

The Effects of Aging on Reaction Time in a Signal Detection Task

Roger Ratcliff
Northwestern University

Anjali Thapar
Bryn Mawr College

Gail McKoon
Northwestern University

The effects of aging on response time are examined in 2 simple signal detection tasks with young and older subjects (age 60 years and older). Older subjects were generally slower than young subjects, and standard Brinley plot analyses of response times showed typical results: slopes greater than 1 and (mostly) negative intercepts. R. Ratcliff, D. Spieler, and G. McKoon (2000) showed that the slopes of Brinley plots measure the relative standard deviations of the distributions of response times for older versus young subjects. Applying R. Ratcliff's (1978) diffusion model to fit the response times, their distributions, and response accuracy, it was found that the larger spread in older subjects' response times and their slowness relative to young subjects comes from a 50-ms slowing of the nondecision components of response time and more from conservative settings of response criteria.

In most cognitive tasks, response times slow as age increases. In an effort to understand this slowing, much research has focused on a regularity that emerges when older subjects' response times are plotted against young subjects' response times in what is called a "Brinley" function. The regularity is that the function is always approximately a straight line with a slope in the range of about 1.0 to 2.5, a regularity that has been found across a wide range of experimental paradigms (Brinley, 1965; Cerella, 1985, 1991, 1994; Faust, Balota, Spieler, & Ferraro, 1999; Fisher & Glaser, 1996; Fisk & Fisher, 1994; Hale & Jansen, 1994; Hale, Myerson, & Wagstaff, 1987; Maylor & Rabbitt, 1994; McDowd & Craik, 1988; Myerson & Hale, 1993; Myerson, Hale, Wagstaff, Poon, & Smith, 1990; Myerson, Wagstaff, & Hale, 1994; Nebes & Madden, 1988; Perfect, 1994; Salthouse, 1991, Chap. 8; Salthouse & Somberg, 1982; G. A. Smith, Poon, Hale, & Myerson, 1988; Spieler, Balota, & Faust, 1996).

In most early work the slope of the Brinley function was taken as evidence supporting a general slowing hypothesis. The fact that response times for young subjects can be transformed to response times for older subjects by the multiplicative constant that is the slope of the Brinley plot has been interpreted as showing that the effect of aging on response time is a general slowing that applies indiscriminately to all aspects of cognitive processing. However,

despite this mainstream tradition, there have been a number of calls for a more analytic approach, including questions about the uniformity of slowing effects across various components of cognitive processes and across various cognitive tasks (Allen, Ashcraft, & Weber, 1992; Allen, Madden, Weber, & Groth, 1993; Cerella, 1994; Fisher & Glaser, 1996; Fisk & Fisher, 1994; Hartley, 1992; Hertzog, 1992; Lima, Hale, & Myerson, 1991; Madden, 1989; Madden, Pierce, & Allen, 1992; Myerson, Ferraro, Hale, & Lima, 1992; Myerson et al., 1994; Perfect, 1994).

Recently, Ratcliff, Spieler, and McKoon (2000) presented a theoretical analysis of Brinley plots and showed that their slope is *not* a measure of how much older subjects slow relative to young subjects. Instead, it is a measure of the relative variance in their response times across conditions (so long as the distributions have about the same shape). The slope of a Brinley plot is typically around 1.5 because the standard deviation of the response times for older subjects is about 1.5 times greater than the standard deviation of the response times for young subjects. What this means is that the target for theories of aging effects must be an understanding of the cognitive mechanisms underlying the response times that make up the distribution of response times for a cognitive task. Given an understanding of the mechanisms responsible for response time distributions, we can begin to construct a picture of how the mechanisms might change with age such that the distribution spreads to give greater variance.

With this goal in mind, we conducted two signal detection experiments with young and older subjects and analyzed their data with Ratcliff's diffusion model (Ratcliff, 1978, 1980, 1981, 1985, 1988; Ratcliff & Rouder, 1998, 2000; Ratcliff, Van Zandt, & McKoon, 1999). For a number of cognitive tasks, for young subjects the diffusion model offers an accurate explanation of multiple facets of response time data including the probabilities of correct and error responses and the shapes of the distributions of response times for correct and error responses. The model divides

Roger Ratcliff and Gail McKoon, Department of Psychology, Northwestern University; Anjali Thapar, Department of Psychology, Bryn Mawr College.

Preparation of this article was supported by National Institute on Aging Grant AG17083, National Institute of Mental Health Grant HDMH44640, and National Institute on Deafness and Other Communication Disorders Grant R01-DC01240.

Correspondence concerning this article should be addressed to Roger Ratcliff, Department of Psychology, Northwestern University, Evanston, Illinois 60208. Electronic mail may be sent to r-ratcliff@nwu.edu.

decision processes into several components: the quality of the information from a stimulus that drives the decision process, the variability in the quality of information, the criteria that set boundaries on the amount of information that must be accumulated in order for a decision to be made, and the nondecisional (encoding and response execution) parts of response time. The question for the experiments here was which of these components change with age.

In the sections below, we first review Brinley plots, then present the two experiments. We then describe the diffusion model and finally fits of the model to the experimental data.

Quantile-Quantile Plots

Ratcliff et al. (2000) showed that Brinley plots are what are commonly called in statistics quantile-quantile (Q-Q) plots. A Q-Q plot is a plot of the quantile points of one distribution against the quantile points of another distribution. The quantiles of a distribution are the points that divide the total frequency in the distribution into parts. For example, the median point divides the distribution into halves (.5 quantiles), the three quartile points divide the distribution into quarters (.25 quantiles), and so on. For empirical data, the mean response time for an experimental condition is a point from a distribution of means, and so each mean is a quantile of the distribution. Plotting means for older subjects against means for young subjects in a Brinley function therefore yields a Q-Q plot.

There are three issues that need clarification here. First, what is the distribution that the quantiles come from? In Experiment 1 presented below, 10×10 arrays containing asterisks and spaces were used as stimuli. The number of asterisks can vary from 1 to 100 so there are 100 possible conditions in the experiment; therefore, there are 100 mean reaction times, one for each condition. These can be used to construct a Brinley plot by using the mean reaction time for each condition for older subjects and for young subjects. If an experiment manipulated a continuous variable (e.g., stimulus duration) instead of a discrete one (number of asterisks), then mean response times would come from a continuous distribution of possible conditions. In any experiment the full range of conditions may not be sampled. For example, in one experiment we might sample from a range of stimulus durations that span the range of accuracy values from low to high, and in another experiment we might sample only those durations that produce poor performance. In both cases the conditions sampled would be the same for the older and young subjects, and therefore the quantiles would be the same for the older and young subjects, so long as the scale of difficulty was the same for both groups.

The second issue is how the quantiles are computed. In Q-Q plots, the shortest mean reaction time from older subjects would be plotted against the shortest mean reaction time for young subjects, and the next shortest against the next shortest, and so on. However, in a Brinley plot the mean response times for each condition for older and young groups are plotted against each other, even if the fastest condition for older subjects is not the fastest condition for young subjects. This is because the experimental conditions provide an independent measure of difficulty, which allows us to line up points that correspond between older and young subjects. Then the Brinley plot is a Q-Q plot. In within-task comparisons in a single experiment in which Brinley plots are produced for the

means across conditions, accuracy can be used to check the index of task difficulty. Accuracy should line up in the same ordinal way as response time across conditions for older and young subjects. (An accuracy index would be hazardous to use with between-task experimental designs because the tasks might differ in other ways; accuracy and reaction times might not line up in the same order.)

The third issue is whether the relationship between Brinley plots and Q-Q plots can be carried through to meta-analyses in which results from different experiments are combined (so condition means from different experiments are all plotted together). For a meta-analysis Brinley plot to be accurately called a Q-Q plot, the different conditions in the different experiments would have to correspond to the same relative levels of difficulty for older and young subjects. When similar tasks are combined, then Brinley plots are reasonably good approximations to Q-Q plots. However, as some analyses have shown, slopes of Brinley plots can be different for different subsets of experiments (e.g., perceptual vs. computational, Cerella, 1994; Fisk & Fisher, 1994), and in such cases we would not want to identify Brinley plots with Q-Q plots. However, the mathematics of the Q-Q analysis for slope and intercept presented will often still apply, and the slope of the Brinley plot will be the ratio of standard deviations in the mean reaction times for the two groups (so long as the distributions have approximately the same shape).

Q-Q plot theory provides insights into the regularities that have been observed in Brinley plots. Equation 1 shows the relationship between the quantiles of the response time distributions of older and young subjects. A quantile for older subjects, Q_o , is defined in terms of a quantile for young subjects, Q_y , and the means and standard deviations of the distributions of means across experimental conditions for older and young subjects.

$$Q_o = (\sigma_o/\sigma_y)Q_y + \mu_o - \mu_y(\sigma_o/\sigma_y). \quad (1)$$

This equation holds for a wide range of distributions, for example, normal, logistic, Cauchy, gamma (with a fixed "number" parameter), exponential, Weibull (with fixed exponent), uniform, and ex-Gaussian (with the parameter of the exponential component a constant multiplied by the standard deviation of the normal component; Ratcliff & Murdock, 1976). However, the distributions across conditions for the older and young subjects have to have the same shape. If the distribution for the group with the larger spread is more skewed than that for the group with the smaller spread, an upwardly accelerated Brinley plot is obtained (see Nebes & Maden, 1988).

The equation shows that when the condition means for the older subjects (the Q_o s) and the young subjects (the Q_y s) are plotted against each other, the slope of the function is the ratio of the standard deviations of the distributions, and, so long as the distributions have the same shape, the function is linear. Thus, the slope of a Brinley plot is greater than 1 because older subjects have a greater spread in mean response time across conditions than young subjects. Also, the Q-Q analysis (i.e., Equation 1) explains why, given means and standard deviations with values typical of those found in experimental data, the intercept of the function is usually negative when the slope is greater than 1 (Ratcliff et al., 2000, Table 2) and becomes more negative as the slope increases (cf. Cerella, 1985, 1991; Maylor & Rabbitt, 1994, p. 224). Equation 1 also shows that it is the intercept of the Brinley plot that provides

information about the relative speeds of the older and young subject groups.

The most important point is that because the slope of a Brinley plot is a ratio of standard deviations, it does not measure general slowing. Instead, it is a measure of the relative spreads in the older versus the young subjects' mean response times across experimental conditions. So, for example, mean response time may have a 200-ms range across conditions for young subjects and a 300-ms range for older subjects.

Perhaps the easiest way to see that the intercept carries information about slowing is to consider a standard Brinley plot with slope 1.5 and intercept -200 ms. Suppose that all the older subjects were speeded up by 500 ms so they were faster than the young subjects. This would result in a slope of 1.5 (the same as the first case when older subjects were slower than young subjects) but with an intercept of -700 ms. The reason that the misinterpretation of Brinley plots has not been noticed before is that the means and the standard deviations in the distributions of means across conditions are usually correlated: The slower a group of subjects, the more spread are their mean response times across conditions.

Given a correct understanding of Brinley plots, the aim is to develop models that can explain why variance in response times increases with age. However, Ratcliff et al. (2000) showed that it is relatively easy for models to do this. For example, any model of the sequential sampling class (see Luce, 1986, chaps. 8 and 9) could produce larger variances in older subjects' response times relative to young subjects' response times by assuming lower rates of accumulation of evidence for older subjects or more conservative response criteria for older subjects. Critical tests of models must involve aspects of the data other than the Brinley plot regularities. For the diffusion model, these tests include jointly fitting response time data and accuracy data, explaining how the shapes of the response time distributions change as a function of accuracy, and explaining the relative speeds of correct and error responses.

Diffusion Model

The aim of the experiments described below was to collect data to which an explicit model of processing, Ratcliff's diffusion model, could be fit and then to use the parameters of the model to understand which of the model's components change between the young and the older subjects. The experiments used a relatively simple signal detection paradigm. For young subjects, the diffusion model fits data from this paradigm very well (Ratcliff & Rouder, 1998; Ratcliff et al., 1999), and the fits of the model allow coherent interpretations of processing in terms of the parameters of the model.

