Available online at www.sciencedirect.com
ScienceDirect
Cognitive Psychology xxx (2008) xxx–xxx
Cognitive Psychology
www.elsevier.com/locate/cogpsych

Recognition memory: A review of the critical findings and an integrated theory for relating them

Kenneth J. Malmberg

Department of Psychology, PCD 4118G, University of South Florida, 4202 East Fowler Avenue, Tampa, FL 33620, USA

Accepted 20 February 2008

Abstract

The development of formal models has aided theoretical progress in recognition memory research. Here, I review the findings that are critical for testing them, including behavioral and brain imaging results of single-item recognition, plurality discrimination, and associative recognition experiments under a variety of testing conditions. I also review the major approaches to measurement and process modeling of recognition. The review indicates that several extant dual-process measures of recollection are unreliable, and thus they are unsuitable as a basis for forming strong conclusions. At the process level, however, the retrieval dynamics of recognition memory and the effect of strengthening operations suggest that a recall-to-reject process plays an important role in plurality discrimination and associative recognition, but not necessarily in single-item recognition. A new theoretical framework proposes that the contribution of recollection to recognition depends on whether the retrieval of episodic details improves accuracy, and it organizes the models around the construct of efficiency. Accordingly, subjects adopt strategies that they believe will produce a desired level of accuracy in the shortest amount of time. Several models derived from this framework are shown to account the accuracy, latency, and confidence with which the various recognition tasks are performed.

Keywords: Recognition; Memory; Memory models

1. Introduction

The nature of recognition memory has been hotly debated (e.g., Macmillan & Rotello, 2006; Malmberg, Holden, & Shiffrin, 2004; Murdock, 2006; Park, Reder, & Dickison,

E-mail address: malmberg@cas.usf.edu

0010-0285/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.cogpsych.2008.02.004

2005; Rotello, Macmillan, & Reeder, 2004; Parks & Yonelinas, 2007; Wixted, 2007), and several reviews have recently appeared (Diana, Reder, Arndt, & Park, 2006; Dunn, 2004; Wixted & Stretch, 2004; Yonelinas, 2002). The main questions are whether recognition is based on one random variable or two random variables and whether these variables differ in how they are generated. In all cases, the goal of the reviews has been to advocate for either a single-process or a dual-process model of recognition. To whit, some reviews argue for the sufficiency of a dual-process model: *Models of recognition: A review of arguments in favor of a dual-process account* (Diana et al., 2006). Other reviews defend the single-process position: *In defense of the signal detection interpretation of remember/know judgments* (Wixted & Stretch, 2004). Still other reviews assumed which is the correct model: *The nature of recollection and familiarity: A review of 30 years of research* (Yonelinas, 2002).

Many of the conclusions drawn from these reviews undoubtedly have merit, but the arguments have not produced a consensus on the "correct" model. A fair reading of them produces a sense that each side can argue points that are consistent with their position. Therefore, we need a broader assessment in order to highlight the weaknesses of these positions. This review focuses primarily on what many would refer to as "long-term" recognition memory, which involves the testing of memory for a relatively large number of stimuli after a filled delay. The review focuses on the relationship between single-item recognition, plurality discrimination, and associative recognition because these tasks place strong constraints on the models. The review also considers both the accuracy and the retrieval dynamics of these tasks because this combination places even stronger constraints on the models. Strong constraints are valuable because they expose the limitations of the models and inspire one to organize the models themselves. In the end, I will propose a framework that in some ways blurs traditional distinctions and can probably account for a wider range of empirical findings than any existing approach. I will begin with a short discussion that lays a foundation for the review of the models and data, which is to follow.

1.1. Levels of explanation and the categorization of data

Modeling recognition is conducted at two levels of understanding. At the higher level, measurement modeling assesses changes in the state of memory versus changes in response bias or assesses the contributions of the different types of information to performance. Measurement models make no assumptions about how memories are acquired, retained, or retrieved. Indeed, some of them are useful when describing the performance of perception tasks. Therefore, they make relatively few testable predictions about the nature of memory, but when they do, there is an opportunity to disconfirm a potentially large number of models that they supersede.

The lower level of modeling is process modeling. Process models describe how memories are acquired, represented, and retrieved. They are classified based on the measurement model with which they are associated given their specific assumptions about encoding processes, representational structures, and retrieval processes. With these assumptions, it is sometimes possible to make *a priori* predictions about how different factors will affect performance and how they will interact with other factors.

The two modeling approaches compliment each other insofar as measurement models address the question of "What is possible?" and process models address the question "How is it possible?" (cf. Batchelder & Riefer, 1999). For instance, a signal-detection measurement model might propose that recognition is based on a continuous random variable corresponding to the strength with which an event is represented, and a global-matching process model might specify how that strength is generated from memory. If recognition cannot be based on a continuous random variable, then process models that assume so are disconfirmed. On the other hand, evidence disconfirming a specific global-matching model does not necessary disconfirm the class of continuous-state models because the process models themselves can be classified. If, however, there are no viable process models, the lack of understanding of the memory system that it seeks to describe undermines the measurement model. Thus, our levels of understanding are linked hierarchically.

In this review, I will consider how both measurement models and process models describe recognition memory. I specifically seek the boundary conditions on the viability of different models vis-à-vis the factors and the tasks for which they readily account. Estes (1975), for instance, advocated the use of models to classify empirical phenomena as a means for understanding the relationships between the phenomena. Phenomena that are explained by one class of models are assigned to one category, and the phenomena that are more readily explained by a different class of models are assigned to another category. The different classes of phenomena are understood by specifying a relationship between the models that describe them.

While not a novel approach to reviewing the literature (cf. Gillund & Shiffrin, 1984), this approach is somewhat different from what has become the convention. In recent years, model development and comparisons have often focused on the ability to maximize the fit of a model to the data. The winning model is the one that best fits the data from a particular experiment. However, the focus on goodness-of-fit tests can lead to narrow models with respect to the phenomena that they explain, and hence the relationships between different phenomena could become murky as the relationship between the models appear overly extreme. For instance, Lewandowsky and Heit (2006) admonished, "Notwithstanding the repeated [quantitative] tests of those models within their particular domain, there appears to be little systematic categorization of phenomena into those that one theory can handle versus those that are best accommodated by another" [p. 442]. There is no question that such endeavors are important, but the abundance of recent review articles attest to the fact this isolationist approach has not led to a consensus view.

I believe it is worth reconsidering the more expansive approach to model development and comparison advocated by Estes (1975),

> "We might...cultivate just a bit more dissatisfaction with the assumption that if we attend devotedly enough to fitting of models to data the problem of generality will take care of itself. Or with the assumption that, if we devote most of efforts to dealing in isolation with measurement or with substance, with structure or with process, these strands will magically come together to form harmonious theories" [p. 279].

Estes accurately predicted the often-fractured state of the recognition memory literature.

Here, the models of three tasks partition the findings: single-item recognition, plurality discrimination, and associative recognition. I will also consider how the models handle free response recognition and signal-to-respond recognition performance. What emerges is that some models are better suited for explaining some phenomena than others. This is the case at both the measurement level and the process level of modeling. Finally, I will propose that different models are related via the construct of *efficiency*. It assumes that there are multiple recognition strategies available to the subject, and that the efficient subject chooses the strategy that achieves a desired level of accuracy in the shortest amount of time. Thus, I seek a general framework for relating models, and I therefore place less emphasis in this review on the importance of precise, quantitative fits of specific results.

1.2. Recognition as the detection of a signal embedded in noise

Several models of are based on signal-detection theory (Green & Swets, 1966; e.g., Banks, 1970; Bernbach, 1970; Kintsch, 1967; Lockhart & Murdock, 1970): the recognition judgment is based on a comparison of a continuous random variable obtained from memory to a criterion (the upper panel of Fig. 1). The random variable is often conceptualized as *familiarity*, and the average familiarity of a target is greater than the average familiarity of a foil since targets were studied in the specified context. The subject determines the placement of the criterion. When it is set at strict levels, hits and false-alarms are lower when compared to those observed when the criterion is set at levels that are more lenient.

A dynamic version of the signal-detection model was developed by Ratcliff (1978; also see Ratcliff and Murdock, 1976; Murdock, 1985) to discriminate between the effects of bias and sensitivity on the accuracy and latency of recognition judgments. According to the random walk model, positive and negative evidence from the underlying distributions accumulates over time at a variable rate. The evidence is compared to two decision criteria: a higher old criterion and a lower new criterion. When the evidence exceeds one of the criteria, the subjects makes a response. Increasing the difference between the amount of positive and negative evidence improves sensitivity and decreases latency. On the other hand, increasing the distance between the old and new decision bounds improves accuracy but with a time cost. Bias is modeled by the assumption that the subject determines the point at which evidence begins to accumulate. Increases in bias to respond "old" increase the HR and the FAR, decrease the latency of "old" responses, and increase the latency of "new" responses. With a few notable exceptions (Diller, Nobel, & Shiffrin, 2001; Hockley & Murdock, 1987; Murdock, 1985; Ratcliff, 1978; Ratcliff & Murdock, 1976), the recognition memory literature has been concerned only with how different factors affect the accuracy of recognition memory. As such, the recognition literature has virtually ignored the random walk model.

Signal-detection models have been successful at measuring how different factors influence accuracy (or bias), and the number of signal-detection models of recognition is growing. They have been extended to discriminate between the contributions of item and associative familiarity (Kelley & Wixted, 2001) and specific and global strength (Rotello et al., 2004). Other models account for the effects of item similarity on recognition memory (Heathcote, Raymond, & Dunn, 2006), and they have been extended to account for source memory (Qin, Raye, Johnson, & Mitchell, 2001; Slotnick, Klein, Dodson, & Shimamura, 2000) and the relationship between recognition memory and source memory (Banks, 2000).

In the 1980s, researchers developed models of the processes and representations that produce familiarity. These models are referred to as global-matching models (Clark & Gronlund, 1996; Gillund & Shiffrin, 1984; Hintzman, 1988; Humphreys, Bain, & Pike, 1989; Murdock, 1982). Each is an instance of the signal-detection class of models because the basis of the recognition judgment is a continuous random variable. Accordingly, a familiarity value is obtained by comparing a temporary representation of the test item (i.e., retrieval cue) to the contents of memory. This typically involves a large number of traces. The familiarity that is associated with a retrieval cue is a positive function of the

Please cite this article in press as: Malmberg, K. J., Recognition memory: A review of the..., *Cognitive Psychology* (2008), doi:10.1016/j.cogpsych.2008.02.004

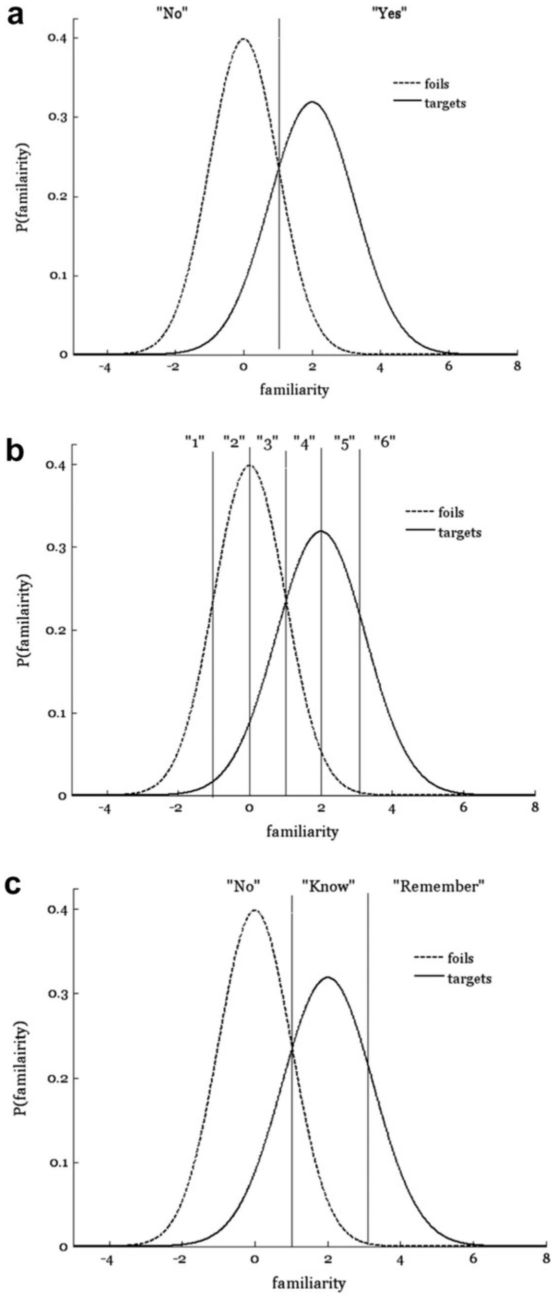

Fig. 1. Signal-detection models of various recognition memory tasks. *Note:* From top to bottom, the yes–no, ratings, and RK signal-detection models are displayed.

similarity between it and traces stored in memory. Because targets are relatively similar to at least one trace in memory, targets tend to be more familiar than foils.

One of the advantages of the global-matching models is that they are part of a larger global-memory framework that accounts for a variety of memory phenomena with just a few assumptions. Over time, however, a number of findings proved to be problematic for the global-matching models. These included mirror effects (Glanzer & Adams, 1985); the shape of the recognition memory receiver operating characteristic (Ratcliff, Sheu, & Gronlund, 1992), and the relationship between list-length and list-strength manipulations (Ratcliff, Clark, & Shiffrin, 1990). In recent years, several new global-matching models were developed including the bind-cue-and-decide model (BCDMEM, Dennis & Humphreys, 2001), the retrieving effectively from memory model (REM; Shiffrin & Steyvers, 1997, 1998), the subjective likelihood model (SLiM; McClelland & Chappell, 1998), and the theory of distributed associative memory model (TODAM; Murdock, 1997, 2006).

BCDMEM, REM, and SLiM belong to a subclass of global-matching models that are motivated by the assumption that cognitive systems have evolved to be optimal and adapt over time to their particular surroundings. Optimality is cast in a Bayesian system that seeks to maximize accuracy given the available information on which to make a decision. Their computations of familiarity dictate that strengthening operations simultaneously increase the similarity between a target cue and its trace in memory and decrease the similarity between the target trace and other traces in memory.[1] This is referred to as *differentiation* (Criss, 2006), and in one way or another, it allowed the new generation of global-matching models to account for the findings that had proven difficult for the prior generation of models. Moreover, the Bayesian approach provides a principled location of the subjective criterion, and thus, the new global-matching models can account for more data than their predecessors can even as more restrictions are placed on the manner in which they do so.

A variant of the single-process model is the *dual-cue model* (Humphreys, 1976, 1978; Humphreys & Bain, 1983; Humphreys et al., 1989; Kelley & Wixted, 2001; Murdock, 2006). It emphasizes the assumption that access to memory depends on the nature of the information used to probe or cue memory (Tulving & Thomson, 1973). One version, assumes that different cues are used to access different types of information, and at the measurement level, the goal is to measure the relative contributions of item and associative (or relational) familiarity to performance (Humphreys, 1976, 1978; Humphreys & Bain, 1983; Kelley & Wixted, 2001; Murdock, 2006). At the process level, dual-cue models are often used to account for the relationship between single-item recognition and associative recognition (Criss & Shiffrin, 2005; Humphreys et al., 1989; Murdock, 1982, 1997). When item and associative information are independent, these models are referred to as *independent cue models*, and different cues are used to access the different types of information.

1.3. Threshold models of recognition memory

Threshold models of recognition memory come in variety of forms, but they only differ from signal-detection models in two important ways (Krantz, 1969; Macmillan & Creelman, 1990). Threshold models assume that the basis of a recognition decision is the state in which memory resides subsequent to a probe of memory. There are only a small number

[1] Glanzer & Adams (1990) took a different Bayesian approach. Their ALT model accounted for a number of findings, including mirror effects, but the local access assumption proved to be a problem (Malmberg & Murnane, 2002).

of mutually exclusive memory states, and one of them is sufficient for describing the state of memory at a given point in time. The second difference between signal-detection and threshold models is that bias reflects the probability that the subject guesses "yes" in the absence of information that the test item was studied, and a variety of different guessing strategies can be assumed (Erdfelder & Buchner, 1998; Krantz, 1969; Malmberg, 2002).

The number of memory states that are assumed defines threshold models (Macmillan & Creelman, 1990). The high-threshold model assumes that there are two internal states. One state represents that an item has been detected in memory and other represents that item has not been detected in memory. The critical assumption is that there is high threshold that must be achieved in order for an item to be detected and only targets may achieve that threshold with a certain probability. When target items do not achieve the threshold, memory falls into a not-detected stated. All foils end up in the not-detected state, but the subject may nevertheless guess "yes" because not achieving the high threshold is an insufficient reason to determine that a test item was not studied (i.e., some targets end up in the not-detected state even though they were studied). Hence, false alarms are the result of the tendency to guess "yes".

The double high-threshold model assumes that there are three internal states: detect-old, detect-new, and an indeterminate state. Two high thresholds exist; one threshold only targets achieve and one threshold only foils achieve, which results in memory being in the detect-old or the detect-new state, respectively. When targets and foils do not achieve their threshold, memory falls into an indeterminate state. Varieties of response strategies are available in the double high-threshold model. The simplest model assumes that correct responses are made when memory is in the detect-old state for targets or the detect-new state for foils, and the false-alarms and misses are the result of guesses that occur when memory is in the indeterminate state (for other response strategies see Erdfelder & Buchner, 1998; Krantz, 1969; Malmberg, 2002).

Threshold models are commonly used to describe the performance of recall tasks (e.g., Gillund & Shiffrin, 1984; Murdock, 1982). Threshold models have also been used occasionally to account for associative recognition and source memory performance (Diller et al., 2001; Humphreys, 1976; Yonelinas, 1999). For instance, recognition is sometimes viewed as a less demanding task than recall (e.g., Bahrick, 1965). Support for this assumption primarily comes from the fact that recognition usually is more accurate than recall, perhaps because accurate recognition does not necessarily require the production of specific information from memory and recall does. However, today threshold models are not widely accepted as viable models of single-item recognition, and they have not often inspired the development of process-level models. We now know, for instance, that there are some conditions under which items that are not recognized are recalled (e.g., Humphreys, 1978; Mandler, 1980; Tulving & Thomson, 1973). This suggests that recognition does not simply place lesser demands on memory than recall, and the question of whether or not recognition is based on a generative process is in fact one the bones of contention in the current literature.

