Chapter 1

# Aging and response times: a comparison of sequential sampling models

Roger Ratcliff, Anjali Thapar,
Philip L. Smith, and Gail McKoon

## Abstract

Ratcliff and colleagues have examined the effects of aging on cognitive processes in a number of two choice tasks. They fit the diffusion model to the response time and accuracy data for each task and interpreted the effects of aging in terms of the components of processing identified by the model. The question addressed in this chapter is whether the interpretations are specific to the diffusion model. To address this question, we fit two other models, the accumulator model and the leaky competing accumulator model of Usher and McClelland, to the data from young and older subjects for six experiments. We found that, although the diffusion model fit the data better than the other models for most of the experiments, the models' explanations of how aging affects components of processing do not differ significantly.

## Introduction

A central finding in the literature on aging is that people's response times in cognitive tasks increase with age. Along with the increase in response times, performance sometimes shows a decrease in accuracy. Recently, Ratcliff, Thapar, and McKoon (2001, 2003, 2004), Ratcliff, Thapar, Gomez, and McKoon (2004), and Thapar, Ratcliff, and McKoon (2003) examined the effects of aging on performance in a number of two-choice decision tasks: signal detection-like tasks, a masked brightness discrimination task, a recognition memory task, a lexical decision task, and a masked letter discrimination

Preparation of this article was supported by NIA grant R01-AG17083, NIMH grants R37-MH44640 and K05-MH01891.

task. These tasks were chosen because they span a range of cognitive processes that might be expected to show deficits, including perceptual processing, lexical processing, and memory, with the signal detection task representing a case where only general deficits might occur. Ratcliff and his colleagues applied a sequential sampling model, the diffusion model (Ratcliff, 1978, 1981, 1985, 1988; Ratcliff, Van Zandt, & McKoon,1999; Ratcliff & Rouder, 1998, 2000), to the data to identify the effects of aging on several of the components of processing that determine performance, separating from each other such factors as the quality of stimulus information available to the processing system and the amount of information required before making a decision.

The older subjects in these studies adopted more conservative criteria for their decisions than the young subjects and they were also slower in components of processing outside the decision process (e.g. encoding and response execution). In all of the tasks except letter discrimination, the quality of the stimulus evidence driving the decision process was not significantly lower for the older subjects than the young ones. For the brightness and letter discrimination tasks, the deficits occurred exactly as would be predicted from psychophysical research on the effects of aging on visual discrimination (Coyne, 1981; Fozard, 1990; Owsley, Sekuler, & Siemsen, 1983; Spear, 1993): a deficit occurred with the high spatial frequencies of letters in the letter discrimination task but not with the low spatial frequencies of the stimuli in the brightness discrimination task.

The advantage of models like the diffusion model is that they provide insights into performance that allow the response time (RT) and accuracy data from two-choice tasks to be decomposed into components of processing. This theoretical approach also has the advantage that it deals with all aspects of the data: correct and error RTs and their relative speeds, the shapes of the RT distributions for correct and error responses, and accuracy values.

This chapter has two aims. First, we fit two other sequential sampling models to the 12 sets of data from Ratcliff et al. (2001, 2003, 2004), Ratcliff et al. (2004), and Thapar et al. (2003) in order to determine whether the conclusions from the diffusion model about the components of processing that differ between young and older subjects are the same for the other models. The two other models, reviewed by Ratcliff and Smith (2004), are the accumulator model (Smith & Vickers, 1988; Vickers, 1970, 1979; Vickers, Caudrey, & Willson, 1971) and the leaky competing accumulator model (the LCA model, Usher & McClelland, 2001). Second, the relative qualities of the fits across the three models allows a moderately comprehensive comparison of them that serves to extend the comparisons in Ratcliff and Smith.

## The diffusion model

The diffusion model is a model of the cognitive processes involved in making simple twochoice decisions. It separates the quality of evidence entering the decision from the decision criteria and from other, nondecision processes such as encoding the stimulus and response execution. The diffusion model, like the other two models as they are considered here, applies only to relatively fast two-choice decisions (mean RTs less than about 1000 to 1500 ms) and only to decisions that are a single-stage decision process (as opposed to the multiple-stage processes that might be involved in, for example, reasoning tasks or card sorting tasks). Other models in the class of diffusion models have been applied to other types of decision making (Busemeyer & Townsend, 1993; Diederich, 1995, 1997; Roe, Busemeyer, & Townsend, 2001) and to simple RT (Smith, 1995).

The diffusion model assumes that decisions are made by a noisy process that accumulates information over time from a starting point toward one of two response criteria or boundaries, as in Figure 1.1, where the starting point is labeled $z$ and the boundaries are labeled $a$ and 0. When one of the boundaries is reached, a response is initiated. The rate of accumulation of information is called the drift rate ($v$), and it is determined by the quality of the information extracted from the stimulus. There is noise (variability) in the process of accumulating information from the starting point toward the boundaries so that processes with the same mean drift rate do not always terminate at the same time (producing RT distributions) and do not always terminate at the same boundary (producing errors). This source of variability is called within trial variability. Empirical RT distributions are positively skewed and in the

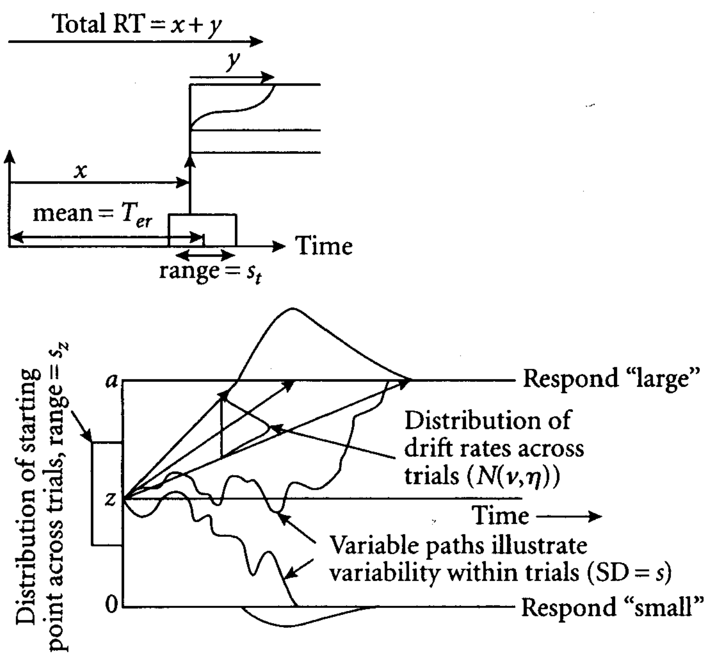


**Fig. 1.1** An illustration of the diffusion model. The top panel shows the combination of the nondecision component of RT, $x$, and the decision component, $y$. The bottom panel shows the diffusion decision process.

diffusion model, this is naturally predicted by simple geometry (see figure 1.1, Ratcliff & Rouder, 1998).

Components of processing are assumed to be variable across trials. From a theoretical perspective, one would not expect subjects to be able to achieve identical settings of the various components of processing from trial to trial (e.g. Van Zandt & Ratcliff, 1995). From a practical perspective, the assumption of across-trial variability allows the model to account for differences in RTs between correct and error responses (Luce, 1986). Variability in drift rate across trials leads to slow errors and variability in starting point leads to fast errors (Ratcliff et al. 1999; Ratcliff & Rouder, 1998), and the relative values of the two control the pattern of error compared to correct RTs that is obtained in an experiment. Drift rate is assumed to be normally distributed with standard deviation $\eta$ and starting point is assumed to be uniformly distributed with range $s_z$. In addition, the nondecision components, which are combined into one component with mean $T_{er}$, are assumed to have variability across trials that is uniformly distributed with range $s_t$ (Ratcliff, Gomez, & McKoon, 2004; Ratcliff & Tuerlinckx, 2002).

In four of the six experiments discussed in this chapter, subjects are sometimes instructed to respond as quickly as possible and sometimes to respond as accurately as possible. Speed-accuracy tradeoffs are modeled by altering the boundaries of the decision process—wider boundaries require more information before a decision can be made and this leads to more accurate and slower responses.

For each stimulus condition in an experiment, it is assumed that the rate of accumulation of evidence is different and so each has a different value of drift, $v$. A zero point, the drift criterion, separates stimuli into those with positive drift rates and those with negative drift rates, functioning in the same way as the criterion in signal detection theory. Like the signal detection criterion, the value of the drift criterion can vary with experimental manipulations such as payoffs or the proportions of stimuli for which one versus the other of the responses is correct (Ashby, 1983; Link, 1975; Link & Heath, 1975; Ratcliff, 1978, 1985, 2002; Ratcliff et al. 1999). Changing the drift criterion from one block of trials to another is equivalent to adding or subtracting a constant to the drift rates for all stimuli in one block relative to another (Ratcliff, 2002).

In sum, the parameters of the diffusion model correspond to the components of the decision process as follows: $z$ is the starting point of the accumulation of evidence, $a$ is the upper boundary and the lower boundary is set to 0; $\eta$ is the standard deviation in drift rate across trials; $s_z$ is the range of the starting point across trials; $T_{er}$ is the mean time taken up by the nondecision components of processing, and $s_t$ is the range of the values of $T_{er}$ across trials. For each stimulus condition in an experiment, there is a different value of drift, $v$. Within-trial variability in drift rate ($s$) is a scaling parameter for the diffusion process (i.e. if it were doubled, other parameters could be multiplied or divided by two to produce exactly the same fits of the model to data). $s$ is set to 0.1 in fits to the data as it has been in other applications of the model to data.

