Are "remember" And "know" The Same Process? —a Perspective From Reaction Time Data

ARE “REMEMBER” AND “KNOW” THE SAME PROCESS?
—A PERSPECTIVE FROM REACTION TIME DATA

A Thesis Presented

by

MIN ZENG

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE

September 2007

Psychology

ARE “REMEMBER” AND “KNOW” THE SAME PROCESS?
—A PERSPECTIVE FROM REACTION TIME DATA

A Thesis Presented

by

MIN ZENG

Approved as to style and content by:

Caren Rotello, Chair

Neil Macmillan, Member

Andrew Cohen, Member

Neil Berthier, Member

Melinda Novak, Chair
Department of Psychology

ACKNOWLEDGMENTS

I wish to thank Caren, for giving me the opportunity to do this project and for her steadfast advice throughout the process. From Caren, I learned to become more meticulous, more patient and more critical. However, what I cannot learn is her passion for research.

In this thesis, I borrowed two figures (Figures 1 and 2) from Caren, Neil and John's article (2004). In addition, I appropriated some ideas from Caren and Neil.

I'm also grateful to my thesis committee members. I would like to thank Neil Berthier for his instruction and help extended in my graduate life, from ABCD to OPT. I was able to understand signal detection theory deeper from all the illuminating talks Neil Macmillan gave us at cognitive seminar and our lab meetings. Thanks to Andrew's modeling seminar, I learned much about modeling and refreshed my rusty Matlab programming skills.

Luckily I took day seminars from almost every faculty in Cognitive program, Andrew, Chuck, Keith, Caren, Lisa and Kyle. That was invaluable experience for me.

My gratitude also goes to Mungchen for his unceasing help and support. He showed me the skills to start with and the attitude to finish with. I would like to thank Aycan, Kris, and Chad for making the lab a cozy place.

Finally, I would like to thank Zixu and my family. Were not for their love and support, I wouldn't have been this healthy and happy.

ABSTRACT

ARE "REMEMBER" AND "KNOW" THE SAME PROCESS?
—A PERSPECTIVE FROM REACTION TIME DATA

SEMPTEMBER 2007

MIN ZENG, B.E., UNIVERSITY OF ELECTRONIC SCIENCE & TECHNOLOGY OF CHINA

M.E., UNIVERSITY OF ELECTRONIC SCIENCE & TECHNOLOGY OF CHINA

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Caren Rotello

The remember-know paradigm is widely used in recognition memory research to explore the mechanisms underlying recognition judgments. The most intriguing question about the paradigm that needs to be answered is: Are the processes that underlie "remember" and "know" responses the same or different? The extant remember-know models provide different answers. The dual-process model (Yonelinas, 1994) assumes that "remember" and "know" judgments are made with qualitatively different underlying processes. The one-dimensional Signal Detection Theory (SDT) model (Donaldson, 1996; Hirshman & Master, 1997) and the Sum-difference Theory of Remembering and Knowing (STREAK) model assume that "remember" and "know" judgments are made with same underlying processes but different response criteria. In this thesis, three experiments were conducted to evaluate these models. The remember-know models were fit to the accuracy data to see which model provides the best account for the ROC data. In

addition, the reaction time data were fit with ex-Gaussian distributions and the best-fit skew parameters were used to reveal whether the underlying strategic processes for “remember” and “know” judgments are same or not.

The results of the remember-know model fit were mixed: In the first experiment with list length manipulation, 6 out of 8 cases were best fit with the one-dimensional models and the other 2 cases were best fit with the dual-process models; in the second experiment with list strength manipulation, 11 out of 18 cases were best fit with the one-dimensional models, another 6 cases were best fit with the dual-process models and the rest one case were best fit with the STREAK model; in the third experiment with response bias manipulation, 6 out of 16 cases were best fit with the one-dimensional models and the other 10 cases were best fit with the dual-process models.

The results of ex-Gaussian fit to RT data supported the one-dimensional model better: for the subjects who provide enough overlapping data in comparison of the distributions of hits followed by “remember” and “know” judgments, the values of skew parameter did not differ for “remember” and “know” responses in 7 out of 8 cases. This indicates that the same process underlies “remember” and “know” responses.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

You have probably had an experience like this one: You meet a man and feel that his face looks so familiar that you are sure that you have met him before, but unfortunately you just cannot recall who he is or when or where you met him before. Cautiously, you greet each other. You turn around, keeping recalling. Then suddenly, it occurs to you that you met him in a symposium and he is a friend of your colleague. The implication of such experiences is that recognizing a person or an event can consist of the assessment of familiarity and the recollection of the specific episode during which the person or the event was encountered previously.

To investigate the subjective experience accompanying retrieval in recall and recognition tasks, Tulving (1985) proposed the remember-know paradigm. During a typical remember-know experiment, subjects are presented with a list of words or pictures for study. Later, the test list includes studied items (targets) mixed with unstudied items (lures). The subjects are asked to respond "old" for the items they think they have already studied and "new" for the items they think were not in the study list. Further, the subjects make "remember" or "know" responses for those they consider to have been studied. "Know" means that the subject feels that the test item looks familiar, and is sure that this word was studied, but does not recollect specific details. In contrast, "remember" means that the subject can recall some details of the actual occurrence of the item in the study list (Rajaram, 1993).

The remember-know paradigm allows subjects to report the subjective basis of their recognition judgment, thus it provides a possible window onto their memory retrieval processes. The question in focus in remember-know research is the underlying basis of "remember" and "know" responses.

1.1 General findings from behavioral experiments

Previous research has found that "remember" and "know" responses can be dissociated under many circumstances, which has been taken as evidence that they are functionally independent (See Gardiner & Richardson-Klavehn, 2000 for a review). The manipulations in these experiments have included impoverished versus elaborated encoding such as divided versus full attention, shallow versus deep level of processing and reading versus generating (Gardiner, 1988; Gardiner et al, 1996). Such manipulations influence the proportion of "remember" responses but do not affect the "know" response rate. In contrast, manipulations of perceptual fluency affect "know" but not "remember" responses (Rajaram, 1993). Nonword versus word presentation and massed versus spaced repetition of study list items increase "know" responses and decrease "remember" responses (Gardiner & Java, 1990; Parkin & Russo, 1993). Other manipulations, such as long versus short response deadlines have been shown to cause parallel changes on "remember" and "know" responses (Gardiner et al, 1998).

"Remember" and "know" have also been dissociated by varying the ages of the participants. For older adults, "remember" responses are reduced compared with young subjects, whereas "know" response rates are similar (Perfect et al., 1995). Patients with amnesia, Alzheimer's disease, or schizophrenia also make fewer "remember" responses than control subjects, but their "know" response rates are similar (Dalla Barba, 1993, 1997; Knowlton & Squire, 1995; Schacter et al., 1997; Huron et al., 1995).

Such functional dissociations have been thought to reflect the independence of "remember" and "know" processes (Kelley & Jacoby, 1998; Reder et al., 2000; Yonelinas, 2001; Murdock, 2006). However, a single process can also account for such dissociations (Dunn & Kirsner, 1988; Dunn & Kirsner, 2003). Recent research shows that, without a model, it is hard to decide whether such dissociations are due to sensitivity change, as is generally assumed, or due to response bias, so that dissociation effects are difficult to interpret (Rotello et al., 2006).

1.2 Brain imaging research on remember-know

Like most of the behavioral research, the research conducted with brain imaging technologies generally assumes that "remember" responses tap recollection and "know" responses tap familiarity. This is also the claim of the dual-process model (Yonelinas, Dobbins, Szymanski, Dhaliwal, & King, 1996), as will be described later. Using this dual-process interpretation, event-related potential (ERP) components have been found that correspond to familiarity and recollection respectively (Allan, Wilding, & Rugg, 1998; Curran, 2000; Curran, 2003; see Wixted, 2007 for alternative interpretation). Two components are related to recollection: A left parietal component with post-stimulus onset of 400ms lasting around 400-600ms, and a right frontal sustained positive component with the same onset of 400ms lasting for about 1.5s. The former component is thought to be a correlate of the successful retrieval of episodic information and the latter is hypothesized to generate or maintain a representation of the study episode. In contrast, the familiarity process is associated with a frontal negativity old/new effect (FN400 with post-stimulus onset of 300ms lasting about 200ms, Curran, 2000).

In ERP studies, the conclusion that recollection and familiarity are differentiated in latency and brain topography is quite common. However, that interpretation is rooted in the dual-process model on the strong assumption that recollection alone contributes to “remember” responses and familiarity alone contributes to “know” responses. If the dual-process model is incorrect (Wixted, 2007), the interpretation of these data is less clear.

While ERPs are useful for their excellent temporal resolution, fMRI is used for its excellent spatial resolution to localize the underlying neural substrates of different mental processes. The general findings of fMRI studies led to two points of view: One is that “remember” and “know” responses dissociate in the loci of neural responses (Henson, Rugg, Shallice, Josephs, & Dolan, 1999; Yonelinas, Otten, Shaw, & Rugg, 2005); the other is that “remember” and “know” associate with the same processes but that “remember” additionally recruits regions specific to retrieved content (Wheeler & Buckner, 2004).

In the study done by Wheeler and Buckner (2004), the subjects studied a list of words with half of the words followed by pictures (each picture was uniquely related to its preceding word), and the other half followed by sounds (each sound was uniquely related to its preceding word). Later these words were tested mixed with new words. The authors found that both remembering and knowing associated with enhanced responses in regions that are activated by “old” judgments relative to “new” judgments, and remembering is additionally associated with the same regions that were activated by the sensory specific (visual or auditory) information during study phase. These findings are consistent with the viewpoint assuming the relationship between “know” and “remember” responses is one of redundancy: “know” processes are also active during remembering (Knowlton, 1998). This is in contrast with the independence of "remember" and "know" processes in the dual-process model.

Though a lot of ERP and fMRI studies have been done to investigate the time course and underlying neural substrates of "remember" and "know" responses, the results are mainly interpreted with the framework of the dual-process model. If the credibility of the dual-process model is in question, these conclusions are also suspicious. Neuroimaging studies could be useful for investigating processes that underlie "remember" and "know" judgments if the researchers examine different accounts of the remember-know paradigm rather than adopting a single framework to interpret results.

1.3 Remember-know models

A different approach to explicate the process or processes that underlie "remember" and "know" judgments is mathematical modeling. There are three extant models, each with different underlying assumptions about what "remember" and "know" mean: the dual-process model (Yonelinas et al., 1994; Reder et al., 2000; Murdock, 2006), the one dimensional Signal Detection Theory (SDT) model (Donaldson, 1996; Hirshman & Master) and a two-dimensional SDT model, STREAK (Sum-Difference Theory of Remembering and Knowing, Rotello, Macmillan & Reeder, 2004).

In the dual-process theory (Yonelinas et al., 1994; Reder et al., 2000; Murdock, 2006), recognition judgments are based on either recollection or a familiarity process. Yonelinas (1994, 1996, 2001) assumes that recollection is an all-or-none, high-threshold process while familiarity works according to an equal variance signal detection model. If the test words evoke recollection, subjects respond "remember" and if the test words provoke feelings of familiarity, they respond "know". Therefore, "remember" judgments are direct measures of recollection and "know" judgments are based on familiarity. Dual-process models are essentially quantitative versions of the naïve process-pure view described earlier. In contrast to Yonelinas' dual-process model, the Sources of Activation Confusion (SAC) model of Reder et al. (2000) and Murdock's extension of TODAM (2006) both assume that recollection and familiarity work according to an equal-variance signal detection model. In the studies presented in this thesis, Yonelinas' dual-process model (1994) is of particular interest because it has dominated behavioral and neuroscience research on recognition memory.

In the one-dimensional signal detection model, all recognition decisions are based on the familiarity or global memory strength of the test item (Donaldson, 1996). Studied items generally have higher memory strengths than the lures, though the distributions overlap as is shown in Figure 1. To make a recognition decision, the participant establishes a criterion of memory strength, calling all items with strength greater than that criterion "old" and all weaker items "new". To respond "remember" or "know", the subject establishes a higher criterion along the strength axis and responds "remember" when the strength of the test word exceeds that higher criterion and "know" otherwise. To reconcile this single dimension with the dual-process account, a variant of this SDT model was proposed by Wixted and Stretch (2004) in which the strength axis is the sum of recollection and familiarity instead of familiarity alone.

Wixted and Stretch (2004) noted that the average confidence ratings for "remember" false alarms are generally higher than for "know" false alarms and even than for "know" hits. This finding is nicely consistent with the one-dimensional model because the set of items that elicit "remember" responses are further away from the old-new decision criterion than those that elicit “know” responses and thus receive higher confidence ratings. This result is more problematic for Yonelinas’ dual-process model because Yonelinas’ version assumes that “remember” is based on an all-or-none recollection process and “remember” false alarms are dismissed as guesses, and guesses are unlikely to be of high confidence. In addition, subjects often report “remember” responses at several levels of confidence (Rotello et al., 2004; Rotello et al., 2006).

More evidence favoring the one-dimensional model over the dual-process model comes from a study done with one-stage (the subjects are asked to make remember-know-new responses for each test word) and two-stage (the subjects first make old-new response for each test word and then make remember-know judgment for each word that receives “old” response) recognition judgments. Hicks and Marsh (1999) showed that these tasks yielded different results. The sensitivity of the one-stage procedure was similar to that of the two-stage procedure, whereas the response bias for the one-stage procedure was more liberal. Yonelinas’ dual-process model cannot account for response bias in the “remember” decisions: This model assumes that “remember” responses tap on recollection, and recollection is a high-threshold process which cannot lead to memory-based false alarms (Rotello et al., 2006). In contrast, response bias can be explained easily by signal detection models: While the distance of the means of the distributions of old and new items remains the same, the response criteria shift with bias manipulation along the memory strength axis. Further, the quantitative fits of the models show that the signal detection model provides the best overall description of the data with manipulation of response bias (Rotello et al., 2006).

However, there is one challenge for the one-dimensional signal-detection model (Rotello et al., 2004). The ratio of the standard deviations of the New and Old item distributions can be estimated from the slope of *z*ROC derived from confidence ratings. Equivalently, it can be estimated from the two-point *z*ROC that results from treating the "remember" and "old" false-alarms and hit rates as two points in *z*ROC space. According to the one-dimensional model, these two slopes should be equal. However, a meta-analysis found that this prediction is violated (Rotello et al., 2004). To resolve this problem, Wixted and Stretch (2004) assumed a variable remember-know criterion in the one-dimensional model. Monte Carlo simulation showed that this variable-criterion model produced two-point *z*ROC slopes of the observed magnitude.

A different solution to the two-point *z*ROC slope issue is to extend the one-dimensional signal model to two dimensions, as in the STREAK model (Sum-Difference Theory of Remembering and Knowing, Rotello et al., 2004). In the STREAK model, the old-new and remember-know decisions are made against a decision plane instead of a decision line, as is shown in Figure 2. The old-new decision bound divides the targets from the lures: items are called "old" if the sum of global and specific information exceeds the decision bound. An R-K decision bound divides items called "old" into those for which the specific memory strength is relatively great compared with global strength ("remember") and those for which specific strength is relatively weak ("know"). Thus the weighted sum of global and specific strengths produces old-new decision and the weighted difference of strengths produces a "remember" or "know" judgment (Rotello et al., 2004).

In STREAK, the “remember/know” criterion is independent of the “old/new” criterion. This independence leads STREAK to predict that the proportion of “old” responses that receive “remember” responses will be constant over manipulations of old-new response bias. Using a between-subjects bias manipulation, Rotello et al. (2006) observed this pattern in data, which they termed *the response-ratio invariant.* However, the experiments using within-subjects manipulations show that “remember” responses are generally correlated with recognition confidence. This correlation is problematic for STREAK (Rotello et al., 2004).

In sum, both the one dimensional model and STREAK assume that “remember” and “know” are based on the same underlying processes and therefore the two judgments are only quantitatively different. In contrast, dual-process models assume that “remember” and “know” responses are based on entirely different processes and thus the two processes are qualitatively different. An insightful review of these mathematical models can be found in Macmillan and Rotello (2006).

These models generally account for the response probability data well, although each has difficulty in explaining particular empirical data. Quantitative fits show that the SDT models explain the data better than the dual-process model (Rotello et al., 2006). However, with the challenges each model faces, it is still hard to determine which model provides better description of the data and which makes the most realistic assumptions.

Since the “remember/know” judgments (or “old” judgments followed by “remember/know” judgments) are made over time, a remember-know model with true assumptions should be able to account for both response accuracy and response times, that is, the speed-accuracy tradeoff (Townsend & Ashby, 1983; Ratcliff & Rouder, 1998).

Thus in the following sections, reaction times in remember-know paradigm will be discussed.

1.4 Reaction times in the remember-know paradigm

In this thesis, the focus is on the reaction time distribution of “remember” and “know” responses. The data analysis is from a bottom-up perspective: what will reaction time distributions tell us about how “remember” and “know” judgments are made? If “remember” and “know” responses are based on the same processes, then examination of their reaction time distributions should reveal a similar pattern; otherwise, we should see reaction time distributions that differ in some way.

Reaction time is considered to be very valuable for revealing the mechanism of cognitive processes (Luce, 1986; Van Zandt, 2000). However, there is not much research on reaction times for “remember” and “know” responses except some studies that reported mean RT for each type of judgment.

Dewhurst and Conway (1994) first reported mean RTs to “old” judgments categorized by whether a subsequent “remember” or “know” response was made. The mean RT to “old” judgments followed by a “remember” response was significantly faster than that followed by a “know” response. Yonelinas (2002) explained this finding as a demand characteristic: the subjects were instructed to respond “know” only when they failed to recollect specific details, suggesting that those responses should be slower. Dewhurst et al. (2006) instructed subjects to make “old/new” and “remember/know” responses in two different test lists. Again, when “old” response RTs were categorized by the subsequent “remember/know” judgments, the mean “old” RT was shorter when followed by a “remember” judgment than a “know” judgment. Thus, Yonelinas’ demand characteristics can be ruled out. Dewhurst et al. (2006) proposed an alternative account that "remember" responses are based on a fast and automatic process that retrieves detailed contextual information while "know" responses require an evaluative process that compares the relative familiarity of the item to others on the test. This account will be examined in the data analyses presented in later chapters.

Relating the mean RT findings to the remember-know models, these results are consistent with the one-dimensional model given a strength-latency relationship: The farther away the item strength is from the decision criterion, the faster the response is (Murdock, 1985). Because the strengths of remembered items are greater than known items, they are further from the old/new criterion. Therefore "remember" responses should be faster than "know" responses. Strong evidence was provided in Wixted and Stretch's (2004) study: they found that both "remember" hits and "remember" false alarms were significantly faster than "know" hits.

The issue is not settled, however. When doing reaction time analyses, researchers generally use the mean RT as the dependent variable, which ignores distributional information that could reflect theoretically interesting processes. Moreover, the mean RT can be a misleading measure (Heathcote et al., 1991), because the same mean RT can occur for a variety of differently shaped RT distributions. Also, models that accurately predict the ordering of mean RTs may nonetheless fail to account for other aspects of the RT data, such as skew (Mewhort, Braun, & Heathcote, 1992). We will give examples in the next section.

Moment-based estimators and cumulants such as skew and kutosis have been shown to be unsuitable for characterizing the shape of empirical distributions because they suffer from lack of efficiency (i.e. very large sample size are required) and robustness (i.e., higher order moments are too sensitive to outliers, Ractliff, 1979). Since RTs can be considered to be a random variable and all the information about a random variable is contained in its probability density function (PDF or density function) or cumulative distribution function (CDF), the PDF or CDF can be used to explicate the reaction times (Luce, 1986). There are many different distribution models to characterize the shape of the RT density function and CDF. The most common model is the ex-Gaussian distribution, which is characterized by three parameters— μ , σ and τ .

1.5 The ex-Gaussian distribution and its applications

The ex-Gaussian is a positively skewed distribution produced by the convolution of a normal and an exponential distribution. The ex-Gaussian has three parameters, the mean (μ) and the standard deviation (σ) of the normal component and the mean of the exponential component (τ). The mean of an ex-Gaussian distribution is $\mu+\tau$ and its variance is $\sigma^2+\tau^2$.

