

**Aging, Practice, and Perceptual Tasks: A Diffusion Model Analysis**

**Roger Ratcliff, Anjali Thapar, and Gail McKoon**

**The Ohio State University, Bryn Mawr College, and The Ohio State University**

Running Head: Aging, Practice, and Perceptual Tasks

Address correspondence to:

Roger Ratcliff,

Department of Psychology,

Ohio State University,

1827 Neil Ave.,

Columbus, Ohio 43210

Phone number: 614 292 7916

Fax number: 614 688 3984

email [ratcliff.22@osu.edu](mailto:ratcliff.22@osu.edu)

### Abstract

Practice effects were examined in a masked letter discrimination task and a masked brightness discrimination task for college age and 60-75 year old subjects. The diffusion model (Ratcliff, 1978) was fit to the response time and accuracy data and used to extract estimates of components of processing from the data. Relative to young subjects, the older subjects began the experiments with slower and less accurate performance, but across sessions their accuracy improved because the quality of the information upon which their decisions were based improved and this, along with reduced decision criteria, led to shorter response times. For the brightness (but not the letter) discrimination task, the older subjects' performance matched the young subjects' by the end of four sessions, except that their nondecision components of processing were slightly slower. These analyses illustrate how a well-specified model can provide a unified view of multiple aspects of data that are often interpreted separately.

Research that has examined the effects of practice in simple perceptual tasks such as signal-detection, motion and orientation discrimination, letter discrimination, and visual search has found that practice improves performance for both young and older adults (for reviews, see Gibson, 1953, 1969; Kausler, 1994). Some studies report greater levels of improvement in older adults (e.g., Rogers & Fisk, 1991), while others show equivalent levels of improvement in young and older adults (e.g., Ball & Sekuler, 1986; Hertzog, Williams, & Walsh, 1976; McDowd, 1986; Plude et al., 1983; Salthouse & Somberg, 1982). However, currently little is known about the mechanisms that underlie improvement in these tasks as a function of aging (Salthouse & Somberg, 1982; Welford, 1985).

In this article, we use the diffusion model (Ratcliff, 1978, 1981, 1985, 1988; 2002; Ratcliff, Gomez, & McKoon, 2004; Ratcliff & Rouder, 1998, 2000; Ratcliff & Smith, 2004; Ratcliff, Van Zandt, & McKoon, 1999; Smith, 2000, Smith, Ratcliff, & Wolfgang, 2004) to characterize age differences in practice effects in two perceptual tasks: a masked letter discrimination task and a masked brightness discrimination task. The diffusion model is a model of the cognitive processes that underlie decision making in two-choice response time tasks. It separates processing into several components: the quality of the information from the stimulus that is available to the decision system (drift rate), the decision criteria that determine the amounts of information that must be accumulated before a decision can be made, and the nondecision components of processing such as stimulus encoding and response execution. The model has been successfully applied to experimental data from a variety of two-choice tasks, including the masked letter and brightness discrimination tasks used here (Ratcliff & Rouder, 2000; Ratcliff, 2002), and it provides the most comprehensive account of data in its domain currently available (Ratcliff & Rouder, 1998; Ratcliff & Smith, 2004; Ratcliff, Van Zandt, & McKoon, 1999).

In aging research, the model has provided insights into age differences in letter discrimination (Thapar, Ratcliff, & McKoon, 2003) and brightness discrimination (Ratcliff, Thapar, & McKoon, 2003), as well as signal detection-like tasks (Ratcliff, Thapar, & McKoon, 2001), recognition memory (Ratcliff, Thapar, & McKoon, 2004), and lexical decision (Ratcliff, Thapar, Gomez, & McKoon, 2004). By applying the diffusion model to the data, Ratcliff and colleagues have been able to separate the effects of aging on the quality of stimulus information

from the effects of aging on criterion settings and on the non-decision components of processing. In all of the experiments, older subjects were slower than young subjects in the non-decision components of processing (with response time differences between 40 and 100 ms). In most of the experiments, older subjects adopted more conservative decision criteria than young subjects. In all of the experiments except masked letter discrimination, the quality of the information driving the decision process was not significantly different between young and older subjects. The finding that the quality of the information was better for young than older subjects in masked letter discrimination but not masked brightness discrimination is consistent with findings in the psychophysical literature using accuracy and threshold measures (e.g., Elliot, Whitker, & MacVeigh, 1990; Higgins, Jaffee, Caruso, DeMonasterio, 1988; Owsley, Sekuler, & Siemsen, 1983; Spear, 1993). These studies show a deficit as a function of age for high spatial frequency stimuli (e.g., letters), but not for low spatial frequency stimuli (e.g., brightness patches). Together, the patterns of deficits that have been found across the tasks pose a challenge to theories that have postulated some type of general slowing mechanism to account for age-related deficits in cognition (e.g., the generalized slowing hypothesis, Brinley, 1965; Cerella, 1985, 1994; Myerson, Adams, Hale, & Jenkins, 2003; Ratcliff, Spieler, & McKoon, 2000, 2004; Salthouse, 1985). These accounts are hard pressed to explain the similarity in slowing of the non-decisional components of processing observed across tasks coupled with task-specific deficits in the quality of stimulus information and response criteria. In contrast, the diffusion model not only accounts for all the data from the tasks, it provides interpretations of the effects of aging on performance that corroborate the conclusions reached by researchers in the psychophysical literature.

In the experiments in this article, masked brightness discrimination and masked letter discrimination, performance improved with practice for both young and older subjects. The aim was to investigate whether the improvements in performance were due to changes in the criteria that determine the amounts of information that must be accumulated before a decision, changes in the quality of stimulus information driving the decision process, changes in the non-decision components of response time, or some combination of these.

Masked brightness and masked letter discrimination are similar to some of the tasks that have been studied in research on perceptual learning. For example, Fine and Jacobs (2002)

compared the effects of practice on a number of perceptual tasks and found that learning was modest with simple stimuli but more pronounced with complex stimuli. In particular, naming common objects and spatial frequency discrimination (the two tasks most like the letter discrimination and brightness discrimination tasks used here) showed relatively small practice effects. From a theoretical perspective, some research on perceptual learning has focussed on separating out components of processing that might be responsible for improvements in performance (e.g., Doshier & Lu, 2004), but few studies have examined response time and none have used the behavior of both accuracy and response time, as the diffusion model does, to jointly constrain theoretical interpretations of data (in fact, Fine & Jacobs explicitly excluded tasks using response time measures).

In this article, the performance of young and older subjects was compared across three one-hour sessions of masked letter discrimination in Experiment 1 and across four one-hour sessions of masked brightness discrimination in Experiment 2. In the sections below, the two experiments and their data are presented first, then the model is described, and then the interactions of practice effects, age, and the components of processing extracted by the model from the data are discussed.

When Ratcliff et al. (2003) investigated age differences in brightness discrimination, they found that older subjects were slower than young subjects in the nondecision components of response time, but this was the only significant difference in the components of processing between the two groups. Older subjects did not set significantly more conservative decision criteria than young subjects, nor was the stimulus information upon which their decisions were based of significantly lower quality. However, Ratcliff et al. reported only data from experimental sessions for which subjects' performance had stabilized, and in general, this was the first two sessions of performance for young subjects but the second and third or third and fourth sessions for older subjects. The question we take up in this article is how practice contributed to the performance of the older subjects. In other tasks, older subjects often set more conservative decision criteria than young subjects, but they also move their settings significantly depending on whether the task instructions emphasize speed or accuracy. Thus, older subjects might reasonably be expected to move to less conservative settings with practice. Criteria settings are highly correlated with response time (averaged across conditions, Ratcliff, Thapar, Gomez, & McKoon, 2004; Ratcliff,

Thapar, & McKoon, 2003; Thapar et al., 2003), so less conservative settings across sessions would mean shorter response times across sessions. More speculative are hypotheses about the effects of practice on older subjects' nondecision components of processing and on the quality of stimulus information entering the decision process. Practice might speed up some nondecision components of processing such as response execution. It also might-- or might not-- improve subjects' abilities to obtain discriminative information about brightness versus darkness from a stimulus. The quality of the information from a stimulus is correlated with response accuracy, so better information upon which to base decisions would mean more accurate responses. Issues like these about how to isolate the effects of practice on specific, separable components of processing have not previously been addressed in the aging literature.

