Dissociating Recollection and Familiarity
in
Recognition Memory

by

Andrew Peter Yonelinas, B.Sc.

A Thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree
Doctor of Philosophy

McMaster University

DOCTOR OF PHILOSOPHY MCMASTER UNIVERSITY

(Psychology) Hamilton, Ontario

TITLE: Dissociating Recollection and Familiarity in Recognition Memory

AUTHOR: Andrew P. Yonelinas, B.Sc. (University of Toronto)

SUPERVISOR: Dr. Larry L. Jacoby

NUMBER OF PAGES: x, 49

Abstract

The processes underlying recognition memory were examined using the process dissociation procedure (Jacoby, 1991). Recognition judgements based on recollection were found to differ in several fundamental ways from those based on familiarity. Increasing list length interfered with recollection but left familiarity intact. Response time distributions and the results of a response signal procedure showed that familiarity was faster as a basis for recognition judgements than was recollection. The analyses of receiver operating characteristics (ROCs) showed that recollection led to high confidence memory judgements that remained relatively constant as response criterion was varied. Familiarity, on the other hand, increased gradually as response criterion was relaxed. Results support a dual-process model of recognition in which a discrete recollection process operates independently of a continuous familiarity process.

ACKNOWLEDGMENTS

I thank Maddy and Ed for their endless love and support, Ian and Lee for their valuable guidance, and Larry for showing me that there is a difference between seeing and just looking. Joanna, thank you for your inspiration and love, your help in thinking and writing about this work, and for sharing those chocolate brownies.

TABLE OF CONTENTS

Chapter 1

Chapter 2

Chapter 3

LIST OF FIGURES

Chapter 2

Chapter 3

LIST OF TABLES

Chapter 2

Chapter 3

Introduction

"I recognize him as familiar, but I can't recollect having met him before." Statements such as this have been important for theories of memory in suggesting that there are two different basis for recognition judgements. Several researchers have argued that recognition judgements can be based on assessments of familiarity or on recollection (i.e., retrieval) of qualitative information about the study event (e.g., Atkinson & Juola, 1974; Jacoby & Dallas, 1981; Mandler, 1980). Although recognition memory has been studied quite extensively, relatively little is known about the processes that underlie performance. The aim of this dissertation is to examine the nature of these memory processes using the process dissociation procedure (Jacoby, 1991) to derive qualitative estimates for recollection and familiarity. I begin with a discussion of the current dual process theories of recognition and the procedures used to measure the influence of recollection and familiarity. This is followed by a discussion of the process dissociation procedure (Jacoby, 1991) and a modification of that procedure which I used to estimate recollection and familiarity. Finally, I discuss several fundamental issues about the processes underlying human memory that were examined in Chapter 2 (Yonelinas and Jacoby, 1994), and Chapter 3 (Yonelinas, 1994).

Dual Process Theories and Measurement Procedures

One of the central claims of dual process theory is that familiarity is a relatively fast process. Atkinson and Juola (1974) argued that when subjects are making recognition judgements, they first make a quick assessment of familiarity. High levels of familiarity lead to a fast yes response, and low levels of familiarity lead to fast no responses. Items with intermediate levels of familiarity require recollection (i.e., a controlled search of memory). In this way, fast responses should be based on familiarity and slower responses on recollection.

Mandler also claimed that familiarity was the faster of the two processes. He argued that familiarity could be assessed quickly and led to fast recognition judgements, but recollection, a process similar to that used in free recall tests, was considerably slower. In support of this claim, Mandler and Boeck (1974) examined response-time distributions and found that only the slowest recognition judgments were affected by the organizational variables that influenced recall performance. For example, they found that the number of categories that subjects sorted the study items into influenced slow recognition judgements, as well as recall performance, but did not influence fast recognition judgements.

Mandler and Boeck's study shows that familiarity is faster than recollection in some cases. However, response speed does not always dissociate the two processes. For example, Gillund and Shiffrin (1984) used a deadline procedure, and either required subjects to respond before a 900 ms deadline or to wait until after a 1 or 1.5 s delay. They compared performance across a number of variables (levels of processing, list length, and number of study presentations) and found that none of the variables interacted with response speed. Thus, even with variables that might be expected to differentially affect recollection and familiarity, the time allowed for responding does not always dissociate the two processes.

Mandler (1980) proposed another method for examining the processes underlying recognition; he argued that recollection was a recall-like process, and thus could be estimated as the probability of producing study words in a recall test. He argued that recollection could be algebraically removed from overall recognition performance to attain a probability estimate for familiarity. To do this, he assumed that recollection and familiarity contributed independently to recognition performance, and thus, the probability of recognizing an item that was previously studied (i.e., old) was equal to the probability that the item could be recalled (Rc) plus the probability it was familiar (F) minus the probability that the item was recalled and familiar [Recognition("old"/old) = Rc + F - Rc*F]. Mandler measured the probability of recognition and recall, and used these scores to calculate familiarity.

A difficulty with this procedure is that performance in a recall test may not provide a pure measure of recollection in a recognition test, and thus the procedure may produce biased estimates. The probability that recollection occurs in a free recall test, where there are no retrieval cues, will likely be less than that in a recognition test where the entire target item is used as the retrieval cue. This will lead the procedure to underestimate recollection. Moreover, it will lead to biased estimates of familiarity as well, because recall is used in the calculation of familiarity.

An alternative approach was proposed by Jacoby (1991), who introduced the process dissociation procedure. The procedure provides estimates for recollection (R) and familiarity (F), but does not rely on response speed or performance in a recall test to do so. Jacoby derived estimates by contrasting performance in an _inclusion_ condition where both processes act in concert, to performance in an _exclusion_ condition where the two processes act in opposition. In phase one of Jacoby's (1991) initial study (Experiment 3) subjects read a list of words aloud. In phase two, subjects heard a different list of words and were instructed to remember them for a later recognition test. At test, subjects were presented with a list containing a mixture of earlier seen, earlier heard, and new words, and were given either inclusion or exclusion instructions. In the inclusion condition, they were instructed to call a word _old_ if it was in either the seen or heard list. In the exclusion condition they were instructed to call a word old only if it was in the heard list. Furthermore, they were told that if they could recollect that the word was seen they could be sure the word was not heard and they should call it new. In this way, subjects included seen words in the inclusion condition and excluded those words in the exclusion condition.

Performance in the inclusion and exclusion conditions for the seen words was used to derive estimates for recollection and familiarity in the following way. If the two processes are independent, the probability of responding old to a seen item in the inclusion condition can be written:

$$P(\text{"old"}/\text{seen})\text{inc} = R + F - RF$$

the probability that the item is recollected (R) plus the probability that it is familiar (F), minus the probability that the item is recollected and familiar (RF). That is, a seen item can be accepted as old if it is recollected as having been seen, or if it is sufficiently familiar to be judged old.

The probability of responding old to a seen item in the exclusion condition can be written:

$$P(\text{"old"}/\text{seen})\text{exc} = F - RF$$

the probability the item is familiar, minus the probability it is familiar and recollected. That is, subjects will call a seen word old only if it is familiar but they cannot recollect that it was seen. Recollection was calculated by subtracting the exclusion score from the inclusion score (R

+ F - RF - (F - RF) = R). Having solved for R, either of the two equations could be used to solve for familiarity (e.g., P("old"/seen)exc / (1-R) = F).

The procedure has been used in several studies of recognition memory. In the study just described (Jacoby, 1991), dividing attention was found to selectively decrease recollection, leaving familiarity unaffected. In contrast, solving words as anagrams at study led to greater recollection and familiarity than simply reading at study. Verfaellie and Treadwell (1993) used the same procedure to examine the effects of amnesia on recollection and familiarity. They found that amnesics exhibited a severe deficit in recollection, but familiarity was functioning at a normal level. A similar, although less dramatic effect, was found with aging (Jennings & Jacoby, 1994). Recollection was found to decrease with age but familiarity remained unaffected.