Because the mean of an ex-Gaussian distribution equals $\mu+\tau$, both μ , the mode of the distribution, and τ , the right tail of the distribution, influence the mean RT. Because μ and τ may vary independently, finding equal mean RTs over conditions does not imply that the underlying RT distributions are the same: the shape of the distributions may be quite different. This is illustrated by Figure 3(a). Even when the means and variances of the reaction times are the same in two conditions, the distributions can differ as is shown in Figure 3(b). When there is a difference in the means of two conditions, the difference may be caused by a shift of the normal distribution (μ) as shown in Figure

3(c), by a difference in the tail (τ) as shown in Figure 3(d), or by combination of these differences.

Early application of the ex-Gaussian distribution to RTs assumed that the parameters of the ex-Gaussian distribution reflected different cognitive processes. For example, Hohle (1965) suggested that the normal component of the RT represents the time for organization and execution of motor responses and that the exponential component reflects the decision time. Though this claim has not been fully supported empirically, the ex-Gaussian distribution has been shown to fit reaction time well (Ratcliff & Murdock, 1976). More recently, the ex-Gaussian distribution has been used for two purposes. First, it has been used to explore the reaction time characteristics of cognitive processes and to differentiate different types of processes (Spieler, Balota, & Faust, 2000; Balota & Spieler D., 1999; Wixted & Rohrer, 1994; Heathcote, Popiel, & Mewhort, 1991; Hockley, 1984; Ratcliff & Murdock, 1976). Second, the ex-Gaussian has been used to evaluate models that account for reaction time distribution characteristics (Mewhort et al., 1992; Ratcliff & Murdock, 1976).

Ratcliff and Murdock (1976) conducted four experiments to investigate the effects of study and test positions, presentation rate, list length and presentation frequency on response accuracy and latency. Three functional distributions, gamma, lognormal and ex-Gaussian, were fit to the latency data. The ex-Gaussian distribution provided the best results. More importantly, examination of the parameters (μ , σ , and τ) showed that different empirical factors were reflected in different RT parameters. For example, study and test positions, as well as presentation frequency effects were mainly reflected by values of τ , whereas the manipulation of list length affected both μ and τ .

Hockley (1984) extended Ratcliff and Murdock's findings to a variety of cognitive tasks. Again the parameters μ and τ behaved differently across conditions. In a visual search task, mean RT increased with search set size, and that effect was mainly reflected by the μ parameter. In contrast, a memory search task revealed that mean RT increased with search size as a result of increase in the τ parameter. A change in μ as a function of visual search size is consistent with a serial search process assuming fixed comparison rate, whereas a change in τ and a constant μ as a function of memory search size rules out the serial search process that assumes a constant comparison rate.

More recently Balota and Spieler (1999) fit the ex-Gaussian distribution to word recognition RTs and found that, in a lexical decision task, the word frequency effect was primarily reflected by changes in τ; in contrast, in a word-naming task, the word frequency effect was mainly reflected by changes in μ. Balota and Spieler pointed out that a greater τ for low-frequency words than for high-frequency words in a lexical decision task indicated that an extra attention demanding check process was engaged for low-frequency words. In contrast, word-frequency effect was reflected in automatic process in word naming task, because the same type of check process in the lexical decision task is not necessary in word naming task.

Later Spieler et al. (2000) used ex-Gaussian distributions to investigate the interference effect in Stroop task and other selective attention tasks (local/global task and flanker task). The interference effect in Stroop task is reflected in both μ and τ, whereas the interference effect in local/global and flanker tasks is reflected in the μ parameter. Thus the parameters of ex-Gaussian distributions can differentiate the mechanisms that underlie Stroop task, where a stimulus attribute was selected, and local/global and flanker tasks, where a spatial location was selected.

In summary, the ex-Gaussian distribution has been used in many studies to explore various cognitive processes and to evaluate and develop models. The μ and τ parameters consistently display different patterns across conditions. As argued in previous research, the τ parameter in ex-Gaussian distributions is a more sensitive indicator of attention-demanding cognitive processes than the μ parameter. In contrast, μ appear to reflect more automatic process (Balota & Spieler, 1999; Rouder et al., 2005). Changes in μ and τ over experimental manipulations can therefore indicate changes in automatic process and strategic process respectively (Ratcliff & Murdock, 1979).

Applying this general descriptive approach to the remember-know paradigm, the distributions of “remember” and “know” RTs could be examined specifically. The parameters of best-fitting ex-Gaussian distribution to “old/new” responses could be found for responses that were followed by “remember” or “know” judgments. As argued in Section 1.3, the one-dimensional and two-dimensional SDT models assume that the same strategic processes underlies both “remember” and “know” judgments, whereas the dual-process model assumes that “remember” and “know” judgments are made based on qualitatively different processes. Therefore comparison of values of the τ parameter fitted to “remember” and “know” judgments can be used to evaluate the remember-know models. Similar τ values suggest that same strategic process underlies “remember” and “know” judgments, and therefore could support the one-dimensional and the two-dimensional SDT models. In contrast, finding different τ values would indicate that different processes underlie “remember” and “know” judgments and would therefore support the dual-process model.

In addition to the τ parameter, comparison of the μ parameter will be conducted to see whether the automatic processes that underlie “remember” and “know” judgments are same. From the dual-process account, “know” judgment taps on familiarity which is an automatic process and “remember” judgment taps on recollection which is an attention demanding process. Thus, μ may display different values for “remember” and “know” judgments. Dewhurst et al. (2006) explained faster “remember” judgments than “know” judgments by assuming that “know” judgments involved an evaluative (attention demanding) process whereas “remember” judgments are more automatic due to the available vivid contextual information. Dewhurst et al.’s account can also be examined by compare values of parameter μ and τ for “remember” and “know” judgments: “remember” judgments should have a small μ and τ relative to “know” judgments.

To explore the “remember” and “know” processes with the descriptive approach mentioned above, and also to test Yonelinas and Dewhurst et al.’s accounts of mean RTs for “remember” and “know” responses, three experiments were conducted: a list length experiment, a list strength experiment and a speed-accuracy experiment that manipulate the response bias by instructing subjects to emphasize speed or accuracy. The first two experiments were conducted because list length and list strength effects are well established and investigated, and therefore can serve as a touchstone of basic performance in our tasks. The third experiment was designed to examine how speed-accuracy tradeoff may influence the RT distributions of “remember” and “know” responses. When accuracy is emphasized, a large value of the τ parameter is expected presumably because subjects could engage complete strategic process before they make response; when speed is emphasized, a small value of the τ parameter is expected, because under time pressure, only the most available information would be retrieved to make decision, and therefore the process could be more automatic and less strategic.

1.6 General data analysis methods

The data analytic approach we adopted in each experiment consists of two steps: First, we conduct an accuracy analysis of the data which includes calculating sensitivity (i.e., the ability to discriminate studied items from novel items in the recognition experiment), and fitting the remember-know models to ROC data. Second, we consider reaction time data, which includes an analysis of both mean RTs and examination of the RT distributions using the descriptive (ex-Gaussian) approach outlined above.

The remember-know model fits to the ROC data will follow the exact procedures in Rotello et al. (2006). In that study, five different models were tested: the one-dimensional model, a variable criterion extension of the one-dimensional model (Wixted and Stretch, 2004), the dual-process model and STREAK plus an extension of the dual-process model, in which "remember" hits were assumed to occur at several levels of confidence. This change makes the extended dual-process model more like the one-dimensional model (Rotello et al., 2006).

All the models will be fit using maximum likelihood estimation (MLE). Akaike Information Criterion (AIC, Akaike, 1974; Myung et al., 2004) will be used to adjust for differing numbers of free parameters. AIC is the sum of two terms: one that reflects the log-likelihood of the model given the data and another that penalizes the model for free parameters. Smaller values of AIC indicate better performance of the model.

To fit the ex-Gaussian distribution to reaction time data, we use Heathcote et al.'s QMPE program (Quantile Maximum Probability Estimation). QMPE exploits quantile maximum probability estimation which minimizes the effects of outliers and measurement noise (Cousineau et al., 2004). A Monte Carlo study shows that estimation performance is fairly good even when the sample size is as small as 40 (Heathcote et al., 2004). The method's performance is not known for samples size less than 40, so (with one exception) we only fit ex-Gaussian distribution to data sets with at least 40 observations.

The output of QMPE includes the estimated values for μ, σ, and τ as well as their estimated standard errors. The output also includes the observed quantiles, the estimated quantiles, and the contribution of each quantile to the log-likelihood sum. Most interesting to us is the estimated parameters of μ, σ, and τ.

We fit ex-Gaussian distributions to hits that were followed by "remember" and "know" judgments in two ways[1]. For each subject, we fit RTs regardless of confidence level. Because reaction times are correlated with decision confidence (Petrusic & Baranski, 2003) and "remember" responses often receive higher confidence ratings than "know" responses, controlling confidence levels is important. We also fit the RTs for each subject at a particular confidence level that was selected to maximize sample size.

After fitting the ex-Gaussian to the RTs, the most important thing is to compare the parameter values of μ, σ, and τ for hits followed by "remember" and "know"

[1] In this thesis, only "remember" and "know" hits were fit by ex-Gaussian distributions. The initial plan is to fit ex-Gaussian distributions to "old" responses followed by "remember" and "know" judgments. However, there were few "remember" false alarms. In addition, hits are cleaner than false alarms and so we chose to only fit hits.

responses. To decide whether the values are the same or different for "remember" and "know" responses, a bootstrap procedure was used to construct 95% confidence intervals of the difference between the parameter values for "remember" and "know" responses. The RTs in each cell (i.e., the RTs for hits followed by "remember" and "know" judgments in each condition for each subject) were resampled 2000 times, and then each sample set was fit with an ex-Gaussian distribution using QMPE. These 2000 resample data sets yielded 2000 parameter estimates of μ, σ, and τ for hits followed by "remember" responses and 2000 for hits followed by "know" responses. We subtracted the parameter values for "know" response from those for "remember" responses (*r*-*k*). The 50th smallest and 50th largest values of the difference of the three parameters form the boundaries of the 95% confidence interval for the difference in μ, σ, and τ for hits followed by "remember" or "know" judgments.

CHAPTER 2

EXPERIMENT 1: LIST LENGTH

A list length manipulation is chosen for Experiment 1 because effects associated with list length have been replicated in many experiments (Ratcliff & Murdock, 1976; Yonelinas, 1994; Cary & Reder, 2003), and therefore list length effect can be used to examine the basic performance of the subjects.

In memory research, long study-test lists result in low performance compared with shorter lists. More specifically, longer lists produce lower hit rate and higher false alarm rate than shorter lists; this is called a *mirror effect*. In the remember-know paradigm, longer lists elicit fewer “remember” responses than shorter lists (Cary & Reder, 2003). Therefore, in this experiment, lower sensitivity and fewer “remember” responses are expected for long study-test lists than for short lists. Enough data were collected for individual-level distributional RT analysis.

2.1 Method

1. Participants

Four University of Massachusetts students participated in exchange for $40 cash payment.

2. Material and design

A pool of 3072 nouns was created. The words varied widely in written frequency (10-4500, Kucera-Francis) and length (3-15 letters). They were randomly assigned to roles for each subject: 1) Study words on short (30 words each) or long (90 words each) list. There were 720 words assigned to each length, yielding 24 Short and 8 Long lists. 2) An equal number of words (30 or 90) served as lures on the recognition test followed each study list. 3) Finally, 3 untested words appeared at the beginning and end of each study list to absorb primacy and recency effects. These primacy and recency words were not tested.

Subjects participated individually in 8 sessions. Each session consisted of three Short study-test blocks and 1 Long study-test block, presented in random order. Following either type of study list there was a recognition test on all of the studied words from that block plus an equal number of New words, in a random order. Responses were made on a Cedrus response box RB 610 with six buttons. With the response box, RTs are accurate to within 1ms.

3. Procedure

Subjects were told to study the words in each list for a recognition test. Each study word was presented for 1750ms with an inter-stimulus interval of 250ms. In each block, after the subjects studied the word list, they immediately proceeded to the test. A practice block composed of 10 study words and 20 test words was presented before the formal blocks in the first session (only) so that subjects could become familiarized with the response procedure.

During the test phase, subjects were instructed to put their ring finger, middle finger and index finger of each hand on the response pad, placing their left ring finger on leftmost button (button 1) and their right ring finger on the rightmost button (button 6). For each test word, the subject made three responses. First, they made an Old/New decision using their middle fingers pressing button 2 ("new") or 5 ("old"). Second, they rated their old-new confidence (1="sure new", 6="sure old"); they were told that they could reverse their judgment (from "new" to "old" or the opposite) by using the appropriate confidence rating. Finally, for the items subjects gave confidence ratings 4 ("maybe old"), 5 ("probably old") or 6 ("sure old"), they made remember-know judgments by pressing button 6 to indicate "remember" and button 1 to indicate "know". For the items that they gave confidence ratings 1 ("sure new"), 2 ("probably new") or 3 ("maybe new"), they were instructed to press button 1 for the third response.

Subjects were instructed to make their responses as quickly as possible without sacrificing accuracy. Reaction times to all three responses were recorded. However, only the reaction times to the "old/new" decision are evaluated. Those responses are cleanest (i.e., no carry-on effects from prior judgments) and analogous RTs have been used in previous studies (Dewhurst & Conway, 1994; Dewhurst et al., 2006).

2.2 Results

As described earlier, the data analysis results will be presented in two sections: first accuracy (or response probability) based analyses and then reaction time analyses. The accuracy analysis results include sensitivity estimates for each condition, the response proportions of "remember", "know" and "new" judgments, and comparison of the remember-know model fits to ROC data. The results of reaction time data analysis include the mean RT analysis and the fits of ex-Gaussian distribution to the data. Further discussion will be based on the values of the parameters of best-fit ex-Gaussian distribution.

Subjects reversed their old/new responses on 2.3% of trials by giving confidence ratings that contradicted their initial response. These reversed responses were included in the accuracy analyses but those trials were excluded from the reaction time analyses. The confidence ratings determine the final judgment. For example, if "new" was the correct response and the subject initially responded “old” (an error) but specified either “1”, “2”, or “3” (the three new ratings) as the confidence level, we counted the judgment to be correct.

2.2.1 Accuracy analysis

The average hit rate for the four subjects is 0.80 for short lists and 0.70 for Long lists; the average false alarm rate is 0.13 for Short lists and 0.20 for long lists. A one-way analysis of variance for hits showed significant effect of list length, $F(1, 3) = 13.294$, $p < .05$. For false alarms, marginal significance was found for list length, $F(1, 3) = 8.347$, $p = .063$. This generally replicates the list length mirror effect observed in previous studies (Ratcliff & Murdock, 1976; Yonelinas, 1994; Cary & Reder, 2003) which leads to higher sensitivity for short lists than for long lists.

The confidence ratings were used to create ROCs and *z*ROCs (see Macmillan & Creelman, 2005 for technical details). The slopes of *z*ROCs are less than 1.0 (see Table 1), indicating that the Old item distribution is more variable than the New item distribution. When the slope is not 1.0, d' is no longer an accurate measurement of sensitivity: it confounds sensitivity with response bias (Macmillan & Creelman, 2005). Therefore, a more accurate index, d_a, is adopted for sensitivity measurement which takes *z*ROC slope into account (Macmillan & Creelman, 2005). d' measures the distance of the means of distributions of old and new items normalized by the standard deviation of either old or new items (because they are assumed to be equal); d_a also measures the distance of the means of distributions of old and new items, however the distance is normalized by the root mean square of the standard deviations of old and new items.

Thus d' is a special case of d_a: when the standard deviations of the old and new items are the same, d' is equal to d_a.

In this experiment, the overall d_a is 1.75 for Short condition and 1.18 for Long condition. The d_a values for individual subjects are shown in Table 1. A paired t-test revealed a significant difference in sensitivities for the Long and Short lists ($t(3) = 5.66$, $p < .05$).

1. "Remember" and "know" response proportions

Table 2 shows the overall proportions of "old" and "new" responses that were followed by "remember" and "know" judgments. Long lists elicited significantly fewer "remember" hits than Short lists ($t(3) = 5.218$, $p < .05$), whereas "remember" false alarms are not different for Long and Short lists ($t(3) = .195$, n.s.), replicating Cary and Reder (2003). As for "know" responses, both "know" hits and false alarms are similar across Long and Short condition (For hits, $t(3) = 2.191$, n.s. and for false alarms, $t(3) = 2.095$, n.s), which is different from what Cary and Reder observed: they found that there were more "know" hits and false alarms for longer lists than for shorter lists.

As discussed by Hirshman and Master (1997), difference combinations of establishment of criteria ("old"/"new" criterion and "remember"/"know" criterion) by subjects could result in three situations under experimental manipulations: "remember" hit rate is larger for Condition Two than for Condition One whereas "know" hit rate is larger for Condition One than Condition Two, as observed in Cary and Reder and illustrated in Panel 2 of Figure 2 in Hirshman and Master (1997, note that "know" false alarm rate is larger for Condition One than for Condition Two, consistent with what Cary and Reder found); "remember" hit rate is larger for Condition Two than for Condition

One whereas “know” hit rates are similar across the two conditions, as observed in our Experiment 1 and illustrated in Panel 1 of Figure 2 in Hirshman and Master (1997, note that “know” false alarm rates are similar across the two conditions, consistent with what is found in Experiment 1); “know” hit rate is larger for Condition One than for Condition Two whereas “remember” hit rates are similar, as illustrated in Panel 3 of Figure 2 in Hirshman and Master (1997).

2. Remember-know model fits

The “old/new” responses at each confidence level that received “remember” and “know” judgments were fit by five models described in Section 1.6.

The one-dimensional model was implemented with 8 parameters: one sensitivity estimate (*d*), 6 decision criteria (C_1-C_5 for old-new confidence ratings and C_r for remember-know judgment), and the slope of *z*ROC (*s*). One additional parameter was required to fit the one-dimensional model with variable criterion: the standard deviation of the remember-know criterion distribution (1/*t*).

The dual-process model was implemented with 8 parameters. There are two sensitivity parameters: R_0 and d'. R_0 reflects the fraction of trials in which true recollection occurs whereas d' is the sensitivity measure for the familiarity process. There are 5 decision criteria (C_1-C_5) to divide the old-new responses into confidence ratings. Finally, the fraction of trials on which false recollection operated is ε. Therefore, familiarity is effective on a fraction, $1-R_0$, of the target trials and on $1-\varepsilon$ of the lure trials. In the extended dual-process model, two more parameters were added to allow “remember” responses to distribute over all “old” ratings.

STREAK also has 9 parameters: 6 criteria (C_1-C_5 and C_r), 2 sensitivity parameters (d_x and d_y) and the slope of the *z*ROC (*s*). C_1-C_5 represent the five criteria that divide responses into confidence ratings and these five decision bounds are parallel to the old/new decision bound and perpendicular to the remember/know decision bound. C_r is the remember/know criterion that divides the "old" responses into "remember" and "know" judgments. d_x measures the memory sensitivity for global information and d_y measures sensitivity for specific information.

Table 3 shows the results of AIC statistics for the five models[2]. Small numbers indicate better fits. The one-dimensional model in one of its two forms fits 6 out of the 8 conditions. In two conditions (B Short and D Short), the dual-process model fits best. Overall, the one-dimensional models provide the best fit. However, it should be noted that some differences are quite small.

To compare the performance of the one-dimensional models and the dual-process model, Figure 4 illustrates the best and worst model fits[3] for four representative cases: Long and Short conditions for Subject A (best fit with the variable-criterion one-dimensional model), Short condition for Subject B and D (best fit with the dual-process model). For the first two cases, the one-dimensional model with variable remember-know criterion provided the best fits whereas the dual-process model provided the worst fits; for the latter two cases, the dual-process model provided the best fits whereas one-

[2] We also calculated BIC and Akaike weights for the fitting results. The results agree with AIC statistics.

[3] The illustration of the best and worst model fit is confined within the one-dimensional models and the dual-process models, since STREAK provided poorest fits to the ROC data in the three experiments conducted. This is due to that it does not predict the correlation between the confidence ratings and "remember/know" responses observed in the data (see Rotello et al., 2004; Rotello et al., 2006).

dimensional models (variable remember-know criterion version for B and fixed remember-know criterion version for D) provided the worst fits.