Thapar et al. (2003) found that masked letter discrimination differs from brightness discrimination in that older subjects were disadvantaged in evidence available from the stimuli and in setting more conservative response criteria, not just in the nondecision components of processing. Like the Ratcliff et al. (2003) study, Thapar et al.'s letter discrimination data were collected after subjects' performance had stabilized, and for older subjects, this was usually after two sessions of practice. Here, we examine performance as it changes across sessions. For the nondecision components of processing and for decision criteria settings, the same expectations can be generated as for brightness discrimination: a possible speed up for nondecision components and possibly less conservative decision criteria. However, the findings in the psychophysical literature that high spatial frequency stimuli are especially difficult for older subjects suggest the hypothesis that practice does not yield improvements in the quality of information on which letter discrimination is based.

### Experiment 1: Letter Discrimination

On each trial, one of two letters was displayed on the screen and then masked and a subject's task was to indicate which letter was presented. Stimulus duration was manipulated such that performance varied from near chance at the shortest duration to near ceiling at the longest duration. Speed instructions alternated with accuracy instructions across blocks of trials. The speed instructions stressed that subjects should respond as quickly as possible, and the accuracy

instructions stressed that they should respond as accurately as possible. The speed-accuracy instructions were included for two reasons: first, to provide more complex patterns of data and therefore more stringent measures of the components of processing identified by the diffusion model; and second, to investigate to what extent and through what components of processing the performance of older subjects can be pushed toward the performance levels of young subjects.

### Method

Subjects. Twenty-seven young subjects and 27 older subjects completed three 1-hour sessions, each receiving a \$45 honorarium for their participation. The young subjects were traditional-aged college students enrolled at Haverford College and Bryn Mawr College, recruited from fliers posted on campus.

The data for the older subjects in the experiment came from the study described by Thapar et al. (2003). The subjects included in the analyses presented here were those for whom all three of their first three sessions of data were available. There were other subjects in the Thapar et al. study for whom the first session had been discarded (it had served as practice). An analysis of the second sessions for the 27 subjects included here and the 11 subjects discarded showed no significant differences between them.

The older subjects were healthy, active, community-dwelling individuals 60 to 75 years old living in the suburbs of Philadelphia and were recruited from advertisements placed in local newspapers. All subjects, both older and young, had to meet the following inclusion criteria to participate in the study: a score of 26 or above on the Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975), and no evidence of disturbances in consciousness, medical or neurological disease causing cognitive impairment, head injury with loss of consciousness, or current psychiatric disorder. Subjects had normal or corrected-to-normal vision (20/30 or better) as measured by a Snellen "E" chart. All subjects completed the Vocabulary subtest and the Picture Completion subtest of the Wechsler Adult Intelligence Scale-III (Wechsler, 1997). Means and standard deviations for the background characteristics are presented in Table 1.

Insert Table 1 about here

Stimuli and Procedure. Stimuli were presented on a PC computer and responses were

collected on the keyboard. The stimuli were white letters displayed in the center of the computer screen against a dark background. Letters were paired so as to be dissimilar to each other. The pairs were F/Q, P/L, W/K, B/N, T/X, and G/R. The same pair was used for all the trials of a block of trials. For each block, the two letters were displayed one to the left of the center of the computer screen and one to the right, and remained on the screen throughout the block. Each test trial began with a + sign fixation point in the center of the screen, displayed for 500 ms, then the target letter was displayed, followed by a variable delay (one of six stimulus durations), and a mask. The mask remained on the screen until the subject made a response. Subjects were instructed to press the '/' key on the keyboard if the right alternative had been presented and the 'Z' key if the left alternative had been presented. Six stimulus durations were used: 10, 20, 30, 40, 50 and 60 ms. The mask consisted of a square outline, larger than the letter stimuli, filled with randomly placed horizontal, vertical, and diagonal lines that varied from trial to trial. Each block consisted of 96 trials. The target letter corresponding to the correct response for each trial was determined randomly with the restriction that each alternative be used equally often. Different test lists were used for each subject within an age group, the same test lists for the young and the older subjects.

Each session consisted of 12 blocks of letter identification trials. There were 6 blocks of trials with speed instructions and 6 blocks of trials with accuracy instructions, with speed and accuracy instructions alternating. 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 displayed for 700 ms, and responses shorter than 250 ms were followed by a "TOO FAST" message displayed for 1500 ms. For the accuracy blocks, subjects were instructed to respond as accurately as possible. Incorrect responses were followed by an "ERROR" message displayed for 300 ms. No feedback was provided for correct responses. Each test block lasted approximately 2 minutes and subjects were encouraged to take brief rest breaks between blocks.

## Results

For the young subjects, correct response times less than 300 ms and greater than 3000 ms were considered outliers and eliminated from analyses; for the older subjects, the cutoffs were 300 ms and 3500 ms. For older subjects, 3.7%, 1.1%, and 0.8% of the data were eliminated for the three sessions (one subject contributed 1.2% of the outliers for the first session). For young subjects,

2.3%, 3.7% (one subject contributed 1.3%), and 2.9% (the same subject contributed 1.2%) were eliminated for the three sessions.

For the first sessions, most of the outliers were fast guesses, identified by response times shorter than 300 ms with accuracy at chance. Over sessions and with instructions, these were largely eliminated for the older subjects, but the young subjects continued to produce 2-4% fast outliers. The fast outliers were due, at least to some extent, to a misunderstanding of the speed instructions; some subjects initially interpreted them as requiring fast guesses (despite instructions to the contrary) and some of these subjects continued to produce fast guesses across all the sessions.

Insert Figures 1 and 2 about here

Figure 1 (left panels) shows the effects of practice on mean response times for correct and error responses and on mean accuracy values. Overall, as expected, young subjects were faster and more accurate than older subjects, and responses were slower and more accurate with accuracy instructions than with speed instructions. Error responses were slower than correct responses and their response times show about the same patterns across conditions as correct responses. The finding of interest is that the performance of older subjects improved more with practice than the performance of young subjects; this is apparent in both response times and accuracy. Averaging over speed and accuracy instructions, older subjects' mean correct response times were 266 ms longer and 14 percent less accurate than those of the young subjects in the first session, and only 138 ms longer and 8 percent less accurate in the third session. The data displayed in Figure 1 are averaged over stimulus duration. Generally, responses were faster and more accurate with longer durations (an effect that was larger for the older subjects), and performance improved with practice more for shorter than longer durations. Analyses of variance of the mean response times and accuracy values are presented in Appendix 1.

To examine the effects of practice on response time distributions, quantile probability functions were generated. A quantile probability function is a plot of the quantiles of the response time distribution for each experimental condition as a function of response probability. In Figure 2, the .1, .3, .5 (median), .7, and .9 quantiles are plotted for each of the six stimulus duration

conditions (for young and older subjects, and for the speed and accuracy conditions, separately). The x's are the data points and the lines are the best fitting functions derived from the diffusion model (discussed in detail later). The data for the left and right hand response alternatives were combined because there were no significant differences between them. The six right hand points for each quantile represent correct responses at each of the six stimulus durations. For example, the upper left panel of Figure 2 shows young subjects' data for the first testing session with speed instructions. The probability of a correct response varies from .97 (at the longest stimulus duration) to .68 (at the shortest stimulus duration). The left-hand points for each quantile represent error responses, with further left points representing errors in the higher accuracy conditions (which correspond to the longer stimulus durations). In many cases for errors, fewer than six stimulus duration conditions are displayed because there were fewer than five errors for many of the subjects and so quantiles could not be computed; these were the cases for which stimulus duration was longest.

The quantile probability functions in Figure 2 show the same trends as the means in Figure 1 for the two groups of subjects, speed and accuracy instructions, and practice. They also show faster and more accurate responses for the longer than the shorter stimulus durations. In addition, the quantile probability functions allow examination of how each of the variables affected response time distributions.

First, for stimulus duration, the longest durations have the fastest correct responses and the highest accuracy (the distributions on the far right of each of the Experiment 1 panels in Figure 2). As duration decreases (moving from the far right toward the center), the leading edges of the response time distributions increase by only a small amount while the tails of the distributions increase by a greater amount. The older subjects' accuracy values generally lie a little closer to .5 probability correct than the young subjects' values, reflecting the lower accuracy of the older subjects. For each stimulus duration condition, error responses are slower than correct responses, but otherwise error responses (the distributions to the left of center) tend to mirror correct responses, with mean response times decreasing as stimulus duration increases mainly because of decreases in tail than because of decreases in leading edge.

Second, the differences in mean response times between young and older subjects come

mostly from larger values in the tails for the older subjects' distributions; the leading edges of the older and young subjects' distributions differ by only about 100 ms, on average, while the mean response times differ by about 200 ms. Similarly, the longer mean response times for conditions with accuracy instructions than for conditions with speed instructions come mostly from increases in the tails.