The diffusion model describes how stimulus information drives a decision process over time. The key feature of the diffusion model that we planned to exploit is that the model allows for the separation of the information that drives the decision process from the other components of the process. Although the model has not previously been used to study the effects of aging on response time, Ratcliff et al. (2000) illustrated the potential usefulness of the approach.

Ratcliff's diffusion model is a member of the class of sequential-sampling models, which are designed to account for response time and accuracy in experimental paradigms using two-

choice tasks. The diffusion model has done a good job of explaining data for recognition memory tasks (e.g., the Sternberg, continuous memory, prememorized list, and study-test paradigms). With no important modifications, it has also been applied to perceptual matching of letter strings (Ratcliff, 1981), to the varied and consistent mapping procedures with the Sternberg paradigm (Strayer & Kramer, 1994), to several perceptual paradigms (Ratcliff & Rouder, 1998), and to new paradigms such as the speed-accuracy decomposition procedure (Meyer, Irwin, Osman, & Kounios, 1988; Ratcliff, 1988). Models of this class have also recently received strong support from data from single cell recordings in monkeys (Hanes & Schall, 1996).

We postpone a detailed description of the diffusion model until after the experimental data have been presented. To anticipate, the model fit the data quite well and yielded interpretations of the effects of age on processing that are considerably different from the common, general slowing interpretation.

Experiment 1

Experiments 1 and 2 both used simple signal-detection paradigms (e.g., Espinoza-Varas & Watson, 1994; Lee & Janke, 1964; Ratcliff & Rouder, 1998; Ratcliff et al., 1999; Vickers, 1979). In Experiment 1, on each trial an array of asterisks was presented on a computer screen, and a subject was asked to decide whether the number of asterisks presented in the display was "high" or "low" (see Figure 1). The number of asterisks that was presented was chosen from one of two distributions of numbers, a high distribution ($M = 56$) and a low distribution ($M = 38$), each distribution

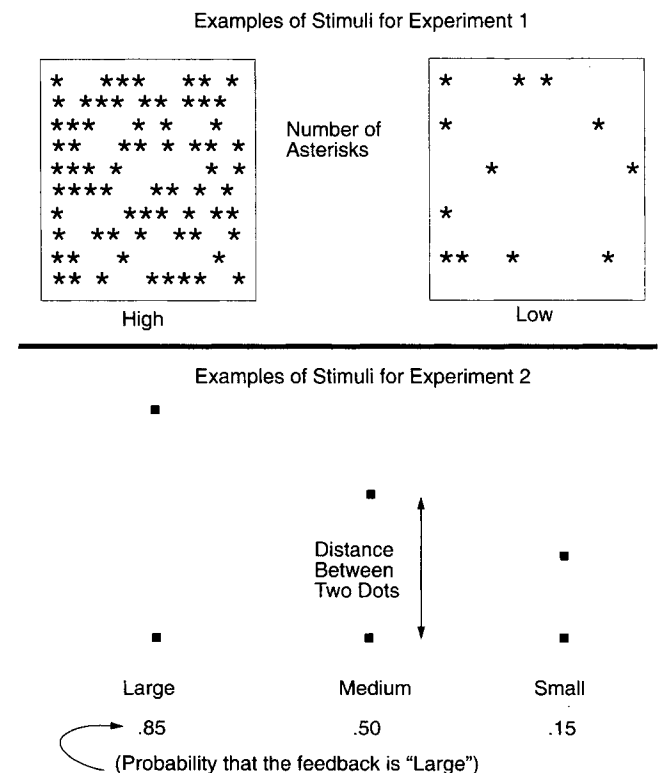


Figure 1. Examples of stimuli for Experiments 1 and 2.

with a fixed mean and standard deviation, and all the numbers were between 0 and 100. Feedback was given after each trial to tell the subject whether his or her response had correctly indicated the distribution from which the stimulus had been chosen. Other than this feedback, the subject had no information about the distributions. The distributions overlapped substantially, so that even after many trials of feedback, the observer could not be highly accurate. A display of 50 asterisks, for example, might have come from the high distribution on one trial and the low distribution on another.

This type of signal detection paradigm has one particular feature that makes it extremely useful. It allows the probabilities of the two responses, "high" and "low," to be varied in small steps from a high probability of one of the responses to a high probability of the other response. For example, a display of 5 asterisks will have a high probability of a "low" response, a display of 90 asterisks will have a high probability of a "high" response, and a display of 50 asterisks will have medium probabilities of both responses. This feature forces a model to account for the relationships between response probabilities and response speeds at all levels of response probability.

Method

Subjects. There were three groups of young adults, one group of 39 adults, one of 33, and one of 36 (15 men and 24 women, 10 men and 23 women, and 11 men and 25 women, respectively). There was also one group of 40 older adults (16 men and 24 women). The young adults were college students who participated in return for course credit in an introductory psychology course at Northwestern University or were paid for their participation at a rate of \$8 per hour. The older adults were healthy, active, community-dwelling individuals aged 60 years or older. The older adults were recruited from advertisements posted at local senior citizen centers, and they received \$15 for their participation. To participate in the study, the older adults had to meet the following inclusion and exclusion criteria: at least 60 years old at the time of entry into the study, a score of 26 or above on the Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975), and a score of 15 or less on the Center for Epidemiological Studies–Depression Scale (CESD; Radloff, 1977). In addition, all subjects completed a medical history form and were excluded if there was evidence of disturbances in consciousness, medical or neurological disease causing cognitive impairment, head injury with loss of consciousness, or current psychiatric disorder. Finally, all participants completed the Picture Completion subtest and the Vocabulary subtest of the Wechsler Adult Intelligence Scale—Revised (Wechsler, 1981). Estimates of full-scale IQ

were derived from the scores of the two subtests (Kaufman, 1990). The means and standard deviations for standard background characteristics are shown in Table 1.

We did not collect the background data presented in Table 1 for the young subjects tested in Experiment 1, though we did collect it for the young subjects in Experiment 2. To provide an additional check on the young subject characteristics, a new group of young subjects was given these tests in the fall quarter (they were performing a memory experiment). They were mainly freshmen and so had a smaller number of years of education, but otherwise their characteristics were similar to those of the young group from Experiment 2 and the older subjects (apart from age).

Subjects in the first group of young adults came from an introductory psychology class subject pool. They were tested at the end of a spring quarter, and they were relatively unmotivated (by our informal assessment, these subjects often produce poor results from psycholinguistic experiments where careful reading is required and often produce large standard deviations in response times). The term "motivated" here is meant to carry only an informal, observational, connotation: If subjects are unmotivated in this sense, at some points in the experiment they are more likely to ignore the stimuli and hit one response key all the time, hit response keys as quickly as possible without processing the stimuli, take long breaks within trials, and so on. Subjects of the second group were also from the introductory psychology subject pool, but they were tested at the beginning of a quarter. In other experiments, these subjects almost always provide good data with small standard deviations in response times. Subjects from the third group were recruited by advertisement and were paid for their participation. Because subjects in this group often want to participate in other experiments for pay, they are usually motivated to produce good data. The older subjects seemed eager to participate and perform well in the experiment. The three different groups of young subjects were tested in an effort to ensure that any differences between older and young were not simply the result of a lack of effort of young subjects in performing the task.

Stimuli. The asterisks were displayed in a 10×10 grid in the upper left corner of a video graphics adaptor (VGA) monitor, subtending a visual angle of 4.30° horizontally and 7.20° vertically. They were clearly visible, light characters presented against a dark background. The VGA monitors were driven by IBM 486-style microcomputers that controlled stimulus presentation time and recorded responses and response times.

The number of asterisks for presentation on a given trial was selected by randomly sampling from one of two discrete, approximately normal distributions with means 38 and 56 and standard deviation 14.4 (following Espinoza-Varas & Watson, 1994). The discriminability (d') between these distributions was therefore approximately 1.25. The probability of sampling each of the two distributions was .5. The feedback on a trial indicated from which of the two distributions the number of asterisks had been

Table 1
Subject Background Characteristics

Measure	Experiment 1: Older adults		Experiment 2: Older adults		Experiment 2: Young adults		Northwestern: Young adults	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
<i>M</i> age	67.7	6.7	68.1	4.4	20.6	1.7	19.1	9.5
Years of education	16.5	2.2	16.5	2.3	14.4	1.1	12.0	0.2
MMSE	28.9	1.3	29.4	0.77	29.5	0.64	28.7	1.4
WAIS-R Vocabulary	14.7	2.6	15.5	2.3	15.5	1.7	14.2	2.1
WAIS-R Picture Completion	12.1	2.2	12.1	1.4	10.9	2.2	10.7	2.4
IQ Estimate	121.9	10.4	122.4	9.6	118.9	7.9	112.8	7.2
CESD: Total	7.6	4.2	7.8	3.8	8.6	3.8		

Note. MMSE = Mini-Mental State Examination; WAIS-R = Wechsler Adult Intelligence Scale—Revised; CESD = Center for Epidemiological Studies–Depression Scale.

selected. This means that feedback to a display with 40 asterisks, for example, would be "low" about twice as often as "high." The two distributions crossed at the number 47; this number is referred to as the *crossover* point for the two distributions. The display positions of the asterisks for a given trial were selected randomly from the possible 100 positions in the 10×10 character grid.

Because there were 100 different possible numbers of asterisks that could be displayed, there were 100 possible different experimental conditions. The number of trials for each of the numbers of asterisks was determined by selecting randomly from the high or the low distribution, and each subject participated in one 30-min session consisting of 550 trials. Therefore, the number of observations for each condition varied from about 2 (for numbers of asterisks less than 5 or greater than 95) to about 25 (for numbers of asterisks around the crossover point 47). In data analyses the 100 conditions were grouped such that each group contained the data for conditions for which responses had about the same response times and probabilities. For one group there were about 350 data points, and for the other groups there were between 80 and 180 data points.

Procedure. Subjects were instructed that the number of asterisks on each trial was selected at random from one of two groups of numbers, a "low" group and a "high" group, and that the low group had fewer asterisks on average than the high group. A subject's task was to decide whether the number of asterisks presented came from the low group, in which case they were to press the "Z" key on the computer keyboard, or the high group, in which case they were to press the "?" key. If a response did not match the distribution from which the stimulus was generated, an error message was displayed immediately after the response. The subjects understood that they could not be completely accurate, that numbers from the middle of the range (e.g., 50) could have come from either distribution, and that their task was to give their best judgment. It is important to note that, for the subjects, an error was defined according to the distribution from which the stimulus was chosen. So a high response to 47 asterisks was correct half the time and was an error half the time. Subjects were fully informed about this ambiguity, using X-ray diagnosis as an illustration of the ambiguity in signal-detection tasks of this kind.

A trial began with the presentation of a 10×10 grid with asterisks. The asterisks remained on the screen until the subject responded, at which point the screen was erased. If the response was correct, a 700-ms waiting period ensued, and then the asterisks for the next trial were presented. If the response was in error, the message "ERROR" appeared on the screen for 500 ms, followed by the next trial 700 ms later. Each block of 50 trials was completed in less than 2 min. Between each two blocks, subjects were encouraged to take a brief rest if desired.

Results

We present the data analyses in two parts: (a) standard Brinley plots and (b) the full range of correct and error response probabilities and their associated response time distributions displayed in quantile probability functions. Theoretical analyses with the diffusion model are presented after Experiment 2. Mean response times less than 300 ms and greater than 3,500 ms were eliminated from the data analyses.

The three groups of young subjects differed little in reaction time and accuracy, so we combined the data for the three groups in some analyses (in which we did not compare the three groups). We fit the diffusion model to the three groups separately, and the parameters of the model were similar across groups.