The List Discrimination Procedure

Typically, the process dissociation procedure is implemented by contrasting performance under two different sets of instructions (i.e., inclusion and exclusion). However, the procedure was modified slightly for the current studies. Rather than using inclusion and exclusion instructions, a list discrimination task was used, and inclusion and exclusion measures were extracted from within that task. By this formulation, recollection is defined rather strictly as the ability to determine list membership.

In most of the experiments reported here, subjects were presented with two study lists. A recognition test followed in which subjects received one of two different test instructions: "Was this item in List 1?" or "Was this item in List 2?". They were told that none of the items were in both lists, and, so, they should respond _yes_ if the item was from the target list, and respond _no_ if the item was new or if they could recollect that the item appeared in the nontarget list. Responses to items under these two sets of instructions provided inclusion and exclusion measures that were used to derive estimates for recollection and familiarity.

Inclusion performance was measured as the probability of correctly accepting items from a target list (i.e., list 1 items accepted under list 1? instructions, and list 2 items accepted under list 2? instructions). Under these conditions, subjects could accept a target item either if it was recollected (i.e., they recollected list membership) or if it was familiar. If the two processes are independent, then the probability of accepting an item is equal to the probability that the item is recollected (R) plus the probability that the item is familiar (F) minus the probability the item is recollected and familiar (RF).

Exclusion performance was measured as the probability of accepting an item from the nontarget list (i.e., list 1 items accepted under list 2? instructions, and list 2 items accepted under list 1? instructions). Under these conditions, items would be accepted only if they were familiar and not recollected. Otherwise, recollection that the item was from the nontarget list would lead to a _no_ response. Thus, the probability of accepting an item in the exclusion conditions is equal to the probability the item is familiar minus the probability the item is both recollected and familiar.

As with the standard procedure, recollection was estimated by subtracting the exclusion measure from the inclusion measure, and familiarity was estimated by dividing the exclusion measure by the probability that the item was not recollected.

The list discrimination procedure was chosen over the standard procedure because it eliminated a possible confound. Under standard inclusion conditions (i.e., respond _yes_ to old items) subjects are not required to recollect information that discriminates between words from the first and second lists. However, under exclusion conditions (i.e., respond _yes_ only to the items from the second list), subjects must recollect information about list membership. Subjects may be more likely to engage in recollection in the exclusion condition where it is required to avoid accepting nontarget items than in the inclusion condition where the task could be completed without recollection. If the difference in inclusion and exclusion instructions leads subjects to rely differentially on recollection and familiarity in the two conditions then the estimates produced by the procedure may be somewhat distorted. This is because the procedure assumes that the two processes are the same in the two conditions.

Under list discrimination instructions, subjects must always monitor for items from the target and nontarget lists. Because the task demands are more similar in the two list discrimination conditions, which provide inclusion and exclusion measures, it is more likely that the processes in the two conditions will be the same.

However, one will note that in Experiment 2 of Chapter 3, inclusion performance in the list discrimination task was not found to differ from that of the recognition (i.e., standard inclusion) condition. Thus the concern with differences in task demands between inclusion and exclusion conditions may have been unwarranted.

In any case, one useful property of the list discrimination procedure is that it provides a larger number of observations, that can be used to derive estimates for recollection and familiarity, than does the original procedure. In the initial procedure, it is possible to derive estimates for recollection and familiarity only for items in the first list (i.e., seen words). Responses to the items from the second list are collected, but because subjects include those items under both sets of instructions, there is no measure of exclusion performance for those items. Thus estimates for those items cannot be derived.

With the list discrimination procedure, on the other hand, estimates can be derived for items from both lists. This essentially doubles the number of observations available to derive estimates. The list discrimination procedure was useful in the current context because both the latency distribution analysis in Chapter 2 and ROC analysis in Chapter 3 required a large number of responses from each subject.

Dissociating Recollection and Familiarity

The aim of this dissertation is to make use of the process dissociation procedure to examine the nature of recollection and familiarity. The experiments focus on three fundamental questions. First, is it the case, as several researchers have claimed, that familiarity is faster that recollection? Second, is recollection, like free recall, highly susceptible to interference effects? Third, can recollection and familiarity be described as discrete or continuous processes?

Most people have had the experience of finding a person familiar, and only later recollecting who the person is. This experience suggests that familiarity may be a faster process than recollection. As already discussed, there is some empirical evidence that familiarity is faster than recollection (Mandler & Boeck, 1974). However, not all attempts to separate the two processes by examining response speed have been successful (Gillund & Shiffrin, 1984).

One of the aims of Chapter 2 was to use the process dissociation procedure to determine if familiarity is faster than recollection. This was done by examining estimates of recollection and familiarity as a function of response time. Response times were collected while subjects responded under inclusion and exclusion conditions. Responses were then partitioned into separate response time bins, and estimates of recollection and familiarity were derived. The estimated familiarity and recollection latency distributions were examined to determine the time course of the two processes.

A separate, response signal procedure was also use to examine process speed. Using this procedure, subjects' response times were controlled by requiring subjects to respond either very rapidly or more slowly at the time of test. Examining the effect of response signal on recollection and familiarity would show if the recollection process took longer to complete than familiarity.

Results from the latency distribution analysis, as well as the response signal procedure, showed that familiarity was considerably faster that recollection. Familiarity's contribution to performance peaked prior to that of recollection. Additionally, forcing subjects to respond quickly had a greater detrimental effect on recollection than familiarity. These results provide strong support for the claim that familiarity is the faster of the two processes.

A second question that was examined in the current studies is whether the two processes differ in terms of their susceptibility to interference. As previously discussed, Mandler argued that recollection is a recall-like process. One striking property of free recall performance is that it is very susceptible to interference. For example, increasing the number of items that are to be remembered leads to a dramatic reduction in the proportion of items recalled (e.g., Roberts, 1972). If recollection is a recall-like process, then, like recall, it should be very susceptible to interference. This claim was tested in several experiments by examining the effect of increasing list length on recollection and familiarity. Results from three experiments in Chapter 2, and two experiments in Chapter 3 support the prediction that recollection is highly susceptible to list length interference. Moreover, familiarity was found to be relatively unaffected by such a manipulation.

One final question that was examined is whether recollection and familiarity reflect discrete or continuous processes. Introspection suggests that recollection is a discrete (i.e., all-or-none) process, but familiarity is a continuous process. Most often, one either succeeds or fails at recollecting who a person is - only rarely do we half-way recollect having encountered a person. This of course is not to say that we cannot recollect many different things about a person, only that for any given feature, we either succeed or fail to recollect it. However, familiarity does not seem to be all-or-none. Rather, it is experienced as graded or continuous in nature. These observations lead to the intriguing possibility that recollection reflects a discrete retrieval process, but familiarity reflects the assessment of a continuous dimension.

These claims were tested in Chapter 3 by examining performance as a function of response criterion and plotting receiver operating characteristics (ROCs). The idea behind the ROC analysis was that if recollection is discrete then varying response criterion should not greatly influence the process. That is, because there are no intermediate states, shifts in criterion should not change recollection. In contrast, if familiarity reflects the assessment of a continuous dimension, and is not associated with any discrete states, then familiarity should increase gradually as response criterion is relaxed. Thus, whether a given level of familiarity is sufficient for a positive recognition response will depend on the subject's response criterion.

The results of the ROC analysis in Chapter 3 supported the claim the recollection and familiarity are discrete and continuous processes respectively. Moreover, in contrast to familiarity, recollection was found to lead to highly confident responses, and when it occurred it was almost always correct. A dual-process signal-detection model was proposed that incorporated these notions and was found to account for ROC results that previous memory models could not.

The results of the experiments presented in the next two chapters provide general support for dual process theories of recognition memory by showing that judgements based on recollection differ in a number of fundamental ways from those based on familiarity. Furthermore, the results provide several further constraints on models of memory, and provide a basis for a more comprehensive understanding of the processes underlying human memory.

