To appear in the *Psychology of Learning and Motivation, Vol. 48, Memory Use as Skilled Cognition*, B. Ross and A. Benjamin (Eds.)

Towards an Understanding of Individual Differences in Episodic Memory: Modeling the Dynamics of Recognition Memory

Kenneth J. Malmberg
University of South Florida

For Correspondence:

Kenneth J. Malmberg
malmberg@cas.usf.edu or 813-974-7641

Outline

I. A Possible Relationship between the Speed-Accuracy Tradeoff and Individual Differences in Associative Recognition

II. Traditional Recognition Testing Procedures

III. Classical Models of Associative Recognition

- A. The Compound Cue Model (e.g., Gronlund & Ratcliff, 1989)
- B. Independent Cue Models (e.g., Murdock, 1982)
- C. Dual-Process Models (e.g., Xu & Malmberg, in press)

IV. Modeling the Accuracy and Latency of Associative Recognition.

- A. Encoding of Associative Information
- B. Familiarity-Based Retrieval
  1. Sampling with Replacement
  2. Sampling without Replacement
- C. Decision
- D. Familiarity-Based Performance
  1. Signal to Respond
  2. Free Response
- E. Recollection
- F. Dual-Process Decision Strategies
  1. Signal to Respond Performance
  2. Free Response Performance

V. Conclusions

A Possible Relationship between the Speed-Accuracy Tradeoff and Individual Differences in Associative Recognition

Viewing the use of memory as skilled cognition emphasizes the interaction between the nature of the task, the nature of memory, and prior knowledge. The major implication is that some people can be better rememberers than others based on their skill as rememberers or other strategic factors; memory performance can change based on differences in one's goals and in one's meta-level knowledge of how to achieve one's mnemonic goals, even when memory is intact and the nature of the memory task is static. While individual differences are usually treated as random factors and conveniently washed away in the averaging process, differences in the strategic use of memory almost certainly characterize different populations and discriminate among individuals of a given population. In some cases, the systematic differences in goals and prior knowledge are likely to be the major source of variance in memory performance.

Understanding individual differences in memory is one of the most important goals for future memory research (cf. Lewandowsky & Heit, 2006). Over the past 125 or so years, memory research has focused on explaining the central tendencies in our observations, largely ignoring the variability in our observations. This approach to memory research served Ebbinghaus (he served as his only subject) and others well, but it puts severe limitations on the scope and usefulness of our understanding of memory. A long-term goal should be to account for different sources of variance in our observations, including individual differences owing to structural and strategic factors. The short-term goal of this chapter is to identify different potential sources of variability in recognition memory performance and to show how they differentially impact performance.

Are individual differences due to random error or systematic factors that influence the manner in which associative recognition is performed? While most would accept the proposition that both random and systematic factors influence performance, little is known about the extent to which each is a contributor or how to discriminate between these sources of variability in memory performance. There are also likely to be different types of systematic influences on variability, and it is particularly important to consider that individuals might use different strategies to perform the same task in the same nominal situations. This could occur because there are differences in one's goals, in the state of memory, and/or because subjects are more or less able to select an adaptive strategy for the task or situation.

If some individual differences are due to systematic factors, are they due to differences in the structural aspects of memory or the strategies or goals adopted? Associative recognition, for instance, requires the discrimination of pairs that were studied together (i.e., *targets* or intact pairs) from pairs that were not studied together (*foils* or rearranged pairs), and younger adults perform associative recognition better on average than older adults (e.g., Light, Patterson, Chung, & Healy, 2004). Even within a given population, say a younger-adult or older-adult population, some individuals are better at associative recognition than others. Should the differences in performance found in younger versus old populations be attributed to the same factors that characterize the differences in performance within either of these populations? The answer to this question might affect how we view the memory deficits of those within a given population, and thus discriminating the impact of different sources of variability should be a clinically important goal of future memory research. (The performance of

other memory tasks are also likely to be systematically affected by individual differences, but the focus of this chapter will be associative recognition.)

One way to assess the impact of structural changes in memory versus the effectiveness of different memory strategies is to implement changes in memory and different strategies in a formal model and observe its behavior. Indeed, I suppose that it would be difficult to discriminate among different sources of variability outside the framework of a formal model. For instance, the patterns of performance observed in the models can be compared to the performance of different individuals or groups. To the extent that different models (or sets of parameters) are required to describe the performance of the individuals, one might be able to attribute the differences in performance to different strategies that were adopted versus differences in the structural aspects of memory. In subsequent sections, I will attempt to use a formal model to tease apart these sources of variance or individual differences.

Because associative recognition can be performed in multiple ways and different populations show deficits in performance, it is an ideal task to explore as a means for understanding the various sources of individual differences in episodic memory performance. Such an endeavor, of course, requires a framework that acknowledges that memory performance can be affected by strategic factors. While most traditional models of discrimination can describe the effects of a variety factors that influence memory, the flexibility in how a given task is performed is limited by their scope. Since these models often do not make assumptions about how memories are encoded, represented, or retrieved, their flexibility is limited to adopting different decision criteria or thresholds in response to the statistical properties of the environment (e.g., Green & Swets, 1966;

Macmillan & Creelman, 1990; Rotello & MacMillan, this volume; Rotello, Macmillan, & Reeder, 2004).

Some have argued that such impoverished models can explain recognition performance in all its varieties (e.g., Dunn, 2004; Rotello et al., 2004). Growing evidence suggests, however, that recognition does not necessarily involve only a simple discrimination between events that occurred and events that did not occur (e.g., Malmberg & Xu, in press; Van Zandt & Maldonado-Molina, 2004). Rather, the manner in which recognition tasks are performed also depends on the similarity between the targets and the foils (Malmberg, Holden, Shiffrin, 2004; Malmberg, Zeelenberg, & Shiffrin, 2004; Malmberg & Xu; in press), task demands (Jacoby, 1991; Malmberg & Xu, 2006), available retrieval cues (Criss & Shiffrin, 2005), the composition of the study lists (Hockley & Newandomski, in press), and the expertise one has in processing different stimuli (Xu & Malmberg, in press). Other findings suggest that the basis for yes-no recognition is different from that used to make a discrimination based on confidence ratings (Malmberg & Xu, in press; Van Zandt & Maldonado-Molina, 2004; also see Baranski & Petrusic, 2001), that yes-no recognition is based on different information than judgments of frequency (Curran, Cleary, & Greene, 2001; Hintzman, Curran, & Oppy, 1992; Malmberg, Holden, & Shiffrin, 2004), and that confidence ratings and judgments of frequency are based on different information (Hintzman, 2005).

Accounting for the aforementioned dissociations is difficult without some degree of model flexibility (cf. Malmberg, Holden, & Shiffrin, 2004), and several newer models have been devised as a result (Criss & Shiffrin, 2005; Kelley & Wixted, 2001; Malmberg, Holden, & Shiffrin, 2004; Reder et al., 2000). One important advance being made is that

these newer models are more comprehensive in their description of memory, and additional strategies emerge from them as the representations and processes involved in recognition memory are fleshed out. Nevertheless, a potential problem arises when additional assumptions are made in order to account for the performance of different tasks and situations. The danger of proposing different models for different tasks and situations is that the theory might become rather *post hoc* in its organization of the data.

To mitigate this danger, we have proposed that there are different ways to perform recognition tasks, and one selects (or at least should select) a recognition strategy that will achieve a subjectively determined level of accuracy in the shortest amount of time (Malmberg & Xu, in press). Such a strategy is said to be *efficient* with respect to the subject's goals, and the construct of efficiency organizes the different models under a single theoretical framework.

The efficiency hypothesis assumes that subjects have meta-level and/or perhaps implicit knowledge of how different tasks can be performed and how the processes involved in performing these tasks are constrained by time pressure and other situational factors. The ability of a subject to efficiently perform a recognition task depends on the ability to adapt to the current situation by adopting an appropriate strategy. This will depend on the quality of the subject's meta-level knowledge of his memory, the task, and the situation. In some cases, the subjects may perform at a suboptimal level of accuracy because speed is deemed to be more important or vice versa (Malmberg & Xu, in press; Van Zandt & Maldonado-Molina, 2004). In these situations, performance is suboptimal with respect to the level of accuracy that can potentially be achieved, but performance would nevertheless be considered efficient with respect to the subjective goal.

The tradeoff between speed and accuracy is well documented in the literature of cognitive psychology. However, the extent to which this relationship is used to affect how learning and memory tasks are performed is only now beginning to be understood. Nelson and Narens (1990; also Nelson & Leonesio, 1990) proposed that subjects adopt a standard to which they seek to learn new material, and they allocate various amounts of time to learning the different elements of the material in order to achieve their standard. For instance, the amount of self-paced study time allocated to a given item is correlated with a metacognitive ease-of-learning judgment, but the fact that subjects tend to underestimate the amount of learning needed to achieve a desired level of performance is referred to as the "labor-in-vain" effect.

What is less understood is the extent to which the speed-accuracy tradeoff is taken into account when performing various memory tasks. Recent associative recognition memory experiments suggest that subjects select a strategy that they determine allows them to achieve a desired level of accuracy in shortest time possible. For instance, Malmberg and Xu (in press) varied the composition of the lists used to test memory, such that in some cases pairs formed from unstudied items were also tested (i.e., *XY pairs*). Since the average similarity between the targets and foils is lower, the discrimination of XY pairs and rearranged pairs from intact pairs is relatively easy compared to the discrimination of only rearranged from intact pairs. In these experiments, the overall accuracy was similar across conditions (Malmberg & Xu, in press). That is, the probability of endorsing rearranged pairs (*hit rate*) was similar, and of course the probability of incorrectly endorsing foils (*false-alarm rate*) was lower for XY pairs than for rearranged pairs. Critically, the false-alarm rate for rearranged pairs was greater

when XY pairs were tested than when XY pairs were not tested. The false-alarm rates were statistically similar across testing conditions, even though the overall time it took to perform the task was much less on average in the condition were XY pairs were tested.

Thus, the ability to discriminate intact from rearranged pairs was dependent on testing conditions that could not be attributable to factors such as bias, interference, attentional load, stress, or delay. Such findings suggested to us that perhaps metacognitive judgments are made prior to and/or during the retrieval of information from memory in order to maximize a subjective speed-accuracy tradeoff (Malmberg & Xu, in press). Understanding the impact on recognition performance of implementing different strategies can help us to understand the effect of different sources of variability in recognition performance and differentiate differences in strategy selection from differences in structural changes in memory.

It is quite possible that the appropriate selection of a strategy that meets one's goals vis-à-vis the speed-accuracy tradeoff is a major source of individual differences. In order to achieve the level of specificity necessary to differentiate strategic versus structural sources, it is necessary to understand how these factors interact with the dynamics of recognition memory. Most memory models have focused on accounting for the accuracy of recognition and not the latency of recognition, and as we will see the latency with which one performs a task places constraints on the models. Other models are used to investigate speed-accuracy tradeoffs without reference to any explicit assumptions about the nature of memory (e.g., Dosher, 1984; Gronlund & Ratcliff, 1989). More rarely, models have focused on the latency of recognition memory under conditions where accuracy was nearly perfect (Atkinson & Juola, 1974). Models that have

addressed both the speed and accuracy of recognition memory have been subsequently disconfirmed (e.g., Diller et al., 2001; Xu & Malmberg, in press).

In this chapter, I will describe several new models of the accuracy and latency of associative recognition performance. Although I will discuss some important findings that constrain model development, quantitative model selection is not the primary goal. Nor is the immediate goal to identify individuals who have different tendencies to emphasize speed versus accuracy. Rather, the present purpose is to develop a better understanding of the potential sources of individual differences by formally describing the speed-accuracy tradeoff within a recognition memory framework and how the models' behavior are influenced by various structural versus strategic factors.

The models that I will describe are illustrative of many that have been verbally described in the past or implemented without reference to the nature of memory (e.g., Yonelinas, 1997), which is usually described as how information is encoded, represented, and retrieved. There are, however, a number of issues that need to be addressed in order to implement them, and their complexity does not readily lend to understanding based on intuition or casual reasoning.

When working with relatively complex models, it is rarely the case that the initial set of assumptions is that which is ultimately adopted and usually only the final outcome of a modeling endeavor is reported. However, there is much to be gained by observing the modeling process, as it is perhaps more important to understand what implementations do not work than to identify an implementation that is sufficient. For instance, we might expect based on intuition that a particular shift in criteria will produce a speed-accuracy tradeoff only to find out that after it is implemented that our intuitions

were incorrect. Thus, the purpose of this chapter is to identify issues that are important for understanding the dynamics of associative recognition and how structural versus strategic influences affect performance, and hence how individual differences might map onto these sources of variability. Before doing so, I will briefly describe traditional recognition memory procedures and several classical models of associative recognition in the context of findings that constrain them.