Overall, all models provided decent fit to the data as shown in Figure 4, and the difference in the goodness-of-fit for the one-dimensional model and dual-process model mainly present in the “remember” and “know” response proportions at the two highest confidence rating. Table 4 displays the observed response proportions for “remember” judgments at each “old” rating and the corresponding proportions predicted by the best-fit and worst-fit models. For the Long and Short condition of Subject A, the problem faced by the dual-process model is that it assumes “remember” responses are made with highest confidence, but subjects spread the responses across the ratings for “old” items and therefore, the one dimensional SDT model fits these data better. In the Short condition for Subjects B and D, more than 98% “remember” judgments are made with confidence level 6; therefore the dual-process model fits better.

2.2.2 Reaction time data analysis

1. Mean RT analysis

We began by using the RTs to “Old/New” judgments as the primary dependent measure (See also Dewhurst & Conway, 1994; Dewhurst et al., 2006).

The mean RTs for hits and false alarms are presented in Table 5. Since there are only four subjects in this study, we do not have enough power to detect the difference in group mean RTs for “remember” and “know” responses. However, the study was designed to test effects at the level of the individual.

For all four subjects, independent two-sample t-tests revealed that the mean RTs to “old” response followed by “remember” judgments are significantly faster than those followed by “know” judgments (minimum $t_{\text{observed}} = 7.223$, $p = .000$). This result replicates Dewhurst and Conway (1994; see also Dewhurst et al., 2005).

The difference between the mean RTs to hits followed by “remember” and “know” judgments is also significant (minimum $t_{\text{observed}} = 6.321$, $p = 0.000$), whereas there is no difference between the mean RTs to false alarms (maximum $t_{\text{observed}} = 1.143$, $p > .2$). When Long and Short conditions are considered separately, the effects of mean RTs to hits are still significant (minimum $t_{\text{observed}} = 3.549$, $p = .000$). However, “remember” judgments tend to be given at higher confidence ratings, and “know” judgments tend to be given at lower confidence levels. Reaction times increase as confidence rating decreases (Petrusic & Baranski, 2003). In Experiment 1, the average correlation between “old” confidence ratings and mean RTs for hits is -.90 (The more confident subjects were, the less time it took for them to make responses) for Long condition and -.95 for Short condition. Therefore, it is indispensable to control confidence level when doing RT data analysis.

When controlled for confidence levels, the difference between mean RTs is greatly reduced. After conducting all possible t-tests for hits in long and short conditions for each subject[4], the difference in mean RTs for hits followed by “remember” and “know” responses is only significant for one condition: hits for Subject C in the Long condition at confidence rating 5 ($t(26.8) = 3.678$, $p = .001$). However, the mean RT for hits followed by “remember” is larger than that followed by “know”, which is opposite to what was observed regardless of confidence level.

[4] There should be 24 t-tests (4 subjects × 2 list lengths × 3 “old” confidence ratings) in total; however, due to the insufficient number (<10) in most cells, only 6 t-tests were conducted

When the confidence ratings were collapsed, the mean RTs for hits followed by "remember" are significantly faster than those followed by "know" responses; when the confidence levels are controlled, the differences in the mean RTs for hits followed by "remember" and "know" disappeared (in one case, the mean RTs for hits followed by "remember" is significantly slower than that followed by "know" responses). Therefore, we can conclude that effect of mean RTs for hits followed by "remember" and "know" responses are mainly due to the difference in response confidence for "remember" and "know" judgments. When confidence level is controlled, the difference is greatly reduced.

2. Distributional RT analysis

The parameter values of the ex-Gaussian fits as well as mean and standard deviations for each subject are shown in Table 6 when confidence level is ignored and when it is controlled. When confidence level is controlled, only Subject A and D provided enough hits (>40, with one exception) followed by "remember" and "know" judgments.

As has been presented, the mean RTs are significantly different for hits followed by "remember" and "know" judgments for each subject in both Long and Short conditions when confidence ratings are collapsed. It is an intriguing question whether the differences in mean RTs presented in the μ parameter or the τ parameter or both.

Using the bootstrapping procedure described in Section 1.6, the 95% confidence intervals of the differences of the best-fit ex-Gaussian parameter values of hits followed by "remember" and "know" judgments were calculated and presented in Table 7. Also presented in Table 7 are 95% confidence intervals of the differences for Subject A at confidence rating 5 and for Subject D at confidence 6. If the 95% confidence intervals include 0, the differences in the parameter values are not significant; otherwise if the 95% confidence intervals do not include 0, the differences in the parameter values are significant.

From Table 7, significant differences in the μ parameter, the τ parameter or the combinations of these two parameters result in the significant differences in mean RTs for hits followed by “remember” and “know” judgments. For Subject B in the Short condition and Subject C in the Long condition, only values of the μ parameter are significantly different; for Subject A in the Long condition and Subject D in the Short condition, only values of the τ parameter are significantly different; for the rest four cases, Subject A in the Short condition, Subject B in the Long condition, Subject C in the Short condition and Subject D in the Long condition, both the differences in the μ and τ parameters are significant.

However, when the confidence level is controlled, no difference existed in the μ, σ and τ parameters. This result accords well with the mean RT data analysis results, in which no difference in mean RTs was found between hits followed by a “remember” and those followed by a “know” response for Subject A and D when confidence level is controlled.

It should be noted that the 95% confidence intervals in Table 7 are quite wide, which is due to the large variances of the reaction times. When the bootstrapping procedure is implemented, the samples could vary dramatically and therefore the best-fit ex-Gaussian parameters are quite different. The wide 95% confidence intervals made the results not compelling enough for a strong conclusion. However, the confidence intervals are smaller for larger and more balanced "remember" and "know" sample sizes.

Figure 5 illustrates the RT distributions for hits that were given "remember" and "know" judgments in the Long condition for Subject A at Rating 5 and Subject D at Rating 6. The best-fit ex-Gaussian distribution was superimposed on each empirical distribution. Figure 6 shows the analogous data for the short condition. The figures showed that the fits of the ex-Gaussian distribution to the reaction times are generally good.

As is argued in Section 1.5, τ is an indicator of strategic processing. When the values of τ are similar across two conditions, the two conditions involve similar strategic processes, i.e., similar retrieval and decision processes. In addition, when the confidence level is controlled, the values of the μ parameter, which indicate the automatic processes involved in decision, do not differ. Therefore, we conclude here that "remember" and "know" do not differ in their underlying processes when the confidence level is controlled and this supports the one-dimensional and two-dimensional signal detection models.

2.3 Discussion

The application of remember-know models to the accuracy data shows that the one-dimensional SDT model (with either fixed or variable remember-know criterion) explains the most data. The dual-process model fit the data better than one-dimensional model for the short condition of Subject B and D. In these two cases, the "remember" judgments concentrated at confidence 6 and this is consistent with the assumption of the dual-process model that "remember" responses are of highest confidence (Yonelinas, 1994). However, for the other six cases, the one dimensional SDT model provided best fits. The implication is that a continuous process underlies “remember” judgments in this experiment, consistent with findings in previous research (Rotello et al., 2004; Rotello et al., 2005; Rotello et al., 2006).

Dewhurst et al. (2006)’s mean RT results are replicated here: mean RTs to “old” responses followed by “remember” judgments are faster than those followed by “know” judgments. In addition, the mean RT for hits followed by “remember” is significantly faster than those followed by “know” judgments. However, this is mainly due to the different confidence levels: “remember” judgments tend to concentrate at confidence level 5 or 6 and “know” judgments at confidence level 4 or 5. When the confidence level is controlled, there is no difference in mean RTs to hits followed by “remember” and “know” judgments. In one case, the mean RT is faster to hits followed by “know” judgments than by “remember” judgments.

The ex-Gaussian distribution fits to RT data shows that when confidence level is controlled, the skew parameter τ, which is the indicator of the strategic processes, shows no difference in best-fit values for hits followed by “remember” judgments and “know” judgments. This is consistent with the signal detection interpretations of remember-know judgments (Donaldson, 1996; Hirshman & Master, 1997; Wixted & Stretch, 2004; Rotello et al., 2004) and inconsistent with the dual-process model (Murdock, 2006; Yonelinas, 2001; Reder et al., 2000).

In Dewhurst et al. (2006), the authors argued that the difference in mean “old” RTs for “remember” and “know” judgments is because the “remember” judgment is a rapid automatic process and a “know” judgment is a conscious, controlled process that involves an evaluative process. The type of information available at test determines the speed of the "old" responses. This predicts larger τ values for hits followed by "remember" judgments than those followed by "know" judgments. In this experiment, when confidence level is collapsed, this prediction holds. However, when the confidence level is controlled, the difference in mean RTs for "remember" and "know" judgments disappears. Neither μ nor τ differ for hits followed by "remember" and "know" judgments. Therefore, we argue that it is the confidence level that caused the difference in mean RTs to hits followed by "remember" and "know" judgments.

In addition, Dewhurst et al. (2006) assumes that when subjects make "remember" or "know" judgments, either contextual information or familiarity is used to make the decision. This is consistent with the dual-process model (Yonelinas, 1994): a "remember" judgment taps recollection and a "know" judgment taps familiarity. In contrast, the variable-criterion one-dimensional model (Wixted and Stretch, 2004; Wixted, 2007) and STREAK (Rotello et al., 2004) both assume that recollection (or specific information) and familiarity (or global information) are used together to make "remember" and "know" decisions. In this experiment, the one-dimensional models provided best fits to the data, and the one-dimensional model assumes that both recollection and familiarity are used to make "remember" and "know" decisions. Therefore, the conclusion here is that subjects did not use either contextual information or familiarity to make "remember" and "know" judgments, both information are used to make the decision.

For Subjects B and D, the remember-know model fits are similar: one-dimensional models provide best fits for Long condition (standard for Subject B and extended version with variable remember-know criterion for Subject D). This may be due to the high performance in the Short condition for Subject B and D (d_a is 2.30 for Subjects B and 2.55 for Subject D). When study list only consists of 30 words, the memory strengths could be so strong that "remember" responses are mainly made with highest confidence and display a high-threshold-like pattern in ROC data.

For Subject D in Short condition, the remember-know model fits and ex-Gaussian fit are in conflict: dual-process model fits the ROC data best while none of the values of the μ , σ or τ parameters differ for "remember" and "know" responses. The conflicting implications may reflect the limitation in AIC statistics. As pointed out in Myung et al. (2004), model fit results consist of two parts: generalizability and overfitting. It may be that dual-process model overfit the noise and the one-dimensional models have better generalizability. Another possibility is that the two models mimic each other and thus both can describe the data well (See Figure 4, also see Cohen, Rotello & Macmillan, in preparation).

To summarize, the results of the list length experiment support the one-dimensional SDT models better from both response accuracy and response times. .

In this study, reaction time data are used together with accuracy data to evaluate remember-know models. Again, reaction time data prove to be useful information to reveal underlying cognitive processes. The ex-Gaussian distribution fit the data well for most of the conditions and the skew of the ex-Gaussian distribution differentiate strategic processes.

CHAPTER 3

EXPERIMENT 2: LIST STRENGTH

In Experiment 1, list length was manipulated within subjects. Two subjects contributed enough data to evaluate their RT distributions for “remember” and “know” judgments while controlling confidence. In Experiment 2, word frequency and presentation frequency were manipulated within subjects. Word frequency effect and presentation frequency effect have been well established and can be used to test subjects’ performance. For the subjects’ accuracy performance, prior results lead us to expect a word-frequency mirror effect (more hits and fewer false alarms for low frequency items than for high frequency items) and a list strength effect (higher sensitivity for more frequently presented items, Glanzer & Adams, 1985; Stretch & Wixted, 1998; Cary & Reder, 2003).

Two levels of word frequency (high and low) and three levels of presentation frequency (study items presented one time, three times or five times) were employed, and therefore there were 6 within-subject conditions. It was expected that each subject would contribute more overlapping RT data for “remember” and “know” responses binned by the confidence ratings as the number of conditions increases. However, due to the limited amount of words in the word pool, there may not be sufficient words in each condition for distributional RT data analysis.

3.1 Method

1. Participants

Four University of Massachusetts students participated in exchange for an $80 cash payment.

2．Stimuli and design

A pool of 2188 nouns is created: 1094 nouns are of Kucera-Francis written frequency 41-10600 per million (high-frequency words) and the other 1094 nouns are of Kucera-Francis written frequency 3-5 per million (low-frequency words). Nouns are 3-15 characters in length.

Subjects participated individually in 8 sessions. The first two sessions are practice sessions to familiarize the subjects with the response routines. Low and high-frequency adjectives and verbs were used in the practice sessions. None of the words presented in the practice sessions were repeated in the formal sessions.

Each session consisted of four blocks. In each block, the study list was composed of 21 low frequency words and 21 high frequency words. Seven of the 21 low frequency words were presented once, seven were presented three times and seven were presented five times; so were the 21 high frequency words: 7 presented once, 7 presented three times and 7 presented five times. In total there were 132 items in each study list including 3 words at the beginning and 3 words at the end to absorb primacy and recency effects. The words used to absorb primacy and recency effects were not tested. Previous study shows that spaced presentations have larger effect on response accuracy and result in larger “remember” rates compared with massed presentation (Parkin & Russo, 1993). Therefore in this experiment, the study lists were generated in a way to make sure that there are at least 7 words between the presentations of the same word.

In each test list, there were 42 low frequency words (half old and half new) and 42 high frequency words (half old and half new. In total there were 84 test items in each test list.

3． Procedure

The procedure was the same as in Experiment 1.

3.2 Results

The data analysis of will be presented in the same way as for Experiment 1: first the accuracy analysis results will be presented including sensitivity, response proportions of “remember” and “know” judgments, and remember-know model fits; then the reaction time data analysis will be presented, including the mean RT data analysis and the fits of the ex-Gaussian distribution to the RTs.

Subjects reversed their old/new responses on 1.5% of trials by giving confidence ratings that contradicted their initial response. As in Experiment 1, these reversed responses were included in the accuracy analyses but were excluded from the reaction time analyses. The confidence ratings determined the final judgment.

3.2.1 Accuracy analysis

A repeated-measures ANOVA for hits rates with word frequency and presentation frequency as within-subjects factors showed significant main effect for presentation frequency, $F(2, 6) = 30.390$, $p < .01$. No word frequency effect was found ($F(1, 3) = 2.318$, n. s.) and the interaction of these two factors was not significant ($F(2, 6) = .688$, n. s.). The null word frequency effect on hits rates could be due to a ceiling effect: the subjects had good performance when the words were presented five times, and this resulted in small difference in hit rates across word frequencies (average .80 for low-frequency words and .85 for high-frequency words).

A one-way ANOVA done for false alarm rates with word frequency as within-subjects factor showed no effect ($F(1, 3) = 3.612$, n. s.). Analogous to the null effect for hit rates, the null effect for false alarm rates could be due to a floor effect (average .02 for high-frequency words and .01 for low-frequency words).

As in Experiment 1, the *z*ROC slope in each condition is smaller than 1 (see Table 8), and therefore d_a is calculated to measure sensitivity and the d_a values are shown in Table 8 for each subject and each condition. Subject A had exceptionally high performance (average hit rate was .89 and average false alarm rate was .01) which made some analyses difficult.

A two-way repeated analysis of variance with word frequency and presentation frequency as within-subject factors revealed significant main effects of both word frequency ($F(1, 3) = 18.426$, $p < .05$), and presentation frequency ($F(2, 6) = 26.877$, $p < .01$). Low-frequency words resulted in higher d_a (2.63) than high-frequency words (2.06), and higher presentation frequency resulted in higher d_a (d_a is 2.97 for words presented 5 times, 2.32 for words presented 3 times, and 1.74 for words presented once). The interaction between these two factors was not significant ($F(2, 6) = .356$, n. s.).

1. "Remember" and "know" response proportions

The average "remember" and "know" response proportions for the four subjects are presented in Table 9.

A two-way repeated-measures ANOVA was conducted for "remember" hits with word frequency and presentation frequency as within-subject factors. The proportion of "remember" hits is significantly increased with the increase of presentation frequency ($F(2, 6) = 13.791$, $p < .01$). Word frequency did not affect "remember" hits ($F(1, 3) = 5.469$, n. s.) and there is no interaction of word frequency and presentation frequency ($F(1, 3) = .533$, n. s.). A one-way ANOVA was conducted for "remember" false alarms with word frequency as within-subject factor. Word frequency had no effect on "remember" false alarm rate ($F(1, 3) = .744$, n. s.).

The same analyses were repeated for "know" hits and false alarms. Neither word frequency nor presentation frequency influenced "know" hit rates. For word frequency, $F(1, 3) = 1.023$, n. s.; for presentation frequency, $F(2, 6) = 0.219$, n. s.. In addition, no effect of word frequency was found for "know" false alarm rates ($F(1, 3) = 2.439$, n. s.).

2. Remember-know model fits

The same five models that were depicted in Section 1.3 were fit to the response accuracy data: one-dimensional model, one-dimensional model with variable remember/know criterion, dual-process model, extended dual-process model and STREAK.

The models failed to fit Subject A's data well, presumably because there were many empty cells in the ROC data: Subject A's responses were concentrated in the two extreme ratings, 1 and 6. Therefore the remember-know model fits were only conducted for Subject B, C, and D.

Subjects in Experiment 2 were quite consistent according to the remember-know model fits. From the AIC values shown in Table 10, Subject B's data were best fit with the dual-process model (in one case, extended dual-process model); Subject C and Subject D were best fit with one-dimensional SDT model with variable criterion (except in one case for Subject D, STREAK provided best fit).

To summarize, the one-dimensional model with variable remember-know criterion provided the best account of the data for two subjects and the dual-process model fit one subject's data best.

3.2.2 Reaction time analysis

1. Mean RT analysis

The mean RT for hits and false alarms followed by "remember" and "know" judgments are shown in Table 11.

As Experiment 1, this experiment is designed for individual RT data analysis. Therefore, the comparisons of mean RTs for "remember" and "know" judgments were conducted within subjects with independent two-sample t-tests in each condition.

For Subject A and C, the hits were mostly followed by "remember" judgments. Due to the imbalanced "remember" and "know" response proportions in each condition for Subject A and C, only Subject B and D provided enough data (>10) for t-tests on RTs for hits followed by "remember" and "know" judgments (In one condition for Subject D, there was not enough data).

For Subject B, hits followed by "remember" judgments were made faster than those followed by "know" in every condition (minimum $t_{\text{observed}} = 5.544$, $p = .000$); for Subject D, the mean RTs were shorter for hits followed by "remember" than "know" in two conditions: high-frequency words presented 1 time ($t(20.7) = 3.084$, $p < .01$) and low-frequency words presented 3 times ($t(85.2) = 2.761$, $p < .01$[5]). No differences were found for other three conditions for Subject D: high-frequency words presented 5 times, low-frequency words presented once and 5 times.

For Subject B and D, hits followed by "remember" judgments were made with confidence rating 6, whereas those followed by "know" judgments were made with confidence rating 4 and 5. In addition, as mentioned before, hits were mostly followed by

[5] Since 5 tests were done for Subject D, the criterion p is set as .01 to control the family-wise error

"remember" judgments for Subject A and C. Therefore, there was not enough data to do the mean RT analyses when the confidence level was controlled.

2. Distributional RT analyses

As argued in Experiment 1, due to the correlation between reaction times and confidence level (.91 in Experiment 2), the distributional RT analyses are supposed to be done when confidence level is controlled. However, there were not enough data for distributional RT analyses when confidence level is controlled. Therefore, distributional RT analyses were conducted on reaction times collapsed across confidence ratings.

Subject B provided enough data (>40) for ex-Gaussian distribution fit in four conditions: high-frequency words presented once, three times, five times and low-frequency words presented once. Table 12 showed that faster hits followed by "remember" judgments than those followed by "know" judgments for Subject B were due to the differences in the μ parameter. Figure 7 illustrates the RT distributions for hits that were given "remember" and "know" judgments for high and low-frequency words presented once for Subject B. The best-fit ex-Gaussian distribution was superimposed on each empirical distribution. Comparing the distributions of hits followed by "remember" and "know" judgments, the skews (reflected by τ) are similar; however, the distribution were shifted to right (reflected by μ) for hits followed by "know" judgments relative to that followed by "remember" judgments.