Another reason threshold models are not widely embraced is because they often predict that factors that affect recognition performance should also affect recall performance in the same way. In some cases, the prediction holds. However, it is interesting and perhaps important to note that almost all the factors that have similar effects on recall and single-item recognition affect the encoding of information in memory. For instance, manipulations of item strength (Ratcliff et al., 1990), the distribution of study (Hintzman, 1974), orienting tasks (Craik & Tulving, 1975), context variability (Cook, Marsh, & Hicks, 2006; Hicks, Marsh, & Cook, 2005; Steyvers & Malmberg, 2003), and maybe list-length (Murnane & Shiffrin, 1991; Strong, 1912; but see Dennis & Humphreys, 2001) all have similar effects on recognition and recall.

There are, however, factors that differently affect single-item recognition and recall. Perhaps the most well known of these is normative word-frequency (Balota & Neely, 1980; Clark & Burchett, 1994; Deese, 1960; Dorfman & Glanzer, 1988; Gorman, 1961; Gregg, 1976; Maddox & Estes, 1997; Malmberg & Murnane, 2002; Schulman, 1967; Shepard, 1967; Wixted, 1992). In addition, aging (Balota, Dolan, & Duchek, 2000; Craik & McDowd, 1987), alcohol intoxication (Nelson, McSpadden, Fromme, & Marlatt, 1986; Soderlund, Parker, Schwartz, & Tulving, 2005), mnemonic organization (Mandler, Pearlstone, & Koopmans, 1969), emotionality (Hamann, 2001; Hertel & Parks, 2002; Windmann & Kutas, 2001), serial position (Lehman & Malmberg, submitted for publication; Murdock & Anderson, 1975), incidental learning (Estes & DaPalito, 1967), and list strength (Ratcliff et al., 1990) all differentially affect recognition and recall. Other findings show an interaction between single-item recognition versus recall and category size manipulations, such that interference is observed for cued recall but free recall and recognition accuracy are unaffected by category size (Gillund & Shiffrin, 1984; Mandler et al., 1969; Raaijmakers, 1979; Shiffrin, Huber, & Marinelli, 1995; Tulving & Pearstone, 1966). Thus, the interactions of variables with single-item recognition versus recall are large in number and diverse in nature. To date, no one has ever offered a coherent threshold-based explanation of any of these interactions.

Recently some have proposed that associative recognition tasks rely on the same processes and representations involved in recall. For instance, Yonelinas (1999) & Diller et al. (2001) proposed associative recognition is based on the same recall-like representations and mechanisms that underlie cued recall (also Humphreys, 1978; Mandler, 1980). Their arguments are not without a certain amount of merit. Yonelinas observed that the pattern of confidence judgments for associative recognition is consistent with a threshold process. Diller et al. observed the latencies of associative recognition judgments are similar to those observed for cued recall but not to those observed for single-item recognition. Thus, a couple of recent findings are consistent with an account of associative recognition that relies on the same model as cued recall.

It is important to note, however, that the finding that some forms of interference affect cued recall but not associative recognition (Anderson & Watts, 1971; Dyne, Humphreys, Bain, & Pike, 1990; Postman & Stark, 1969) does not support the recall-like account of associative recognition. Moreover, other forms of interference affect free recall and single-item recognition but not associative recognition (Criss & Shiffrin, 2005; Gillund & Shiffrin, 1981). Another well-established finding is that increases in retention intervals negatively affect cued recall (Krueger, 1929; Mandler et al., 1969) and single-item recognition (Slamecka & McElree, 1983; Wixted & Ebbesen, 1997) more than associative recognition (Hockley & Murdock, 1987; Hockley, 1992; Weeks, Humphreys, & Hockley, submitted for publication). Last, the recall-like models of associative recognition are disconfirmed by findings that show that the ability to correctly reject rearranged foils remains steady as the pairs are strengthened via spaced repetitions (Cleary, Curran, & Greene, 2001; Kelley & Wixted, 2001; Malmberg & Xu, 2007; Xu & Malmberg, 2007).

In summary, there is relatively little support for a recall-like model of either single-item recognition or associative recognition. In fact, only occasionally a new finding emerges that seems to be consistent with a recall-like model of recognition. Threshold models, however, play an important role in many hybrid models.

1.4. Dual-process models of recognition

By the early 1990s, many single-process global-matching models had been disconfirmed, and the time seemed ripe for a radical change. While some embraced the new Bayesian global-matching approach, others adopted a dual-process model of recognition memory.[2] Accordingly, there are two ways to perform recognition: recognition is based on the familiarity of the retrieval cue or it is based on recollecting episodic details that are informative to the recognition decision. The former process is often modeled as a signal-detection process and the later process is often modeled as a threshold process, and researchers were particularly interested in measuring the relative contributions of recollection and familiarity to performance (e.g., Jacoby, 1991).

Dual-process investigations of recognition memory have usually been concerned with somewhat different issues than single-process investigations. Atkinson & Juola (1974) investigated a dual-process model that was particularly adept at accounting for the response latencies of over-learned items. Hence, the foundations of dual-process theory were in accounting for the dynamics of recognition memory, while single-process models have often ignored reaction times. Mandler (1980, 1991; Mandler et al., 1969), on the other hand, advocated the dual-process approach based on his observation that the organization of episodic memories influenced both recognition and recall and on the correlations between cued-recall and associative recognition performance. Single-process investigations, in contrast have typically been less interested in memory organization, and more interested in memory strength.

Given these findings in the very early literature that did not immediately support the single-process familiarity-based model, why was the debate not settled at that point and why in fact was the global-matching paradigm dominating? To some extent, the equivocation was due to attrition; several leading figures on the dual-process side of the debate moved on to other interests. Another reason is that dual-process models were not as convenient as the single-process model. For instance, the dual-process model did not provide simple ways of distinguishing between sensitivity and bias (Buchner & Erdfelder, 1996; Yonelinas & Jacoby, 1996). Perhaps most importantly, however, dual-process models were only implemented at the measurement level, and hence their explanations suffered from a degree of descriptiveness or circularity.

Unlike the random walk model, however, many explicitly rejected the dual-process model. Some questioned the validity of the methodology used in several influential experiments or the necessity of dual-process model to explain the data (Broadbent, 1973; Broadbent & Broadbent, 1975, 1977). For instance, Gillund & Shiffrin (1984) evaluated the arguments in favor of and against the dual-process model. They found no compelling reason to pursue the more complicated dual-process approach. One of the justifications for this conclusion might have been the historical lack of attention paid to the temporal

[2] Mandler et al. (1969) cite Muller (1913) as the originator of the dual-process model.

dynamics of recognition memory (cf. Atkinson's work). Nevertheless, the relationship between accuracy and latency was important to Murdock (1982; cf. Ratcliff & Murdock, 1976), and he too preferred the global-matching model for his TODAM.

Years later, Yonelinas (1994, 1997, 1999) capitalized on the rediscovery of ROC analyses that was central to disconfirming the first generation of global-matching models (Ratcliff & McKoon, 1991; Ratcliff et al., 1992, Ratcliff, McKoon, & Tindall, 1994). In doing so, he developed a new paradigm for distinguishing between single-process and dual-process models. The production of ROC curves has since become a cottage industry (e.g., Glanzer, Kim, Adams, & Hilford, 1999; Heathcote, 2003; Hilford, Glanzer, Kim, & DeCarlo, 2002; Rottelo, Macmillan, & Van Tessel, 2000; Van Zandt, 2000; Yonelinas, 2002), and thus many of the arguments used to support one view versus the other hinge upon them. The ROC analysis also has enormous implications for how one interprets data from another important dual-process paradigm, remember–know recognition, which I will discuss shortly.

2. Comparing ratings and remember–know measures of recollection

A goal of dual-process theory is to measure the contribution of recollection to recognition performance. In this section, we will discuss two measures of recollection, compare them, and consider the strengths and weaknesses of the alternative single-process interpretations.

One dual-process measure is based on a receiver operating characteristic, or ROC, analysis, which involves the collection of pairs of HRs and FARs that correspond to a fixed level of accuracy at different levels of response bias. The most common method for constructing an ROC utilizes a confidence-rating task (Macmillan & Creelman, 1991). The interpretation of a ratings ROC depends on the assumptions of a theoretical framework. The fundamental single-process assumption is that confidence is a proxy for bias (Egan, 1958): high levels of confidence require greater amounts of evidence than do lower levels of confidence (see the middle panel of Fig. 1). In dual-process models, ratings based on familiarity are made in the same manner. However, the highest confidence "old" response is sometimes reserved for recognition decisions based on recollection (Joordens & Hockley, 2000; Reder et al., 2000; Yonelinas, 1994). This is the *process-pure* assumption. Thus, interpreting the basis for the highest confidence "old" response is a primary difference between single-process and dual-process ratings models.

Yonelinas (2002) related ROC analyses to another dual-process measure of recollection based on the subject's awareness of the current state of memory. If a recognition judgment is based on the recollection of a prior encounter with the stimulus, the answer is, "remember". If a judgment is based on a feeling of knowing that an item was encountered in the specified context, the answer is, "know". This has become known as the *remember–know (RK) procedure* (Gardiner, 1988; Tulving, 1983). Like the ratings ROC, however there is a different way to interpret RK performance. For instance, Donaldson (1996) proposed a signal-detection model in which the subject uses two criteria. Familiarity values exceeding the strictest criterion result in a "remember" response, and familiarity values falling between the two criteria lead to a "know" response (lower panel of Fig. 1). According to the signal-detection model, RK responses reflect the amount of evidence on which the decision was based.

Even though the limitations of introspections are well known, and under most conditions they are not very accurate measures of mnemonic abilities or states (Nelson & Narens, 1990), the empirical question is whether the RK measure is a reliable and valid measure of

Please cite this article in press as: Malmberg, K. J., Recognition memory: A review of the..., *Cognitive Psychology* (2008), doi:10.1016/j.cogpsych.2008.02.004

recollection (Nelson, 1996). One way to address this issue is to compare the abilities of the single-process and dual-process models to fit RK data and the recognition ROC. Unfortunately, such efforts have not to lead to a consensus. Dunn (2004) found that the signal-detection RK model is consistent with a large number of findings, and Wixted & Stretch (2004) argued that many findings from the RK procedure actually favor the single-process model. Others argue that the dual-process model is sufficient (Ardnt & Reder, 2002), and still others argue that neither does a good job (Rotello et al., 2004). Likewise, Yonelinas (1994) noted that the single-process and dual-process models predict different shapes of the recognition ratings ROC, and that the dual-process model provides a better fit. In opposition, others argue that the single-process model provides a superior fit (e.g., Glanzer et al., 1999; Wais, Wixted, Hopkins, & Squire, 2006). Still others conclude that neither model fits the ratings ROC well (Sherman, Atri, Hasselmo, Stern, & Howard, 2003).

2.1. Assessing the measures of recollection

The debate over which model fits the RK and ratings data better is an example of where relying on goodness-of-fit tests has produced accurate models but has not lead to a consensus. A different tactic is to focus on the goals of the models, and determine whether the models achieve them. For instance, dual-process models seek to measure the contributions of recollection and familiarity to performance, and a critical question is whether the dual-process measures of recollection derived from the ratings and RK tasks are valid. Note that the ratings measure is derived from a fit of the model to the ROC, whereas the RK measure is derived from a transformation of metacognitive introspections. If the ratings and RK measures tap the same construct, they will produce the same result. If they are not comparable, we can infer that something is amiss in the dual-process framework, and arguments based on these measures are undermined.

In the process-pure version of the dual-process, ratings model (Rotello et al., 2004), the strictest old rating is reserved for responses based on recollection, but the probability of assigning the highest confidence rating to a foil is usually greater than zero. Therefore, a fitting procedure locates where the ROC intersects the old-item axis of ROC space, and this is an estimate of recollection, R_{ratings} (Yonelinas, 1994). Fig. 2 plots R_{ratings} and three HR–FAR pairs, and $R_{\text{ratings}} \sim .20$. If the dual-process theory measures of recollection are consistent, then the RK measures should equal R_{ratings}.

The RK task produces an estimate of recollection based on the probabilities of responding "remember" and "know". These probabilities correspond to a process-pure measure in the same way that the probability of responding "old" with the highest degree of confidence measures recollection in the ratings procedure. Again, however, the process-pure standard is rarely achieved because subjects tend to respond "remember" in response to both targets and foils (P_{Rem}-target, P_{Rem}-foil), and the fundamental question is: why do subjects say that they remember studying items that were not studied? The response of many investigators is to assume that there is some error in RK decisions and/or that the instructions given to the subjects were not sufficiently clear, and they developed methods for transforming raw data in order to provide RK estimates of recollection and familiarity.

Two RK estimates of recollection have been used. One is the difference between P_{Rem}-target and P_{Rem}-foil (Yonelinas, 2002). Is this RK measure and R_{ratings} (i.e., .20) the same? Consider the middle point in ROC space (Fig. 2), where P_{Rem}-target and P_{Rem}-foil are about .57 and .10, respectively, and the RK estimate of recollection is .47.

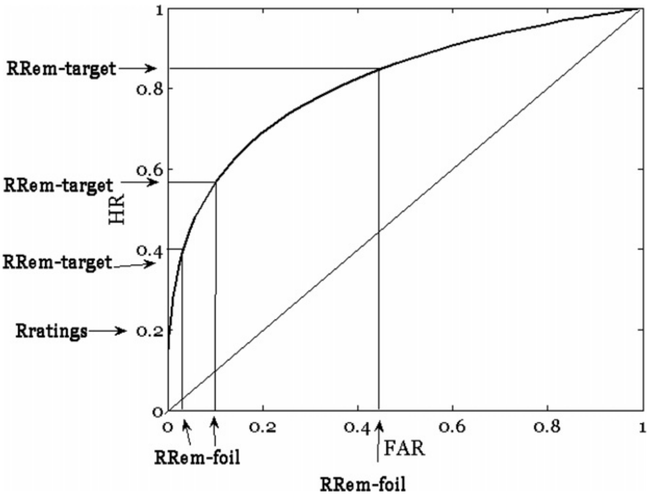

Fig. 2. A typical asymmetrical three-point ROC from which dual-process measures of recollection are computed. *Note:* A three-point receiver operating characteristic (ROC) is displayed. HR and FAR are hit rate and false-alarm rate, respectively. R_{Rem}-target and R_{Rem}-foil are the HRs and FARs from which various measures of RK recollection are derived. R_{rating} is the dual-process estimate of recollection derived from confidence ratings. If the various dual-process measures of recollection measure the same thing, then the measures derived from the RK methods should be the same as the measure derived from the ratings method.

This is about two and half times the estimate obtained from the rating method. However, Yonelinas, Kroll, Dobbins, Lazzara, & Knight (1998) advocated normalizing the difference between P_{Rem}-target and P_{Rem}-foil by $(1 - R_{\mathrm{Rem}}$-foil). In this case, the RK estimate of recollection is: $.47/(1 - .10) = .52$, which is even less in line with R_{ratings}.

The dual-process estimates of recollection derived from the ratings and RK task are not the same, and they are therefore unlikely to measure the same thing. One reason for the difference might be the degree to which each method achieves the process-pure assumption. For instance, Rotello et al. (2004) argued against the process-pure assumption because it predicts that the use of the high-confidence rating and the remember response should be immune to shifts in yes–no response bias, but use of the remember response is affected in predicable ways by factors that influence yes–no bias (Hirshman & Henzler, 1998).[3]

At this point, it is unknown whether one dual-process estimate of recollection is an accurate measure or whether both measures are inaccurate. The ratings measure might be preferred because it produces the same estimate of recollection regardless of which points are used. As for the RK measures, the Yonelinas et al. (1998) method should be preferred over the Yonelinas (2002) method because it converges on the ratings estimate

[3] To see how violations of the process-pure assumption resulting from shifts in bias affect the correspondence between the ratings and RK estimates of recollection, consider the more extreme points in ROC space depicted in Fig. 2: (.85, .45) and (.40, .05). They fall on the same ROC, and hence they should produce the same RK estimates of recollection. However, the estimates of recollection using the Yonelinas (2002) method are .40 and .35, which are very different from the estimate of .20 obtained from the rating method, and they are also different from the estimate of .47 obtained from the middle point on the ROC. The Yonelinas et al. (1998) method does not fair much better; it produces estimates of recollection for the two extreme points of .37 and .72.

as the remember FAR decreases. However, the remember FAR is never truly zero in practice, and thus the RK measures overestimate recollection relative to the dual-process ratings measure.

While there is little support for the assumptions underlying the dual-process measures of recollection, not all arguments against the dual-process model are correct. For instance, Wixted & Stretch (2004) showed that high confidence ''old'' responses and ''remember'' responses are faster than low-confidence ''old'' responses and ''know'' responses for both targets and foils. They suggested that it was difficult for the dual-process model to explain why a slower recollective process would produce faster responses, but Diana et al. (2006) proposed that recollection is always attempted before responding, and therefore recollective responses should be faster than familiarity-based responses. Wixted and Stretch also suggested that signal-detection models easily explain these findings. However, the random walk model is a model of yes–no decisions, and therefore it has nothing to say about the latencies of ratings or RK judgments. Thus, some of these criticisms can be accommodated by dual-process models, and it is unclear what signal-detection model was used to generate the latency predictions.[4]

2.2. Challenges to the signal-detection model

Rotello et al. (2004) concluded that the ratings ROC was asymmetrical and the RK ROC was symmetrical. If true, these findings are equally problematic for single-process and dual-process models. However, Dunn (2004; also Malmberg, Zeelenberg, & Shiffrin, 2004) showed that the single-process signal-detection provided an excellent fit to the same data, and Wixted & Stretch (2004) argued that the difference between ratings and RK ROCs might have been due to differences in how decision noise contributes to their construction. More importantly, the RK and ratings ROCs used by Rotello et al. were based on data averaged over subjects, and the shapes of the ratings and RK ROCs can be differentially distorted by averaging. When properly constructed, the shapes of the ratings and RK ROCs are the same (Malmberg & Xu, 2006).

The single-process model has also been challenged on the grounds that it cannot explain the differences in the shape of the ratings ROCs produced by subjects with selective hippocampal damage. For instance, Yonelinas et al. (1998) observed the well-documented asymmetrical ROC in normal subjects (e.g., Fig. 2), but a more symmetrical ROC in the hippocampus-impaired subjects. This finding is consistent with a dual-process model interpretation; the hippocampus is responsible for recollection and the impairment of the hippocampus caused patients to rely more on a single-process familiarity-based recognition strategy. However, the unequal-variance signal-detection model makes the same prediction as accuracy decreases, and performance of the patients was, of course, lower than performance of the normal subjects. When Wais et al. (2006) controlled for the level of accuracy, the shapes of the control-subject ROCs and hippocampal-impaired ROCs were indistinguishable from each other. Last, the findings of Manns, Hopkins, Reed, Kitchener, & Squire (2003) conflicted with those of Yonelinas et al., and Wixted & Squire

[4] Ratcliff (1978) speculated that the random walk model could be used to model confidence ratings by assuming that the subject sets a set of internal time deadlines, and those test trials that produce relatively fast yes–no decisions are also given relatively high confidence-judgments. If so, confidence is now a proxy for latency and not for bias, and the single-process interpretation of the ratings and RK data is fundamentally altered.