Traditional Testing Procedures

In single-item recognition experiments, subjects discriminate items that were studied from items that were not studied. Usually, items are randomly selected from a large corpus and assigned to either target or foil conditions for each subject. Thus, the targets and foils are only randomly similar to each other.

In comparison, pairs of items are studied in most associative recognition experiments (*A-B*, *C-D*, *E-F*, etc.). Traditionally, the pairs consist of words, but there is growing and increasingly important literature of the associative recognition of nonverbal and novel stimuli (e.g., Criss & Shiffrin, 2006; Curran et al., 2006; Greene, 1996, Hockley, 1994, Xu & Malmberg, in press). At test, associative recognition requires the subject to discriminate between intact (e.g., *A-B*, *C-D*) and rearranged pairs (e.g., *A-D*, *C-B*). Thus, the discrimination of intact from rearranged pairs involves discriminating between pairs that are similar to each other.

There are two traditional procedures for assessing the relationship between accuracy and latency. The *signal-to-respond* procedure forces subjects to report their recognition judgment within a narrow window subsequent to the presentation of the test stimulus (Diller et al., 2001; Dosher, 1984; Gronlund & Ratcliff, 1989; Light et al.,

2006). The *free-response* procedure allows subjects to respond when they choose. The advantage of the signal-to-respond procedure is that it allows the researcher to assess the relationship of speed and accuracy at different points during retrieval. The disadvantage of the signal-to-respond procedure is that the severe constraints placed on the report might influence how subjects make recognition judgments.

Classical Models of Associative Recognition

Models are often characterized by the information used to make a recognition decision. Some models assume that recognition is based on a continuous random variable often conceptualized as familiarity. The *compound cue model* (Gronlund & Ratcliff ,1989; also see Gillund & Shiffrin, 1984; Hinztman, 1986; Murdock, 1982) attempts to account for the time course of recognition accuracy derived from the signal-to-respond procedure by positing a simple random walk or diffusion process (e.g., Ratcliff, 1978). Accordingly, associative recognition involves a comparison of different types of cues with the contents of memory: concurrent cues and compound cues. Concurrent cues represent the individual items of a test stimulus, and they provide a measure of the familiarities of the individual test items when used as memory probes. A compound cue is comprised of the two test items, and when used as a probe it provides a measure of how familiar the test pair is.

Item familiarity tends to provide positive evidence for both intact and rearranged pairs, whereas associative evidence tends to provide positive evidence when an intact pair is tested and negative evidence when a rearrange pair is tested. According to the continuous-time version of the compound-cue model (i.e., a diffusion model), positive and negative discrete evidence (i.e., familiarity) is accumulated subsequent to the

memory probe starting at point Z at a mean rate of u with a variance of s^2. Positive endorsements are made when the accumulated familiarity associated with the cue is greater than a subjective "old" criterion and negative endorsements are made when it is less than a subjective "new" criterion.

One can think of the random walk model as dynamic version of signal detection models (e.g., Green & Swets, 1966; Ratcliff, 1978), and they account for the latencies of a wide variety of binary decisions (Ratcliff & Smith, 2004). For instance, the random walk model provides a natural explanation for the speed-accuracy tradeoff. Speeded decisions are made by decreasing the difference between the "yes" and "no" criteria, which leads to faster but more error-prone performance. A compound-cue random walk model can also predict a non-monotonic relationship between false-alarm rates and the delay of the signal-to-respond judgment. Responses made at relatively short delays are influenced by the item familiarities to a greater degree than the associative familiarity, but responses made at relatively long delays are influenced by the associative familiarity to a greater degree than the item familiarities. Thus, false-alarm rates initially increase with response delay and then decrease (Dosher, 1984; Gronlund and Ratcliff, 1989).

In addition to the non-monotonic signal-to-respond false-alarm rate function, the effect of repetitions on false-alarm rates from the free response procedure is a critical finding to be explained. Most random walk models do not describe how memories are represented or retrieved and therefore they make very few if any useful predictions concerning the accuracy of recognition memory unless they are implemented in a model of memory. Because they assume that the evidence on which a decision is based is a continuous random variable, it makes the most sense to implement a compound-cue

model of recognition in a global-matching model of familiarity (e.g., Shiffrin & Steyvers, 1997). Simple versions of these models predict that increasing the number of times that pairs are studied should increase false-alarm rates (e.g., Shiffrin & Steyvers, 1997; Xu and Malmberg, in press; see Criss & McClelland, 2006, for an analysis of the effect of list composition). In contrast, there is often little or no effect of repetitions on false-alarm rates and sometimes false-alarms decrease (Cleary et al., 2001; Kelley & Wixted, 2001; Xu & Malmberg, in press).

A related class of models assumes that associative recognition involves a computation of associative familiarity that is independent of the familiarity of the items comprising the test pair, and hence they are often referred to as *independent-cue models* (Criss & Shiffrin, 2005; Kelley & Wixted, 2001; Murdock, 1997). Independent-cue models can predict little or no affect of target presentations on false-alarm rates because the associative cue is only randomly similar to the contents of memory. That is, the familiarities of rearranged pairs are indistinguishable from the familiarities of the XY pairs, which consist of items that were not studied (see above). Independent-cue models that assume that associative information is the basis of all responses (Murdock, 1982) are disconfirmed by findings that show the false-alarm rates for rearranged pairs are greater than those for XY pairs.

Other independent-cue models assume that associative recognition is based on a combination of item and associative information (Criss & Shiffrin, 2005; Kelley & Wixted, 2001). These models can predict that the false-alarm rates are greater for rearranged pairs than for XY pairs, but they cannot explain a null effect of repetitions on false-alarm rates. Moreover they predict that the effect of repetitions on the latencies of

the hit rates and correct rejections of rearranged pairs should be similar (Criss & Shiffrin, 2005; Kelley & Wixted, 2001), and numerous experiments in my laboratory have found that hits tend to be faster than correct rejections, especially when pairs are relatively well encoded. Thus, compound and independent cue models are challenged to explain the accuracy and the latency of associative recognition.

While the compound cue models come up short in some important respects, they nevertheless have had a large impact on theory. The assumption that more that one type of information contributes to recognition performance increases the complexity of the models and introduces the possibility of different retrieval strategies that are characterized by the type of information that is used.

Dual-process models (Xu & Malmberg, in press; Yonelinas, 1999) are related to independent cue models insofar as both assume that associative recognition involves more than one source of information. It is often emphasized that the primary difference between these models is that the dual-process model assumes that associative information has a discrete form whereas the independent cue models assume that associative information has a continuous form. Many models assume that the retrieval processes involved in producing the different forms of information are themselves distinct. However, these assumptions do not necessarily have to be the case (e.g., Rotello et al., 2004), and these points tend to obscure the more important functional roles that associative information plays in memory performance.

Perhaps a more functionally important distinction between independent-cue and dual-process models is the independence assumption. Independent-cue models assume that associative information provides neutral evidence concerning the status of rearranged

pairs, and hence there is no relationship between the familiarity of the items comprising rearranged pairs and the ability to reject rearranged pairs based on associative familiarity. Dual-process models, on the other hand, assume that associative information is in the form of episodic details recalled from memory. The episodic details provide negative information concerning the status of rearranged pairs. Because there is a positive relationship between the familiarity of the items and the ability to reject otherwise familiar rearranged pairs, the more familiar the rearranged pair, the more likely it will be rejected based on recollection. Thus, dual-process models can account for the relationship between pair repetitions and false-alarm rates if it is assumed that some responses are based on familiarity and some responses are based on the episodic details retrieved from memory.

Dual-process models can also account for the non-monotonic relationship between false-alarm rates and response delay that is observed using the signal-to-respond procedure (Gronlund & Ratcliff, 1989). Accordingly, familiarity begins to accumulate shortly after the presentation of the stimulus. For both intact and rearranged pairs, this tends to be positive evidence. Most models assume that the ability to recall episodic details requires an additional processing stage, and hence takes longer than the production of familiarity. Once the details are recalled, however, they may be used to reject rearranged pairs, producing lower false-alarm rates at longer delays. Individual differences might arise based on the extent to which the outcome of familiarity process versus the recall process is emphasized. Thus, dual-process models provide a means for describing many facets of associative recognition performance, and effects of strategic factors naturally arise from their framework.

There are, however, different ways to implement a dual-process model, and not all of the implementations can account for both speed and accuracy of recognition performance. For instance, a REM dual-process model that was devised by Malmberg and Xu accounts for a variety of recognition accuracy findings (Malmberg & Xu, in press; Xu & Malmberg, in press; also see Malmberg, Holden, & Shiffrin, 2004). When applied to the free-response procedure it assumes that subjects initially evaluate the familiarity of the test pair, and if it does not exceed a subjective criterion the pair is called "new". However, familiarity is not assumed to be a sufficient basis for "old" status because the subject is most likely aware that both intact and rearranged pairs yield reasonably high levels of familiar. In order to call a pair "old", we assume that the subject first attempts to recall episodic details, which may be used to accept an intact pair or correctly reject a rearranged pair. If the attempt to recall episodic details fails, a guess is made.

While this dual-process model does a good job accounting for the accuracy of associative recognition, it has problems accounting for the latency of associative recognition. Notice that before responding "old", an attempt to recall is made, but "new" responses can be made without an attempt to recall. On this assumption, the latencies of hits should be slower than the latencies of correct rejections. Figure 1 shows the data from Experiment 2 of Malmberg & Xu (in press). These data are representative of a large number of similar experiments conducted in my lab, where we have shown that the dual-process model provides a reasonable account of the accuracy of associative recognition. However, the latencies of the hits are shorter than the latencies of correct rejections for rearranged pairs, and hence these findings indicate that the decision model

needs to be modified in order to account for both the accuracy and the latency of free-response associative recognition.

Modeling the Accuracy and Latency of Associative Recognition.

We have assumed that recognition can be performed in different ways, and that the strategy adopted depends on the subject's goals, the task, and subject's ability to implement a strategy that allows for a desired level of accuracy to be achieved in the shortest time possible. Differences in the goals and abilities of subjects are likely to be major sources of individual differences in memory performance. Therefore in order to understand the sources of individual differences, it is critical to understand the dynamics of associative recognition memory. In this section, I will describe several new models of the accuracy and latency of associative recognition that lead to a better understanding of the time course of associative recognition.

Encoding of Associative Information. As in other REM models (Shiffrin & Steyvers, 1997), this dynamic version of the Xu and Malmberg model of associative recognition assumes that words are represented by lexical/semantic traces consisting of w geometrically distributed features, where the probability of each feature value j is: $P(j) = g(1-g)^{j-1}$. These traces are assumed to be complete and accurate representations of the general knowledge about the words that has been built up over a lifetime.

When a list of word pairs is studied, an episodic trace is stored for each pair. Episodic encoding is a concatenation of incomplete and error prone copies of the lexical/semantic traces that represent the words. For each attempt, t, made to store a lexical/semantic feature in the episodic trace, there is a probability u^* that a non-zero

feature value will be stored. If a non-zero feature value is stored, it is copied correctly with probability *c*. If encoding of a feature does not occur, a zero is stored. In all the simulations, the encoding parameters are set to: w =10, g = .4, t = 8, u^* = .04, and c = .7.

The episodic association between the two words is the concatenation of two episodic vectors. These episodic traces also contain context features (cf. Criss & Shiffrin, 2005), but for the sake of simplicity allow me to omit this consideration for the present. Repetition of studied items has been modeled in REM in different ways (Malmberg et al, 2004; Malmberg & Shiffrin, 2005; Xu and Malmberg, in press; Shiffrin & Steyvers, 1997). For present purposes, I will assume that feature storage accumulates in a single-trace for a given pair of words because the more complex assumptions will not differentially affect the qualitative predictions of the models that we will consider.

Familiarity-Based Retrieval. The familiarity-based process operates in a fashion similar to other REM models (Malmberg et al., 2004; Shiffrin & Steyvers, 1998; Xu & Malmberg, in press). The critical difference is that the retrieval cue is compared to only a subset of the features that comprise the traces in the activated memory set, and that retrieval occurs over time in successive cycles on this subset of activated memory. For each unit of retrieval time, $T = 1 \ldots m$, a set of features, F_T, is sampled from an activated memory set with a probability of *s*. In the simplest simulations, context does not play an important role in associative recognition, and hence the activated set consists of the traces stored during study.

