was set at $d' = 0.42$). However, the lower curve included a contribution of recollection of 0.33 and the upper curve reflected a contribution of 0.67.

As can be seen in Figure 3 (top), as the probability of recollection increased, the curve moved up and became more skewed. The intercept of the *z*-transformed curve increased from 0.80 to 1.28, and the slope dropped from 0.79 to 0.58. However, a different pattern of results was produced when both recollection and familiarity increased together. The lower curve in Figure 3 (bottom) represents performance when familiarity was $d' = 0.42$ and the probability of recollection was 0.33. The upper curve represents performance when familiarity increased to $d' = 0.94$ and the probability of recollection increased to 0.50. In this case, the intercept increased from 0.80 to 1.39 while the slope remained constant at approximately 0.79. Thus, if both recollection and familiarity increase together, the result will be an increase in the intercept but the slope of the ROC may remain constant.

In fact, an examination of the model's predictions across a wide range of recollection and familiarity values suggested that it would be relatively easy to obtain curves with similar ROC slopes. It was only when the magnitude of the change in recollection was much greater than that on familiarity that large slope differences arose. For example, to obtain a difference in slope of .10 (anything less than this would be difficult to detect), the increase in recollection had to be three to four times greater than the increase in familiarity (familiarity was measured as a probability at a false-alarm rate of .20).

Figure 3 shows that the introduction of a recollection process could account for the pattern of results described earlier. If recollection contributes to recognition performance, then one would expect the slope of the ROC to be less than 1.0, as is often the case. Furthermore, if interference and material manipulations were to increase the probability of recollection and leave familiarity relatively unchanged, then this would lead to the observed increase in the intercept and a decrease in the slope. If the strength manipulations were to lead to similar increases in both the probability of recollection and familiarity,

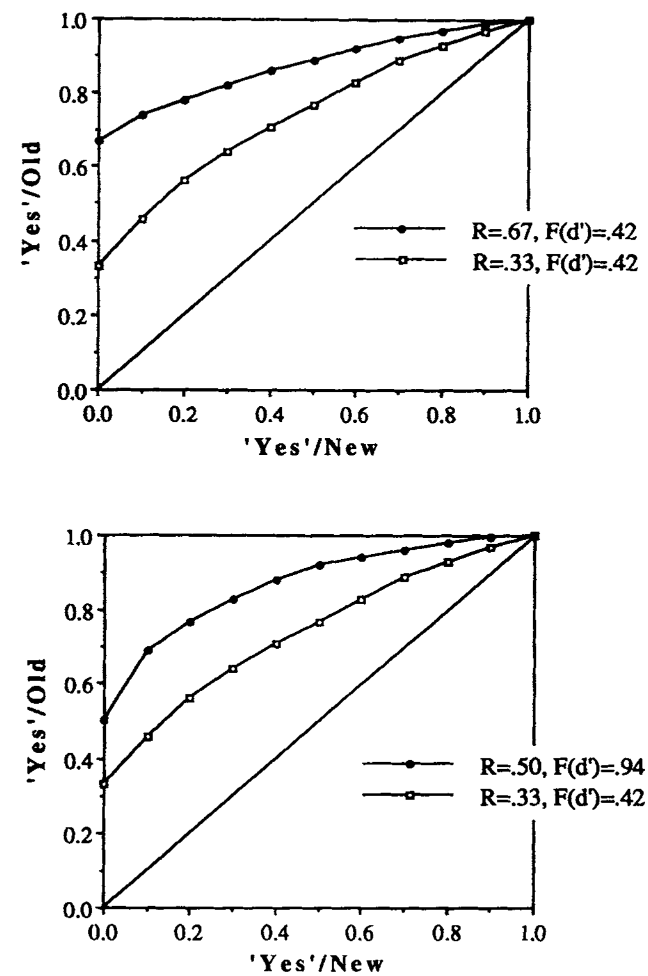


Figure 3. Hypothetical receiver-operating characteristics generated by the dual-process signal detection model. The top portion represents performance in two conditions that differ only in the probability of recollection. The bottom portion represents performance in two conditions that differ both in the probability of recollection and familiarity.

then the slope of the ROC should remain relatively constant while the intercept increases.

Furthermore, there is some evidence that the interference and material manipulations do have the expected effects on the two memory processes. Using a procedure that I describe in the next section, recent studies have found that both interference manipulations and material manipulations affect recollection to a greater extent than familiarity. For example, recollection was found to be more susceptible to list length interference (Yonelinas & Jacoby, in press) than was familiarity. Similarly, changes in word frequency produced greater changes in recollection than familiarity (Jacoby, 1994).

To assess the model more fully, I conducted three ROC experiments. In all three experiments, it was necessary to examine how recollection and familiarity contributed to overall recognition performance. In all experiments, the process dissociation procedure was used to estimate the contribution of these two processes. The procedure provides estimates by comparing performance on one condition in which the two processes act in concert to produce the same response with another condition in which they act in opposition to produce different responses. The procedure is described as it was used in this study.

The Process Dissociation Procedure

In all three experiments, subjects performed a yes–no list-discrimination task. They began by studying two lists of words, one list immediately after the other. Following this, they were presented with a mixture of List 1, List 2, and new items and were asked either "Was this word in List 1?" or "Was this word in List 2?" They were instructed to respond yes if the item was in the appropriate list and to respond no if it was a new word. Furthermore, they were informed that no word would be in both lists, so if they recollected that a word was in the inappropriate list they should respond no.

To begin, consider the fate of the List 1 items. When asked if the item was in List 1, recollection and familiarity both lead to a correct yes response. That is, subjects could respond yes because the word was sufficiently familiar or because they recollected that the word was in List 1. If the two processes are independent, then the probability of correctly responding yes is equal to

$$P(R) + P(F > cr) - [P(R) \times P(F > cr)],$$

which is the probability that an item is recollected P(R) plus the probability that its familiarity exceed some criterion $P(F > cr)$ minus the intersect of the two. This is referred to as the *inclusion condition* because subjects included items for which they recollected list membership.

Under the other set of instructions, subjects were asked to respond yes if the item was in List 2. In this case, familiarity would lead the subject to respond yes, but recollection would lead them to respond no. That is, a subject who recollects that the item was in List 1 would exclude the item and respond no. Again, given that the two processes are independent, the probability of incorrectly responding yes to a List 1 item is equal to

$$P(F > cr) - [P(R) \times P(F > cr)],$$

which is the probability that the item is familiar $(F > cr)$ minus the probability that the item is both familiar and recollected. This is referred to as the *exclusion condition* because subjects excluded the items that they recollected.