To display the data compactly, they were grouped in several ways. First, we examined response time and accuracy to "low" responses to low stimuli and "high" responses to high stimuli and found that they did not differ from each other systematically. Therefore we collapsed them together. That is, responses for each

stimulus for which the probability of a "high" response was larger than .5 were collapsed with the corresponding stimulus for which the probability of a "low" response was greater than .5. For example, the probability of responding "high" to 57 asterisks was the same as the probability of responding "low" to 37 asterisks (and their response times were not significantly different), so responses for 57 asterisks and 37 asterisks were collapsed. We designate these greater-than-.5-probability responses as correct responses. Similarly, responses for each stimulus for which the probability of a "high" response was less than .5 were collapsed with the corresponding stimuli for which the probability of a "low" response was less than .5, and these responses were designated as errors. Note that these designations for describing the results are not the same as the correct and error feedback given to subjects during the experiment.

After collapsing in the way just described, responses for the various experimental conditions (i.e., for the various numbers of asterisks) were grouped such that responses in each group had approximately the same response times and response probabilities. The 100 different experimental conditions were combined into four groups of experimental conditions. The four groups were defined as follows: "low" responses for stimuli less than 35 asterisks and "high" responses for stimuli greater than 61 were grouped, "low" responses to Stimuli 36–40 were grouped with "high" responses to Stimuli 56–60, and in the same way (high combined with low as in the last example), Stimuli 41–45 were grouped with Stimuli 51–55, and Stimuli 46–48 were grouped with Stimuli 49–50. Because the stimuli were obtained by selecting the numbers of asterisks from normal distributions, the numbers of responses in the four groups varied. For the young late-term subjects, for example, the total numbers of data points in the four groups were 8,900, 3,700, 4,000, and 2,100.

Brinley plot. Figure 2 shows a Brinley function with the data

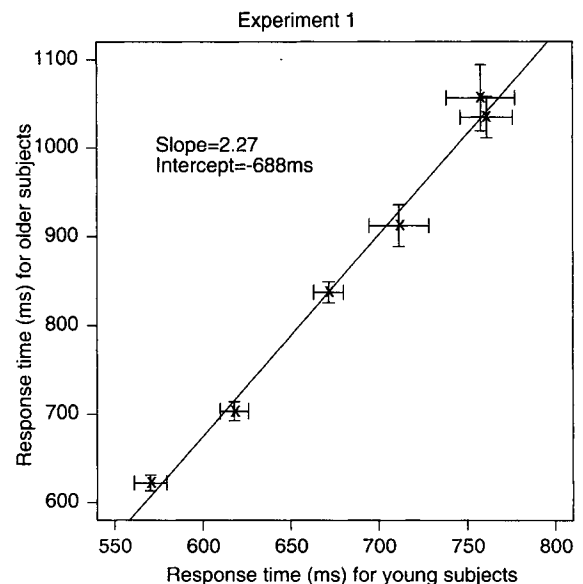


Figure 2. A Brinley plot for Experiment 1. The points on the graph represent the same conditions for older and young subjects. The error bars represent 2 standard errors.

for the older subjects plotted against the data for the first group of young subjects (later term, subject pool); the other young groups produced similar results. The four points corresponding to the fastest responses on the function are the mean response times for correct responses for each of the four groups of experimental conditions, and the two slowest points are the mean response times for error responses for two of the four groups with lowest accuracy. We did not display mean error response times for the two groups with higher accuracy because the data are a mixture of responses from some subjects with many errors (e.g., 10%) and some subjects with zero or very few errors (e.g., 1% or 2%); thus the mean error response times would not be representative of the group.

The error bars in Figure 2 are 2 standard errors in length. The diagonal line is the best fitting straight line. The function is linear, replicating the usual finding in the literature. The slope is 2.27 with an intercept of -688 ms, both in the normal range of results (see Cerella, 1985, Figure 3; Ratcliff et al., 2000, Figure 3). The fact that the slopes are greater than 1 means that the spread in response times for older subjects is greater than that for young subjects.

Latency probability functions and quantile probability functions. A difficulty in dealing with two-choice response time data is the number of dependent variables: the probabilities of correct and error responses and the distributions of response times for correct and error responses. Each of these must be described for each experimental condition. Plotting each dependent variable separately for each experimental condition would make the relative behaviors of the dependent variables extremely difficult to grasp.

Latency probability functions help solve this problem because they present the joint behavior of response probability and mean correct and error response times. A latency probability function is constructed by plotting the probabilities of responses on the x -axis and mean response times on the y -axis. Responses with probability greater than .5, on the right-hand end of the axis (in the asterisk task, these are correct "high" responses to stimuli from the high distribution for stimuli greater than the crossover point 47 asterisks and correct "low" responses to stimuli from the low distribution for stimuli lower than the crossover point) are the correct responses. Responses with probability less than .5, on the left-hand end of the axis, are error responses.

Latency probability functions were used 20 to 30 years ago to test among various response time models (e.g., Audley & Pike, 1965; Vickers, 1979; Vickers, Caudrey, & Willson, 1971), but after that they were largely ignored. Perhaps the main reason for this was that a latency probability function clearly displays error response times, and few models of the past 30 years deal adequately with error response times. Recently, however, new models have become available that do deal with error response times, and so the latency probability function is once again a useful tool for displaying data and evaluating models.

Unfortunately, latency probability functions do not provide information about the shapes or spreads of response time distributions. To show distributional information along with response probability, we introduce a new type of parametric plot, the quantile probability function. For this function, quantiles of the response time distribution for each experimental condition are plotted as a function of response probability. Examples are shown in Figure 3, for which the .1, .3, .5 (median), .7, and .9 quantiles are

plotted for each of the four (groups of) experimental conditions for the data from Experiment 1. The Xs are the data points, and the lines are the best fitting functions from the diffusion model, which we discuss later. As a specific example, consider the second panel, data for the later-term subject pool group of young subjects. For both correct and error responses for all four conditions, response times at the fastest quantile are about the same, about 450 ms. Responses in the middle of the response probability range, those with probabilities of .3 to .7, are slower than responses with higher or lower probabilities because the response time distributions spread. Responses in the .9 quantiles, for example, are much slower for responses in the middle of the probability range than at the ends of the range.

Quantile probability functions contain information about all the data from the experiment: the probabilities of correct and error responses and the shapes of the response time distributions for both correct and error responses. The information about the response time distributions is as detailed as the choice of quantiles; more detail can be plotted by using a larger number of quantiles (e.g., 20 quantiles instead of the 5 in Figure 3).

The four quantile probability functions in Figure 3 show data for the older subjects and for the three groups of young subjects. As is expected (Espinoza-Varas & Watson, 1994; Ratcliff & Rouder, 1998; Ratcliff et al., 1999) and was just mentioned, response times for correct responses increase as correct response probability moves from near 1 to near .5. The change in the shape of the response time distribution is in the tail, the .7 and .9 quantiles (i.e., the distribution skews out rather than shifting; e.g., see Ratcliff & Murdock, 1976). The .1 quantile changes little across conditions, whereas the .9 quantile changes considerably. Error responses are slower than correct responses, as shown in the quantile probability functions by comparison of mirror image points on opposite sides of the .5 probability point. For example, the quantiles at probability .2 are higher (slower) than the quantiles at their mirror image, probability .8. The error response times are longest when error response probability is in the .2 to .4 range, and they are shorter when error response probability is less than .1. However, error response times are never shorter than their mirror image correct response times.

There is one aspect of the results in Figure 3 in which the older subjects differ from the young subjects, which is that the distribution of error response times shifts for older subjects (all error response times become longer as accuracy increases; in particular, the response time for the .1 quantile increases from about 600 to 700 ms as error rate decreases from .4 to .1). However, there is a problem here, which is that, unlike the young subjects, some of the older subjects produced too few errors in the extreme conditions to allow computation of quantiles. Twenty six out of 40 older subjects had fewer than five errors in the highest accuracy (lowest error rate) condition. So the error response times for error rates with probabilities between .2 and 0 are based on the data from some but not all of the older subjects. In addition, some of the older subjects produced some very slow response times, of the order of several seconds (in pilot work, some older subjects actually tried to count the number of asterisks before we instructed them to go with a global sense of the number). These extra long response times were masked in the accurate conditions by many faster response times, but in the error conditions, these were the only responses produced. For these reasons, the fits of the model

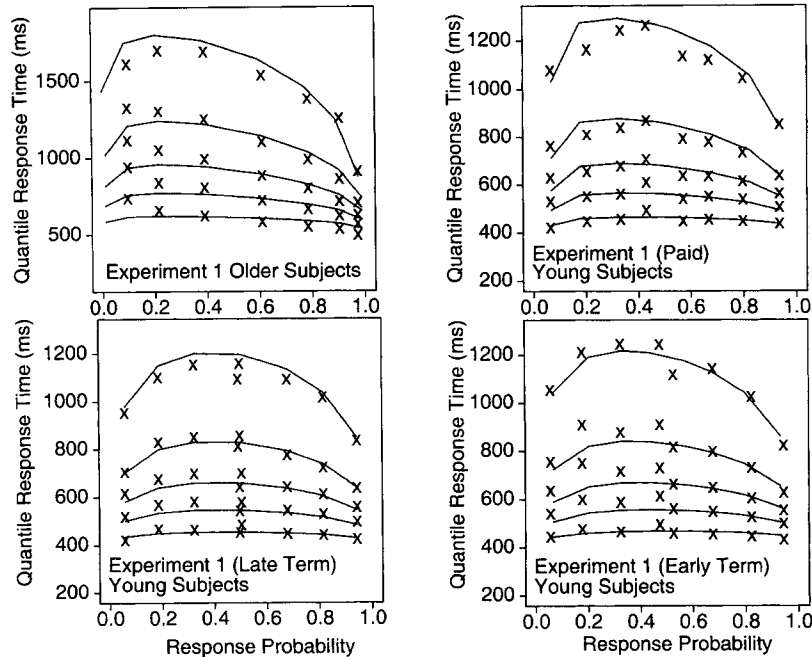


Figure 3. Quantile probability plots for older and young subjects for Experiment 1. The lines represent the theoretical fits of the diffusion model, and the Xs represent the data. The lines in order from the bottom are for the .1, .3, .5, .7, and .9 quantile response times. For older subjects, there are seven points displayed because the highest accuracy error condition had too few observations for several of the subjects to produce quantiles. Correct responses are to the right of the .5 response probability point, and the corresponding error responses are to the left (if the correct response probability is p , the error response probability is $1 - p$).

to the average data miss fitting the extreme error response times (i.e., for low response probabilities).

In sum, the experimental data show qualitatively similar trends for older and young subjects. The locus of the effects of age on response time will be apparent through analyses of the data with the diffusion model.

Experiment 2

There were two goals for Experiment 2. The first was to generalize the signal detection task to a different kind of stimulus. Instead of judging whether some number of asterisks was "high" or "low," subjects were asked to decide whether the distance between two dots was "large" or "small." The second goal was to add a speed-accuracy manipulation. For some blocks of trials subjects were instructed to respond as quickly as possible, and for other blocks of trials they were instructed to respond as accurately as possible. With this manipulation, mean response time can vary by as much as 500 ms (Ratcliff & Rouder, 1998, Experiment 1). The speed-accuracy manipulation provides a strong test of the diffusion model because only one parameter of the model can vary between the speed and accuracy conditions, yet the model must accommodate all differences in response time and accuracy.

Method

Stimuli. The stimuli were 32 pairs of dots presented on the screen of a personal computer monitor, the same as was used in Experiment 1. The bottom dot was in a fixed position on the screen, and the upper dot varied

in vertical distance above it (see Figure 1). There were 32 different distances (and thus 32 different experimental conditions), varying between $1\frac{1}{16}$ in. (1.7 cm; "small") and $1\frac{15}{16}$ in. (2.4 cm; "large") in equal-size steps.

On each trial, a distance was chosen from the 32 possible distances, each with probability $1/32$ of being chosen. Which response, "large" or "small," was considered the correct response was chosen according to a probability associated with each stimulus: For Stimuli 1 through 6, "small" was chosen with probability .999. For Stimuli 7 through 15, "small" was chosen with probabilities .913, .888, .856, .819, .774, .722, .664, .601, and .534, respectively. For Stimuli 16 through 25, "large" was chosen with the same probabilities of "small" Stimuli 15 through 7. For Stimuli 26 through 32, "large" was chosen with probability .999.

