

A Diffusion Model Analysis of the Effects of Aging in the Lexical-Decision Task

Roger Ratcliff
The Ohio State University

Anjali Thapar
Bryn Mawr College

Pablo Gomez
DePaul University

Gail McKoon
The Ohio State University

The effects of aging on response time (RT) are examined in 2 lexical-decision experiments with young and older subjects (age 60–75). The results show that the older subjects were slower than the young subjects, but more accurate. R. Ratcliff's (1978) diffusion model provided a good account of RTs, their distributions, and response accuracy. The fits show an 80–100-ms slowing of the nondecision components of RT for older subjects relative to young subjects and more conservative decision criterion settings for older subjects than for young subjects. The rates of accumulation of evidence were not significantly different for older compared with young subjects (less than 2% and 5% higher for older subjects relative to young subjects in the 2 experiments).

Across a wide variety of cognitive tasks, research has shown that processing slows with age. For some tasks, especially those like letter discrimination that depend heavily on peripheral processes, this is not surprising (e.g., Thapar, Ratcliff, & McKoon, 2003). However, for other tasks it might be expected that performance would improve with age. One such task is lexical decision, the task of interest in this article. Over a lifetime of 60 to 70 years, the number of encounters with many words must greatly exceed the number of encounters in the first 20 years. Yet despite so many years of practice, lexical-decision response times (RTs) increase with age. For example, Allen, Madden, and Crozier (1991) found average RTs of 800 ms for older adults compared with 500 ms for young adults. Word frequency effects, longer RTs with lower frequency words, are also larger for older adults (see Allen et al., 1991; Allen, Madden, Weber, & Groth, 1993; Allen, Sliwinski, & Bowie, 2002; Lima, Hale, & Myerson, 1991).

Recently, Ratcliff, Gomez, and McKoon (2004) have applied the diffusion model for two-choice decisions (Ratcliff, 1978, 1981, 1985, 1988, 2002; Ratcliff & Rouder, 1998, 2000; Ratcliff & Smith, 2004; Ratcliff, Van Zandt, & McKoon, 1999) to lexical-decision data. The model allows processing to be separated into several components: the rate at which information about the stimulus string of letters accumulates in the decision system (which reflects the goodness of match between the test string and lexical memory), the criteria that determine the amounts of information

that must be accumulated before a decision can be made, nondecision components of processing such as encoding and response execution, and variabilities in the various components. Ratcliff, Gomez, and McKoon applied the model to nine experiments in which word frequency, type of nonword, number of repetitions, and proportions of different types of stimuli were manipulated. The model fit the data well, accounting for the effects of the experimental variables on RTs for correct and error responses, shapes of the RT distributions, and accuracy values. All of the variables affected only the rate of accumulation of evidence and none of the other components of processing.

Our aim for this article is to use the diffusion model to compare lexical-decision performance for older subjects, 60–75 years, and young, college-age, subjects. Two experiments manipulated word frequency and type of nonword. If, following Ratcliff, Gomez, and McKoon's (2004) results, these variables affect only the rates of accumulation of evidence, then it might be expected that the older subjects, given their many years of experience with words, would have larger rates of accumulation of evidence than the young subjects. Longer RTs for the older subjects relative to the young subjects might come from slower nondecision components of processing and from more conservative decision criteria, as has been found in earlier studies (Ratcliff, Thapar, Gomez, & McKoon, 2004; Ratcliff, Thapar, & McKoon, 2001, 2003; Thapar et al., 2003).

Across several tasks, including a brightness discrimination task, signal detection-like tasks, a recognition memory task, and a letter discrimination task (following Ratcliff, Thapar, et al., 2004; Ratcliff et al., 2001, 2003; Thapar et al., 2003), older subjects have consistently adopted more conservative decision criteria than did young subjects and they have been 40 to 100 ms slower in the nondecision components of processing. In brightness discrimination, signal detection using arrays of asterisks with a large/small numerosity judgment, and recognition memory, their rates of accumulation of evidence have not been significantly different than

Roger Ratcliff and Gail McKoon, Department of Psychology, The Ohio State University; Anjali Thapar, Department of Psychology, Bryn Mawr College; Pablo Gomez, Department of Psychology, DePaul University.