The probability of recollection can be estimated as the difference between the probability of accepting an old item in the inclusion condition from the probability of accepting an old item in the exclusion condition:

$$P(\text{Recollection}) = P(\text{Inclusion}) - P(\text{Exclusion}).$$

After calculating the contribution of recollection, one can solve either of the previous equations to estimate the contribution of familiarity. For example,

$$P(\text{Familiar}) = P(\text{Exclude})/[1 - P(R)].$$

The probability of recollection provides a measure of consciously controlled processing defined in terms of selective responding. To the extent that subjects were able to recollect the list in which a word was earlier presented, they should be able either to include or to exclude that word, in line with the instructions. For example, if recollection were perfect ($R = 1$), subjects would always respond yes to List 1 words when instructed to select words from List 1 (an inclusion test) and never call those words old when instructed to select words from List 2 (an exclusion test). In contrast, familiarity does not support such selective responding. The contribution of familiarity as a basis for responding old is the same on an exclusion test as on an inclusion test. That is, familiarity has the same effect regardless of whether that effect results in correct responses (an inclusion test) or errors (an exclusion test).

The previous example illustrated how recollection and familiarity values were calculated for List 1 words. However, the same was done for List 2 words. In fact, the necessary inclusion and exclusion conditions were nested within the previously described test conditions. The instructions "List 1?" served as exclusion instructions for the List 2 words and inclusion instructions for the List 1 words. Similarly, the instructions "List 2?" served as inclusion instructions for the List 2 words and exclusion instructions for the List 1 words. Because list number was not of immediate interest, performance was collapsed across the two lists, yielding an inclusion and an exclusion score, which were used to calculate the contribution of recollection and familiarity.

Subjects responded on a 6-point confidence scale ranging from *sure yes* (6) to *sure no* (1). In this way, it was possible to examine performance as a function of response confidence and to plot ROCs. In Experiments 1 and 2, the effect of list length was examined. Increasing the length of the study list has been found to interfere with recognition performance (Ratcliff & Murdock, 1976; Strong, 1912). Furthermore, Yonelinas and Jacoby (in press) found that the effect of list length was restricted to the recollection process, leaving the use of familiarity unchanged. Because the effect was restricted to recollection, it was expected that the manipulation would influence the intercept as well as the slope of the ROC. In Experiment 3, item strength was varied in a manner similar to that of Ratcliff et al., (1992), and it was expected that increases in strength would result in an increase in the intercept but that the slope of the ROC would remain constant. Experiment 3 tested the prediction that the constant slope could be attained only if the strength manipulation influenced both the recollection and familiarity components.

Beyond these general predictions, the process dissociation procedure allowed two more tests of the model. First, on the basis of estimates for recollection and familiarity gained by examining inclusion and exclusion scores collapsed across levels of confidence, it was possible to make predictions about the shape of the ROC curve. I made predictions for the slopes and intercepts of the ROCs and these were compared with the observed data. Second, the procedure allowed an examination of the two processes as a function of confidence. Thus, the overall ROC was decomposed into the recollection and familiarity components. If familiarity is well described by an equal-variance signal detection theory, then the use of familiarity should increase in a curvilinear fashion as predicted by that theory (i.e., slope = 1). Furthermore, if recollection is acting as an all-or-none retrieval process, then the estimates for recollection should not change as a function of false-alarm rate but should remain constant.

Experiment 1

Method

Subjects and materials. Six subjects from the psychology department at McMaster University participated in the experiment. All subjects (except one, the author of this article), were graduate students who were paid for their participation. Eight-hundred items were randomly selected from the Toronto word pool at the beginning of each session.

Design and procedure. Materials were presented and responses collected on PC-compatible computers. The character size of the stimuli was approximately 5 × 5 mm and the viewing distance was approximately 0.5 m. Stimuli were presented in lowercase letters in the center of the screen.

Each subject completed six sessions, each taking approximately 40 min. Each session consisted of eight study-test blocks. Each block consisted of two study lists followed by two test lists. Half of the blocks contained two short lists (10 words in each of the two lists) and the other half contained two long lists (30 words each). The first list of study items was presented one at a time on the computer screen at a 2-s rate. After a 5-s delay, a second list of words was presented at the same rate. The same word was never in both lists and the two lists were always of equal length.

Immediately following the study phase, subjects completed two recognition tests. For half of the test blocks, the first test required subjects to respond yes if the item was studied in List 1, and the second test required subjects to respond yes if the item was studied in List 2. For the other half of the tests, the instructions were reversed. Each of the two test lists contained 10 items from List 1, 10 items from List 2, and 10 new items. These items were presented one at a time in a random order. Subjects were instructed that if they remembered the word from the appropriate list, they were to respond yes. If the word was new, then they were to respond no. Furthermore, if the word was presented in the inappropriate list, they should respond no. For each test item, subjects responded on a 6-point confidence scale ranging from *sure yes* (6) to *sure no* (1). The experiment was based on a 2 × 2 design, List Length (short vs. long) × Instructions (inclusion vs. exclusion). List number (1 vs. 2) was counterbalanced across conditions as was test order, and all factors were varied within subjects. The significance level for all statistical tests was $p < .05$.

Results and Discussion

An analysis was performed to determine whether the list length manipulation had the desired effects on recollection and familiarity. Table 1 presents the proportion of old and new items receiving a yes response (a response of 4, 5, and 6 was counted as a yes) across experimental conditions as well as the estimates of recollection and familiarity for long and short lists. A statistical analysis of the raw inclusion and exclusion condition scores is not presented, but the pattern of results is worth noting. As list length increased, the probability of correctly accepting an item under inclusion instructions decreased (accepting a List 1 item under List 1? instruction or accepting a List 2 item under List 2? instruction). The effect was small but is typical of those found for list length experi-

Table 1
Proportion of Items Accepted and Parameter Estimates for Short and Long Lists in Experiment 1

Condition/ parameter	Short list		Long list	
	M	*SD*	*M*	*SD*
Condition				
Inclusion	.78	.10	.70	.11
Exclusion	.22	.07	.30	.08
New	.09	.07	.14	.05
Parameter				
Recollection	.56	.16	.40	.13
Familiarity	.50	.10	.50	.10

ments in recognition memory (see Ratcliff & Murdock, 1976). The opposite effect was seen for the exclusion condition (accepting a List 1 item under List 2? instruction or accepting a List 2 item under List 1? instruction). As list length increased, the probability of incorrectly accepting an item from the wrong list increased. Similarly, as list length increased, the probability of incorrectly accepting a new item increased.

Of main interest are the estimates for recollection and familiarity. Recollection was calculated by subtracting the probability of accepting an old item under exclusion instructions from the probability of accepting an old item under inclusion instructions. Familiarity was calculated by dividing the exclusion score by one minus the estimated probability of recollection. I performed an analysis of variance (ANOVA) on recollection as well as familiarity on the basis of estimates derived from each subject. As can be seen in Table 1, increasing list length reduced recollection from .56 to .40, $F(1, 5) = 39.00$, $MS_e = .001$. However, familiarity was not significantly influenced by list length (.50 for both lists, $F < 1$).