Subjects. Seventeen young adults (5 men and 12 women) and 13 older adults (4 men and 9 women) participated in the experiment. The young adults were college students who were recruited from the Northwestern University student body by advertisements and received \$16 for their participation. The older adults were healthy, active, community-dwelling adults aged 60 years or older. They were recruited from advertisements posted at local senior citizen centers and received \$30 for their participation. All subjects completed the CESD and a medical history form. Subjects were excluded if there was evidence of any of the following: disturbances in consciousness, medical or neurological disease causing cognitive impairment, head injury with loss of consciousness, current psychiatric disorder, a score of 25 or below on the Mini-Mental State Examination, or a score of 16 or more on the CESD. The older adults also completed the Vocabulary subtest of the Wechsler Adult Intelligence Scale—Revised (Wechsler, 1981). Estimates of full-scale IQ were derived from the scores of the two subtests (Kaufman, 1990). The means and standard deviations for these standard background characteristics are shown in Table 1.

Procedure. Each subject participated in two 45-min sessions. There were 12 lists of stimuli per session, and each list was made up of three presentations of each of the 32 stimuli in random order. A subject's task was to decide whether the distance between the dots was large or small, pressing the "F" key on the keyboard if large and the "Z" key if small. Of the 12 lists per session, half were speed lists and half were accuracy lists. In both list types subjects were given feedback 200 ms after the response: "CORRECT" or "ERROR" presented for 500 ms followed by a blank screen for 200 ms. In addition to the accuracy feedback, in the speed blocks responses longer than 700 ms had a "TOO SLOW" message presented for 500 ms followed by a 200-ms blank screen. In the accuracy blocks, after the accuracy feedback, a "large" response to Stimuli 1–6 or a "small" response to Stimuli 26–32 had the message "BAD ERROR" presented for 500 ms followed by a 200-ms blank screen.

Results

The results are organized in the same way as for Experiment 1; first we present Brinley plots and then quantile probability functions. Also as in Experiment 1, the experimental conditions were collapsed into four groups for analysis. Along with grouping by similar response times and accuracy (e.g., Distances 1–8 grouped together), we grouped first, high-probability responses ("large" responses to large distances and "small" responses to small distances) and second, low-probability responses ("small" responses to large distances and "large" responses to small distances). Specifically, the stimulus groupings were (a) "small" responses to Distances 1–8 were grouped with "large" responses to Distances 21–32, (b) "small" responses to Distances 9 and 10 were grouped with "large" responses to Distances 19 and 20, (c) "small" responses to Distances 11 and 12 were grouped with "large" responses to Distances 17 and 18, and (d) "small" responses to Distances 13 and 14 were grouped with "large" responses to Distances 15 and 16. There was a slight bias toward "small" responses that made these grouping choices optimal.

Brinley plots. Figure 4 shows Brinley plots for older versus young subjects for the speed blocks and for the accuracy blocks. As is typical, the functions are linear. For the speed blocks the slope is 1.46. If this were an index of the internal speed of processing for the older versus the young subjects, then the slope should be the same for the accuracy blocks. However, the slope for the accuracy blocks is 2.62, almost twice as large. If this were put in terms of the general slowing hypothesis, it would mean that older subjects' processing is only 1.46 times slower than young subjects' under speed instructions but 2.62 times slower under accuracy instructions. This is inconsistent with the widely held view that processing speed is a characteristic of the individual and not a characteristic of combinations of the individual with task instructions. It is also especially inconsistent with the view that the slope of the Brinley plot is an index of the speed of neural processes.

If we think of this slowing due to speed–accuracy instructions as reflecting a change in speed–accuracy criteria (see Pachella, 1974, and the application of the diffusion model below), then the reason for the change in the slope of the Brinley plot is that older subjects change their criteria more with accuracy instructions than do young subjects. It could be that the criterion change is greater for older subjects, or it could be that the same criterion shift has a larger effect for old subjects at their position on the speed–accuracy trade-off function (see Pachella, 1974). We discuss cri-

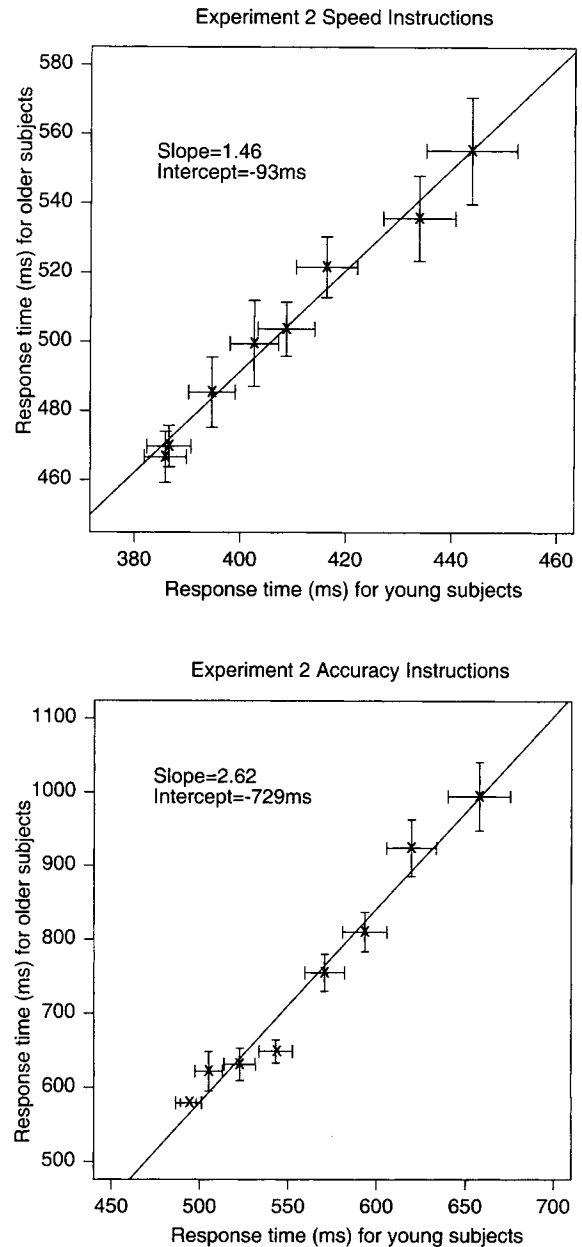


Figure 4. Brinley plots for the speed and the accuracy conditions for Experiment 2. The points on the graph represent the same conditions for older and young subjects. The error bars represent 2 standard errors.

terion changes in more detail below with application of the diffusion model.

For the speed and accuracy blocks for Experiment 2, the intercepts of the Brinley plots are negative, consistent with typical data and with Experiment 1. The fact that the slopes are greater than 1 means that the spread in response times for older subjects is greater than for young subjects, but this difference is reduced for speed conditions relative to accuracy conditions.

If the speed condition for older subjects is plotted against the accuracy condition for young subjects, the slope is 0.54 with a positive intercept of 200 ms. A result like this might be obtained

if subjects had participated in prior experiments with older subjects participating in deadline or response signal experiments, where rapid responses are required and young subjects participating in difficult tasks in which accurate responding is required. If the speed-accuracy criteria carried over from the prior task to the new task, then this pattern of results (a Brinley slope less than 1) could be obtained. In terms of general slowing, this would mean that the mental processes for older subjects trained to perform quickly are twice as fast as the mental processes of young subjects trained to be accurate.

The point is that the changes that subjects set in their speed-accuracy criteria can alter the slopes of Brinley plots a great deal. Thus, the slope can only be interpreted in terms of the criteria subjects set.

Quantile probability functions. Figure 5 shows quantile probability functions for responses under speed and accuracy instructions for older and young subjects. The functions all look similar to each other and similar to the functions for Experiment 1. The change in distribution shape over conditions is displayed in the behavior of the quantile response times: Response times at the .1 quantile vary little across conditions, whereas response times at the .9 quantile vary the most across conditions. In other words, the changes in response time are due to the distributions skewing out. The shapes of the functions for the speed and accuracy conditions are similar, although the scales on the y-axes are different (varying by about 3:1 for the older subjects' accuracy responses to the young subjects' speed responses). This means that speed-accuracy

instructions had a large effect on response time, but the overall shapes of the response time distributions stayed about the same. At the same time, the effect of speed versus accuracy instructions on accuracy was small: The largest difference in accuracy between the speed and accuracy conditions was about 4%. In the quantile probability functions, this appears as a horizontal shift in the five quantiles for each condition for speed versus accuracy conditions (see also Ratcliff & Rouder, 1998).

The Diffusion Model

The data from Experiments 1 and 2 provide a complete set of tests for modeling. The data spread the range in accuracy probabilities from ceiling to floor, and a model must accommodate the shapes of the response time distributions for correct and error responses at all levels of accuracy. Also, with the speed-accuracy manipulation, the range of data is replicated at two different speed-accuracy criterion settings. A model should be able to accommodate the effects of speed versus accuracy instructions on both response time distributions and accuracy probabilities with only criteria changing across the two kinds of instruction. In the paragraphs below, we present the diffusion model in detail and then fit it to the data. As we elaborate later, the model fits the data well and, in doing so, provides a comprehensive picture of the differences in processing between young and older subjects for the simple signal detection tasks used in the experiments.

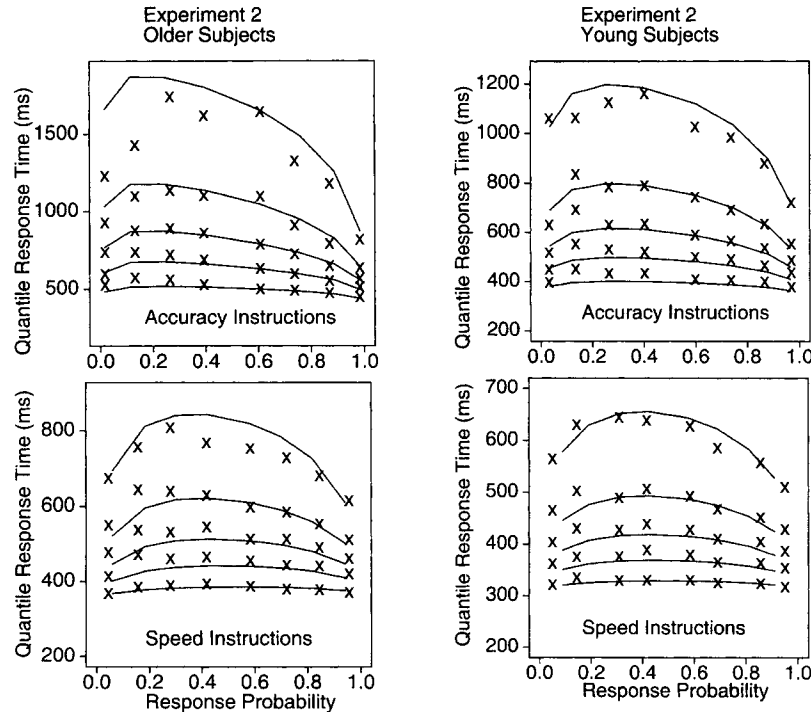


Figure 5. Quantile probability functions for older and young subjects for speed and accuracy conditions for Experiment 2. The lines represent the theoretical fits of the diffusion model, and the Xs represent the data. The lines in order from the bottom are for the .1, .3, .5, .7, and .9 quantile response times. Correct responses are to the right of the .5 response probability point, and the corresponding error responses are to the left (if the correct response probability is p , the error response probability is $1 - p$).

Ratcliff's diffusion model (Ratcliff, 1978, 1981, 1985, 1988; Ratcliff & Rouder, 1998, 2000; Ratcliff et al., 1999) is designed to account for response time and accuracy in experimental paradigms in which subjects are asked to make two-choice decisions. The model has been successful in dealing with both averaged group data and with data from individual subjects across a range of experimental paradigms, but it has not previously been applied to the effects of aging on response time.