(2004) found that source of the difference was the inclusion of a single outlying recognition score in the Yonelinas et al. experiment.

In summary, both the single-process model and the dual-process measurement model fit the data from RK and ratings experiments reasonably well. The attempts to disconfirm the single-process model have so far failed, and the signal-detection model provides simple, parsimonious measures of sensitivity and bias. Although several arguments against the dual-process model come up short, the dual-process measures of recollection lack the consistency to be convincingly valid. While this does not disconfirm the dual-process theory, it does mean that the dual-process measures do not provide strong evidence for or against any model. In the next two sections, I will describe how the single- and dual-process assumptions fair when implemented in process-level models.

3. Remembering and knowing in global-matching models

Malmberg et al. (2004) described a REM model that is based on Donaldson's (1996) signal-detection model. It was developed in response to the findings of Hirshman, Fisher, Henthorn, Arndt, & Passannante (2002), who varied normative word-frequency and study time in subjects who were under the influence of midazolam or saline during study but not during test. The subjects in the saline condition produced better performance as study time increased and a mirror effect for high-frequency versus low-frequency words was obtained. In addition, higher levels of study time produced greater probabilities of remembering, as did low-frequency words. Under the influence of midazolam, the findings were different: study time had very little influence on recognition accuracy, the low-frequency hit rate advantage was reversed, but the low-frequency FAR advantage was unaffected by midazolam. In addition, the tendency to respond based on remembering was equivalent for high- and low-frequency words. Similar findings concerning the effect of aging on recognition memory have also been reported (Balota, Burgess, Cortese, & Adams, 2002), and these findings were interpreted by many to be consistent with the dual-process assumptions that low-frequency targets are recollected more often than high-frequency targets, and the low-frequency FAR advantage is based on the pre-experimental familiarity of the foils.

However, the REM model accounted for the effects of word-frequency and study time on RK performance. In addition, Malmberg et al. (2004) provided four different explanations for the effect of midazolam on RK performance, the simplest of which is that midazolam affects encoding accuracy, as opposed to the extent of encoding. More recently, Criss & McClelland (2006) speculated that REM might also account for these findings on the later assumption, but this model was rejected because it predicts that midazolam affects both HRs and FARs. On the other hand, Curran, DeBuse, Woroch, & Hirshman (2006) found that midazolam affected both HRs and FARs. Thus, the effect of midazolam on recognition memory is not well-established, but five single-process REM models might account for it, and therefore it poses no special problem.

A TODAM global-matching RK model (Murdock, 2006) is based on a two-dimensional continuous-state model described by Rotello et al. (2004). It extends the TODAM global-matching model to account for the two sources of evidence that such models assume exist. Murdock assumed that recognition judgments based on associative information produce "remember" responses and that recognition judgments based on item information resulted in "know" responses. Murdock compared the results of several TODAM model fits to those obtained by Dunn (2004), and he found that they were nearly perfect.

In summary, REM and TODAM account for RK performance. However, it is perhaps more noteworthy that the global-matching models are quite different, and yet they are successful. To this point, no data have been reported that are inconsistent with them, and it is not possible to disconfirm these models of familiarity at this point. Thus, there are two very viable single-process global-matching models that can account for a wide variety of RK results and that support the single-process measurement model interpretation of RK performance.

4. Remembering and knowing in dual-process models

It is not possible to reject the dual-process account of remembering and knowing solely based on the unreliability of the dual-process theory measures of recollection. Indeed, these measurement issues are not necessarily a major problem for dual-process models like the source-activation confusion model (SAC; Reder et al., 2000). SAC is a process model that is somewhat different from the dual-process measurement model, but different estimates of recollection and familiarity can be derived from it nevertheless.

According to SAC (Reder et al., 2000), there exists for every item a concept node, and the base line activations of these nodes are affected by how frequently and recently a word has been encountered. The level of activation of a concept node is also influenced by the levels of activation of other nodes via the spread of activation on connections that exist between them. The activation of one node spreads to the other nodes, and the amount of spreading activation is positively related to the strength of connections between two nodes and negatively related to the strength and number of the connections between the source node and all other nodes. The activation of a concept node is the basis for familiarity, and hence "know" responses. In addition, context nodes represent the situation in which a concept is encountered. Event nodes encoded during study mediate the connections between concept nodes and context nodes. They represent the occurrence of a concept in a given context, and their activation is the basis for "remember" responses.

The SAC models are able to account for word-frequency effects in continuous recognition (Reder et al., 2000) and study-test recognition (Reder, Angstadt, Cary, Erickson, & Ayers, 2002), list-length and item strength effects (Cary & Reder, 2003), and the list-strength effect (Diana & Reder, 2005). To account for incidental context effects generated with a manipulation of the font in which a word occurs at study and at test, Reder, Donavos, & Erickson (2002) proposed that font nodes connect event nodes and concepts nodes. Last, Park et al. (2005) proposed that plurality nodes represent the plurality of studied words. This allows SAC to account for interaction of word-frequency and word similarity.

While SAC accounts for a wide range of RK findings, SAC may have a difficult time accounting for the recognition memory ROC. According to Diana et al. (2006), "Current attempts to use the SAC model to model ROC data involve fitting each confidence category, using successively more lenient thresholds for both episode and concept nodes. These fits reveal a deficiency in fitting the highest threshold conditions" [p. 17]. Ardnt & Reder (2002) found that the source of the SAC deficiency was in the SAC assumption that both recollection and familiarity are discrete-state processes, and often times such models produce linear ROCs that do not fit the decidedly non-linear recognition ROC. Specifically, SAC has trouble in its current form predicting the use of the highest confidence old rating in response to a foil because there are no links between foil concept nodes and event nodes, which are the source of recollection.

Thus, SAC cannot fit the ratings ROC, and at this point, it is unknown whether SAC can fit the RK ROC (e.g., Rotello et al., 2004). On the obvious assumption that the same model accounts for both the ratings and RK ROC, it might prove to be difficult, and it is noteworthy that the SAC dual-process model comes up short in the very same way that the dual-process measures of recollection come up short.

5. The retrieval dynamics of recognition memory

Reed (1973) introduced to recognition memory research a method for measuring accuracy at various latencies subsequent to the presentation of a test item. Accordingly, subjects make a recognition decision within a small, experimentally controlled time window following a prompt (~200 ms). This procedure is known as the signal-to-respond procedure, and it generates a function relating the latency of responses and the accuracy of those responses.

The pioneering signal-to-respond experiments by Dosher (1981, 1984; also see Ratcliff, 1978) utilized an associative recognition procedure that requires the discrimination of intact and rearranged pairs of items. Intact pairs are items that were studied together, and rearranged pairs are items that were studied but not studied together. The task of the subject is to respond "yes" to intact pairs and respond "no" to rearranged pairs.

A major difference between associative recognition and single-item recognition is that intact and rearranged pairs are similar to each other, and the similarity of the targets and the foils influences the predictions of the global-matching models. First, intact and rearranged pairs should be relatively difficult to discriminate. The more similar the retrieval cue is to the contents of memory, the more familiar the item(s) associated with the retrieval cue will seem, and hence, rearrange pairs will, on average, seem more familiar than new items. Second, HRs and rearranged FARs should steadily increase with increases in response lag as more matching information is accumulated, and FARs for new items should decrease as more mismatching information is accumulated.

Dual-process models usually assume that generating episodic details takes more time than the accumulation of evidence that is the basis for familiarity. The rejection of foils based on recalling some aspect of a study event is often referred to as *recalling-to-reject*. For associative recognition, for instance, one might recall that one item of a rearranged pair was studied with a different item, and use that information to reject the test pair. Thus, an initial increase in FARs over short lags due to increases in familiarity should be followed by a decrease in FARs due to the rejection of rearranged pairs.

Independent cue models make a similar prediction. They assume that associative recognition is based on associative information that is independent of the item information that comprises the pair. When combined with a dual-cueing approach that assumes the formation of an associative cue takes longer than the formation of an item cue, the models predict that FARs should initially increase and then decrease (e.g., Humphreys & Bain, 1991).

A test of these predictions was carried out by Gronlund and Ratcliff (also see Nobel & Shiffrin, 2001; Rotello & Heit, 2000). Subjects studied pairs of words and single-words, and they were tested with intact pairs, single-word targets, rearranged pairs, single-word foils, and pairs consisting of two unstudied words. HRs increased and the FARs decreased with the latency of the responses for all but one type of foil. For rearranged pairs, the

FARs initially increased until about 600 ms subsequent to the probe, after which the FARs decreased, converging to an asymptote at about 1400 ms subsequent to the probe. These results are consistent with both a dual-process model and dual-cue model of the time course of recognition. The finding that older adults produce a monotonically increasing FAR functions further supports these models, suggesting that their ability to recall-to-reject rearranged pairs is impaired (Light, Patterson, Chung, & Healy, 2004) or that their ability to form new associations is impaired (Naveh-Benjamin, 2000).

The findings from associative recognition generalize to other recognition tasks. The inclusion–exclusion procedure requires subjects to study two lists of words (Jacoby, 1991). In the inclusion condition, subjects answer "yes" to any studied item. In the exclusion condition, subjects must only answer "yes" to items that appeared on one of the lists and answer "no" to all other items. Hence, the inclusion condition is a traditional study-test recognition task, and the exclusion condition is a standard list-discrimination (e.g., Winograd, 1968) or source memory task (e.g., Bayen, Murnane, & Erdfelder, 1996).

McElree, Dolan, & Jacoby (1999) combined the inclusion–exclusion procedure with the signal-to-respond procedure. In the inclusion condition, HRs increased and FARs decreased as the response lag increased. In the exclusion condition, HRs increased, but FARs were non-monotonically related to response latency, initially increasing and then decreasing. Moreover, older adults are not as able as younger adults to effectively reject foils in the exclusion task (Hay & Jacoby, 1999). These data suggest that exclusion performance relies more heavily on a recall-to-reject process than inclusion performance, and older subjects might rely more on familiarity than younger subjects.

The dual-cue model has a bit more difficulty explaining the difference in the time course of FARs observed in the inclusion and exclusion conditions. That is, the subjects must use a cue representing the association between the study context and the item to perform both tasks, and thus one might expect the time courses of the FARs in the inclusion and exclusion conditions to be similar. One possible extension is that subjects initially probe with context representing the experiment to decide if the item is old or new, and then they probe with context representing a particular study list in order to perform the exclusion task (cf. Dennis & Humphreys, 2001).

Thus, it is difficult to distinguish between dual-process models that assume that recollection of episodic details is relatively slow and dual-cue models that assume that the formation of an associative retrieval cue takes longer than the formation of item cues. Both models predict the non-monotonic FAR function found for associative recognition and list-discrimination.

Plurality discrimination (Hintzman & Curran, 1994; Hintzman & Curran, 1995; Hintzman, Curran, & Oppy, 1992) requires subjects to discriminate between targets (e.g., *RIVER*) and foils of the opposite plurality (e.g., *RIVERS*). Thus, plurality discrimination and associative recognition both require the discrimination of targets and foils that are similar. Indeed, Hintzman and Curran found that plurality discrimination performance is very similar to associative recognition performance. The FAR for plurality-reversed foils is much greater than FAR for new foils, and the retrieval dynamics are similar: HRs increase and new-item FARs decrease steadily with response latency, but the FARs for plurality-reversed foils initially increase and then decrease. The increase in the plurality-reversed FARs peaks around 700 ms subsequent to the probe and then the FARs flatten out about 1000 ms subsequent to the probe. In addition, Light, Chung, Pendergrass, &

Van Ocker (2006) found that older subjects produce a plurality-reversed FAR function that increases steadily with processing time.

While the plurality discrimination findings are consistent with the dual-process predictions, they challenge the dual-cue models more. Accurate plurality discrimination does not require associative information. Indeed, it is somewhat counterintuitive to assume that independent associative information is the basis for plurality discrimination especially since the nature of task is not typically revealed until after study (Hintzman & Curran, 1994; Hintzman et al., 1992; Malmberg et al., 2004). Weeks et al. (submitted for publication) proposed that subjects might probe twice; one probe is with a singular cue and one is with a plural cue, and the cue is chosen that produces the greatest amount of familiarity. In essence, this is a two-alternative forced-choice task between highly similar cues, and hence their familiarities will co-vary to a high degree. In such cases, performance is predicted by global-matching models to be very high (i.e., the similar distractor advantage, Clark, 1997; Tuiving, 1981). However, the accuracy of plurality discrimination is actually quite low.

6. The effects of strengthening operations on the discrimination of similar items

The signal-to-respond procedure might encourage a recall-to-reject or a dual-cue strategy to discriminate between targets and similar foils because the subject often needs to wait to respond (Ratcliff, 2006). In a free responding condition, subjects might not be as willing to delay their responses, and hence recollection or associative information might play less of a role. To address the ability of different models to account for associative recognition and plurality discrimination under free response conditions, let us first consider the interaction between target–foil similarity and strengthening operations. Here, strengthening operations are variations in study time or repetitions.

The global-matching models predict that strengthening operations will have two effects on recognition memory. First, they increase the match between the memory trace of a studied item and the corresponding retrieval cue (cf. Clark & Gronlund, 1996). Second, strengthening operations induce differentiation (McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997); traces representing different items in memory become less similar to each other as the strengthening operation increases in magnitude. As a result, HRs naturally increase, but the effects on FARs is variable (Criss, 2006). When strengthening occurs between lists, FARs decrease, which produces a mirror effect. When strengthening occurs within lists, there is obviously no effect on FARs, assuming the criterion remains steady (cf. Hirshman & Arndt, 1997; Stretch & Wixted, 1998).

The predictions hold when targets and foils are randomly similar, but what about when they are similar, as in associative recognition and plurality discrimination? According to the differentiation models, it again depends on the whether the manipulation occurs within or between lists (Criss, 2006). When the strengthening manipulation occurs between lists, differentiation is effective; the extent to which this is so is negatively related to the similarity between the targets and foils. At the limit, one cannot differentiate two identical traces. However, when the strengthening operations occur within lists, the global-matching models predict that FARs should increase as the magnitude of the strengthening operation increases and as target–foil similarity increases (Dyne et al., 1990; Malmberg et al., 2004; Xu & Malmberg, 2007). This prediction is confirmed for associative recognition and plurality discrimination in older adults, but it is disconfirmed in younger adults: repetitions increase HRs but not FARs, and increases in study time actually decrease FARs (Cleary et al., 2001; Hintzman et al., 1992; Kelley & Wixted, 2001; Light et al., 2004, 2006; Malmberg et al., 2004; Xu & Malmberg, 2007).

The null effect of repetitions on FARs has been explained on the assumption that recall-to-reject strategy is increasingly used to reject similar foils despite their high level of familiarity. In accordance with this proposal, Dyne et al. (1990) created a situation that rendered the recall-to-reject strategy useless to younger subjects by pairing items at study with multiple partners. Thus, the retrieval of an item at test that does not correspond with to test pair would provide insufficient evidence to conclude that the test items were not studied together. Based on their results, Dyne et al. concluded that a single-process global-matching model best captured performance.

Another constraint on the empirical generalization is that flat FAR functions only occur when the stimuli are pairs of words or pairs of novel faces (Xu & Malmberg, 2007). When pairs of pseudowords or Chinese characters are used, FARs increase with increases in both repetitions and study time. Thus, the form of the function relating repetitions to FARs depends (a) on the nature of study lists, (b) on the age of the subjects, (c) on the nature of stimuli that are tested, and (d) whether pairs are strengthened via increases in repetitions or via increases in study time.

Malmberg & Xu (2007) noted that some of the Xu & Malmberg (2007) three-point FAR functions appeared non-linear (i.e., inverted U-shaped), and this suggested that the null effects of repetitions that were previously observed were the result of varying repetitions at only two levels. By parametrically increasing 1.5-s repetitions from 1 to 12, Malmberg and Xu observed that the form of this FAR function was variable and dependent on the testing conditions. When a ratings task was used, the FAR function was slightly inverted U-shaped, but the FAR function increased steadily when a yes–no task was used.

Malmberg & Xu (2007) related these findings to the findings from the signal-to-respond procedure (Gronlund & Ratcliff, 1989) by suggesting that subjects sometimes respond "yes" or "no" before information could be recollected from memory (cf. Baranski & Petrusic, 1998; Van Zandt & Maldonado-Molina, 2004). The ratings task is more complex than the yes–no task and therefore it takes more time; during the additional decision time, recollective information becomes available and could be used to reject the arranged pairs. Malmberg and Xu tested this hypothesis by delaying yes–no responses by 2 s and observed that the function relating FARs and repetitions was inverted U-shaped. Parameter estimates from fits of an extension of a dual-process plurality discrimination model supported their interpretation of the findings.

The different forms of the FAR function challenges some models. The effect of study time on associative recognition is particularly problematic for global-matching models like TODAM (Murdock, 1982), which assumes that the item information and associative information are independent, and associative recognition is based on a match between the associative information in memory and the associative information used as a retrieval cue. Because the associative information is independent of the items from which it was generated, TODAM predicts no effect of strengthening operations on FARs, and there is no sensible means to decrease the strength of associative information stored in memory via increases in study time.

In contrast, the some-or-none model (SON) assumes that studying a pair results in the storage of item and associative information, and that the strengths of the item and associative information are sometimes combined at test into a single source of evidence. The associative information provides positive evidence for targets and negative evidence for foils, whereas item information provides positive evidence for both targets and foils. In the remaining cases, memory only produces item strength. On the assumption that memory produces associative evidence as often as it does not, SON predicts no effects of repetitions on FARs. SON can also predict variability in the FAR functions depending on the tradeoff between item strength and associative strength. Last, SON can account for the steadily increasing non-delayed yes–no FAR function if the retrieval of associative information is slower than the retrieval of item information (cf. Humphreys & Bain, 1991).

A problem for SON, however, is how it characterizes item familiarity, which would presumably be the basis for plurality discrimination. If strengthening operations simply increase item strength, then one would expect steadily increasing plurality-reversed FAR functions, but they are never steadily increasing (Hintzman & Curran, 1995; Hintzman et al., 1992; Malmberg et al., 2004). SON is an example of a model that is limited in scope insofar it only produces the correct predictions for tasks that involve the use of associative information.