Note that the probability of a false alarm was higher in the long lists (.14) than for the short lists (.09). Thus, relative to the base rate, the familiarity can be seen as decreasing from .41 to .36 from short to long lists, which is a difference of .05. However, even with this correction, the change in familiarity was small in comparison with that found for recollection (.16).

The list length manipulation interfered primarily with recollection. Moreover, for both list lengths there was a sizable contribution of recollection. If recollection is responsible for the skew in the ROC, then the ROCs for both short and long lists should exhibit a skew such that the slope is less than 1.0. Furthermore, because there is a higher probability of recollecting an item in the short lists than in the long lists, the skew should be greater in the short lists (the slope should be shallower).

Figure 4 shows the group ROC curves for inclusion and exclusion conditions for long and short lists in Experiment 1. For the inclusion condition, the curve for the long lists fell slightly below that of the short lists, showing that overall discrimination was greater for the short lists. For the exclusion condition, the curves for long and short lists did not differ greatly, but both fell far below the inclusion curves. The difference between the inclusion and exclusion curves is expected because recollection leads to a yes response in the inclusion condition, but to a no response in the exclusion condition.

Linear regression analysis was performed on the z-transformed ROC curves for the inclusion conditions to estimate the slopes and intercepts for each subject. The average goodness of fit for the linear regressions (R^2) was .98. The average slope and intercept of the long and short lists are presented in Table 2. Discriminability as measured by the intercept was greater in the short lists than in the long lists, $F(1, 5) = 8.399$, $MS_e = .028$. Most important, as predicted, the slope of the transformed ROC curve was greater for the long lists than for the short lists, $F(1, 5) = 8.40$, $MS_e = .011$. Thus, as list length increased, the intercept decreased and the slope increased. Moreover, the slope for both curves was considerably less than 1.0.

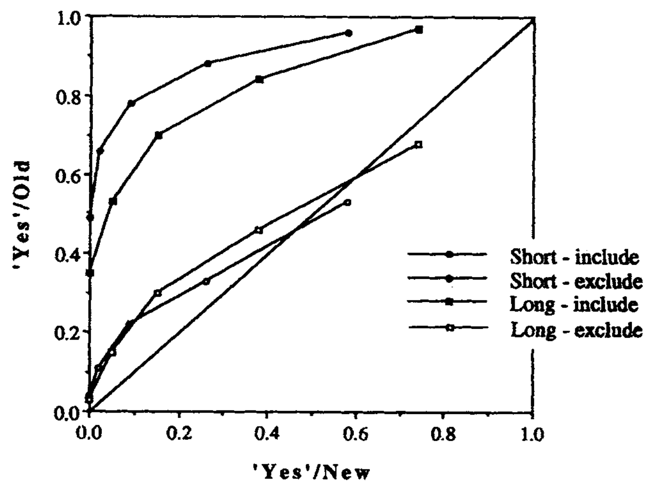


Figure 4. Receiver-operating characteristics for the inclusion and exclusion conditions for short and long study lists in Experiment 1.

Analyses were carried out to examine the effects of study position, study list, test position, and session number. The pattern of results did not show any systematic change across any of these factors. Furthermore, several alternate techniques were used to estimate the slopes and intercepts of the z-transformed ROC curves. Linear regression on the overall scores, as well as linear regression on subjects with outlying points removed, produced similar results. Moreover, the estimation algorithm of Ogilvie and Creelman (1968) was used and again produced similar results.

Predictions of the Dual-Process Model

The pattern of results is in agreement with the dual-process model; recollection contributed to performance for long and short lists, thus the slope of the ROC for both list lengths was less than 1. Furthermore, as list length became shorter, recollection increased, leading to a decrease in the slope of the

Table 2
Slopes and Intercepts for Short and Long Lists in Experiment 1

ROC measure	Short		Long	
	M	*SD*	*M*	*SD*
Slope	.57	0.10	.74	0.12
Intercept	1.60	0.30	1.33	0.34

Note. ROC = receiver-operating characteristic.

ROC. However, the dual-process model allows one to make more specific predictions about the ROCs. Using the estimates for recollection and familiarity (Table 1), hypothetical ROCs were calculated and compared with the observed ROC data. For example, in short lists, the probability of accepting an item on the basis of familiarity was .50 when base rate was .09, which corresponds to a d' of 1.34. Using standard d' tables, one can find the probability that an item will be accepted on the basis of familiarity for any false-alarm rate. Assuming that recollection remains constant across levels of false-alarm rate, the predicted probability of a hit for any level of false-alarm rate will be as follows:

$$P(R) + P(F > cr) - [P(R) \times P(F > cr)].$$

Estimates were derived for each of the five observed levels of false-alarm rates on the ROC. The observed false-alarm rates were used in the current calculations so that the slopes and intercepts could be compared with the observed values. Because the model predicts that the transformed ROC will exhibit a slight U shape, the slope of the ROC will depend on the range of false-alarm rates. However, the nonlinearity turns out to be quite small, as the linear fits to the predicted z-transformed ROCs were good (R^2 = .98).[1] It is important to note that the model predicts the shape of the ROC on the basis of two points: an inclusion score and an exclusion score for old and new items. The estimates were derived using the observed false-alarm rates only because the slope and intercept depend on the points chosen.

The predicted intercepts of the z-transformed ROCs were almost identical to the observed values. The predicted and observed intercepts were 1.70 and 1.60, respectively, for the short lists and 1.33 and 1.33, respectively, for the long lists. More important, the slopes were also as expected. The predicted and observed slopes were .64 and .57 for the short lists and .76 and .74 for the long lists. So, not only did the model successfully predict the overall pattern that was found for the ROCs but it provided close approximations to the observed slopes and intercepts.

An examination of the highest confidence response category suggested that recollection led to very confident and accurate responses. If recollection is a retrieval process, then one might expect recollected items to lead to highly confident responses. That is, if one can recollect something about a study event, then one can be relatively sure that the event did occur, in contrast to the case in which judgments are based on an assessment of familiarity. This suggests that the proportion of high-confidence responses should be close to the estimate for recollection. In fact, these proportions were close to the predicted recollection estimates in the short and long lists, but fell slightly short in both cases. For the short and long lists, the proportion of hits in the highest confidence response category were .49 and .35 where the estimated values of recollection were .56 and .40. This discrepancy could arise if a small proportion of the recollected items were placed in lower confidence categories. An alternative explanation is that all of the recollected items led to the highest confidence responses but that the estimation procedures produced a slight overestimation of recollection at the intermediate-confidence levels.