The diffusion model is meant to apply only to two-choice decisions that are relatively fast and composed of a single-stage decision process (as opposed to the multistage decision processes that might be involved in, e.g., reasoning tasks, card-sorting tasks, and so on; for such tasks, other modeling techniques are available, e.g., see Fisher & Glaser, 1996). As a rule of thumb, we would not want to apply the diffusion model to experiments in which mean response times are much longer than 1 to 1.5 s (although this is a rough guideline rather than an absolute rule). With a multistage decision process, it could be assumed that the diffusion model was the decision process used in each of the individual stages.

The diffusion model assumes that decisions are made by a noisy process that accumulates information over time toward one of two response criteria or boundaries. The mean rate of accumulation of information is called *drift rate*, and it is determined by the quality of the information driving the decision process. For example, a large number of asterisks would have a large value of drift rate toward the large boundary. Within each trial, there is noise or variability in the process of accumulating information so that processes with the same mean drift rate will terminate at different times (this produces distributions of response time) and sometimes at different response boundaries (this is how errors occur). This source of variability is called *within-trial variability*. When one of the criteria is reached, a response is initiated. Speed-accuracy instruction manipulations are modeled by altering the boundaries: Wider boundaries require more information before a decision can be made, and this leads to more accurate and slower responses.

In the past, response time models had considerable difficulty in dealing with the relative speeds of correct versus error responses. Empirically, response times for errors are sometimes longer than response times for correct responses, sometimes shorter, and sometimes the relationship between error and correct response times varies for an individual subject across the conditions of an experiment (Ratcliff & Rouder, 1998; Ratcliff et al., 1999; P. L. Smith & Vickers, 1988). The diffusion model can account for these varying patterns by using variability in parameter values, specifically *across-trial variability* in mean drift rate and across-trial variability in the position of the starting point. In memory tasks, across-trial variability in drift rate means, for example, that the 10th word in a list of words to memorize is not encoded with exactly the same strength for each list (Ratcliff, 1978). In the signal detection tasks used in Experiments 1 and 2, across-trial variability in drift rate means that a stimulus of 34 asterisks, for example, would not always be encoded in exactly the same way. This variability across trials in drift rate produces slow errors relative to correct responses (see Ratcliff & Rouder, 1998; Ratcliff et al., 1999). Without across-trial variability in drift rate, the diffusion model (like any random walk or diffusion model with constant mean drift rate across trials) predicts that error response times should be the same as correct response times with the starting point of the process equidistant from both response bound-

aries. The opposite pattern, error responses faster than correct responses, is produced by variability in starting point across trials (Laming, 1968; see also Ratcliff, 1981). With both of these two sources of across-trial variability, crossover patterns of error versus correct response times are exactly predicted by the model (errors faster than correct responses when accuracy is high, errors slower than correct responses when accuracy is low). Currently, no other model is capable of producing these patterns of results (Ratcliff & Rouder, 1998; Ratcliff et al., 1999, see also P. L. Smith & Vickers, 1988, Van Zandt & Ratcliff, 1995).

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 0 is the lower boundary. For the simulations described in this article, the boundaries were assumed to be symmetric about the starting point so that $z = a/2$ (because the data are symmetric, "high" responses to high stimuli produce the same values of accuracy and response time as do "low" responses to low stimuli). The amount of variability in the mean drift rate across trials is assumed to be normally distributed with standard deviation η , and the variability in starting point is assumed to have a uniform distribution with range s_z . 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 . In the fits of the diffusion model to the data from both Experiments 1 and 2, there were four values of v , one for each group of experimental conditions. Finally, there is a parameter T_{er} that represents the nondecisional components of response time such as encoding and response execution. Within-trial variability in drift rate (s) was kept constant in all the simulations because it 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 components of the model that are the most likely candidates for explaining age-related differences in response times are drift rate, boundary position, and T_{er} . Older subjects might extract less information from the stimulus, and so their drift rates might be lower than young subjects'. They might also set boundary positions wider to make accurate performance more likely. Alternatively, they might be slower in the nondecisional components of processing. Across-trial variability might also differ between young and old adults; this would have only a small effect on correct response times but a large effect on error response times.

Fitting the Diffusion Model to Data

The diffusion model has four parameters that are free to vary to fit the shape of the quantile probability function for a single data set: the boundary separation, a ; the nondecision component of response time, T_{er} ; the variability in drift across trials, η (the standard deviation of a normal distribution); and the variability in starting point, s_z (the range of a rectangular distribution). With four experimental conditions (four groupings of numbers of asterisks), there are four drift rates. It is assumed that the response criteria (represented by the boundary separation a) are constant across different levels of the independent variable, that is, the number of asterisks. This assumption is made because it seems unlikely that subjects can determine how many asterisks are presented, adjust decision criteria, and then use the information about the number of

asterisks to make a decision within 400 to 800 ms (and it makes no sense—if they had information about the stimulus to adjust criteria, they could use this to make their decision). With these eight parameters (a , T_{er} , η , s_z , and four drift rates), the model has to fit the probabilities of correct and error responses and their response times for all of the conditions in an experiment. To capture the shapes of the response time distributions for each group of conditions, the model is fit to the quantiles for that group (five quantiles for the data from Experiments 1 and 2). The parameters of the model are not free to vary independently of each other because they jointly affect predictions for all conditions. If one or two data points are not in line with the model, it is not possible to adjust just one parameter to bring these two points into line; all eight parameters must be adjusted. For example, adjusting boundary separation a alters both response times and accuracy values for all conditions (and alters the location of the .1 quantile reaction time).

The diffusion model was fit to the data from Experiments 1 and 2 by minimizing a chi-square value with a general SIMPLEX (Nelder & Mead, 1965) minimization routine that adjusts the parameters of the model to find the parameters that give the minimum chi-square value. The data entered into the minimization routine for each group of experimental conditions were the response times for each of the five quantiles for correct and error responses (i.e., the values shown in Figures 3 and 5). The quantile response times were fed into the diffusion model, and for each quantile the cumulative probability of a response by that point in time was generated from the model. Subtracting the cumulative probabilities for each successive quantile from the next higher quantile gives the proportion of responses between each quantile. For the chi-square value, these are the expected values, to be compared with the observed proportions of responses for each quantile (multiplied by the number of observations). The observed proportions of responses for each quantile are the proportions of the distribution between successive quantiles (i.e., the proportions between 0, .1, .3, .5, .7, .9, and 1.0 are .1, .2, .2, .2, .2, and .1) multiplied by the probability correct for correct response distributions or the probability of error for error response distributions (in both cases, multiplied by the number of observations). Summing over $(\text{observed} - \text{expected})^2 / (\text{expected})$ for all conditions gives a single chi-square value to be minimized.

In the fits to the data for the experiments, the data from each subject were fitted individually to obtain the best fitting parameter values, and then these values were averaged across subjects. These averages are used in interpreting the effects of age and speed-accuracy instructions.

For displaying in figures the quality of the fits to the data, these averages might be used to generate predicted fits for average data. However, there would be an averaging problem because the diffusion model is nonlinear. If perfect fits were produced for each individual, then averaging the parameter values from the individual fits would not produce the same parameter values as fitting the model to the data averaged across the individuals. This means that displaying in figures model fits generated from averaged parameter values will not show how well the model fits the individual subjects' data. Faced with this problem, we chose to illustrate the quality of the fit of the model in Figures 3 and 5 by using the fit of the model to the average data (for averages across the quantiles, see Ratcliff, 1979) rather than generating predicted values from the

average parameters. It is important to note that the differences between the average of the parameters for the individual fits and the parameters from the fits to the average data are not large; both sets of parameter values lead to the same interpretations for the effects of aging on parameters of the model.

The predictions of the diffusion model for latency probability functions have interesting properties. It is a general characteristic of the model that the predicted functions are parametric plots for which the shape is completely determined by only three parameters. To explain this we first review receiver operating characteristics (ROCs) from signal detection theory. The top panel of Figure 6 shows an ROC function, which is a parametric plot, for normally distributed signal and noise distributions with equal standard deviations. The parameter that sweeps out the function is the position of the decision criterion between the signal and noise distributions. Thus, the shape and position of the ROC function is independent of the criterion; the criterion setting determines the position on the function for a particular experiment.

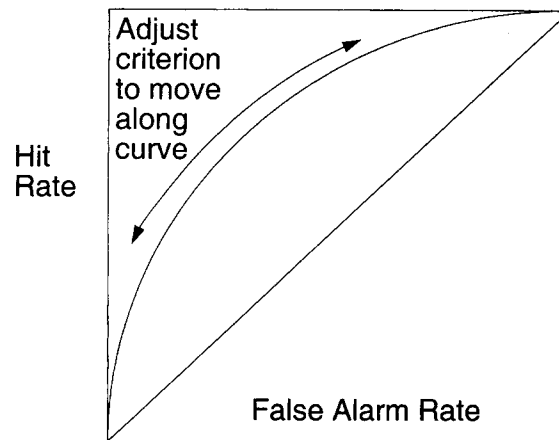
The bottom panel of Figure 6 shows how the latency probability function predicted from the diffusion model is also a parametric plot. (We use the latency probability function here instead of the quantile probability function for simplicity.) The drift rate parameter sweeps out the function. The shape of the function is determined by just three parameters: the boundary separation a , the standard deviation in drift across trials η , and the range of starting point values s_z . The displacement of the function in the vertical direction is determined by T_{er} . The function represents a stringent constraint on the model. There are only three parameters to fit the shape of the function and only the single parameter drift rate to sweep it out across experimental conditions. This applies to all the quantiles of the response time distribution; the three parameters determine all of these. These constraints apply in fitting data for all experimental conditions for which subjects cannot change their response criteria (or other parameters) from one experimental condition to the next, that is, the constraints apply in within-subjects designs (see Ratcliff, 1978, Experiment 2). For example, in Experiments 1 and 2 subjects cannot change their response criteria as a function of the number of asterisks, but in Experiment 2 they can change their response criteria as a function of speed versus accuracy instructions.

An even stronger constraint on the model is that the shapes of the response time distributions and how they change across experimental conditions are also completely determined by these same three parameters. The diffusion model allows only one particular kind of shape change as drift rate varies across experimental conditions: As response times slow across experimental conditions, the leading edges of the distributions must shift only slightly, and the main changes in response times must come from the distributions spreading in the higher quantiles. This is exactly what the quantile probability functions in Figures 3 and 5 show: virtually no change in the fastest quantiles across conditions accompanied by spreading in the slower quantiles.

Fits of the Diffusion Model to Experiments 1 and 2

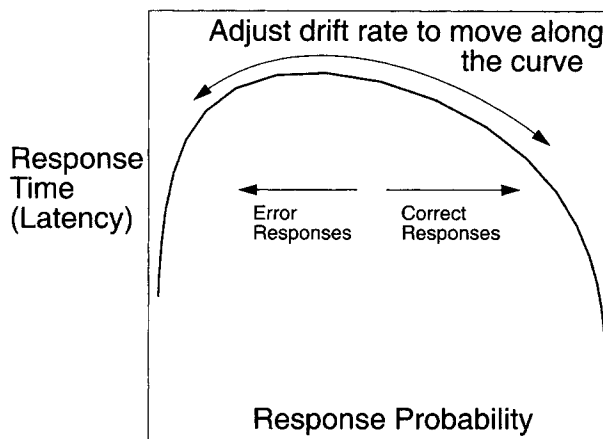
The fits of the diffusion model are shown in Figures 3 and 5. The Xs are the data points, and the solid lines are the best fitting lines to those data points. In general the fits are quite good. The only serious misses (greater than 20 ms) are in the .9 quantiles or

Signal Detection Theory



Criterion is the parameter of the parametric plot (of Hits vs. False Alarms) used to sweep out the ROC curve.

Latency Probability Function



Drift rate is the parameter of the parametric plot (of RT vs. response probability) used to sweep out the latency probability function in the diffusion model.