Dissociations of Processes in Recognition Memory: Effects of Interference and of Response Speed

ANDREW P. YONELINAS and LARRY L. JACOBY
McMaster University

Abstract Effects on two bases for recognition-memory judgements were examined using a process dissociation procedure (Jacoby, 1991). In three experiments it was found that increasing the length of a study list interfered with conscious recollection but left familiarity in place. Furthermore, an examination of reaction time distributions as well as results from a response-signal procedure showed that familiarity was faster as a basis for recognition judgements than was conscious recollection. However, both bases contributed to performance on the fastest as well as the slowest responses, suggesting that the two processes were acting in parallel.

For a test of recognition memory, subjects must decide whether a test item was presented in a previously studied list. At least in principle, subjects could base their recognition judgements solely on the familiarity of the test items because, on average, an item that was presented in the study list would be more familiar than one that was not. A rationale of this sort underlies single-factor theories such as signal detection theory (see Swets, 1964; Wickelgren, 1972). However, subjects may not be limited to assessments of item familiarity when making recognition-memory judgements. If some aspect of the study event could be consciously recollected (e.g., 'I remember seeing that word... it was the first word in the list') this could serve as a second basis for responding.

Several researchers have proposed a dual-process view of recognition memory along with criteria that can supposedly be used to distinguish between the two bases for responding. Familiarity is assumed to be a fast basis for responding (Atkinson & Juola, 1974; Jacoby, 1991; Mandler, 1980) that relies on perceptual characteristics (Jacoby & Dallas, 1981; Mandler, 1980) or item-specific information (Humphreys & Bain, 1983), and reflects an automatic or unconscious use of memory (Jacoby, 1991) that is largely spared by amnesia (e.g., Piercy & Huppert, 1972; Verfaellie & Treadwell, 1993). In contrast, the use of recollection is described as a slow, search-like

Canadian Journal of Experimental Psychology, 1994, 48:4, 516-534
Reprinted with permission of the Canadian Psychological Association

process that relies on conceptual processing or associative information and requires attention. Furthermore, recollection is said to be absent or reduced in amnesic patients.

In this paper we examine the two bases for recognition memory by focussing on differential effects of interference, and differences in response speed. Although the examination of interference effects has a long history in recognition experiments, it has received little attention in the context of dual-process theories. Thus we know that manipulations such as increasing the length of the study list (e.g., Strong, 1912) or increasing the delay between study and test (e.g., Strong, 1913) interferes with recognition memory performance, but we do not know whether just one or both of the two bases for recognition are influenced.

To examine interference effects, we varied the length of study lists with the expectation that increasing list length would increase interference. The list length effect was first reported by Strong (1912) who found that as he increased the number of advertisements a subject studied, the probability of later recognizing a particular advertisement decreased. Since then the list length effect has been demonstrated numerous times in recognition (e.g., Atkinson & Shiffrin, 1971; Gillund & Shiffrin, 1984; Ratcliff & Murdock, 1976), free recall (Murdock, 1960; Roberts, 1972) and word-fragment completion (Sloman, Hayman, Ohta, Law & Tulving, 1988).

A number of observations led us to believe that recollection might be more susceptible to interference than familiarity. First, the magnitude of the list-length effect is greater in free recall than in recognition. Doubling list length produces a dramatic drop in free recall (Roberts, 1972) but only a small decrease in recognition (Ratcliff & Murdock, 1976). If the recollective process involved in recognition is similar to that of free recall, then it might be this process alone that is producing the list length effect in recognition. Second, Jacoby (1991) found that dividing attention reduced recollection but had no effect on the use of familiarity as a basis for recognition decisions. The greater interference produced by increasing list length might have effects similar to those of dividing attention: reducing recollection but leaving familiarity in place.

In contrast to the effects of interference, differences in processing speed have attracted considerable attention. Because familiarity is generally believed to be the faster of the two processes (see Atkinson & Juola, 1974; Mandler & Boeck, 1974), an obvious test of the dual process theory is to look for dissociations between fast and slow recognition responses. If some variable differentially affects the fast and slow responses, this would lend support to the dual-process claim.

There are a number of studies that suggest that the two processes do differ in terms of speed. For example, Mandler and Boeck (1974) examined response-time distributions and found that slow recognition judgements were more affected by organizational variables than were faster judgements. They argued that this was because recollection, like recall, was influenced more by organizational variables than was familiarity. That is, because recollection was slower than familiarity, the effects of list organization were greater for the slower responses.

Another method that has proved useful for assessing differences in processing speed is the response-signal procedure. With this procedure, the experimenter controls the time allowed for memory retrieval by requiring subjects to respond at a given speed. Each test item is followed by a signal to respond that occurs anywhere from 200 ms to 3 s after its presentation, and, typically, subjects are required to respond within 300 ms of the signal. Using this procedure, Dosher (1984) found that in a word-pair-recognition task, subjects tended to first accept then reject semantically-related lures. This suggests that fast responses were predominately based on familiarity, which did not discriminate between prior semantic associations and associations formed during the study phase; only as the products of the recollection process became available were those discriminations possible. Using a similar procedure, Gronlund and Ratcliff (1989) found that judgements requiring recollection of associative information were more time consuming than were judgements that could be based upon item familiarity.

However, in contrast to the results of the above experiments, Gillund and Shiffrin (1984) found no evidence that fast and slow recognition responses had qualitatively different bases. Using a deadline procedure, they required subjects either to respond before a 900 ms deadline or to withhold responding until after a 1 or 1.5 s delay. They compared performance for fast and slow recognition judgements across a number of variables (levels of processing, list length, and number of study presentations) and found that none of the variables interacted with response deadline. Thus, even with variables that might be expected to differentially affect the two processes, the amount of time allowed for judgements did not always dissociate the two processes. For this reason, we chose to rely on an alternative method.

We made use of a process-dissociation procedure (Jacoby, 1991) that allowed us to derive quantitative estimates for the contribution of familiarity and recollection. The procedure involves comparing performance in a condition in which the two processes act to produce the same response, to a condition in which they act in opposition to produce different responses. The procedure, which we will describe shortly, is based on the assumption that recollection and familiarity serve as *independent* bases for responding. Recollection is assumed to be a consciously controlled process that provides qualitative information about the remembered event. This could include such information as the source of the memory, the perceptual aspects of the stimuli, or the conceptual processing associated with the event. Familiarity, on the other hand, is assumed to be an automatic process providing only an overall strength measure of memory. Familiarity is assumed to allow simple memory discriminations (i.e., old vs new), but not to support discriminations requiring qualitative information about the study event. Let us describe the process-dissociation procedure by outlining its use to examine the effects of list length on the different bases for recognition judgements.

In the first experiment, subjects were presented with a list of words, some visually and others auditorily. In one condition (the inclusion condition), subjects were presented with a recognition memory test for the words that they had *seen* earlier. Subjects were warned that the test list would also contain words that they had heard earlier, and that none of the words were both seen and heard. Consequently, if they could recollect earlier hearing a tested word, they could be certain that the word was not seen and that they should respond "no". Put simply, they should include seen words and exclude heard words.

Seen words could be selected either on the basis of recollection (subjects recollect having seen the word) or, alternatively, on the basis of their familiarity. If we assume that recollection and familiarity serve as two independent bases for recognition judgements, then the probability of selecting an old item in the inclusion condition is:

$$R + F - RF$$

the probability that an item is recollected (R) plus the probability that it is familiar (F) minus the intersect of the two.

In another condition (the exclusion condition), subjects were presented with a recognition test for the words that they *heard* earlier (they should exclude seen words and include heard words). For that test condition, seen words that were accepted must have been sufficiently familiar to be selected as old but not been recollected as earlier seen. That is, the probability of responding "yes" to a seen item in the exclusion condition is:

$$F - RF$$

the probability that an item is familiar (F) minus the probability of the item being familiar and recollected (RF).

The contribution of recollection to recognition can be estimated by subtracting the probability of selecting an item in the exclusion condition from that in the inclusion condition ($(R + F - RF) - (F - RF)$). Having found the contribution of recollection, we solve either of the previous equations to estimate the contribution of familiarity (i.e., $EXC/(1 - R)$).