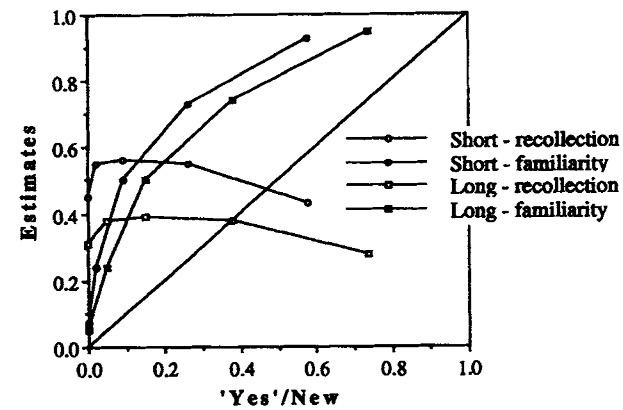


Figure 5. Estimates for recollection and familiarity for long and short lists as a function of response confidence in Experiment 1.

Both of these possibilities are considered in more detail in the General Discussion section. In any case, the discrepancy between the estimate of recollection and the proportion of highest confidence responses was relatively minor, and the results provide general support for the claim that recollection does lead to very confident recognition judgments.

A final point worth noting about the high-confidence response category is that the false-alarm rate was 0.00 for every subject. If the high-confidence category does include primarily recollected items and the false-alarm rate is essentially zero, then it would seem that false recollection of new items is rare. That is, when recollection does occur it is highly accurate.

A further test of the model involved deriving estimates for familiarity and recollection at each point on the ROC; thus the overall ROC was decomposed into its constituent parts. If familiarity does operate as a signal detection process, then the familiarity function should increase gradually with false-alarm rate and should produce a symmetrical ROC. If recollection is an all-or-none retrieval process, then it should not change as a function of false-alarm rate but should remain constant. Figure 5 presents the estimates for familiarity and recollection for long and short lists as a function of confidence. The familiarity functions for both list lengths fell close to the predicted signal detection curves. When the estimates were z-transformed, they were found to approach unity: For the short lists, estimates were .93, for the long lists, estimates were 1.00. This supports the claim that the use of familiarity reflects a simple signal detection process.

The estimates for recollection show a considerably different pattern. As with the initial estimates, recollection was greater

[1] Although the model does predict a slight U-shaped z-ROC, detection of such a nonlinearity is difficult. As pointed out, the predicted nonlinearity in the current experiments is very small. Generating noticeable U-shaped curves would require a large contribution of recollection, a small contribution of familiarity, and a false-alarm rate that varies across much of the range from 0 to 1.0. Furthermore, the end points, which critically affect any linearity analysis, are based on the high-confidence responses, and these points often contain few responses, making the detection of the nonlinearity even more difficult.

Table 3
Proportion of Items Accepted and Parameter Estimates for Short and Long Lists for Experiment 2

Condition/ parameter	Short list		Long list	
	M	*SD*	*M*	*SD*
Condition				
Recognition				
Old	.85	.05	.76	.07
New	.08	.05	.16	.08
Inclusion	.79	.06	.69	.04
Exclusion	.17	.04	.30	.09
New	.04	.02	.11	.05
Parameter				
Recollection	.62	.09	.39	.09
Familiarity	.45	.05	.49	.07

for the short lists than for the long lists. This was apparent across the range of false-alarm rates. Moreover, as was expected, the estimates for each list length remained constant across much of that range. There was, however, a tendency for the estimates to decrease at the extremes. This decrease in recollection was likely due in part to floor and ceiling effects. Because recollection was estimated by subtracting the exclusion score form the inclusion score, when either score approaches extreme values this can lead to underestimates of recollection. For example, at the most strict criterion the exclusion score approached 0.0 and at the most lax criterion the inclusion score approached 1.0 (see Figure 4). The same pattern of results was also seen in Experiment 2, in which performance again approached floor and ceiling. However, in Experiment 3, in which overall performance was reduced and scores did not approach these extremes, the estimates for recollection remained constant across the full range of false alarms.

Experiment 2

In Experiment 1, subjects were asked to make list discrimination judgments; thus subjects were required to recollect information about the study event. However, in standard recognition experiments, subjects are required only to make old–new discriminations. In those experiments, it is possible that judgments could be based on the assessment of familiarity alone. Although it seemed unlikely that subjects would not make use of recollection if they had this process available to them, Experiment 2 was designed to assess this possibility. The experiment was a replication of the first experiment, with the addition of one more test condition: a standard recognition test. As well as receiving "List 1?" and "List 2?" test instructions, subjects also were given a simple recognition task: "Was this item presented in either list?" If subjects were using the same processes in the list discrimination task as they were in the recognition task, then one would not expect the ROC for these two conditions to differ: The intercept and slope should be the same. However, a difference one might expect to see is that subjects may be more confident in their use of familiarity in the recognition test than in the list discrimination task because there is less reason to be suspect of the source of item familiarity. This would not change the shape of the ROC. However, it would tend to push the points along the curve such that each point would be shifted to the upper right.

Method

Subjects and materials. Four subjects from the previous experiment along with two new graduate students participated. Four-hundred and eighty words were randomly selected from the Toronto word pool at the beginning of each session.

Design and procedure. The design and procedure were the same as the previous experiment except for the following changes. In the short-list condition, two lists of 6 words each were presented. In the long-list condition, two lists of 18 words were presented. Each test was broken into three short sections. For each section there were two items from List 1, two items from List 2, and two new items mixed in a random order. For one section of the test, subjects were required to respond yes if the item was in List 1. For another section, they were required to respond yes if the item was in List 2. For the remaining section, they were required to respond yes if the item was in either study list. The order of the test instructions was randomized. Each session contained eight pairs of short lists and eight pairs of long lists. Each subject completed six sessions, each taking approximately 40 min.

Results and Discussion

Table 3 presents the proportion of old and new items receiving a yes response (Responses 4, 5, and 6 were counted) across inclusion and exclusion conditions, as well as the estimates of recollection and familiarity for long and short lists. The pattern of results in the inclusion and exclusion conditions, as well as the estimates for familiarity and recollection, was similar to that of Experiment 1. Increasing list length significantly reduced recollection from .62 to .39, $F(1, 5) = 51.58$ $MS_e = .003$, but left familiarity unaffected, $F(1, 5) = 2.26$, $MS_e = .003$. The estimates for familiarity were slightly less in the short lists that in the long lists (.45 vs. .49). However, the false-alarm rate was higher with longer lists. Relative to the base rate, familiarity decreased from .41 to .38 from short to long lists—a decrease of .03 compared with a decrease in recollection of .23.