Third, the large speedups in the older subjects' response times from the first to the second and third sessions come mainly from decreases in the tails of their response time distributions. The changes in the leading edges of their distributions are much smaller than the changes in the tails. The young subjects' response times change little across sessions, but their increases in accuracy are shown by the shifts of the distributions toward higher probabilities of correct responses and lower probabilities of error responses from the first to the second and third sessions. The older subjects' distributions also shift toward higher probabilities of correct responses and lower probabilities of error responses across sessions, although the effect is not as obvious in the figure because their accuracy rates are lower.

The quantile probability plots show the complete patterns of data that the diffusion model must explain: accuracy values, correct and error response times, and the shapes of the response time distributions across all the conditions of stimulus duration, instructions, practice, and age group. The task for the model is to identify the components of processing responsible for the effects of each independent variable. For example, the model should explain what components of processing are responsible for the longer response times and higher error rates for older subjects than young subjects, and why there is a larger decrease in response times for the older relative to the young subjects across sessions (differences from about 100-200 ms for older subjects compared to about 20-50 ms for young subjects) but only a relatively small increase in accuracy (differences from about 6-12 percent correct compared to about 1-6 percent correct). We hold off on reporting the results from the fits of the diffusion model until after Experiment 2.

### Experiment 2: Brightness Discrimination

The stimuli were squares of 64x64 pixels presented on a computer monitor. The difficulty of the stimuli was varied by stimulus duration and brightness; brightness was manipulated by

varying the proportions of white versus black pixels. On each trial, a square was presented and then masked, and the subjects' task was to decide whether the square was "bright" or "dark." As in Experiment 1, speed instructions alternated across blocks of trials with accuracy instructions. For each subject, four sessions of data were collected.

### Method

Subjects. There were 26 older subjects and 24 young subjects recruited in the same manner as described for Experiment 1. All subjects were paid \$15 per session for four sessions and all had to meet the inclusion criteria described in Experiment 1. Means and standard deviations for the background characteristics are presented in Table 1.

The data for the older subjects came from the study described by Ratcliff et al. (2003). Data for seven of the subjects from the earlier study were not included because they were unavailable to complete four sessions, and data from two other subjects were not included here because substantial proportions of their data were classified as outlier response times.

Stimuli and Procedure. Stimuli were presented on a PC computer and responses were collected on the keyboard. The 64x64 squares of black and white pixels were presented on a grey background (the whole display was 320x200 pixels). There were six levels of brightness, achieved with six values for the probability of a pixel being white: .350, .425, .475, .525, .575, and .650. These were crossed with three stimulus durations: 50, 100, and 150 ms. Four checker board patterns, each 64x64 pixels, were used to mask each stimulus; presented sequentially, they were: a checker board with 2x2 black and white squares, a checker board the same as the first but with the black and white squares reversed, a checker board with 3x3 black and white squares, and its reverse. The checker boards were designed to mask both smaller and larger random features of a stimulus that might have remained visible through only one or two of the masks. The smaller checker board seemed to eliminate the smaller random patterns in a stimulus and the larger checker board seemed to eliminate the larger random patterns.

Each trial began with a + sign fixation point presented for 250 ms. Then the stimulus was displayed, followed by the four checkerboard masks displayed for 17 ms each. Then a grey background was presented until a response was made. In accuracy blocks, if a response was

correct, there was a 500 ms pause and then the next trial; if a response was incorrect, the word "ERROR" was displayed for 300 ms, then erased, followed by a 500 ms pause before the next trial. Responses of "bright" to stimuli with fewer than .5 white pixels and responses of "dark" to stimuli with more than .5 white pixels were defined as errors. In speed blocks, there was no accuracy feedback. If a response was shorter than 250 ms, a message "TOO FAST" was displayed; if a response was longer than 700 ms, "TOO SLOW" was displayed. Then there was a 500 ms pause before the next trial.

In each session, five blocks of accuracy trials alternated with five blocks of speed trials, with 144 trials per block presented in random order. There were a total of 40 trials for each brightness, duration, and speed versus accuracy condition in each session. Subjects were encouraged to take brief rest breaks between blocks.

## Results

For the young subjects, response times less than 280 ms and greater than 3000 ms were considered outliers and eliminated. This corresponded to 2.2%, 1.0%, 0.7%, and 0.6% of the data for sessions 1 through 4, respectively. For the older subjects, the cutoffs were 280 ms and 3500 ms and this corresponded to 1.8%, 0.6%, 0.4%, and 0.1% of the data for sessions 1 through 4, respectively. As with Experiment 1, most of the outliers were fast guesses, and these were gradually eliminated for most subjects across sessions.

Figure 1 (right panels) shows performance across speed and accuracy instructions and sessions for the older and young subjects. Analyses of variance for the data are reported in Appendix 1. Responses were faster and less accurate with speed than accuracy instructions. Error responses were slower than correct responses, more so for the older subjects than the young subjects. Older subjects improved across the test sessions in speed (from an average of 717 ms to 603 ms) and accuracy (from .71 to .82), while young subjects improved mainly in accuracy (from .75 to .83) and less in speed (from 624 ms to 586 ms). By the last test session, the older subjects' performance closely matched the young subjects' except that correct response times were about 50 ms longer with speed instructions. The data presented in the figure are averaged over the stimulus brightness and duration variables. Performance was better with the stimuli that were easier to

classify as "bright" or "dark", and the effect of practice was to improve performance for the darker stimuli more than the lighter stimuli.

In Figure 3, the .1, .3, .5, .7, and .9 quantiles are plotted for young and older subjects and for speed and accuracy instructions. The x's are the data points and the lines are the best fitting functions from the diffusion model. Eighteen quantiles (six brightness values by three durations) are plotted for correct responses in each panel of Figure 3, but there are fewer than 18 for error responses because many subjects had fewer than five errors in the conditions with highest accuracy (very bright or very dark stimuli), especially, for example, young subjects with accuracy instructions.

Insert Figure 3 about here

The effects of the independent variables on the shapes of the response time distributions were similar to those in Experiment 1. For correct responses, as the difficulty of the stimuli increased across the brightness and duration conditions, the slowdown in the leading edges of the distributions was small whereas the increase in the tails of the distributions was much larger. As in Experiment 1, the effects of instructions and practice were larger on the tails of the response time distributions than on the leading edges.

### The Diffusion Model

The diffusion model is a model of the processes involved in making simple two-choice decisions. The model is designed to apply only to relatively fast two-choice decisions that are composed of a single-stage decision process (as opposed to the multiple-stage process that might be involved in reasoning tasks or card sorting tasks). The model assumes that decisions are made by a noisy process that accumulates information over time from a starting point toward one of two response boundaries, as in Figure 4, where the starting point is labeled  $z$  and the upper and lower boundaries are labeled  $a$  and  $0$ , respectively.

Insert Figure 4 about here

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. For example, if the letter A was

displayed for a long time prior to masking, information quality would be good and the mean value of the drift rate toward the boundary for an "A" response would be large. Within each trial, there is noise (variability) in the process of accumulating information so that processes with the same mean drift rate do not always terminate at the same time (producing response time distributions) and do not always terminate at the same boundary (producing errors). This source of variability is called within trial variability. The bottom panel in Figure 4 shows two processes with the same mean drift rate toward the top boundary. One terminates at the correct boundary, the other terminates at the incorrect boundary. Besides the decision process, there are non-decision components of processing such as encoding and response execution. These are combined in the diffusion model into one component with mean  $\underline{T}_{er}$  and range  $\underline{s}_t$  (shown in the top panel of Figure 4).

In the experiments presented in this article, subjects are instructed either to respond as quickly as possible or 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.

Empirical response time distributions are positively skewed and spread out more as drift rate increases. The diffusion model naturally predicts this by simple geometry; equal size decreases in the rate of approach to the boundary lead to increases in response time smaller for faster processes than for slower processes.