Modeling how the features (or evidence) are sampled from memory is usually ignored in random walk models. However, the issue is both theoretically interesting and pragmatically important, and therefore I will consider two approaches. In the simplest

model, sampling occurs with replacement. That is, each feature in the activated memory set is sampled with probability *s* for each *T*, and the probability of sampling a given feature for F_T is independent of sampling that feature for F_{T+1}. For each *T*, global matching is performed over F_T producing a familiarity value Φ_T. The logarithm of Φ_T is computed in order to produce either positive or negative evidence that the test pair was studied. This is the evidence that will be used to drive the random walk, and this evidence accumulates over *T* such that when *T* = *m* the accumulated values of odds ($\phi(m)$) is:

$$\phi(m) = \sum_{T=1}^{m} \log \Phi_T = \sum_{T=1}^{m} \log\left(\frac{1}{N}\sum_{i=1}^{j} \lambda_j\right), \text{ where } \lambda_j = (1-c)^{n_{jq}} \prod_{i=1}^{\infty}\left[\frac{c+(1-c)g(1-g)^{i-1}}{g(1-g)^{i-1}}\right]^{n_{ijm}} .$$

Thus, $\phi(m)$ represents the evidence accumulated to time *T* = *m* in response to a probe with a compound cue consisting of the lexical/semantic traces representing the words comprising the test stimulus.

An intuitively unappealing characteristic of the sampling-with-replacement model is that it assumes that there is potentially an infinite amount of information to be gained from any probe of memory. Given that the amount information stored about an event is assumed to be finite, how can the information retrieved from memory be potentially infinite? The answer to this question is unclear, even though sampling with replacement is the standard assumption in the literature. This assumption is, however, what guarantees that a probe will result in the evidence crossing one of the decision criteria.

As an alternative to the sampling-with-replacement model, the sampling-without-replacement model assumes that once a feature has been sampled as the result of a probe of memory, it remains a part of F_T. Thus, features accumulate in F_T and the matching

process is computed over a more and more complete version of the activated memory set, as F_T will become identical to activated memory set at sufficiently long values of T:

$$\phi(m) = \log \Phi_T = \log\left(\frac{1}{N}\sum_{i=1}^{j}\lambda_j\right), \text{ where } \lambda_j = (1-c)^{n_{jq}} \prod_{i=1}^{\infty}\left[\frac{c+(1-c)g(1-g)^{i-1}}{g(1-g)^{i-1}}\right]^{n_{ijm}}.$$

In the following, section I will compare the performance of these sampling models.

Decision. We already discussed that the nature of the decision process has important implications for the dynamics of recognition memory. For instance, the assumption that a recall process must be invoked before an "old" response can be made suggested that "old" responses should be slower on average than "new" responses.

For the signal-to-respond procedure, I assume that a response is based on the evidence accumulated at any given T. If $\phi(m) \geq 0$, the response is "old", otherwise it is "new". That is, the subject is not free to determine the amount of evidence required to make a response.

The nature of the free-response decision process depends on the sampling model. First, consider the sampling-with-replacement model. For the free response procedure, I assume that there is an old criterion, K_O, and a new criterion, K_N. If $\phi(m)$ is greater than K_O, then the response is "old", and if $\phi(m)$ is less than K_N, the response is "new". If $K_N < \phi(m) < K_O$, then a new sample of features is drawn from the activated memory set, matched against the retrieval cue, and added to $\phi(m)$. (Another way of viewing the signal-to-respond model is to assume that $K_N = K_O = 0$.) Because $\phi(m)$ is a summed accumulation of evidence based on a series of independent samples from the activated memory set, this model guarantees that at some T the evidence will exceed one of the

decision criteria, unless of course the log odds of the match between the compound cue and the activated memory set is 0, which is highly unlikely.[1]

A different assumption is that $\phi(m)$ is based on sampling-without-replacement at each *T*. That is, the sample drawn from the activated memory system becomes more complete with respect to *T*, and the evidence is evaluated anew after each sample. Because *T* is bounded by the amount of information stored during study, it is therefore possible (even likely) under sampling without replacement that neither decision criterion will ever be exceeded. Thus, a sampling-without-replacement model is inherently more complex than a sampling-with-replacement model because there must be a stopping rule that is utilized when neither decision criterion has been met and there is little or no evidence left to sample or the subject is unwilling to search any longer.

For instance, I assume that for free response recognition K_{Max} is the maximum amount of time the subject is willing to search memory (cf. Raaijmakers & Shiffrin, 1980). If the $K_N < \phi(m) < K_O$ and $T = K_{Max}$, then a guess is made. This stopping rule is similar to the guessing assumption made in earlier models of recognition accuracy (Malmberg, Holden, & Shiffrin, 2004; Malmberg & Xu, in press; Xu & Malmberg, in press). Accordingly, I usually assume that $K_{Max} = 10$ and that guessing is biased to respond "old", which I assume is equal to .90 in the present simulations. The guessing bias might seem unreasonably high, but its value is consistent with prior applications of the accuracy model where it was determined based on attempts to quantitatively fit data and justified on the assumption that even when recall fails that items seem relatively

[1] One might be tempted to assume that information accumulates based on a match of the cue against the complete activated memory set for each *T* because this would allow for an optimal basis for a decision, but it would obviate the need for an accumulation of evidence since there would be no need to sample more than once.

familiar (i.e., there was no evidence to indicate that the item or pair was new; Malmberg, Holden, & Shiffrin, 2004; Malmberg & Xu, in press; Xu & Malmberg, in press).

Familiarity-Based Performance. Before moving to a discussion of the recall process, it is worth exploring the assumptions that concern the accumulation of evidence based on the global-matching process. Other REM models have focused on only the accuracy of recognition memory. Those models used a single decision criterion with a default value of $\log(\Phi) = 0.0$, which can be considered Bayesian optimal because the odds are equivalent to the probability that test stimulus is old divided by the probability that the test stimulus is new given the matching and mismatching features (Shiffrin & Steyvers, 1997). In the dynamic model, however, only a sample from the activated memory set is compared to the retrieval cue at time T, and thus the evidence accumulated at any given T is not likely to provide an optimal basis for a recognition decision. The sample-without-replacement model comes closer to optimality in the sense that it is possible that the entire set of activated features would be involved in the global-matching process at a sufficiently long T.

Figure 2 shows the behavior of the sampling-with-replacement model (left panel) and the sampling-without-replacement (right panel) model derived from the signal-to-respond assumptions. It shows how hit rates and false-alarm rates are affected over time by the number of times that targets are presented. Note that the models make identical predictions when the signal to respond comes at the earliest delay because replacement occurs or does not occur only after the first sample is evaluated. Hit rates and false-alarm rates then tend to increase with repetitions and delay in both models.

Like other implementations of the compound-cue model in a global-matching framework, performance based on familiarity cannot alone account for the non-monotonic relationship between the response delay and false-alarm rates (Gronlund & Ratcliff, 1989). Nevertheless, this analysis tells us something useful about how to efficiently model the relationship between speed and accuracy: The sampling-without-replacement model produces a higher level of accuracy than the sampling-with-replacement at any $T > 1$. Therefore, sampling without replacement is inherently more efficient than sampling with replacement, and if we want a model of familiarity within a Bayesian system of memory this is an important factor to consider.

One structural factor that influences the rate at which information is accumulated is the sample size, *s*. It plays a role in the present model that is similar to the role drift rate plays in a random walk model. Variability in the drift rate has been shown to be critical in order to achieve realistic levels of accuracy (Ratcliff, 1978). In the present models, variability in the drift rate is the result of the variability in the familiarity of the pairs. That is, the variability occurs in Φ_T for each sample, which is determined by how well pairs are encoded and *s*, and therefore it is likely to differ between individuals and populations.

Figure 3 shows how the sample size affects free-response recognition, where s = .10, .25, and .40 and pairs were studied 1, 2, 3, 6, or 12 times. Figure 3 also illustrates the point concerning the optimality of recognition memory when the decision rule is based on partial information. The top panels of Figure 3 show how the sample size affects the accuracy and latency of the sample-with-replacement model and bottom panels show the same for the sample-without-replacement model. In both models, the latencies decrease

as the sample size increases and that the latencies of the correct rejections are much longer than the latencies of the hits regardless of the value of *s*. Moreover, increases in the sample size improve accuracy and decrease latencies. Because optimal decisions are based on all the information in memory, but the samples are incomplete representations of what is stored, there is less noise in the accumulation of evidence as the sample size increases, especially in the sample-without-replacement model. The sampling-without-replacement model achieves greater accuracy in far less time than the sampling-with-replacement model, despite the fact that guessing is a necessary component of the sampling-without-replacement model. Thus, as a structural component of memory, a component whose operations are outside the control of the subject, the sampling without replacement is to be preferred in a free-response model as well as in a signal-to-respond model.

In the context of this chapter on memory as skilled cognition, it is also important to explore how a strategically initiated speed-accuracy tradeoff might be achieved based on global matching. This will also help us to understand the behavior of the more complex dual-process model. Under most conditions, the sampling rate is not likely to be influenced by the goals or strategies of the subject, at least after encoding is complete. Thus, let us assume that differences in sampling rate reflect differences in a structural aspect of memory. On the other hand, the decision criteria are usually assumed to be under the control of the subject (Egan, 1958; Green & Swets, 1966). In simple versions of a random walk model, for instance, increasing the difference between K_O and K_N should improve accuracy and increase latencies, but Figures 4 and 5 show that this is not

necessarily the case for the present model. Let us consider that changes in the location of the decision criteria reflect strategies with different emphases on speed versus accuracy.

Let us first consider the effect of implementing a symmetrical change in the decision criteria. Three sets of criterion locations were chosen for the simulation whose results are shown in Figure 4: K_O = 1.0, 2.0, and 3.0, where K_N = -K_O. The right panels of Figure 4 show that latencies increase with increases in the spread of decision criteria for both sampling models, and that accuracy actually decreases. Increasing the spread has a positive effect on both hit rates and false-alarm rates. However, the increase is smaller for hit rates than for false-alarm rates. This is because the rearranged pairs are relatively familiar due to their similarity to the studied pairs, and therefore lowering K_N makes it less likely that enough negative evidence will accumulate to allow for a negative response. Hence, a symmetrical change in the decision criteria does not produce a speed-accuracy trade off.

Upon further analysis, a symmetrical change in the decision criteria does not make much sense as a means of slowing performance in order to increase accuracy given the nature of the associative recognition task. Both the targets and the foils are relatively familiar, and it is in the interest of the subject to maximize the likelihood of a correct rejection. To achieve a goal of greater accuracy, the subject could instead asymmetrically alter the decision criteria by increasing K_O relative to K_N. The result of a simulation of this model is shown in Figure 5, where K_O = 1.0, 2.0, and 3.0 and K_N = -1.0. Here, the latencies increase with increases in the spread of the decision criteria, and accuracy increases. Once again, the spread of criteria has little effect on hit rates, but false-alarm rates decrease as the spread increases. Thus, an asymmetrical change in the

location of the criteria with respect to $\phi(m) = 0$ is one way to achieve a speed-accuracy tradeoff for associative recognition based on familiarity.

So far the models that we have considered cannot alone account for the accuracy and latency of associative recognition. However, we have made some critical observations that will allow us to construct a model that can better account for associative recognition performance. Based on these analyses, one way to achieve a speed-accuracy tradeoff for free-response associative recognition in a global-matching model is to assume that there is an asymmetrical spread in the decision criteria. In addition, there is little qualitative difference between the sampling-with-replacement model and the sampling-without-replacement model. Therefore, because the sampling-without-replacement model is the more efficient, less noisy model and because it seems unprincipled to assume that an infinite amount of evidence can be obtained from a finite amount of information stored in memory, I will assume that sampling occurs without replacement from now on.

Recollection. Recall in REM involves the same mechanisms of sampling and recovering traces as in the SAM model (Malmberg & Shiffrin, 2005; Raaijmakers & Shiffrin, 1980; Shiffrin & Steyvers, 1998). The sampling process follows a Luce choice rule, whereby the probability of sampling trace, i, from the set of sampled features, F_T, given retrieval cue, Q, is positively related to the similarity of Q and trace i, and negatively related to the similarity of Q and other traces in memory:

$$P(i \mid Q) = \frac{\lambda_i}{\sum_{j=1}^{n} \lambda_j} .$$

Note that the sampling involved in recollection involves retrieving traces from an activated set of memory traces in addition to first sampling a set of features from memory, whereas the sampling model considered earlier only involves the latter.

Recovery is assumed to be a special case of a threshold process. Recovery of trace *i* is successful if and only if there are at least K_R non-zero features in trace *i* sampled from activated memory set. If the number of non-zero sampled features does not exceed K_R, then recollection fails. If K_R is exceeded, then the individual features are used to make a decision by comparing them to the test stimulus. Note that it is possible for K_R to be exceeded by varying amounts of evidence recalled from memory, and hence it is possible that this information might provide more than an all-or-none source of information. While this assumption will not play an important role in the present modeling, it might help to address some of the issues involving the nature of information recalled from memory and how it influences confidence judgments (e.g., Rotello et al., 2004).