Subject D provided enough data (>40, with one exception) in two conditions: low-frequency words presented three times and five times. For Subject D, faster hits followed by "remember" than those followed by "know" judgments for low-frequency words presented three times are due to the difference in τ . Note that for the low-frequency words presented five times for Subject D, there was no difference in mean RTs for hits followed by "remember" and "know" judgments; however, the differences in μ and τ parameters are in opposite directions and therefore the mean RTs (which equals the sum of μ and τ) were the same across "remember" and "know" judgments. Figure 8 shows the RT distributions for hits that were given "remember" and "know" judgments for low-frequency words presented three and five times for Subject D. Comparing the distributions of hits followed by "remember" and "know" judgments for low-frequency words presented three times, the skew (reflected by τ) is large for hits followed by "know" judgments relative to that followed by "remember" judgments, whereas the mode positions are similar. For low-frequency words presented five times, the distribution were shifted to left (reflected by μ) for hits followed by "know" judgments relative to that followed by "remember" judgments, whereas the skew (reflected by τ) is large for hits followed by "know" judgments relative to that followed by "remember" judgments. Therefore, the mean RTs are similar.

3.3 Discussion

The remember-know model fits to ROC data indicated individual differences. The one-dimensional model with variable remember-know criterion provided best fits for Subjects C and D (except for one case, STREAK fit best) and the dual-process model provides best account for Subject B. These results suggest that different subjects could adopt different strategies to encode and retrieve information and that these differences were reflected in the model fits. For Subject C and D, the one-dimensional model with variable remember-know criterion provided significantly better fits than the standard one-dimensional model (The average AIC value is 1472 for the one-dimensional model with variable remember-know criterion and 1490 for the standard one-dimensional model, $t(11) = 10.751$, $p = .000$), suggesting that subjects could change the remember-know criterion from time to time (Wixted & Stretch, 2004).

The analysis of the mean RTs generally replicated what was found by Dewhurst et al. (2006): hits followed by "remember" judgments are made faster than "know" judgments. However, in three conditions for Subject D, no differences were found for mean RTs for hits followed by "remember" and "know" judgments.

The ex-Gaussian distribution fit could only be conducted for Subjects B and D for the conditions that provided enough data when confidence ratings were collapsed. The results showed that for Subject B, the differences in the mean RTs for hits followed by "remember" and "know" judgments were due to the μ parameter. For Subject D, in one condition (low-frequency words presented three times), the difference in the mean RTs for hits followed by "remember" and "know" judgments was due to the τ parameter; in another condition (low-frequency words presented five times), there was no difference in the mean RTs for hits followed by "remember" and "know" judgments. Examination of the μ and τ parameters showed that these two parameters were different in opposite directions.

Experiment 2 did not provide useful data for distributional RT analysis when confidence level is controlled. In this study, there were 2016 test words and 12 conditions (3 presentation frequencies $\times$ 2 word frequencies $\times$ 2 old/new conditions) and therefore in each condition there were 84 words. If the hit rate is 1 and the subjects uniformly distributed confidence ratings, there were only 28 words in each rating. In addition, "remember" and "know" judgments were imbalanced at confidence levels: The confidence ratings followed by “remember” judgments were mostly 6 and the confidence ratings followed by “know” judgments were mostly 4 or 5.

Therefore, in next experiment, the main goal is to increase the word amount in each condition so that there will be enough data for distribution RT data analysis when confidence level is controlled. In addition, appropriate difficulty should be obtained for the task to elicit balanced “remember” and “know” judgments when binned in each confidence level.

CHAPTER 4

EXPERIMENT 3: ACCURACY-SPEED BIAS

In Experiment 2, due to insufficient amount of words in each condition and high performance of the subjects, there were not enough data for distributional RT analyses on hits followed by "remember" and "know" judgments when confidence level was controlled.

In this experiment, a pilot study was done to decide the lengths of study lists and test lists, so that there were enough words in each condition and the task would not be too easy. The results showed that 56 words in study lists and 112 words in test lists may result in appropriate difficulty and elicit more balanced "remember" and "know" responses. Therefore, the study list length and test list length were set as 56 and 112 words respectively.

It is an intriguing question how Accuracy-Speed response bias will influence the μ and τ parameter values of best-fit ex-Gaussian distribution for hits followed by "remember" and "know" responses. Therefore, an Accuracy-Speed response bias was manipulated within subjects by asking the subjects to make responses "as fast as possible" or "as accurately as possible". Another within-subjects manipulation is word frequency and therefore word frequency effect on sensitivity is expected. In addition, as mentioned in Section 1.5, large skews in RT distributions are expected in the Accuracy condition relative to the Speed condition because compared with the Speed condition, strategic process should be thoroughly exerted in the Accuracy condition and therefore producing large skews due to the prolonged reaction times.

4.1 Method

1．Participants

Four University of Massachusetts students participated in exchange for an $80 cash payment.

2．Material and design

The stimuli were 3304 words including nouns, verbs and adjectives where 1652 words were of Kucera-Francis written frequency 40-10600 per million (high-frequency words) and the other 1652 words were of Kucera-Francis written frequency 1 per million (low-frequency words). Words were 3 to 15 letters in length.

High-frequency words and low-frequency words were randomly assigned to roles: 1) Study lists. 28 high frequency and 28 low frequency words forms one study list; in total 784 high frequency words and 784 lower frequency words were used to yield 28 study lists; 2) An equal number of words (28 high-frequency words and 28 low-frequency words) served as lures on the recognition test that followed each study list. 3) Finally, 3 untested words appeared at the beginning and end of each study list to absorb primacy and recency effects.

Subjects participate individually in 7 sessions. Each session consisted of 4 blocks. In two blocks, the subjects were told to make responses as fast as possible and we name it *Speed condition*; in the other two blocks, the subjects were told to make responses as accurately as possible and we name it *Accuracy condition*. Speed conditions and Accuracy conditions alternated within a session. The conditions to start with were counterbalanced across the subjects over the 7 sessions. For two subjects, four sessions were Accuracy-Speed-Accuracy-Speed and three sessions were Speed-Accuracy-Speed-

Accuracy; for other two subjects, four blocks were Speed-Accuracy-Speed-Accuracy and three blocks were Accuracy-Speed-Accuracy-Speed.

Before each block, there was a practice block composed of 10 study words and 20 test words to prepare the subject for that condition. For each test word in a practice block in the Accuracy condition, an error response would elicit a feedback: “Wrong! Please make response as ACCURATELY as possible!” For each test word in a practice block in the Speed condition, a response longer than 1200ms will elicit a feedback: “Too slow! Please make response as FAST as possible!” Subjects were asked to apply the bias, either toward speed or accuracy, to “old/new” responses throughout the block.

3． Procedures

The procedure was same as in Experiment 1 and 2.

4.2 Results

The data analysis were presented in the same way as for Experiment 1 and 2: First the accuracy analysis results were presented including sensitivity, response proportions of “remember” and “know” judgments, and remember-know model fits; then the reaction time data analysis were presented, including the mean RT data analysis and the distributional RT data analysis.

Subjects reversed their old/new responses on 0.4% of trials by giving confidence ratings that contradicted their initial response. As in Experiment 1 and 2, these reversed responses were included in the accuracy analyses but were excluded from the reaction time analyses. The confidence ratings determined the final judgment.

4.2.1 Accuracy analysis

Repeated-measures ANOVAs were conducted for hit rates and false alarm rates with word frequency and Accuracy-Speed bias manipulations as within-subjects factors.

Low-frequency words resulted in marginally significantly higher hit rate (mean = .85) than high-frequency words (mean = .74, $F(1, 3) = 8.899$, $p = .058$). Accuracy condition resulted in higher hit rate (mean = .84) than Speed condition (mean = .76, $F(1, 3) = 33.924$, $p < .05$). No interaction of these two factors was found, $F(1, 3) = 0.997$, *n. s.*.

Low-frequency words resulted in lower false alarm rate (mean =.09) than high-frequency words (mean = .20, $F(1, 3) = 103.383$, $p < .01$). Accuracy condition resulted in lower false alarm rate (mean = .13) than Speed condition (mean = .16, $F(1, 3) = 19.879$, $p < .05$). The interaction of these two factors was not significant, $F(1, 3) = 0.783$, *n. s.*.

As in Experiment 1 and 2, because the *z*ROC slopes were less than 1 (see Table 13), d_a was used to measure sensitivity. Table 13 shows d_a values for each subject in each condition.

Low-frequency words resulted in higher d_a values (mean = 2.46) than high-frequency words (mean = 1.64) and the Accuracy condition elicited higher d_a values (mean = 2.23) than the Speed condition (mean = 1.87). A repeated-measures ANOVA with word frequency and Accuracy-Speed bias as within-subjects factors yielded reliable main effects for both word frequency ($F(1, 3) = 38.178$, $p < .01$) and Accuracy/Speed bias manipulation ($F(1, 3) = 25.75$, $p < .05$). There was no interaction of these two factors ($F(1, 3) =.026$, *n. s.*).

1. “Remember” and “know” response proportions

The manipulations of both word frequency and Accuracy-Speed bias affected “remember” hit rates but not “remember” false alarms. Word frequency mirror effect was found for “know” judgments. Accuracy-Speed bias affected “know” false alarms but not “know” hits. Table 14 shows the proportions of hits and false alarms followed by “remember” and “know” judgments in each condition collapsed over the subjects.

Repeated-measures ANOVAs were done for “remember” hits and false alarms as well as “know” hits and false alarms with word frequency and Accuracy-Speed bias as within-subjects factors.

Low-frequency words received more “remember” hits (mean = .65) than high-frequency words (mean = .45, $F(1, 3) = 19.615$, $p < .05$), and more “remember” hits were made in the Accuracy condition (mean = .57) than in the Speed condition (mean = .53, $F(1, 3) = 280.131$, $p = .000$). No interaction of these two factors was found ($F(1, 3) = 1.464$, *n. s.*). “Remember” false alarm rates were not affected by either word frequency ($F(1, 3) = 6.482$, *n. s.*) or Accuracy/Speed bias manipulation ($F(1, 3) = .123$, *n. s.*).

High-frequency words received more “know” hits (mean = .30) than low-frequency words (mean = .20, $F(1, 3) = 14.202$, $p < .05$). Accuracy-Speed bias manipulation did not affect “know” hits ($F(1, 3) = 6.019$, *n. s.*). There was no interaction of the two factors on “know” hit proportion.

“Know” false alarm rates were significantly higher for high-frequency words (mean = .19) than for low-frequency words (mean = .09, $F(1, 3) = 100.347$, $p < .01$), and were higher in the Speed condition (mean = .16) than in the Accuracy condition (mean

= .16, $F(1, 3) = 13.882$, $p < .05$). No interactions of these two factors were found ($F(1, 3) = .206$, *n. s.*).

2. Remember-know model fits

Overall the ROC data were best accounted for by the extended dual-process model as is shown in Table 15. Ten out of 16 cases were best fit with the dual-process models (the standard and the extended versions) and other 5 cases were best fit with the one-dimensional models (with fixed or variable remember-know criterion). For Subjects A, B and D, the Speed and Accuracy conditions led different conclusions: The Speed condition was best fit with the extended dual-process model whereas the Accuracy condition was best fit with a one-dimensional model except one case (for Subject B with high-frequency words as stimuli, the dual-process model fit this case best). All four conditions for Subject C were best fit with the extended dual-process model. Note that the differences between the model fits were quite small, sometimes less than 1. Therefore model selection was not compelling.

In Figure 9, the best and worst fit to each condition for Subject A are plotted. In the Accuracy condition, the variable-criterion one-dimensional model provided best fits and the dual-process model provided worst fits; in the Speed condition, the extended dual-process model provided best fits and the dual-process model provided worst fits. For each condition, the dual-process model misses the first point. The discrepancy between the dual-process model prediction and the observed data is due to an assumption made by the dual-process model: “remember” taps recollection and recollection is a high-threshold process. Therefore the dual-process model predicts a high percentage of “remember” responses at the highest rating, rating 6. Table 16 shows that, though the dual-process model predictions of the proportions of "know" response were similar to other model predictions or even closer to the observed data than other model predictions, the predicted "remember" response proportions at rating 6 were larger than what was observed and what was predicted by other models. Consequently the dual-process model missed the first point on the observed ROCs of Subject A. Similar situation was found for Subjects C and D. For Subject C, the extended dual-process model provided best fits and the dual-process model provided worst fits to the ROC data. As shown in Figure 10, the extended dual-process model hit the first points on observed ROCs for Subject C, whereas the predictions of the dual-process model were slightly off. For Subject D, the dual-process model slightly missed the first point on ROCs whereas the best-fit models hit on the spot. However, both the best-fit and worst-fit models missed the third and fourth points in the Accuracy condition for Subject D. The remember-know model fits for Subject B were quite decent, and also there was not much difference between the best fits and worst fits as illustrated in Figure 11.

4.2.2 Reaction time analysis

1. Mean RT analyses

Mean reaction time data are shown in Table 17. As in Experiment 1 and 2, comparisons of mean RTs for hits followed by "remember" and "know" judgments were conducted within subjects with independent two-sample t-tests.

For all four condition for Subject A, C, and D, mean RTs for hits followed by "remember" judgments were significantly shorter than those followed by "know" judgments (the smallest $t_{\text{observed}} = 3.963$, $p = .000$). For Subject B, in the Accuracy condition, mean RTs for hits followed by "remember" judgments were shorter than those followed by "know" judgments (the smallest $t_{\text{observed}} = 4.487$, $p = .000$); however, in the Speed condition, there was no difference (the largeset $t_{\text{observed}} = 1.531$, *n. s.*).

When the confidence was controlled[6], 4 tests out of 11 revealed significant differences (the smallest $t_{\text{observed}} = 2.627$, $p = .010$): mean RT was smaller for hits followed by "remember" judgments than those followed by "know" judgments. These four conditions were low-frequency words in the Accuracy and the Speed condition for Subject A, low-frequency words in the Accuracy condition for Subject B, and high-frequency words in the Speed condition for Subject C. For the other seven tests, the largest $t_{\text{observed}} = 2.003$, *n. s.*. So after controlling the confidence level, the differences between mean RTs for hits followed by "remember" and "know" disappeared for seven cases.

2. Distributional RT analysis

The distribution RT analyses were done for RT data collapsed over confidence ratings and when confidence level was controlled.

The mean RTs for hits followed by "remember" judgments were shorter than those followed by "know" judgments for all conditions of all subjects except for the Speed condition of Subject B. As presented in Table 18, the ex-Gaussian distribution fits showed that the differences in mean RTs could be due to both μ and τ parameters (8 out of 14 cases), μ parameter alone (4 cases), or τ parameter alone (1 case). In one condition (high-frequency words in the Speed condition for Subject D), neither μ or τ parameter values were different for hits followed by "remember" and "know" judgments. However, the mean RTs for hits followed by "remember" was shorter than that followed by "know"

[6] There are supposed to be 72 t-tests (4 subjects × 2 word frequencies × 3 presentation frequencies × 3 "old" confidence ratings) in total; however, due to the insufficient number (<10) in most cells, only 11 t-tests were conducted

judgments. This may be due to the same trend of the difference in μ and τ parameters: both were larger for “know” judgments than for “remember” judgments. In the Speed condition for Subject B, there were not differences in the mean RTs for hits followed by “remember” and “know” judgments. Examination of the best-fit ex-Gaussian distribution parameter values showed that in one case, neither μ or τ parameter was different for hits followed by “remember” and “know” judgments. In the other case, there was difference in the μ parameters whereas the difference was in opposite direction from the difference in τ parameters.

Among the four subjects, only Subject A provided enough data for comparison of best-fit ex-Gaussian distribution parameters for hits followed by “remember” and “know” judgments when confidence level was controlled. Table 19 showed the parameter values of the best-fit ex-Gaussian distribution for hits followed by “remember” and “know” judgments in the four conditions. The results of the mean RT analysis for Subject A at rating 5 revealed differences in mean RTs for hits followed by “remember” and “know” judgments for low-frequency words. Examination of the best-fit ex-Gaussian distribution parameter values showed that the difference for low-frequency words in the Accuracy condition was due to the τ parameter, and the difference in the Speed condition was due to the μ parameter. This indicated that for low-frequency words, different strategic processes underlie the “remember” and “know” judgments in the Accuracy condition. In contrast, the automatic process was different for “remember” and “know” judgments in the Speed condition, with same underlying strategic process involved in both “remember” and “know” judgments.

Examination of the 95% confidence intervals of the τ parameters showed that for the Accuracy condition with low frequency words as stimuli, the 95% confidence interval did not include 0, so that τ values were different for hits followed by “remember” and “know” judgments. As discussed in Section 1.5, different τ values suggested different strategic processes underlie “remember” and “know” judgments, which supported the dual-process model. For other three conditions, the 95% confidence intervals included 0, so that τ parameter values were similar for hits followed by “remember” and “know” judgments, thus supporting the one-dimensional models.

Figure 13 illustrates the RT distributions for hits followed by “remember” and “know” judgments with high-frequency words as stimuli in the Accuracy and Speed conditions for Subject A. The best-fit ex-Gaussian distribution is superimposed on each empirical distribution. Figure 14 illustrates analogous data with low frequency words as stimuli. As is expected, in the Accuracy condition, because the subjects had sufficient time to engage strategic processes, the skews of the RT distributions are larger than those in the Speed condition.

4.3 Discussion

The remember-know model fit results showed that the extended dual-process model provided best fits for 10 out of 16 cases and the dual-process model fit one case best; the one-dimensional model fit one case best and the one-dimensional model with variable remember-know criterion fit 4 cases best. The remember-know model fit results indicated that recollection is not an “all-or-none” threshold process, since the dual-process model only fit one case best, and all the other models assume a continuous recollection process.

The mean RTs generally replicated Dewhurst et al. (2006): the mean RT for hits followed by “remember” was faster than “know” judgments, except for the Speed condition for one subject. When reaction times were analyzed separately for each confidence rating, the differences in mean RTs for hits followed by “remember” and “know” judgments disappeared for seven cases.

Examination of the best-fit ex-Gaussian distribution parameter values without controlling confidence level revealed that the differences in the mean RTs for hits followed by “remember” and “know” judgments could be due to both μ and τ parameters, μ parameter alone, or τ parameter alone.

Only Subject A provided enough data for ex-Gaussian distribution fit when confidence level was controlled. For three out of four conditions, there was no difference between the best-fit values of the τ parameter for hits followed by “remember” and “know” responses, which indicated that same strategic process underlies the “remember” and “know” responses and thus supported the one-dimensional models. For another condition, low-frequency words in the Accuracy condition, the difference between τ values of hits followed by “remember” and “know” responses was significant. Different τ values suggested that “remember” and “know” judgments engaged different strategic processes and therefore supported the dual-process model.

The conclusions from ex-Gaussian distribution fits were not consistent with the remember-know model fits. The remember-know model fit results showed that the one-dimensional model with variable remember-know criterion provided best fits for the Accuracy condition for Subject A and the extended dual-process model provided best fits for the Speed condition for Subject A. Therefore, only in the Accuracy condition with high-frequency words as stimuli, the remember-know model fit and the ex-Gaussian distribution fit provided convergent evidence: the one-dimensional model provided best account for the data. In the other conditions, the results of the remember-know model fit contradicted the results of the ex-Gaussian distribution fit.

As pointed out in the remember-know model fit results section, the differences of AIC values between the one-dimensional model and the dual-process models were quite small. The average difference in the AIC values of the one-dimensional model and the dual-process model was 12. Whether such small difference could determine the best-fit model remains a question. Meanwhile, the 95% confidence intervals of the difference in the τ parameter values for hits followed by “remember” and “know” judgments were wide, which caused uncertainty about the difference values of the τ parameter for hits followed by “remember” and “know” judgments.

CHAPTER 5

GENERAL DISCUSSION

5.1 Summary of the experiments

Remember-know paradigm has been extensively used to investigate recognition memory. Five remember-know models have been proposed to explain the data. Since all the responses are made over time, both response accuracy and response time data should be accounted for by the true model. Therefore, in this thesis, both response accuracy data and reaction time data were used to evaluate the performance of the five remember-know models.