Figure 6 presents the ROC curves for the inclusion, exclusion, and recognition conditions for long and short lists for

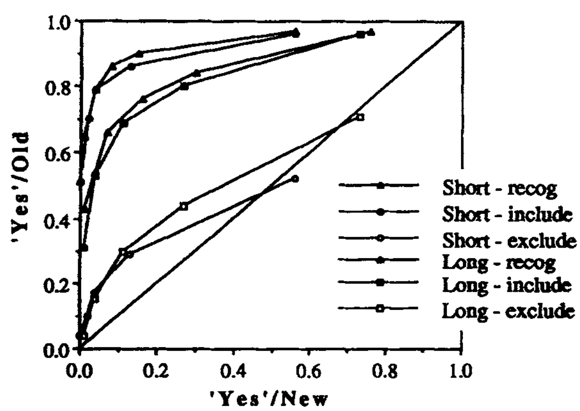


Figure 6. Receiver-operating characteristics for the inclusion, exclusion, and recognition (recog) conditions for short and long study lists in Experiment 2.

Experiment 2. The figure shows that the inclusion and recognition curves were remarkably similar. An analysis of the slopes and intercepts supported this observation. Linear regression analysis was performed on the z-transformed scores for the inclusion condition and recognition scores to estimate the slopes and intercepts. The average goodness of fit for the linear regressions (R^2) was .98. The average slope and intercept of the long and short lists are presented in Table 4. An analysis was conducted on the slopes of the z-transformed ROCs for the inclusion and the recognition conditions. A similar analysis was conducted on the intercepts. As in the previous experiment, as list length became shorter, the intercept increased and slope decreased. The same pattern was found for inclusion as well as recognition instructions. For ROC slope, there was a significant effect of list length, $F(1, 5) = 7.00$, $MS_e = .025$. There was no effect of instructions (inclusion vs. recognition), nor did the type of instructions interact with list length ($Fs < 1$). For the intercepts, there was an effect of list length, $F(1, 5) = 19.221$, $MS_e = .050$. Again, there was no effect of instructions nor was there an instruction by list-length interaction ($Fs < 1.1$). Furthermore, Figure 6 shows that the points for the recognition condition are shifted to the right along the ROC relative to the inclusion condition, suggesting that subjects were more confident about the use of familiarity in the recognition condition.

Predictions of the Dual-Process Model

The pattern of results in the current experiment was in agreement with that of the previous experiment and supports the predictions of the dual-process model: Recollection contributed to performance for long and short lists, thus the slopes of the ROCs were less than 1.0. Furthermore, as the list length became shorter, recollection increased leading to a decrease in ROC slope. As a further test of the model, estimates for recollection and familiarity were used to derive predictions for the slope and intercept of the ROC curves. Curves were calculated as described in Experiment 1. Replicating results of Experiment 1, the predicted values were close to the observed data. For the intercepts, the predicted values for the short and long lists were 1.93 and 1.50, respectively, compared with the observed values in the recognition condition of 1.88 and 1.49 and the observed values of 1.78 and 1.37 in the inclusion condition. The predicted slopes for the short and long lists were .64 and .73, respectively, compared with the observed values of .53 and .71 in the recognition conditions and the observed values of .53 and .69 in the inclusion condition.

Table 4
Slopes and Intercepts for Recognition and Inclusion Conditions for Short and Long Lists in Experiment 2

ROC measure	Short list *M*	Short list *SD*	Long list *M*	Long list *SD*
Slope				
Recognition	0.53	0.21	0.71	0.10
Inclusion	0.53	0.20	0.69	0.14
Intercept				
Recognition	1.88	0.42	1.49	0.26
Inclusion	1.78	0.44	1.37	0.37

Note. ROC = receiver-operating characteristic.

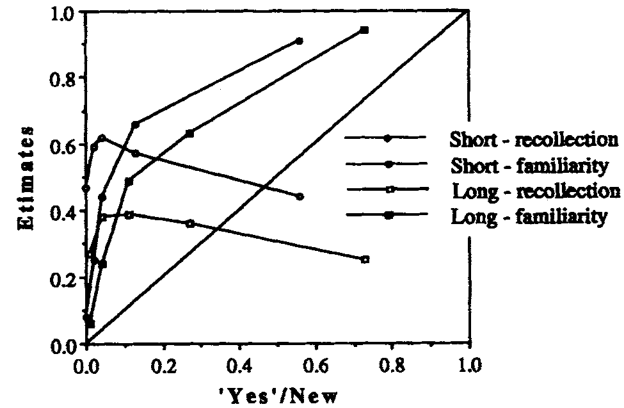


Figure 7. Estimates for recollection and familiarity for long and short lists as a function of response confidence in Experiment 2.

As in Experiment 1, the proportion of high-confidence hits came close to the estimates of recollection. For the short lists, the proportions of old items in the highest confidence category were .51 and .65 for the inclusion and recognition conditions, respectively, which were close to the estimated value of recollection (.62). For the long lists, the proportions of old items in the highest confidence category were .31 and .43 in the inclusion and recognition conditions, respectively, compared with the estimated value of recollection (.39). Thus, as in the previous experiment, recollection would seem to lead to highly confident responses. Similarly, the false-alarm rate in this category was low (0.00). This suggests that the high-confidence category included primarily recollected items, and that false recollection of new items was rare.

Figure 7 presents the estimates for familiarity and recollection for long and short lists as a function of confidence. As in the previous experiment, the use of familiarity increased gradually as false-alarm rate increased, producing an ROC of the form predicted by the equal-variance signal detection model. The slopes of the z-transformed curves approached unity: .94 for the short lists and 1.00 for the long lists. The estimates for recollection across confidence were similar to those of the previous experiment. Unlike the estimates for familiarity, they remained relatively constant across much of the range. There was a tendency for the estimates to decrease for the extreme levels of confidence. However, as in the first experiment the inclusion and exclusion condition scores approached ceiling and floor and this may have produced underestimates in recollection at these extremes.

The results of Experiments 1 and 2 provide strong support for the dual-process model. Recognition judgments were found to be based on assessments of familiarity as well as recollection, and it was the recollection process that produced the skew in the ROC curves. For both long and short lists, recollection produced ROC slopes that were less than 1.0. Moreover, as the lists became shorter, recollection increased, causing the slope of the ROC to decrease. Estimates for the slope and intercepts of the ROCs were found to closely

Table 5
Proportion of Weak and Strong Items Accepted and Parameter Estimates for Experiment 3

	Weak		Strong	
Condition/parameter	*M*	*SD*	*M*	*SD*
Condition				
Inclusion	.51	.13	.64	.13
Exclusion	.40	.09	.42	.10
Parameter				
Recollection	.10	.19	.22	.15
Familiarity	.45	.11	.54	.12

Note. The probability of accepting a new item was .24.

approximate the observed data. Furthermore, when the ROC was decomposed into its constituent parts, it was found that the use of familiarity was well described as a simple signal detection process. Recollection, however, remained relatively constant across levels of confidence, as would be expected if recollection was a retrieval process that produced highly confident responses. Finally, even when the subjects were not required to recollect, as with the standard recognition instructions, performance still reflected a combination of familiarity and recollection. However, as expected, subjects were more confident in their use of familiarity in standard recognition conditions.