The decision based on recalled information assesses the matching and mismatching features of the trace sampled from F_T and the test stimulus. I assume that the target trace must be sampled from memory in order to recall that an intact pair was studied, and I assume that one of two traces that correspond to rearranged pairs must be sampled in order to recall that a rearranged pair was not studied. Therefore, the recovery process only leads to veridical responses. That is, if an intact pair is tested, recollection can only lead to a positive response if it is successful (and vice versa for rearranged pairs).

In a Bayesian system, the comparison of the sampled features to the test stimulus would ideally be informed by the extent and accuracy of encoding. However, this might not always be known. Moreover, recovery most likely involves an interaction between the features sampled from memory and general knowledge which is in the form of lexical/semantic traces in REM. The additional complexity involved in implementing these models probably would not affect the behavior of the model vis-à-vis our goal of understanding the dynamics of recognition memory, however. In the present simulations, I have therefore allowed for one mismatching feature between the sampled features of a target trace and an intact stimulus in order to take into account the errors in encoding. For rearranged pairs, I have required that there be at least two mismatching features in order for recollection to succeed in rejecting the otherwise familiar foils.

These assumptions provide a more complete account of recovery than that which has been provided previously (e.g., Xu & Malmberg, in press). Nevertheless, the recovery model is still overly simplistic and therefore the selection of different parameter values is necessarily somewhat arbitrary. For now, however, these assumptions capture much of what needs to be explained in a formal manner. Perhaps the most important aspect of this model is the assumption that only one mismatching feature is allowed in order to judge that a test stimulus was studied based on recollection, whereas at least two mismatching features must be observed in order to judge that a test stimulus was not studied. This makes recalling-to-reject less demanding than recalling-to-accept given that there are w features possible to sample from F_T. In essence, this means that given the same levels of encoding that responses to targets are more likely to be based on familiarity than responses to foils and that responses to foils are more likely to based on

recollection than responses to target (all else being equal). Hence, these assumptions will contribute to the model's predictions concerning the relationship between the latencies of hits and correct rejections.

The recollection process is assumed to be slower than the familiarity-based process (cf. Dosher, 1984; Gronlund & Ratcliff, 1989). For simplicity, it is arbitrarily assumed that the result of the recall attempt begins to be available at $T = 4$. Presumably, this accounts for the amount of time it takes to construct a cue, probe, sample, recover, and evaluate the contents of a trace. Some of these operations must also be completed prior to completing the first global-matching comparison. Therefore, I assume that no responses are made until $T = 2$ in the free response model.

Once recollection succeeds no further attempts to recall information is made. This is true even in the case when the recollection succeeds prior to $T = 4$. When recollection fails, a new attempt is made on the subsequent probe of memory. At times, a trace will be sampled, but its contents do not exceed K_R. Because of the sampling-without-replacement model, these contents will still be available in F_T when the trace is sampled on a subsequent attempt to recall. In this sense, the sampling-without-replacement model is simpler than the sampling-with-replacement model that I have rejected.

Dual-Process Decision Strategies. First, let us consider the signal-to-respond model. Assume that the subject responds at the moment the signal is made (regardless of T). If the signal occurs before recall is complete, then a response is made based on familiarity in the fashion described earlier. If the signal occurs, after recall is complete and it is successful, then the response is "old" if the test stimulus is a target (i.e., the

recovered features tend to match the intact pair) and "new" if the test stimulus is a foil (i.e., the recovered features tend to mismatch the rearranged pair). If recall fails, then a decision is based on familiarity in the usual fashion: If $\phi(m) \geq 0$, the response is "old", otherwise it is "new".

Figure 6 shows the performance of the dual-process signal-to-respond model. It is instructive to compare this performance with the performance of the familiarity-based sample-without-recovery model in Figure 2. Both models predict an increase in hit rates with increases in delay and presentations. Unlike the familiarity-based model, however, the dual-process model predicts the observed non-monotonic relationship between response delay and false-alarm rates if the initial encoding of the trace is relatively strong (i.e., the target pairs were presented more than 1 time).

The contribution of recollection to signal-to-respond performance is greater for targets than for foils when pairs are presented infrequently, but as the number of presentations increases the contribution is greater for foils than for targets. In addition, the contribution of recollection increases for both targets and foils as the number of presentations and delay increases. Before recovery and trace comparisons are completed, performance is based on familiarity only, and hence false-alarm rates initially increase with delay if the targets are sufficiently well encoded. After recollective processing is assumed to be completed (i.e., $T = 4$), the evidence obtained tends to indicate that rearranged pairs are familiar and that the items comprising the test pair were not studied together. Both sources of evidence tend to increase in probability as the delay increases, and thus false-alarm rates decrease with delay thereafter.

The performance of the signal-to-respond procedure is assumed to be relatively immune to strategic factors, as the response is demanded at a specific point in time. In free-response recognition, however, the choice of how long one waits before making a response is up to the subject. Therefore, it is in the free-response procedure where we would expect to find greater influences on strategic differences in performance.

One way to affect the length of time before a response is initiated is by increasing the spread between the familiarity-based decision criteria. Figure 5 shows that asymmetrically increasing the spread of the old and new criteria by increasing the old criterion relative to new criterion produces an increase in accuracy and a decrease in speed in the familiarity-based model. Figure 7 shows how the same manipulations of the decision criteria affect the dual-process model free-response performance.

The left panel of Figure 7 shows that accuracy increases as the spread of the criteria increases, primarily by reducing false-alarm rates. In addition, when the spread is relatively small, false-alarms increase with increases in the number of target presentations. In this case, performance is very similar to what was observed in Experiment 2 of Malmberg & Xu (in press, Figure 1).

When there is a relatively moderate spread in the decision criteria, there is little or no effect of pair strength on false-alarm rates and contribution of recollection to performance increases. In several experiments, Kelley & Wixted (2001) observed a similar pattern of false-alarm rates. It is important to note, however, that in their experiments Kelley and Wixted only varied pair strength at two levels. Thus, the function relating their two observations is unknown and it is possible that a non-linear relationship exists between pair strength and false-alarm rates (e.g., Experiment 3 and 4

in Malmberg & Xu, in press). We will return to this matter shortly. As the spread of the criteria increases, fewer responses to rearranged pairs are based on familiarity, and hence the false-alarm rates can even decrease as the number of target presentations increases. This can explain the variability in the observed patterns of false-alarm rates between several experiments in the literature if one assumes that different groups of subjects adopt sets of decision criteria that lead to more or less responses based on recollection (e.g., Cleary et al., 2001; Malmberg & Xu, in press; Kelley & Wixted, 2001; Xu & Malmberg, in press). This could be due to factors such as motivation, instructions, fatigue, etc.

The middle panel of Figure 7 shows that asymmetrically increasing the spread of decision criteria increases the latencies of the responses. When the spread of the criteria is relatively small the latencies of the correct rejections are slower than the latencies of the hits, as was observed in Experiment 2 of Malmberg and Xu (in press) and corresponding to the rise in false-alarm rates with increases in target presentations. As the spread increases, hits tend to become slower than correct rejections because it becomes less and less likely that familiarity will lead to an “old” response. This is illustrated in the right panel of Figure 7 by increases in the amount recollective-based responding as the spread of the decision criteria increases. In addition, the contribution of recollection to performance tends to decrease for targets and increase for foils as the number of times targets are presented increases. Thus, repetitions have the opposite effects on the contribution of recollection to hit rates for free-response and signal-to-respond performance.

The increase in latencies associated with the increase in the spread of the decision criteria leads to an increase in accuracy for foils but not for targets, which is to say that

overall accuracy increases. The increase in latencies is much greater for targets than for foils, such that latencies of hits is greater than the latencies of correct rejections when K_O = 5. This model is not necessarily disconfirmed by the data shown in Figure 1, since those findings are that false-alarm rates increase as the number of target presentations increase, which is what we observe in Figure 7 when K_O = 1. More data are therefore required to in order to evaluate this prediction of the model. Specifically, it would be of interest to devise a manipulation that affects the subject's tendency to respond based on recollection and determine if the latencies of hits and correct rejections are differentially affected.

Distinguishing between strategic and structural sources of systematic variability in recognition performance is an important component to understanding individual differences. The prior analysis shows differences in the contribution of recollection to performance can be achieved by controlling the spread between the familiarity-based decision criteria. How would a change in some structural aspect of memory affect associative recognition? Figure 8 shows the effect of an increase in the sample size, *s*, on dual-process free-response performance. As the sample size increases, latencies tend to decrease, and there is little effect of sample size on accuracy: Both hit rates and the false-alarm rates increase with sample size. This is because increasing the sample size decreases the latencies of the familiarity-based responses. For rearranged pairs this generally leads to false-alarms, and these tend to occur prior to when recollection is assumed to be complete in this model. For targets, the larger sample sizes are more likely to make their familiarity exceed K_O and they tend to exceed K_O more quickly, prior to the completion of recollection. This is illustrated in the right panel of Figure 8 where it

shows that the increase in sampling rate has little effect on the contribution of recollection to correct rejections, but the contribution of recollection to hits decreases as the sampling rate increases. This is because more responses are based on familiarity prior to when recollection is assumed to be complete. Thus, there appear to be patterns of accuracy and latency data that can distinguish between strategic and structural differences in the manner in which associative recognition is performed. Changes in the location of "old" decision criterion can possibly account for differences in the patterns of false-alarm rates observed in the literature, whereas difference in sampling rate probably cannot.

So far the dual-process model performs fairly well with the possible exception of the longer response latencies for hit than for correct rejections when the spread of the decision criteria is relatively large. While it is unknown if this is a truly problematic prediction, it is still worth exploring a different way to delay responding as a means of increasing accuracy. One way this might be accomplished is by leaving the old and new criteria fixed, but delaying the responses until a subjective metacognitive estimate of when recollection should be completed is met. That is, subjects may have meta-level beliefs about the amount of time recollection takes to succeed, and responses could be delayed until that time even though the familiarity-based evidence is already sufficient for making a response. This could, of course, be affected by the nature of the task or difficulty of the task. In this case, I will assume that subjects delay their responding until $T = 4$.

Figure 9 reports a simulation in which responses are made as soon as either familiarity-based or recollective-based evidence is sufficient for a response versus the situation where all responses are delayed until $T = 4$. The latter simulation is similar to

Experiment 3 in Malmberg and Xu (in press). In all respects that experiment was the same as the Experiment 2 from Malmberg and Xu (in press) which produced the data in Figure 1. The only difference is that responses were delayed by 2 s. That is, the subjects were to respond after 2 s. had elapsed since the presentation of the test stimuli. The accuracy data from both experiments are shown in middle-left panel of Figure 9. The primary effect of delaying the yes-no response was to induce a statistically reliable increase then a decrease in false-alarm rates as the number of target presentations increased. The left panel of Figure 9 shows a similar pattern of data derived from the computer simulation. In addition, the relationship between the latencies of the hits and the correct rejections is maintained, even as accuracy increases as the result of delaying the response.

How does a change in a structural aspect of memory affect the performance of this model? Figure 10 shows the effect of sampling rate on associative recognition performance based on the free-response model that assumes that all responses are delayed until some $T = m$. Unlike the free-response model that assumes that responses will be made as soon as either the familiarity-based evidence or the recollective-based evidence provides a sufficient reason to respond (Figure 8), increasing the sampling rate produces a robust increase in accuracy and a decrease in latencies. Comparing Figures 8 and 10 shows that increasing the sampling rate has large effects on the contribution of recollection to correct rejections when responses are delayed but not when the subject is free to respond based on familiarity prior to recollection being made available. This is because the present model responds based on recollective evidence when it is available. Thus, the effect of a structural change in memory can have different effects on

recognition memory performance that depend on the nature of the decision invoked at the time of retrieval.

It is important to note that these predictions and all of the predictions derived for the purposes of these analyses were generated without formally obtaining a “best fit” of the models to the data. It is therefore quite probable that more accurate representations of the data can be generated by the models. The point being that the qualitative predictions of different models can be used to distinguish between them, which is an encouraging sign, given the goal of understanding the sources of individual differences in recognition memory performance.

Thus, there are at least two models for achieving a speed accuracy tradeoff for associative recognition in a dual-process framework. One model assumes that subjects vary their familiarity-based decision criteria. The greater the distance between them the less likely responses will be based on familiarity and therefore accuracy increases. The other model assumes that subjects simply delay responding until a meta-cognitively determined estimate of the amount of time one should wait in order to achieve their goals. The longer one waits, the more evidence there is on which to base a decision and the more accurate performance becomes.