Three experiments were conducted with broadly used manipulations. For the response accuracy data from each experiment, the five remember-know models were fit to the ROC data and AIC statistics were used to compare the performance across the models. Ex-Gaussian distribution was fit to reaction times of hits followed by "remember" and "know" responses when confidence level is controlled. The best-fit values of the τ parameter were used to examine whether the underlying strategic processes are same or not.

The remember-know model fits showed prominent individual differences and condition differences. Overall, Experiment 1 supported the one-dimensional model: the one-dimensional model (or the one-dimensional model with variable remember-know criterion) provided best fits to the ROC data for 6 out of 8 subject-condition combinations. The RT distributions also implied that the one-dimensional model could better account for the data. For Subject A and Subject D who provided enough data to fit ex-Gaussian distribution to both hits followed by "remember" and "know" judgments when controlled for confidence ratings, the similar τ values indicated that the same strategic processes underlie the two responses, which is consistent with the one-dimensional models.

The remember-know model fit results of Experiment 2 were mixed. Two subjects' ROC data were best fit with the one-dimensional SDT model with variable remember-know criterion and one subject was best fit with the dual-process model. There were not enough data for distributional data analysis when confidence level was controlled.

In Experiment 3, the Accuracy condition for Subjects A, B and D were best fit with the one-dimensional model with fixed or variable remember-know criterion with one exception: The Accuracy condition with high frequency stimuli for Subject B was best fit with the dual-process model. In contrast, the Speed condition for Subjects A, B and D were best fit with the extended dual-process model (except for one case). Subject C was consistently best fit with the extended dual-process model. The differences in AIC statistics between models were quite small, and so model selection was not compelling. The distributional RT analysis results were mixed. In three conditions, similar τ values were found for hits followed by "remember" and "know" judgments, and thus supported the one-dimensional models. In one condition, τ values were different for hits followed by "remember" and "know" judgments, which supported the dual-process models.

In summary, the three experiments provided mixed results for model evaluation. The remember-know model fit results (except for two subject-condition combinations) and distributional RT analysis results in Experiment 1 provided convergent evidence that supported the one-dimensional models. The remember-know model fit results in Experiment 2 were mixed: two subjects were best fit with the one-dimensional models and one subject was best fit with the dual-process models. The remember-know model fits in Experiment 3 favored the extended dual-process model, whereas the distributional RT analyses favored the one-dimensional models.

The current study confirmed that mean RT is not a reliable measure. Same mean RTs could come from different shaped distributions and conclusions based on mean RTs overlook the useful distributional information and could be misleading. Take an example from Experiment 2. For the low-frequency words presented five times for Subject D, there was no difference in the mean RTs for hits followed by "remember" and "know". However, examination of the best-fit ex-Gaussian distribution parameter values showed that both the μ and τ parameters were significantly different. However, since μ and τ parameters were different in opposite directions and mean RT equals the sum of μ and τ, there was not difference in mean RTs.

If conclusion is drawn based on same mean RTs, it could be claimed that same process underlies manipulation levels A and B. However, this is not true if the RT distributions for the two manipulation levels are different. Therefore, the ex-Gaussian distribution provides a more accurate and informative approach for reaction time data analysis than mean RTs, and should be used more often in research that involves RT data analysis.

5.2 Implications for the remember-know models

The one-dimensional models provided the best account for the ROC data for two subjects and the Long condition for the other two subjects in Experiment 1, two out of three subjects in Experiment 2, and the Accuracy condition for three subjects (except for one case) in Experiment 3.

The one-dimensional models assume that in the remember-know paradigm same qualitative process underlies "remember" and "know" responses, and the only difference is that the criterion for "remember" responses is higher than that for "know" responses (Donaldson, 1996; Wixted & Stretch, 2004). Projecting this assumption onto reaction time data analysis, one dimensional SDT model predicts that similar τ parameter values of the best fit ex-Gaussian distribution for hit followed by "remember" and "know" judgments, because similar τ parameter values suggest similar strategic processes involved in the task. This prediction was confirmed in Experiment 1: when confidence level was controlled, no difference was found in the τ values for hits followed by "remember" and "know" responses. In Experiment 3, when confidence level was controlled, in three out of four conditions for Subject A, no difference was found in τ values, although in the Accuracy condition with low frequency words as stimuli, the τ values are significantly different for hits followed by "remember" and "know" .

The dual-process models provided best fits for the ROC data in Short condition for two subjects in Experiment 1, one subject in Experiment 2 and one subject in Experiment 3 as well as the Speed condition for all four subjects.

The dual-process model assumes that "remember" responses tap recollection and "know" responses tap familiarity. Recollection is a high threshold process while familiarity works according to an equal variance SDT model (Yonelinas et al., 1996). Therefore, "remember" and "know" are two qualitatively different processes with qualitatively different information engaged. Projecting this onto reaction times, the dual-process models predict that τ parameter in the ex-Gaussian distribution should have different values for hits followed by “remember” and “know” judgments. This is supported by one condition for Subject A in Experiment 3.

In some conditions where the dual-process model provided inferior fits to other models, it was due to that the dual-process model predicts a high percentage of “remember” responses at the highest confidence rating, which was not consistent with what was observed. However, the dual-process model did provide best fit for the ROC data of some subjects. This suggested that for some subjects in certain conditions, “remember” judgments were based on a high-threshold process.

STREAK provided poor fit for most data except for one case in Experiment 2 presumably because “remember” and “know” judgments were highly correlated with the “old/new” confidence ratings (the average correlation between the “remember” response rates and the “old” confidence ratings is .99), which is not predicted by STREAK.

5.3 Challenges confronting the current study

Inconsistencies were found in the remember-know model fits and the ex-Gaussian distribution fits. In Experiment 1, Subject D was best fit with the dual-process model. However, the ex-Gaussian distribution fit to the reaction time of hits followed by “remember” and “know” at confidence rating 5 shows that the strategic process indicator, τ, was similar for hits followed by “remember” and “know” and therefore supported the one-dimensional models. Similar contradictory situation was found for Subject A in the third experiment. The Speed condition for Subject A was best fit by the extended dual-process model, whereas the τ parameter values were not different for hits followed by “remember” and “know” and therefore supported the one-dimensional models. The Accuracy condition for Subject A was best fit by the one-dimensional model with variable remember-know criterion. For high-frequency words in the Accuracy condition, the τ parameter values were not different and supported the one-dimensional models, providing consistent evidence with the remember-know model fit. For low-frequency words in the Accuracy condition, τ values were significantly different for hits followed by “remember” and “know” and therefore supported the dual-process models, inconsistent with the remember-know model fit.

One possible explanation is that the differences in AIC statistics for the remember-know model fits were too small to claim real differences. Therefore, the τ parameter values of ex-Gaussian distribution provide more reliable evidence for model evaluation. Another possibility is that the reaction times were contaminated by lapses of attention due to the fatigue of the subjects. Such contamination changed the parameter values of the best-fit ex-Gaussian distribution, especially τ since there would be more outliers contributing to the tail of the distribution. Therefore for contaminated data, the difference in the τ parameter values is no long a reliable index of the strategic processes of “remember” and “know” judgments.

5.4 Future research directions

The common issue for the current three experiments is that none of them provided enough data to do the reaction time data analysis when confidence is controlled. Therefore, to design an experiment that can accomplish this goal is the next step.

One attempt is to dissociate confidence ratings with “remember/know” judgments since the subjects may always give “remember” judgment for items that receive high confidence ratings and give “know” judgment for items that receive low confidence ratings. If during the test, the subjects make confidence ratings in the first test phase and then make “remember/know” judgments in the second test phase in which the same test items are presented randomly, “remember” and “know” responses may be correlated with common confidence ratings more often.

Another attempt is to utilize a continuum of confidence ratings instead of six discrete intervals. Chance to obtain an overlapping range of confidence level related to both “remember” and “know” judgments is much larger.

In addition to using different experiment design, distributional RT data analysis can be conducted with different reaction time models. The diffusion model is one good option since it is capable of investigating response bias and accuracy-speed tradeoff. The diffusion model has been applied in various two-choice tasks such as perception, recognition memory and lexical decision and successfully accounted for the relationship between reaction time distribution and response accuracy (Ratcliff, 2004). In the diffusion model, information accumulation process begins from the starting point and stops until the amount of accumulated information reaches either the positive or the negative response boundary and therefore a decision was made. Response time equals the time of accumulation process plus a constant encoding and decision-execution time. To explicate the mechanism that underlies the cognitive tasks, the diffusion model fits the probabilities of the correct and errors responses at their five quantile[7] reaction times in each condition.

To understand “remember” and “know” judgments with the diffusion model, five quantiles are needed for hits, misses, correct rejections and false alarms. However, for the current three experiments, due to the extremely low “remember” false alarm rates, there

[7] A quantile means the fraction (or percent) of data below the given value. For example, the .1 (or 10%) quantile is the point at which 30% percent of the data fall below and 70% fall above that value.

was not enough data to calculate five quantiles for “remember” false alarms. In addition, diffusion model required that subjects only made a single attempt to complete the two-choice task, which generally takes less than 1 second. However, in the current three experiments, reaction times exceeding 1 s were quite common. Therefore it is not appropriate to use diffusion model to interpret the reaction data for our experiments. If subjects receive training to make responses quickly within 1s before the formal experiment, diffusion model could provide insights into the “remember” and “know” judgments.

To evaluate the remember-know model with RT distributions of “remember” and “know” judgments, another attempt is to compare the predicted distributions from a specific remember-know model with the observed distributions. Each remember-know model has definite predictions on response accuracy which make it feasible to compare their performances in fitting ROC data. Meanwhile, it is possible to infer reaction time distributions of “remember” and “know” responses from these models. To simulate the distributions from the predictions of the three models and then compare them with the real RT data distributions will give us some hints whether these models have the potential to predict the correct reaction time distributions for “remember” and “know” responses.

5.5 Conclusions

Three experiments were conducted to evaluate the remember-know models from response accuracy and reaction time data. Overall, the one-dimensional model provided best account for the response accuracy data. In most conditions, no difference was found in the skew parameter values of the best-fit ex-Gaussian distribution to the reaction times of hits followed by “remember” and “know” judgments. Same skew parameter values suggested that same strategic process underlie “remember” and “know” judgments. This is consistent with the assumption of the signal detection models.

Individual difference and condition difference were found in the remember-know model fit results, suggesting that different subjects could use different encoding or retrieval strategies and for different tasks.

Table 1. The z-ROC slopes and d_a values of the Long and Short conditions for each subject in Experiment 1.

Subject	List length	d_a	z-ROC slope
A	long	1.34	0.70
	short	1.77	0.52
B	long	1.79	0.53
	short	2.30	0.59
C	long	0.75	0.63
	short	1.33	0.67
D	long	1.63	0.69
	short	2.55	0.67
Overall	long	1.18	0.54
	short	1.75	0.53

Table 2. Overall "remember/know" proportion for old and new words in the Long and Short conditions in Experiment 1.

Word type	List length	Proportion	
		Remember	Know
Old	Long	0.40	0.31
	Short	0.51	0.31
New	Long	0.03	0.17
	Short	0.02	0.12

Table 3. AIC values for the fits of each remember-know model for each subject in the Long and Short conditions of Experiment 1. "Extended one-dimensional" denotes the one-dimensional model with variable remember-know criterion. The shading number is the smallest number in each row, and therefore it indicates the best fit in that condition.

Subject	Condition	Model				
		One-dimensional	Extended one-dimensional	Dual-process	Extended dual-process	STREAK
A	Long	4474	4419	4576	4430	4963
	Short	4454	4399	4640	4407	5022
B	Long	4037	4123	4087	4073	4613
	Short	3485	3507	3478	3482	3973
C	Long	3310	3342	3438	3327	3588
	Short	3304	3341	3340	3317	3742
D	Long	4877	4863	4868	4866	5092
	Short	4219	4201	4187	4190	4395

Table 4. The observed response proportions and predicted proportions for “remember” and “know” judgments across three “old” confidence ratings in Long and Short condition for Subject A (best fit by extended one-dimensional model), Short condition for Subject B and D (best fit by dual-process model) in Experiment 1. For the Short conditions for Subject B and Subject D, the “remember” judgments mostly concentrated at confidence rating 6. “Ext One-dim” denotes the one-dimensional model with variable remember-know criterion.

Subject	Conditions	Models	Response					
			“Remember”			“Know”		
			6	5	4	6	5	4
A	Long	observed	0.811	0.189	0.000	0.000	0.287	0.713
		Ext One-dim predicted	0.817	0.184	0.000	0.000	0.285	0.715
		Dual-process predicted	0.980	0.010	0.010	0.000	0.267	0.733
	Short	observed	0.799	0.201	0.000	0.000	0.347	0.653
		Ext One-dim predicted	0.805	0.195	0.000	0.000	0.348	0.652
		Dual-process predicted	0.980	0.010	0.010	0.000	0.338	0.662
B	Short	observed	0.988	0.009	0.004	0.000	0.512	0.488
		Ext One-dim predicted	0.969	0.030	0.000	0.037	0.622	0.341
		Dual-process predicted	0.980	0.010	0.010	0.000	0.557	0.443
D	Short	Observed	0.983	0.010	0.007	0.533	0.219	0.248
		Ext One-dim predicted	0.953	0.024	0.024	0.522	0.220	0.258
		Dual-process predicted	0.980	0.010	0.010	0.512	0.218	0.270

Table 5. Mean reaction times (in ms) for hits and false alarms followed by "remember" and "know" judgments at each confidence rating in the Long and Short conditions of Experiment 1. "RT" means mean reaction time and the numbers in the parentheses are standard errors of the mean RTs. N indicates the number of observations that contributed to the mean RTs.

Confidence level	List length	Item Type							
		Old				New			
		Remember		Know		Remember		Know	
		RT	N	RT	N	RT	N	RT	N
All "old" responses	Long	913 (12)	1140	1096 (15)	860	1427 (119)	76	1284 (27)	470
	Short	833 (24)	1433	1020 (15)	860	1191 (68)	40	1276 (30)	333
6	Long	886 (11)	1061	983 (25)	136	1295 (170)	50	1035 (198)	9
	Short	828 (25)	1356	852 (19)	187	1054 (65)	28	1074 (112)	3
5	Long	1156 (58)	62	1100 (26)	263	1630 (146)	11	1402 (65)	96
	Short	909 (23)	70	1027 (31)	273	1514 (173)	7	1212 (67)	64
4	Long	1657 (180)	17	1128 (22)	461	1720 (165)	15	1259 (29)	365
	Short	1051 (152)	7	1094 (22)	400	1503 (216)	5	1293 (33)	266

Table 6. The parameter values of best-fit ex-Gaussian for hits followed by "remember" and "know" responses in the Long and Short conditions for the four subjects collapsed over all "old" ratings, for Subject A at rating 5, and for Subject D at rating 6 in Experiment 1.

Subject	Condition	Remember/Know	Mean	SD	μ	σ	τ
A	Long	Remember	882	212	679	60	203
		Know	964	285	700	108	264
	Short	Remember	823	180	647	39	176
		Know	908	243	679	82	229
B	Long	Remember	873	324	550	34	322
		Know	1380	585	805	110	575
	Short	Remember	1188	780	409	38	779
		Know	1391	729	668	96	723
C	Long	Remember	972	353	622	50	349
		Know	1213	499	722	92	491
	Short	Remember	814	226	595	54	220
		Know	1111	414	701	52	411
D	Long	Remember	981	313	669	26	312
		Know	1127	401	732	71	395
	Short	Remember	886	261	628	37	258
		Know	1011	375	638	38	373
A/5	long	Remember	970	245	734	64	236
		Know	998	276	742	105	256
	Short	Remember	918	205	728	78	190
		Know	862	190	691	81	171
D/6	Long	Remember	971	304	669	26	302
		Know	977	286	695	47	283
	Short	Remember	881	255	628	37	253
		Know	850	223	629	33	221

Table 7. Confidence intervals (95%) for the differences in best-fit ex-Gaussian distribution parameters for hits followed by "remember" and "know" responses for the four subjects collapsed over all "old" ratings, for Subject A at confidence rating 5, and for Subject D at confidence rating 6 in Experiment 1. "R" represents "remember" judgments and "K" represents "know" judgments.

Subject/ confidence level	List length	Sample size		μ	95%CI		σ	95%CI		τ	95%CI	
		R	K	$\mu_R - \mu_K$	Lower bound	Upper bound	$\sigma_R - \sigma_K$	Lower bound	Upper bound	$\tau_R - \tau_K$	Lower bound	Upper bound
A	Long	179	353	-20.9	-61.2	22.7	-47.7	-79.3	-19.2	-61.4	-121.0	-4.2
	Short	262	329	-32.1	-61.6	-1.7	-42.6	-64.3	-19.3	-53.1	-95.5	-9.2
B	Long	532	98	-255.0	-311.0	-189.5	-76.0	-117.3	-12.8	-252.7	-426.3	-97.6
	Short	571	85	-259.7	-320.8	-89.1	-57.8	-134.4	39.3	56.4	-679.3	206.6
C	Long	207	82	-100.0	-160.6	-33.2	-42.2	-74.8	8.1	-141.4	-298.7	8.7
	Short	304	105	-105.9	-154.1	-53.1	2.5	-27.9	52.8	-191.4	-287.4	-89.7
D	Long	222	327	-62.6	-94.8	-27.2	-45.4	-70.6	-16	-82.9	-157.9	-15.6
	Short	296	341	-9.8	-34.9	13.9	-1.0	-19.2	17.7	-114.7	-177.8	-54.7
A/5	Long	34	102	-8.0	-92.0	139.7	-40.5	-104.8	81.6	-19.6	-185.5	108.6
	Short	53	113	37.2	-56.1	112.5	-3.5	-79.1	55.3	18.6	-71.1	128.9
D/6	Long	214	134	-25.8	-63.1	4.6	-21.0	-52.5	10.5	20.0	-55.3	95.1
	Short	291	185	-0.8	-25.8	20.4	4.6	-11.8	22.9	31.7	-24.0	87.1

Table 8. The d_a values and $z$slopes in each condition (word frequency × presentation frequency) for the four subjects in Experiment 2.

Subject	Word frequency	Presentation frequency					
		1		3		5	
		d_a	z-slope	d_a	z-slope	d_a	z-slope
A	High	1.89	0.28	3.00	0.35	3.96	0.97
	Low	3.04	0.72	3.67	0.89	4.36	2.08
B	High	1.36	0.57	1.37	0.61	2.26	0.51
	Low	1.91	0.65	2.23	0.62	2.65	0.84
C	High	1.41	0.63	2.16	0.61	2.87	0.66
	Low	1.56	0.35	2.17	0.34	3.27	0.50
D	High	1.08	0.82	1.49	0.96	1.84	0.72
	Low	1.67	0.65	2.48	0.78	2.54	0.59

Table 9. Response proportions of "know" and "remember" judgments in each condition (high frequency/low frequency, presentation frequency 1/3/5/0) for old and new words in Experiment 2.

Word type	Word Frequency	Presentation frequency	Response	
			"Remember"	"Know"
Old	High	1	0.47	0.22
	Low		0.55	0.21
	High	3	0.55	0.26
	Low		0.67	0.22
	High	5	0.68	0.25
	low		0.75	0.20
New	High	0	0.02	0.15
	Low		0.01	0.06

Table 10. AIC values of the remember-know model fits to each condition for three subjects in Experiment 2. The shading number in each row indicates the best fit. "DP" denotes the dual-process model; "DP-ext" denotes the extended dual-process model: "One-dim" denotes the one-dimensional model and "one-dim ext" represents the one-dimensional model with variable remember-know criterion.

Subject	Word frequency	Presentation frequency	Remember-know models				
			One-dim	One-dim ext	DP	DP-ext	STREAK
B	High	1	2009	1987	1973	1974	2236
		3	2003	1979	1968	1971	2268
		5	1812	1789	1783	1784	2011
	Low	1	1975	1953	1949	1946	2215
		3	1879	1856	1850	1852	2067
		5	1774	1751	1748	1749	1960
C	High	1	1610	1597	1763	1612	1992
		3	1460	1447	1587	1476	1366
		5	1313	1301	1419	1337	1559
	Low	1	1031	1020	1062	1041	1301
		3	944	933	980	952	1723
		5	800	784	825	811	939
D	High	1	1860	1836	1840	1841	1908
		3	1813	1789	1787	1790	1731
		5	1822	1800	1812	1815	1910
	Low	1	1787	1762	1781	1783	1874
		3	1740	1716	1743	1723	1957
		5	1701	1680	1701	1701	1845

Table 11. The numbers in the parentheses are the SEs of the mean RTs. “Rating” means confidence rating.