## The accumulator model

The accumulator model (Smith & Vickers, 1988; Vickers, 1970, 1978, 1979; Vickers et al. 1971) assumes that evidence in favor of one response is accumulated in one accumulator, evidence in favor of the other response is accumulated in a second accumulator, and the decision is determined by the first accumulator to reach its criterion (see Figure 1.2). Evidence is accumulated at discrete time steps. The amount of evidence accumulated on each step is variable, normally distributed with standard deviation 1.0 and a mean, $\mu$, that depends on the quality of the information from the stimulus. With this variability, information in the wrong accumulator can reach its criterion first, leading to an error. A criterion, termed the "sensory referent," is set on the underlying evidence dimension such that if the amount of evidence sampled at a time step falls above the criterion, an amount equal to the difference between that amount and the criterion is added to one accumulator. If the amount falls below the criterion, the difference is added to the other accumulator. Like the drift criterion in the diffusion models, this criterion represents a point of zero stimulus information. Because evidence is accumulated at discrete time steps, a parameter, $\lambda$, is required to converts time steps to continuous time.

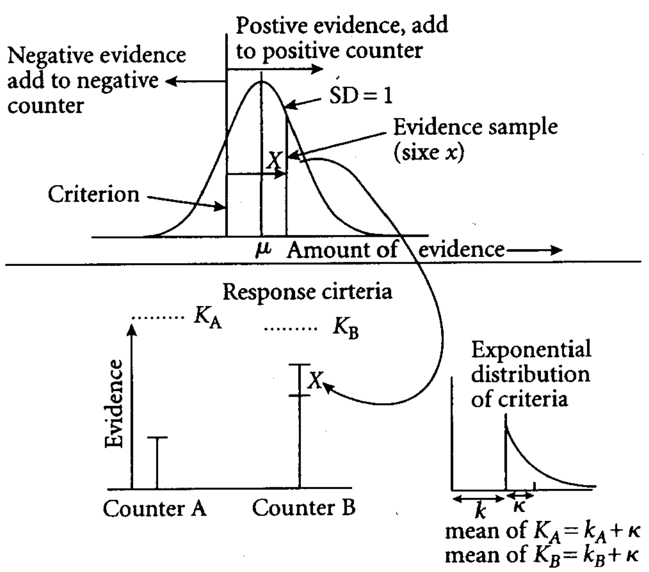


**Fig. 1.2** An illustration of the accumulator model.

Ratcliff and Smith (2004) assumed that there is variability across trials in three components of the accumulator model (*cf.* Smith & Vickers, 1988) for the same reasons as for the diffusion model. First, the means of the evidence distributions are assumed to vary randomly across trials according to a normal distribution with mean $\mu$ and standard deviation $\sigma_{\mu}$. This variability is the counterpart to variability in drift rates across trials in the diffusion model. Second, the nondecision component of RT varies across trials with a rectangular distribution with mean $T_{er}$ and range $s_t$, exactly as in the diffusion model. Third, the values of the response criteria vary across trials. The values of the criteria on each trial (Figure 1.2) are calculated by adding a value obtained from an exponential with mean $\kappa$ to two base values, $k_A$ and $k_B$ (the same value added to each), to obtain the values of $K_A$ and $K_B$ for each trial (i.e. the mean values of the criteria are $k_A + \kappa$ and $k_B + \kappa$). Without this variability, RT distributions are not skewed enough to match empirical data. Also, in order to accommodate differences in performance, it is necessary for the mean of the exponential to be larger when subjects are instructed to respond accurately than when they are instructed to respond quickly.

This model produces error responses that are slower than correct responses, a pattern that is often found in experimental data, but not always. An important problem for the model is that no way has been found for it to produce errors faster than correct responses, a pattern often obtained experimentally especially in paradigms such as choice RT (Ratcliff & Smith, 2004). The problem arises because the evidence sample is larger for evidence added to the positive accumulator (in Figure 1.2) than evidence added to the negative accumulator, because the average amount of evidence above the criterion is larger (e.g. above $\mu$) than the average amount below (e.g. just a little below the criterion).

## The leaky competing accumulator model

The LCA model (Usher & McClelland, 2001) was developed as an alternative to the diffusion model with the aim of implementing neurobiological principles that the authors felt should be incorporated into RT models, especially mutual inhibition mechanisms and decay of information across time.

The LCA model is similar to the accumulator model in that evidence is accumulated in separate accumulators for the two responses (see Figure 1.3), but the accumulation processes themselves are modeled as diffusion processes. Evidence is continuously distributed and accumulates in continuous time, just as in other diffusion process models. The rate of accumulation is a combination of three components. The first is the input from the stimulus, $v$, with a different value of $v$ for each experimental condition. If the input to one of the

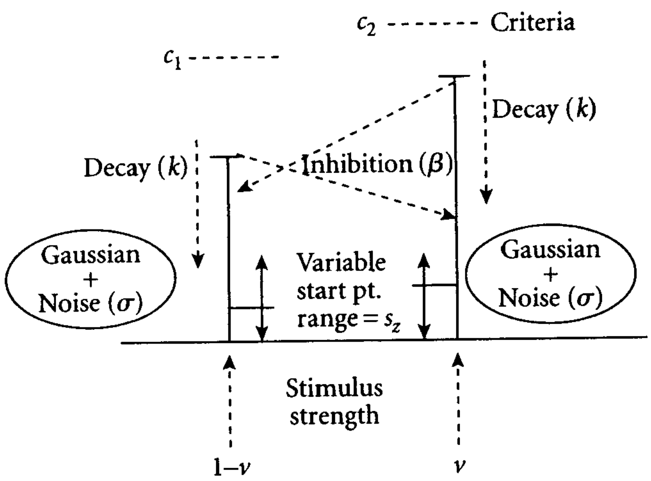


**Fig. 1.3** An illustration of the leaky competing accumulator model.

accumulators is $v$, the input to the other is $1 - v$ so that the sum of the two rates is 1. The second component is decay in the amount of accumulated information, $k$, with decay growing as the amount of information in the accumulator grows, and the third is inhibition from the other accumulator, $\beta$, with the amount of inhibition growing as the amount of information in the other accumulator grows. Combining the three components, the equivalent of drift rate in the diffusion model for accumulator $i$ (where $j$ is the competing accumulator) is $v - kx_i - \beta x_j$, where $x_i$ and $x_j$ is the amount of evidence already accumulated in accumulator $i$ and $j$ respectively.

If the amount of inhibition is large, the model exhibits features similar to the diffusion model because an increase in accumulated information for one of the response choices produces a decrease for the other choice. The assumption of inhibition between accumulators makes the model similar to an earlier, discrete-time model proposed by Heuer (1987).

The rate of accumulation of information is variable; the amount of evidence added to an accumulator on each trial includes Gaussian variability with standard deviation $\sigma$. Because of this variability, accumulated information can reach the wrong criterion, resulting in an error. Because of the decay and inhibition in the accumulation rates, the tails of RT distributions are longer than would be produced without these factors (*cf.* Vickers, 1970, 1979; Vickers et al. 1971), which leads to good matches with the skewed shape of empirical RT distributions.

The expression for the increment to the amount of accumulated information at time $t$ in accumulator $i$, is:

$$dx_i = \left[ v_i - kx_i - \beta \sum_{j \neq i} x_j \right] \frac{dt}{\tau} + \sigma \sqrt{\frac{dt}{\tau}},$$

where $dt/\tau$ is set to 0.1 to correspond to 10 ms steps as in Usher and McClelland (2001). The amount of accumulated information is not allowed to take on values below zero, so if it is computed to be below zero, it is reset to zero; this constraint is written as $x_i \rightarrow \max(x_i, 0)$ and it introduces nonlinearity into the model.

The LCA model without across-trial variability for any of its components predicts errors slower than correct responses. To produce errors faster than correct responses, Usher and McClelland assumed variability in the accumulators' starting points, just as is assumed for the diffusion model. Also, we made the same assumption about nondecision components of processing as for the diffusion and accumulator models, that they vary with a rectangular distribution with range $s_t$ and mean $T_{er}$ (Ratcliff & Smith, 2004).

## Displaying data and fitting the models to the data

In fits of any model to RT data, there are two dependent variables to consider, accuracy and RT. The proportions of correct and error responses and the relationships between their RTs, as well as the distributions of the RTs, must all be considered when assessing the fit of a model. Traditionally, accuracy, mean RTs, and RT distributions have all been plotted separately as a function of experimental condition. Here, we display them all together in quantile probability functions (QPFs). This method of displaying the data has the advantage that the joint behaviors of the dependent variables can be more easily examined. The QPF derives from the latency probability function (LPF), which was used to display the joint behavior of mean RT and accuracy in early work on sequential sampling models by Audley and Mercer (1968), Audley and Pike (1965), LaBerge (1962), Pike (1973), Pike and Ryder (1973), and others.

A QPF is constructed by plotting the quantiles of the distribution of RTs for each experimental condition on the $y$-axis and the probability of the response on the $x$-axis. For the data presented in this chapter, we used five quantiles, with the plotted quantile points representing the RTs below which fall 0.1, 0.3, 0.5, 0.7, and 0.9 of the total probability mass in the distribution. The lines that connect the quantiles across experimental conditions, as in Figures 1.4 and 1.5, form the QPF, and the shape of this function must be explained by the models. For each experimental condition, the quantile points plotted on the $y$-axis show the shape of the RT distribution. Because there is 0.2 probability mass between each pair of quantiles (e.g. the 0.1 and 0.3 quantiles), equal area rectangles can be constructed between the quantiles and these approximate the RT histograms that conventional analyses would produce (see Ratcliff & Smith, 2004, figure 1.5).