Other single-process models, like SLiM (McClelland & Chappell, 1998) and the cue-matching model (Heathcote et al., 2006) account for plurality discrimination. However, the cue-matching model does not account for strengthening operations, and SLiM comes up short for precisely the opposite reason that the SON model does; it does not account for the variable effects of strength on false-alarms for associative recognition (Criss & McClelland, 2006). Indeed, SLiM might not even provide reasonable accounts of the effect of repetitions on plurality discrimination. According to SLiM, repetitions induce differentiation, and this accounts for the lack of effect of repetitions on FARs. In order to achieve accurate predictions (McClelland & Chappell, 1998), SLiM must assume that, say, *CAR* and *CARS* only overlap in over half of their features (~60%). Intuition suggests that plurality-reversed items must overlap to a greater degree than this, and if intuitions are correct, and they might not be, then the SLiM's account is questionable. Last, it is unclear from these single-process models how to account for the retrieval dynamics of plurality discrimination, especially why the time course of differentiation would be non-monotonic (cf. Hintzman & Curran, 1994).

7. The relationship between accuracy and judgments of frequency

In Hintzman and Curran's plurality discrimination experiments (1992, 1994; also Malmberg et al., 2004; Sheffert & Shiffrin, 2003), subjects were asked to make a frequency judgment at test. Frequency judgments estimate the number of times that an item was studied (JOFs). Viable models of plurality discrimination should be able to account for the HRs and FARs and the JOFs.

Hintzman (1988) showed that MINERVA2 accounted for recognition accuracy and JOFs on the assumption that both judgments were based on familiarity. Hintzman et al. (1992) noted, however, that the model breaks down when targets and foils are similar because repetitions have different effects on recognition accuracy and JOFs. Repetitions increase JOFs, but the ability to discriminate targets from similar foils is unchanged after about two repetitions. Thus, the increase in JOFs indicates that memory registered additional item presentations, but they did not help the subjects to learn the features that were necessary for more accurate plurality discrimination. Cleary et al. (2001) extended this phenomenon to associative recognition, and Hintzman et al. referred to it as *registration-without-learning*.

The challenge is to explain how the occurrences of items are registered in memory and the details that allow for accurate discrimination are not learned. Most single-process plurality discrimination models take the convenient path and ignore the JOF data (e.g., Dennis & Humphreys, 2001; Heathcote et al., 2006; McClelland & Chappell, 1998; Park et al., 2005). However, the challenge is account for all of the data (Hintzman et al., 1992). To do so, Malmberg et al. (2004) assumed that subjects initially make an old–new decision based on a combination of familiarity and recollection, and then they make a JOF based on the familiarity of the test item, if it was judged old. This allowed the model to predict both increasing JOFs functions and the relatively flat FAR function.

8. The relationship between recognition accuracy and confidence

We noted above that the effects of strengthening operations on associative recognition have been tested with a ratings procedure (Kelley & Wixted, 2001; Xu & Malmberg, 2007). Like JOFs, ratings provide more information than simply HRs and FARs. For instance, the tendency to use the high-confidence response increases with increases in study time and repetitions. This holds for both hits and false-alarms, even as the FARs remain steady or even decrease. It also holds for different types of stimuli, including pairs of non-words and pairs of Chinese characters. Thus, different types of stimuli produce different patterns of accuracy, but they produce the same patterns of confidence.

Much like the case of JOFs, the challenge for the single-process models is to describe how false-alarms remain steady or even decrease, while confidence in the responses always increases. This combination of accuracy data and confidence data is very difficult for extant single-process models to explain. In addition, increases in high-confidence false-alarms with increases in strength violate the process-pure assumption. Of course, a violation of the process-pure assumption is not particularly problematic for the wider class of dual-process models; indeed some of the earliest dual-process models assumed that high-confidence responses were based at times on very high or very low levels of familiarity (e.g., Rabinowitz, Mandler, & Patterson, 1977).

9. The cognitive neuroscience of recognition memory

The Complimentary Learning Systems (CLS) model (Norman & O'Reilly, 2003) advances the modeling of recognition by grounding its framework in what is known (or at least believed) about how the brain supports memory. It assumes that there are three main structures involved in recognition memory: the hippocampus, the neocortex, and a medial temporal lobe complex. The neocortex is slow learning and represents abstract knowledge, whereas the hippocampus is fast learning and supports the representation of episodic details. The medial temporal lobe (MTL) complex mediates the interaction of the contributions of the neocortex and the hippocampus to recognition performance.

The CLS model assumes that events are stored in the neocortical network in a very dense manner, much like the distributed representation of TODAM. In this sense, it can be compared to some of the global-matching models discussed earlier. The neocortex (i.e., the perirhinal cortex, *personnel communication* between the author and Ken Norman, October, 2007) is responsible for input to the medial temporal lobe, which in turn creates a transformed representation that is used to generate a familiarity signal. Thus, the familiarity signal is a product of abstract, concept-like representations; much like the familiarity signal in SAC is generated from the activation of concept nodes.

The hippocampus receives input from the entorhinal cortex, and through an interaction of area CA3 and the dentate gyrus it produces a sparse representation of the input. This allows the hippocampus to represent the input in a manner that preserves its details. To recover these details, the detailed representation in CA3 feeds into CA1, which has well-established connections with an entorhinal output network allowing the hippocampus to reinstate the episodic details of the original learning event. In this manner, the hippocampus supports recollection.

The goal of the CLS model is different from those of the models that we discussed so far. Measurement models seek to measure the effects of different factors on recognition performance, and process models seek to describe the processes and representations in involved in recognition. The CLS model seeks to describe how different brain region interact to support recognition. For instance, CLS departs from SAC and measurement level dual-process models by assuming that the recollective output of the hippocampus is a continuous variable and not a discrete variable and false-alarms are sometime the result of this output. Thus, it should be better equipped to handle high-confidence false-alarms, which is currently problematic for SAC.

On the other hand, CLS places less emphasis on drawing quantitative predictions about behavioral performance. For instance, CLS is neutral with respect to the issue of process purity because it is not interfaced with a decision model that requires that a position be taken. Rather, the model has so far been satisfied with stating that distinct brain regions support familiarity and recollection and that both types of information are used to perform recognition tasks. Thus, the findings that are critical for the CLS model are those that identify the areas of the brain that are responsible for recognition memory.

For single-item recognition, there is mixed support for CLS. CLS assumes that the hippocampus and other MTL regions are involved in the encoding and the retrieval of episodic memories. However, Schacter & Wagner (1999) concluded that there was only support in the fMRI literature for the assumption that posterior areas of the hippocampus (particularly the parahippocampal gyrus) were involved in encoding, and this region corresponds to the neocortical network in CLS and not the MTL. Moreover, Schacter and Wagner did not conclude that CA3, the dentate gyrus, or the entorhinal cortex was involved in encoding or that there was enough evidence from fMRI studies to make strong conclusions about whether the hippocampus was involved in retrieval. In contrast, Lepage, Habib, & Tulving (1998) reviewed the PET literature. They concluded that the parahippocampus supports encoding and that the hippocampus was involved in retrieval.

Taking the fMRI and PET reviews together, Schacter & Wagner (1999) suggested that methodological differences between the fMRI and PET studies lead to the inconsistent findings. There were variations across studies concerning the orienting tasks used during study, the design of study and test lists, etc. Importantly, some of the inconsistencies in the literature might have been attributable to the different memory tasks used between studies. Studies in which the hippocampus has been consistently implicated in recognition performance use associative recognition (Davachi, Mitchell, & Wagner, 2003; Jackson & Schacter, 2004; Kirwan & Stark, 2004; Ranganath et al., 2004; Sperling et al., 2003; Staresina & Davachi, 2006) or source memory tasks (Cansino, Maquet, Dolan, & Rugg, 2002; Gold et al., 2006; Rugg, Henson, & Robb, 2003). There is far less support for the assumption that hippocampus plays a critical role in single-item recognition. For instance, Wixted & Squire (2004) showed that the hippocampal damage is associated with deficits in familiarity-based and recollective-based decisions, but according to the CLS model, the hippocampus is only involved in recollection.

Eldridge, Knowlton, Furmanski, Bookheimer, and Engel (2000) proposed that one reason why it had been difficult to observe activation in the hippocampal region during single-item recognition was because the data analysis included all responses, but only some responses were based on recollection. They proposed that remaining responses were based on familiarity, and hence they would not be associated with hippocampal activation. Thus, the effect of collapsing over responses based on recollection and familiarity was the inability to observe the activation of the hippocampus during single-item recognition.

Today, fMRI studies often utilize a RK or confidence-rating procedure as a means for identifying brain regions responsible for recollective versus familiarity-based responses. For instance, Eldridge et al. (2000) utilized the RK task and observed greater hippocampal activity associated with "remember" responses than with "know" responses. In contrast, others found that "remember" and "know" responses were not associated with differences in hippocampal activation during test, but hippocampal activation was greater at study when items subsequently elicited a "know" response than when the items elicited a "remember" response (Henson, Rugg, Shallice, Josephs, & Dolan, 1999). Still others concluded that the perirhinal cortex was associated with the production of the familiarity signal (Henson, Cansino, Herron, Robb, & Rugg, 2003). Thus, it does not appear that the RK procedure has led to a better understanding of how different brain regions support recognition.[5]

It is, of course, possible than more than one region of the brain is responsible for familiarity and/or that more than one area of the brain is responsible for recollection. Yonelinas, Otten, Shaw, and Rugg (2005), for instance, used a procedure that combined the RK procedure with the ratings procedure in order to reveal the brain areas responsible for recollective-based responses and familiarity-based responses. Upon the presentation of the test stimulus, subjects indicated if they could recollect information that indicated that the item was a target. Brain activity related to these responses was assumed to reflect recollective-based acceptance. If subjects indicated that they could not report episodic details, they rated on a scale of 1–4 their confidence that an item was studied. The ratings were assumed to reflect familiarity-based responses.

Yonelinas et al. (2005) further noted that simple differences in activations among different brain regions would not allow for meaningful conclusions about whether one area is responsible for producing familiarity and the another recollection. One could interpret simple differences in activation of given brain region as a reflection of the differences along a single-dimensional scale. According to them, "To overcome this objection it is *necessary* to demonstrate that increases in familiarity confidence are associated with a pattern of neural activity that is qualitatively different from that revealed by the contrast between recollection and high-confidence familiarity" [p. 3002, italics added].

[5] There are additional clear methodological differences between imaging studies that might influence one's conclusions. For instance, in the Eldridge et al. (2000) experiment in which the hippocampus was implicated in remember responses, 80% of the test items were targets, and hence the trial did not emphasize the discrimination of targets and foils insomuch as it emphasized the ability to recollect episodic details.

Their results revealed two patterns neural activation. One set of regions (Broadman's area, posterior cingulate, lateral parietal/temporal cortex, hippocampus, and parahippocampus) produced a non-linear U-shaped activation function, with high levels of activation being observed for recollection responses, high-confidence "yes", and high-confidence "no" responses. Lower levels of activation were observed for moderate levels of "yes" and "no" confidence. A somewhat different set of regions (precuneus, left prefrontal cortex, and left parietal cortex) produced more linear functions with the highest levels of activation related to recollection and high-confidence "yes" responses and the lowest level of activation related to high-confidence "no" responses.

If Yonelinas et al. (2005) are correct and the areas of the brain that produce the U-shaped activation are involved in recollection, their findings suggest that no single area of the brain is uniquely responsible for recollection and that these areas are broadly distributed. Likewise, no single area of the brain is associated with the production of a familiarity signal. On the other hand, the number of brain regions implicated in this experiment might simply reflect the complexity of the task used. A different question is whether the two brain states are uniquely associated with recollective and familiarity-based responses. Since both the U-shaped pattern and the linearly decreasing pattern of brain pattern show heightened activation for recollective responses and high-confidence "yes" responses, it is difficult to determine which areas are involved in recollection and not in familiarity, and vice versa. Thus, showing that recollective and familiarity-based responses are associated with qualitatively different patterns of neural activity may be necessary to infer that the brain regions are providing qualitatively different types of information on which to make recognition decision, but such evidence might not be sufficient for drawing strong conclusions.

In summary, there is consistent support of the hypothesis that the hippocampus plays an important role in associative recognition and source memory. On the assumption that the hippocampus is also involved in the encoding of events in such a way that supports recollection, the dual-process models of associative recognition and source memory are also supported. In contrast, there is only mixed support for the hypotheses that hippocampus is critically involved in the single-item recognition. In this sense, the cognitive neuroscience literature aligns with the behavioral literature.

10. Conclusions

Toward a classification of data, I have four main conclusions. At the measurement level, the critical question concerns the ability of the models to measure what they are designed to measure, and the dual-process measures of recollection are unreliable. Because the RK literature is so vast (Rotello et al., 2004), the fact that the RK based arguments are not overly compelling is in itself damning for models that assume that the RK measures of recollection tell us something important about the nature of recognition memory. In contrast, the signal-detection and global-matching models remain viable, if somewhat incomplete, accounts of single-item recognition.

The retrieval dynamics of associative recognition and plurality discrimination are highly similar, and they suggest that there are two sources of information occurring over different time courses. While a dual-cue model might explain the dynamics of associative recognition, the plurality discrimination task seems less amenable to that approach. However, a dual-process model in which a recall-to-reject mechanism plays a prominent role can explain the dynamics of both tasks.

The function relating false-alarms to strengthening operations is systematically variable and dependent on strengthening operations, stimulus properties, the age of the subject, and testing conditions. Moreover, the FAR functions are often different from the functions relating strengthening operations and confidence or JOFs. Although several single-process models account for the accuracy of associative recognition or plurality discrimination, without exception they fail to account for both tasks or for the observed interactions. On the other hand, the dual-process approach captures the basic effects and the complex interactions, even if there is no evidence that recognition decisions are "process pure".

Based on these conclusions, there are three obvious ways to partition the data. The first one creates a single category; it assumes that the dual-process model applies to single-item recognition, associative recognition, and plurality discrimination. Specifically, a recall-like process always plays a significant role by providing information in the form of episodic details that are used to make a decision. A different partitioning of the data assumes that recollection never plays an important role in recognition, and that a dual-cue approach explains the differences in task performance. The final classification assumes that a recall-like process plays a significant role only when performance would benefit from the retrieval of episodic details. Accordingly, single-item recognition is based only on a global-matching process, but associative recognition and plurality discrimination also utilize a process that supports a recall-to-reject strategy. In my view, the dual-cue approach is the least supported due to its inability to so far provide a reasonable explanation of plurality discrimination. The lack of evidence, behavioral, cogneuro, or otherwise, in support of the dual-process model of single-item recognition is also striking. Thus, the latter approach is the one that I have pursued.

11. Recognition memory: An integrated approach

Now I will describe a number of REM models and demonstrate in several novel ways how they account for the critical findings that I have reviewed within a single coherent global-memory framework.

11.1. Single-item recognition

Shiffrin & Steyvers (1997) created the REM global-matching model to account for benchmark findings, including item strength effects, list-length effects, list-strength effects, mirror effects, and the form of the recognition memory ROC. For RK performance, the model adopts the conventional single-process assumption that RK responses reflect different levels of familiarity, and that the RK responses are generated by comparing an item's familiarity to two decision criteria. This RK model provides a quantitative account for item strength effects, mirror effects, the effect of midazolam (Malmberg et al., 2004), and the RK and ratings ROCs (Malmberg & Xu, 2006). Many have stated that it is critical to account for list-length and list-strength effects, but the REM RK model has not yet done so (Dennis & Humphreys, 2001; Glanzer, Adams, Iverson, & Kim, 1993; Malmberg & Murnane, 2002; Murdock & Kahana, 1993; Shiffrin & Steyvers, 1997).

11.1.1. Encoding in REM

The same encoding assumptions are made in all of the models that I will describe. According to the REM, a lexical/semantic trace is activated when an item is studied. These consist of vectors of w geometrically distributed feature values. Variability in the feature values is determined by the geometric distribution parameter, g. When single-items are studied incomplete and error-prone copies of a lexical/semantic trace are stored (Malmberg et al., 2004). When pairs are studied the concatenation of two episodic traces is stored (Xu & Malmberg, 2007).

Encoding of the item in an episodic trace produces an incomplete and error prone version of its lexical/semantic representation. Specifically, one or more attempts at storing a lexical/semantic feature is made, t, and on each attempt a feature is stored with probability u^*. If a feature is stored in the episodic trace, it is copied correctly from the lexical–semantic trace with probability c. If the feature is stored incorrectly, a feature value is randomly drawn from the geometric distribution. Thus, the probability that a given lexical/semantic feature is correctly stored in an episodic trace is: $c(1-(1-u^*)^t)$.

11.1.2. Global-matching

For single-item recognition, responses are based on the familiarity value generated by a global-matching process. It compares a test item's lexical/semantic trace to each trace stored during study, j, by noting for each trace the features that match and mismatch the features of the cue and their values. Based on this information, a likelihood ratio is computed for each trace, j, λ_j:

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

where i corresponds to features values drawn from the geometric distribution, m is the number of those that match and q is the number that mismatch. n_{jq} is the number of mismatching features in the jth concatenated trace and n_{ijm} is the number of features in the j^{th} concatenated trace that match the features in the compound retrieval cue.

The recognition decision is based on the odds, $\Phi = \frac{1}{n}\sum_{j=1}^{n}\lambda_j$. If the odds exceed a criterion, the response is positive. To make RK judgments, an additional, stricter RK criterion is set. If the odds exceed the RK criterion, the response is "remember", otherwise the response is "know".

11.1.3. The list-length RK effect

Jing Xu and I generated global-matching REM predictions for Cary & Reder's (2003) list-length data.[6] In this experiment, subjects studied lists that consisted of 16, 32, 48, and 64 words, and each list was followed by RK testing. Fig. 3 shows an accurate fit of the model to the data, but what is more remarkable is that the fit was generated using the same parameter values used by Shiffrin & Steyvers (1997) to account for yes–no recognition performance. A better fit is possible, but this one is sufficient to demonstrate the robustness of the RK model's predictions.

11.1.4. The list-strength RK effect

The critical question concerns the relationship between strengthening operations and interference. We will refer to items studied relatively often as *strong items*, and items studied less often as *weak items*. In the mixed–pure paradigm, subjects study lists consisting of

[6] I thank Melanie Cary and Lynne Reder for providing their data.