In the diffusion model, variability across trials in the components of processing is explicitly represented (Ratcliff, 1978, 1981; Ratcliff & Rouder, 1998; Ratcliff et al., 1999). Across trial variability in drift rate is assumed to be normally distributed with standard deviation  $\eta$ . Across trial variability in starting point is assumed to have a uniform distribution with range  $\underline{s}_z$ . The across trial variabilities in drift rate and starting point, in conjunction with boundary positions and drift rates, determine the relative speeds of correct versus error responses. If variability in drift rate dominates, then errors are slower than correct responses, and if variability in starting point dominates, then errors are faster than correct responses. There is also variability across trials in the nondecision component of processing, which is assumed to have a uniform distribution with mean  $\underline{T}_{er}$  and

range  $s_t$  (Ratcliff, Gomez, & McKoon, 2004; Ratcliff & Tuerlinckx, 2002). The effect of this variability depends on drift rate; with a large value of drift rate, variability in the nondecision component of processing can shift the leading edge of the response time distribution shorter than it would otherwise be (by as much as 10% of  $s_t$ ; Ratcliff & Tuerlinckx, 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,  $s_t$  is the variability in the starting point,  $a$  is the upper boundary, the lower boundary is set to 0,  $\eta$  is the variability in the mean drift rate across trials,  $T_{er}$  is the mean value of the non-decision components of processing, and  $s_t$  is the variability in the non-decision components. For each different 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,  $v$ . Within-trial variability in drift rate,  $s_v$ , 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).

The two experiments presented here use a masking manipulation. This means that stimulus information is nonstationary, that is, it is available for the time for which the stimulus is displayed and then at the mask, it is terminated. Ratcliff and Rouder (2000) examined two classes of models for how drift rate in the diffusion model relates to the availability of information from the stimulus. For one, drift rate is constant (stationary) over time. This could come about if the information upon which drift rate depends is information from the stimulus that is maintained in a short term store. For the other class of models, drift rate is nonstationary. Two nonstationary models were examined, one in which drift rate increased until the onset of the mask and then decreased, and one in which drift rate was constant until mask onset and then fell to zero. The nonstationary models both predict slow error responses, responses much slower than found in experimental data. In contrast, the stationary model fit the data well for the letter discrimination task examined by Ratcliff and Rouder and in a number of other studies (Ratcliff, 2002; Ratcliff & Rouder, 2000; Ratcliff & Smith, 2004; Ratcliff, Thapar & McKoon, 2003; Smith, Ratcliff, & Wolfgang, 2004; Thapar, Ratcliff, & McKoon, 2003), so we assume stationarity of stimulus information for the

current studies.

It should be stressed that the two experimental manipulations, speed-accuracy instructions and difficulty (presentation duration in the letter discrimination task and presentation duration and brightness in the brightness discrimination task), are modeled by changes in one and only one parameter each. For instructions, the diffusion model has to account for the changes in accuracy and response time distribution shape (the relative shifts and spreads in the distributions) for both correct and error responses with only the separation between the boundaries varying. Similarly, changes in drift rate alone must account for changes in accuracy and response time distributions for both correct and error responses as a function of difficulty. Both of these manipulations produce changes that involve many degrees of freedom, and the model must account for these with changes in just one degree of freedom.

### Fitting the Diffusion Model to Data

The diffusion model is fit to data under several general assumptions about which components of processing can change across levels of an independent variable and which components cannot change. First, as just discussed, subjects can adjust their response criteria according to whether instructions emphasize speed or accuracy, but they are assumed not to change drift rates. Drift rates are affected only by the quality of the information from a stimulus. In other words, the values of drift rate can increase with stimulus duration but they are kept constant across the two types of instructions. Second, when subjects adjust response criteria, they can separately adjust the distances from the starting point to the two criteria. However, for the data presented here, the model fit well with the two distances set equal to each other (i.e., the amount of evidence needed to make one response was the same as the amount of evidence needed to make the other response). Third, it is assumed that  $\underline{T}_{er}$ , the mean value of the nondecision components of processing, is constant across levels of stimulus difficulty and instructions. Although small differences in  $\underline{T}_{er}$  might reasonably be assumed between, for example, speed versus accuracy instructions, little would be added to the quality of the fits of the model to data (except a somewhat better fit to the .9 quantile response times in the conditions with accuracy instructions) and nothing significant would change in interpretations of the data. Finally, for the fits presented here, it was

assumed that all of the across trial variability parameters were constant across levels of stimulus difficulty and instructions.

Under these assumptions, the model must capture all the trends in the data for mean response times for correct and error responses, for the shapes of the response time distributions, and for the accuracy values-- in other words, the data displayed in the quantile probability functions. The structure of the model places strong constraints on how this can be achieved. In the model,  $T_{er}$  determines the placement of the quantile probability functions vertically, that is, on the response time axis. The shapes of the quantile probability functions are determined by just three parameters: the distance between the two response criteria  $a$ , the standard deviation in drift across trials  $\eta$ , and the range of the starting point across trials  $s_z$ . The drift rates for the different levels of stimulus difficulty sweep out the function across response probabilities, with the parameter  $a$  being the main determinant of the spread of the response time distribution at each level of stimulus difficulty.

For the data from Experiments 1 and 2, short and long outlier response times were eliminated from the analyses, as described above in the Results sections. Ratcliff and Tuerlinckx (2002) showed that the influences of further contaminant responses (e.g., from momentary lapses of attention) can be excluded by including a parameter to represent them in the model.  $p_o$  is the probability of a contaminant in each condition of the experiment, where its value comes from a uniform distribution that has maximum and minimum values corresponding to the maximum and minimum response times in the condition. For the experiments reported here, the value of  $p_o$  was assumed to be the same across all experimental conditions (speed and accuracy instructions, level of difficulty of the stimulus, and session) for each subject group. Its values were less than .008 for Experiment 1 and less than .002 for Experiment 2.

The diffusion model was fit to the data by minimizing a chi-square value with a general SIMPLEX minimization routine that adjusts the parameters of the model until it finds the parameter estimates that give the minimum chi-square value (see Ratcliff & Tuerlinckx, 2002, for a full description of the method). The data entered into the minimization routine for each

experimental condition were the mean response times over subjects for each of the five quantiles for correct and error responses and the corresponding mean accuracy values. The quantile response times and the diffusion model were used to generate the predicted cumulative probability of a response by that quantile response time. Subtracting the cumulative probabilities for each successive quantile from the next higher quantile gives the proportion of responses between adjacent quantiles. For the chi-square computation, these are the expected values, to be compared to the observed proportions of responses between the quantiles (i.e., the proportions between 0, .1, .3, .5, .7, .9, and 1.0, which are .1, .2, .2, .2, .2, and .1) multiplied by the number of observations. Summing over  $(\text{Observed-Expected})^2/\text{Expected}$  for all conditions gives a single chi-square value to be minimized.

The reason that the model was fit to means across subjects instead of to the data for individual subjects was that there were too few responses per condition. Furthermore, quantile response times could not be computed for some conditions for many subjects, in particular, the quantiles for error responses in the most accurate conditions. For a condition in which accuracy was very high, there were often fewer than the five error responses needed to compute five quantiles. To deal with this problem, cutoffs were set on accuracy values. If the mean accuracy value across subjects for an experimental condition was greater than .9 for the letter discrimination task or greater than .78 for the brightness discrimination task, then quantiles for error responses for that condition were not computed. For these error conditions, a single value (instead of the six values that would be obtained from five quantiles) was added to the chi square computation, i.e., a value of  $(\text{Observed-Expected})^2/\text{Expected}$  based on the proportion of errors. The cutoff for the brightness task was lower than for the letter task because there were fewer observations per condition. The use of these cutoffs explains why quantiles for error responses from high accuracy conditions are not plotted in Figures 2 and 3.

Fitting the model to group averages rather than individual subjects means that the chi square goodness of fit index minimized in fitting is not a meaningful statistic for measuring absolute goodness-of-fit (Ratcliff & Smith, 2004). However, in prior studies in which parameter values obtained from fits to group average data were compared with averages over parameter

values obtained from fits to individual subject data, there was little difference between the two sets of parameter values (Ratcliff et al., 2001, 2003, 2004, 2004; Thapar et al., 2003).

### Experiment 1: Masked Letter Discrimination

Table 2 shows the values of the model's parameters that best fit the data and Figure 5 shows them graphically. These parameter values were used to generate the model's predictions that are shown by the solid lines in Figure 2.

Insert Table 2 and Figure 5 about here

The fits of the model to the quantile probability functions in Figure 2 are generally as good as those reported in the study that previously examined the same task (Thapar et al., 2003). For each group of subjects and each session, the model accurately accounts for the large changes in spread and smaller changes in leading edge of the response time distributions as stimulus difficulty increases and as instructions shift from speed to accuracy. Between groups, the model accurately accounts for the relatively small differences in the leading edges of the response time distributions between the young and older subjects and the much larger differences in the response time distributions' spread. Across sessions, the model fits the much larger changes in performance for the older than the young subjects.