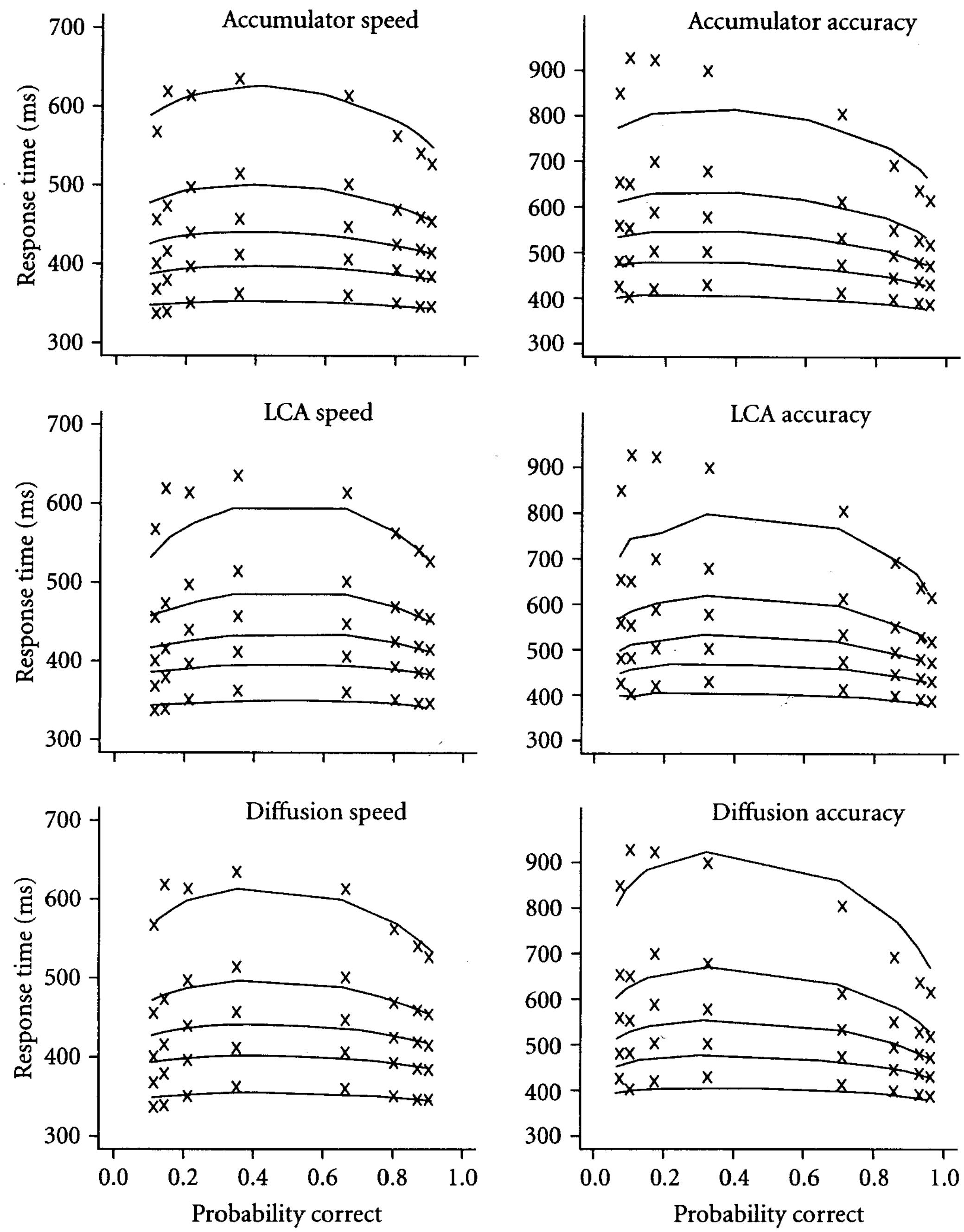


**Fig. 1.4** Sample fits for the accumulator, LCA, and diffusion models to data from the letter discrimination experiment with young subjects. $X^2$ values are 22.4, 15.7, and 7.5, respectively. (a) Letter discrimination: young subjects; (b) diffusion model recognition memory older subject data; (c) LCA model recognition memory older subject data.

A full representation of the data from an experiment requires two QPFs, one for each response, as were plotted for the data from Experiments 4, 5, and 6 below. However, if the data are symmetric for the two responses (i.e. RTs and accuracy values for the two choices are about the same), they can be averaged across responses to give a single QPF with error responses plotted to the left of 0.5 and correct responses to the right. Experiments 1, 2, and 3 below yielded symmetric data of this kind.

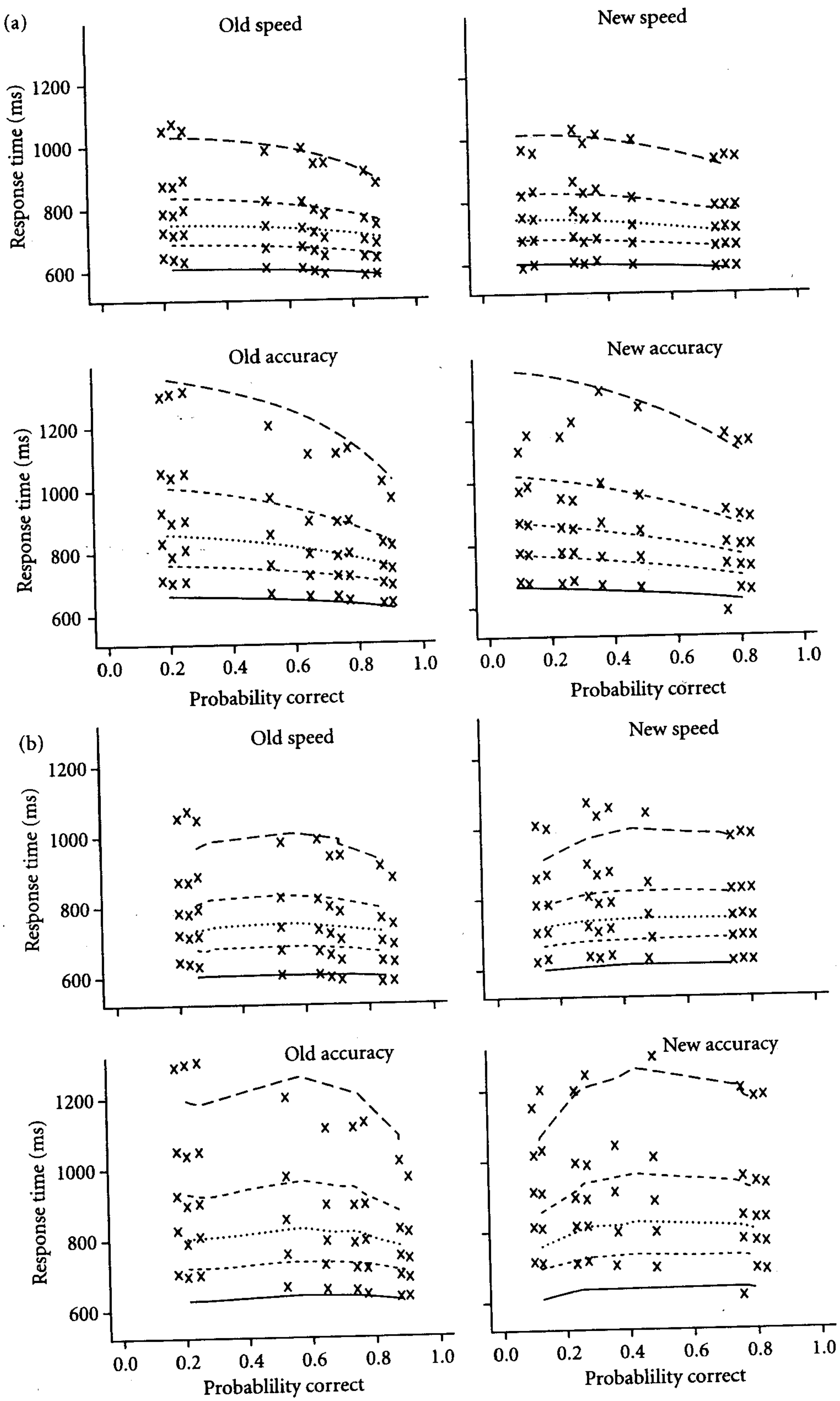


**Fig. 1.5** Sample fits for the accumulator, LCA, and diffusion models to data from the recognition memory experiment with older subjects. $X^2$ values are 60.9, 151.9, and 64.9, respectively. (a) Letter discrimination: young subjects; (b) diffusion model recognition memory older subject data; (c) LCA model recognition memory older subject data.

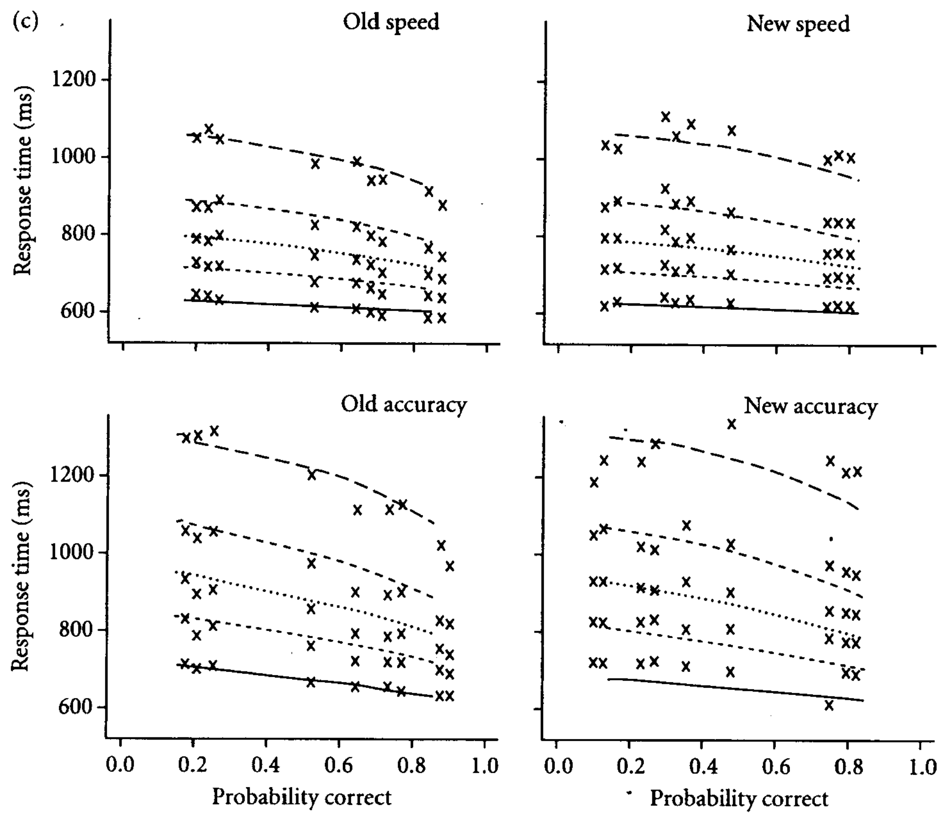


**Fig. 1.5** *(Continued)*

When difficulty is varied across experimental conditions in such a way that subjects cannot know at the time a stimulus is presented which condition it is in (e.g. when high and low frequency words are presented in random order in a lexical decision experiment), there are two important constraints on the models. First, the effects of difficulty are determined by only one parameter, the rate of accumulation of evidence: drift rate in the diffusion model and accumulation rate in the accumulator and LCA models. With only drift or accumulation rate varying, accuracy rates plus mean RTs and RT distributions for both correct and error responses must be explained.