Familiarity, as estimated by this means, reflects both the effect of memory (M) from the study phase and the base rate familiarity (B) of the items. If we assume that the two effects are additive:

$$F = M + B$$

Given that assumption, one can estimate the increase in familiarity due to the study phase (M) by simply subtracting base rate as measured by responding to "new" test words. An alternative approach would be to apply signal-detection theory (Swets, 1964) to analyze differences in familiarity. For the experiments reported here, the choice between approaches does not influence the conclusions drawn because the base rates remained constant across the experimental conditions. Consequently, subtracting base rate amounts to subtracting a constant and would not change the pattern of results, nor would the application of signal-detection theory.

The estimate of familiarity will also reflect the subjects' willingness to use familiarity in the recognition task. For example, subjects may be less willing to use familiarity as a basis for recognition judgements in some experimental conditions than in others. This could be represented as a bias factor or a response criterion. However, doing so would not influence the conclusions of the current experiments because, as already mentioned, the base rates remained constant across all experimental conditions.

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 modality in which a word was earlier presented, they should be able to either include or exclude that word, in line with instructions. For example, if recollection were perfect (R = 1), subjects would always respond "yes" when asked whether words that were earlier heard had been earlier heard (an inclusion test) and never respond "yes" when asked whether those words were earlier seen (an exclusion test). In contrast, familiarity is assumed not to support such selective responding. The contribution of familiarity as a basis for responding is the same on an exclusion test as on an inclusion test. That is, familiarity serves as a basis for responding "yes" regardless of whether that response is correct (an inclusion test) or an error (an exclusion test).

We estimated the contributions of recollection and familiarity to recognition of words from long and short lists. Half of the study lists were short (30 words); 15 of those words were seen and 15 were heard. The remaining lists were long (60 words). Again, half of those words were presented visually and half were presented auditorily. For one type of test, subjects were instructed to select seen words; for the other type of test, subjects were instructed to select heard words. Inclusion and exclusion scores on the short and long lists were used to calculate the contributions of recollection and familiarity to recognition performance for each of the two list lengths.

In the previous example, we focussed on calculating recollection and familiarity values for seen words. However, the same was done for heard words. In fact, the necessary inclusion and exclusion conditions were nested within the previously described test instructions. The instruction "select the heard words" served as exclusion instructions for the seen words and inclusion instructions for the heard words, whereas "select the seen words" served as inclusion instructions for the seen words and exclusion instructions for the heard words.

In Experiment 1, we defined recollection as the ability to remember study modality, and examined the influence of list length on recollection and familiarity. Experiment 2 was the same except that recollection was measured as the ability to remember in which of two study lists an item had been presented. For both experiments, it was expected that increasing list length would interfere with conscious recollection but leave familiarity unchanged. Response time distributions were used to examine the time course of the two bases for decisions. This was done by deriving estimates for recollection and familiarity in the same way as we did for overall performance except that responses were broken into bins in terms of response time so as to map out the contribution of the two bases for responding as a function of response speed. In Experiment 3, we further examined differences in the speed of the two bases for recognition by using a response-signal procedure similar to that used by Gronlund and Ratcliff (1989).

Experiment 1

METHOD

Subjects

Twenty-one subjects, who were students enrolled in an introductory psychology course at McMaster University, participated in the experiment.

Materials

Three sets of 240 words were randomly selected from the Toronto word pool. During study, each set was either presented visually, presented auditorily, or not presented at all. The sets were rotated through each condition such that each set appeared equally often in each condition. The study words were divided into eight short lists (30 words each) and 4 long lists (60 words each). The modality in which the words were presented was alternated from one word to the next. Thus half of the words in each list were presented visually and half were presented auditorily. All the study items were tested. Each test list contained an equal number of heard, seen and new words mixed in a random order.

Design and Procedure

Materials were presented and responses collected on a PC compatible computer with a monochrome monitor. The character size of the stimuli was approximately 5 × 5 mm and the viewing distance was approximately .5 m. Stimuli were presented in lower case letters in the centre of the screen.

Each subject was tested individually. At the beginning of the test session, subjects were informed that they would receive a number of recognition tests.

They were told that for each study list, half the words would be presented visually and half would be presented auditorily. Each list was followed by a recognition test for which subjects were instructed to select words that had been presented in one modality (e.g., heard) and avoid selecting those that had been presented in the other modality (e.g., seen). Subjects were also informed that the number of items in each study list would vary.

The study lists were presented at a rate of 1 word per second, which was approximately the time required to read a word aloud. The presentation modality was alternated from one item to the next, with the first item of each list always presented visually. Each visually presented word appeared on the screen for 1 s followed by 1 s of blank screen, during which time the experimenter read the next word aloud. Eight short lists (30 words) and 4 long lists (60 words) were presented in a random order.

A yes-no recognition test immediately followed each study list. All of the seen and heard words from the list, as well as an equal number of new words, were presented one at a time on the screen in a random order. For example, for the short lists the test list contained 15 seen words, 15 heard words, and 15 new words. For half the lists subjects were instructed to select the seen words and avoid selecting the heard words. The experimenter informed the subjects that they were to respond "yes", by pressing a designated key, only if they had seen the word in the prior list. If it was a new word, then they were to respond "no". Further, they were informed that words were never presented in both modalities and that if they remembered hearing an item they should respond "no", meaning that they had not seen the word. For the other half of the lists, subjects were given the converse instructions: Select the heard words and avoid selecting the seen words. The type of test instruction was randomized such that subjects did not know at the time of study which type of item they would be asked to select for at test.

List length (short vs long) was crossed with study modality (seen vs heard) which was crossed with type of test (seen? vs heard?). All were within-subject factors. Subjects' responses as well as response times were recorded. The significance level for all statistical tests was set at $p < .05$.

RESULTS AND DISCUSSION

An initial analysis revealed that study modality did not significantly influence performance so responses were collapsed over this variable. Because the number of items tested from each long list was greater than the number tested from each short list, only the first half of the items from each long test list was scored.

Response Accuracy

An analysis of variance was performed on the probabilities of responding "old" (Table 1) for items in long and short lists under inclusion instructions

TABLE 1
Proportion of Items Selected and Parameter Estimates for Long and Short Lists for Experiment 1

	List Length	
	Short	Long
Inclusion	.58	.50
Exclusion	.23	.25
New	.18	.20
Recollection	.35	.25
Familiarity	.35	.33

(heard words accepted as heard and seen words accepted as seen) and under exclusion instructions (heard words accepted as seen and seen words accepted as heard). The analysis revealed that list length interacted with test condition, F (1,20) = 25.04, MS_e = .005. For the inclusion condition, more items were selected in the short lists than in the long lists, F (1,20) = 22.21, MS_e = .003. For the exclusion conditions, an equal number of items were accepted in the long and short lists, F (1,20) = 2.11, MS_e = .002. Similarly, for new items, an equal number were accepted in the short and long lists, F (1,20) = 2.71, MS_e = .004.

Of main interest are the contributions of recollection and familiarity to recognition performance (see Table 1). Recollection values were calculated by subtracting the probability of selecting an item under exclusion instructions from the corresponding probability under inclusion instructions. Familiarity values were calculated by dividing the exclusion scores by one minus the estimated recollection scores. An analysis of variance was performed on the recollection values as well as the familiarity values, which were calculated for each subject.

There was a significant list length effect on the recollective component of recognition, F(1,20) = 25.64, MS_e = .009. Familiarity, on the other hand, was not significantly affected by list length, F (1,20) = 3.083, MS_e = .002. The familiarity of old items (.34) was significantly higher than the false alarm rate to new items (.19). For twenty of the twenty-one subjects, old items were more familiar than new items, showing that familiarity did significantly contribute to performance.