Figure 6. Parametric plots for receiver operating characteristic (ROC) functions in signal-detection theory and for latency probability functions. RT = response time.

in the conditions for which the probability of errors is low (i.e., the conditions with few observations). The chi-square values for the fit of the model to the group data from Experiment 1 for the four groups of subjects were 75.9, 62.8, 43.0, and 41.8 (old, young, paid, and early-term subjects, respectively), with $df = 35$ and critical $\chi^2 = 49.7$. The first two groups had significant chi-square values, whereas the last two did not. Because we computed chi-square values for groups, there were averaging problems that would lead to inflated chi-square values. These results, however, show that the fits are only slightly different from the predictions. For Experiment 2 the chi-square values were 105.5 and 99.2, with $df = 78$ and with the critical chi-square value 99.8 for $p = .05$. One value was barely significant and the other barely not significant, which shows that the model fits the group data reasonably well.

The means of the parameter values from the fits of the model to individual subjects are displayed in Table 2, and the standard errors in the mean parameter values are displayed in Table 3.

The three groups of young subjects in Experiment 1 differed little in the parameters a , T_{er} , η , and s_z . The parameter values among the groups were all within two standard errors of each other. Where the three groups did differ was that the drift rates for the experimental conditions with highest accuracy were smaller for the end-of-term subjects than for the other two groups. The end-of-term subjects were extracting poorer information from the stimulus displays than the other two groups, probably because they were less motivated (t tests for drift rates v_1 , v_2 , and v_3 between the end-of-term young subjects and the other two groups combined were each significant at the .05 level).

Table 2
Parameter Values From Fits of the Diffusion Model to Experiments 1 and 2

Experiment and group	a	T_{er}	η	s_z	v_1	v_2	v_3	v_4
Experiment 1								
Older	0.1705	0.3938	0.0928	0.0365	0.3507	0.2040	0.1152	0.0425
Young late term	0.1413	0.3409	0.0846	0.0513	0.2727	0.1472	0.0861	0.0037
Young early term	0.1475	0.3459	0.0992	0.0458	0.3152	0.1950	0.1188	0.0399
Young paid	0.1416	0.3549	0.0959	0.0561	0.3008	0.1801	0.1163	0.0410
Experiment 2								
Older accuracy	0.1805	0.3234	0.1081	0.0453	0.3302	0.2063	0.1357	0.0364
Older speed	0.1096	0.3234	0.1081	0.0453	0.3302	0.2063	0.1357	0.0364
Young accuracy	0.1483	0.2846	0.1040	0.0396	0.3300	0.2050	0.1077	0.0425
Young speed	0.0914	0.2846	0.1040	0.0396	0.3300	0.2050	0.1077	0.0425

Note. a = boundary separation; T_{er} = nondecision component of response time; η = standard deviation in drift across trials; s_z = range of the distribution of starting point (z); v = drift rates.

The older subjects in Experiment 1 differed from the three young groups combined in three main ways. First, the older subjects spread their boundary positions wider than the young subjects (larger values of the parameter a , $t(142) = 4.07$, $p < .05$), indicating that they adopted a more conservative response criterion in order to achieve greater accuracy. The standard error in the mean values of a across subjects was also larger for the older subjects, meaning that there were wider differences in this parameter across the older subjects than across the young subjects.

Second, the nondecision component of response time T_{er} was about 50 ms larger for the older subjects than for the young subjects, $t(142) = 4.53$, $p < .05$. The standard error in this parameter value was also larger across older subjects than young subjects.

Third, and most surprising, was that the drift rates in the two highest accuracy conditions were a little larger for the older subjects than the three groups of young subjects combined, $t(142) = 3.44$ and 2.74 , $p < .05$ (the difference is also significant with the late-term young subjects excluded for the highest drift rate). This means that the older subjects were extracting better information from the stimuli. We attribute this to higher motivation for the older subjects; that is, they were more likely to try to perform well over the whole experimental session than were young subjects. Accuracy values also suggest that older subjects are

extracting better information because older subjects are more accurate than the young subjects for equivalent conditions (this can be seen in the quantile probability functions in a slight shift to the right for older subject conditions compared with young subject conditions). The main conclusion here is that the older subjects do not have worse drift rates in this signal detection task. The evidence that they extract from the stimuli is not of lower quality than that of the young subjects.

It should also be pointed out that the across-trial variability in drift rate (η) was about the same for older and young subjects. It might have been thought that older subjects would show more variability, but they did not. There were also small, but not significant, differences in starting point variability for the four groups of subjects.

In Experiment 2, two of the same differences between older and young subjects emerged. The parameter values and the standard errors in them are shown in Tables 2 and 3. First, the boundary separations and their standard errors were larger for the older than the young subjects (the speed and accuracy boundary separations were not separately significant, but the average of the two was), $t(28) = 2.26$, $p < .05$. Second, there was about a 40-ms difference between older and young subjects in the nondecision component of response time, $t(28) = 2.71$, $p < .05$. However, in this experiment, unlike Experiment 1, drift rates for the older subjects were not

Table 3
Standard Errors Across Subjects in Parameter Values From Fits of the Diffusion Model to Experiments 1 and 2

Experiment and group	a	T_{er}	η	s_z	v_1	v_2	v_3	v_4
Experiment 1								
Older	0.0073	0.0125	0.0075	0.0054	0.0155	0.0114	0.0062	0.0037
Young late term	0.0048	0.0074	0.0108	0.0054	0.0101	0.0080	0.0066	0.0085
Young early term	0.0065	0.0078	0.0121	0.0061	0.0158	0.0109	0.0083	0.0042
Young paid	0.0047	0.0070	0.0109	0.0058	0.0164	0.0091	0.0075	0.0035
Experiment 2								
Older accuracy	0.0100	0.0129	0.0171	0.0080	0.0170	0.0170	0.0078	0.0127
Older speed	0.0042	0.0129	0.0171	0.0080	0.0170	0.0170	0.0078	0.0127
Young accuracy	0.0095	0.0044	0.0138	0.0044	0.0129	0.0137	0.0091	0.0050
Young speed	0.0035	0.0044	0.0138	0.0044	0.0129	0.0137	0.0091	0.0050

Note. a = boundary separation; T_{er} = nondecision component of response time; η = standard deviation in drift across trials; s_z = range of the distribution of starting point (z); v = drift rates.

significantly larger than for the young subjects (except for the third condition in which drift rate was a little larger for the older subjects), $t(28) = 2.74, p < .05$. As in Experiment 1, the standard deviation in drift across trials and the range of starting points did not differ from older to young subjects, and the values were about the same as those in Experiment 1.

The speed-accuracy manipulation in Experiment 2 had a large effect on performance. For both the young and older subjects, boundary positions are set wider with accuracy instructions than with speed instructions, but the difference is much larger for the older subjects, consistent with the interpretation that they were more motivated. The diffusion model provides good fits to all the data despite having only one parameter (boundary position a) that can vary between the speed and accuracy conditions (see Ratcliff & Rouder, 1998, Experiment 1). Across these conditions, all the other parameters must remain constant. The success of the model at fitting such large differences in speed, accuracy, and response time distributions between the two conditions is strong support for the model.

Comparisons of parameter values across Experiments 1 and 2 show that the values are about the same for the two experiments, and this also provides compelling support for the model. First, the boundary separation parameters for the young and older subjects are about the same for the accuracy condition of Experiment 2 as for Experiment 1. For young subjects the parameters are .144 and .148, and for older subjects the parameters are .171 and .181, for Experiments 1 and 2, respectively. This suggests that, in this simple signal detection task when no explicit instructions are given, subjects tend to stress accuracy in their performance.

Second, the difference between the T_{er} values for young and older subjects is about the same in the two experiments, about 50 ms. This means that the motor output and other nondecision components of processing are slower for the older subjects by about the same amount in the two experiments. (It should be noted that this 50-ms difference is not nearly as large as the slowing factor of 1.5 or more that the Brinley plot analyses would suggest.)

Third, the two variability parameters, standard deviation in drift across trials (η) and range of starting points (s_z), differ little across the experiments; they also differ little between the groups of subjects (see Table 2). The differences between the averages for the groups are much less than differences among the individuals.

Finally, the fits of the diffusion model to the data provide an explanation for the Brinley plot results. Ratcliff et al. (2000) showed that the diffusion model could produce slopes greater than 1 with drift rates, boundary positions, or both differing between young and older subjects. In Experiments 1 and 2 it is clearly differences in boundary positions that produce the observed differences in the spreads of mean response times across conditions. The boundary separation is larger for older subjects than young subjects, and there is more variability across subjects for older subjects than young subjects. This is what produces the larger spread in mean response time across conditions for older subjects, and this in turn produces the greater than 1 value of slope for the Brinley plot.

Individual Differences in the Diffusion Model

For most models, there have been no systematic studies of how the parameters of the model vary across individual subjects, and

this is true of the diffusion model. It could be that some parameters vary little across individuals. Or it could be that the distributions of parameter values are highly skewed with most subjects having, for example, fairly low values of drift rates, but a few (very motivated) subjects having much higher values. Another possibility is that the values of two or more parameters might be correlated across individuals. For example, motivated subjects might set their boundaries far apart under conditions for which they realize that drift rates are low (as they might be in experimental conditions for which the task was very difficult). For these subjects, drift rates and boundary positions would be correlated.

To provide a large sample size to examine parameter values across individuals, we combined the three groups of young subjects from Experiment 1 and plotted histograms for each of the diffusion model parameters. Figure 7 shows these histograms, and Figure 8 shows the same plots for the older subjects from Experiment 1. The main result was that the distributions of parameter values are about normal (with the exception of the leftmost peaks for the η and s_z parameters, Figures 7 and 8), perhaps with a little more skew to the right than a normal distribution. The reason for the slight skew is that for each of the parameters, there is a minimum of zero but no maximum. The finding of roughly normal distributions of parameter values across subjects shows that the

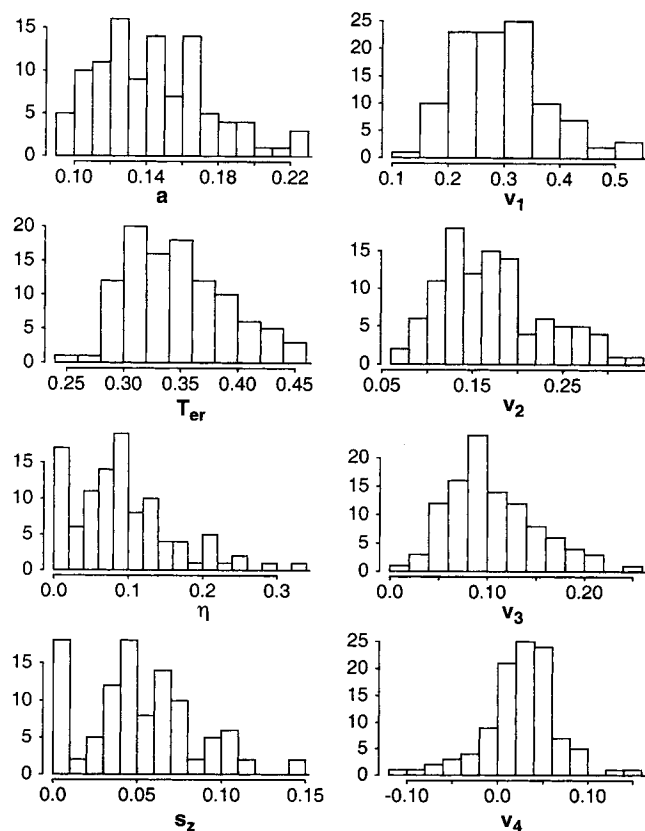


Figure 7. Histograms for the parameter values across young subjects (the three groups combined) for fits of the diffusion model for Experiment 1. a = boundary separation; T_{er} = nondecision component of response time; η = standard deviation in drift across trials; s_z = range of the distribution of starting point (z); v = drift rates.