The second constraint is that the shape of the QPFs is determined by only a few parameters of the models. For example, in the diffusion model with starting point equidistant from the two boundaries, the form of the QPF is determined by three parameters: $a$ (boundary separation), $\eta$ (across-trial variability in drift rate), and $s_z$ (across-trial variability in starting point). With starting point half way between the decision boundaries, then: When $\eta$ and $s_z$ are zero, RTs for correct and error responses for each experimental condition are equal and the QPF is symmetric with an inverted U shape. When $\eta$ is high and $s_z$ is low, error responses are slower than correct responses and the QPF has a peak to the left of the 0.5 probability point. When $\eta$ is low and $s_z$ is high, error responses are faster than correct responses and the QPF has a peak to the right of the 0.5 probability point. Thus the shape of the QPF allows the relative speeds of correct and error responses to be determined by visual inspection. The vertical location of the QPF is determined by the nondecision component of response time, $T_{er}$

We chose to evaluate the models against group data obtained by averaging quantiles of the RT distributions and response probabilities across subjects (Ratcliff, 1979; Thomas & Ross, 1980). Fits to individual subjects and fits to quantile-averaged group data exhibit very similar features and the parameter values obtained from group fits are in good agreement with the average parameter values obtained from fits to individual subjects (see Ratcliff, Thapar & McKoon, 2001, 2003, 2004; Ratcliff, Thapar, Gomez & McKoon, 2004; and Thapar et al. 2003, for examples with about 40 subjects per group, and Smith, Ratcliff, & Wolfgang, 2004, with six subjects). Fitting individual subjects would have required months of computer time for the LCA and accumulator models.

We used a minimum $\chi^2$ statistic to assess how well a model fit experimental data. For $N$ observations grouped into six bins between the five quantile RTs and outside the two extreme quantiles, this statistic has the form

$$\chi^2 \quad \Sigma\, N\,(p_i \quad \pi_i)^2\, /\, \pi_i,$$

where $p_i$ is the observed proportion of responses in the $i$th bin, $\pi_i$ is the theoretical (expected) proportion of responses in the $i$th bin, and $N$ is the number of observations per condition.

Because our fits were carried out on group data, obtained by averaging quantiles across subjects, it was not appropriate to weight the observed and predicted proportions in the $\chi^2$ statistic by the total sample size $N$ as is done in the usual Pearson $\chi^2$ test. Instead, we calculated the statistic from the observed and predicted proportions instead of frequencies and multiplied the values by 100 for readability. We use this statistic as a relative rather than absolute measure of fit and denote it by the symbol $X^2$ to emphasize that it is not a proper $\chi^2$ because it has been calculated from quantile-averaged data. In order to compare the models, later we divide the $X^2$ values by the number of experimental conditions in each experiment to provide $X^2$ values that are approximately in the same range across the experiments (e.g. Experiments 5 and 6 had 4 conditions whereas Experiments 3 and 4 had 18 conditions).

The diffusion model was fit to the experimental data by minimizing the $X^2$ value with a general Simplex minimization routine that adjusts the parameters of the model to find the parameters that give the minimum $X^2$ value (see Ratcliff & Tuerlinckx, 2002, for a full description of the methods). The data entered into the minimization routine for each experimental condition were the RTs for each of the five quantiles for correct and error responses and the accuracy values. The quantile RTs and the diffusion model were used to generate the predicted cumulative probability of a response occurring at or before the given quantile RT. Subtracting the cumulative probabilities for each successive quantile from the next higher quantile gives the proportion of responses between each quantile. These expected values are compared to the observed proportions of responses between the quantiles and the observed proportions of responses for each quantile are the proportions of the distribution between successive quantiles (i.e. the proportions between 0, 0.1, 0.3, 0.5, 0.7, 0.9, and 1.0 are 0.1, 0.2, 0.2, 0.2, 0.2, and 0.1).

The expected (theoretical) values of the probabilities were generated from the models in two different ways. For the diffusion model, an explicit expression for the cumulative distribution function is available (Ratcliff, 1978; Ratcliff et al. 1999). This involves an infinite series that must be summed numerically, with numerical integration over the distributions of drift rate, starting point, and the nondecision component of RT that represent across-trial variability in these components of processing. This combination was used to produce the cumulative probabilities at the quantile RTs. From these, the proportions between successive quantiles needed for $X^2$ can be computed.

For the accumulator model, a simulation method was used to compute accuracy and RT distributions. Because we wished to allow independent variability in criteria, it was more efficient to generate predictions by simulation than to use the exact numerical methods used by Ratcliff and Smith (2004) Smith and Vickers (1988). Exact methods require an additional numerical integration for each new source of parameter variability, which appreciably slows program execution. In contrast, simulations make it easy to add sources of across trial variability in parameters of the models because all that is required is for a random value to be selected from the appropriate distribution on each trial. The simulation method allowed us to avoid the need to do extensive checking on the four numerical integrations that are needed to handle variability across trials in accumulation rate, the nondecision component of RT, and the two decision criteria. One hundred thousand simulations of the accumulator process for each experimental condition were used, which provided highly accurate predictions (repeated runs provided the same values of all the quantile RTs except the 0.9 quantile for errors that varied by a few ms from run to run).

No explicit expressions have yet been obtained for the RT distributions predicted by the LCA model when the amount of accumulated evidence is constrained to be positive. Because of this, Usher and McClelland (2001) obtained predictions from the model by simulation and we followed their method. In the fits of the model to the data described here, 20,000 simulations of the decision process per condition were used to compute the accuracy values and the RT distributions for the two responses (fewer simulations were used than for the accumulator model because the fitting program took 4 to 6 h for the larger data sets for one fit).

Each of the models was fit to the data using a Simplex algorithm (Nelder & Mead, 1965) along with a set of starting values that were a reasonable guess based on other fits. The Simplex algorithm is given ranges (we used 0.1 times the parameter values) for each of the starting values and the algorithm evaluates the $\chi^2$ function for a range of values of the parameters based on these initial points and ranges. One hundred iterations of the Simplex algorithm were run and then the final parameters were used as initial points with the same ranges around these values. This was repeated for five sets of runs of the Simplex algorithm with the last set running for 400 iterations. If the fits were moderately poor, new starting values were tried and the fitting procedure run again until different and better fits could not be found.

## Experiments

The data against which the models were tested came from six experiments each with one data set for young subjects and one data set for older subjects, previously published by Ratcliff and colleagues. The experimental tasks, all two-choice tasks, were chosen to allow comparison of the performance of young and older subjects across several different kinds of cognitive processing. In two of the tasks, one with high spatial frequency letters and one with low spatial frequency brightness arrays, the stimuli were masked in order to look at the effect of limited availability of stimulus information. Another of the tasks was recognition memory, chosen to investigate the availability of newly learned information, and another was lexical decision, chosen to investigate the availability of well-known information. For the first five experiments, there was a speed-accuracy manipulation: In half the blocks of trials, subjects were instructed to respond quickly, while in the other half of the trials, subjects were instructed to be as accurate as possible. For all the experiments, sufficient data were collected per subject to provide reliable estimates of the differences in components of processing between young and older subjects. The young subjects were all Bryn Mawr or Northwestern University students and the older subjects were all between the ages of 60 and 75 and they were matched to the young subjects on standard characteristics.

### Experiment 1: Signal detection

In this experiment (Ratcliff et al. 2001), two vertically-aligned dots were displayed on each trial and subjects were asked to decide whether the separation between them was "large" or "small." Stimulus difficulty was varied via the amount of separation: There were 32 possible separations, labeled 1 through 32 with 1 being the smallest separation, ranging from 1.75 cm to 3.33 cm in equal intervals. After each trial, subjects were given feedback such that the response was designated as "correct" or "error." The response classification was probabilistic, so it was not possible for subjects to be perfectly correct. Feedback was determined by a probability associated with each stimulus: For stimuli 1–7, "small" was designated correct with probability 0.999. For stimuli 8–16, "small" was designated correct with probabilities 0.913, 0.888, 0.856, 0.819, 0.774, 0.722, 0.664, 0.601, and 0.534, respectively. For stimuli 26–32, "large" was designated correct with probability 0.999, and for stimuli 25 through 17, "large" was correct with the same probabilities as for "small" for stimuli 8 through 16. Subjects understood that they could not be completely accurate, that for separations in the middle of the range, either response might be designated as correct, and that their task was to give their best judgment. There were 12 blocks of stimuli in each session, with 3 presentations of each of the 32 stimuli in each block. In 6 of the blocks, subjects were given accuracy instructions and in the other 6, they were given speed instructions. Speed versus accuracy instructions alternated between blocks. Subjects were asked either to respond as quickly as possible or to make as few errors as possible. In the speed blocks, responses longer than 700 ms were followed by a "Too slow" message. In the accuracy blocks, "large" responses to stimuli 1–6 and "small" responses to stimuli 26–32 were followed by a "Bad error" message. There were 17 young and 13 older subjects, and each participated in two 45 min sessions.