Preparation of this article was supported by National Institute on Aging Grant R01-AG17083, and National Institute of Mental Health Grants R37-MH44640 and K05-MH01891.

Correspondence concerning this article should be addressed to Roger Ratcliff, Department of Psychology, The Ohio State University, 1827 Neil Avenue, Columbus, OH 43210.

young subjects'. However, when visual acuity was critical, in masked letter discrimination, their rates show a significant deficit. This deficit for a task for which high spatial frequency visual information is necessary, but not for tasks requiring only low spatial frequencies, is consistent with findings in the psychophysical literature on aging (Coyne, 1981; Fozard, 1990; Owsley, Sekuler, & Siemsen, 1983; Spear, 1993).

In another study comparing lexical-decision performance across populations of subjects, Ratcliff, Perea, Coleangelo, and Buchanan (in press) reanalyzed lexical-decision data collected by Moreno, Buchanan, and Van Orden (2002) from aphasic patients and normal control subjects. There were large RT differences between the two groups, but when the diffusion model was fit to the data, it was found that the rates of accumulation of evidence did not differ significantly between the aphasic patients and the normal subjects (except for two severely impaired aphasic patients). The large differences in RTs between the groups were accounted for by the aphasic patients' much slower nondecision components of RT and their more conservative decision criteria. This pattern of results is consistent with the view that aphasic patients suffer from a failure of inhibition that leads to difficulties with lexical selection rather than damage to the lexical representations themselves (Buchanan, McEwen, Westbury, & Libben, 2003).

In all of the studies just reviewed, the diffusion model fits the data well, giving a complete account of correct and error RTs, RT distributions, and accuracy values. In giving an account that includes both RTs and accuracy, the model provides an interpretation of RT effects of whether accuracy is the same for older and young subjects or different. It also allows differences in criteria settings between older and young subjects to be factored out of interpretations of differences in performance. In contrast, under the more common approach to the effects of aging on processing, exemplified by Myerson, Ferraro, Hale, and Lima's (1992) meta-analysis of a number of lexical-decision experiments with priming manipulations, experimental results are interpreted only in terms of correct mean RTs. Averaging over a large number of lexical-decision experiments (22), Myerson et al. found that error rates were about the same for older and young subjects. However, there were large RT differences—older subjects' lexical-decision RTs were 250 to 300 ms longer than those of young subjects'. By applying the diffusion model to these data, it should be possible to determine whether the similar accuracy values indicate similar rates of accumulation of lexical evidence or slower rates of accumulation offset by more conservative response criteria.

The Diffusion Model

The diffusion model is a model of the processes involved in making simple two-choice decisions. It separates the rate of evidence accumulation in the decision process from the decision criteria and from nondecision components of processing. The model is designed to apply only to relatively fast two-choice decisions and to decisions that are composed of a single-stage decision process (as opposed to the multiple-stage decision processes that might be involved in, for example, reasoning tasks or card-sorting tasks). As a rule of thumb, the model would not be applied to experiments in which mean RTs are much longer than about 1 to 1.5 s. Models in the class of diffusion models have also been applied to decision making (Busemeyer & Townsend, 1993;

Roe, Busemeyer, & Townsend, 2001) and simple reaction time (Smith, 1995). A comprehensive comparison between the diffusion model, other diffusion models, and other models of the sequential sampling class is presented in Ratcliff and Smith (2004). This comparison includes fits of the various models to a lexical-decision experiment presented in Wagenmakers, Ratcliff, Gomez, and McKoon (2004). Some of these other models might also apply to the data from the experiments presented here.

The model assumes that decisions are made by a noisy process that accumulates information over time from a starting point toward one of two response criteria or boundaries, as in Figure 1, where the starting point is labeled z and the boundaries are labeled a and 0 . When one of the boundaries is reached, a response is initiated. The rate of accumulation of information is called the "drift rate" (v), and it is determined by the quality of the information extracted from the match between the test word and lexical memory. In other words, word knowledge is represented in the drift rate of the model. For example, a high-frequency word will produce a high match and therefore a high drift rate, while a low-frequency word will produce a low match and a low drift rate. Within each trial, there is noise (variability) in the process of accumulating information so that processes with the same mean drift rate do not always terminate at the same time (producing RT distributions) and do not always terminate at the same boundary (producing errors). This source of variability is called "within-trial" variability. Empirical RT distributions are positively skewed, and this is naturally predicted by simple geometry (see Ratcliff & Rouder, 1998, Figure 1).