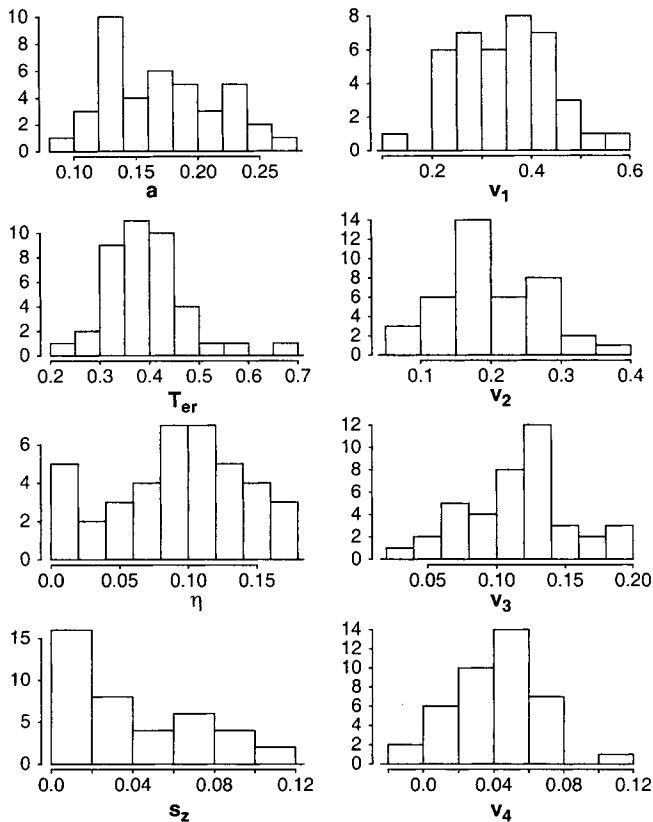


Figure 8. Histograms for the parameter values across older subjects for fits of the diffusion model for Experiment 1. a = boundary separation; T_{er} = nondecision component of response time; η = standard deviation in drift across trials; s_z = range of the distribution of starting point (z); v = drift rates.

diffusion model produces a reasonably uncontroversial interpretation of individual differences. Some subjects have small parameter values, whereas others have large parameter values, but most are in the middle.

A number of values of the η and s_z variability parameters for individual subjects were estimated to be zero (the left peak in the histograms for these parameters). This is caused by the fitting method. When the true η or s_z parameter value is small, and other parameter values are fixed at their minima, then there is no minimum in the chi-square function as η or s_z is increased from zero. This occurs because of the variability in the data that enters the chi-square computation. Without a minimum, the fitting method produces a best fit at zero. There may also be cases in which these parameters are genuinely small (e.g., 0.001), but the fitting method cannot distinguish these cases because the variability in the data is too large to allow such precise location of the parameters.

The other aspect of individual differences that we examined is whether some of the parameter values are correlated. For example, a subject could take less time to encode the stimuli (smaller T_{er}) and as a result have smaller drift rates. To examine covariations among the parameters, we computed correlations across all the young subjects from Experiment 1 for all the possible pairs of

parameter values. Table 4 shows these correlations, and Figure 9 shows scatter plots for each pair.

Many of the correlations in Table 4 are of moderate size (.2–.4), but most would be considerably smaller if a few extreme values were eliminated (e.g., if 5–10 points in the top right-hand corner of the panels in Figure 9 for the correlations between a and the drift rates v were removed, the correlations would be close to 0). Figure 9 shows that the main systematic correlations between parameters are, first, the drift rates v_1 , v_2 , and v_3 . If one drift rate value is high, the others are also high (v_4 is near 0 and so does not correlate with the other drift rates). Second, the standard deviation in drift across trials correlates with drift rates and with boundary separation. As the drift rates increase, the variability in drift across trials may increase simply as a scaling effect (i.e., the standard deviation increases because the range above floor increases). The variability in drift across trials may increase with boundary separation because the more variable the drift rate across trials, the more conservative the subjects will be, and this leads to them setting a higher value of boundary separation. Boundary separation does not covary with drift rate as much as standard deviation in drift across trials covaries with each of them. This can be interpreted as meaning that a higher drift rate does not lead to more conservative response criteria except through the intermediate factor of variability in drift across trials. Although some of the other correlations may be significant (a value of $r = .32$ is barely significant at the .028 level corrected for the 28 possible comparisons; for $N = 100$ –104 data points from 104 subjects and a p level of .001, $1 - [1 - .001]^{28} = .028$), this is not caused by a strong linear relationship between the variables (see Figure 9) but rather by a few scores being extreme in one direction or another.

The analyses of individual differences show no surprises. The distributions of parameter values are reasonably normal, and there are few large correlations among the parameter values. This means that the conclusions that we draw from fitting the model to the data are not distorted by extreme individual differences.

General Discussion

Overall, the diffusion model provides a novel interpretation of aging effects on response time. Ratcliff et al. (2000) showed that the Brinley plot regularities that have been obtained in aging research can be explained by the diffusion model (and other models of the sequential sampling class, e.g., accumulators, other random walks, counter models, and so on; see Luce, 1986). The larger variance of the older subjects' distributions of response times and the slower mean response times for older subjects can be produced in a variety of different ways from the diffusion model, for example, by differences between the older and young subjects in drift rates, boundary positions, the nondecisional components of response time, or combinations of these.

The fact that the data from both the older and young subjects are fit by the same diffusion model suggests the same conclusion as was drawn by G. A. Smith and Brewer (1995), that the "older group used the same processing mechanisms as the younger group... but... respond more carefully" (p. 246).

The experiments described in this article examined the effects of aging in simple response time tasks. When the diffusion model was used to interpret the data, three main effects of aging were found. First, the older subjects set wider, more conservative response

Table 4
Correlations Among Parameter Values for Individual Subjects for All Young Subjects for Experiment 1

	a	T_{er}	η	s_z	v_1	v_2	v_3	v_4
a	—	.086	.675	.353	.487	.354	.310	.207
T_{er}	.086	—	.223	-.079	.350	.281	.165	.168
η	.675	.223	—	.408	.694	.664	.541	.279
s_z	.353	-.079	.408	—	.285	.318	.303	.251
v_1	.487	.350	.694	.285	—	.822	.682	.363
v_2	.354	.281	.664	.318	.822	—	.770	.449
v_3	.310	.165	.541	.303	.682	.770	—	.403
v_4	.207	.168	.279	.251	.363	.449	.403	—

Note. a = boundary separation; T_{er} = nondecision component of response time; η = standard deviation in drift across trials; s_z = range of the distribution of starting point (z); v = drift rates.

criteria than the young subjects. This means that older subjects preferred more evidence than the young subjects to make a decision, and so they were acting more conservatively than the young subjects. Second, the nondecisional components of processing were slower for the older subjects by about 50 ms. Third, drift rates were never smaller for the older subjects than for the young subjects, and in some conditions of Experiment 1 they were significantly larger than for the young subjects. This means that the quality of evidence obtained from the stimulus displays is just as good (sometimes better) for the older subjects than for the young subjects.

In Experiment 2 both older and young subjects were able to adjust their speeds of processing according to speed versus accuracy instructions. This is consistent with the diffusion model and other models of the sequential sampling class. The differences in

performance under speed versus accuracy instructions are explained by movement of the criteria, which represent the amount of evidence required for a decision (a greater change for older subjects than younger subjects).

We can speculate why older subjects adopt more conservative response criteria. First, fit the diffusion model to pilot data from a letter identification task with masking (cf. Ratcliff & Rouder, 2000). Overall, the older subjects responded about 150 ms slower than the young subjects. The nondecisional component of processing was larger for the older subjects, and the response criteria were more conservative for the older subjects than for the young subjects (the same qualitative pattern as we found in Experiments 1 and 2 above). However, in contrast with the experiments presented here, drift rates were lower for older subjects than young subjects, by nearly a factor of two. This means that the older subjects are getting perceptual information at a lower rate than young subjects. If lower rates of accumulation of information occur in many tasks for older subjects, then, as a result, they might adopt more conservative decision criteria across all tasks that require rapid one-process decisions. This would explain why the older subjects are more conservative in the signal detection tasks even when their information accumulation is just as good as for young subjects.

Although we would not want to apply the diffusion model to tasks in which mean response times were much over 1 to 1.5 s, the finding that older subjects adopt more conservative criteria may also apply in tasks where response time is much greater than a second. For example, Hertzog, Vernon, and Rypma (1993) examined the effect of speed-accuracy instruction in a mental rotation task. They found the same kinds of qualitative effects we obtained here. Older subjects were less willing to trade accuracy for speed than were young subjects. It may be that criteria that determine how much information is needed for a decision or how much information is needed before the processing system can proceed to another stage of processing are set globally so that they apply across a range of experimental tasks.

Patterns the Diffusion Model Cannot Fit

Roberts and Pashler (2000) have criticized the use of goodness-of-fit measures as the sole method for evaluating a model and have suggested that knowing what a model cannot fit is almost as important as knowing that it can fit a data set. This leads to the suspicion that a many-parameter model such as the diffusion

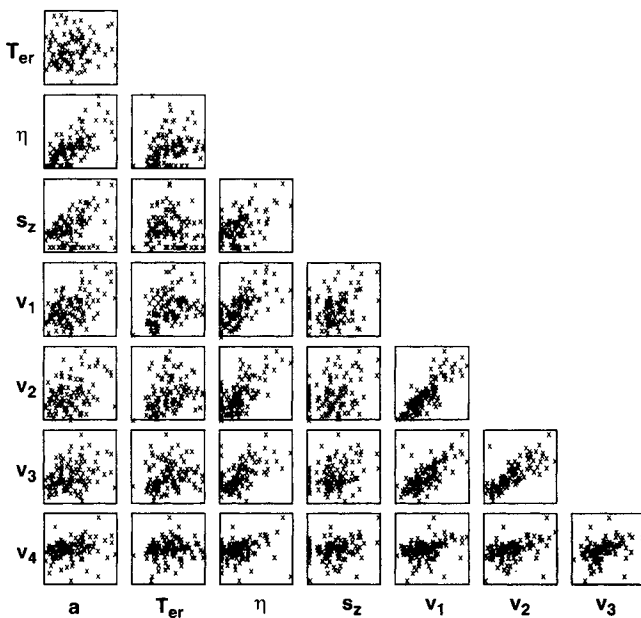


Figure 9. Scatter plots of parameter values for fits of the diffusion model to individual subjects for young subjects (three groups combined) for Experiment 1. a = boundary separation; T_{er} = nondecision component of response time; η = standard deviation in drift across trials; s_z = range of the distribution of starting point (z); v = drift rates.

model can fit any pattern of results (a reviewer of another article recently made this claim). Ratcliff (2001) has presented examples of empirical data that could not be fit by the diffusion model. He showed several possible quantile probability functions that could not be obtained with the diffusion model (under the assumption, appropriate for the relevant applications, that all that changes in the model across experimental conditions is drift rate). Quantile probability functions that the diffusion model cannot fit include U-shaped functions; for example, if the inverted U-shaped quantile probability functions in Figures 3 and 5 were flipped upside down, the diffusion model could not fit them. Other examples that the diffusion model could not fit include quantile probability functions with large changes in the .1 quantile across conditions, functions with normal distributions of reaction times, functions with tails significantly longer than exponential, and so on. The point from these examples is that the model is highly constrained. The patterns of data it can explain are produced from experiments, and the patterns that it cannot explain are not produced from the two-choice experiments that have been addressed by the diffusion model so far.

Alternative Models

No alternative models are presented in this article for contrast with the diffusion model. The reason is that it is only in the last few years that attempts have been made to explicitly fit accuracy and reaction time distributions for correct and error responses with stochastic processing models. Before this, the focus was on only some aspects of the experimental data, for example, accuracy and correct mean response time. Some competitor models that are promising are Ornstein Uhlenbeck diffusion models (Busemeyer & Townsend, 1993; P. L. Smith, 1995), accumulator models, and counter models (LaBerge, 1962; P. L. Smith & Vickers, 1988; Vickers, 1979). These models should be able to account for accuracy rates and for correct and error response time distributions. However, none of these models has yet been shown to successfully account for the full range of experimental data across more than a single experimental paradigm. When these models (e.g., Ratcliff & Smith, 2001; Van Zandt, Colonius, & Proctor, 2000) have been more fully evaluated, they can be applied to the data presented in this article, and their parameters can be used to interpret the effects of aging by using the application of the diffusion model in this article as a guide.