Examining the changes in shape of the response time distributions more specifically (Figure 2 for Experiment 1; see also Figure 3 for Experiment 2), increases in response times from the easier to the more difficult conditions show proportional increases for each of the response time quantiles. That is, if the increase in difference in response times between the first and second quantiles is 20%, then the increase in the difference between each of the other quantiles is also 20%. Plotting these increases against each other would give a straight line function (as in Ratcliff et al., 2000, Figure 12; see also Rabbitt & Banerji, 1989; Smith & Brewer, 1995).

The largest misses for the model are for the data with the most variability: The quantiles for error response times are more variable than for correct responses because there are fewer error responses. The model also tends to miss the data at the higher quantiles, especially the .9 quantiles. These misses possibly represent subjects aborting processing for test items for which response time would become very long (Ratcliff et al., 2003; Thapar et al., 2003). The misses for the higher

quantiles show up where the most variability would be expected: more for the older subjects than the young subjects, and more in the first than the later test sessions.

The interesting questions concern what components of processing change with practice and how these components change differently for older compared to young subjects. The values of the model parameters that represent the various processing components are shown in Table 2 and Figure 5. In Figure 5, the drift rate values are shown for stimulus durations of 10, 20, and 30 ms, and the average of the 40, 50, and 60 ms durations because drift rate values were almost the same for the latter three durations. Figure 5 shows two standard error bars around the parameter values. The standard error bars were derived from Monte Carlo simulations of the model that were used to determine the average variability that would be expected from repetitions of the experiment with the same parameter values (Ratcliff & Tuerlinckx, 2002).

For letter discrimination, Thapar et al. (2003) showed that older subjects set their criteria (a) farther apart than young subjects, that the nondecision components of processing ( $T_{er}$ ) are longer for older subjects, and that drift rates are smaller for older subjects. These findings were replicated here: older subjects set their criteria farther apart by about 40 percent, their value of  $T_{er}$  was larger by about 70 ms, and their drift rate values were about half those of young subjects. However, the older subjects improved more with practice than the young subjects in both response times and accuracy.

In the diffusion model, response time is largely determined by the settings of the response criteria and by the amount of time taken up by the nondecision components of processing. For the findings presented here, the time for the nondecision components changed little across sessions for either group of subjects. In contrast, the distance between the response criteria decreased for both groups of subjects, and it decreased more for the older than the young subjects.

The finding that is perhaps surprising is that the quality of the stimulus information underlying the older subjects' decisions improved with practice. Older subjects' drift rates increased by small amounts for the 10 and 20 ms stimulus durations and by larger amounts for the longer durations. Information quality, that is, drift rate, largely determines accuracy (Ratcliff, Thapar, Gomez, & McKoon, 2004; Ratcliff, Thapar, & McKoon, 2004). Changes in drift rates

accounted for the older subjects' increase in overall accuracy from about .72 to about .80. For the young subjects, drift rates changed almost not at all with practice, except with stimulus durations of 20 ms.

For the across trial variability parameters of the model, there were no large differences between the young and older subjects and no systematic changes across sessions for either group, although variability in the nondecision components of processing in early sessions is somewhat larger for the older subjects. However, these findings for the across trial variability parameters are qualified by the large amounts of variability in their values (see Ratcliff & Tuerlinckx, 2002).

### Experiment 2: Masked Brightness Discrimination

Table 3 shows the values of the best fitting parameters for the model, Figure 3 shows the predictions from them for the quantile probability functions, and Figure 6 show the parameters graphically (with two standard error bars). For Figure 6, "bright" responses to bright stimuli and "dark" responses to dark stimuli were combined, and also the data were collapsed across stimulus duration conditions because duration had a small effect on performance while brightness had a large effect. The fits of the model to the quantile probability functions are generally good (Figure 3), as good as those in Ratcliff et al. (2003), with the theoretical functions intersecting the data points for most of the predicted quantile functions. However, the same qualifications apply as for the letter discrimination data: The model misses where the data are more variable; specifically there is more variability for the longer than the shorter quantiles, for errors than correct responses, and for older than young subjects.

Insert Table 3 and Figure 6 about here

In tasks like the brightness discrimination task in Experiment 2, subjects often have a slight bias toward one or the other of the response alternatives. In other words, subjects put a criterion on the brightness dimension to determine which stimuli will be called "bright" and which will be called "dark." For example, a subject might make a "bright" response to all stimuli for which the proportion of white pixels was above .55, and "dark" to all the other stimuli. In the diffusion model, this criterion is implemented as a bias on drift rates (see Ratcliff, 1985, 2002; Ratcliff et al., 1999). For modeling the brightness data presented here, the amount of bias was allowed to vary across the

stimulus duration conditions.

For masked brightness discrimination, Ratcliff et al. (2003) found that, after practice, response time differences between older and young subjects were due almost entirely to slower nondecision components of processing. The data from the experiment presented here show that the older subjects' nondecision components of processing averaged about 40 ms longer than those of the young subjects in the first session and about 30 ms longer in the fourth session. With speed instructions, the older subjects decreased their response times from the first to the fourth sessions more than young subjects did, mainly by decreasing their criteria settings more than the young subjects, especially from the first to the second sessions. By the fourth session, their criteria settings were similar to those of the young subjects. With accuracy instructions, criteria settings were more variable and less reliable, as we have found elsewhere (Ratcliff et al., submitted).

Accuracy also improved for the older subjects more than the young subjects with practice, and the component of processing responsible for this was mainly drift rate. The older subjects' drift rates increased significantly across sessions, at least doubling from the first to the fourth sessions. In contrast, the increase in drift rates for young subjects was less than 30% from the first to the fourth sessions.

Finally, for the across trial variability parameters, variability in the nondecision components of processing decreased across all sessions for the older subjects and from the first to the second sessions for the young subjects. The older subjects' variability in drift rates was somewhat larger than that of the young subjects, although just as for Experiment 1, the results for the variability parameters are qualified by the large variability in their estimates.

## General Discussion

Experiments 1 and 2 examined the effects of practice on performance in perceptual tasks for young and older subjects. In particular, by applying a model that estimates the contributions to performance of various components of processing, it was possible to determine which components changed with practice for older compared to young subjects. Much of the model's power to accomplish this derives from the constraints that are placed on it by the requirement that it account for all aspects of accuracy and response time measures.

Our informal observations had suggested that older subjects' performance improved substantially across one to two or three sessions, whereas the performance of young subjects did not. The results of the two studies reported here confirmed these observations. We found that performance improved with practice on both tasks for both young and older subjects, with older adults' performance benefiting more. Our findings are consistent with results reported in the broader literature on skill-acquisition where differences in young and older adults' learning functions have been found in a variety of perceptual and cognitive tasks (e.g., Charness & Campbell, 1988; Fisk & Rogers, 1991; Hertzog, Cooper, & Fisk, 1996; Jenkins & Hoyer, 2000; Rabbitt, 1979; Rogers, Hertzog, & Fisk, 2000; Strayer & Kramer, 1994; Touron, Hoyer, & Cerella, 2001; Verhaeghen & Kliegl, 2000). In our experiments, the numbers of trials per session were large, over 1000, and so subjects had more practice than in other studies. For example, in the influential early study of practice effects by Salthouse and Somberg (1982), there were only about 375 trials per session, and the numbers of trials per session were also lower than ours in the studies reviewed by Fine and Jacobs (2002). Because of these differences, any comparisons between our results and earlier research need to take into account the different numbers of trials per session in the studies.

The experiments presented here allowed analyses of the factors that contribute to older subjects' improvements across sessions in letter and brightness discrimination. First, the main factor contributing to the speed up in their performance was a decrease in the amount of evidence required before making a decision. For both letter and brightness discrimination and with both accuracy and speed instructions, the older subjects' criteria decreased across sessions, to the point where, for brightness discrimination, the values were about the same as for the young subjects. In contrast, for the young subjects, there were smaller decreases in criteria.

Second, the contribution of nondecision components of processing to the older subjects' faster performance in later sessions was negligible. The older subjects' nondecision components of processing were slower than the young subjects', and they stayed slower across sessions (although the older subjects did speed up by about 20 ms in the brightness discrimination task).

Third, the older subjects' decisions were based on better information as practice increased, and this led to higher accuracy. In letter discrimination, drift rates increased across sessions for the

older subjects, although their values for the more difficult conditions were much lower than those of the young subjects. In brightness discrimination, drift rates increased for the older subjects sufficiently to match the young subjects' drift rates in the fourth session. The young subjects showed much smaller increases in drift rates across sessions.