Word frequency/ rating	Presentation frequency															
	0				1				3				5			
	Remember		Know		Remember		Know		Remember		Know		Remember		Know	
	RT	N	RT	N	RT	N	RT	N	RT	N	RT	N	RT	N	RT	N
High/ overall	895 (108)	35	1105 (22)	302	671 (10)	315	1105 (30)	147	658 (12)	357	1062 (23)	173	656 (13)	452	1115 (45)	168
High/6	738 (53)	16	710	1	655 (9)	289	--	0	649 (12)	338	--	0	644 (10)	438	1078	1
High/5	1187 (375)	5	965 (21)	79	818 (80)	13	970 (39)	57	833 (60)	12	956 (20)	79	760 (63)	10	1022 (71)	85
High/4	971 (228)	14	1157 (29)	222	894 (93)	13	1190 (39)	90	789 (145)	7	1151 (38)	94	1664 (862)	4	1211 (54)	82
Low/ overall	940 (73)	17	1209 (41)	120	707 (11)	362	1197 (37)	134	713 (10)	429	1176 (33)	142	670 (10)	506	1050 (25)	131
Low/6	998 (98)	12	--	0	694 (10)	349	694	1	706 (11)	412	--	0	668 (10)	494	1018	1
Low/5	910	1	1012 (27)	36	905 (117)	7	1040 (26)	66	888 (31)	10	1012 (22)	78	757 (60)	11	985 (23)	91
Low/4	774 (54)	4	1293 (55)	84	1255 (229)	6	1360 (64)	67	845 (124)	7	1375 (58)	64	538	1	1202 (58)	39

Table 12. Confidence intervals for the differences in best-fit ex-Gaussian distribution parameters for hits followed by “remember” and “know” responses in Experiment 2. For Subject B, the conditions that provided enough data (>40) are high-frequency words presented once, three times, five times, and low-frequency words presented once; for Subject D, the conditions that provided enough data are low-frequency words presented three times and five times. “PF” denotes “presentation frequency”. “HF” denotes “high-frequency” and “LF” denotes “low-frequency”. “R” represents “remember” judgments and “K” represents “know” judgments.

Subject/ Word frequency	PF	Sample size		μ	95%CI		σ	95%CI		τ	95%CI	
		R	K	$\mu_R - \mu_K$	Lower bound	Upper bound	$\sigma_R - \sigma_K$	Lower bound	Upper bound	$\tau_R - \tau_K$	Lower bound	Upper bound
B/HF	1	77	60	-222.8	-318.4	-168.4	-40.6	-97.7	1.2	31.5	-53.3	145.1
	3	71	67	-193.9	-315.4	-148.0	21.7	-71.8	64.6	-12.1	-78.4	98.3
	5	122	42	-205.8	-373.7	-166.4	-15.5	-112.5	33.6	-14.3	-86.6	162.8
B/LF	1	101	36	-261.5	-402.6	-184.7	-34.2	-109.0	36.0	-10.2	-98.7	115.4
D/LF	3	35	113	29.8	-51.5	116.0	2.6	-65.0	63.3	-202.8	-362.6	-33.5
	5	45	107	85.6	21.4	201.6	57.1	-7.1	130.6	-159.4	-281.0	-73.1

Table 13. The d_a values and *z*slopes in Accuracy and Speed condition for high and low-frequency words for each subject in Experiment 3.

Subject	Word frequency	Condition	d_a	*z*slope
A	High	Accuracy	2.07	0.70
		Speed	1.50	0.66
	Low	Accuracy	2.97	0.51
		Speed	2.69	0.67
B	High	Accuracy	1.01	0.89
		Speed	0.67	0.89
	Low	Accuracy	1.66	0.60
		Speed	1.43	0.82
C	High	Accuracy	2.68	0.89
		Speed	2.31	0.92
	Low	Accuracy	3.35	0.82
		Speed	2.63	0.88
D	High	Accuracy	1.52	0.49
		Speed	1.37	0.53
	Low	Accuracy	2.61	0.39
		Speed	2.32	0.51

Table 14. Proportions of hits and false alarms followed by "remember" or "know" judgments in each condition of Experiment 3.

Word type	Word frequency	Condition	Response proportion	
			Remember	Know
Old	High	Accuracy	0.474	0.315
		Speed	0.421	0.277
	Low	Accuracy	0.667	0.216
		Speed	0.632	0.188
New	High	Accuracy	0.010	0.172
		Speed	0.006	0.203
	Low	Accuracy	0.002	0.067
		Speed	0.004	0.109

Table 15. AIC values of the remember-know model fits in Experiment 3. Small values (the shading numbers) indicate better fits. "One-dim" denotes the one-dimensional model and "one-dim ext" represents the one-dimensional model with variable remember-know criterion. "DP" denotes the dual-process model; "DP-ext" denotes the extended dual-process model:

Subject	Word frequency	Condition	Model				
			One-dim	One-dim ext	DP	DP-ext	STREAK
A	High	Accuracy	2666	2652	2904	2662	2801
		Speed	2852	2844	3202	2820	3143
	Low	Accuracy	1828	1808	2014	1820	2058
		Speed	2046	2064	2256	2040	2315
B	High	Accuracy	2716	2704	2697.5	2698.3	3000
		Speed	2540	2516	2522	2514	2748
	Low	Accuracy	2581	2570	2579	2579	3014
		Speed	2536	2540	2528	2512	2926
C	High	Accuracy	2108	2092	2174	2086	2426
		Speed	2362	2368	2478	2346	2815
	Low	Accuracy	1688	1680	1708	1678	1923
		Speed	2060	2056	2110	2052	2416
D	High	Accuracy	2258	2266	2404	2292	2994
		Speed	2484	2486	2578	2458	2835
	Low	Accuracy	1652	1632	1724	1642	1970
		Speed	1694	1676	1774	1672	2055

Table 16. The observed response proportions and predicted proportions for “remember” and “know” judgments across three “old” confidence ratings 4, 5 and 6 in four conditions for Subject A in Experiment 3. “WF” denotes “word frequency”, “HF” denotes high-frequency and “LF” denotes low-frequency. “Ext One-dim” denotes the one-dimensional model with variable remember-know criterion and “Ext dual-process” denotes the extended dual-process model. In each condition, the dual-process model provided worst fit and one-dimensional model or variable-criterion one-dimensional model provided best fit.

Subject	WF/Condition	Observation/model prediction	Response					
			“Remember”			“Know”		
			6	5	4	6	5	4
A	HF/Accuracy	Observed	0.805	0.189	0.007	0.000	0.620	0.380
		Ext One-dim predicted	0.797	0.198	0.004	0.005	0.597	0.398
		Dual-process predicted	0.980	0.010	0.010	0.000	0.564	0.436
	HF/Speed	Observed	0.594	0.377	0.029	0.000	0.480	0.520
		Ext dual-process predicted	0.594	0.377	0.029	0.000	0.412	0.588
		Dual-process predicted	0.980	0.010	0.010	0.000	0.412	0.588
	LF/Accuracy	Observed	0.842	0.158	0.000	0.029	0.632	0.338
		Ext One-dim predicted	0.860	0.135	0.005	0.016	0.646	0.338
		Dual-process predicted	0.980	0.010	0.010	0.029	0.619	0.352
	LF/Speed	Observed	0.822	0.175	0.003	0.032	0.587	0.381
		Ext dual-process predicted	0.822	0.175	0.003	0.069	0.477	0.454
		Dual-process predicted	0.980	0.010	0.010	0.069	0.477	0.454

Table 17. Mean RTs of “old” responses and the sample size followed by “remember” and “know” judgments in Accuracy and Speed conditions with high and low frequency words as stimuli in Experiment 3. The numbers in the parentheses are the SEs of the mean RTs.

Confidence level	Word frequency	Condition	Item Type							
			Old				New			
			Remember	N	Know	N	Remember	N	Know	N
6	High	Accuracy	884(18)	648	1386(247)	11	997(82)	7	1818(392)	5
		Speed	635(7)	539	719(208)	4	661(90)	6	631	1
	Low	Accuracy	836(14)	989	1315(146)	21	962	1	1745(611)	4
		Speed	647(5)	896	691(44)	3	723(87)	5	641(100)	4
5	High	Accuracy	1091(49)	115	1223(52)	235	1347(536)	2	1490(192)	51
		Speed	710(16)	141	673(12)	181	681(105)	5	708(27)	52
	Low	Accuracy	983(36)	98	1204(42)	188	1102(171)	2	1383(105)	15
		Speed	722(16)	128	717(15)	146	761	1	719(40)	26
4	High	Accuracy	1584(158)	10	1523(52)	270	1877(396)	8	1698(80)	228
		Speed	882(99)	13	768(18)	266	--	0	758(14)	273
	Low	Accuracy	1337(206)	2	1438(54)	143	--	0	1553(94)	91
		Speed	809(77)	8	724(20)	155	--	0	728(21)	144

Table 18. 95% CIs for the differences of ex-Gaussian parameters for hits followed by "remember" and "know" in Experiment 3. "A" represents the Accuracy condition and "S" represents the Speed condition. "R" represents "remember" judgments and "K" represents "know" judgments.

Subject/ confidence level	WF/ Condition	Sample size		μ	95%CI		σ	95%CI		τ	95%CI	
		R	K	$\mu_R - \mu_K$	Lower bound	Upper bound	$\sigma_R - \sigma_K$	Lower bound	Upper bound	$\tau_R - \tau_K$	Lower bound	Upper bound
A	HF/A	195	136	-103	-151	38	-31	-74	55	-431	-778	-197
	HF/S	175	127	-72	-118	-21	-32	-55	-5	-50	-135	17
	LF/A	302	67	-138	-315	-15	-72	-154	49	-275	-551	49
	LF/S	314	63	-162	-274	-85	-61	-108	-7	19	-54	123
B	HF/A	61	207	-107	-178	-11	-37	-112	35	-159	-107	-178
	HF/S	39	150	89	4	184	2	-57	45	-54	-141	29
	LF/A	156	155	-102	-167	-43	-52	-104	0	-236	-102	-167
	LF/S	121	133	38	-22	92	-24	-68	13	-29	-92	44
C	HF/A	272	103	-126	-225	-10	-105	-169	23	-216	-360	-94
	HF/S	271	111	-44	-74	-17	-26	-51	2	-60	-102	-16
	LF/A	302	79	-133	-225	-37	-72	-121	36	-173	-323	-45
	LF/S	307	69	-22	-116	-12	26	-55	32	-115	-141	-10
D	HF/A	241	65	-306	-452	-244	-4	-148	61	-344	-558	-101
	HF/S	198	62	-78	-217	16	-92	-178	-16	-104	-225	54
	LF/A	323	50	-236	-389	-87	-158	-247	-32	-522	-870	-220
	LF/S	282	39	-292	-381	-122	-204	-265	-103	110	-42	131

Table 19. 95% Confidence intervals for the differences of best-fit ex-Gaussian distribution parameters of hits followed by “remember” and “know” responses for Subject A at confidence rating 5 in Experiment 3. “A” represents the Accuracy condition and “S” represents the Speed condition. “R” represents “remember” judgments and “K” represents “know” judgments.

Subject/ Confidence level	WF/ Condition	Sample size		μ	95%CI		σ	95%CI		τ	95%CI	
		R	K	$\mu_R - \mu_K$	Lower bound	Upper bound	$\sigma_R - \sigma_K$	Lower bound	Upper bound	$\tau_R - \tau_K$	Lower bound	Upper bound
A/5	HF/A	54	83	-15.0	-89.0	155.5	35.2	-33.5	92.5	-221.2	-634	19.7
	HF/S	66	61	-67.9	-142.0	-0.2	-45.2	-90.3	-9.0	73.8	-3.8	158.7
	LF/A	47	43	46.4	-55.1	131.1	9.8	-74.7	77.2	-455.9	-735.2	-151.7
	LF/S	54	36	-113.0	-239.6	-12.0	-48.5	-115.3	32.1	24.2	-81.6	142.1

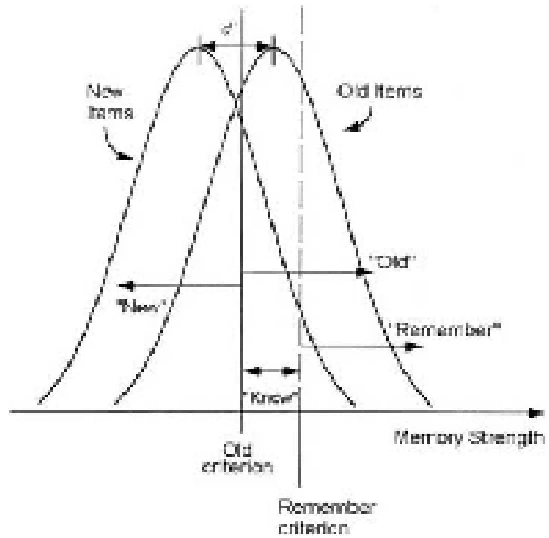

Figure 1. The one-dimensional SDT model . Old and new items differ in average strength and two criteria are used to determine responses. Observations above the old criterion receive “old” responses and those below the old criterion receive “new” responses; observations above the remember criterion lead to “remember” responses while the observations between the two criteria lead to “know” responses.

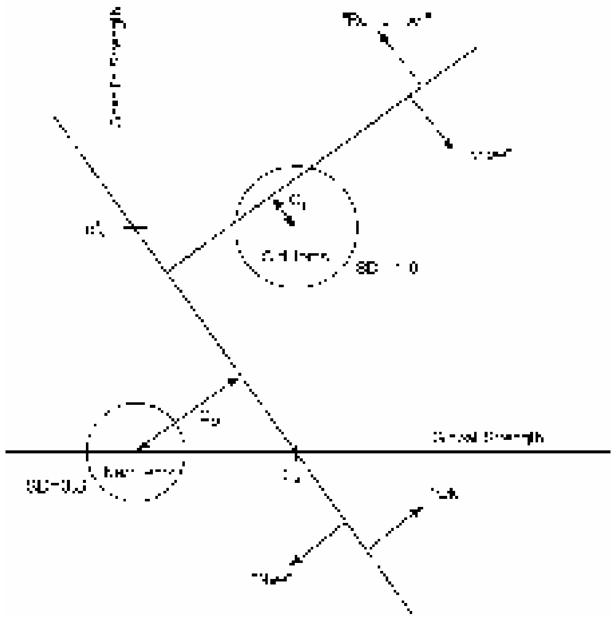

Figure 2. The STREAK model from Rotello et al. (2004). Old and new items differ in both specific and global strength. One decision bound distinguishes “old” from “new” responses on the basis of the weighted sum of the two axes; a second bound distinguishes “remember” from “know” responses on the basis of a weighted difference. Circles represent equal-likelihood contours from bivariate distributions. d_x = diagnosticity of global information; d_y = diagnosticity of specific information; C_o =distance of the old-new decision bound from the mean of the New distribution; C_r =distance of the remember-know decision bound from the mean of the Old distribution.

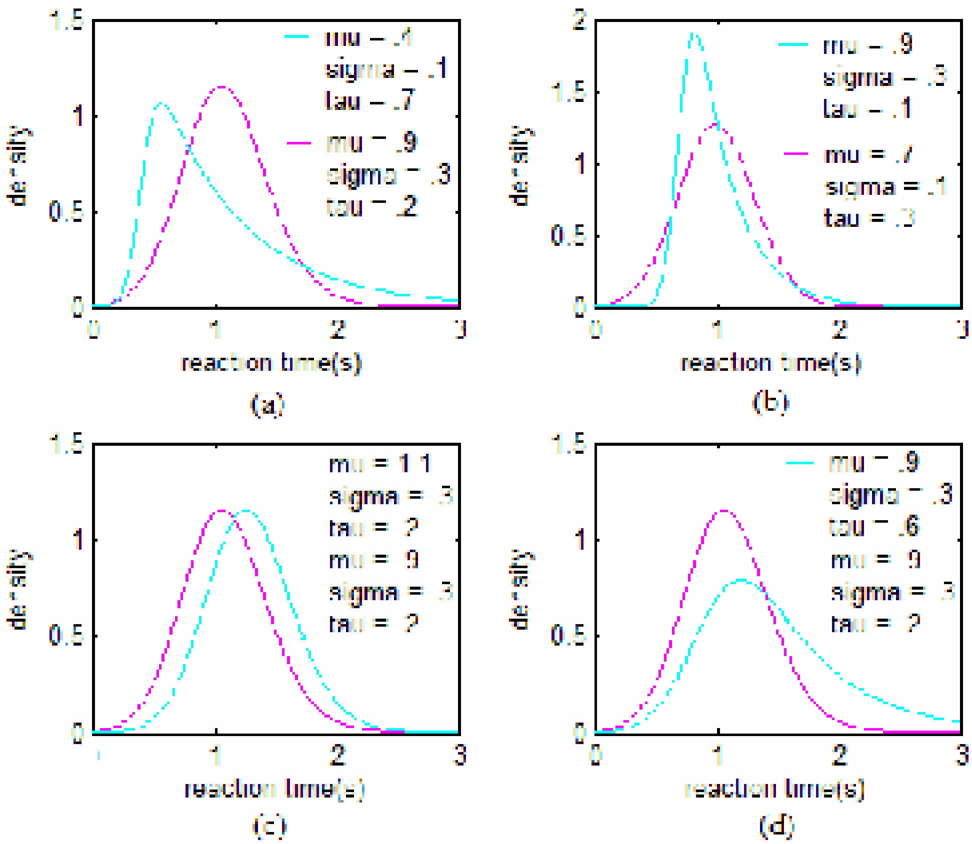

Figure 3. (a) Two ex-Gaussian distributions with the same mean (mean=1.1) but different variance (0.13 and 0.5 respectively); (b) Two ex-Gaussian distributions with the same mean (mean=1.0) and variance (variance=0.10); (c) Two ex-Gaussian distributions with different means (1.1 and 1.3 respectively); (d) Two ex-Gaussian distributions with different means (1.5 and 1.1) because of difference in the size of the tails.

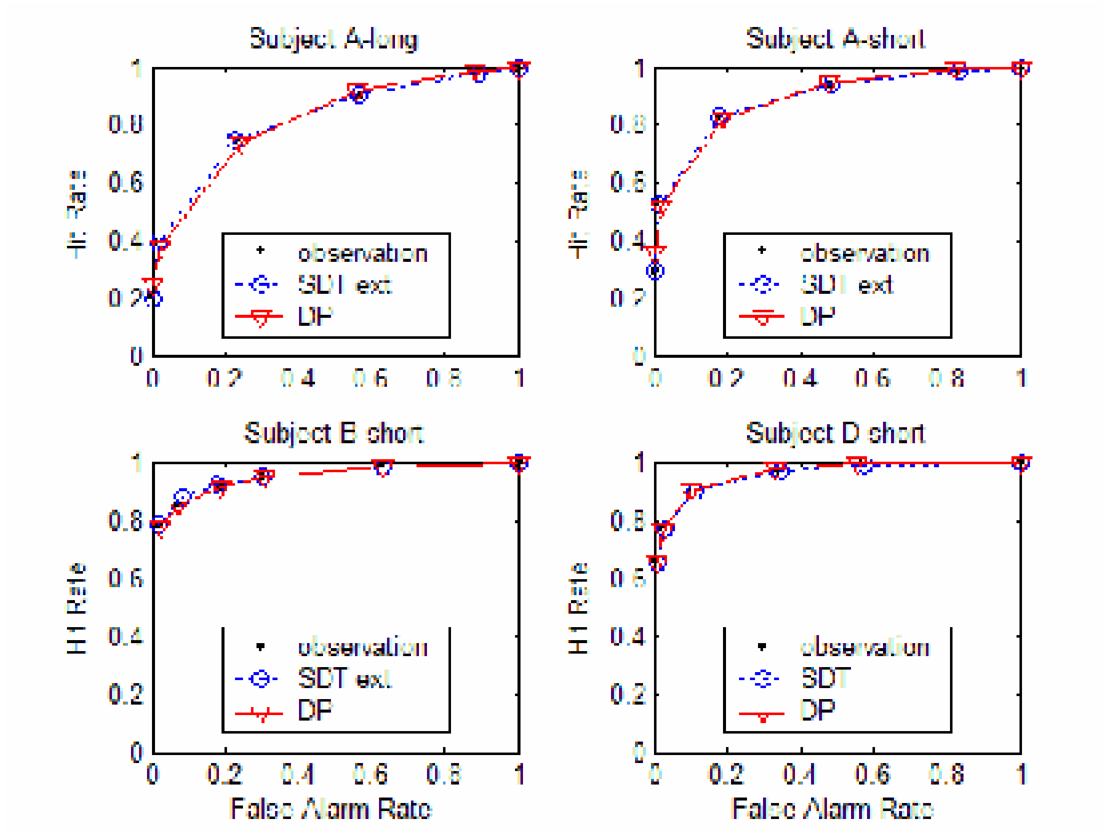

Figure 4. Illustration of the best and worst remember-know model fits to the response accuracy data in the Long and Short conditions for Subject A and in the Short condition for Subject B and D in Experiment 1. The upper left panel illustrates the fits to the Long condition of Subject A. The upper right panel illustrates the fits to the Short condition of Subject A. The bottom left panel illustrates the fits to the Short condition of Subject B and the bottom right panel illustrates the fits to the Short condition of Subject D. In Long and Short conditions for Subject A, the variable remember-know criterion one-dimensional model provided best fits and the dual-process model provided worst fits. For the Short condition of Subject B and D, the dual-process model provided best fits, and one-dimensional model or one-dimensional model with variable remember-know criterion provided worst fits. However, all the model fits are decent.