Interestingly, there appears to be a way to empirically distinguish between these models because the patterns of accuracy and latency data are different. According to the model that assumes that subjects vary their decision criteria (see Figure 7), increases in the number of times targets are presented should correspond to small decreases in the contribution of recollection to hit rates, and the latencies of the hits should increase relative to the latencies of correct rejections. According to the response-delay model, by

way of comparison, when responses are relatively slow (see Figure 9), increases in the number of times targets are presented should increase the amount that recollection contributes to hits, but the latencies of hits should be less than latencies of correct rejections.

Thus, it is possible to distinguish between two models of the speed-accuracy tradeoff by observing the patterns of accuracy and latency data. At present, there are no data that can help us make an empirically informed decision about which model of speed-accuracy tradeoffs is correct. Doing so could have important implications for how fMRI and EEG data are interpreted. That is, the important implication is how we interpret the effect of an increase in accuracy and latency. According to one model, increases in pair strength produces an increase in the contribution of recollection to hits and correct rejections, and according to the other model increases in pair strength only increases the contribution of recollection to correct rejections. Thus, such modeling endeavors when used in combination with neuroscientific methods can produce conclusions based on a combination of our understanding of different brain structures and our understanding of the dynamics of memory systems.

In recent years, an important means for evaluating models of response latencies is to observe not only the central tendency of responses but also the distribution of response latencies. A critical finding in many areas of research is that latency distributions are skewed such that the leading edge is relatively steep compared to the tail of distribution, as opposed to being normally distributed. As a final test of whether the current model is viable, I conducted a simulation of the free-response model from which the latency distributions for hits and correct rejects could be obtained. Figure 11 shows the hit and

correct rejection latency distributions of the dual-process model for two levels K_O. In all cases, the distributions are skewed. The correct rejection latency distribution is clearly bimodal, owing to an early mode corresponding to an initial contribution of familiarity to performance and a later mode corresponding to a later contribution of recollection. The later mode diminishes as the spread between the old and new criteria decreases, as this reduces the contribution of recollection to performance.

Conclusions

Over the past few decades, we have learned a lot about the nature of human memory. Perhaps the best examples of the advances in our understanding that have been made are the formal models that have been developed. A shortcoming of these models, however, is that their scope has traditionally been limited to providing accounts of how the average individual remembers. The reason for this has not only to do with the incremental nature of the scientific method, but also the empirical limitations of conducting memory research and the relegation of individual differences to the error term in our description of the data. If, however, formal modeling of memory is to have an impact on everyday lives of people, we must apply what we know about how the average individual remembers to understanding why some people or populations are better rememberers than other.

Other areas of memory research have promoted methodologies that are more amenable to identifying individual differences. The recent upsurge in interest in cognitive neuroscience is due in large part to the saliency of the data that provide a glimpse into the brain of the rememberer. Such glimpses not only can tell us about where

the brain supports human memory, but they can also tell us about how individual brains differ when they perform a memory task. In my experience, subjects and laypeople have little interest in how memory works. Their primary interest is whether they are as bad at remembering as they think they are. Indeed, the greatest prize one often receives when visiting a cognitive neuroscience laboratory is a photograph of their brain. As such, the cognitive neuroscience approach to memory is often more appealing to the layperson, than the intrinsically dense mathematical formulations that increasingly characterize behavioral research.

A limitation of current cognitive neuroscience approach is that extant theories have more to say about where or when memory occurs and has relatively little to say about how or why memory occurs. Thus, we can point to an area of the brain that discriminates between good and not-so-good remembers and not have a very good understanding of why one person remembers better than another. For instance, are differences in the patterns of brain activation due to structural abnormalities of the brain or due to differences in the utilization of different memory strategies?

Viewing memory as skilled cognition provides an opportunity to take advantage of the explanatory power of the modeling approach to memory and the specificity of cognitive neuroscience approach to understanding individual differences. In this chapter, I have described a theory that assumes that there is a variety of ways to perform a given memory task, and the strategy adopted is assumed to be efficient with respect to the subjects goal to achieve a given level of accuracy in the shortest time possible. The extent to which one is a “good rememberer” depends (a.) on the operations of the

structural components of memory and (b.) on the quality of one's meta-level understanding of the nature of the task and nature of memory and decision.

Discriminating between structural and strategic sources of variance depends on having a model of how they affect performance, and doing so might allow us to discriminate between those with serious structural memory impairments and those who are not effectively choosing efficient strategies for remembering. Such a theory could be further extended to include assumptions about the effects of implementing different memory strategies on brain activity, and these assumptions could be tested using traditional behavioral and cognitive neuroscience methodologies. To the extent that these joint efforts are successful, we may in the future use this knowledge to support diagnostic tools for evaluating human memory performance in order to more effectively characterize and treat memory impairments.

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, and P. Suppes (Eds.), *Contemporary developments in mathematical psychology, Vol. 1, Learning, memory, and thinking* (pp. 243-293). San Francisco: Freeman.

Baranski, J. V. & Petrusic, W. M. (1998). Probing the locus of confidence judgments: Experiments on the time to determine confidence. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 929-945.

Cleary, A. M., Curran, T., & Greene, R. L. (2001). Memory for detail in item versus associative recognition. *Memory & Cognition, 29,* 413-423.

Criss, A. H. and McClelland, J.L. (2006). Differentiating the differentiation models: A comparison of the retrieving effectively from memory model (REM) and the subjective likelihood model (SLiM). *Journal of Memory & Language: Special Issue on Computational Models of Memory*

Criss, A.H. and Shiffrin, R.M. (2005). List discrimination and representation in associative recognition. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 31(6), 1199-1212.

Curran, T., Schacter, D. L., Norman, K. A., Galluccio, L. (1997). False recognition after right frontal lobe infarction: Memory for general and specific information. *Neuropsychologia, 35,* 1035-1049.

Diller, D.E., Nobel, P.A., & Shiffrin, R.M. (2001). An ARC-REM model for accuracy and response time in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27,* 414-435.

Dosher, B. A. (1984). Discriminating Preexperimental (Semantic) from Learned (Episodic) Associations: A Speed-Accuracy Study. *Cognitive Psychology, 16,* 519-555.

Dunn, J. A. (2004). Remember-Know: A matter of confidence. *Psychological Review, 111,* 524-542.

Egan, J. P. (1958). *Recognition memory and the operating characteristic.* Indiana University Hearing and Communications Laboratory, AFCRC-TN-58-51.

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

Greene, R. L. (1996). Mirror effect in order and associative information: Role of response strategies. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 22(3), 687-695.

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

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

Hintzman, D. L., Curran, T., & Oppy B. (1992). Effects of similarity and repetition on memory: Registration without learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 667-680.

Hockley, W. E. (1992). Item versus associative information: Further comparisons of forgetting rates. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1321-1330.

Hockley, W. E. (1994). Reflections on the mirror effect for item and associative recognition. *Memory & Cognition, 22,* 713-22.

Hockley, W. E. & Niewiadomski, M. W. (in press). Strength-based mirror effects in item and associative recognition: Evidence for within-list criterion changes *Memory & Cognition*.

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

Lewandowsky, S. & Hiet E. (2006). Some targets for memory models. *Journal of Memory & Language*, 55, 441-446.

Light, L. L., Patterson, M. M., Chung, C. Healy, M. R. (2004). Effects of repetition and response deadline on associative recognition in young and older adults. *Memory & Cognition*, 32(7), 1182-1193.

Macmillan, N. A. & Creelman, C. D. (1990) Response bias: characteristics of detection theory, threshold theory, and nonparametric indexes. *Psychological Bulletin*, 107, 401-413.

Malmberg, K. J., Holden, J. E., & Shiffrin, R. M. (2004). Modeling the effects of repetitions, similarity, and normative word frequency on judgments of frequency and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 319-331.

Malmberg, K. J., & Shiffrin, R. M. (2005). The "one-shot" hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 322-336.

Malmberg, K. J. & Xu, J. (in press). On the flexibility and on the fallibility of associative memory. *Memory & Cognition*.

Malmberg, K. J., Zeelenberg, R., & Shiffrin, R.M. (2004). Turning up the Noise Or Turning Down the Volume? On the Nature of the Impairment of Episodic Recognition Memory by Midazolam. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 540-549.

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

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review, 104,* 839-862.

Murdock, B. B. (1993). TODAM 2: A model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review, 100,* 183-203.

Nelson, T. O. & Narens, L. (1990). Metamemory: A Theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory*. New York: Academic Press.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59-108.

Ratcliff, R. & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111,* 333-367.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In Bower, G. H. (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 14, pp. 207-262). New York: Academic Press.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 294-320.

Rotello, C. R., Macmillan, N. A., & Reeder, J. A. (2004) Sum-difference theory of remembering and knowing: a two-dimensional signal-detection model. *Psychological Review*, 111(3), 588-616.

Shiffrin, R.M., & Steyvers, M. (1997). A model for recognition memory: REM – retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145-166.

Shiffrin, R.M., & Steyvers, M. (1998). The effectiveness of retrieval from memory. In M. Oaksford & N. Chater (Eds.), *Rational models of cognition* (pp. 73-95). Oxford, England: Oxford University Press.

Van Zandt, T, & Moldonado-Molina, M. M. (2004). Response reversals in recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30, 1147-1166.

Xu, J. & Malmberg, K. J. (submitted). Modeling the effects of verbal- and nonverbal-pair strength on associative recognition.

Yonelinas, A.P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition, 25,* 1397-1410.

Acknowledgments

Many thanks to Harry Bahrick and Aaron Benjamin for their comments on prior versions of this chapter and to Roger Ratcliff for a very generous and enlightening conversation concerning the random walk model. Please address correspondence to: malmberg@cas.usf.edu.

Figure Labels

Figure 1. Data from Malmberg and Xu (in press) illustrating the relationship between the speed and accuracy associative recognition in a free-response paradigm.

Figure 2. Hit rates and false-alarm rates as a function of response deadline and pair presentations for the familiarity-based signal-to-respond model under the assumptions of sampling-with-replacement versus sampling-without replacement.

Figure 3. A comparison of the effect of different sampling rates on the speed and accuracy of familiarity-based free-response recognition under the assumptions of sampling-with-replacement versus sampling-without replacement.

Figure 4. A comparison of the effect of a symmetrical shift in the decision criteria on the speed and accuracy of familiarity-based free-response recognition under the assumptions of sampling-with-replacement versus sampling-without replacement.

Figure 5. A comparison of the effect of an asymmetrical shift in the decision criteria on the speed and accuracy of familiarity-based free-response recognition under the assumptions of sampling-with-replacement versus sampling-without replacement.

Figure 6. Hit rates and false-alarm rates as a function of response deadline and pair presentations for the dual-process signal-to-respond model.

Figure 7. The effect of an asymmetrical shift in the decision criteria on the accuracy and latency of the dual-process model and the contribution of recollection to its performance.

Figure 8. The effect of sampling rate on the accuracy and latency of the dual-process model and the contribution of recollection to its performance.

Figure 9. The effect of delaying responses on the accuracy and latency of the dual-process model and the contribution of recollection to its performance.

Figure 10. The effect of sampling rate on the accuracy and latency of the dual-process model and the contribution of recollection to its performance when responses are delayed.

Figure 11. The distribution of the latencies of hits and correct rejections for different levels of bias to respond “old” based on familiarity in the dual-process delayed-response model.

Figure 1.