One hypothesis concerning the effects of aging on processing is that aging increases noise (cf., Allen, Madden, Weber & Crozier, 1992; Cremer & Zeef, 1987; Li, 2002; Li, Lindenberger & Sikstrom, 2001; Welford, 1981). In the diffusion model, this would correspond to an increase in the parameter representing variability in drift rate within trials ( $\underline{s}$ ). However, the hypothesis that  $\underline{s}$  increases from the young to the older subjects in the tasks we have studied does not appear to be reasonable. This is because  $\underline{s}$  is a scaling parameter, fixed to the same value (.1) for all groups, conditions, and experiments. If it were fixed at some different value, then all the other parameters of the model would change to compensate. For example, if  $\underline{s}$  were increased, the other parameters would increase by the same ratio (drift rate, boundaries, and variability in these parameters). Thus,  $\underline{s}$  is not identifiable. However, suppose that  $\underline{s}$  is, in fact, larger for older subjects than young subjects. If this were true, that is, if  $\underline{s} = .1$  is, in fact, too small relative to the  $\underline{s} = .1$  for young subjects, then the values of drift rate that have been calculated in fits of the diffusion model to older subjects' data (e.g., Ratcliff et al., 2001; Ratcliff et al., 2003; and the studies reported here) have all also been too small; they should be larger. This is what we find unreasonable-- with  $\underline{s} = .1$ , the drift rates for the older and young subjects were not significantly different in Ratcliff et al. (2001), Ratcliff et al. (2003), or the studies reported here; with  $\underline{s}$  larger than .1 for older subjects, the drift rates for the older subjects would be larger, meaning that subjects would be getting better evidence from the stimulus than young subjects, and this seems unlikely.

An important conclusion to be drawn from the two studies reported here concerns the comparison of response criteria between young and older subjects. It appears that young subjects are more willing to set their response criteria at lower levels than older subjects in response to speed instructions. The accuracy lost by such liberal setting of criteria is small (a few percent) and the gain in speed can be quite large (tens or hundreds of ms). Young adults may either come into an experiment with a much lower concern for accuracy or they may be more willing to switch response strategies than older subjects. For example, work by Touron and colleagues on noun-pair

associative learning (Touron & Hertzog, 2004a, 2004b) and arithmetic problems (Touron, Hoyer, & Cerella, 2004) shows that older adults apply more conservative criteria for selecting strategies, opting for slower, nonoptimal strategies over faster but riskier, memory retrieval strategies. However, the important point is that both young and older subjects can change their criteria such that, at least in some tasks, older subjects can match the speeds of young subjects, although they may require significant amounts of practice to do so. The implication is that performance levels can be compared between young and older subjects only once the malleability of the various components of processing is understood (Touron & Hertzog, 2004b). In the context of a fully specified model, manipulations of such variables as speed versus accuracy instructions and amounts of practice allow examination of the dimensions on which older subjects change their performance relative to young subjects, and the dimensions on which they do not change.

In an applied sense, the tradeoffs between speed, accuracy, and response criteria settings are somewhat surprising. In the diffusion model, criteria settings have relatively small effects on accuracy. Thus, if older subjects set their criteria at smaller values than they otherwise would, even considerably smaller values, the small cost in accuracy might be outweighed by the large gain in speed. For example, in tasks like the perceptual discrimination tasks used for this article, the sacrifice in accuracy might be only 2 to 3 percent with a gain in speed of as much as 100 ms, about 20 percent of total response time. From a practical perspective, older people might benefit more from a large increase in speed of processing relative to a small loss in accuracy.

It should be stressed that the diffusion model is a model of the decision process. It separates the evidence that enters the decision process from the decision process but beyond that, it does not provide an analysis of the processes that produce that evidence. For each condition in an experiment, the model extracts an estimate of the amount of evidence, namely drift rate, from the dependent variables (accuracy, correct and error response times, and their distributions). The drift rates serve as the contact point between the diffusion decision process and models of how perceptual information is extracted from stimuli (e.g., Ratcliff, 1981; Smith et al., 2004) and how information extraction might change with practice.

Other approaches to perceptual learning have separated components of processing by examining the effects of experimental manipulations on accuracy measures alone. For example,

Dosher and Lu (2000a, 2000b, 2004; Lu & Dosher, 2004) propose that perceptual learning involves three different mechanisms: better processing of the stimulus itself (stimulus enhancement, in their terms), increased ability to exclude irrelevant information (exclusion of external noise), and improved setting of thresholds for responding (gain control). Dosher and Lu have been able to explain a variety of perceptual learning effects in terms of only stimulus enhancement and external noise exclusion. To integrate their approach with the diffusion model, and explain response times as well as accuracy, it can be speculated that stimulus enhancement and external noise exclusion affect drift rates whereas gain control affects decision criteria.

For two-choice perceptual tasks, the diffusion model offers a vehicle for combining the effects of the two dependent variables, response time and accuracy, in order to provide a contact point between factors that determine the evidence entering the decision process and the decision process. The greater theoretical leverage that comes from modeling both dependent variables simultaneously can lead to a better understanding of how perceptual processes limit performance, and therefore to a clearer picture of what it is that a model of perceptual processing should explain. The critical contribution of this study is that it demonstrates that the diffusion model can be successfully applied to investigate the effects of practice on a task as well as age differences in practice-related improvements. An important next step will be to isolate the mechanisms underlying the improvements in performance associated with cognitive training programs (Ball et al., 2002; Loewenstein et al., 2004).

Author Note

Preparation of this article was supported by NIA grant R01-AG17083, NIMH grants R37-MH44640 and K05-MH01891 and NIDCD grant R01-DC01240.

## References

- Allen, P. A., Madden, D. J., Weber, T. A., & Crozier, L. C. (1992). Age differences in short-term memory: Organization or internal noise? Journal of Gerontology: Psychological Sciences, *47*, 281-288.
- Ball, K., Birch, D.B., Helmers, K.F., Jobe, J.B., Leveck, M.D., Marsiske., et al. (2002). Effects of cognitive training interventions with older adults. The Journal of the American Medical Association, *288*, 2271-2281.
- Ball, K., & Sekuler, R. (1986). Improving visual perception in older observers. Journal of Gerontology, *41*, 176-182.
- Brinley, J.F. (1965). Cognitive sets, speed and accuracy of performance in the elderly. In A.T. Welford & J.E. Birren (Eds.), Behavior, aging and the nervous system (pp.114-149). Springfield, IL: Thomas.
- Cerella, J. (1985). Information processing rates in the elderly. Psychological Bulletin, *98*, 67-83.
- Cerella, J. (1994). Generalized slowing in Brinley plots. Journal of Gerontology: Psychological Sciences, *49*, 65-71.
- Charness, N. & Campbell, J.I. (1988). Acquiring skill at mental calculation in adulthood: A task decomposition. Journal of Experimental Psychology: General, *117*, 115-129.
- Cremer, J., & Zeef, E. J. (1987). What kind of noise increases with age? Journal of Gerontology, *42*, 515-518.
- Dosher, B.& Lu, Z.-L. (2000a). Mechanisms of perceptual attention in precuing of location. Vision Research, *40*, 1269-1292.
- Dosher, B. & Lu, Z.-L. (2000b). Noise exclusion in spatial attention. Psychological Science, *11*, 139-146.
- Dosher, B., & Lu, Z.-L. (2004). Mechanisms of perceptual learning. In Itti, L. & Rees, G. (Eds.), (pp., 471-476). Neurobiology of Attention. MIT Press.

Elliot, D., Whitker, D., & MacVeigh, D. (1990). Neural contribution to spatiotemporal contrast sensitivity decline in healthy ageing eyes. Vision Research, 30, 541-547.

Fine, I., & Jacobs, R. A. (2002). Comparing perceptual learning tasks: A review. Vision Research, 2, 190-203.

Fisk, A.D. & Rogers, W.A. (1991). Toward an understanding of age-related memory and visual search effects. Journal of Experimental Psychology: General, 120, 131-149.

Folstein, M.F., Folstein, S.E., & McHugh, P.R. (1975) Mini-mental state: A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research, 12, 189-198.

Gibson, E. J., (1953). Improvement in perceptual judgments as a function of controlled practice or training. Psychological Bulletin, 50, 401-431.

Gibson, E. J., (1969). Principles of perceptual learning and development. East Norfolk, CT. US: Appleton Century Crofts.

Hertzog, C., Cooper, B.P. & Fisk, A.D. (1996). Aging and individual differences in the development of skilled memory search performance. Psychology and Aging, 11, 497-520.

Hertzog, C. K, Williams, M. V., Walsh, D. A., (1976). The effect of practice on age differences in central perceptual processing. Journal of Gerontology, 31, 428-433.