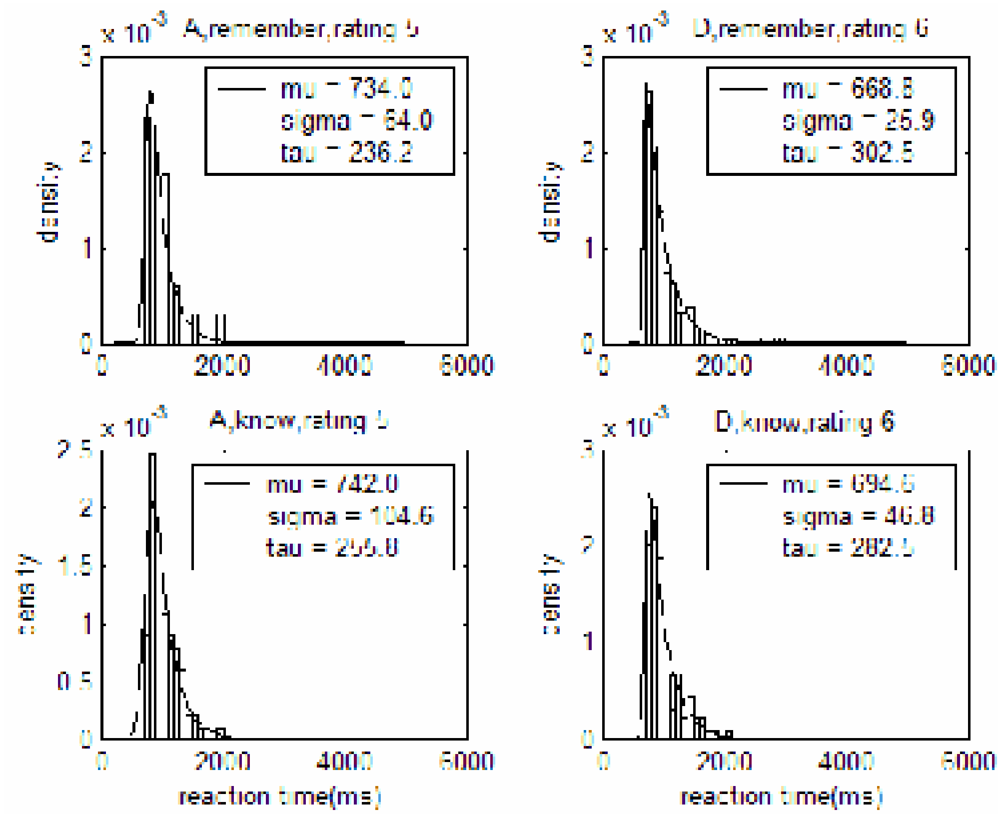

Figure 5. Reaction time distributions for hits followed by "remember" (top row) or "know" judgments (bottom row) in the Long List condition, separately for Subject A (left column) and D (right column) in Experiment 1. Subject A's data are shown for confidence level 5; subject D's data are shown for confidence level 6. In all panels, the empirical data are shown as a histogram and the best-fitting ex-Gaussian distribution is superimposed with a solid contour. Parameters of the ex-Gaussian are shown.

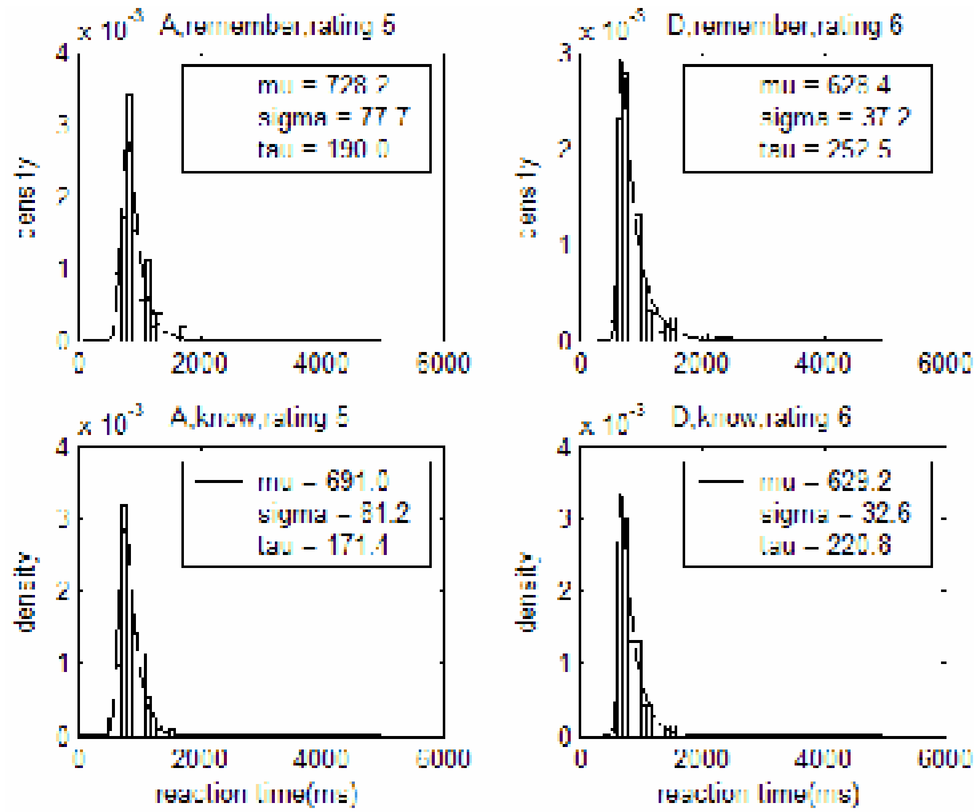

Figure 6. Reaction time distributions for hits followed by "remember" (top row) or "know" judgments (bottom row) in the Short List condition, separately for Subject A (left column) and D (right column) in Experiment 1. Subject A's data are shown for confidence level 5; subject D's data are shown for confidence level 6. In all panels, the empirical data are shown as a histogram and the best-fitting ex-Gaussian distribution is superimposed with a solid contour. Parameters of the ex-Gaussian are shown.

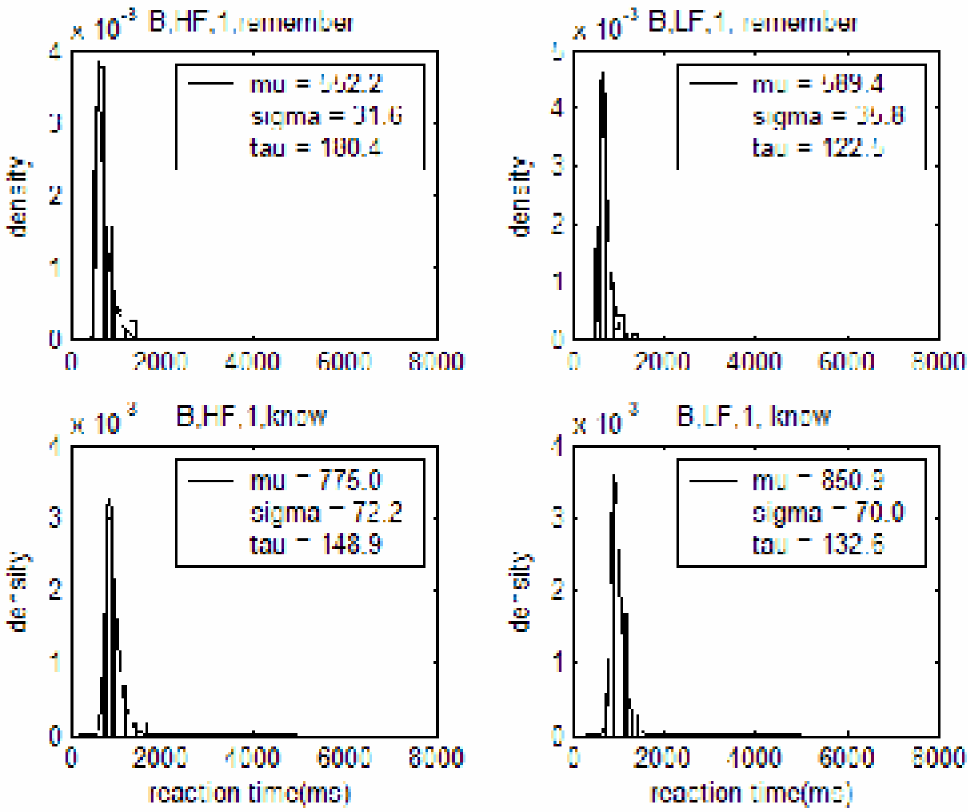

Figure 7. Reaction time distributions for hits followed by “remember” (top row) or “know” judgments (bottom row) for Subject B in Experiment 2, separately for high-frequency condition (left column) and low-frequency condition (right column) presented once. In all panels, the empirical data are shown as a histogram and the best-fitting ex-Gaussian distribution is superimposed with a solid contour. Parameters of the ex-Gaussian are shown.

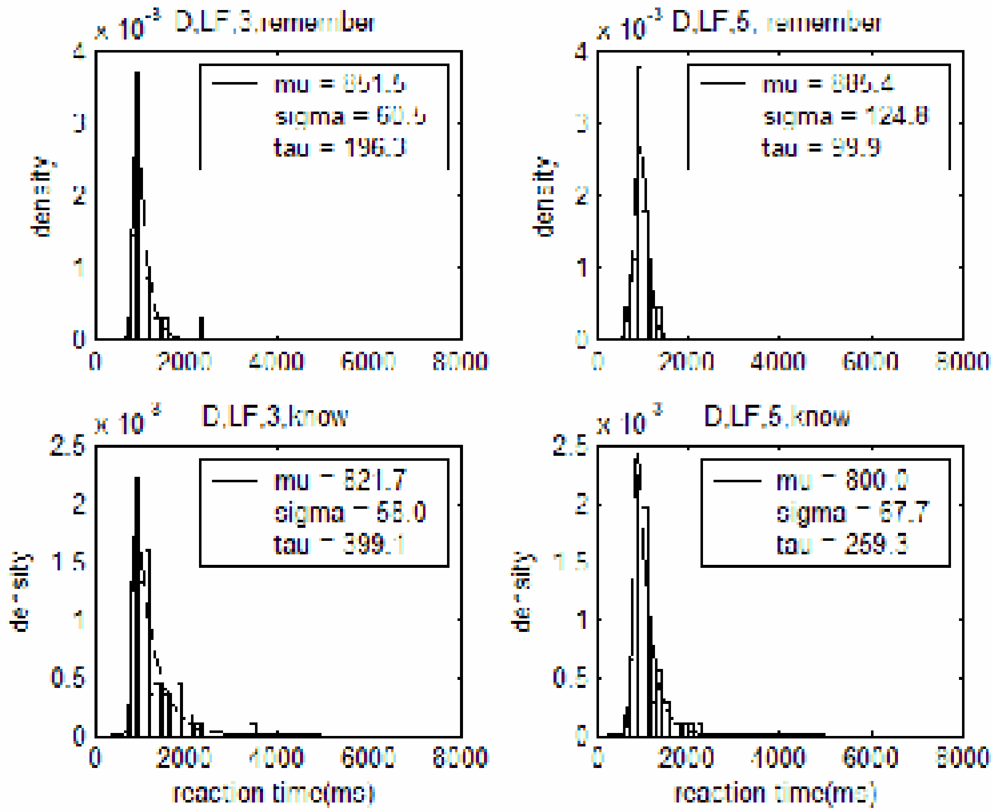

Figure 8. Reaction time distributions for hits followed by “remember” (top row) or “know” judgments (bottom row) in low-frequency word condition for Subject D in Experiment 2, separately for presentation frequency as 3 (left column) and 5 (right column). In all panels, the empirical data are shown as a histogram and the best-fitting ex-Gaussian distribution is superimposed with a solid contour. Parameters of the ex-Gaussian are shown.

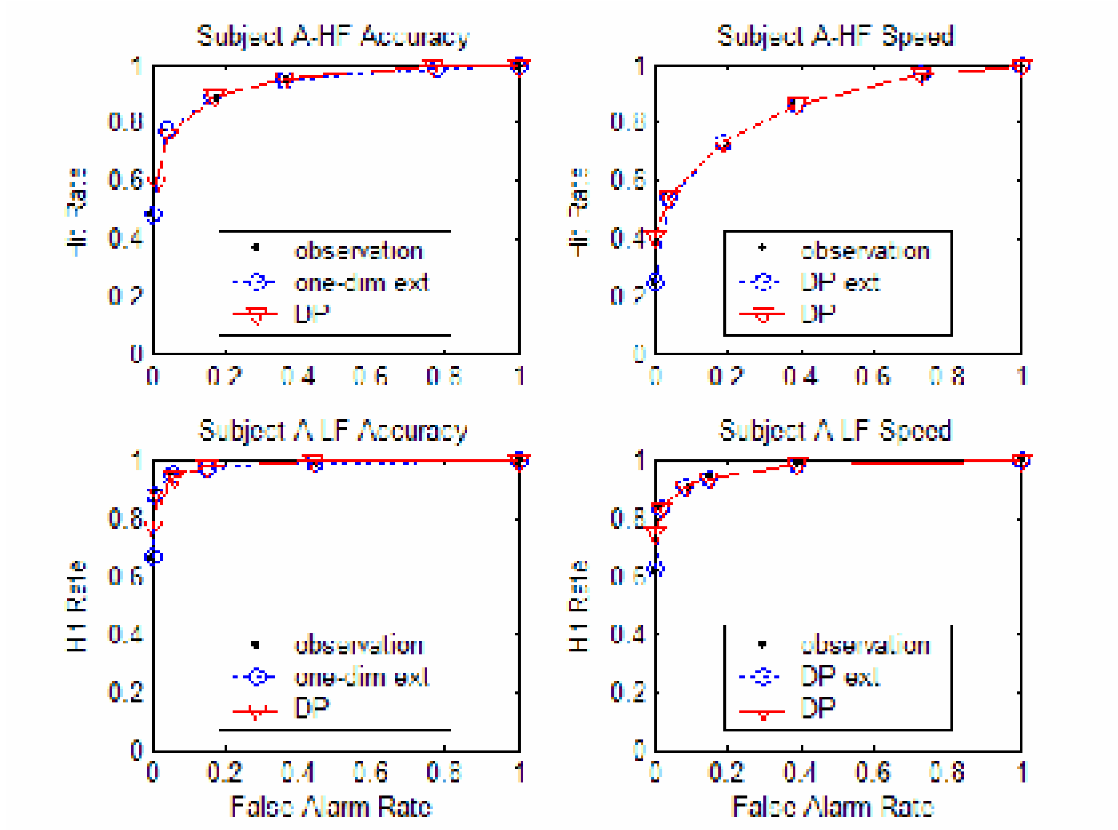

Figure 9. The fits of remember-know models to Subject A in Experiment 3, where "one-dim ext" represents the one-dimensional SDT model with variable remember-know criterion, "DP" represents the dual-process model and "DP ext" represents the extended dual-process model. Each panel illustrates the best and worst fit models for that condition. Circles illustrate the best fits and triangles illustrate the worst fits. Panel (A) illustrates the fits to high frequency words in Accuracy condition. Panel (B) illustrates the fits to high frequency words in Speed condition. Panel (C) illustrates the fits to low frequency words in Accuracy condition and Panel (D) illustrates the fits to low frequency words in Speed condition.

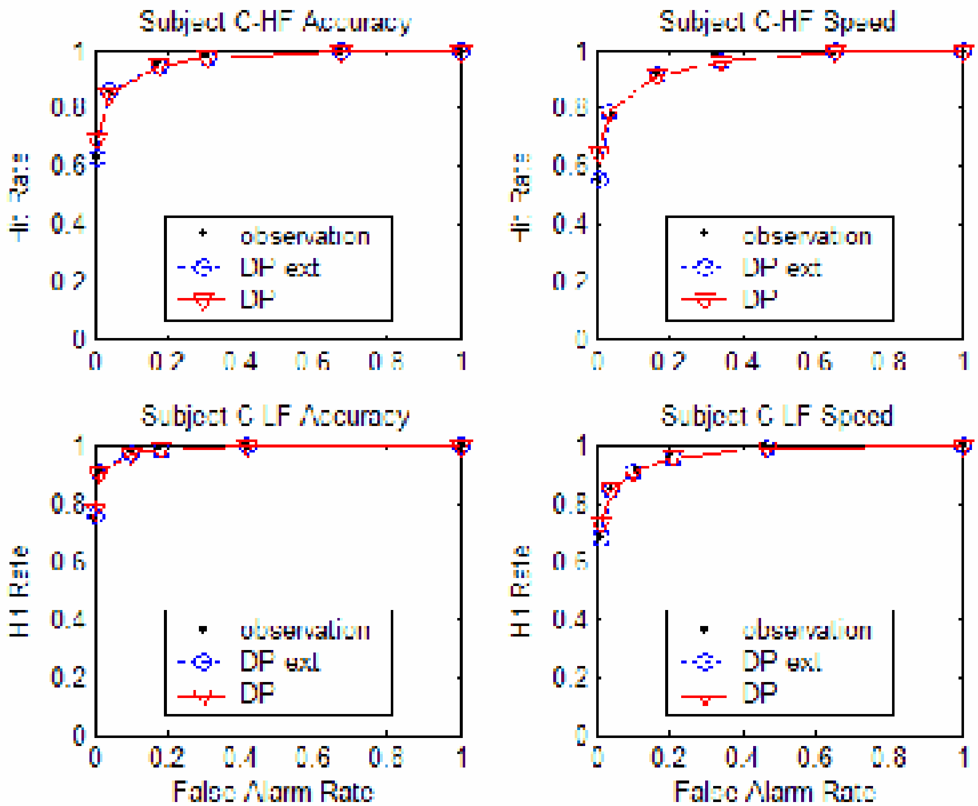

Figure 10. The fits of remember-know models to Subject C in Experiment 3, where "DP-ext" represents the extended dual-process model was as. Each panel illustrates the best and worst fit models for that condition. Circles illustrate the best fits and triangles illustrate the worst fits. Panel (A) illustrates the fits to high frequency words in Accuracy condition. Panel (B) illustrates the fits to high frequency words in Speed condition. Panel (C) illustrates the fits to low frequency words in Accuracy condition and Panel (D) illustrates the fits to low frequency words in Speed condition.