Experiment 3

In the two previous experiments, it was shown that, when recollection increased while familiarity remained relatively unaffected, there was an increase in the intercept as well as a decrease in ROC slope. This could account for the pattern seen for other interference manipulations such as lag and delay (e.g., Donaldson & Murdock, 1968; Gehring et al., 1976) as well as the material manipulations (e.g., Glanzer & Adams, 1990) in which increases in ROC intercept are accompanied by decreases in slope. However, can the dual-process model account for the pattern that is seen across levels of item strength (i.e., Ratcliff et al., 1992)? That is, can the model account for increases in intercept not accompanied by increases in slope. If the dual-process model is correct, the only way that pattern could occur is if the strength manipulation influenced both familiarity and recollection. Experiment 3 was designed to test this prediction. In a design similar to that of Ratcliff et al. (1992), subjects studied pairs of items that were presented for either 1 or 3 s per pair. Two lists were presented, and each list contained a mixture of weak and strong pairs. Subjects were tested for recognition of single items using the same list discrimination instructions as in the previous experiments, and again estimates for recollection and familiarity were derived.

Method

Subjects and materials. Sixteen subjects participated in the experiment for an extra credit in an undergraduate psychology course. Four-hundred and eighty words were randomly selected from the Toronto word pool for each subject.

Design and procedure. The design and procedure were similar to those of the previous experiments with the following changes: Each subject participated in a single one-hour session. Each session consisted of 10 study-test blocks. For each block, subjects studied two lists of word pairs, each list containing 8 pairs. Half of the pairs in each list were presented for 1 s (weak) and half were presented for 3 s (strong). Within each list, the presentation rate was randomized. The test list contained all 32 studied items plus 16 new items mixed in a random order. Subjects were tested on one word at a time. The first 2 items of each study list were treated as buffer items; they were tested but they were not included in the analysis. For half of the recognition tests, subjects were instructed to respond yes if the word was from List 1. For the remaining tests they were instructed to respond yes for words from List 2. Test instructions were randomized such that subjects did not know what list would be tested. Subjects were instructed to distribute their responses across the response keys, using all of the keys. The experiment was based on a 2 × 2 design, Item Strength (weak vs. strong) × Instructions (inclusion vs. exclusion). List number (1 vs. 2) was counterbalanced across conditions and all factors were varied within subjects.

Results and Discussion

Table 5 presents the proportion of old and new items receiving a yes response (Responses 4, 5, and 6 were counted) across experimental conditions, as well as the estimates of recollection and familiarity for weak and strong items. The probability of correctly accepting an item in the inclusion condition increased with strength from .51 to .64. A similar but smaller effect was seen for the exclusion condition; the probability of incorrectly accepting an item under exclusion conditions increased from .40 to .42. By increasing the presentation duration, the recollection rate increased from .10 to .22, $F(1, 15) = 13.76$ $MS_e = .008$. Familiarity also increased with item strength from .45 to .54, $F(1, 15) = 21.97$, $MS_e = .003$. Thus as predicted, the strength manipulation had a sizable effect on both recollection and familiarity. Increasing the strength increased the probability of recollection by .12 and the probability of familiarity by .09.

Figure 8 shows the group ROC curves for the inclusion and exclusion conditions for Experiment 3. The figure shows that although performance was higher for the strong items than for the weak items, the ROC curve for the strong items was no

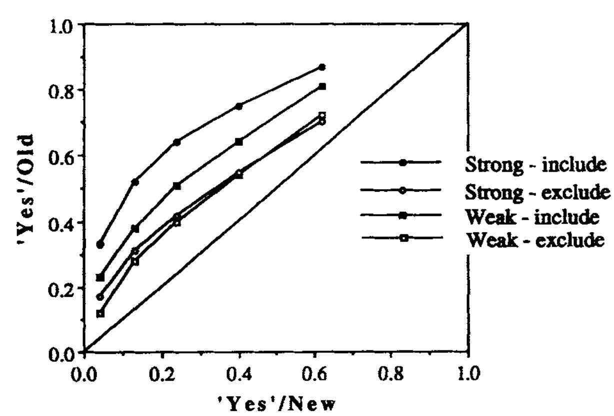


Figure 8. Receiver-operating characteristics for the inclusion and exclusion conditions for weak and strong items in Experiment 3.

more skewed than was that for the weak items. Linear regression analysis was performed on the z-transformed data to estimate the slopes and intercepts for each subject. The average slopes and intercepts of the weak and the strong items are presented in Table 6. The average goodness of fit for the linear regressions (R^2) was .97. The slope of the transformed ROC curves did not differ from weak (.79) to strong (.80) items, ($F < 1$). However, the intercept was greater for the strong items than for the weak items, $F(1, 15) = 12.651$, $MS_e = .069$. These results replicate those of Ratcliff et al. (1992), who found the slope of the ROCs to remain constant at approximately .80 across changes in item strength.

Predictions of the Dual-Process Model

The strength manipulation was found to increase both recollection and familiarity, as predicted by the model. To test the model further, the estimates for recollection and familiarity were used to derive predictions for the intercept and the slope of the ROC curves. Curves were calculated as described in Experiment 1. As in the previous experiments, the predicted values were close to the observed data. The predicted and observed intercepts were .66 and .65 for the weak items and .97 and .98 for the strong items. The predicted and observed slopes were .86 and .79 for the weak items and .81 and .80 for the strong items.

Unlike the previous experiments, the proportions of old items in the highest confidence category were greater than the estimated values of recollection. The proportions of high-confidence hits in the weak and strong conditions were .23 and .33, respectively, compared with the respective recollection estimates of .10 and .22, which is an average difference of .12. In Experiment 1, the proportions of hits that were in the high-confidence category fell below the estimates for recollection by approximately .06. In Experiment 2, the average again fell slightly below the recollection estimate by .03. However, the increase in Experiment 3 reflects a higher false-alarm rate. In the first two experiments, the false-alarm rate in the high-confidence category was .00, compared with a false-alarm rate of .04 in Experiment 3. In fact, examination of the ROCs in Experiment 3 (Figure 8) suggests that if the curves were extended to the point where false-alarms would be equal to .00, the inclusion scores would closely approximate the derived estimates for recollection. It would seem that in this experiment, subjects were more lenient with the treatment of familiarity and were responding with the highest level of confidence to the more familiar items as well as the recollected items.

Figure 9 presents the estimates for familiarity and recollection for weak and strong items as a function of false-alarm rate.

Table 6
Slopes and Intercepts for Weak and Strong Items in Experiment 3

ROC measure	Weak *M*	Weak *SD*	Strong *M*	Strong *SD*
Slope	0.79	0.14	0.80	0.27
Intercept	0.65	0.26	0.98	0.35

Note. ROC = receiver-operating characteristic.