Higgins, K. E., Jaffee, M. J., Caruso, R. C., & DeMonasterio, F. M. (1988). Spatial contrast sensitivity: Effects of age, test-retest, and psychophysical method. Journal of the Optical Society of America, 5, 2173-2180.

Jenkins, L. & Hoyer, W.J. (2000). Instance-based automaticity and aging: Acquisition, reacquisition, and long-term retention. Psychology and Aging, 15, 551-565.

Kausler, D.H. (1994). Learning and memory in normal aging. San Diego, CA: Academic Press.

Li, S.-C. (2002). Connecting the many levels and facets of cognitive aging. Current Directions in Psychological Science, 11, 38-43.

Li, S.-C., Lindenberger, U., & Sikstrom, S. (2001). Aging cognition: From neuromodulation to representation. Trends in Cognitive Sciences, 5, 479-486.

Loewenstein, D.A., Acevedo, A., Czaja, S.J., & Duara, R. (2004). Cognitive rehabilitation of mildly impaired patients on cholinesterase inhibitors. American Journal of Geriatric Psychiatry, 12, 395-402.

Lu, Z.-L. & Doshier, B. (2004). Perceptual learning retunes the perceptual template in foveal orientation identification. Journal of Vision, in press.

McDowd, J. M. (1986). The effects of age and extended practice on divided attention performance. Journal of Gerontology, 41, 764-769.

Myerson, J., Adams, D. R., Hale, S., & Jenkins, L. (2003) Analysis of group differences in processing speed: Brinley plots, Q-Q plots, and other conspiracies. Psychonomic Bulletin and Review, 10,

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

Plude, D. J. et al. (1983). Aging and visual search under consistent and varied mapping. Developmental Psychology, 19, 508-512.

Rabbitt, P. (1979). How old and young subjects monitor and control responses for accuracy and speed. British Journal of Psychology, 70, 305-311.

Rabbitt, P., & Banerji, N. (1989). How does very prolonged practice improve decision speed. Journal of Experimental Psychology: General, 118, 338-345.

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

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 two choice brightness discrimination task: Fitting real data and failing to fit fake but plausible data. Psychonomic Bulletin and Review, 9, 278-291.

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of lexical decision task. Psychological Review, 111, 159-182.

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

Ratcliff, R., Spieler, D. & McKoon, G. (2000). Explicitly Modeling the Effects of Aging on Response Time. Psychonomic Bulletin and Review, 7, 1-25.

Ratcliff, R., Spieler, D. & McKoon, G. (2004) Analysis of group differences in processing speed: Where are the models of processing? Psychonomic Bulletin and Review, 11, 755-769.

Ratcliff, R., Thapar, A., Gomez, P. & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical-decision task. Psychology and Aging, 19, 278-289.

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., 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. (2003). A diffusion model analysis of the effects of aging on brightness discrimination. Perception and Psychophysics, 65, 523-535.

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

Rogers, W.A., Hertzog, C. & Fisk, A.D. (2000). An individual differences analysis of ability and strategy influences: Age-related differences in associative learning. Journal of Experimental Psychology: Learning, Memory, and Cognition, 26, 359-394.

Rogers, W. A., & Fisk, A. D. (1991). Are age differences in consistent-mapping visual search due to feature learning or attention training? Psychology and Aging, 6, 542-550.

Salthouse, T.A. (1985). A theory of cognitive aging. Amsterdam: North-Holland.

Salthouse, T. A., & Somberg, B. L. (1982). Skilled performance: Effects of adult age and experience on elementary processes. Journal of Experimental Psychology: General, 111, 176-207.

Smith, G.A., & Brewer, N. (1995). Slowness and Age: Speed-Accuracy Mechanisms. Psychology and Aging, 10, 238-247.

Smith, P.L. (2000). Stochastic dynamic models of response time and accuracy: A foundational primer. Journal of Mathematical Psychology, 44, 408-463.

Smith, P. L., Ratcliff, R., & Wolfgang, B. J. (2004). Attention orienting and the time course of perceptual decisions: response time distributions with masked and unmasked displays. Vision Research, 44, 1297-1320.

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

Strayer, D.L. & Kramer, A.F. (1994). Strategies and automaticity: I. Basic findings and conceptual framework. Journal of Experimental Psychology: Learning, Memory, & Cognition, 20, 318-341.

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.

Touron, D. R., & Hertzog, C. (2004a). Distinguishing age differences in knowledge, strategy use, and confidence during skill acquisition. Psychology and Aging, 19, 452-96466.

Touron, D. R., & Hertzog, C. (2004b). Strategy shift affordance and strategy choice in young and older adults. Memory & Cognition, 32, 298-96310.

Touron, D.R., Hoyer, W.J., & Cerella, J. (2001). Cognitive skill acquisition and transfer in younger and older adults. Psychology and Aging, 16, 555-563.

Touron, D. R., Hoyer, W. J., & Cerella, J. (2004). Cognitive Skill Learning: Age-Related Differences in Strategy Shifts and Speed of Component Operations. Psychology and Aging, 19, 565-580.

Verhaeghen, P., & Kliegl, R. (2000). The effects of learning a new algorithm on asymptotic accuracy and execution speed in old age: A reanalysis. Psychology and Aging, 15, 648-656.

Wechsler, S. (1997). Resultative predicates and control. In Proceedings of the 1997 Texas Linguistics Society Conference, 307-321. Texas Linguistic Forum 38, University of Texas, Austin.

Welford, A.T. (1981). Signal, noise, performance, and age. Human Factors, 23, 97-109.

Welford, A. T. (1985). Practice effects in relation to age: A review and a theory. Developmental Neuropsychology, 1, 173-190.

## Appendix 1

Experiment 1: Letter Discrimination

The experimental design was a 2 x 2 x 6 x 3 (Age x Speed-Accuracy Instruction x Stimulus Duration x Session) mixed-design analysis of variance (ANOVA), with Age as the only between-subjects variable. The dependent measures were response time and response accuracy. An alpha level of .05 is used throughout the paper. Sphericity tests were carried out and where significant, Greenhouse-Geiser adjusted significance levels are reported. Significant results were analyzed with an analysis of simple interactions and post hoc comparisons using either the Tukey HSD test (for equal variance among groups) or the Games-Howell test (for unequal variance).

Response Time. The Age x Instruction x Stimulus Duration x Session mixed-factor ANOVA revealed significant main effects of Age,  $F(1, 52) = 101.34, p < .001$ , Instruction,  $F(1, 52) = 79.59, p < .001$ , Stimulus Duration,  $F(1.35, 70.34) = 77.74, p < .001$ , and Session,  $F(1.51, 78.38) = 40.44, p < .001$ . Young adults responded more quickly than older adults (response times 479 and 666 ms, respectively). Responses to the speed-stress condition were shorter than responses to the accuracy-stress condition (514 and 631 ms, respectively). Performance improved with practice (630, 549 and 539 ms for sessions 1, 2, and 3, respectively). Responses were slower to stimuli presented for shorter durations than to stimuli presented for longer durations (mean response times ranged from 631 ms at the 10 ms SOA to 535 ms at the 60 ms SOA).

As expected, these main effects were qualified by several higher-order interactions. Specifically, young and older adults were differentially influenced by decreasing stimulus duration; older adults' mean response times for stimuli presented at 10 ms increased by 147 ms relative to stimuli presented at 60 ms, whereas young adults' mean response times increased by 47 ms, resulting in a significant Age x Stimulus Duration interaction,  $F(1.35, 70.34) = 21.35, p < .001$ . On average, older adults improved about 158 ms from Session 1 to Session 3, whereas young adults improved about 27 ms, resulting in a significant Age x Session interaction,  $F(1.51, 78.38) = 18.95, p < .001$ . The effects of practice were bigger for stimuli presented at shorter durations than longer durations, resulting in a significant Stimulus Duration x Session interaction,  $F(2.31, 120.20) = 12.36, p < .001$ . Although this was true for both age groups, the effect was more pronounced in

young adults, resulting in a significant Age x Stimulus Duration x Session interaction,  $F(2.31, 120.20) = 4.06, p < .05$ . Moreover, the effect was more pronounced in the accuracy-stress condition, resulting in a significant Instruction x Stimulus Duration x Session interaction,  $F(3.64, 189.31) = 2.67, p < .05$ . Across the three test sessions, for both young and older adults, mean response times in the speed-stress condition were shorter than mean response times in the accuracy-stress condition. This finding was confirmed by the non-significant Age x Instruction, Instruction x Session, and Age x Instruction x Session interactions (all  $ps > .10$ ). Lastly, the four way Age x Instruction x Stimulus Duration x Session interaction was not significant,  $F(3.64, 189.31) = 1.08, p = .37$ .