The data were grouped into four conditions such that high probability responses were grouped together ("large" responses to large separations and "small" responses to small separations) and low probability responses were grouped together ("small" responses to large separations and "large" responses to small separations). Specifically, the stimulus groupings were: "small" responses to distances 1–8 were grouped with "large" responses to distances 21–32; "small" responses to distances 9 and 10 were grouped with "large" responses to distances 19 and 20, "small" responses to distances 11 and 12 were grouped with "large" responses to distances 17 and 18, and "small" responses to distances 13 and 14 were grouped with "large" responses to distances 15 and 16.

## Experiment 2: Letter discrimination with masking

In this experiment (Thapar et al. 2003), subjects were presented on each trial with a letter that was masked after 10, 20, 30, or 40 ms. The task was to decide whether the masked letter was one of two target letters, which were presented in the top corners of the display screen and changed after every block of

96 trials. The mask consisted of a square outline, larger than the letter stimuli, filled with randomly placed horizontal, vertical, and diagonal lines that were different on every trial. Subjects participated in six blocks of speed trials alternating with six blocks of accuracy trials in each session for either two or three sessions each. There were 40 young and 38 older subjects.

For the speed blocks, subjects were instructed to respond as quickly as possible. Responses longer than 650 ms were followed by a "Too slow" message, and responses faster than 250 ms were followed by a "Too fast" message. For the accuracy blocks, subjects were instructed to respond as accurately as possible. Incorrect responses were followed by an "Error" message. No feedback was provided for correct responses.

## Experiment 3: Brightness discrimination with masking

The task in this experiment (Ratcliff, Thapar, & McKoon, 2003) was to decide whether 64 × 64 arrays of black and white pixels were "bright" or "dark." On each trial, an array was presented for 50, 100, or 150 ms, then masked by four different checkerboard patterns, each 64 × 64 pixels, presented sequentially for 17 ms each. There were six brightness conditions, determined by the probability of a pixel being white equal to 0.350, 0.425, 0.475, 0.525, 0.575, or 0.650. Thirty six young and 35 older subjects participated in two or three sessions each, with five blocks of speed trials alternating with five blocks of accuracy trials (144 trials per block) in each session. In accuracy blocks, if a response was an error, the word "Error" was displayed and in speed blocks, there was no accuracy feedback. In speed blocks, if a response was longer than 700 ms, "Too slow" was displayed.

## Experiment 4: Recognition memory

In this experiment (Ratcliff et al. 2004), subjects studied lists of single words. Each list had 9 words presented once and 9 words presented 3 times for a study list of 36 words, presented at a rate of 1 s per word. Immediately following each study list, subjects were presented with 36 single test words, deciding for each word whether it had been in the study list or not (half of the 36 had been in the study list, half had not). Within lists, accuracy was manipulated by varying the number of times a word appeared in the study list (one or three) and by using high, low, and very low frequency words (mean frequency values of 325, 4.4, and 0.37, Kucera & Francis, 1967).

Subjects received alternating speed and accuracy blocks of trials. In accuracy blocks, if a response was incorrect, the word "Error" was displayed before the next test word was presented. In speed blocks, there was no accuracy feedback, and if a response was longer than 800 ms for young subjects and 900 ms for older subjects, "Too slow" was displayed.

In each session, there were 10 blocks of accuracy trials and 10 blocks of speed trials. A minimum of two sessions per subject were used in data analyses.

Thirty nine young adults and 41 older adults participated in the experiment.

### Experiments 5 and 6: Lexical decision

On each trial of these experiments (Ratcliff et al. 2004), a letter string was presented and subjects were asked to judge whether it was a word or a nonword. In each session, there were 70 blocks of 30 trials each, half words and half nonwords. In each block, there were equal proportions of high, low, and very low frequency words, the same pools of words as in Experiment 4. In Experiment 5, the nonwords were constructed from words by randomly replacing all the vowels in each word with other vowels and in Experiment 6, the nonwords were random letter strings. Subjects were instructed to respond quickly and accurately and error feedback was given.

Fifty four young adults and 44 older adults participated in Experiment 5 and 54 young adults and 40 older adults participated in Experiment 6.

## Components of processing

When the models were fit to the data from the six experiments, the diffusion model fit 30% better on average than the accumulator and LCA models, as is shown in the next section. Despite this difference in the qualities of the fits, the differences in performance between the young and older subjects were generally ascribed to the same sources by all three models. The parameter values corresponding to each component of processing are shown in the Tables 1.1–1.6 for the three models. Sample fits for Experiments 2 and 4 are presented in Figures 1.4 and 1.5. Here, the main findings are reviewed.

### Rates of accumulation of information

Figure 1.6 (bottom right panel) shows the differences in average drift and accumulation rates between young and older subjects for the three models for each experiment. The rates of accumulation of evidence for the older subjects were substantially lower than those for the young subjects only for masked letter discrimination, for which the stimuli are high spatial frequency. For all the other tasks, the rates for older subjects were at least as high as those for the young subjects; in other words, the quality of the information they obtained from the stimuli was not measurably worse than for the young subjects. (The

**Table 1.1** Diffusion model parameters for fits to the 12 experiments

| Experiment | a speed | a accuracy | z speed | z accuracy | $T_{er}$ | $\eta$ | $s_z$ | $s_t$ |
|---|---|---|---|---|---|---|---|---|
| SDT young | 0.082 | 0.146 | 0.041 | 0.073 | 305 | 0.151 | 0.004 | 122 |
| SDT older | 0.097 | 0.181 | 0.048 | 0.090 | 358 | 0.151 | 0.017 | 173 |
| Letter young | 0.074 | 0.111 | 0.037 | 0.056 | 337 | 0.119 | 0.004 | 121 |
| Letter older | 0.109 | 0.178 | 0.054 | 0.089 | 403 | 0.244 | 0.009 | 132 |
| Bright young | 0.073 | 0.137 | 0.036 | 0.066 | 409 | 0.142 | 0.044 | 179 |
| Bright older | 0.072 | 0.126 | 0.036 | 0.063 | 456 | 0.167 | 0.040 | 163 |
| Recogn young | 0.076 | 0.140 | 0.040 | 0.065 | 488 | 0.172 | 0.053 | 181 |
| Recogn older | 0.100 | 0.135 | 0.046 | 0.062 | 589 | 0.203 | 0.004 | 213 |
| Lex pseud young | — | 0.126 | — | 0.064 | 439 | 0.101 | 0.062 | 159 |
| Lex pseud older | — | 0.190 | — | 0.096 | 518 | 0.090 | 0.032 | 171 |
| Lex rand young | — | 0.154 | — | 0.086 | 406 | 0.183 | 0.115 | 140 |
| Lex rand older | — | 0.199 | — | 0.108 | 458 | 0.137 | 0.078 | 96 |

**Table 1.2** Drift rates for the diffusion model for the 12 experiments

| Experiment | $v_1$ | $v_2$ | $v_3$ | $v_4$ | $v_5$ | $v_6$ | $v_7$ | $v_8$ | $v_9$ |
|---|---|---|---|---|---|---|---|---|---|
| SDT young | 0.370 | 0.230 | 0.128 | 0.051 | — | — | — | — | — |
| SDT older | 0.397 | 0.217 | 0.123 | 0.048 | — | — | — | — | — |
| Letter young | 0.328 | 0.283 | 0.211 | 0.095 | — | — | — | — | — |
| Letter older | 0.260 | 0.185 | 0.096 | 0.033 | — | — | — | — | — |
| Bright young | 0.322 | 0.205 | 0.089 | 0.373 | 0.248 | 0.098 | 0.375 | 0.265 | 0.083 |
| Bright older | 0.320 | 0.198 | 0.065 | 0.371 | 0.229 | 0.075 | 0.384 | 0.244 | 0.093 |
| Recogn young | −0.185 | −0.240 | −0.256 | 0.003 | 0.122 | 0.161 | 0.168 | 0.304 | 0.398 |
| Recogn older | −0.190 | −0.200 | −0.220 | −0.003 | 0.081 | 0.126 | 0.152 | 0.267 | 0.313 |
| Lex pseud young | 0.442 | 0.249 | 0.161 | −0.231 | — | — | — | — | — |
| Lex pseud older | 0.374 | 0.242 | 0.173 | −0.218 | — | — | — | — | — |
| Lex rand young | 0.573 | 0.454 | 0.396 | −0.409 | — | — | — | — | — |
| Lex rand older | 0.475 | 0.372 | 0.306 | −0.363 | — | — | — | — | — |

*Note*: Drift criteria for the brightness discrimination experiments are for young subejcts: −0.058, 0.004, and 0.048, and for older subjects: −0.039, 0.023, and 0.061. Order of conditions for experiments. Dot separation, extreme to intermediate separation. Letter discrimination: large stimulus presentation to small. Brightness discrimination, intermediate brightness, long to short stimulus presentation; moderate brightness long to short stimulus presentation, extreme brightness long to short stimulus presentation. Recognition memory, H, L, VL frequency new words, H, L, VL once presented words, and H, L, VL three times presented words. Lexical decision, H, L, VL frequency words and nonwords.