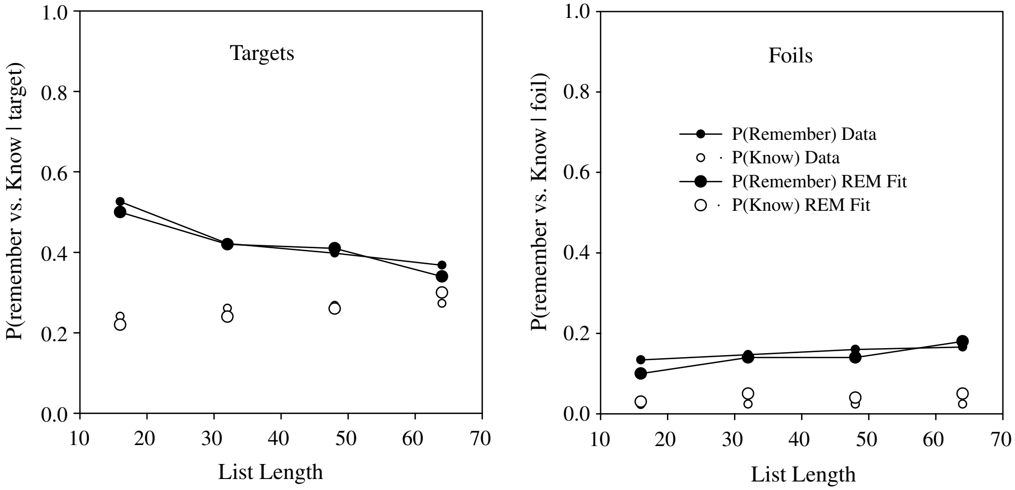

Fig. 3. A REM single-process model to Cary and Reder's (2003) data. *Note:* The REM parameters values generating this fit were $u = 0.04$, $w = 20$, $t = 10$, $g = 0.4$, $c = 0.7$. The old–new and RK criteria were 1.0 and 6.9.

all strong items (i.e., pure-strong lists), all weak items (pure-weak lists), or lists consisting of some strong and some weak items (i.e., mixed lists). If adding relatively strong items to memory causes more interference than adding relatively weak items to memory, memory for weak items should be worse on mixed lists than on pure lists, and vice versa for strong items. This is a *positive list-strength effect*. For recognition memory, however, adding relatively strong traces to memory does not affect, or might slightly benefit, recognition (Murnane & Shiffrin, 1991; Ratcliff et al., 1990; Shiffrin, Ratcliff, & Clark, 1990). This is a null or slightly *negative list-strength effect* (Dennis & Humphreys, 2001; Estes, 1994; Glanzer et al., 1993; McClelland & Chappell, 1998; Ratcliff et al., 1990; Shiffrin & Steyvers, 1998).

To evaluate the single-process RK model Jing Xu and I conducted a list-strength experiment. (The design of is presented in Appendix A). Strong and weak items are operationally defined as words studied three times or once, respectively, for 2 s. This level of strength manipulation is known to produce positive list-strength effects for free recall (Malmberg & Shiffrin, 2005), but for RK recognition we again find a null list-strength effect. Fig. 4 shows a fit of the RK model. The only free parameters were the number attempts at encoding features in the weak and strong conditions and the criterion locations. The fit is good, and hence these data do not cause one to reject the model.[7]

11.2. A model of retrieval dynamics

We have shown here and elsewhere that the global-matching RK model accounts for all the effects that it was designed to handle. However, this model does not account for

[7] Only between list manipulations produce a mirror effect. This interaction is sometimes explained by assuming different yes-no decision criteria for weak and strong items when strength is varied between lists (Hirshman, 1995). However, such comparisons are the same as list-strength comparisons, and REM accounts for the findings based on the principle of differentiation.

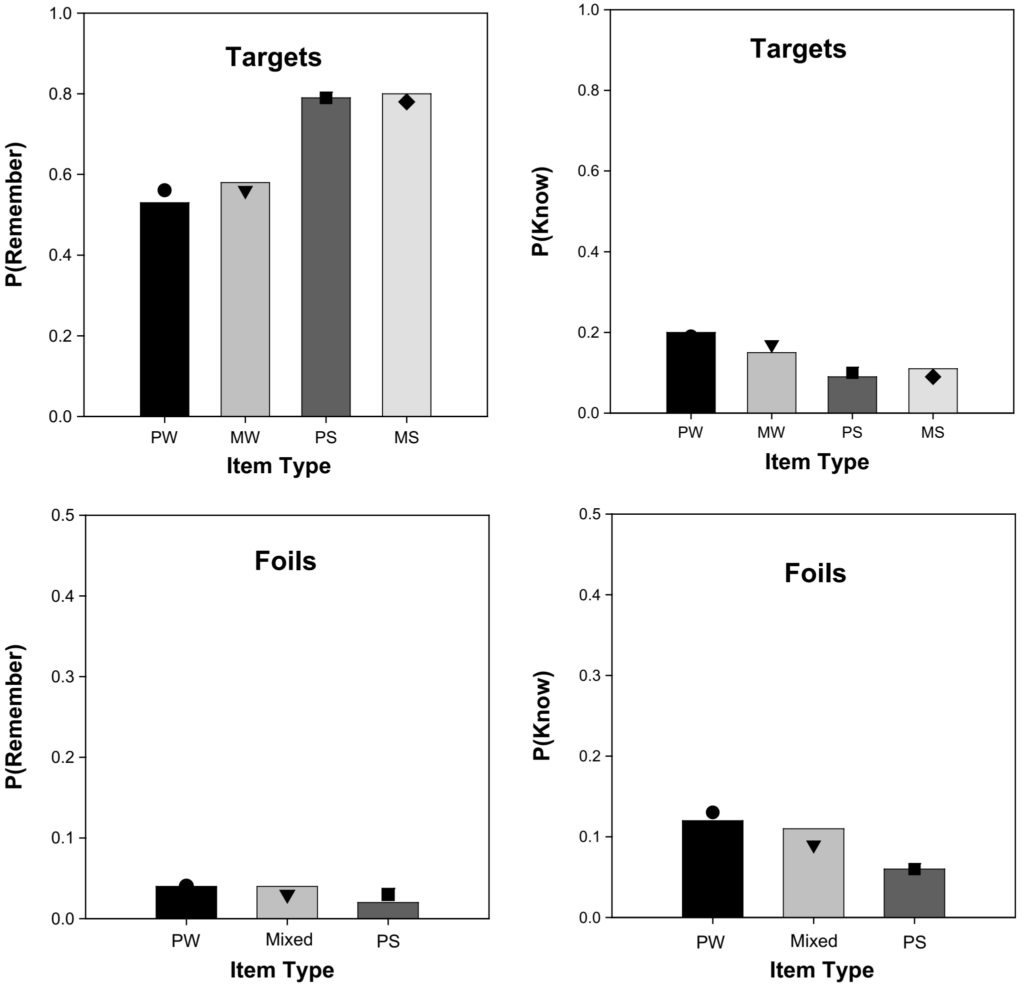

Fig. 4. Fit of the REM single-process model to the list-strength RK recognition data. *Note:* The REM parameters values generating this fit were $u = 0.04$, $w = 20$, $t_{weak} = 9$, $t_{strong} = 17$, $g = 0.4$, $c = 0.7$. The old–new and RK criteria were .8 and 3.8.

response latencies. It also does not explain the performance of tasks where targets and foils are similar (Malmberg, 2008; Malmberg et al., 2004; Xu & Malmberg, 2007). To account for these tasks, the dual-process REM models of plurality discrimination and associative recognition were developed. According to these models, recollection involves the retrieval of episodic details using the REM/SAM sampling and recovery processes (Malmberg & Shiffrin, 2005; Raaijmakers & Shiffrin, 1980), and the output of these processes is used to reject items when the task involves discriminating similar items. Aside from minor encoding and cueing assumptions (see above), the associative recognition and plurality discrimination models are identical, and the qualitative predictions of the dual-process model hold for both tasks. Here, I will compare the performance of a dynamic global-matching model and a dynamic dual-process model of associative recognition.

Please cite this article in press as: Malmberg, K. J., Recognition memory: A review of the..., *Cognitive Psychology* (2008), doi:10.1016/j.cogpsych.2008.02.004

The core of the dynamic model is a sequential-sampling process, whereby a buffer accumulates features stored during study. For each unit of time, T, the probability of sampling a feature is s, and the most recently sampled features are added to the features that were sampled in the prior time steps. Thus, the buffer maintains an incomplete representation of the traces stored during study that increases in completeness over time. Global-matching and retrieval is performed over this buffered representation.

Features are sampled without replacement. A feature remains in the buffer until the subject makes a decision, and the set of sampled features at time T is F_T. Thus, it is possible to sample all of the features stored during study given an infinite amount of retrieval time. This assumption has the advantage of guaranteeing the possibility of an optimal decision based on all of the evidence stored in memory.

For each T, Eq. 1 produces a familiarity value, Φ_T, by global-matching over F_T. For single-item recognition, a single lexical/semantic trace representing the test item serves as the retrieval cue. For associative recognition, a compound cue consisting of the concatenation of two lexical/semantic traces is compared to the traces stored during study. The log of Φ_T is computed in order to preserve the signs of the positive and negative evidence. Evidence accumulates such that when $T = m$ the value of the odds, $\phi(m)$, is

$$\phi(m) = \log \Phi_m = \log\left(\frac{1}{N}\sum_{i=1}^{j}\lambda_j\right). \tag{2}$$

In parallel, a single-trace is sampled from F_T (Malmberg & Shiffrin, 2005). The probability of sampling trace j is

$$P(j) = \frac{\lambda_j}{\sum \lambda_j}. \tag{3}$$

The probability of sampling j increases as the similarity between it and the retrieval cue increases and the similarity between the retrieval cue and other traces in F_T decreases.

When a trace is sampled, an attempt to recover its contents is made. Recovery is delayed, which affects the temporal relationship between global-matching and retrieval. Successful recovery requires sampling a minimum number of stored features, K_R (e.g., Malmberg & Shiffrin, 2005). Thus, recovery chances improve as encoding improves and as F_T becomes more complete. If recovery is successful, the contents of the recovered trace are compared to the test pair (see below).

11.2.1. Familiarity-based recognition decisions

For free response recognition, $\phi(m)$ is compared to two decision criteria, K_O and K_N (e.g., Atkinson & Juola, 1974; Hockley & Murdock, 1987; Smith, Shoben, & Rips, 1974). If $\phi(m)$ does not exceed K_N, a “new” response is made (i.e., $\phi(m) < K_N$). If it exceeds K_O, an “old” response is made. If $K_N < \phi(m) < K_O$, then the process begins anew by sampling a new set of features from memory and adding them to F_T.

The major difference between the dynamic global-matching model and the classical random walk model concerns the accumulation of the evidence. The classical random walk model (Ratcliff, 1978) assumes that there is an infinite amount of information that can be sampled from memory, whereas the dynamic global-matching model assumes the amount of information that can be sampled is limited by the amount of information that was stored during encoding. Thus, sampling is carried out without replacement of features in the present model. Because of sampling without replacement, it is possible even at very long lags that the evidence will not cross one of the decision boundaries. Therefore, the model requires a stopping rule.

In some of the prior models (Malmberg & Shiffrin, 2005; Raaijmakers & Shiffrin, 1980) we proposed that subjects halt a search of memory after K_{MAX} iterations. The question for the present models is when K_{MAX} is reached, what happens? We have assumed that guessing takes place in situations where familiarity is relatively uninformative, as in associative recognition and plurality discrimination (Malmberg et al., 2004; Xu & Malmberg, 2007; cf. Anderson, 1983). Indeed, the threshold-like nature of the sampling and recovery processes suggests that guessing would play a more significant role when the recall-to-reject strategy is used. In other situations, where targets and foils are randomly similar, the familiarity of the cue is more diagnostic, and it is possible that these decisions are based on the accumulated evidence (Ratcliff, 2006). In keeping with the continuous-state framework, therefore, it seems likely that guessing occurs less often for single-item recognition.

11.2.2. Free response recognition

According to the global-matching model used for single-item recognition, the latencies of hits and correct rejections decrease as the encoding improves, and this is what has been observed (e.g., Ratcliff, 1978; Ratcliff & Murdock, 1976). Increases in list-length also negatively impact latencies, and this has been reported several times, as well (e.g., Koppell, 1977; Ratcliff, 1978; Ratcliff & Murdock, 1976). Last, latencies decrease as the sampling rate, s, increases and as K_{MAX} decreases, but to the best of my knowledge, neither of these variables has been systematically explored for recognition memory.

This free response yes–no model can be extended to the RK and ratings tasks. Due to the sampling-without-replacement assumption, some items will cross K_O or K_N before K_{MAX} iterations and some will not. Hence, those items that produce enough evidence to cross a decision boundary are given a "remember" response, and those that do not are given a "know" response. This is fully in keeping with the single-process framework insofar as RK judgments reflect the familiarity of the retrieval cue. Also, recall that "remember" responses tend to be faster than "know" responses regardless of whether the test item is a target or a foil (Wixted & Stretch, 2004). The top panel of Fig. 5 shows the dynamic RK model predicts faster "remember" than "know" responses because these items produce enough evidence to cross K_O or K_N before K_{MAX} is exceeded. The lower panel of Fig. 5 illustrates the model for the ratings procedure. High-confidence responses are made more quickly than low-confidence responses because they are made as soon as a yes–no decision is made based on crossing one of the decision boundaries.

Because the RK and ratings models are so similar, there should be a strong relationship between the performances of these tasks (e.g., Malmberg & Xu, 2006). Indeed, high-confidence responses tend to be faster than low-confidence responses (Ratcliff & Murdock, 1976), just as "remember" responses are faster than "know" responses. Moreover, remember responses are typically made with greater confidence than know responses; again this is true regardless of whether a target or foil was tested (Malmberg & Xu, 2006; Wixted & Stretch, 2004). Likewise, high-confidence responses are often reported to be based on remembering episodic details (Yonelinas et al., 2005). Since the dynamic ratings model is nearly identical to the dynamic RK model, they provide a natural account of the strong correspondence between these findings.

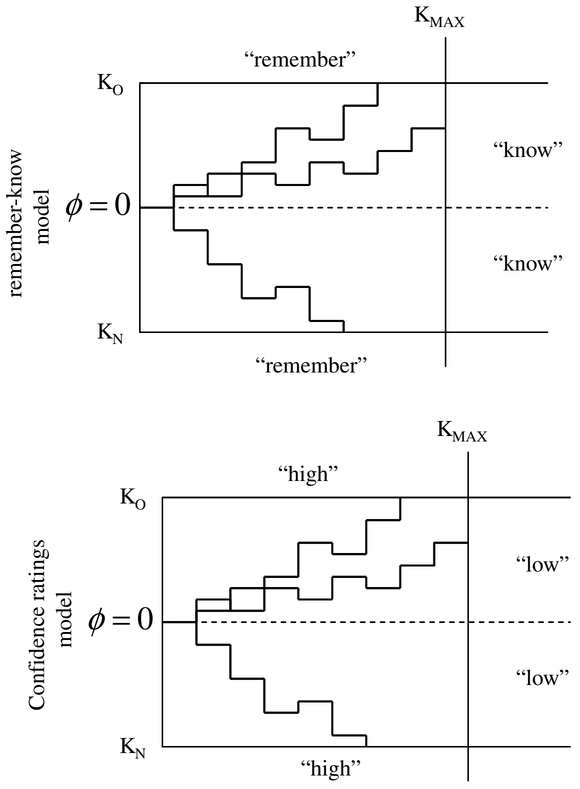

Fig. 5. Dynamic global-matching models of the remember–know and confidence-ratings tasks.

11.2.3. Signal-to-respond recognition

Thus, the dynamic global-matching free response model provides a reasonable account of the latency of yes–no recognition, RK recognition, and confidence ratings when single-item recognition memory is tested. However, much of the data that is critical for testing models concerns the time course of retrieval obtained form the signal-to-respond procedure. In this framework, the signal-to-respond model is a special case of the free response model (cf. Ratcliff, 2006), where decisions are made based on the information that is available when the signal occurs. This is accomplished by setting the higher and lower decision boundaries to the same value. In unbiased conditions, $\phi(m)$ is compared to a criterion optimally set to 0. (i.e., $K_N = K_O = 0$) If $\phi(m) > 0$, the response is "old", otherwise the response is "new".

Fig. 6 shows the behavior of the single-item recognition global-matching model. In this simulation, items were encoded relatively weakly or strongly on separate lists ($u^* = .04$ versus .08), and the model predicts greater HRs and lower FARs for strong items. Most importantly, the relationship between delay and FARs is monotonically decreasing, and hence the model provides a good account of the empirical findings (Gronlund & Ratcliff, 1989; Hintzman & Curran, 1994). I also applied the global-matching associative recognition model to a design in which the number of pair presentations is varied within lists. The only difference between this model and single-item recognition model is that for each rearranged foil there are two similar traces in memory. The left panel of Fig. 7 shows that HRs and FARs are positively related to repetitions and total processing time, and the rate of

Please cite this article in press as: Malmberg, K. J., Recognition memory: A review of the..., *Cognitive Psychology* (2008), doi:10.1016/j.cogpsych.2008.02.004

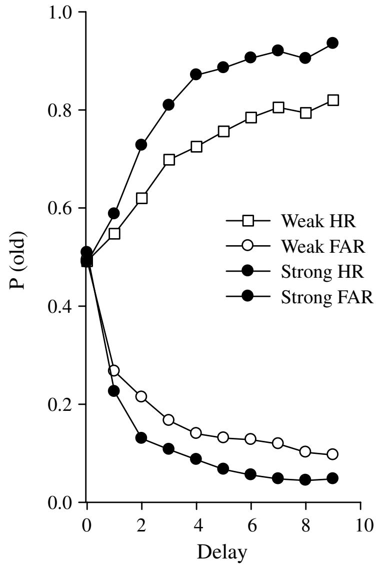

Fig. 6. Signal-to-respond functions generated from the dynamic global-matching model. *Note:* The following parameter values were used $w = 20$, $t = 16$, $c = .7$, $K_O = K_N = 0.0$, $s = .25$, u^* (strong) = .08, and u^* (weak) = .04.

increase is greater for HRs than for FARs. Thus, the global-matching model does not produce the non-monotonic FARs functions that are characteristic associative recognition.

11.2.4. A dynamic dual-process model

This framework assumes that a recall-to-reject strategy is used for recognition tasks involving the discrimination between targets and similar foils. Here, I will introduce the sampling and recovery processes to the global-matching model in order to observe the dynamics of the dual-process model of associative recognition.