Response Accuracy. The results of the Age x Instruction x Stimulus Duration x Session mixed-factor ANOVA revealed significant main effects of Age,  $F(1, 52) = 40.88, p < .001$ , Instruction,  $F(1, 52) = 127.30, p < .001$ , Stimulus Duration,  $F(2.05, 106.50) = 387.34, p < .001$ , and Session,  $F(1.72, 89.61) = 14.65, p < .001$ . Young adults were more accurate than older adults (.87 and .76, respectively). Responses to the speed-stress condition were less accurate than responses to the accuracy-stress condition (.80 and .84, respectively). Performance improved with practice (.78, .82 and .85 for sessions 1, 2, and 3, respectively). Responses to stimuli presented for shorter durations were less accurate than responses to stimuli presented for longer durations (accuracy ranged from .63 to .96).

These main effects were qualified by several higher-order interactions. Young and older adults were differentially influenced by decreasing stimulus duration; the range was .95 to .56 in older adults and .96 to .71 in young adults, resulting in a significant Age x Stimulus Duration interaction,  $F(2.05, 106.50) = 30.73, p < .001$ . Older adults improved more than young adults (.10 and .04, respectively), which produced a significant Age x Session interaction,  $F(1.72, 89.61) = 14.65, p < .001$ , and whereas older adults improved at all stimulus durations, improvements in young adults' performance were limited to the shorter stimulus durations, resulting in a significant Age x Stimulus Duration x Session interaction,  $F(5.79, 300.83) = 7.87, p < .001$ . Lastly, the four way Age x Instruction x Stimulus Duration x Session interaction was not significant,  $F(6.61, 343.56) = 1.14, p = .33$ .

### Experiment 2: Brightness Discrimination

The experimental design was a 2 x 2 x 6 x 3 x 4 (Age x Speed-Accuracy Instruction x Brightness x Stimulus Duration x Session) mixed-factor ANOVA, with Age as the between-subjects variable. The dependent measures were response accuracy and response time. In the interest of presentational clarity, we report only the results that are directly relevant to the hypotheses tested in this manuscript, namely the effects of Age and Session and the higher-order interactions that include these variables.

Response time. The results of the omnibus 2 x 2 x 6 x 3 x 4 (Age x Speed-Accuracy Instructions x Brightness x Stimulus Duration x Session) mixed-factor ANOVA revealed significant main effects of Age,  $F(1, 48) = 6.23, p < .05$ , Speed-Accuracy Instructions,  $F(1, 48) = 199.38, p < .001$ , Brightness,  $F(2.28, 109.44) = 83.41, p < .001$ , and Session,  $F(2.50, 120.20) = 14.15, p < .001$ . Young adults responded more quickly than older adults (response times 599 and 644 ms, respectively). Responses to the speed-stress condition were faster than responses to the accuracy-stress condition (response times 538 and 705 ms, respectively). Responses to stimuli that were more easily categorized as bright or dark were faster than responses to intermediate stimuli. Performance improved with practice (672, 621, 602 and 593 ms for sessions 1, 2, 3, and 4, respectively). The main effect of stimulus duration was not significant. Regarding practice effects, improvements in the accuracy-stress condition were more pronounced than in the speed-stress condition,  $F(2.52, 120.88) = 5.45, p = .001$ , and older adults benefited more from practice than young adults,  $F(2.50, 120.20) = 5.03, p < .01$ . However, these effects were qualified by a significant Age x Instruction x Session interaction,  $F(2.52, 120.88) = 3.08, p < .05$ . Follow-up analyses revealed that older adults' performance improved in both the accuracy-stress and the speed-stress conditions, but their improvements were more pronounced in the accuracy-stress condition. In contrast, young adults' performance was stable across the four test sessions. The remaining higher-order interactions involving Age did not approach significance (all  $ps > .10$ ). Lastly, the four-way and the five-way interactions were also not significant (all  $ps > .10$ ).

Response Accuracy. The results of the omnibus 2 x 2 x 6 x 3 x 4 (Age x Speed-Accuracy Instructions x Brightness x Stimulus Duration x Session) mixed-factor ANOVA revealed significant main effects of Speed-Accuracy Instructions,  $F(1, 48) = 93.36, p < .001$ , Brightness,  $F(1.54, 73.87) = 321.32, p < .001$ , Duration,  $F(1.64, 78.79) = 183.29, p < .001$ , and Session,  $F(1.49,$

71.31) = 43.46,  $p < .001$ . Responses to the accuracy-stress condition were more accurate than responses to the speed-stress condition (.81 and .78, respectively). Subjects were more accurate when responding to stimuli that were more easily categorized as bright or dark than they were to intermediate stimuli. Response accuracy increased as stimulus duration increased (.76, .80, .82, for 50 ms, 100 ms, and 150 ms durations, respectively). Performance improved with practice (.73, .81, .82, .83, for sessions 1, 2, 3, and 4, respectively). The main effect of Age (young = .81; old = .78), approached significance,  $F(1,48) = 3.16$ ,  $p = .08$ , and the Age x Speed-Accuracy Instructions interaction was significant,  $F(1,48)=5.04$ ,  $p < .05$ . Young adults were more accurate than older adults in the accuracy-stress condition (young adults: .83, older adults: .80), and although a similar pattern was observed in the speed-stress condition (young adults: .79, older adults: .77), the difference was not reliable. More importantly, young and older adults were not differentially influenced by practice as indicated by the non-significant Age x Session interaction,  $F(1.49,71.31) = 2.47$ ,  $p > .10$ . Follow-up analyses confirmed that both young and older adults' performance showed reliable improvements for stimuli at brightness levels .350, .425, .575, and .650, but not for stimuli at brightness levels .475 and .525.

Lastly, the Instruction x Session, Brightness x Session, Duration x Session, Brightness x Duration, and the Instruction x Brightness x Duration x Session interactions were all significant (all  $ps < .05$ ). Subsequent analysis of these higher-order interactions revealed that improvements in responses to stimuli of different brightness levels varied by stimulus duration and by speed-accuracy instructions. In addition, the main effect of duration varied for stimuli of different brightness levels and for the speed-accuracy conditions.

## Figure Captions

Figure 1: Mean correct response time, accuracy, and mean error response time as a function of speed accuracy condition, subject group (older vs. young), and session for Experiment 1 (letter discrimination) and Experiment 2 (brightness discrimination).

Figure 2: Quantile probability plots for older and young subjects as a function of speed accuracy condition and session for Experiment 1, letter discrimination. The lines represent the theoretical fits of the diffusion model, and the Xs represent the .1, .3, .5, .7, and .9 quantile response times.

Figure 3: Quantile probability plots for older and young subjects as a function of speed accuracy condition and session for Experiment 2, brightness discrimination. The lines represent the theoretical fits of the diffusion model, and the Xs represent the .1, .3, .5, .7, and .9 quantile response times.

Figures 4: An illustration of the diffusion model with two sample paths, and illustrations of the distributions of parameter values across trials.

Figure 5: Parameter estimates for young and older subjects across sessions for the letter discrimination task. The error bars represent 95% confidence intervals derived using Monte Carlo methods. The horizontal bars are lines with slope zero to aid evaluation of differences across sessions. Parameters of the model are  $a_s$ =boundary separation in the speed condition,  $a_a$ =boundary separation in the accuracy condition,  $T_{er}$ =nondecision components of RT,  $\eta$ =SD in drift across trials,  $s_z$ =range of the distribution of starting point ( $z$ ) across trials,  $v$ =drift rate, and  $s_t$ =range of the distribution of nondecision times across trials. Drift rates  $v_1$  through  $v_4$  correspond to stimulus durations of 10 ms, 20 ms, 30 ms, and 40 ms respectively.

Figure 6: Parameter estimates for young and older subjects across sessions for the brightness discrimination task. The error bars represent 95% confidence intervals derived using Monte Carlo methods. The horizontal bars are lines with slope zero to aid evaluation of differences across sessions. Parameters of the model are  $a_s$ =boundary separation in the speed condition,

$a_a$  = boundary separation in the accuracy condition,  $T_{er}$  = nondecision components of RT,  $\eta$  = SD in drift across trials,  $s_z$  = range of the distribution of starting point ( $z$ ) across trials,  $v$  = drift rate, and  $s_t$  = range of the distribution of nondecision times across trials. Drift rates  $v_1$ ,  $v_2$ , and  $v_3$  correspond to stimuli with .525 & .475 pixels, .575 & .425 pixels, and .650 & .350 pixels respectively averaged over stimulus duration.