Components of processing are assumed to be variable across trials, and the assumption of such variability allows the model to

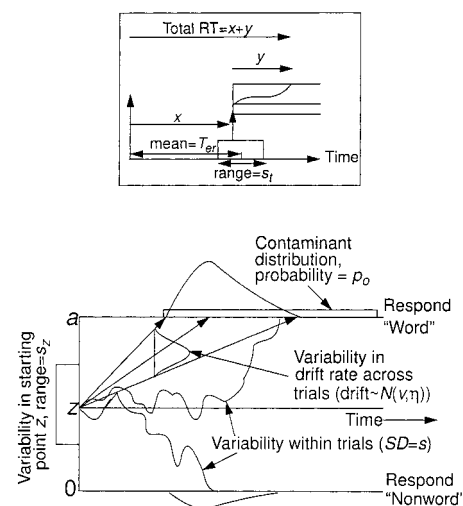


Figure 1. An illustration of the diffusion model. The top panel illustrates the total processing time as the sum of the decision time y and the nondecision time x , T_{er} = mean of the nondecision component of response time (RT), and s_r = range of the distribution of nondecision times across trials (from a uniform distribution). Parameters of the decision process are shown in the bottom panel: a = boundary separation, z = starting point, η = standard deviation in drift across trials (from a normal distribution), s_z = range of the distribution of starting point (z) across trials (from a uniform distribution), v = drift rate, p_o = proportion of contaminants, and s = standard deviation in drift within trials.

account for differences in RTs between correct and error responses (Luce, 1986). Variability in drift rate across trials leads to slow errors, and variability in starting point leads to fast errors (Ratcliff & Rouder, 1998; Ratcliff et al., 1999). Drift rate is assumed to be normally distributed with standard deviation η which means that nominally equivalent stimuli vary in their match to lexical memory from trial to trial (cf. signal detection theory), and starting point is assumed to be uniformly distributed with range s_z .

Besides the decision process, there are nondecision components of processing such as encoding and response execution. These are combined in the diffusion model into one component with mean T_{er} (see the inset in Figure 1). Like drift rate and starting point, the nondecision component of processing is assumed to have variability across trials, and it is assumed to be uniformly distributed with range s_r . The effect of variability in the nondecision component of processing depends on the mean value of drift rate (Ratcliff & Tuerlinckx, 2002). With a large value of mean drift rate, variability acts to shift the leading edge of the RT distribution shorter than it would otherwise be (by as much as 10% of s_r). With smaller values of drift rate, the effect is smaller. The standard deviation in the distribution of the nondecision component of processing is typically less than one quarter of the standard deviation in the decision process, so the combination of the two (convolution) will have little effect on distribution shape and on the standard deviation in the distribution predicted from the decision process. For example, if $s_r = 100$ ms ($SD = 28.9$ ms) and the standard deviation in the decision process is 100 ms, the combination (square root of the sum of squares) is 104 ms. With variability in the nondecision component of processing, Ratcliff and Tuerlinckx showed that the diffusion model could fit data with considerable variability in .1 quantile RTs across experimental conditions.

Variability in the nondecision component of processing (represented by the parameter s_r) also plays an important role in fits of the model to lexical-decision data (see Ratcliff, Gomez, & McKoon, 2004). An important result obtained by Balota and Spieler (1999) is that there is a relatively large shift in the leading edge of the lexical-decision RT distribution as a function of word frequency for correct responses. The leading edge of the distribution for high-frequency words is about 30 ms shorter than the leading edge for low-frequency words. This shift is larger than would be expected from the diffusion model if there were no variability in the nondecision component of processing. Without variability in the nondecision component of processing, the model can accommodate only 19 ms of the 30-ms leading edge shift (assuming parameter values similar to those for the fits of the diffusion model presented later). Ratcliff, Gomez, and McKoon replicated Balota and Spieler's results and showed that with variability in the nondecision component of RT, the large change in the leading edge of the distribution was fit with only drift rate changing.