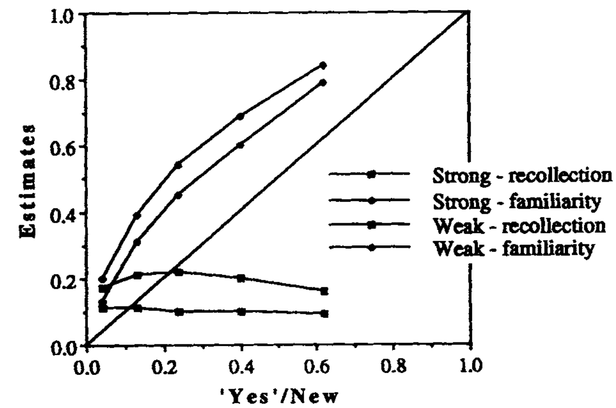


Figure 9. Estimates for recollection and familiarity for weak and strong items as a function of response confidence in Experiment 3.

As in Experiments 1 and 2, familiarity exhibited the symmetrical ROC curve predicted by signal detection theory. The slopes of the z-transformed curves were .92 for the weak items and .91 for the strong items. Recollection was greater for the strong items than for the weak items. The estimates for recollection remained constant across the entire range of false-alarm rates. This differed from the previous experiments in which the estimates showed a drop at the extreme ends of the function. However, unlike the previous experiments, the inclusion and exclusion condition scores in Experiment 3 did not approach ceiling or floor (see Figures 4, 6, and 8). When ceiling and floor effects are avoided, recollection remained constant, as would be expected if it were an all-or-none retrieval process.

General Discussion

The results of the three experiments provided strong support for a dual-process model of recognition memory, whereby recognition judgments were based on an assessment of familiarity, as well as a recollection process such that qualitative information about the study episode was retrieved. Familiarity was found to be a signal detection process, whereby only items exceeding some criterion were judged as old. Recollection, on the other hand, was found to reflect an all-or-none retrieval process that either succeeded or failed. The assessment of familiarity produced a symmetrical ROC, but the recollection process introduced a skew such that the slope of the ROC was less than 1.0.

In Experiments 1 and 2, decreasing the length of the study list increased the probability of recollection but left the use of familiarity relatively unaffected. When the contribution of recollection increased while familiarity remained constant, the ROC became more skewed (the slope decreased); thus in both experiments, increases in intercept were accompanied by decreases in slope. In Experiment 3, increasing the study time increased both recollection and familiarity. When the contribution of both processes increased, the intercept increased while the ROC slope remained constant. In Experiment 2, the ROCs for standard recognition instructions were found to be almost identical to those for the include conditions; as list length decreased, the intercept increased and the slope decreased.

The finding that increasing list length decreased recollection but left familiarity relatively unaffected replicated those of Yonelinas and Jacoby (in press). Furthermore, the finding that changes in list length produced opposite effects on the slope and intercept of the ROC was similar to the effects found for other interference manipulations. As previously mentioned, increases in study-test lag (Donaldson & Murdock, 1968) as well as increases in the delay between study and test (Gehring et al., 1976) produced a decrease in d' accompanied by an increase in slope. In the current study, it was found that the change in slope was due to an increase in recollection. It is likely that the same can be said for the other interference manipulations. Similarly, material manipulations that produce changes in d' accompanied by changes in slope may also arise because of changes in the probability of recollection. Finally, in Experiment 3, it was found that increases in study time increased d' but did not change the ROC slope. The results of that experiment replicate those of Ratcliff et al. (1992) and Egan (1958) in finding that changes in strength did not lead to changes in ROC slope. However, the current study extends those findings by showing that the constant slope arose because both recollection and familiarity increased with strength.

The overall pattern of results is in agreement with the dual-process model. Recollection contributed to performance for long and short lists and for strong and weak items, thus ROC slope was always less than 1.0. Furthermore, when recollection increased while familiarity remained relatively constant (list length), the increase in recollection led to an increase in the intercept and a decrease in the ROC slope. When both recollection and familiarity increased together (strength), then the intercept increased but the slope remained constant. Several other more specific predictions of the model were also supported. Based on the inclusion and exclusion condition scores collapsed across levels of confidence, the model was used to generate hypothetical ROCs. The intercepts and slopes of the predicted ROCs were found to closely approximate those of the observed ROCs in all three experiments. Moreover, estimates for recollection and familiarity were derived as a function of confidence, and these revealed that the two processes were operating in agreement with the model. The use of familiarity was found to increase gradually as response criterion became more lax, producing a symmetrical ROC with slope approaching unity. This supports the notion that familiarity reflects an equal-variance signal detection process. The contribution of recollection, however, did not resemble familiarity but remained constant across changes in criterion, supporting the idea that recollection was an all-or-none retrieval process.

One should note that the slopes of the familiarity curves were slightly less than 1.0 (averaged across experiments, the slope was .95), and it was only for the long lists that the slopes reached unity. One possibility is that the deviation from 1.0 reflects partial recollection. That is, subjects may have occasionally recollected information about a studied word that did not support the list discrimination that was required. For example, the subject may remember that they coughed as a word was studied, but this information would not allow subjects to determine which list the item was in. Such partial recollection would not be captured by the estimate of recollection and might contaminate estimates of familiarity. To the extent that this occurred, the estimates of familiarity would begin to look like the estimates of recollection—manipulations should affect the estimates of both processes in a similar manner. If partial recollection were occurring in the short lists, then one might expect the slope of the familiarity curves to be slightly less than unity. Partial recollection would also explain why there was a small effect of list length on familiarity. Across the two list length experiments, when base rate was taken into account, changes in list length led to a decrease in familiarity of .04 compared with a .20 drop in recollection.

The notion that recollection leads to high-confidence responses was generally supported. The estimates for recollection remained relatively flat across the range of false-alarm rates. Although there was a tendency for the estimates of recollection to decrease at the extreme false-alarm rates, this was considerably reduced in Experiment 3 in which floor and ceiling effects were minimized. Moreover, an examination of the proportion of hits in the highest confidence response category suggested that most of the recollected items did lead to high-confidence responses.

However, at least in the first two experiments there was a tendency for the derived estimates of recollection to be slightly greater than the observed proportion of high-confidence hits (by .06 and .03 in Experiments 1 and 2, respectively). One possibility is that some small proportion of recollected items were not assigned to the highest confidence category. This could occur if there were different types of list discrimination information recollected that led to higher or lower levels of confidence, or if more list-relevant information was retrieved for some items than others. To see how this could occur, consider the case in which a 20-point confidence scale were used, rather than the 6-point scale that was used in the current studies. It is likely that under these conditions, subjects would begin to spread out their recollected items rather than using just the two extreme points.