A conditional analysis was carried out equating for study-test lag (the number of items encountered between the presentation of the item in the study list and the presentation of the item at test), to determine if the list length effects were due to differences in the average delay between study and test. For the long lists, only items that were studied in the second half of the study list (study position > 30) and that were tested in the first half of the test (test position < 46) were included in the analysis. However, the estimates for

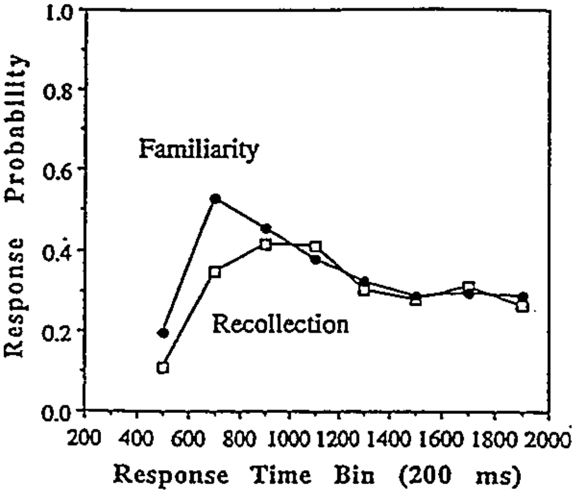

Fig. 1 Response latency distributions for familiarity and recollection for Experiment 1. Each point reflects estimates based on 200 ms wide response time bins.

recollection and familiarity did not differ from those of the previous analysis, suggesting that the list length effect was not due to differences in the number of items that intervened between study and test.

Response Latency

Responses in the inclusion and exclusion conditions were grouped into bins on the basis of response time. From those scores, estimates for recollection and familiarity were derived in the same manner as was done for the overall data. Each bin was 200 ms wide with the first bin beginning at 400 ms. Thus the first bin contained responses between 400 and 600 ms. All responses were used in the response time analysis. Responses were collapsed over subjects as well as study modality and list length because the response time distributions did not differ greatly from one condition to the next.

Figure 1 presents the estimates for recollection and familiarity as a function of response time. An examination of that figure showed that the contribution of familiarity was at its greatest between 600-800 ms, whereas recollection reached its peak between 800-1100 ms. Beyond these points the contribution of both of these processes decreased similarly. This pattern was seen for the majority of the subjects: For only 2 of the 21 subjects did recollection peak before familiarity. Thus familiarity had a speed advantage over recollection. However, recollection and familiarity contributed to recognition performance throughout the range of response times. This was true of the fastest as well as the slowest responses. This would not be expected if recollection was initiated only after an assessment of familiarity had failed, as was suggested by Atkinson and Juola (1974). Rather, it would seem that both processes are contributing to performance in parallel.

Experiment 2

Experiment 2 was designed to test the generality of the findings from Experiment 1. In Experiment 1, subjects were required to select for or against items previously presented in a particular modality. Thus recollection was measured as subjects' ability to respond selectively on the basis of the modality in which words were initially presented. In the current experiment, all items were presented visually but the study session consisted of two separate lists. Recollection was measured as the ability to determine list membership and to respond accordingly. As another test of the generality of the findings in Experiment 1, recognition was measured using a forced-choice procedure rather than a yes-no procedure.

METHOD

Subjects and Materials

The subjects were from the same pool and materials were the same as those used in the previous experiment. Twenty-four subjects participated in the experiment. For each subject 400 words were randomly partitioned into 8 short lists (10 words each) and 8 long lists (20 words each). The remaining words served as test lures.

Design and Procedure

Both the design and procedure of the current experiment were similar to those of the prior experiment with the following changes: At the beginning of the experimental session subjects were instructed to think of the meaning of each of the words that they studied as this would promote better recognition performance. The study phase began with the presentation of a cue identifying the list ("List # 1") which was followed by the first list of words presented one word at a time at a rate of one word per second. Immediately following this was the cue for the second list ("List # 2") followed by the second list of words. Half of the time the two lists were long, and half the time the two lists were short.

The presentation of study lists was followed by a two-alternative, forced-choice recognition test. A pair of words appeared on the screen and remained there until the subject made a response, at which time the next pair appeared. Subjects responded by pressing one of two designated keys on the keyboard, selecting either the right or the left word. For the first 10 test pairs (containing 5 words from List #1 and 5 from List #2, in a random order) of half of the test lists, subjects were instructed to select items from the first study list and avoid selecting items from the second study list. That is, they were told that if they could recollect an item as presented in List 1, they were to select the item; whereas if they could recollect an item as presented in List 2, they were to select the other item in a pair (which, unbeknownst to subjects, was always a new word). For the next 10 test pairs, subjects were

TABLE 2
Proportion of Items Selected and Parameter Estimates for Long and Short Lists for Experiment 2

	List Length	
	Short	Long
Inclusion	.77	.72
Exclusion	.39	.45
Recollection	.38	.27
Familiarity	.61	.61
	(.63)	(.62)

Note. Parameter estimates in parentheses were calculated using the overall means rather than the subject averages.

instructed to select words from the second list and avoid selecting words from the first list. The order of test instructions was reversed for the other half of the test lists.

The test order was randomized such that subjects did not know which list they would be asked to select for first. This procedure was repeated 8 times, 4 times with short lists and 4 times with long lists in a random order. Unlike the previous experiment, an equal number of items were tested in long and short lists. List length (short vs long) was crossed with study list (List 1 vs List 2) which was crossed with type of test (List 1? vs List 2?). All were within-subject factors.

RESULTS AND DISCUSSION

Response Accuracy

An initial analysis showed that list number (list 1 vs list 2) did not significantly influence performance so responses were collapsed over this variable. Table 2 presents the probabilities of responding "old" for items in long and short lists under inclusion instructions (list 1 items accepted as from list 1 and list 2 items accepted as from list 2) and under exclusion instructions (list 1 items accepted as from list 2 and list 2 items accepted as from list 1). An analysis of variance revealed that list length interacted with test condition, F (1,23) = 13.55, MS_e = .005. For the inclusion condition, slightly more items were selected in the short lists than in the long lists, however, the effect did not reach significance, F (1,23) = 3.90, MS_e = .006. For the exclusion conditions, fewer items were selected in the short lists than in the long lists, F (1,23) = 6.64, MS_e = .004.

Recollection and familiarity values were calculated in the same manner as in the previous experiment. An analysis of variance was performed on the recollection values as well as on the familiarity values. As in the previous experiment, increasing list length interfered with recollection but left

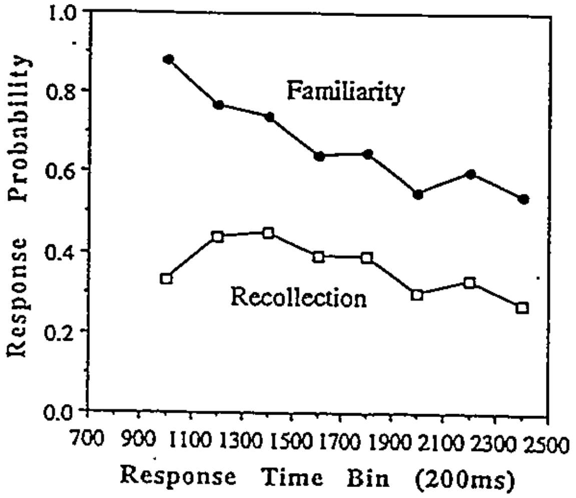

Fig. 2 Response latency distributions for familiarity and recollection for Experiment 2. Each point reflects estimates based on 200 ms wide response time bins.

familiarity in place. Recollection was greater in the short lists than in the long lists, F (1,23) = 13.41, $MS_e = .010$. In contrast, the estimates for familiarity were equal in short and long lists, $F < 1$. As in the previous experiment, familiarity contributed to the recognition of old items. For 21 of the 24 subjects, the estimated probability of selecting an item on the basis of familiarity was above chance (.61 compared to .50).

Response Latency

Response time data were used to examine the speed of the two processes. Beginning at 900 ms, responses in the inclusion and exclusion conditions were grouped into 200 ms wide bins. From those scores, estimates for recollection and familiarity were derived as was done for the overall scores. Responses were collapsed over subjects as well as study list and list length because the response time distributions did not differ greatly between conditions.

Figure 2 presents the estimates for recollection and familiarity as a function of response time. The functions were similar to those of the previous experiment. Familiarity produced its greatest contribution to performance very early – having peaked in the first response bin (900-1100 ms). Recollection, on the other hand, did not peak until 1300-1500 ms after stimulus presentation. This supports the claim that familiarity is the faster of the two processes. Furthermore, both processes contributed to the fastest as well as the slowest responses, suggesting that the two processes were operating in parallel.