Brinley Plots

The study of the effects of aging on response times in cognitive tasks has focused in large part on the degree to which cognitive processes slow with age. The reason for this has been the common interpretation of Brinley plots, that their slope shows the degree to which the mental processes of older people are slowed relative to young people (see Cerella, 1985, 1991, 1994; Myerson et al., 1990; Nebes & Madden, 1988).

Fisher and Glaser (1996) questioned the use of Brinley plots when they showed that, within a framework that assumes that cognitive processes are actually combinations of processes (operating in serial or parallel fashion or a combination of the two), the conclusions that might be drawn from qualitative analyses of Brinley plots are not unique. For example, response time data for

older versus young subjects that appear to be consistent with a general slowing of all components of processing could instead be the result of different components slowing at different rates. They also showed the opposite, that data that might be interpreted as showing differential slowing of components could also be the result of general slowing of all components. Thus, where it is appropriate to think of processing as combinations of processes (we think of appropriate domains as those where response times are seconds), these warnings imply that the Brinley plot cannot be used by itself to make inferences about processing.

Ratcliff et al. (2000) provided a different attack on the diagnosticity of Brinley plots. As mentioned above, they showed that the slope of the Brinley plot does not measure the slowing of one group of subjects relative to another; rather, it measures the spread in mean response times across conditions of one group relative to another. It is the intercept of the Brinley plot rather than the slope that gives a measure of slowing. The slope of the Brinley plot is an index of the relative spread of the subjects' response times across conditions.

There are also practical problems with the standard Brinley analyses, as exemplified by the data from Experiment 2. For this experiment, the slope of the Brinley plot was affected in almost a 2:1 ratio by the speed-accuracy manipulation. Were the slope of the Brinley plot to indicate the relative speed of mental processes for older versus young subjects, then the results of Experiment 2 would indicate, first, that the older subjects' mental processes speed up in going from accuracy to speed instructions, and second, that the relative speeds of older versus young subjects are different under speed versus accuracy instructions. Most extremely, if the response times of the older subjects in the speed condition are plotted against the response times of the young subjects in the accuracy condition, then the slope is about 0.5, indicating that the speed of mental processes for the older subjects is about twice that of the young subjects. Obviously, the often-invoked hypothesis that Brinley plots show a general slowing of cognitive processes for older people is not consistent with the data from Experiment 2.

The relative processing speeds shown in Experiment 2 also conflict with two prominent models, Cerella's (1990) neural network model and Myerson et al.'s (1990) information loss model. Cerella suggested that processing is performed in a series of stages within a neural network. The effect of aging is to break some of the links, with the result that cognitive computations must take longer neural paths. Myerson et al. suggested that cognitive tasks are performed by series of processes and that the effect of aging is to slow down each process by a fixed amount. The fact that older people can speed their responses to a rate faster than that of young subjects (when the young subjects are trying to be accurate) implies that these models would have to incorporate adjustable processing rates. However, this would conflict with the models' basic hypotheses about information loss or speed of neural processes, the hypotheses that motivated the models.

Summary

The two experiments presented here show that, at least in simple signal detection tasks, there is considerably less difference between older and young subjects in processing speed than has been supposed in prior research. The main reason older subjects are slower than young subjects and have wider distributions of re-

sponse times across conditions is that the older subjects set more conservative response criteria than the young subjects, and older subjects have longer nondecisional components of response time than young subjects. Of course, it is likely that in other rapid, two-choice decision tasks, there is degradation in the quality of the information that enters the decision process for older versus young subjects. For example, there might be decrements in extraction of perceptual information from difficult-to-see stimuli or decrements in information extracted from long-term memory. Along with these decrements, however, our data suggest that much of the slow down observed in response times for older subjects will come from more conservative response criteria.

References

- Allen, P. A., Ashcraft, M. H., & Weber, T. A. (1992). On mental multiplication and age. *Psychology and Aging, 7*, 536–545.
- Allen, P. A., Madden, D. J., Weber, T. A., & Groth, K. E. (1993). Influence of age and processing stage on visual word recognition. *Psychology and Aging, 8*, 274–282.
- Audley, R. J., & Pike, A. R. (1965). Some alternative stochastic models of choice. *The British Journal of Mathematical and Statistical Psychology, 18*, 207–225.
- 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: Charles C Thomas.
- Busmeyer, 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.
- Cerella, J. (1985). Information processing rates in the elderly. *Psychological Bulletin, 98*, 67–83.
- Cerella, J. (1990). Aging and information-processing rate. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., pp. 201–221). San Diego, CA: Academic Press.
- Cerella, J. (1991). Age effects may be global, not local: Comment on Fisk and Rogers (1991). *Journal of Experimental Psychology: General, 120*, 215–223.
- Cerella, J. (1994). Generalized slowing in Brinley plots. *Journal of Gerontology: Psychological Sciences, 49*, 65–71.
- Espinoza-Varas, B., & Watson, C. (1994). Effects of decision criterion on response latencies of binary decisions. *Perception and Psychophysics, 55*, 190–203.
- Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Cognitive slowing, cognitive speed, and response latency. *Psychological Bulletin, 125*, 777–799.
- Fisher, D. L., & Glaser, R. A. (1996). Molar and latent models of cognitive slowing: Implications for aging, dementia, depression, development, and intelligence. *Psychonomic Bulletin & Review, 3*, 458–480.
- Fisk, A. D., & Fisher, D. L. (1994). Brinley plots and theories of aging: The explicit, muddled, and implicit debates. *Journal of Gerontology: Psychological Sciences, 49*, 81–89.
- 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.
- Hale, S., & Jansen, J. (1994). Global processing-time coefficients characterize individual and group differences in cognitive speed. *Psychological Science, 5*, 384–389.
- Hale, S., Myerson, J., & Wagstaff, D. (1987). General slowing of nonverbal information processing: Evidence for a power law. *Journal of Gerontology, 42*, 131–136.
- Hanes, D. P., & Schall, J. D. (1996, October 18). Neural control of voluntary movement initiation. *Science, 274*, 427–430.
- Hartley, A. A. (1992). Attention. In F. Craik & T. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 3–49). Hillsdale, NJ: Erlbaum.
- Hertzog, C. (1992). Aging, information processing, and intelligence. In K. W. Schaie (Ed.), *Annual review of gerontology and geriatrics* (Vol. 11, pp. 55–79). New York: Springer.
- Hertzog, C., Vernon, M. C., & Rypma, B. (1993). Age differences in mental rotation task performance: The influence of speed/accuracy tradeoffs. *Journal of Gerontology, 48*, P150–P156.
- Kaufman, A. (1990). *Assessing adolescent and adult intelligence*. Boston, MA: Allyn and Bacon.
- LaBerge, D. A. (1962). A recruitment theory of simple behavior. *Psychometrika, 27*, 375–396.
- Laming, D. R. J. (1968). *Information theory of choice reaction time*. New York: Wiley.
- Lee, W., & Janke, M. (1964). Categorizing externally distributed by stimulus samples for three continua. *Journal of Experimental Psychology, 68*, 376–382.
- Lima, S. D., Hale, S., & Myerson, J. (1991). How general is general slowing? Evidence from the lexical domain. *Psychology and Aging, 6*, 416–425.
- Luce, R. D. (1986). *Response times*. New York: Oxford University Press.
- Madden, D. J. (1989). Visual word identification and age-related slowing. *Cognitive Development, 4*, 1–29.
- Madden, D. J., Pierce, T. W., & Allen, P. A. (1992). Adult age differences in attentional allocation during memory search. *Psychology and Aging, 7*, 594–601.
- Maylor, E. A., & Rabbitt, P. M. A. (1994). Applying Brinley plots to individuals: Effects of aging on performance distributions in two speeded tasks. *Psychology and Aging, 9*, 224–230.
- McDowd, J. M., & Craik, F. I. M. (1988). Effects of aging and task difficulty on divided attention performance. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 267–280.
- Meyer, D. E., Irwin, D. E., Osman, A. M., & Kounios, J. (1988). The dynamics of cognition: Mental processes inferred from a speed-accuracy decomposition technique. *Psychological Review, 95*, 183–237.
- Myerson, J., Ferraro, F. R., Hale, S., & Lima, S. D. (1992). General slowing in semantic priming and word recognition. *Psychology and Aging, 7*, 257–270.
- Myerson, J., & Hale, S. (1993). General slowing and age invariance in cognitive processing: The other side of the coin. In J. Cerella, J. Rybash, W. Hoyer, & M. L. Commons (Eds.), *Adult information processing: Limits on loss* (pp. 115–141). San Diego, CA: Academic Press.
- Myerson, J., Hale, S., Wagstaff, D., Poon, L. W., & Smith, G. A. (1990). The information-loss model: A mathematical theory of age-related cognitive slowing. *Psychological Review, 97*, 475–487.
- Myerson, J., Wagstaff, D., & Hale, S. (1994). Brinley plots, explained variance, and the analysis of age differences in response latencies. *Journal of Gerontology: Psychological Sciences, 49*, 72–80.
- Nebes, R. D., & Madden, D. J. (1988). Different patterns of cognitive slowing produced by Alzheimer's disease and normal aging. *Psychology and Aging, 3*, 102–104.
- Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal, 7*, 308–313.
- Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41–82). New York: Halstead Press.
- Perfect, T. J. (1994). What can Brinley plots tell us about cognitive aging? *Journal of Gerontology: Psychological Sciences, 49*, 60–64.
- Radloff, L. S. (1977). The CES-D Scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.
- 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. (1980). A note on modelling accumulation of information when the rate of accumulation changes over time. *Journal of Mathematical Psychology*, 21, 178–184.
- 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. (2001). *A diffusion model account of reaction time and accuracy in a two choice brightness discrimination task*. Manuscript submitted for publication.
- Ratcliff, R., & Murdock, B. B., Jr. (1976). Retrieval processes in recognition memory. *Psychological Review*, 83, 190–214.
- 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 letter identification. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 127–140.
- Ratcliff, R., & Smith, P. L. (2001). *A comparison of sequential sampling models for two-choice reaction time*. Manuscript in preparation.
- 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., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, 106, 261–300.
- Roberts, S., & Pashler, H. (2000). How persuasive is a good fit? A comment on theory testing. *Psychological Review*, 107, 358–367.
- Salhouse, T. A. (1991). *Theoretical perspectives on cognitive aging*. Hillsdale, NJ: Erlbaum.
- Salhouse, T. A., & Somberg, B. L. (1982). Isolating the age deficit in speeded performance. *Journal of Gerontology*, 37, 59–63.
- Smith, G. A., & Brewer, N. (1995). Slowness and age: Speed-accuracy mechanisms. *Psychology and Aging*, 10, 238–247.
- Smith, G. A., Poon, L. W., Hale, S., & Myerson, J. (1988). A regular relationship between old and young adults' latencies on their best, average and worst trials. *Australian Journal of Psychology*, 40, 195–210.
- Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review*, 102, 567–591.
- Smith, P. L., & Vickers, D. (1988). The accumulator model of two-choice discrimination. *Journal of Mathematical Psychology*, 32, 135–168.
- Spieler, D. H., Balota, D. A., & Faust, M. E. (1996). Stroop performance in healthy younger and older adults and in individuals with dementia of the Alzheimer's type. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 461–479.
- Strayer, D. L., & Kramer, A. F. (1994). Strategies and automaticity: I. Basic findings and conceptual framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 318–341.
- Van Zandt, T., Colonius, H., & Proctor, R. W. (2000). A comparison of two response time models applied to perceptual matching. *Psychonomic Bulletin and Review*, 7, 208–256.
- 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. (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.
- Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised*. San Antonio, TX: The Psychological Corporation.

Received April 19, 2000

Revision received October 18, 2000

Accepted October 19, 2000 ■