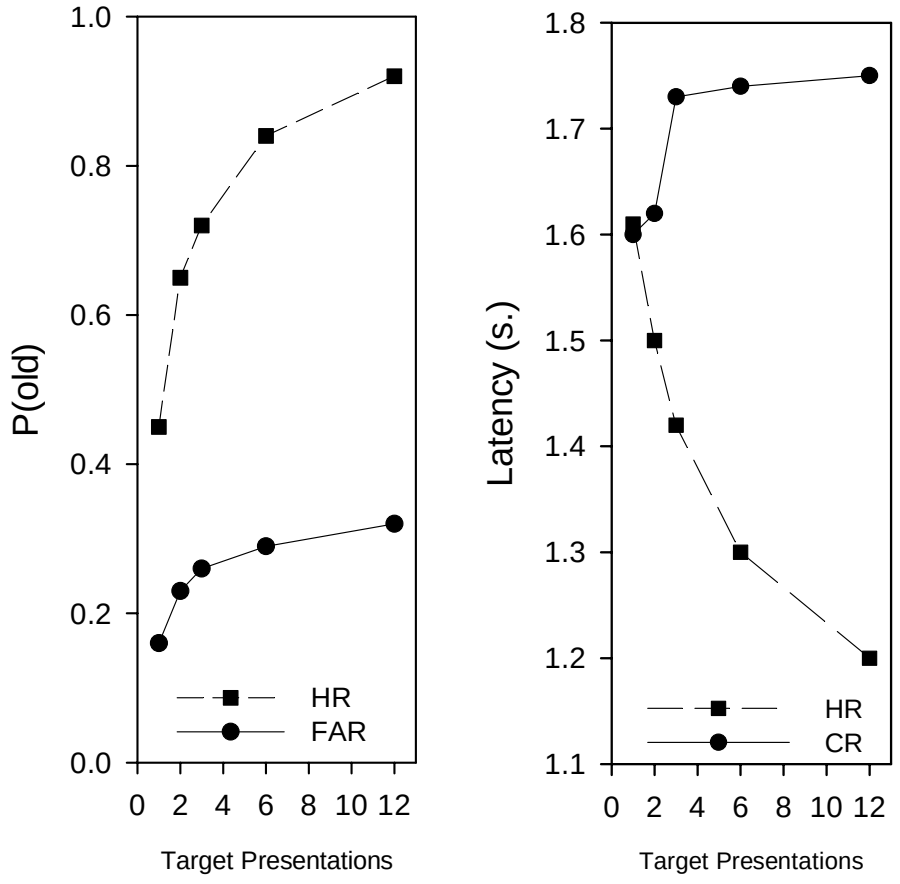
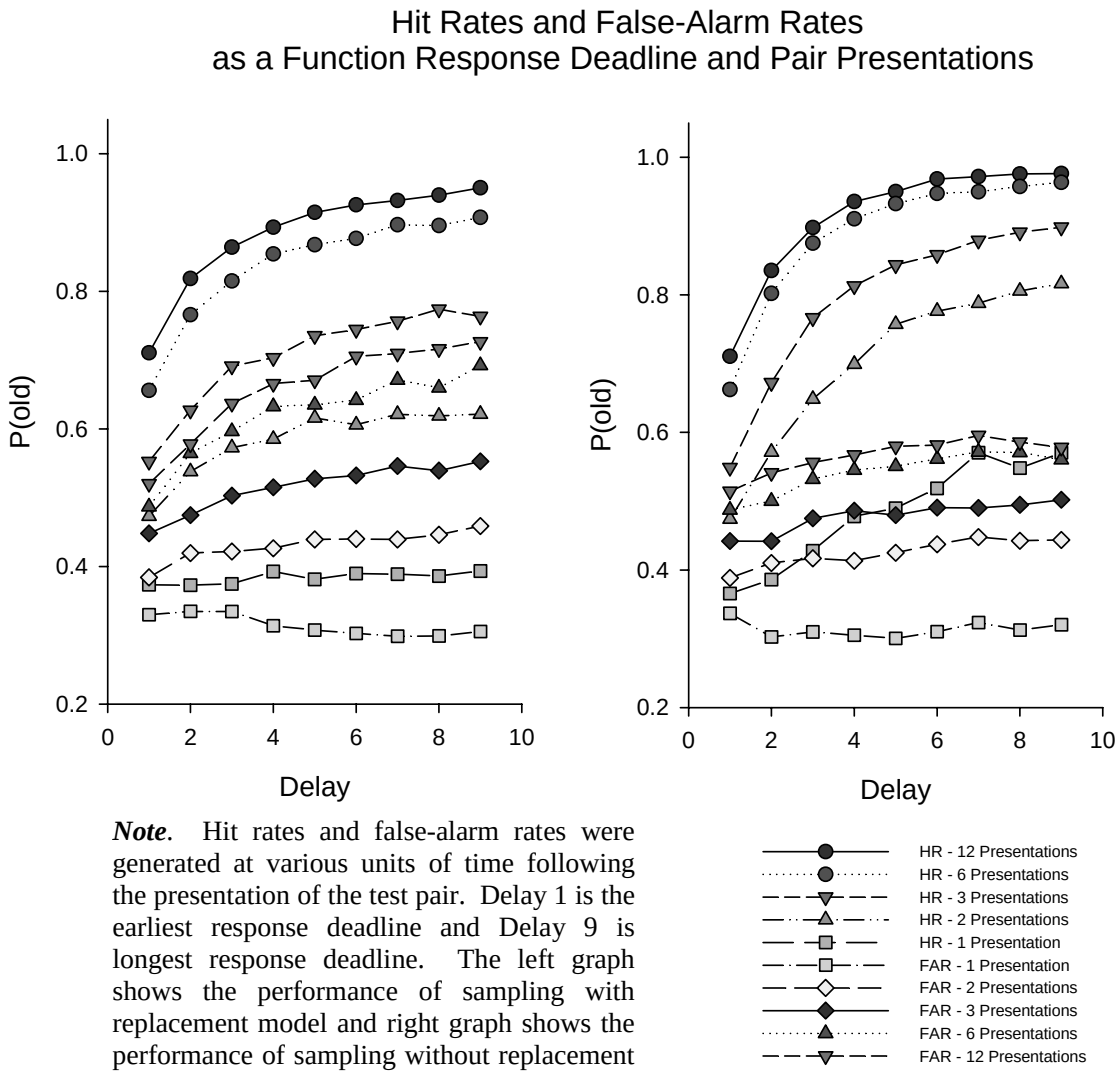
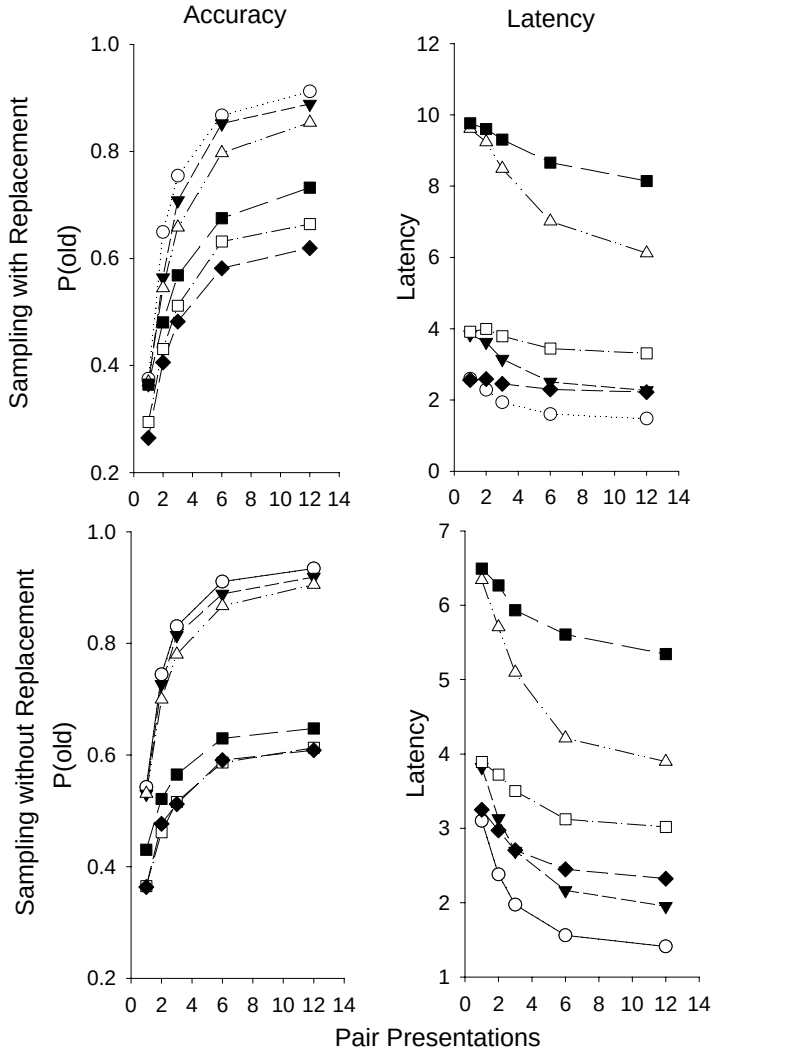

1.0
0.8
0.6
0.4
0.2
0.0
P(old)
HR
FAR
0 2 4 6 8 10 12
Target Presentations
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
Latency (s.)
HR
CR
0 2 4 6 8 10 12
Target Presentations

Figure 2.



Note. Hit rates and false-alarm rates were generated at various units of time following the presentation of the test pair. Delay 1 is the earliest response deadline and Delay 9 is longest response deadline. The left graph shows the performance of sampling with replacement model and right graph shows the performance of sampling without replacement model. The parameter values were: w = 10, t = 8, c = .7, u^* = .04, g = .4, $K_O = K_N$ = 0, s = .25.

Figure 3.



Note. This figure shows the relationship between associative recognition performance and the rate at which evidence accumulates for free response recognition when performance is based on familiarity only. Hit rates and false-alarm rates as a function of the number of target presentations are plotted in the left panel, and responses latencies are plotted in the right panel. The top row shows performance based on a sampling with replacement model and the bottom row shows performance with a sampling without replacement model. Values of *s* are .10, .25, and .40. The parameter values are: w = 10, t = 8, c = .7, g = .4, K_O = 1.0, K_N = -1.0.

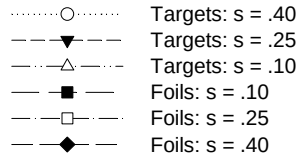
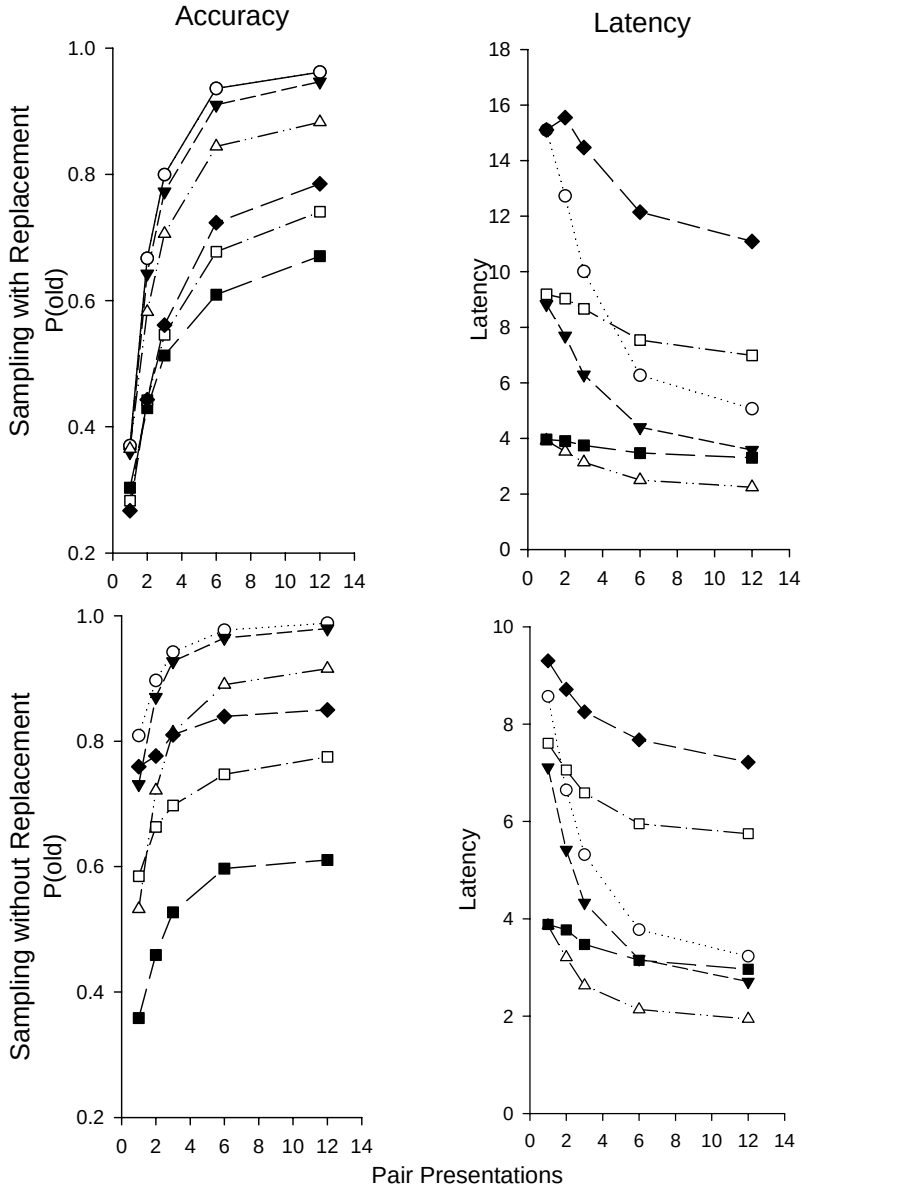


Figure 4.