**Table 1.3** Accumulator model parameters for fits to the 12 experiments

| Experiment | Crit 1 speed | Crit 2 speed | Crit 1 acc | Crit 2 acc | $T_{er}$ | $\sigma$ | $\lambda$ | $\kappa$ speed | $\kappa$ acc | $s_t$ |
|---|---|---|---|---|---|---|---|---|---|---|
| SDT young | 0.565 | 0.565 | 1.594 | 1.594 | 315 | 0.481 | 0.058 | 0.960 | 2.317 | 90 |
| SDT older | 0.894 | 0.894 | 1.936 | 1.936 | 352 | 0.491 | 0.068 | 1.079 | 3.193 | 78 |
| Letter young | 0.609 | 0.609 | 1.370 | 1.370 | 340 | 0.377 | 0.052 | 0.818 | 1.281 | 83 |
| Letter older | 0.807 | 0.807 | 2.085 | 2.085 | 411 | 0.698 | 0.053 | 1.904 | 3.010 | 65 |
| Bright young | 0.487 | 0.487 | 1.804 | 1.804 | 405 | 0.469 | 0.054 | 0.835 | 1.901 | 164 |
| Bright older | 0.657 | 0.657 | 1.968 | 1.968 | 443 | 0.426 | 0.046 | 1.013 | 1.540 | 129 |
| Recogn young | 0.480 | 0.513 | 1.863 | 1.452 | 475 | 0.583 | 0.092 | 0.314 | 0.698 | 163 |
| Recogn older | 0.978 | 0.630 | 1.600 | 1.010 | 596 | 0.907 | 0.126 | 0.317 | 0.592 | 160 |
| Lex pseud young | — | — | 1.406 | 0.976 | 419 | 0.450 | 0.074 | — | 1.745 | 101 |
| Lex pseud older | — | — | 1.153 | 1.114 | 577 | 0.660 | 0.113 | — | 2.787 | 125 |
| Lex rand young | — | — | 1.450 | 0.413 | 389 | 0.286 | 0.083 | — | 1.502 | 105 |
| Lex rand older | — | — | 1.212 | 0.775 | 510 | 0.298 | 0.073 | — | 3.506 | 76 |

**Table 1.4** Accumulation rates for the accumulator model for the 12 experiments

| Experiment | $v_1$ | $v_2$ | $v_3$ | $v_4$ | $v_5$ | $v_6$ | $v_7$ | $v_8$ | $v_9$ |
|---|---|---|---|---|---|---|---|---|---|
| SDT young | 1.177 | 0.708 | 0.368 | 0.074 | — | — | — | — | — |
| SDT older | 1.258 | 0.690 | 0.377 | 0.002 | — | — | — | — | — |
| Letter young | 0.884 | 0.775 | 0.555 | 0.142 | — | — | — | — | — |
| Letter older | 0.773 | 0.543 | 0.263 | 0.120 | — | — | — | — | — |
| Bright young | 0.926 | 0.546 | 0.193 | 1.020 | 0.712 | 0.260 | 0.981 | 0.695 | 0.229 |
| Bright older | 0.971 | 0.435 | 0.178 | 1.034 | 0.644 | 0.272 | 1.020 | 0.628 | 0.243 |
| Recogn young | −0.581 | −0.760 | −0.820 | 0.092 | 0.345 | 0.474 | 0.534 | 0.917 | 1.054 |
| Recogn older | −0.783 | −0.882 | −0.944 | −0.004 | 0.306 | 0.462 | 0.614 | 1.056 | 1.243 |
| Lex pseud young | 1.441 | 0.953 | 0.626 | −0.772 | — | — | — | — | — |
| Lex pseud older | 2.095 | 1.514 | 1.056 | −1.393 | — | — | — | — | — |
| Lex rand young | 1.460 | 1.188 | 0.997 | −0.813 | — | — | — | — | — |
| Lex rand older | 1.718 | 1.341 | 1.011 | −1.389 | — | — | — | — | — |

*Note*: Accumulation rate criteria for the brightness discrimination experiments are for young subjects: −0.174, 0.012, and 0.134 and for older subjects: 0.010, 0.029, and 0.048. Order of conditions for expts. Dot separation, extreme to intermediate separation. Letter discrimination: large stimulus presentation to small. Brightness discrimination, intermediate brightness, long to short stimulus presentation; moderate brightness long to short stimulus presentation, extreme brightness long to short stimulus presentation. Recognition memory, H, L, VL frequency new words, H, L, VL once presented words, and H, L, VL three times presented words. Lexical decision, H, L, VL frequency words and nonwords.

**Table 1.5** LCA model parameters for fits to the 12 experiments

| Experiment | crit 1 speed | crit 2 speed | crit 1 acc | crit 2 acc | $T_{er}$ | $k$ | $\beta$ | $\sigma$ | $s_z$ | $s_t$ |
|---|---|---|---|---|---|---|---|---|---|---|
| SDT young | 1.642 | 1.642 | 2.562 | 2.562 | 278 | 4.127 | 0.188 | 0.595 | 0.754 | 86 |
| SDT older | 1.806 | 1.806 | 2.745 | 2.745 | 323 | 4.967 | 0.234 | 0.618 | 0.583 | 121 |
| Letter young | 1.431 | 1.431 | 2.137 | 2.137 | 326 | 2.680 | 0.125 | 0.667 | 0.560 | 112 |
| Letter older | 1.726 | 1.726 | 2.245 | 2.245 | 360 | 8.931 | 0.245 | 0.509 | 0.507 | 110 |
| Bright young | 1.073 | 1.073 | 1.709 | 1.709 | 372 | 3.693 | 0.331 | 0.433 | 0.281 | 166 |
| Bright older | 1.091 | 1.091 | 1.638 | 1.638 | 414 | 3.061 | 0.296 | 0.397 | 0.329 | 139 |
| Recogn young | 1.329 | 1.251 | 2.020 | 2.002 | 492 | 3.194 | 0.521 | 0.989 | 0.441 | 197 |
| Recogn older | 1.506 | 1.432 | 1.658 | 1.755 | 611 | 3.808 | 1.008 | 1.106 | 0.008 | 213 |
| Lex pseud young | — | — | 1.807 | 1.896 | 315 | 0.447 | 0.087 | 0.537 | 1.105 | 158 |
| Lex pseud older | — | — | 2.281 | 2.563 | 413 | 0.577 | 0.113 | 0.602 | 1.135 | 177 |
| Lex rand young | — | — | 1.957 | 1.755 | 308 | 0.433 | 0.082 | 0.534 | 0.992 | 76 |
| Lex rand older | — | — | 3.313 | 3.161 | 423 | 0.560 | 0.127 | 0.642 | 1.073 | 92 |

**Table 1.6** Accumulation rates for the LCA model for the 12 experiments

| Experiment | $v_1$ | $v_2$ | $v_3$ | $v_4$ | $v_5$ | $v_6$ | $v_7$ | $v_8$ | $v_9$ |
|---|---|---|---|---|---|---|---|---|---|
| SDT young | 0.884 | 0.746 | 0.623 | 0.550 | — | — | — | — | — |
| SDT older | 0.885 | 0.706 | 0.616 | 0.547 | — | — | — | — | — |
| Letter young | 0.615 | 0.729 | 0.811 | 0.878 | — | — | — | — | — |
| Letter older | 0.523 | 0.573 | 0.641 | 0.694 | — | — | — | — | — |
| Bright young | 0.716 | 0.639 | 0.552 | 0.732 | 0.670 | 0.568 | 0.727 | 0.665 | 0.563 |
| Bright older | 0.741 | 0.640 | 0.552 | 0.757 | 0.666 | 0.557 | 0.751 | 0.669 | 0.564 |
| Recogn young | 0.724 | 0.808 | 0.813 | 0.456 | 0.312 | 0.258 | 0.269 | 0.054 | 0.071 |
| Recogn older | 0.699 | 0.729 | 0.760 | 0.401 | 0.314 | 0.226 | 0.219 | 0.000 | 0.008 |
| Lex pseud young | 0.035 | 0.234 | 0.342 | 0.767 | — | — | — | — | — |
| Lex pseud older | 0.065 | 0.213 | 0.308 | 0.784 | — | — | — | — | — |
| Lex rand young | 0.086 | 0.157 | 0.229 | 0.822 | — | — | — | — | — |
| Lex rand older | 0.140 | 0.232 | 0.298 | 0.784 | — | — | — | — | — |

*Note*: Accumulation rate criteria for the brightness discrimination experiments are for young subjects: −0.041, 0.002, and 0.033 and for older subjects: −0.035, 0.016, and 0.039. Order of conditions for expts. Dot separation, extreme to intermediate separation. Letter discrimination: large stimulus presentation to small. Brightness discrimination, intermediate brightness, long to short stimulus presentation; moderate brightness long to short stimulus presentation, extreme brightness long to short stimulus presentation. Recognition memory, H, L, VL frequency new words, H, L, VL once presented words, and H, L, VL three times presented words. Lexical decision, H, L, VL frequency words and nonwords.

difference for the lexical decision task with random letter strings appears to be large, but this task has very high drift and accumulation rates and their estimates have much higher variability than those for the other tasks.)

For the diffusion model, for the signal detection, masked brightness discrimination, recognition memory, and lexical decision tasks, we were able to perform significance tests between the 4–9 drift rates for young and older subjects. But we could not do this for other parameters because only one value was computed. There were no significant differences in drift rates between the older and young subjects. The difference for masked letter discrimination was significant ($t(3) = 6.84, p < 0.05$ for all $t$-tests reported in this chapter unless otherwise stated). For lexical decision with nonwords random letter strings, accuracy rates were very high so drift rates we high and quite variable.