However, another possible explanation is that subjects were relying at least partially on some form of familiarity based list discrimination.[2] If subjects were using familiarity as a basis for list discrimination (e.g., "If the item is very familiar it was probably in List 2"), this would tend to inflate the estimates of recollection because recollection is estimated as the ability to determine list membership, and any familiarity based list discrimination will be included in this estimate. Moreover, it seems likely that such familiarity based list discrimination would not lead to high-confidence responses but rather would contribute to recollection at an intermediate level of confidence. This would lead to a decrease in the estimate of recollection at the highest level of confidence as well as the tendency for the high-confidence hits to be slightly less than the derived estimate of recollection (see Experiments 1 and 2).

If subjects were using familiarity as a basis for list discrimination, then one would expect to see differences in the false-alarm rates under the two test instructions. That is, if subjects were accepting only high-familiarity items under the "List 2?" instructions and medium familiarity items under "List 1?"

[2] The possibility of familiarity based list discrimination was suggested by Janet Metcalfe.

instructions, then the false-alarm rate to new items should be greatest under "List 1?" instructions. Subsequent analysis showed that there was a slight tendency for this to occur; the average difference in false-alarm rates for List 1 and List 2 instruction was .03, .02, and .01 in Experiments 1, 2, and 3, respectively. However, the difference did not approach significance in any experiment. Although any influence of familiarity based list discrimination would seem to be quite small, it could be at least partially responsible for the decreases in recollection seen at the extreme false-alarm rates as well as the observation that the estimate for recollection was occasionally greater than the proportion of high-confidence hits.

A potential criticism of the process dissociation procedure is that the conclusions drawn from this procedure may not generalize to standard recognition tests. Because in standard tests subjects are only required to make old–new judgments, it is possible that they base those judgments on assessments of familiarity alone. In the list discrimination procedure used in the current experiments, subjects were required to recollect. However, the fact that the ROC curves for the inclusion condition test were almost identical to those for the recognition test (Experiment 2) suggests that similar processes support performance in list discrimination and standard recognition tests.

A related question is the generality of the pattern of results to other study conditions. In all three experiments, subjects were highly motivated to encode the items in such a way that they could later recollect them. If recollection were never required, one might expect that subjects' encoding strategies might differ such that the probability of recollection was considerably reduced. This, of course, would lead to slopes closer to 1.0. However, in all of the other ROC studies mentioned earlier, list discrimination was never required, and the slopes were often considerably less than 1. It would seem that although recollection may be reduced, it is difficult to eliminate altogether.

The results of the current study present problems for a number of current global memory models. First, the results of the ROC analysis showed that as list length increased, d' decreased and slope increased. Moreover, as item strength increased, d' increased but slope remained constant. The theory of distributed associative memory (TODAM; Murdock, 1982) on one hand predicts that the slope of the ROC will remain constant and close to 1.0 as d' increases (see Ratcliff et al., 1992). In Experiments 1–3, the slope was considerably less than 1.0. Moreover, list length was found to produce a change in d' that was accompanied by a change in slope. The search of associative memory (SAM; Gillund & Shiffrin, 1984) and MINERVA 2 (Hintzman, 1986), on the other hand, predict that as d' increases, the slope should decrease (see Ratcliff et al., 1992). Although these two models could account for the effects of list length, they cannot account for the effects of item strength.

The inability of the models to account for the diverse pattern of results may lie in the assumption, made by all of the models, that recognition judgments are based solely on the assessment of a single familiarity process. One option would be to drop the single factor assumption by introducing a second mechanism that is qualitatively different from familiarity. In fact, all of these models do possess recall-like search mechanisms that could be incorporated into recognition. However, even if the models could be modified to account for the changes in ROC slope, it is not clear that they could produce the observed pattern of results in both the inclusion and the exclusion conditions.

In Experiments 1–3, the process dissociation procedure was used to examine the processes underlying recognition performance. However, other measurement procedures have been used. For example, Tulving (1985) and Gardiner (1988) have used a procedure in which subjects are asked to report whether they remember an item from the study list or if they just know it was presented. One might expect "remember" responses to reflect the recollection process and "know" responses to reflect the familiarity process. However, there are a number of differences between the two procedures, and work is under way to more carefully examine the relationship between these two procedures.

The two processes found to operate in recognition memory may also support performance in other memory tasks. Using a procedure similar to that used in the current study, Jacoby et al. (1993) found that word-stem completion performance reflected a mixture of conscious and unconscious uses of memory. These different uses of memory may reflect the same processes that support recognition judgments. For example, the unconscious uses of memory in tasks such as stem completion may reflect a signal detection process similar to the one found to operate in recognition.

References

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. 242–293). San Francisco: Freeman.

Donaldson, W., & Murdock, B. B., Jr. (1968). Criterion change in continuous recognition memory. *Journal of Experimental Psychology, 76,* 325–330.

Egan, J. P. (1958). *Recognition memory and the operating characteristic* (Tech. Note No. AFCRC-TN-58-51). Bloomington: Indiana University, Hearing and Communication Laboratory.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16,* 309–313.

Gehring, R. E., Toglia, M. P., & Kimble, G. A. (1976). Recognition memory for words and pictures at short and long retention intervals. *Memory & Cognition, 4,* 256–260.

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

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 5–16.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics.* New York: Wiley.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

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

Jacoby, L. L. (1994). *The effects of word frequency on recollection and familiarity.* Unpublished manuscript.

Jacoby, L. L., Toth, J. P., & Yonelinas, A. P. (1993). Separating conscious and unconscious influences of memory: Measuring recollection. *Journal of Experimental Psychology: General, 2,* 1–16.

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

Mandler, G., & Boeck, W. (1974). Retrieval processes in recognition. *Memory & Cognition, 2,* 613–615.

Murdock, B. B., Jr. (1982). A theory of storage and retrieval of item and associative information. *Psychological Review, 89,* 609–626.

Murdock, B. B., Jr., & Dufty, P. O. (1972). Strength theory and recognition memory. *Journal of Experimental Psychology, 94,* 284–290.

Norman, D. A., & Wickelgren, W. A. (1969). Strength theory of decision rules and latency in retrieval from short-term memory. *Journal of Mathematical Psychology, 6,* 192–208.

Ogilvie, J. C., & Creelman, C. D. (1968). Maximum likelihood estimation of receiver operating characteristic curve parameters. *Journal of Mathematical Psychology, 5,* 377–391.

Ratcliff, R., & Murdock, B. B., Jr. (1976). Retrieval processes in recognition memory. *Psychological Review, 83,* 190–214.

Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review, 3,* 518–535.

Strong, E. K., Jr. (1912). The effect of length of series upon recognition memory. *Psychological Review, 19,* 447–462.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist, 26,* 1–12.

Yonelinas, A. P., & Jacoby, L. L. (in press). Dissociations of processes in recognition memory: Effects of interference and response speed. *Canadian Journal of Experimental Psychology.*

Received July 14, 1993
Revision received November 12, 1993
Accepted November 23, 1993 ■