Note. This figure shows the relationship between associative recognition performance and the rate at which evidence accumulates for response recognition when performance is based on familiarity only. Hit rates and false-alarm rates as a function of the number of target presentations are plotted in the left panel, and responses latencies are plotted in the right panel. The parameter values were: w = 10, t = 8, c = .7, g = .4, s = .25.

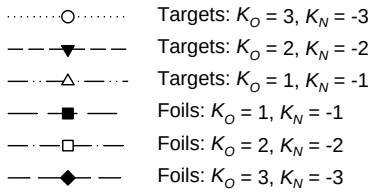
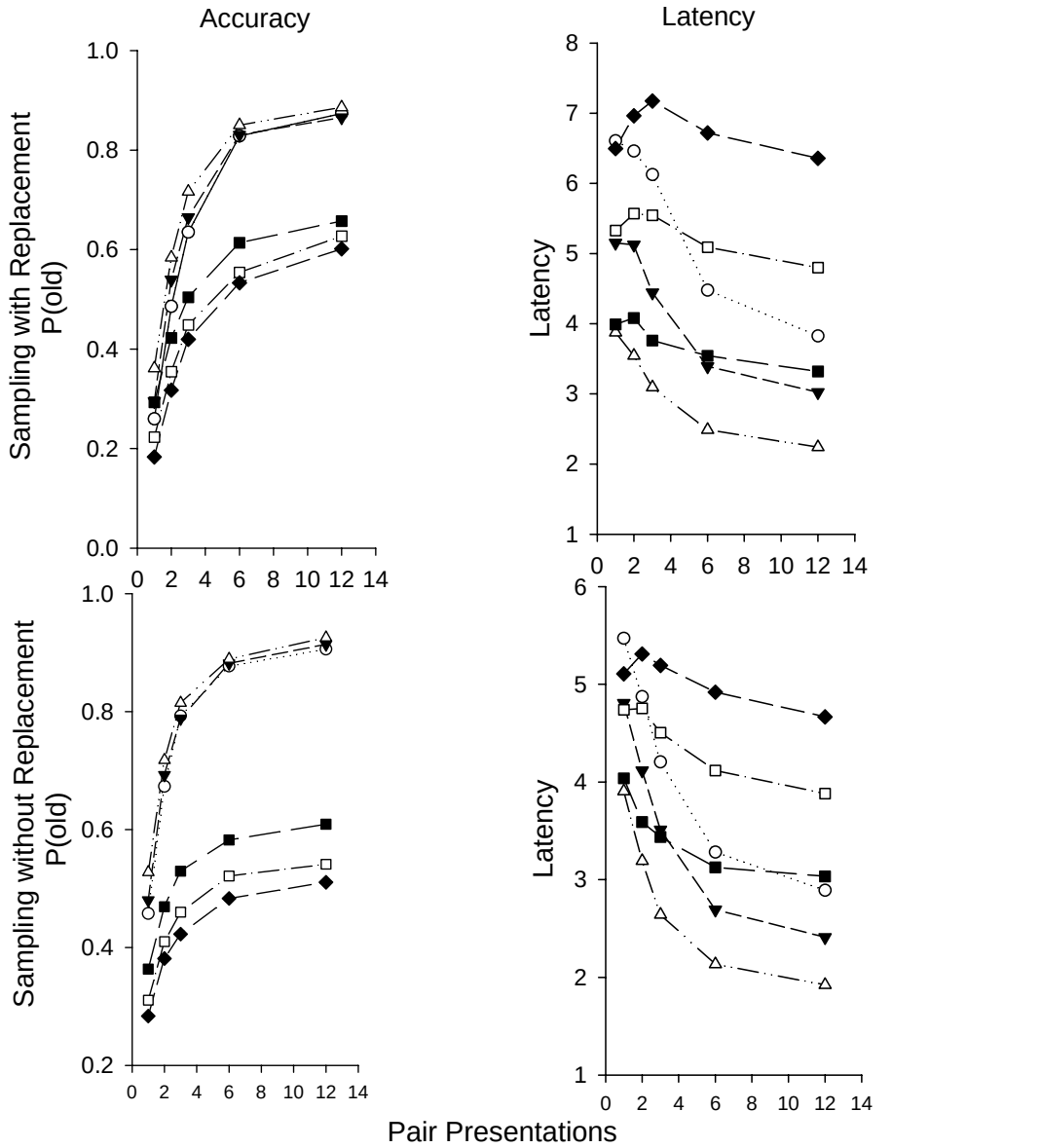
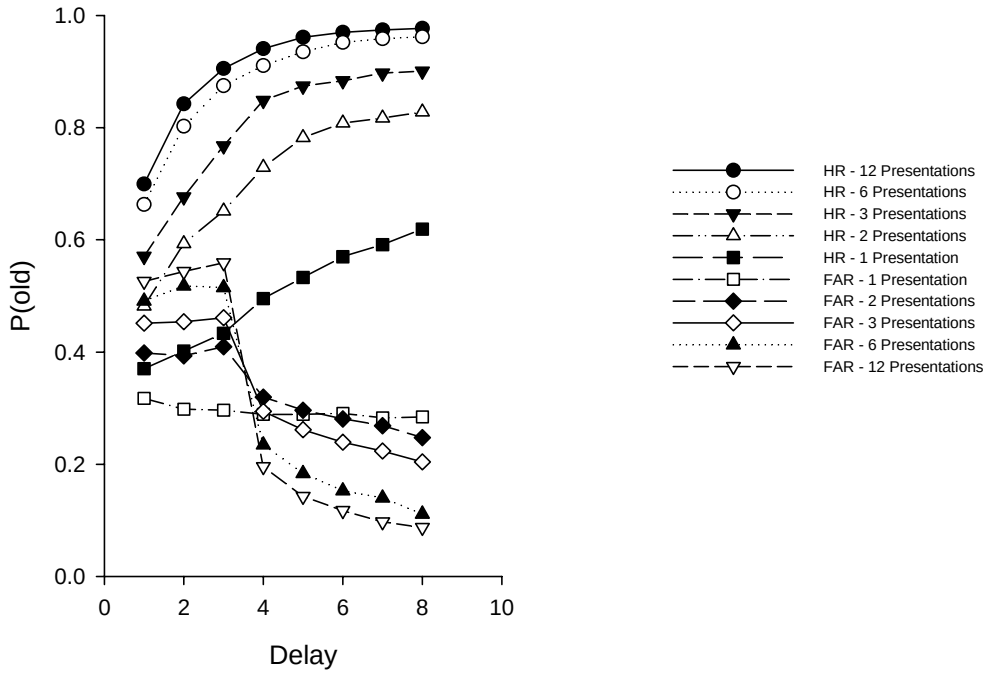


Figure 5.



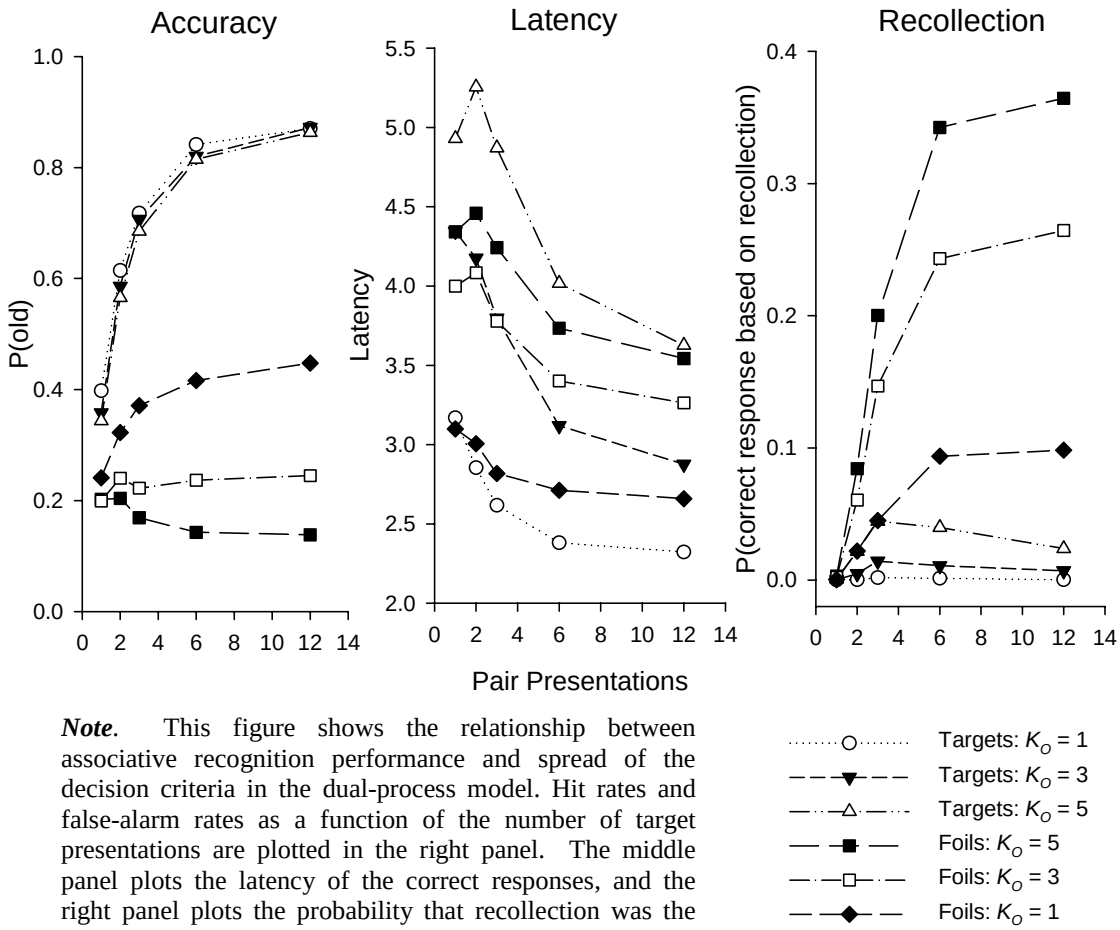
Note. This figure shows the relationship between associative recognition performance and the spread between the old and new criteria. The spread is implemented in an asymmetrical manner. Hit rates and false-alarm rates as a function of the number of target presentations are plotted in the left panel, and responses latencies are plotted in the right panel. The parameter values were: w = 10, t = 8, c = .7, g = .4, K_O = 1.0, K_N = -1.0 s = .25.

Figure 6



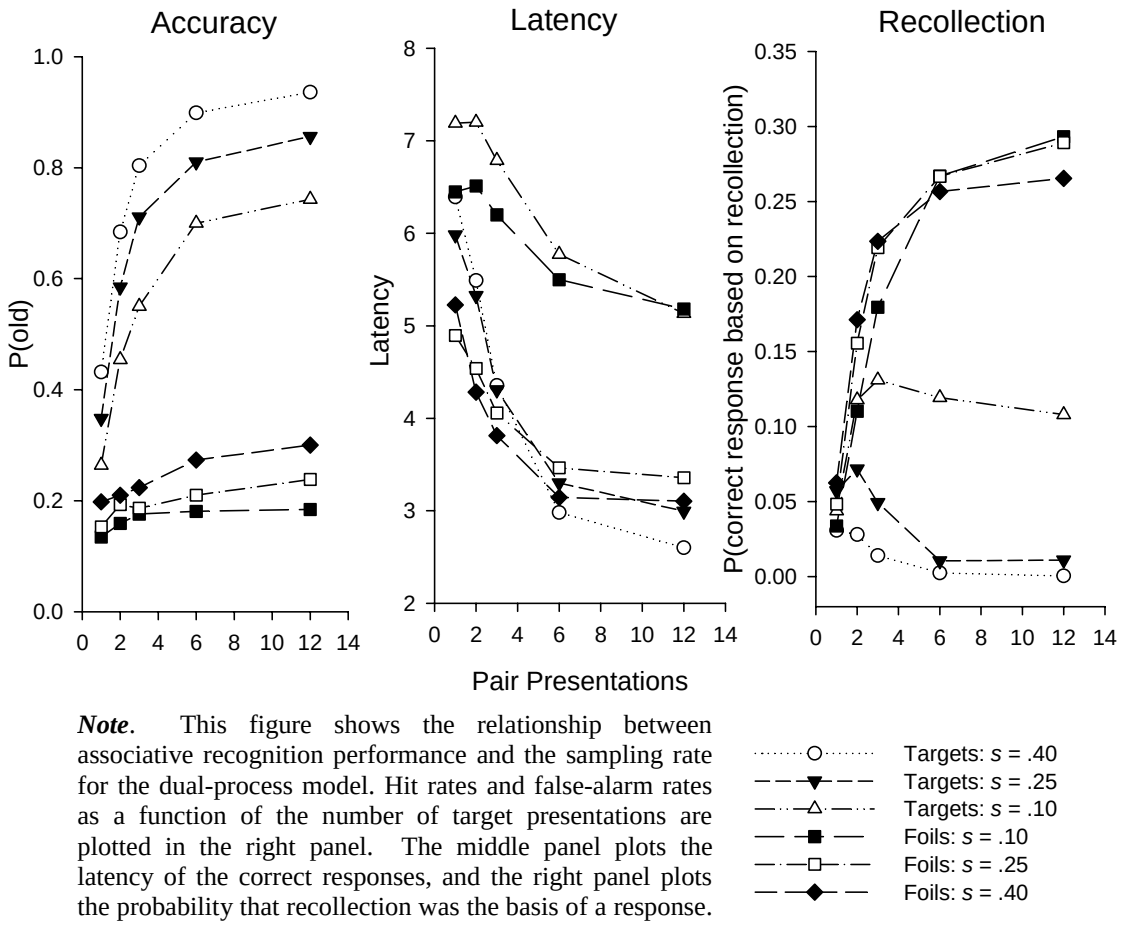
Note. Hit rates and false-alarm rates were generated at various units of time following the presentation of the test pair. Delay 1 is the earliest response deadline and Delay 9 is longest response deadline. This model assumes at the shortest response deadlines (Delays 1-3) responses are based solely on the amount of familiarity accumulated to that point. At the longer response deadlines, hit rates and false-alarm rates are based on a mixture of responses based on the amounts of familiarity and recollected details to that point in time. The parameter values were: w = 10, t = 8, c = .7, u^* = .04, g = .4, $K_O = K_N$ = 0, K_R = 5, s = .25.