The functions in the current forced-choice experiment are "flattened out" and shifted to the right compared to those of the previous yes-no experiment. However, this can be understood as a product of the different demand characteristics of the two types of test procedure. Because the forced-choice procedure required subjects to make a decision about two items, compared to the one item in the yes-no procedure, it is understandable that responses were slowed. Another difference between the two sets of functions was that the familiarity function in the forced-choice experiment was considerably elevated above the recollection function, unlike in the first experiment. This can be understood as reflecting the higher base rate in the forced-choice procedure (i.e., chance performance is .50 compared to the base rate in Experiment 1 of .19).

In both previous experiments, we found that increasing list length interfered with recollection but left familiarity unchanged. Moreover, examination of reaction time distributions revealed that familiarity had a speed advantage over recollection. However, these results would seem to conflict with those of Gillund and Shiffrin (1984) who found that list length did not interact with test speed. If recollection alone is responsible for the list length effect and it is the slower of the two processes, then one might have expected Gillund and Shiffrin to find a larger list length effect for the slower recognition condition. One possible explanation is that the two processes do not differ in retrieval speed but that, under unspeeded conditions, such as in our experiments, subjects were just slower to use recollection than familiarity. Another, more plausible, explanation is that the difference in retrieval speed is not large enough to produce the interactions that Gillund and Shiffrin sought. One should note that the dual process model of Atkinson and Juola (1974) proposed that recollection was only initiated if an assessment of familiarity failed to support a fast response, thus the fastest responses should be based almost entirely upon familiarity whereas the slower responses should be based almost entirely upon recollection. Given such a model, one would expect list length to interact with test speed. However, if as our results suggest, the two processes are not sequential but operate in parallel, the interaction may be considerably reduced. A final experiment was conducted to address this issue.

Experiment 3

In Experiment 3 we examined the effect of list length on familiarity and recollection using a response signal procedure. In a design similar to that of the previous experiments, subjects were tested with a yes-no recognition procedure in which they were required to determine list membership (list 1 vs list 2). However, in this experiment, each test item was followed by a signal to respond. The response signal was presented either 600 ms (fast) or 1600 ms (slow) after the presentation of the test word. Subjects were encouraged to respond as soon as they heard the signal. If the two processes did not differ in terms of processing speed, then the contributions of both processes should increase equally with retrieval time. If, however, familiarity is the faster of the two, then it should be less affected by the response-signal manipulation than recollection.

METHOD

Subjects and Materials

Sixteen subjects from the same pool as the previous experiments participated in the current experiment. Words (598) were randomly selected form the Toronto word pool and were partitioned into 4 pairs of short lists (16 words in each list) and 4 pairs of long lists (32 words in each list) for each subject. The remaining words served as practice items and test lures.

Design and Procedure

The design and procedure of the current experiment were similar to those of the prior experiments. As in Experiment 2, subjects studied pairs of lists: two short lists (16 words in each list) or two long lists (32 words in each list). Each word was presented for 2 s, and there was a 3 s delay between lists. As in Experiment 1, subjects were tested using a yes-no recognition procedure. For half of the tests, subjects were required to respond "yes" if the word was from List #1 and to respond "no" if it was from List #2 or if it was a new item. For the other half of the tests, the instructions were reversed. Subjects were instructed to respond "yes" if they thought the word was presented but could not recollect which list it was in. Subjects completed nine study-test blocks, the first of which was used as a practice phase to familiarize the subjects with the procedure. Responses from this test were not collected. Furthermore, the first 2 items of each test list were also treated as practice items – they were always new items. Apart from the two practice items, each test list included 16 items from list 1, 16 items from list 2, and 16 new items mixed in a random order.

The critical difference in the current experiment was that subjects heard a signal to respond either 600 ms (fast) or 1600 ms (slow) after each test item appeared. If they responded before the signal or later than 300 ms after the signal, an error tone would sound and an error message ("too slow" or "too fast") would appear on the screen. The delay was randomized such that the subjects did not know until they heard the signal how long they would have to respond on any item. List length (short vs long) was crossed with signal speed (fast vs slow) which was crossed with type of test (inclusion vs exclusion). Because list number (list 1 vs list 2) did not influence performance in the previous experiment, this factor was not analyzed. However, it was fully counterbalanced across all experimental conditions. All factors were varied within-subject.

RESULTS AND DISCUSSION

The results of Experiment 3 are presented in Table 3. The values reported represent average performance on all tested items. An initial analysis based only on the responses that fell within the response window (after the response signal and no later than 300 ms after the signal) was conducted but produced

TABLE 3
Proportion of Items Selected and Parameter Estimates for Long and Short Lists with Fast and Slow Response Signals for Experiment 3

	List Length			
	Short		Long	
Response Speed	Fast	Slow	Fast	Slow
Inclusion	.70	.77	.66	.72
Exclusion	.36	.34	.40	.38
New	.14	.15	.16	.15
Recollection	.34	.43	.26	.34
Familiarity	.55	.58	.53	.57
	(.55)	(.60)	(.54)	(.58)

Note. Parameter estimates in parentheses were calculated using the overall means rather than the subject averages.

the same pattern of results as the overall data. An analysis of the probabilities of responding "old" revealed that the instructions (inclusion vs exclusion) interacted with list length, F $(1,15) = 6.662$, $MS_e = .008$, and signal speed, F $(1,15) = 7.986$, $MS_e = .007$. Inclusion scores increased from long to short lists and from fast to slow response speeds. Exclusion scores showed the opposite effects, decreasing from long to short lists and from fast to slow signal speeds. The only other significant effect was that of instructions, F $(1,15) = 75.551$, $MS_e = .049$. More items were correctly called old in the inclusion condition than were incorrectly called old in the exclusion condition. The probability of incorrectly accepting a new item was .15 and was not influence by condition (all effects $F < 1$).

A separate analysis was conducted on the inclusion scores to determine if signal speed interacted with list length. There was no such effect. Although there was an effect of list length, F $(1,15) = 7.725$, $MS_e = .004$, and of signal speed, F $(1,15) = 14.022$, $MS_e = .005$, the interaction was nonsignificant, $F < 1$. This is in agreement with Gillund and Shiffrin (1984) who found that list length did not interact with response speed. However, does this mean that the two processes did not differ in terms of processing speed?

An examination of the estimates for recollection and familiarity showed that despite the lack of an interaction between list length and response speed for the inclusion condition, familiarity was faster than recollection. An analysis of the estimated recollection and familiarity values revealed that recollection increased significantly from fast to slow response conditions, F $(1,15) = 8.048$, $MS_e = .015$, but that familiarity did not, F $(1,15) = 1.988$, $MS_e = .009$. Recollection increased from .30 to .39 but familiarity showed only a modest increase from .54 to .58. This pattern suggests that the process of recollection was still largely unfinished at 600-900 ms, compared to familiarity which did not show a sizable increase after that time. These results are in agreement with a number of previous response-signal studies in finding that familiarity is the faster of the two processes (Dosher, 1984; Gronlund & Ratcliff, 1989). Furthermore, the same pattern of results was reported by Toth (1993) who also used a response signal procedure in conjunction with the process dissociation procedure.

The effects of list length replicate those of the previous two experiments. Increasing list length interfered with recollection, $F(1,15) = 6.649$, $MS_e = .015$, but left familiarity unaffected, $F < 1$. As in the previous experiments, old items were more familiar (.56) than were new items (.15). For 15 out of 16 subjects the estimate for familiarity was greater than the false alarm rate to new items. There were no other significant effects (all Fs < 1).

GENERAL DISCUSSION

The results of the current experiments provide evidence that increasing list length affects only the recollective processes in recognition memory, leaving familiarity largely unaffected. This was found with yes-no and forced-choice recognition test procedures, for fast and slow recognition judgements. Furthermore, it was found when recollection was measured as the ability to selectively respond on the basis of study modality as well as when recollection was measured as the ability to respond on the basis of list membership. Although the list length effect has been found a number of times in recognition memory performance (e.g., Ratcliff & Murdock, 1976; Atkinson & Shiffrin, 1971; Gillund & Shiffrin, 1984), the differential effect on the processes underlying recognition performance has not previously been demonstrated.