In summary, 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, the lower boundary is set to 0, η is the standard deviation in mean drift rate across trials, s_z is the range of the starting point across trials, and s_r is the range of T_{er} across trials. For each different stimulus condition in an experiment, it is assumed that the rate of accumulation of evidence is different, so each has a different value of drift, v . Within-trial variability in drift rate (s) is a scaling parameter for the diffusion process (i.e., if it

were doubled, other parameters could be multiplied or divided by two to produce exactly the same fits of the model to data).

Experiments 1 and 2

The experiments used a standard lexical-decision procedure. On each trial, a single letter string was presented and subjects were asked to decide whether it was a word or a nonword. In Experiment 1, the nonwords were pronounceable; in Experiment 2, they were random strings of letters. In both experiments, there were high-, low-, and very low-frequency words.

Method

Subjects. Fifty-four young adults (36 women and 18 men) and 44 older adults (27 women and 17 men) participated in Experiment 1, and 54 different young adults (38 women and 16 men) and 40 different older adults (26 women and 14 men) participated in Experiment 2. Subjects were tested individually for one practice session of about 20 min that followed collection of data for the inclusion criteria described next. They returned within a week for the experimental session.

The young adults were college students who participated for course credit in an introductory psychology course at Northwestern University. The older adults were healthy, active, community-dwelling individuals age 60–75 years old living in Evanston, in the nearby suburbs or in the suburbs adjacent to Bryn Mawr. The older subjects were paid \$15 for a 1-hr session for their participation. Participants had to meet the following inclusion criteria to participate in the study: a score of 26 or above on the Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975); a score of 15 or less on the Center for Epidemiological Studies—Depression Scale (CES-D; Radloff, 1977); and no evidence of disturbances in consciousness, medical or neurological disease causing cognitive impairment, head injury with loss of consciousness, or current psychiatric disorder. The means and standard deviations for standard background characteristics (including a Wechsler Adult Intelligence Scale—Revised [Wechsler, 1997] IQ estimate) are presented in Table 1.

Stimuli. The stimuli were high-, low-, and very low-frequency words and nonwords. There were 800 high-frequency words with frequencies from 78 to 10,600 per million ($M = 325$, $SD = 645$; Kucera & Francis, 1967); 800 low-frequency words, with frequencies of 4 and 5 per million ($M = 4.41$, $SD = 0.19$); and 741 very low-frequency words, with frequen-

Table 1
Subject Characteristics

Experiment and measure	Older Adults		Young Adults	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1				
Age	68.53	4.83	19.78	1.21
Years of education	16.36	2.77	13.12	1.11
MMSE	28.73	1.22	29.00	0.80
IQ estimate (WAIS-R)	113.37	10.65	117.31	10.52
CES-D	7.81	4.32	8.76	3.95
2				
Age	67.23	4.72	20.18	1.67
Years of education	16.04	2.62	13.58	1.55
MMSE	28.88	0.91	29.04	1.21
IQ estimate (WAIS-R)	115.59	12.97	116.69	11.91
CES-D	8.12	3.81	9.36	4.03

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

cies of 1 per million or no occurrence in the Kucera and Francis's corpus ($M = 0.365$, $SD = 0.48$). All of the very low-frequency words did occur in the *Merriam-Webster's Ninth New Collegiate Dictionary* (Merriam-Webster, 1990), which were then screened by three Northwestern undergraduate students; any words they did not know were eliminated. Stimuli were chosen randomly without replacement from the pools.

From each word, a pseudoword was generated by randomly replacing all of the vowels with other vowels (except for *u* after *q*), giving a pool of 2,341 nonwords, none of which were words. This pool was used for Experiment 1. For Experiment 2, there was a pool of 2,400 random letter strings, created by randomly sampling letters from the alphabet and then by removing those strings that were pronounceable. These pools are described in more detail in Ratcliff, Gomez, and McKoon (2004).

Apparatus. Stimuli were presented on a Pentium II class machine, and responses were collected on the keyboard.

Procedure. The practice session consisted of 30 blocks each with 30 lexical-decision trials (taking about half an hour), and the experimental session consisted of 70 blocks of 30 lexical-decision trials (taking about 50 min). Subjects were given a chance to rest between each block of trials that was self-paced.