Figure 7.



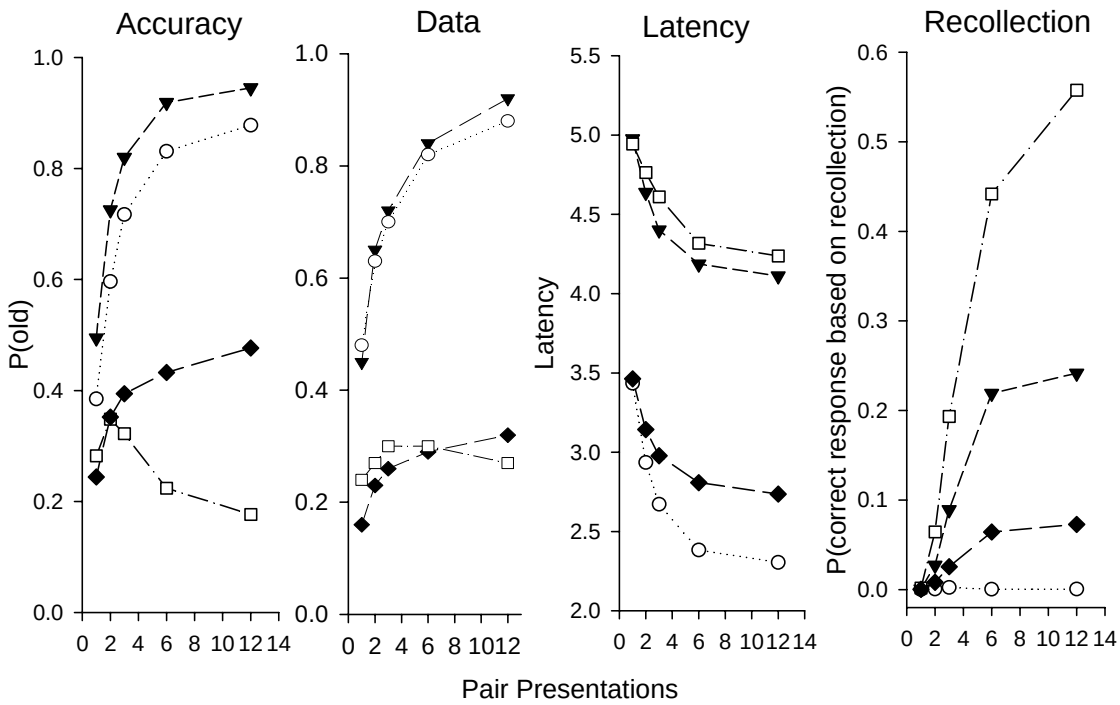
Note. This figure shows the relationship between associative recognition performance and spread of the decision criteria in the dual-process model. Hit rates and false-alarm rates as a function of the number of target presentations are plotted in the right panel. The middle panel plots the latency of the correct responses, and the right panel plots the probability that recollection was the basis of a response. The parameter values were: w = 10, t = 8, c = .7, g = .4, K_N = -.5, K_R = 10, s =.25.

Figure 8.



Note. This figure shows the relationship between associative recognition performance and the sampling rate for the dual-process model. Hit rates and false-alarm rates as a function of the number of target presentations are plotted in the right panel. The middle panel plots the latency of the correct responses, and the right panel plots the probability that recollection was the basis of a response. The parameter values were: w = 10, t = 8, c = .7, u^* = .04, g = .4, K_O = 3, K_N = -.5, K_R = 5.

Figure 9.



Note. This figure shows the relationship between associative recognition performance and the delay in a free response. Hit rates and false-alarm rates as a function of the number of target presentations are plotted in the left panels. The left-most panel shows the performance of the model and the next panel shows the data from Malmberg and Xu (in press). The middle-right panel plots the latency of the correct responses, and the right-most panel plots the probability that recollection was the basis of a response. The parameter values were: w = 10, t = 8, c = .7, g = .4, K_O = 1.0, K_N = -.5, K_R = 10, s =.25.

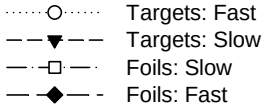
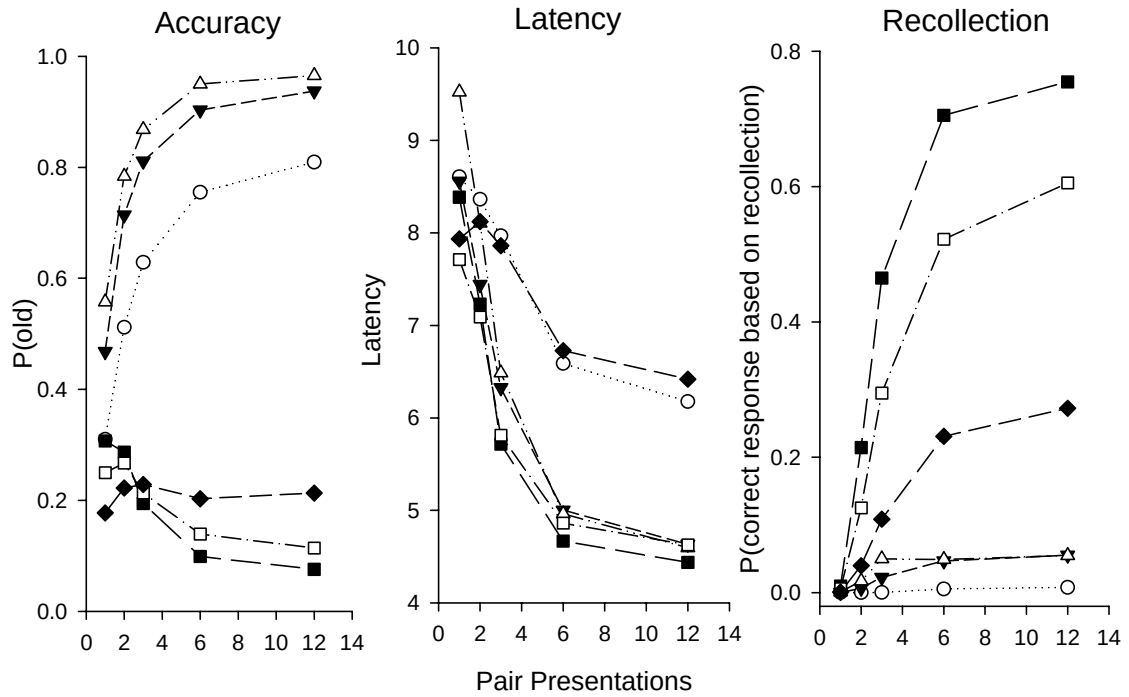
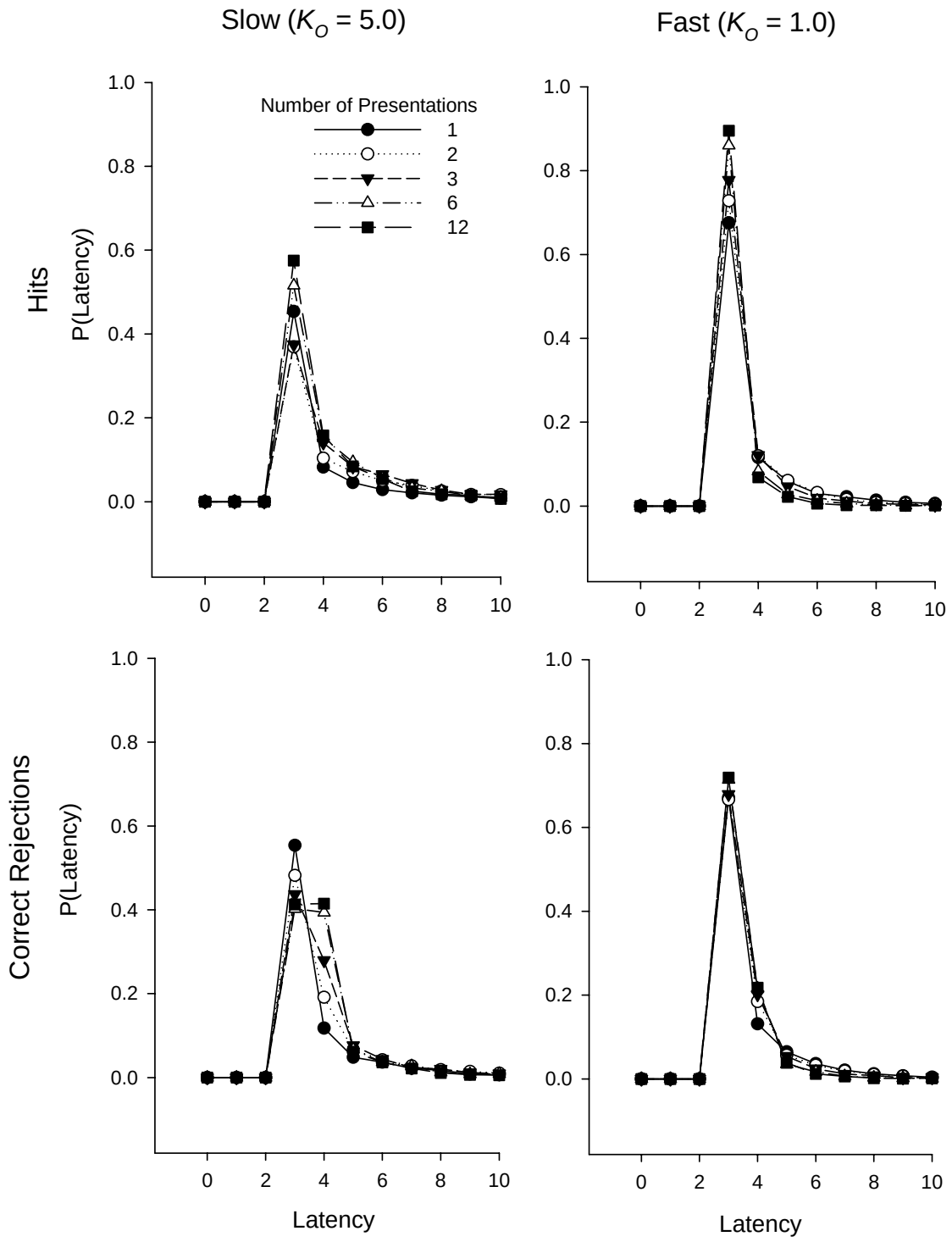


Figure 10



Note. This figure shows the relationship between associative recognition performance and the sampling rate for the dual-process model when responses are assumed to be delayed until the subjective estimate of when recollection should be completed. Hit rates and false-alarm rates as a function of the number of target presentations are plotted in the right panel. The middle panel plots the latency of the correct responses, and the right panel plots the probability that recollection was the basis of a response. The parameter values were: w = 10, t = 8, c = .7, u^* = .04, g = .4, K_O = 3, K_N = -.5, K_R = 10.

Figure 11.


Slow (K_O = 5.0)
Fast (K_O = 1.0)
Number of Presentations
1
2
3
6
12
Hits
Correct Rejections
P(Latency)
Latency
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10