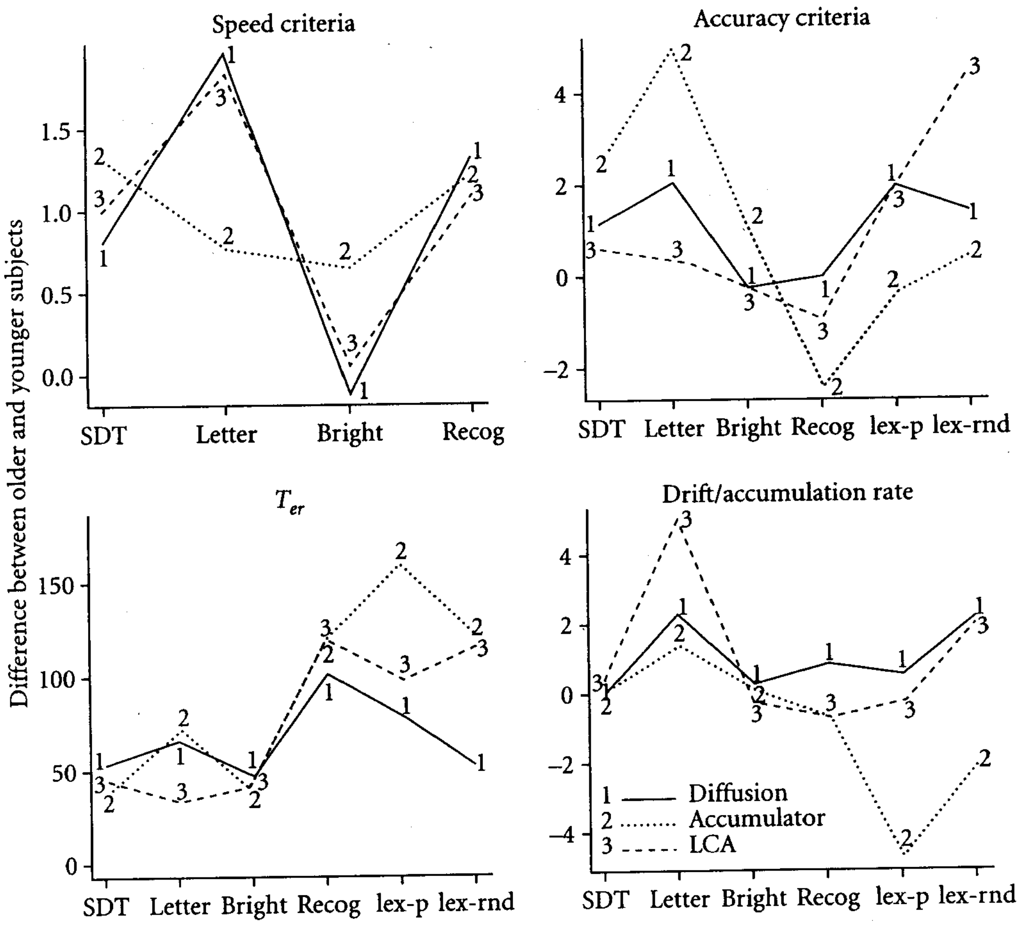


**Fig. 1.6** Differences in parameter values between older and young subjects. SDT represents the signal detection task, "letter" represents the letter discrimination task with masking, "bright" represents the brightness discrimination task with masking, "recog" represents the recognition memory task, and "lex" represents the lexical decision task where "lex-p" is the experiment with pseudoword nonwords and "lex-rnd" is the experiment with random letter string nonwords.

Thus, although the difference was close to two in Figure 1.6, it was not significant. For the LCA model, the results were the same ($t(3) = 7.41$ for the letter discrimination task). For the accumulator model, the results were also the same except that the rate of accumulation was larger in the lexical decision task for the older than the young subjects ($t(7) = 4.82$ for lexical decision, $t(3) = 2.71, p = 0.073$, significant by 1-tailed test for masked letter discrimination). However, the accumulator model did not fit the lexical decision experiments particularly well because it was unable to accommodate error response times shorter than correct response times.

## Response criteria

Figure 1.6 (top two panels) shows the differences in response criteria between older and young subjects with speed instructions in the left-hand panel and accuracy instructions in the right-hand panel. There were both speed and accuracy instructions in the first four experiments. The criteria for the last two experiments, lexical decision experiments, are shown on the right-hand panel along with accuracy criteria for the other experiments because subjects tend to more conservative criteria in the lexical decision task without speed instructions.

For the diffusion model, boundary separation was larger for older than young subjects except with both speed and accuracy instructions in the brightness discrimination task and accuracy instructions in the recognition memory task. For the accumulator model, the decision criteria were higher for older than young subjects with both speed and accuracy instructions in the signal detection, letter discrimination, and brightness discrimination experiments, and with speed instructions in the recognition memory experiment. With accuracy instructions in the recognition memory experiment, the decision criteria were lower for older than young subjects. The differences in criteria were small for the lexical decision experiments, with the older subjects criteria higher in one experiment and lower in the other than the young subjects.

For the LCA model, the decision criteria behaved similarly to the criteria for the diffusion model. The only exceptions were a small difference between the two models for the recognition memory task with accuracy instructions and the much higher criteria for older compared to young subjects in the lexical decision experiment with random letter strings as the nonwords.

## Nondecision component of RT

The bottom left panel of Figure 1.6 shows how much slower older subjects were than young subjects in the nondecision components of RT. The estimates of $T_{er}$ averaged 69, 91, and 78 ms longer for older subjects for the diffusion, accumulator, and LCA models, respectively. The estimate of the nondecision components is a measure of the same quantity for each of the models (unlike the estimates of decision criteria and drift or accumulation rates), and correlations between the models' estimates across the six experiments can show the extent to which a large value of $T_{er}$ for one model corresponds to a large value of $T_{er}$ for another model. Tables 1.1, 1.3, and 1.5 provide the results that were used to compute the correlations. The correlation between the $T_{er}$ values for the diffusion and accumulator models was 0.96, between the diffusion and LCA models 0.88, and between the accumulator and LCA models, 0.84. Correlations were also computed between the ranges of the nondecision component ($s_t$); they were 0.81, 0.77, and 0.59, for the three comparisons, respectively. Larger values of $T_{er}$ would be expected to have larger values of $s_t$ and hence larger correlations, so the correlations between $s_t$ and $T_{er}$ were also computed; they were 0.62, 0.52, and 0.72 for the diffusion, accumulator, and LCA models, respectively. Thus, the models produce qualitatively similar accounts of the duration of the nondecision components of processing (and therefore the duration of the decision process) across experiments.

### Summary

Within the frameworks of the models, the main effects of aging on cognitive processes were that decision criteria were higher for older subjects in most experiments and that the nondecision components of RT were somewhat longer. However, despite the more conservative decision criteria and slower nondecision processes, there was little difference between the older and young subjects in the rate of accumulation of information, that is, little difference in the quality of information available from the stimuli, except with masked letter stimuli that contain high spatial frequency information. Although there were differences in the parameter values that represent the components of processing across the three models, the differences occurred in only a few cases and they were only minor. In general, the behavioral differences in RT and accuracy between young and older adults had the same sources in all three models.

## Goodness of fit

When the models were fit to the experimental data, the chi-square values used as the criteria for goodness of fit used the same weights for each experimental condition because the data were averaged over experiments (see above and Ratcliff & Smith, 2004). The minimum $X^2$ values are shown in Table 1.7 for each model for each experiment. To more easily compare the models across experiments, the $X^2$ values in Table 1.7 were divided by the number of conditions in an experiment so that they were adjusted to approximately the same range (Experiment 1: 8; Experiment 2: 8; Experiment 3: 18; Experiment 4: 18; Experiment 5: 4; Experiment 6: 4) and the results are shown in Figure 1.7. The means of the adjusted $X^2$ values in Figure 1.7 are 2.95, 3.94, and 3.91, for the diffusion, accumulator, and LCA models, respectively. The diffusion model fit 7 of the 12 data sets best, the accumulator model fit 1 best, and the LCA model fit 4 best (see Table 1.7 and Figure 1.7). As mentioned above, no one model fit the data best in every case and every model fit the data best in at least one case.

To show samples of the qualities of the fits, Figure 1.4 shows fits of all three models to the QPFs for young subjects for the letter discrimination experiment and Figure 1.5 shows fits for older subjects for the recognition memory experiment.

For the letter discrimination experiment with speed instructions, each of the models fit well with the accumulator model missing a little in the 0.9 quantile for correct responses and the LCA model missing a little for error RTs for the 0.9 quantile. For accuracy instructions, the diffusion model fit the error RT distributions better than the other models and it fit the RT distributions for correct responses in the 0.9 quantiles a little worse than the other models. The accumulator model missed the 0.9 quantile RTs for both errors and correct responses, and the LCA model fit correct response RTs better than the diffusion model but fit error RTs worse than either of the other models.

For the recognition memory experiment, the diffusion model captures the leading edges of the RT distributions (the 0.1 quantile RTs) and misses only in the 0.9 quantiles for "new" responses. The accumulator model misses all the quantiles for "new" responses with accuracy instructions, especially the 0.1 quantile. The LCA model misses the 0.1 quantiles for "new" responses (and for "old" responses to a lesser extent) with accuracy instructions, and the model

**Table 1.7** $X^2$ values for diffusion, accumulator, and leaky competing accumulator models for signal detection, letter discrimination, brightness discrimination, recognition memory, and lexical decision experiments for older and young subjects

| | SDT young | SDT older | Letter young | Letter older | Bright young | Bright older | Recogn young | Recogn older | Lex pseud young | Lex pseud older | Lex rand young | Lex rand older |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Diffusion | 13.70 | 13.28 | 7.47 | 5.34 | 84.33 | 119.14 | 85.25 | 64.90 | 16.66 | 11.59 | 7.94 | 6.81 |
| Accumulator | 11.31 | 32.02 | 22.38 | 12.99 | 70.97 | 110.41 | 97.76 | 60.92 | 22.64 | 16.21 | 22.38 | 12.74 |
| LCA | 9.25 | 17.09 | 15.70 | 13.59 | 59.98 | 55.12 | 150.44 | 151.90 | 13.91 | 20.09 | 14.15 | 18.96 |