An examination of the response-time distributions of Experiments 1 and 2 suggested that familiarity held a speed advantage over recollection. Results gained in Experiment 3 using the response-signal procedure showed that only recollection was significantly reduced by requiring subjects to respond quickly, providing further support for the claim that familiarity is the faster of the two processes (Atkinson & Juola, 1974; Mandler, 1980). Moreover, in Experiments 1 and 2, both recollection and familiarity contributed to the fastest and the slowest responses, as would be expected if the two bases for responding operate in parallel (cf. Atkinson & Juola, 1974).

An important assumption that underlies the process-dissociation procedure is that recollection and familiarity serve as *independent* bases for judgements. If the two processes are independent then it should be possible to find variables that influence one process while leaving the other invariant. Several such variables have been identified. For example, dividing attention at time of test reduces recollection but leaves familiarity in place (Jacoby, 1991), as does aging (Jennings & Jacoby, 1993), and amnesia (Verfaellie & Treadwell, 1993). List length interference and response speed add to the list of variables that produce such dissociations.

A potential criticism of the procedure used in the current experiments is the possibility of partial recollection. It is likely that subjects occasionally recollect information about a studied word that does not support the source discriminations that we required. For example, the subject may remember that they coughed as the word was studied but this information would not allow the subject to determine which list the item was in. Such partial recollection would not be captured by our estimate of recollection and might contaminate estimates of familiarity. That is, if these items were treated as familiar then our estimate of familiarity would reflect familiarity plus some partial recollection. If this occurred, our estimates of familiarity would begin to look like our estimates of recollection; manipulations like list length should affect the estimates for both processes in a similar manner. However, examination of the data in the current experiments as well as those previously mentioned shows little evidence of such contamination. One possibility is that partial recollection is relatively rare in comparison to the measured recollection. In fact, when subjects are instructed to remember study modality, and are repeatedly tested for just that, it would not be surprising if their ability to recollect task-irrelevant information was quite poor.

Do the results of our experiments generalize to performance on standard recognition tests? In our experiments, subjects could not rely solely on familiarity because they were required to differentially respond dependent on the study modality or the list in which an item had been presented. In traditional recognition experiments, on the other hand, subjects only have to discriminate between old and new items, a task that might be based on assessments of familiarity alone. However, there are a number of reasons to believe that under standard recognition conditions subjects rely on familiarity as well as recollection. First, it is likely that subjects would make use of recollection if they have this basis for responding available to them. Second, if subjects relied solely on familiarity in the standard recognition task, but relied on familiarity and recollection in the current discrimination task, then we would expect to see systematic differences in overall performance of the two types of task, which we do not see. In the current experiments, the magnitude of the list-length effect under inclusion instructions was similar to that found in standard list-length experiments (see Ratcliff & Murdock, 1976). Also, the response time manipulation had effects on performance in the inclusion test condition that were the same as those found for standard recognition by Gillund and Shiffrin (1984). Finally, if performance in the standard recognition experiments was based solely on familiarity, and, as we have shown, familiarity is not affected by list length, then we would not expect to see list length effects in standard recognition tests. However, such effects are consistently found (e.g., Atkinson & Shiffrin, 1971; Gillund & Shiffrin, 1984; Ratcliff & Murdock, 1976; Strong, 1912). Although not logically required to make simple recognition judgements, it seems that recollection serves as a basis for judgements in standard recognition tasks as well as in the tasks used in our experiments.

The finding that recollection and familiarity are differentially affected by list length interference as well as by retrieval speed joins a growing body of literature reporting dissociations within recognition memory. We believe that these results argue against single factor models such as signal detection theory and provide strong support for a dual process view of recognition memory.

This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada. The authors would like to thank Ian Begg, Lee Brooks, William Hockley and Bennet Murdock for their thoughtful suggestions on an earlier version of this paper, as well as Scott Gronlund, Jonathan Schooler and Colin MacLeod for their critical and helpful comments. Correspondence should be sent to Andrew Yonelinas, Department of Psychology, McMaster University, Hamilton, Ontario, L8S 4K1 (E-mail: yonelina@mcmaster.ca).

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, & thinking* (pp. 242-293). San Francisco: W.H. Freeman.

Atkinson, R.C., & Shiffrin, R.M. (1971). The control of short-term memory. *Scientific American, 225*, 82-90.

Dosher, B.A. (1984). Discriminating preexperimental (semantic) from learned (episodic) associations: A speed-accuracy study. *Cognitive Psychology, 16*, 519-555.

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

Gronlund, S.D., & Ratcliff, R. (1989). Time course of item and associative information: Implications for global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 846-858.

Humphreys, M.S., & Bain, J.D. (1983). Recognition memory: A cue and information analysis. *Memory and Cognition, 11*, 583-600.

Jacoby, L.L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology, 3*, 306-340.

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

Jennings, J., & Jacoby, L.L. (1993). Automatic versus intentional uses of memory: Aging, attentio, and control. *Psychology and Aging, 8*, 283-293.

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. (1960). The immediate retention of unrelated words. *Journal of Experimental Psychology, 60*, 222-234.

Piercy, M. & Huppert, F.A. (1972). Efficient recognition of pictures in organic amnesia. *Nature, 240*, 564.

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

Roberts, W.A. (1972). Free recall of word lists varying in length and rate of presentation: A test of total-time hypotheses. *Journal of Experimental Psychology, 92*, 365-372.

Sloman, S.A., Hayman, G., Ohta, N., Law, J., & Tulving, E. (1988). Forgetting in primed fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 223-239.

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

Strong, E.K., Jr. (1913). The effect of time-interval upon recognition memory. *Psychological Review, 20*, 339-372.

Swets, J.A. (1964). Signal detection and recognition by human observers. New York: John Wiley and Sons.

Toth, J.P. (1993). *Deadlines, levels, and automatic processes in recognition memory*. Manuscript in preparation.

Verfaellie, M., & Treadwell, J.R. (1993). The status of recognition memory in amnesia. *Neuropychology, 1*, 5-13.

Wickelgren, W.A. (1972). Trace resistance and the decay of long-term memory. *Journal of Mathematical Psychology, 87*, 149-162.

Date of acceptance: November 22, 1993

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

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 $\approx$ 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.

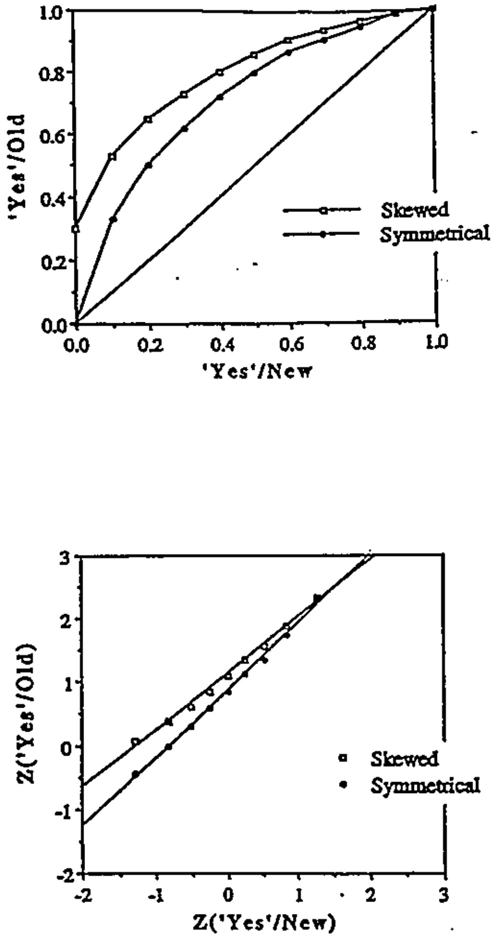

Figure 1. Symmetrical and skewed receiver-operating characteristics plotted on probability coordinates as well as z-coordinates.