Familiarity-based evidence accumulates in the same way as before. In parallel, a trace is sampled from F_T according to Eq. 3 during each time slice after a delay, and an attempt to recover its contents is made. The delay corresponds to the assumption that the sampling and recovery process take longer to produce output than the global-matching process. I also assume that an appropriate trace must be sampled in order for recollected evidence to be informative. For targets, the only appropriate trace is the one that corresponds to the target. For rearranged pairs, there are two appropriate traces, and sampling and recovery of either one of them has the potential to provide sufficient episodic details for the rejection of the pair.

In this model, recollection involves the sampling and recovery of features corresponding to the items that were studied. (In other versions of the model, recollection might involve the sampling and recovery of context features, say for source memory.) Recollection is successful if the number of sampled features exceeds K_R. Here, I assumed that $K_R = 2w/2 = 10$. I further assumed that unstored features did not count for or against a match, that four mismatching features was sufficient for detecting a rearranged pair, and that fewer than two mismatching features allowed the model to detect an intact pair. To simplify matters, I also assumed that recovery is a veridical process, and therefore foils only lead to negative recollective responses, and targets only lead to positive recollective responses. Thus, if a rearranged pair is tested, at least 10 features are sampled from an appropriate

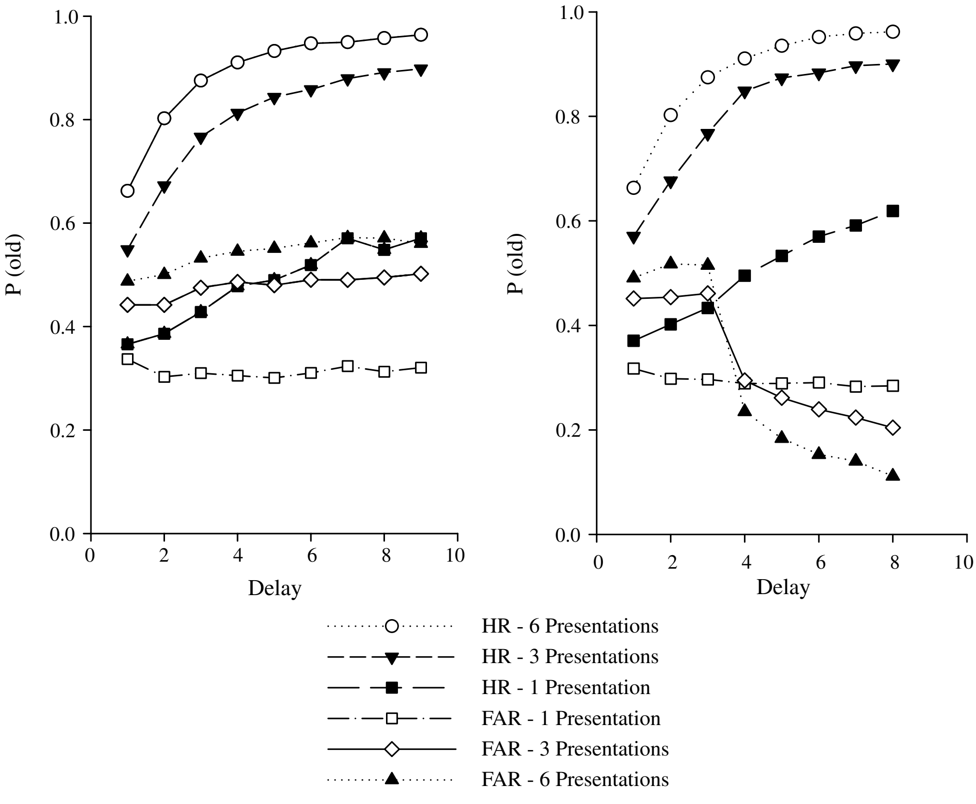

Fig. 7. Accuracy and latency predictions of the dynamic dual-process REM signal-to-respond recognition model of associative recognition. *Note:* Associative recognition hit rates and false-alarm rates were generated from a single-process and a dual-process REM signal-to-respond model at various units 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 familiarity and recollected details. The parameter values were $w = 10$, $t = 8$, $c = .7$, $u^* = .04$, $g = .4$, $K_O = K_N = 0$, $K_R = 5$, $s = .25$.

trace, and at least 4 mismatches are observed, then a "new" response is made. Likewise, if at least 10 features are sampled from the target trace, and 2 or fewer mismatches are observed, then the response is "old". When recovery fails ($K_R < 10$) and the familiarity falls between the two criteria, a new sample is taken from memory, it is added to the prior the sample, and the cycle begins a new. Guessing occurs when K_{MAX} is exceeded.[8]

[8] The recovery process is similar to, but not the same as, a double high-threshold model. When the number of sampled features exceeds K_R and a certain number of matches are observed, the subject has detected an intact pair, and when a certain number of mismatches are observed, the subject has detected a rearranged pair. (When K_R is not exceeded, the subject only has information produced by the global-matching process and may be in the indeterminate state.) There is additional information concerning which features and how many features of the sampled trace match and mismatch the test pair, and this information could be used to generate other responses (cf. Joordens & Hockley, 2000; Rotello et al., 2004). In the present model, such information does not affect the yes–no recognition decision.

11.2.5. Dual-process signal-to-respond model

The right panel of Fig. 7 shows that the dual-process signal-to-respond model produces the non-monotonic FAR function that is characteristic of associative recognition and plurality discrimination (e.g., Gronlund & Ratcliff, 1989). In addition, the asymptotic FAR decreases as the number of target presentations increases (e.g., McElree et al., 1999).

For free response associative recognition, hits tend to be faster than correct rejections (Malmberg & Xu, 2007). According to the model, hits are relatively fast because the global-matching process tends to produce sufficiently positive evidence prior to successful completion of the sampling and recovery processes. The latency of correct rejections tend to be longer because rearranged pairs tend to get caught between K_O and K_N prior to the completion of the slower sampling and recovery processes. Last, the associative recognition latency distributions are skewed with a relatively steep leading edge and a longer tail (Diller et al., 2001), and the model predicts similar shaped latency distributions (Malmberg, 2008).

I conducted simulations of two free response experiments from Malmberg & Xu (2007) in order to assess how the model accounts for the variable forms of the functions relating encoding strength and FARs. The accuracy data from these experiments and the behavior of the free response model is shown in Fig. 8. In one experiment, subjects made free response yes–no decisions and false-alarms tended to increase as the number of presentations increased. In the other experiment, subjects made free response yes–no decisions, but they were required to wait at least 2 s before responding. In this case, a non-monotonic, inverted U-shaped relationship between FARs and target presentations was observed, and of course the response latencies were on average much slower than in the first experiment.

Malmberg & Xu (2007) proposed that the different patterns of false-alarms reflected the use of different decision strategies. When allowed to respond as soon as they wanted, subjects sometimes responded before recollected evidence was available, and hence their performance was harmed by relying too heavily on familiarity-based evidence. On the other hand, when subjects were forced to wait 2 s, it allowed more time for the sampling and recovery of episodic details, which could then be used to make decisions that were more accurate.

I implemented the two free-response strategies in the model, and they produce different patterns of results. The fast strategy assumes the subjects make a response as soon as a familiarity-based judgment can be made based on exceeding the familiarity-based decision criteria. In this case, the model correctly predicts that FARs increase with increases in the number of pair presentations. It also correctly predicts that correct rejections are slow compared to hits. The right panel of Fig. 8 shows that recollection played almost no role in producing hits, and that correct rejections based on recollection occurred more often as the number of pair presentations increased.

In contrast, the slow response strategy assumes that subjects delay responses, perhaps based on a metacognitive judgment of when the sampling and recovery process is most likely to produce episodic details. Here, I assumed that no responses were made until three samples were taken from memory. Prior to that, the model waits for more samples from memory. Fig. 8 shows that the slow decision strategy produces a FAR function that is inverted U-shaped. The response latencies are obviously much slower than those produced

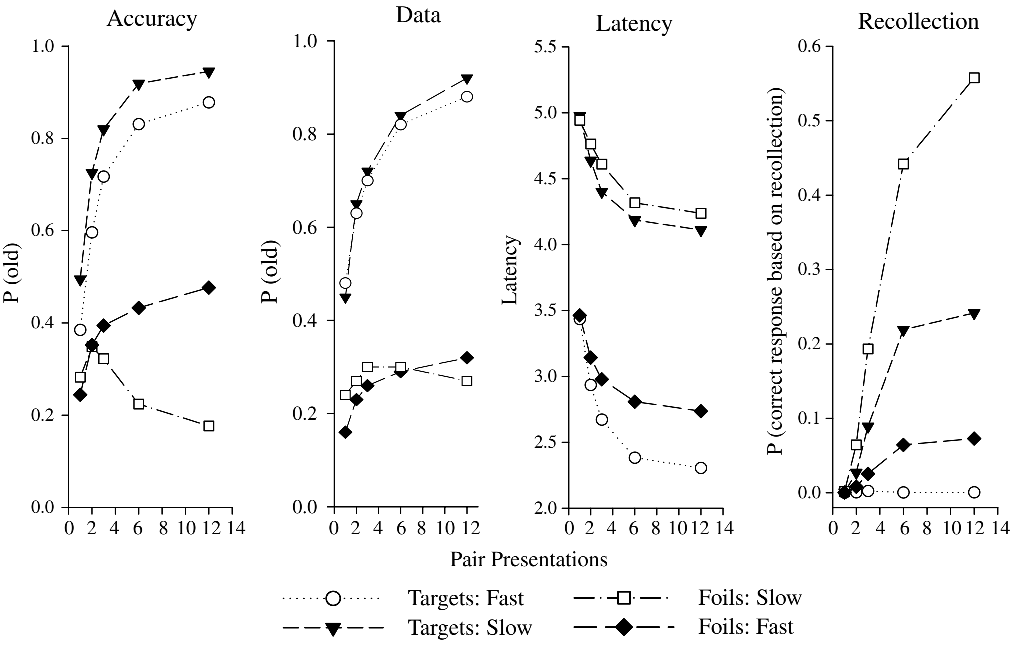

Fig. 8. Accuracy and latency predictions of the dynamic dual-process REM free response recognition model of associative recognition. Reprint from Malmberg (2008). Toward an Understanding of Individual Differences in Recognition Memory. *The Psychology of Learning and Memory* (B. H. Ross and A. S. Benjamin, Eds.), 48p. *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 (2007). 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.

by the fast decision strategy, and recollection contributed much more to performance, especially to correct rejections.[9]

In addition, the difference diminishes between the latencies of hits and correct rejections when a slow retrieval strategy is used. This suggests that the use of familiarity is the basis for hits. Moreover, an emphasis on recollection to reject foils accounts for much of the difference in the latencies of hits and correct rejections. The analysis presented in the far-right panel of Fig. 8 shows that the contribution of recollection to hits is almost nil regardless of the number of times the targets were studied, whereas the contribution of recollection increases steadily for correct rejections. When responses are slowed, the contribution of recollection to performance increases for both hits and correct rejections, which illustrates the role of recollection in the FAR functions and the latency data.

The inverted U-shaped function is also a characteristic of functions obtained from the ratings procedure (Malmberg & Xu, 2006). This suggests that the additional complexity of the rating tasks requires additional time, during which recollective details become avail-

[9] The contribution of recollection to hits in the model is somewhat misleading because at times the familiarity-based evidence was sufficient to warrant a “yes” response and recollection was successful. In such cases, I assumed that the response was only based on recollection, but that assumption can be reasonably disputed.

able and are used by the subject to reject foils (Baranski & Petrusic, 1998; Van Zandt & Maldonado-Molina, 2004). To explore the effect of pair strength on confidence ratings and response latencies, Fig. 9 presents a new analysis of the results of Experiment 1 from Malmberg & Xu (2007). In this experiment, subjects judged on a scale of 1–4 how confident they were that a test pair was studied, where 1 = high-confidence old, 2 = low-confidence old, 3 = low-confidence new, and 4 = high-confidence new, and Fig. 9 plots the tendencies to use the high-confidence versus the low-confidence alternative for each type of response (hit, correct rejection, false-alarm, or miss) and the average latency of each response type.

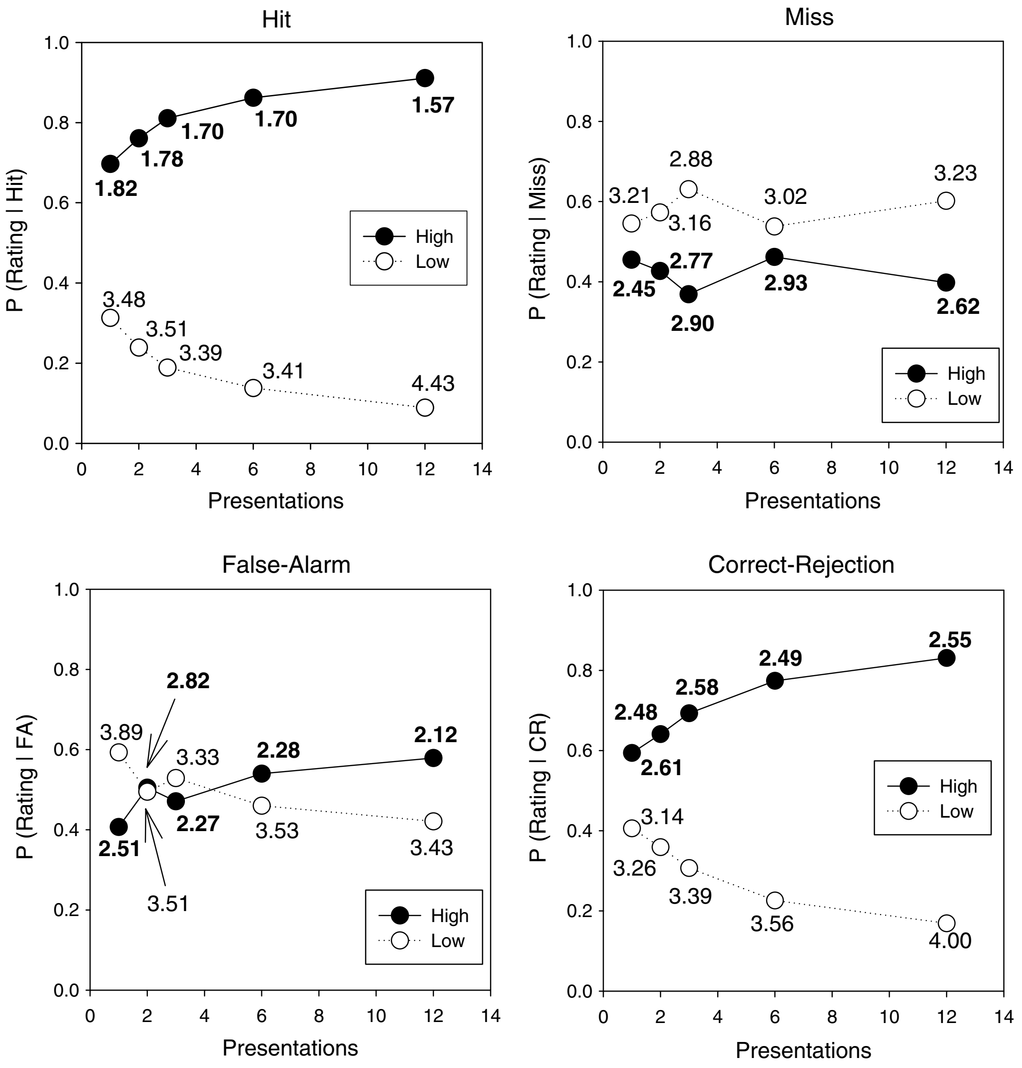

Fig. 9. Confidence and latency data from Experiment 1 of Malmberg and Xu, 2007.

With respect to the ratings model that I have described, there are several important findings to consider. First, high-confidence responses are faster than low confidence responses, the tendency to use the high-confidence response increases with the number of target presentations for hits and false-alarms, and the latencies of these responses decreases. This is consistent with the assumption that high-confidence responses are made when K_O is crossed; the stronger the target items is encoded in memory, the more quickly the boundary is reached. The increasing tendency to use the high-confidence old response when a false alarm occurs even though there was no effect of repetitions on FARs. This finding is consistent with the current assumptions. The steady FARs are due to the increasing effectiveness of the recall-to-reject strategy opposing the increase in pair familiarity, and high-confidence false alarms are based on high levels of familiarity occurring in the absence of recollection of negative evidence.

For correct rejections, there is a pattern very similar to the pattern for hits, except that the latencies of high-confidence correct rejections are much slower. This supports the assumption that the correct rejections are based on the slower sampling and recovery processes, whereas high-confidence hits are much more likely to be based on the familiarity of the pair. It also worth noting the latency of low-confidence responses are little affected by the number of pair presentations, which supports the assumption that these are made when K_{MAX} has been exceeded, and neither the global-matching nor the sampling and recovery processes have been entirely successful in reaching K_O or K_N.

Thus, the present framework is shown to account for all of the critical findings that we reviewed, including the interactions between single-item recognition and associative recognition, between accuracy and latencies, between accuracy and confidence ratings, between latency and confidence ratings, between latencies and RK judgments, and between free response and signal-to-respond procedures. It should be clear, however, that simply implementing a recall-to-reject strategy does not uniquely determine the contribution of recollection to all aspects of performance. Constraints that are imposed by memory also impact performance. For instance, Fig. 10 is a diagram that illustrates the contribution of recollection to the performance of different tasks. Recollection plays almost no role in single-item recognition, and it plays an almost universal role in cued recall. Between these extremes, fall plurality discrimination, source memory, and associative recognition. Recollection plays a more significant role in associative recognition than in plurality discrimination because targets and foils are less similar to each other, and thus the sampling and recovery processes are more likely to produce details that cause one to reject a foil (Malmberg et al., 2004; Xu & Malmberg, 2007).

We know less about source memory within the current framework, but we do know that the similarity of the sources negatively affects performance (Bayen et al., 1996; Gruppuso, Lindsay, & Kelley, 1997; Mulligan & Hirshman, 1997). Hence, the contribution of recollection to source memory performance probably increases as the sources diverge in similarity. Under conditions where the sources are very dissimilar, the subject might rely on a single-process strategy (e.g., Dennis & Humphreys, 2001).