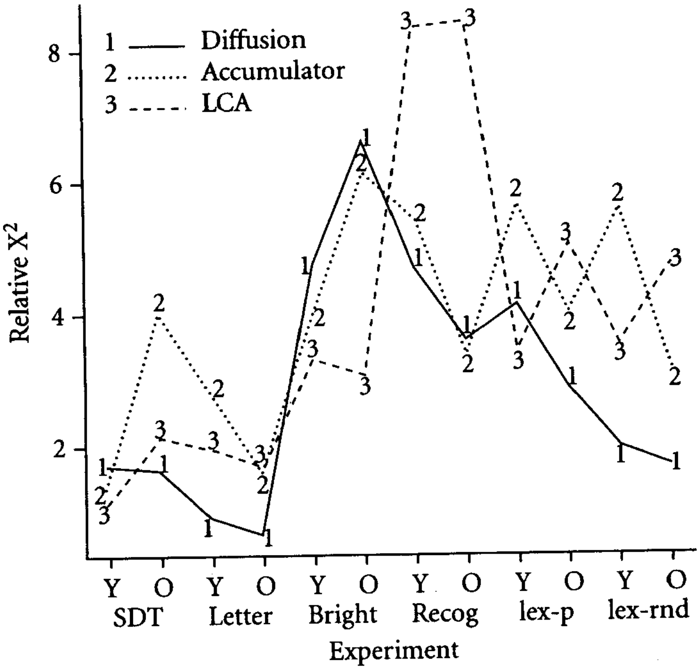


**Fig. 1.7** Relative $X^2$ values for the accumulator, LCA, and diffusion models. The $X^2$ values in Table 7 were divided by the number of conditions in each experiment and plotted as a function of experiment. Y = young subjects, O = older subjects, and SDT represents the signal detection task, "letter" represents the letter discrimination task with masking, "bright" represents the brightness discrimination task with masking, "recog" represents the recognition memory task, and "lex" represents the lexical decision task where "lex-p" is the experiment with pseudoword nonwords and "lex-rnd" is the experiment with random letter string nonwords.

fits the higher quantiles for "new" responses but misses them for "old" responses.

Ratcliff and Smith (2004) provided similar comparisons across three experiments using young subjects only. One was Experiment 1 from this chapter (signal detection for young subjects), the second was a lexical decision experiment with speed and accuracy instructions, and the third was a recognition memory experiment that manipulated the proportion of old and new items. They found similar results: the diffusion model fit somewhat better on average than the accumulator and LCA models.

## General discussion

Theoretical development in the domain of models for two-choice decisions has matured over the last few years. Existing models have been augmented so that they can now overcome some of the seemingly insurmountable theoretical problems of the past. Also, new models have been developed. In parallel with theoretical development, large data sets have become available against which the models can be tested. In this chapter, we applied three models to 12 sets of data from five different tasks, the first comparison of sequential sampling models for two-choice decisions with such a large number of data sets. In past studies, usually only a single model has been evaluated, fit to only one or two sets of data.

What we found is that the diffusion model gave the best overall account of the data. The LCA and accumulator models fit the data about equally well, about 30% worse than the diffusion model (in terms of the $\chi^2$ measure we used). Even though the accumulator model fit about as well as the LCA model, no way has yet been found to allow it to predict RTs for errors shorter than RTs for correct responses. Based on this qualitative problem, it can be rejected. In contrast, even though the LCA model fit worse on average than the diffusion model, there are no qualitative grounds for choosing one over the other.

One concern that might have been raised about the application of the diffusion model to aging data is that any conclusions about the effects of aging on cognitive processes might be specific to the diffusion model. By fitting all three models, we have shown that the general conclusions are the same from all three. For all three models, in five of the six experiments, there was no difference in the rate of accumulation of evidence from the stimuli between young and older subjects. In most cases, the decision criteria were set more conservatively by the older subjects than the young subjects, and this was largely responsible for the longer RTs for the older subjects. There was also an increase in the duration of the nondecision component of processing for older subjects that was quite consistent across the models.

Models for simple two choice decisions have reached a degree of maturity that is rare in modeling in cognitive psychology. The models address most of the phenomena within their domains of study and they fit the phenomena quite accurately. It is now possible to critically test them against each other (Ratcliff & Smith, 2004) and they can be used to interpret the effects of aging, as studied here, and the effects of head injury on cognitive processes (e.g. Ratcliff, Perea, Coleangelo & Buchanan, 2004). In addition to these psychological investigations, the models are also being tested in computational neuroscience as simultaneous accounts of neural processes and behavioral data (e.g. Glimcher, 2003; Gold & Shadlen, 2001; Platt, 2002; Ratcliff, Segraves, & Cherian, 2003; Roitman & Shadlen, 2002; Smith & Ratcliff, 2004). At this point in the evolution of the models, they offer an explanation of processing that encompasses applications to group and individual differences as well as to the neurophysiological underpinnings of cognition.

## References

Ashby, F. G. (1983). A biased random walk model of two choice reaction times. *Journal of Mathematical Psychology*, **27**, 277–97.

Audley, R. J., & Mercer, A. (1968). The relation between decision time and the relative response frequency in a blue-green discrimination. *British Journal of Mathematical and Statistical Psycology*, **21**, 183–92.

Audley, R. J., & Pike, A. R. (1965). Some alternative stochastic models of choice. *The British Journal of Mathematical and Statistical Psychology*, **18**, 207–225.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic–cognitive approach to decision making in an uncertain environment. *Psychological Review*, **100**, 432–459.

Coyne, A. C. (1981). Age difference and practice in forward visual masking. *Journal of Gerontology*, **36**, 730–732.

Diederich, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology*, **41**, 260–274.

Diederich, A. (1997). Dynamic stochastic models for decision making under time constraints. *Journal of Mathematical Psychology*, **41**, 260–274.

Fozard, J. L (1990). Vision and hearing in aging. In J. E. Birren & K. W. Schaie, (Eds.), *Handbook of the psychology of aging* (pp. 150–170). San Diego, CA: Academic Press.

Glimcher, P. W. (2003). The neurobiology of visual-saccadic decision making. *Annual Review of Neuroscience*, **26**, 133–179.

Gold, J. N., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Science*, **4**, 10–16.

Heuer, H. (1987). Visual discrimination and response programming. *Psychological Research*, **49**, 91–98.

Kucera, H., & Francis, W. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

LaBerge, D. A. (1962). A recruitment theory of simple behavior. *Psychometrika*, **27**, 375–396.

Link, S. W. (1975). The relative judgement theory of two choice response time. *Journal of Mathematical Psychology*, **12**, 114–35.

Link, S. W., & Heath, R. A. (1975). A sequential theory of psychological discrimination. *Psychometrika*, **40**, 77–105.

Luce, R. D. (1986). *Response times.* New York: Oxford University Press.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, 7, 308–13.

Owsley, C., Sekuler, R., & Siemsen, D. (1983). Contrast sensitivity through adulthood. *Vision Research*, **23**, 689–699.

Pike, R. (1973). Response latency models for signal detection. *Psychological Review*, **80**, 53–68.

Pike, R., & Ryder, P. (1973). Response latencies in the yes/no detection task: An assessment of two basic models. *Perception & Psychophysics*, **13**, 224–32.

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

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, **86**, 446–461.

Ratcliff, R. (1981). A theory of order relations in perceptual matching. *Psychological Review,* **88**, 552–572.

Ratcliff, R. (1985). Theoretical interpretations of speed and accuracy of positive and negative responses. *Psychological Review,* **92**, 212–225.

Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling the accumulation of partial information. *Psychological Review,* **95**, 238–255.

Ratcliff, R. (2002). A diffusion model account of reaction time and accuracy in a brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin and Review,* **9**, 278–91.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). Diffusion model account of lexical decision. *Psychological Review,* **111**, 159–182.

Ratcliff, R., Perea, M., Coleangelo, A. & Buchanan, L. (2004). A diffusion model account of normal and impaired readers. *Brain & Cognition,* **55**, 374–82.

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

Ratcliff, R., & Rouder, J. F. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance,* **26**, 127–140.

Ratcliff, R., Segraves, M., & Cherian, A. (2003). A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of simple two-choice decisions. *Journal of Neurophysiology,* **90**, 1392–1407.

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

Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on lexical decision. *Psychology and Aging,* **19**, 278–89.

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging,* **16**, 323–341.

Ratcliff, R., Thapar, A., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language,* **50**, 408–424.

Ratcliff, R., Thapar, A., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. *Perception and Psychophysics,* **65**, 523–535.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating the parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin and Review,* **9**, 438–481.

Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review,* **106**, 261–300.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision-making. *Psychological Review,* **108**, 370–392.

Roitman, J. D., & Shadlen, M. N. (2002). Responses of neurons in the lateral interparietal area during a combined visual discrimination reaction time task. *Journal of Neuroscience,* **22**, 9475–9489.

Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review,* **102**, 567–91.

Smith, P. L., & Ratcliff, R. (2004). The psychology and neurobiology of simple decisions. *Trends in Neuroscience,* **27**, 161–168.

Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology,* **32**, 135–168.

Spear, P. D. (1993). Minireview: Neural bases of visual deficits during aging. *Vision Research,* **33**, 2589–2609.

Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging,* **18**, 415–429.

Thomas, E. A. C., & Ross, B. H. (1980). On appropriate procedures for combining probability distributions within the same family. *Journal of Mathematical Psychology,* **21**, 136–152.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review,* **108**, 550–592.

Van Zandt, T., & Ratcliff, R. (1995). Statistical mimicking of reaction time distributions: Mixtures and parameter variability. *Psychonomic Bulletin and Review,* **2**, 20–54.

Vickers, D. (1970). Evidence for an accumulator model of psychophysical discrimination. *Ergonomics,* **13**, 37–58.

Vickers, D. (1978). An adaptive module of simple judgements. In J. Requin (Ed.), *Attention and performance, Part* VII. (pp. 599–618). Hillsdale, NJ: Erlbaum.

Vickers, D. (1979). *Decision processes in visual perception.* New York: Academic Press.

Vickers, D., Caudrey, D., & Willson, R. J. (1971). Discriminating between the frequency of occurrence of two alternative events. *Acta Psychologica,* **35**, 151–172.