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

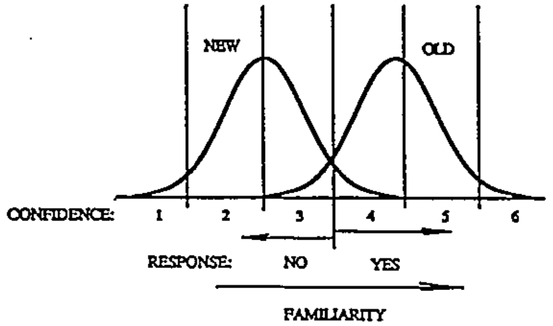

Figure 2. Familiarity distributions representing old and new items in the classical signal detection model.

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:

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

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,

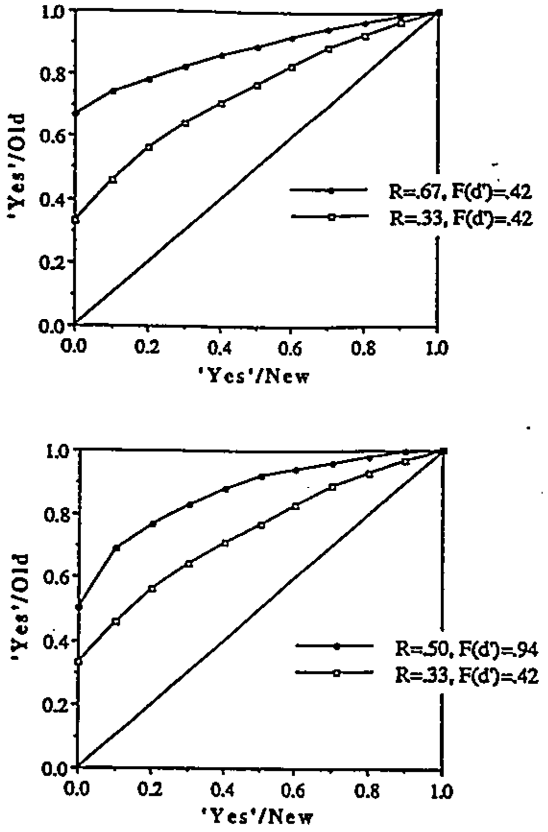

Figure 3. Hypothetical receiver-operating characteristics generated by the dual-process signal detection model. The top portion represents performance in two conditions that differ only in the probability of recollection. The bottom portion represents performance in two conditions that differ both in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

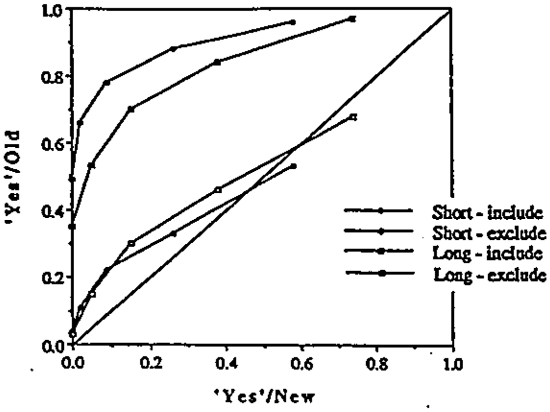

Figure 4. Receiver-operating characteristics for the inclusion and exclusion conditions for short and long study lists in Experiment 1.

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.

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.

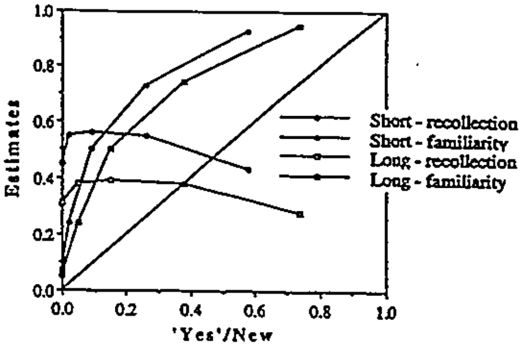

Figure 5. Estimates for recollection and familiarity for long and short lists as a function of response confidence in Experiment 1.

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

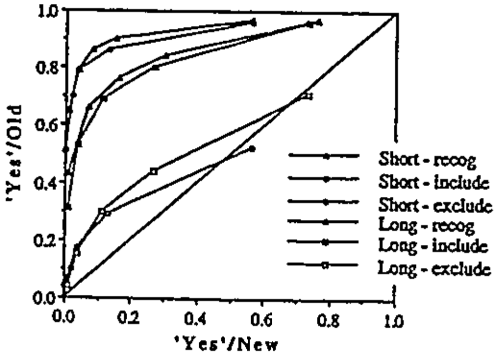

Figure 6. Receiver-operating characteristics for the inclusion, exclusion, and recognition (recog) conditions for short and long study lists in Experiment 2.

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

	Short list		Long list	
ROC measure	*M*	*SD*	*M*	*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.

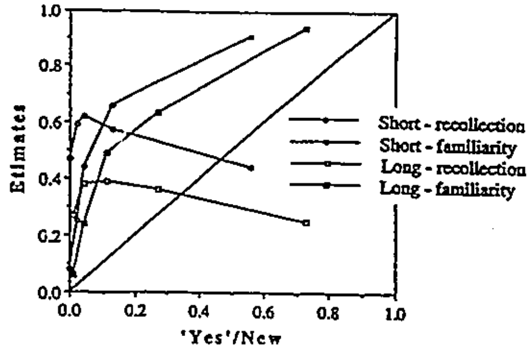

Figure 7. Estimates for recollection and familiarity for long and short lists as a function of response confidence in Experiment 2.

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

Condition/parameter	Weak *M*	Weak *SD*	Strong *M*	Strong *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

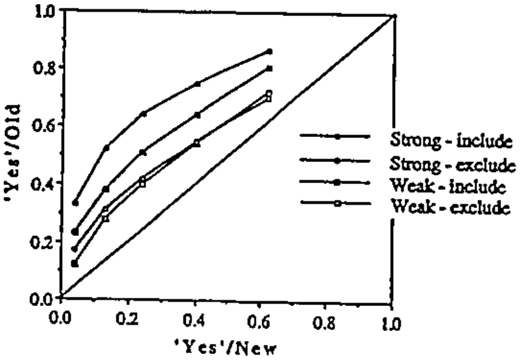

Figure 8. Receiver-operating characteristics for the inclusion and exclusion conditions for weak and strong items in Experiment 3.

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.

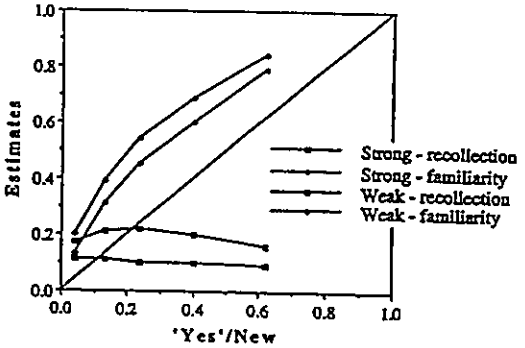

Figure 9. Estimates for recollection and familiarity for weak and strong items as a function of response confidence in Experiment 3.

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 [illegible] 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 ■

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, & thinking (pp. 242-293). San Francisco: W. H. Freeman.

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. Scientific American, 225, 82-90.

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

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., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. Journal of Experimental Psychology: General, 3, 306-340.

Jennings, J., & Jacoby, L. L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. Psychology and Aging, 8, 283-293.

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.

Roberts, W. A. (1972). Free recall of word lists varying in length and rate of presentation: A test of total-time hypotheses. *Journal of Experimental Psychology*, *92*, 365-372.

Verfaellie, M., & Treadwell, J. R. (1993). The status of recognition memory in amnesia. *Neuropychology*, *1*, 5-13.

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

Yonelinas, A. P., & Jacoby, L. L. (1994). Dissociations of processes in recognition memory: Effects of interference and of response speed. *Canadian Journal of Experimental Psychology*, 48:4.