11.2.6. Implementing recognition strategies

There are relatively simple ways to implement recognition strategies (Malmberg, 2008). One way is to assume that subjects simply wait longer to make their responses when their understanding of task indicates that recollection of episodic details would be useful. Another way to reduce the role of recollection is by placing both decision boundaries close

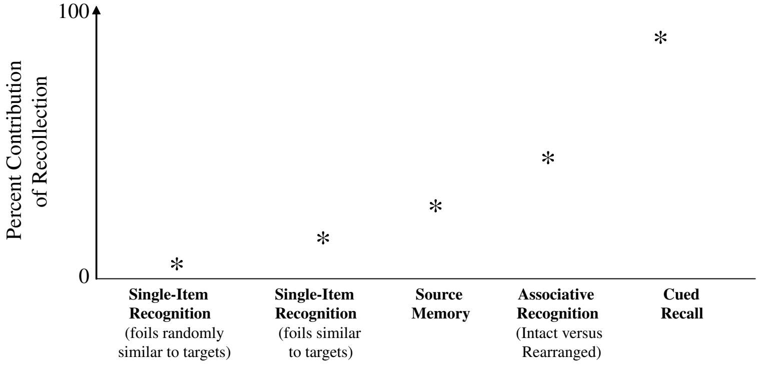

Fig. 10. Qualitative assumptions about the contribution of recollection to the performance of various recognition memory tasks.

to the starting point (Malmberg, 2008). This produces fast responding, like we find in single-item recognition. As the decision boundaries diverge, the contribution of recollection to performance increases. Indeed, single-item recognition is typically faster than associative recognition (Gronlund & Ratcliff, 1989; Nobel & Shiffrin, 2001).

If this model is correct, then we might expect to observe changes in bias as subjects switch from one task to another or if they switch from a situation where a recall-to-reject strategy is useful to one where it is not. There are a small number of findings supporting this prediction. For instance, Humphreys (1976, 1978) noted shifts in bias that occur between different pair-recognition tasks. There were lower FARs (at relatively steady hit rates) when task demands rendered the episodic details more useful. Signal detection analyses indicated that the changes in performance were due to changes in both accuracy and bias across tasks. In addition, Xu and Malmberg (2007) found that fits of the model suggested that recollection contributed more in the conditions in which pairs were comprised of word or faces versus non-words or Chinese characters, and they noted that fits of the model improved when bias varied between these conditions. Thus, there is some support for the assumption that the establishment of decision boundaries influences affects how tasks are performed.

11.3. The efficiency of recognition memory

One of the goals of this review was to classify different phenomena based on the models that can account for them. Another goal was to develop a coherent way to relate the different models. Efficiency is usually defined as a function of work per unit of time in the physical sciences, but it is also characteristic of several cognitive architectures, including SOAR (Laird, Newell, & Rosenbloom, 1987) and ACT-R (Anderson, 1993). In these models, efficiency refers to the capacity of the system to achieve a goal within the constraints of the time, resources, or space available.

Efficiency also plays an important role in metamemory research. For instance, Nelson and Narens (1990) proposed that students set a standard of learning prior to studying a list of items. Items judged to be relatively easy to learn should be given less time to be studied, and the efficient learner will achieve his desired level of mastery without spending any more time than is necessary. Typically, self-paced study time is related to ease-of-learning judgments, but subjects do not allocate enough time to achieve a mastery of the materials, and this is referred to as the "labor in vain effect".

The review of the literature also suggests that recognition memory is an efficient system, and that some models are more efficient than other models in performing specific tasks. The models described here focus on the retrieval of information, and the efficiency retrieval has not been emphasized or even explored in detail. Nevertheless, the construct of efficiency is applicable to recognition memory.

Recognition memory is the performance of a range of tasks that differ in what must be discriminated. There are different recognition strategies available to the subject, and their use reflects the nature of the task, the goals of the subject, and meta-level knowledge about how to achieve them. In the case of a subject faced with a large number of targets and foils to discriminate, for instance, the subject needs to implement a strategy that balances the amount of time spent on the task with the level of accuracy that he desires. The most efficient subject maximizes the probability of a correct response, while minimizing the amount of time devoted to completing the experiment. This is not to say, that the efficient subject will achieve perfect accuracy, as performance is limited by the constraints placed on him during study or even at test. Rather, the efficient subject allocates just enough time to the memory task in order achieve a desired level of accuracy.

11.3.1. Normative versus subjective efficiency

In the case of a computer simulation, one can readily define efficiency as the number of retrieval cycles necessary to achieve a given level of accuracy. The optimal strategy is one that achieves the best possible accuracy with the fewest retrieval cycles. Strategies that require more retrieval cycles to attain the same level of accuracy are less efficient. We can refer to this as *normative efficiency*, and the model can be used as a normative measurement tool.

The delay imposed on recollection in the current model makes the dual-process strategy slower than the global-matching strategy, and hence the dual-process strategy is more efficient only when it increases accuracy relative to the level achieved by the single-process strategy. Recollection does not significantly improve single-item recognition and, therefore, the normatively efficient subject should allocate less time to a single-item recognition task than to associative recognition. One reason might be because a more complex retrieval cue is required for associative recognition, but another reason is because recollection can provide information to improve associative recognition accuracy if the decision boundaries are set far enough apart.

These assumptions fit easily in the REM framework, which assumes that memory processes are optimal vis-à-vis the manner in which information is processed. In fact, the "retrieving effectively from memory" theory was originally named the "retrieving efficiently form memory" theory. There are, however, certain challenges associated with measuring normative efficiency in an absolute sense. The adoption of a retrieval strategy is a control process (Atkinson & Shiffrin, 1968), and the strategy reflects not only what is possible to achieve but also the goals and motivations of the subject. In the laboratory, for instance, we usually observe subjects who want to do well on memory tests, but who are also constrained by the demands of daily life. If the subject is under a great deal of time pressure, then he might not be willing to wait for the retrieval of all the available evidence before making a decision. Clearly, this subject has violated the standard of normative efficiency. However, his strategy is efficient with respect to his goal even if accuracy is

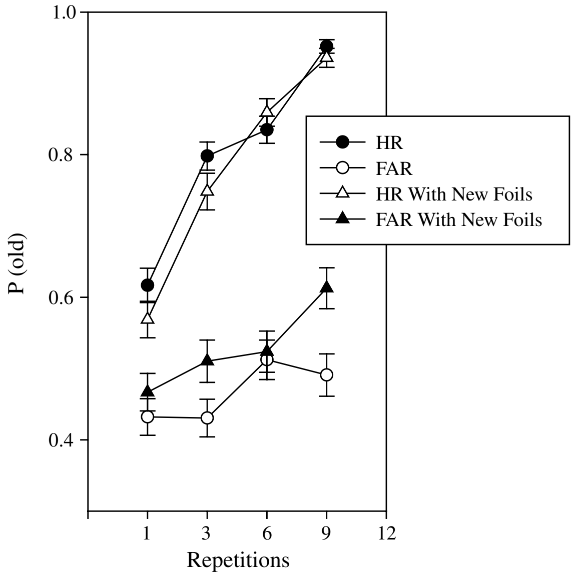

Fig. 11. The accuracy of plurality discrimination is affected by the composition of the test list.

lower than what is normatively possible. We can refer to the subject who attains his accuracy goal in shortest amount of time possible as being *subjectively efficient*.

One way to assess subjective efficiency of recognition memory is to compare the performance in situations where different strategies are available. For instance, the addition of new foils to the test list blurs the line between single-item recognition and associative recognition, and the question is, what strategy should the subject use? Compared to when new foils are not tested, a higher overall level performance is easier to achieve because the FAR for new foils will be less than the FAR for similar foils. Indeed, most of the items can be accurately discriminated based solely on their familiarity. Therefore, if the subject sets a similar level of accuracy as a goal in both situations and if he emphasizes global-matching when new foils are tested, then he will achieve the accuracy goal more quickly. The price paid, however, is an increase in the FAR for similar foils.

Adding new foils to the test list does in fact negatively affect the accuracy with which targets are discriminated from similar foils. Malmberg & Xu (2007) observed a reduction in associative recognition performance due an increase in FARs, and Heathcote et al. (2006) reported that source memory and plurality discrimination accuracy is lower when new items are tested. To assess whether the decrease in plurality discrimination performance is also due to a greater FAR for similar foils, I conducted a new analysis of the plurality discrimination data from Malmberg et al. (2004) and the data from a previously unreported experiment.[10] Fig. 11 shows that the results replicate Heathcote et al.'s (2006) finding insofar as accuracy was lower when new foils were tested, and the difference is due almost entirely to higher FARs [$F(1, 131) = 5.5$, $p < .02$].

[10] The new experiment differs from the published experiment only in that there were no new words on the test list. The unpublished experiment was conducted in 2001 a few days prior to the experiment reported by Malmberg et al. (2004).

According to the subjective efficiency principle, overall accuracy should be *at least* as high when new items are tested. That is, the subject makes no distinction between similar and randomly similar foils when determining the desired level of accuracy. The combined FAR is what determines overall performance, and the adoption of a faster strategy should not come at the expense of harming accuracy, if the subject is operating efficiently. Indeed, Malmberg & Xu (2007) observed almost equivalent levels of overall accuracy even though the patterns of responding were very different at times. Moreover, the plurality discrimination results (Fig. 11) indicate that overall performance improved when new items were tested; there were similar HRs but the difference in overall FARs was about .15. Second, the addition of new items to the test list should decrease response latencies, as subjects rely less on a recall-to-reject strategy. There are no relevant plurality discrimination results in the literature. However, this is exactly what Malmberg and Xu observed for associative recognition.

12. Final discussion

I reviewed a wide variety of models, their abilities to measure what they intend to measure, and their abilities to account for a large number of single-item recognition, associative recognition, and plurality discrimination findings. I chose these tasks because their combined data strains the models. I also considered accuracy and latency data and accuracy and confidence data. Again, by considering a wide range of findings, the models are further strained, and we observed the weaknesses in them.

Given these findings, I then extended several different REM models to account for the latencies of yes–no recognition, RK recognition, and confidence ratings. I also demonstrated the ability of the models to account for the retrieval dynamics of single-item recognition and associative recognition, and I related the models via the construct efficiency. I have no doubt, indeed we have seen it here, that other approaches to recognition can provide accurate accounts of some of the findings that we explored. However, I am equally confident that no existing framework can as readily account for the variety of findings that current one accounts for without significant revisions.

Acknowledgments

I thank Jing Xu for her invaluable contributions to this project and Mike Humphreys and an anonymous reviewer for their informative commentary on a previous version of this article.

Appendix A. Method and results of the list-strength RK experiment

A.1. Subjects

Thirty-two students enrolled in undergraduate psychology courses at Iowa State University participated in exchange for extra course credit.

A.2. Design and materials

Subjects studied three 30-item lists consisting of nouns with a normative frequency of between 20 and 50 per million (Kucera & Francis, 1983). One list consisted of items studied once for 2 s (pure-weak list), one list consisted of items studied three times for 2 s at spaced intervals (pure-strong list), and one list consisted of 15 weak items and 15 strong items (mixed list). Words were randomly assigned to lists and conditions for each subject. The order of the lists was determined randomly for each subject. Prior to testing, a distractor task was performed that consisted of mentally adding a series of digits. The length of the distractor task was varied to control for study-test lag (Murnane & Shiffrin, 1991). The length of distractor period was 140, 80, 20 s for pure-weak, mixed, and pure-strong conditions, respectively.

A.3. Procedure

Subjects were tested in individual subject booths while seated at a desktop computer that controlled the presentation of the stimuli and collected the data. The instructions indicated that their memory for the study items would be tested by showing them a series of test words. Half of the test words were studied and half were not. Subjects were instructed to respond "yes" if they thought they studied a word on the prior list and "no" if they thought that they did not. Subjects were told that sometimes an item could be recognized as having been studied because they could recall some of the details of the study event and sometimes they could recognize a word because it seems relatively familiar to them even though they could not recall some of the details of having studied the word. The former basis for recognition was likened to meeting a person on the street and recognizing that person from a recent party or class they attended. The later basis for recognition was likened to meeting a person on the street and not remembering where this person was met, but nevertheless knowing that they had met that person before.

If a "yes" response was made, subjects then indicated whether they responded "yes" because they remembered details of having studied the word or because they did not but the word seemed relatively familiar to them. In the former case, subjects clicked a "remember" button. In the later case, subjects clicked a "know" button. Subjects were instructed to ask the experimenter if they had any questions regarding the instructions. Otherwise, they were allowed to begin the experiment.

A.4. Results

A significance level of .05 is used. The mean pure-strong and mixed strong HRs were greater than the mean pure-weak and mixed–weak HRs [$t(31) = 6.75$, SEM = .02; $t(31) = 6.02$, SEM = .03, respectively]. The mean FAR was lower on the pure-strong than on the pure-weak list [$t(31) = 4.04$, SEM = .02].

A slightly negative list-strength effect was observed. The FAR was lower for pure-strong lists than for mixed lists [$t(31) = 3.13$, SEM = .02], but the HRs for strong items were not reliably different [$t(31) = .93$]. There was not a reliable difference in HRs [$t(31) = .19$] or FARs [$t(31) = .83$] for weak items on pure versus mixed lists.

The mean remember rate was greater for strong than weak targets [$F(1,31) = 115.6$, MSE = 1.80] but did not differ for mixed versus pure lists [$F(1,31) = 1.12$]. The mean know rate for targets was less for strong than weak items [$F(1,31) = 10.10$, MSE = .16] but did not differ for mixed versus pure lists [$F < 1.0$]. The mean remember [$t(31) = 2.09$, SEM = .02] and know rates [$t(31) = 2.93$, SEM = .02] were greater for mixed versus pure-strong lists (i.e., the FAR was greater for mixed than pure lists), but the means were not reliably different for mixed versus pure-weak lists.

References

Anderson, R. C., & Watts, G. H. (1971). Response competition in the forgetting of paired associates. *Journal of Verbal learning and Verbal Behavior, 10*, 29–34.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22*, 261–295.

Anderson, J. R. (1993). *Rules of the mind*. Hillsdale, NJ: Earlbaum.

Ardnt, J., & Reder, L. M. (2002). Word frequency and receiver-operating characteristic curves in recognition memory: Evidence for a dual-process interpretation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 830–842.

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology. Learning memory and thinking* (Vol. 1, pp. 243–293). San Francisco: Freeman.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed systems and its control processes. In W. E. Spence & J. T. Spence (Eds.). *The psychology of learning and motivation* (Vol. 2). New York: Academic.

Bahrick, H. P. (1965). The ebb of retention. *Psychological Review, 72*, 60–73.

Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. A. (2002). The word-frequency mirror effect in young, old, and early-stage Alzheimer's disease: Evidence for two processes in episodic recognition memory. *Journal of Memory and Language, 46*(1), 199–226.

Balota, D. A., Dolan, P. O., & Duchek, J. M. (2000). Memory changes in healthy older adults. In E. Tulving & F. I. M. Craik (Eds.), *The oxford handbook of memory* (pp. 395–409). NY: Oxford.

Balota, D. A., & Neely, J. H. (1980). Text expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 576–587.

Banks, W. P. (1970). Signal detection theory and human memory. *Psychological Bulletin, 74*, 81–99.

Banks, W. P. (2000). Recognition and source memory as multivariate decision processes. *Psychological Science, 11*(4), 267–273.

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.

Batchelder, W. H., & Riefer, D. M. (1999). Theoretical and empirical review of multinomial processing tree modeling. *Psychonomic Bulletin & Review, 6*, 57–86.

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 197–215.

Bernbach, H. A. (1970). Decision process in memory. *Psychological Review, 74*, 462–480.

Broadbent, D. E. (1973). *In defense of empirical psychology*. London: Methuen.

Broadbent, D. E., & Broadbent, M. H. P. (1975). The recognition of words which cannot be recalled. In P. M. A. Rabbin & S. Dornic (Eds.), *Attention and performance V*. New York: Academic Press.

Broadbent, D. E., & Broadbent, M. H. P. (1977). Effects of recognition on subsequent recall: Comments on "Determinants of Recognition and Recall: Accessibility and Generation" by Rabinowitz, Mandler, and Patterson. *Journal of Experimental Psychology: General, 106*(3), 330–335.

Buchner, A., & Erdfelder, E. (1996). On assumptions of, relations between, and evaluations of 1 some process dissociation measurement models. *Consciousness and Cognition, 5*(4), 581–594.

Cansino, S., Maquet, P., Dolan, R. J., & Rugg, M. D. (2002). Brain activity underlying encoding and retrieval of source memory. *Cerebral Cortex, 12*(10), 1048–1056.

Cary, M., & Reder, L. M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *Journal of Memory and Language, 49*(2), 231–248.

Clark, S. E. (1997). A familiarity-based account of confidence-accuracy inversions in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*(1), 232–238.

Clark, S. E., & Burchett, E. E. R. (1994). Word frequency and list composition effects in associative recognition and recall. *Memory & Cognition, 22*, 55–62.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review, 3*, 37–60.

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

Cook, G. I., Marsh, R. L., & Hicks, J. L. (2006). The role of recollection and familiarity in the context variability mirror effect. *Memory & Cognition, 34*, 240–250.

Craik, F. I. M., & McDowd, J. M. (1987). Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 474–479.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268–294.

Criss, A. H. (2006). The consequences of differentiation in episodic memory: Similarity and the strength based mirror effect. *Journal of Memory & Language: Special Issue on Computational Models of Memory, 55*, 461–478.

Criss, A. H., & 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, 55*, 447–460.

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

Curran, T., DeBuse, C., Woroch, B., & Hirshman, E. (2006). Combined pharmacological and electrophysiological dissociation of familiarity and recollection. *Journal of Neuroscience, 26*, 1979–1985.

Davachi, L., Mitchell, J. P., & Wagner, A. D. (2003). Multiple learning mechanisms: Distinct medial temporal processes build item and source memories. *Proceeding of the National Academy of Science of the United States of America, 100*(4), 2157–2162.

Deese, J. (1960). Frequency of usage and number of words in free recall: The role of association. *Psychological Review, 7*, 337–344.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review, 108*, 452–478.

Diana, R., & Reder, L. M. (2005). The list strength effect: A contextual competition account. *Memory & Cognition, 33*(7), 1289–1302.

Diana, R., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual process account. *Psychonomic Bulletin & Review, 13*, 1–21.

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, and Cognition, 27*(2), 414–435.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24*, 523–533.

Dorfman, D., & Glanzer, M. (1988). List composition effects in lexical decisions and recognition memory. *Journal of Memory and Language, 27*, 633–648.

Dosher, B. A. (1981). The effects of delay and interference: A speed–accuracy study. *Cognitive Psychology, 13*, 551–582.

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

Dunn, J. A. (2004). RK: A matter of confidence. *Psychological Review, 111*, 524–542.

Dyne, A. M., Humphreys, M. S., Bain, J. D., & Pike, R. (1990). Associative effects in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(5), 813–824.

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

Eldridge, L. L., Knowlton, B. J., Furmanski, C. S., Bookheimer, S. Y., & Engel, S. A. (2000). Remembering episodes: A selective role for the hippocampus during retrieval. *Nature Neuroscience, 3*(11), 1149–1152.