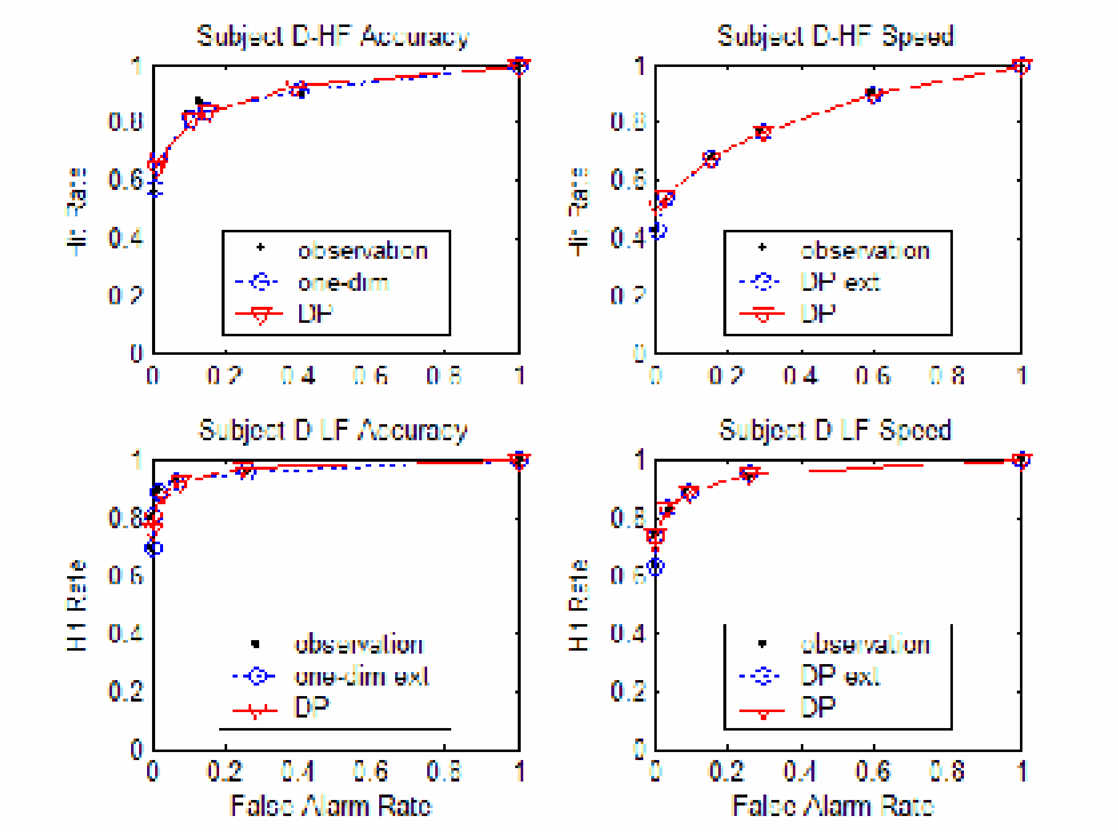

Figure 11. The fits of remember-know models to Subject D in Experiment 3, where "one-dim" represents the one-dimensional SDT model, "one-dim ext" represents the one-dimensional SDT model with variable remember-know criterion, "DP" represents the dual-process model and "DP ext" represents the extended dual-process model. Each panel illustrates the best and worst fit models for that condition. Circles illustrate the best fits and triangles illustrate the worst fits. Panel (A) illustrates the fits to high frequency words in Accuracy condition. Panel (B) illustrates the fits to high frequency words in Speed condition. Panel (C) illustrates the fits to low frequency words in Accuracy condition and Panel (D) illustrates the fits to low frequency words in Speed condition.

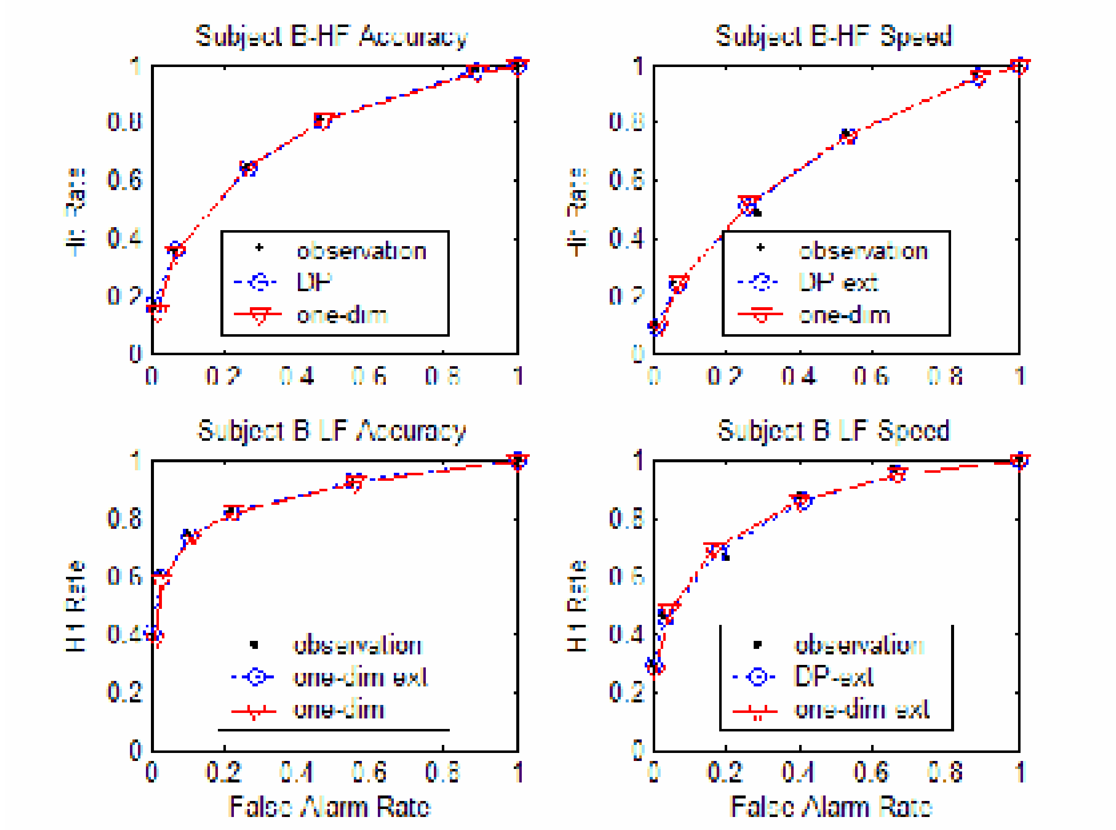

Figure 12. The fits of remember-know models to Subject B in Experiment 3, where “DP-ext” represents the extended dual-process model. Each panel illustrates the best and worst fit models for that condition. Circles illustrate the best fits and triangles illustrate the worst fits. Panel (A) illustrates the fits to high frequency words in Accuracy condition. Panel (B) illustrates the fits to high frequency words in Speed condition. Panel (C) illustrates the fits to low frequency words in Accuracy condition and Panel (D) illustrates the fits to low frequency words in Speed condition.

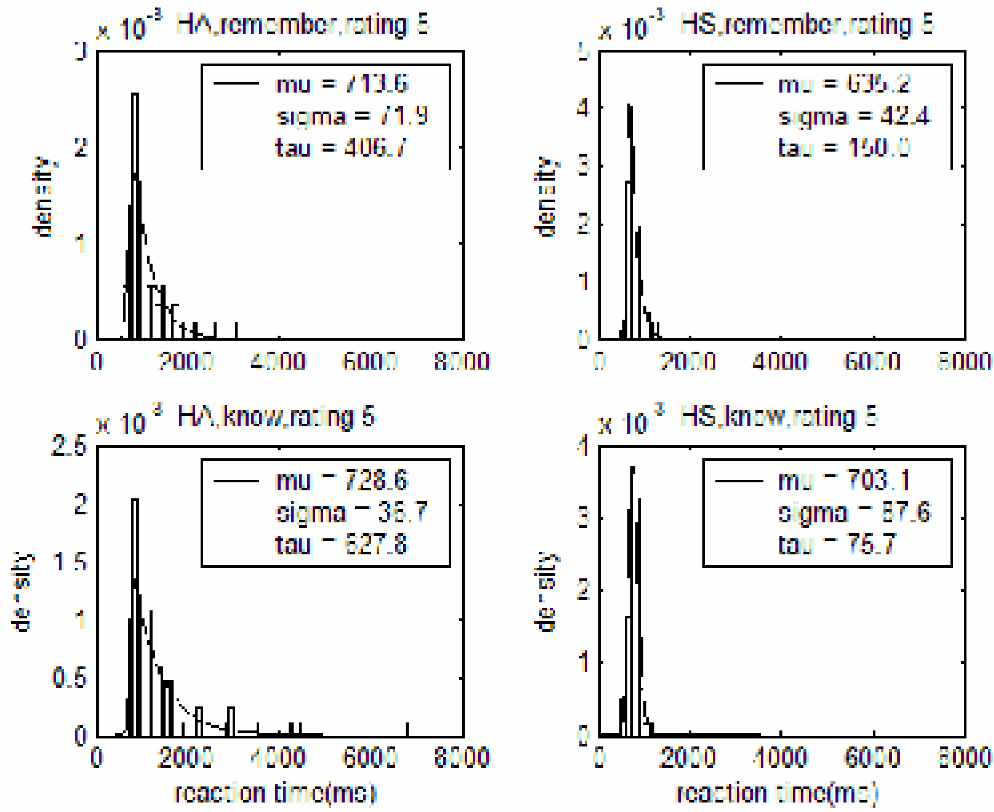

Figure 13. Reaction time distributions for hits followed by "remember" (top row) or "know" judgments (bottom row) with high frequency words as stimuli, separately for Accuracy condition (HA, left column) and Speed condition (HS, right column) for Subject A in Experiment 3. All data are for confidence rating 5. In all panels, the empirical data are shown as a histogram and the best-fitting ex-Gaussian distribution is superimposed with a solid contour. Parameters of the ex-Gaussian are shown.

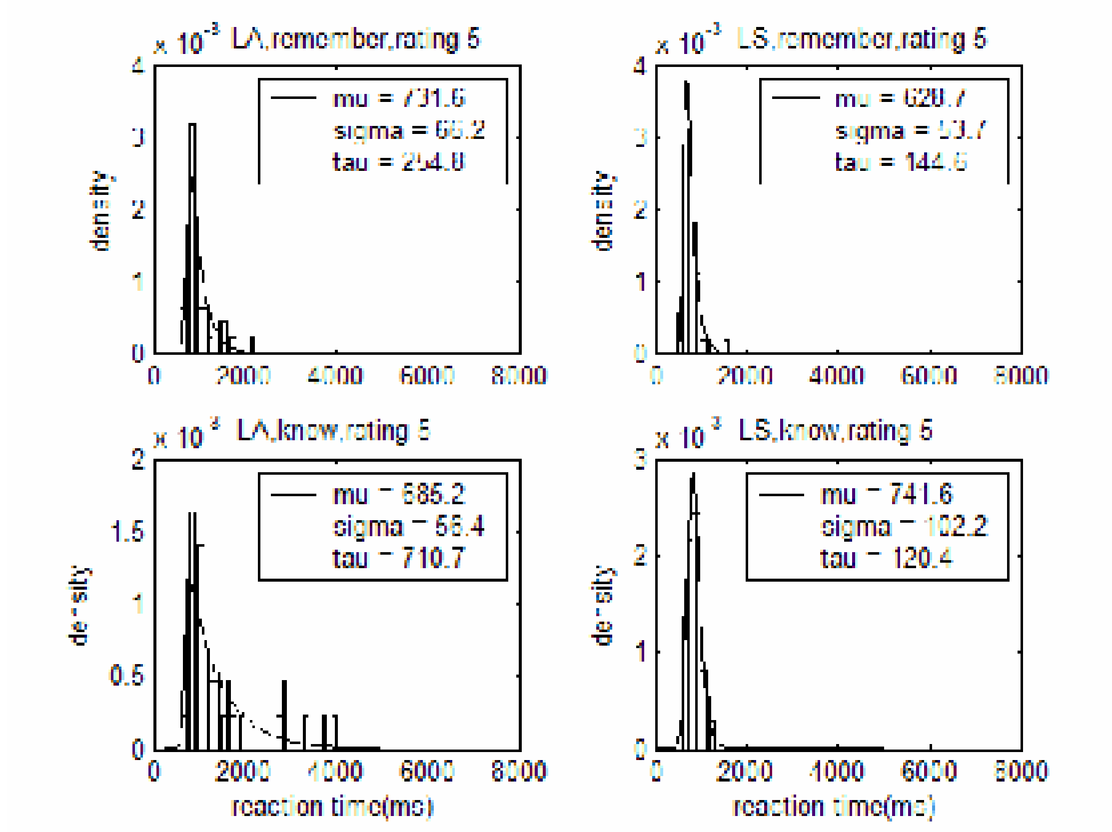

Figure 14. Reaction time distributions for hits followed by "remember" (top row) or "know" judgments (bottom row) with low frequency words as stimuli, separately for Accuracy condition (LA, left column) and Speed condition (LS, right column) for Subject A in Experiment 3. All data are for confidence rating 5. In all panels, the empirical data are shown as a histogram and the best-fitting ex-Gaussian distribution is superimposed with a solid contour. Parameters of the ex-Gaussian are shown.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19, 716-723

Allan, K., Wilding, E. L., & Rugg, M. D. (1998). Electrophysiological evidence for dissociable processes contributing to recollection. *Acta Psychologica, 98*, 231-252.

Balota, D. A., & Spieler D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: beyond measures of central tendency. *Journal of Experimental Psychology: General, 128*(1), 32-55.

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(), 231-248.

Cousineau, D., & Larochelle, S. (1997). PASTIS: A program for curve and distribution analyses. *Behavior Research Methods, Instruments, & Computers*, 29, 542-548.

Cousineau, D., Brown, S., & Heathcote, A. (2004). Fitting distributions using maximum likelihood: Methods and packages. *Behavioral Research, Methods, Instruments & Computer, 36*(4), 742-756.

Curran, T. (2000). Brain potentials of recollection and familiarity. *Memory & Cognition, 28*(6), 923-938.

Curran, T., & Cleary, A. M. (2003). Using ERPs to dissociate recollection from familiarity in picture recognition. *Cognitive Brain Research, 15*(2), 191-205.

Dewhurst, S. A., & Conway, M. A. (1994). Pictures, images, and recollective experience. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 20, 1088-1098

Dewhurst, S. A., Holmes, S. J., Brandt, K. R., & Dean, G. M. (2006). Measuring the speed of the conscious components of recognition memory: Remembering is faster than knowing. *Consciousness and Cognition, 15*(1), 147-162.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition, 24*, 525-533.

Dunn, J. C., & Kirsner, K. (2003). What can we infer from double dissociations? *Cortex, 39*(1), 1-7.

Dunn, J. C., & Kirsner, K. (1988). Discovering Functionally Independent Mental Processes - the Principle of Reversed Association. *Psychological Review, 95*(1), 91-101.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Gardniner, J. M., & Richardson-Klavehn, A. (2000). Remembering and Knowing. In E. Tulving, & Craik, F. I. M. (Eds.), *The Oxford handbook of memory* (pp. 229-244). New York: Oxford university Press.

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. Memory & Cognition, 3, 8-20

Heathcote, A. Brown, S. & Mewhort, D. J. K. (2002) Quantile maximum likelihood estimation of response time distributions. *Psychonomic Bulletin & Review*, 9, 394-401

Heathcote, A., Popiel, S. J., & Mewhort, D. J. K. (1991). Analysis of Response-Time Distributions - an Example using the Stroop Task. *Psychological Bulletin, 109*(2), 340-347.

Henson, R. N. A., Rugg, M. D., Shallice, T., Josephs, O., & Dolan, R. J. (1999). Recollection and familiarity in recognition memory: An event-related functional magnetic resonance imaging study. *Journal of Neuroscience, 19*(10), 3962-3972.

Hicks, J. L., & Marsh, R. L. (1999). Remember-know judgments can depend on how memory is tested. *Psychonomic Bulletin & Review, 6*(1), 117-122.

Hockley, W. E. (1984). Analysis of Response-Time Distributions in the Study of Cognitive-Processes. *Journal of Experimental Psychology-Learning Memory and Cognition, 10*(4), 598-615.

Hohle, R. H. (1965). Inferred Components of Reaction-Times as Functions of Foreperiod Duration. *Journal of Experimental Psychology, 69*(4), 382-386.

Kelley, C. M., & Jacoby, L. L. (1998). Subjective reports and process dissociation: Fluency, knowing, and feeling. *Acta Psychologica, 98*(2-3), 127-140.

Knowlton, B. J. (1998). The relationship between remembering and knowing: A cognitive neuroscience perspective. *Acta Psychologica, 98*(2-3), 253-265.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization.* New York: Oxford University Press.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection Theory: A User's Guide (2nd ed.).* Mahwah, NJ: Lawrence Erlbaum Associates.

Macmillan, N. A., & Rotello, C.M. (2006). Deciding about decision models of remember and know judgments: A reply to Murdock. *Psychological review*, 113 (3): 657-664

Mewhort, D. J. K., Braun, J. G., & Heathcote, A. (1992). Response-Time Distributions and the Stroop Task - a Test of the Cohen, Dunbar, and Mcclelland (1990) Model. *Journal of Experimental Psychology-Human Perception and Performance, 18*(3), 872-882.

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

Murdock, B. B. (2006). Decision-making models of remember/know judgments. *Psychological Review,* 113 (3): 648-655

Myung, J., Pitt M.A. & Kim W. (2004) Model evaluation, testing and selection. In K. Lamberts and R. Goldstone (eds.), *The Handbook of Cognition*. London: Sage Publications.

Petrusic, W. M., & Baranski, J. V.(2003) . Judging confidence influences decision processing in comparative judgments. *Psychonomic Bulletin & Review,10(1),* 177-183

Rajaram, S. (1993). Remembering and knowing: two means of access to the personal past. *Memory & Cognition, 21*(1), 89-102.

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

Ratcliff, R. (1979). Group Reaction-Time Distributions and an Analysis of Distribution Statistics. *Psychological Bulletin, 86*(3), 446-461.

Ratcliff, R. (1978). Theory of Memory Retrieval. *Psychological Review, 85*(2), 59-108.

Ratcliff, R., & Rouder, J.N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*, 347-356.

Ratcliff, R. & Smith, P. L. (2004) A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111(2),* 333-367

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember-know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology-Learning Memory and Cognition, 26*(2), 294-320.

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*(3), 588-616.

Rotello, C.M., Macmillan, N. A., Reeder, J. A. & Wong, M. (2005) The *remember* response: Subject to bias, graded, and not a process-pure indicator of recollection. *Psychonomic Bulletin & Review,12(5),* 865-873.

Rotello, C.M., Macmillan, N.A., Hicks, J. L., & Hautus, M. (2006). Interpreting the effects of response bias on remember-know judgments using signal-detection and threshold models. *Memory & Cognition, 34 (8)*: 1598-1614.

Rouder, J. N., Lu, J., Speckman, P., Sun, D. H., & Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychonomic Bulletin & Review, 12*(2), 195-223.

Spieler, D. H., Balota, D. A., & Faust, M. E. (2000). Levels of selective attention revealed through analyses of response time distributions. *Journal of Experimental Psychology-Human Perception and Performance, 26*(2), 506-526.

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 & Cognition, 24*, 1379-1396.

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.

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

Van Zandt, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review, 7*(3), 424-465.

Wheeler, M. E., & Buckner, R. L. (2004). Functional-anatomic correlates of remembering and knowing. *NeuroImage, 21*(4), 1337-1349.

Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review, 1*, 89-106.

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review, 11*(4), 616-641.

Wixted, J. T. (2007). Dual-Process Theory and Signal-Detection Theory of Recognition Memory. *Psychological Review,* 114 (1): 152-176.

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*(6), 1341-1354.

Yonelinas, A. P., Dobbins, I., Szymanski, M. D., Dhaliwal, H. S., & King, L. (1996). Signal-detection, threshold, and dual-process models of recognition memory: ROCs and conscious recollection. *Consciousness and Cognition, 5*(4), 418-441.

Yonelinas, A. P. (2001). Components of episodic memory: the contribution of recollection and familiarity. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 356*(1413), 1363-1374.

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., 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.