Erdfelder, E., & Buchner, A. (1998). Process-dissociation measurement models: Threshold theory or detection theory? *Journal of Experimental Psychology: General, 127*, 83–96.

Estes, W. K., & DaPalito, F. (1967). Independent variation of information storage and retrieval processes in paired-associate learning. *Journal of Experimental Psychology, 75*, 18–26.

Estes, W. K. (1994). *Classification and cognition*. New York: Oxford.

Estes, W. K. (1975). Some targets for mathematical psychology. *Journal of Mathematical Psychology, 12*, 263–282.

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

Gillund, G., & Shiffrin, R. M. (1981). Free recall of complex pictures and abstract words. *Journal of Verbal Learning and Verbal Behavior, 20*, 575–592.

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

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 12*, 8–20.

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

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review, 100*, 546–567.

Glanzer, M., Kim, K., Adams, J. K., & Hilford, A. (1999). Slope of the receiver operating characteristic in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 500–513.

Gold, J. J., Smith, C. N., Bayley, P. J., Shrager, Y., Brewer, J. B., Stark, C. E. L., et al. (2006). Item memory, source memory, and the medial temporal lobe: Concordant findings from fMRI and memory-impaired patients. *Proceedings of the National Academy of Sciences of the United States of America, 103*(24), 9351–9356.

Gorman, A. M. (1961). Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology, 61*, 23–29.

Gregg, V. H. (1976). Word frequency, recognition, and recall. In J. Brown (Ed.), *Recall and recognition*. London: Wiley.

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

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.

Gruppuso, V., Lindsay, D. S., & Kelley, C. M. (1997). The process-dissociation procedure and similarity: Defining and estimating recollection and familiarity in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 259–278.

Hamann, S. (2001). Cognitive and neural mechanisms of emotional memory. *Trends in Cognitive Sciences, 5*, 394–400.

Hay, J. F., & Jacoby, L. L. (1999). Separating habit and recollection in young and elderly adults: Effects of elaborative processing and distinctiveness. *Psychology and Aging, 12*, 122–134.

Heathcote, A. (2003). Item recognition memory and the receiver operating characteristic. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 1210–1230.

Heathcote, A., Raymond, F., & Dunn, J. (2006). Recollection and familiarity in recognition memory: Evidence from ROC curves. *Journal of Memory and Language, 55*, 495–514.

Henson, R. N., Cansino, S., Herron, J. E., Robb, W. G. K., & Rugg, M. D. (2003). A familiarity signal in human anterior medial temporal cortex. *Hippocampus, 13*, 259–262.

Henson, R. N., Rugg, M. D., Shallice, T., Josephs, O., & Dolan, R. J. (1999). Recollection and familiarity in recognition memory: An event-related fMRI study. *Journal of Neuroscience, 19*, 3962–3972.

Hertel, P. T., & Parks, C. (2002). Emotional episodes facilitate word recall. *Cognition and Emotion, 16*, 685–694.

Hicks, J. L., Marsh, R. L., & Cook, G. I. (2005). An observation on the role of context variability in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 1160–1164.

Hilford, A., Glanzer, M., Kim, K., & DeCarlo, L. T. (2002). Regularities of source recognition: ROC analysis. *Journal of Experimental Psychology: General, 131*(4), 494–510.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 77–99). Potomac, MD: Erlbaum.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace model. *Psychological Review, 95*, 528–551.

Hintzman, D. L., & Curran, T. (1994). Retrieval dynamics of recognition and frequency judgments: Evidence for separate processes of familiarity and recall. *Journal of Memory and Language, 33*, 1–18.

Hintzman, D. L., & Curran, T. (1995). When encoding fails: Instructions, feedback, and registration without learning. *Memory & Cognition, 23*, 213–226.

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.

Hirshman, E. (1995). Decision processes in recognition memory: Criterion shifts and the list strength paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 302–313.

Hirshman, E., & Arndt, J. (1997). Discriminating alternative conceptions of false recognition: The cases of word frequency and word concreteness. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 1306–1323.

Hirshman, E., Fisher, J., Henthorn, T., Arndt, J., & Passannante, A. (2002). Midazolam amnesia and dual-process models of the word frequency mirror effect. *Journal of Memory and Language, 47*, 499–516.

Hirshman, E., & Henzler, A. (1998). The role of decision processes in conscious memory. *Psychological Science, 9*, 61–64.

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., & Murdock, B. B. (1987). A decision model for accuracy and response latency in recognition memory. *Psychological Review, 94*(3), 341–358.

Humphreys, M. S. (1976). Relational information and the context effect in recognition memory. *Memory & Cognition, 4*, 221–232.

Humphreys, M. S. (1978). Item and relational information: A case for context independent retrieval. *Journal of Verbal Learning and Verbal Behavior, 17*, 175–188.

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

Humphreys, M. S., & Bain, J. D. (1991). Episodic strength components in recognition memory. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data; essays in honor of Bennet B. Murdock* (pp. 309–330). Hillsdale, NJ: Erlbaum.

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different way to cue a coherent memory system: A theory of episodic, semantic, and procedural tasks. *Psychological Review, 96*, 208–233.

Jackson, O., & Schacter, D. L. (2004). Encoding activity in anterior medial temporal lobe supports associative recognition. *NeuroImage, 21*, 456–464.

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

Joordens, S., & Hockley, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1534–1555.

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

Kintsch, W. (1967). Memory and decision aspects of recognition learning. *Psychological Review, 74*, 496–504.

Kirwan, C. B., & Stark, C. E. (2004). Medial temporal lobe activation during encoding and retrieval of novel face-name pairs. *Hippocampus, 14*, 919–930.

Koppell, S. (1977). Decision latencies in recognition memory: A signal detection analysis. *Journal of Experimental Psychology: Human Learning and Memory, 3*(4), 445–457.

Krantz, D. H. (1969). Threshold theories of signal detection. *Psychological Review, 76*, 308–324.

Krueger, W. C. F. (1929). The effects of overlearning on retention. *Journal of Experimental Psychology, 12*, 71–78.

Kucera, H., & Francis, W. (1983). *Computational analysis of present day American English*. Providence, RI: Brown University Press.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). SOAR: an architecture for general intelligence. *Artificial Intelligence, 33*(1), 1–64.

Lehman, M. L., & Malmberg, K. J. (submitted for publication). A global memory model of remembering and forgetting from multiple lists.

Lepage, M., Habib, R., & Tulving, E. (1998). Hippocampal PET activations of memory encoding and retrieval: The HIPER model. *Hippocampus, 8*, 313–322.

Lewandowsky, S., & Heit, E. (2006). Some targets for memory models. *Journal of Memory and 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 (12).

Light, L. L., Chung, C., Pendergrass, R., & Van Ocker, J. C. (2006). Effects of repetition and response deadline on item recognition in young and older adults. *Memory & Cognition, 34*(2), 335–343.

Lockhart, R. S., & Murdock, B. B. (1970). Memory and theory of signal detection. *Psychological Review, 74*, 100–109.

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

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge, England: Cambridge University Press.

Macmillan, N. A., & Rotello, C. M. (2006). Deciding about decision models of remember and know judgments: A reply to Murdock. *Psychological Review, 113*, 657–665.

Maddox, W. T., & Estes, W. K. (1997). Direct and indirect stimulus frequency effects in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 539–559.

Malmberg, K. J. (2002). On the form of ROCs constructed from confidence ratings. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 380–387.

Malmberg, K. J. (2008). Towards an understanding of individual differences in episodic memory: Modeling the dynamics of recognition memory. In A. Benjamin & B. Ross (Eds.). *The psychology of learning and motivation: Skill and strategy in memory use* (Vol. 48, pp. 313–349). San Diego, CA: Elsevier.

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., & Murnane, K. (2002). List composition and the word-frequency effect for recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(4), 616–630.

Malmberg, K. J., & Shiffrin, R. M. (2005). The "one-shot" hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(2), 322–336.

Malmberg, K. J., & Xu, J. (2006). The influence of averaging and noisy decision strategies on the recognition memory ROC. *Psychonomic Bulletin & Review, 13*(1), 99–105.

Malmberg, K. J., & Xu, J. (2007). On the flexibility and on the fallibility of associative memory. *Memory & Cognition, 35*(3), 545–556.

Malmberg, K. J., Zeelenberg, R., & Shiffrin, R. M. (2004). Turning up the noise or turning down the volume? On the impairment of recognition memory by midazolam. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

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

Mandler, G. (1991). Your face looks familiar but I can't remember your name: A review of dual-process theory. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data; essays in honor of Bennet B. Murdock* (pp. 207–226). Hillsdale, NJ: Erlbaum.

Mandler, G., Pearlstone, Z., & Koopmans, H. S. (1969). Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning and Verbal Behavior, 8*, 410–423.

Manns, J. R., Hopkins, R. O., Reed, J. R., Kitchener, E. G., & Squire, L. R. (2003). Recognition memory and the human hippocampus. *Neuron, 38*, 127–133.

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological Review, 105*, 724–760.

McElree, B., Dolan, P. O., & Jacoby, L. L. (1999). Isolating the contributions of familiarity and source information in item recognition: A time-course analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 563–582.

Mulligan, N. W., & Hirshman, E. (1997). Measuring the bases of recognition memory: An investigation of the process dissociation framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 280–304.

Muller, G. E. (1913; cited by Mandler et al., 1969). Zur Analyse der Gedichtnistitigkeitund des Vorstellungs-verlaufes. III. Teil. *Zeitschrift fiir Psychologie*, Erganzungsband, 8.

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

Murdock, B. B. (1985). An analysis of the strength-latency relationship. *Memory & Cognition, 13*(6), 511–521.

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

Murdock, B. B. (2006). Decision-making models of remember–know judgments: Comment on Rotello, Macmillan, and Reeder (2004). *Psychological Review, 113*, 648–656.

Murdock, B. B., & Anderson, R. E. (1975). Encoding, storage, and retrieval of item information. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium*. Hillsdale, NJ: Earlbaum.

Murdock, B. B., & Kahana, M. J. (1993). Analysis of the list-strength effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 689–697.

Murnane, K., & Shiffrin, R. M. (1991). Interference and the representation of events in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 855–874.

Naveh-Benjamin, M. (2000). Adult-age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1170–1187.

Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist, 51*, 102–116.

Nelson, T. O., McSpadden, M., Fromme, K., & Marlatt, G. A. (1986). Effects of alcohol intoxication on metamemory and on retrieval from long-term memory. *Journal of Experimental Psychology: General, 115*, 247–254.

Please cite this article in press as: Malmberg, K. J., Recognition memory: A review of the..., *Cognitive Psychology* (2008), doi:10.1016/j.cogpsych.2008.02.004

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.

Nobel, P. A., & Shiffrin, R. M. (2001). Retrieval processes in recognition and cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 384–413.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary learning systems approach. *Psychological Review, 110*(4), 611–646.

Park, H., Reder, L. M., & Dickison, D. (2005). The effects of word frequency and similarity on recognition judgments: The role of recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(3), 568–578.

Parks, C. M., & Yonelinas, A. P. (2007). Moving beyond pure signal detectionmodels: Comment on Wixted (2007). *Psychological Review, 114*, 188–202.

Postman, L., & Stark, K. (1969). Role of response availability in transfer and interference. *Journal of Experimental Psychology, 79*, 168–177.

Qin, J., Raye, C. L., Johnson, M. K., & Mitchell, K. J. (2001). Source ROCs are (typically) curvilinear: Comment on Yonelinas (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1110–1115.

Raaijmakers, J. G. W. (1979). *Retrieval from long-term store: A general theory and mathematical models*. Druk: Stichting Studentenpers, Nijmegen.

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

Rabinowitz, J. C., Mandler, G., & Patterson, K. E. (1977). Determinants of recognition and recall: Accessibility and generation. *Journal of Experimental Psychology: General, 106*, 302–329.

Ranganath, C., Yonelinas, A. P., Cohen, M. X., Dy, C. J., Tom, S. M., & D'Esposito, M. (2004). Dissociable correlates of recollection and familiarity within the medial temporal lobes. *Neuropsychologia, 42*, 2–13.

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

Ratcliff, R. (2006). Modeling response signal and response time data. *Cognitive Psychology, 53*, 195–237.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 163–178.

Ratcliff, R., & McKoon, G. (1991). Using ROC data and priming results to test global memory models. In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data; essays in honor of Bennet B. Murdock* (pp. 279–296). Hillsdale, NJ: Erlbaum.

Ratcliff, R., McKoon, G., & Tindall, M. (1994). Empirical generality of data from recognition memory receiver-operating characteristic functions and implications for the global memory models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 763–785.

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

Ratcliff, R., Sheu, C.-F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review, 99*, 518–535.

Reder, L. M., Angstadt, P., Cary, M., Erickson, M. A., & Ayers, M. A. (2002). A reexamination of stimulus-frequency effects in recognition: Two mirrors for low- and high-frequency pseudowords. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 138–152.

Reder, L. M., Donavos, D. K., & Erickson, M. A. (2002). Perceptual match effects in direct tests of memory: The role of contextual fan. *Memory & Cognition, 30*(2), 312–323.

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 RK judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 294–320.

Reed, A. V. (1973). Speed–accuracy tradeoff in recognition memory. *Science, 181*, 574–576.

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition, 28*(6), 907–922.

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

Rottelo, C. M., Macmillan, N. A., & Van Tessel, G. (2000). Recall-to-reject in recognition memory: Evidence from ROC curves. *Journal of Memory and Language, 43*, 67–88.

Rugg, M. D., Henson, R. N., & Robb, W. G. (2003). Neural correlates of retrieval processing in the prefrontal cortex during recognition and exclusion tasks. *Neuropsychologia, 41*(1), 40–52.

Schacter, D. L., & Wagner, A. D. (1999). Medial temporal lobe activations in fMRI and PET studies of episodic encoding and retrieval. *Hippocampus, 9*, 7–24.

Schulman, A. I. (1967). Word length and rarity in recognition memory. *Psychonomic Science, 9*, 211–212.

Sheffert, S. M., & Shiffrin, R. M. (2003). Auditory "registration without learning". *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*, 10–21.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6*, 156–163.

Sherman, S. J., Atri, A., Hasselmo, M. E., Stern, C. E., & Howard, M. W. (2003). Scopolamine impairs human recognition memory: Data and modeling. *Behavioral Neuroscience, 117*, 526–539.

Shiffrin, R. M., Huber, D., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(2), 267–287.

Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). List-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 179–195.

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). London: Oxford University Press.

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

Slamecka, N. J., & McElree, B. (1983). Normal forgetting of verbal lists as a function of their degree of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 384–397.

Slotnick, S. D., Klein, S. A., Dodson, C. S., & Shimamura, A. P. (2000). An analysis of signal detection and threshold models of source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1499–1517.

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

Smith, E. E., Shoben, E. J., & Rips, L. J. (1974). Structure and process in semantic memory: A featural model for semantic decision. *Psychological Review, 81*, 214–241.

Sperling, R., Chua, E., Cocchiarella, A., Rand-Giovannetti, E., Poldrack, R., Schacter, D. L., et al. (2003). Putting names to faces: Successful encoding of associative memories activates the anterior hippocampal formation. *NeuroImage, 20*, 1400–1410.

Soderlund, H., Parker, E. S., Schwartz, B. L., & Tulving, E. (2005). Memory encoding and retrieval on the ascending and descending limbs of the blood alcohol concentration curve. *Psychopharmacology, 182*, 305–317.

Steyvers, M., & Malmberg, K. J. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(5), 760–766.

Staresina, B. P., & Davachi, L. (2006). Differential encoding mechanisms for subsequent associative recognition and free recall. *The Journal of Neuroscience, 26*(36), 9162–9172.

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1379–1396.

Tuiving, E. (1981). Similarity relations in recognition. *Journal of Verbal Learning and Verbal Behavior, 20*, 479–496.

Tulving, E. (1983). *Elements of episodic memory*. Oxford: Clarendon Press.

Tulving, E., & Pearstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5*, 381–391.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352–373.

Van Zandt, T. (2000). ROC curves and confidence judgments in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 582–600.

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

Wais, P., Wixted, J. T., Hopkins, R. O., & Squire, L. R. (2006). The hippocampus supports both the recollection and the familiarity components of recognition memory. *Neuron, 49*, 459–468.

Windmann, S., & Kutas, M. (2001). Electrophysiological correlates of emotion-induced recognition bias. *Journal of Cognitive Neuroscience, 13*(5), 577–592.

Winograd, E. (1968). List differentiation as a function of frequency and retention interval. *Journal of Experimental Psychology, 76*(2, part 2).

Weeks, C. S., Humphreys, M. S., & Hockley, W. E. (submitted for publication). Buffered forgetting: When targets and distractors are both forgotten.

Wixted, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 681–690.

Wixted, J. T. (2007). Spotlighting the probative findings: A reply to Parks and Yonelinas (2007). *Psychological Review, 114*, 203–209.

Wixted, J. T., & Ebbesen, E. (1997). Genuine power curves in forgetting. *Memory & Cognition, 25*, 731–739.

Wixted, J. T., & Stretch, V. (2004). In defense of the signal-detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review, 11*, 616–641.

Wixted, J. T., & Squire, L. R. (2004). Recall and recognition are equally impaired in patients with selective hippocampal damage. *Cognitive, Affective & Behavioral Neuroscience, 4*, 58–66.

Xu, J., & Malmberg, K. J. (2007). Modeling the effects of verbal- and nonverbal-pair strength on associative recognition. *Memory & Cognition, 35*(3), 526–544.

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

Yonelinas, A. P. (1997). Recognition memory ROCs for item and associative information: The contribution of recollection and familiarity. *Memory & Cognition, 25*, 747–763.

Yonelinas, A. P. (1999). The contribution of recollection and familiarity to recognition and source memory: An analysis of receiver operating characteristics and a formal model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(6), 1415–1434.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441–517.

Yonelinas, A. P., & Jacoby, L. L. (1996). Response bias and the process-dissociation procedure. *Journal of Experimental Psychology: General, 125*, 422–434.

Yonelinas, A. P., Kroll, N. E. A., Dobbins, I. G., Lazzara, M., & Knight, R. T. (1998). Recollection and familiarity deficits in amnesia: Convergence of remember/know, process dissociation, and receiver operating characteristic data. *Neuropsychology, 12*, 1–17.

Yonelinas, A. P., Otten, L. J., Shaw, K. N., & Rugg, M. D. (2005). Separating the brain regions involved in recollection and familiarity in recognition memory. *Journal of Neuroscience, 25*(11), 3002–3008.