A block of trials consisted of 5 high-, 5 low-, and 5 very low-frequency words and 15 nonwords. Each stimulus string of letters was displayed on the PC screen until the subject responded by pressing the / key for a word and the z key for a nonword. If a response was correct, there was a 150-ms pause and then the next trial began; if a response was incorrect, the word *ERROR* was displayed for 750 ms, then erased, and the next trial was presented 50 ms later. Subjects were instructed to respond quickly and accurately.

Results

In the data analyses for Experiment 1, RTs smaller than 300 ms and greater than 3,000 ms were eliminated for young subjects (less than 1.5% of the data), and RTs smaller than 350 ms and greater than 4,000 ms were eliminated for older subjects (less than 1.5% of the data). In Experiment 2, the cutoffs were the same but the proportion of RTs eliminated was less than 0.5% for both older and young subjects. Further discussion of outliers and contaminants is presented in the *Fitting the Diffusion Model to the Data* section.

Many subjects made few errors, so in order to obtain reliable estimates of RT distribution shapes for error responses, we grouped subjects with similar performance in terms of RT to form supersubjects. This is an appropriate procedure for fitting the diffusion model, because in other experiments (Ratcliff, Thapar, et al., 2004; Ratcliff et al., 2001, 2003; Thapar et al., 2003) fits to average data provided similar parameter values to parameter values averaged across fits to individual subjects. Because averaging over widely different levels of RT is more likely to produce discrepancies in fits to data than grouping over widely different levels of accuracy, we grouped subjects into supersubjects on the basis of correct RT for nonwords. Nonwords were used because their estimate had a lower standard deviation—there were three times as many observations as for each word category. For Experiment 1, there were 11 older and 12 young supersubjects; for Experiment 2, there were 10 older and 10 young supersubjects, each supersubject was made up of 2 to 6 individual subjects. For each of the supersubjects, there were at least six error RTs in each condition.

A summary of the results is shown in Table 2. For correct responses to words, high-frequency words had shorter RTs and higher accuracy than low-frequency words, which had shorter RTs

Table 2
Summary of Data From Experiments 1 and 2

Experiment, age group, and condition	Accuracy	Mean correct RT	Mean error RT	No. of observations
1. Old				
HF words	.992	815	819	14,808
LF words	.972	925	1,123	14,703
VLF words	.911	1,015	1,172	14,574
Pseudowords	.975	969	1,137	43,938
1. Young				
HF words	.970	590	577	18,373
LF words	.905	664	688	18,357
VLF words	.811	712	755	18,334
Pseudowords	.929	724	690	54,939
2. Old				
HF words	.993	730	699	13,958
LF words	.987	788	947	13,935
VLF words	.970	855	1,065	13,871
Random letter strings	.989	751	993	41,703
2. Young				
HF words	.966	574	513	18,789
LF words	.953	614	569	18,777
VLF words	.939	644	657	18,770
Random letter strings	.966	590	626	56,300

Note. *SEs* for older subjects are in the range 20–45 ms for error responses and 2–5 ms for correct responses. *SEs* for young subjects are in the range 5–11 ms for error responses and 1–3 ms for correct responses. RT = response time; HF = high-frequency words; LF = low-frequency words; VLF = very low-frequency words.

and higher accuracy than very low-frequency words. Accuracy for nonwords was higher, and RTs were shorter when the nonwords were random letter strings (Experiment 2) than when they were pseudowords (Experiment 1).

The older subjects were slower than the young subjects: Their correct RTs to words were longer by 150 to 300 ms, and their correct RTs to nonwords were longer by 150 to 250 ms (the differences in error RTs were even greater). Error rates were about twice as large for the young subjects as for the older subjects (cf. Allen et al., 1991; Allen, Sliwinski, & Bowie, 2002).

Mean error RTs were sometimes shorter than correct RTs—and sometimes longer. However, the behavior of these means has to be taken with a grain of salt. For some young supersubjects in Experiment 1, for example, errors to very low-frequency words were 40 ms shorter than correct RTs; for other young supersubjects, they were 200 ms longer. Generally, for young subjects, the faster supersubjects had errors faster than correct responses while the slower supersubjects had errors slower than correct responses (see similar results in Ratcliff, Gomez, & McKoon, 2004). For older supersubjects, error responses were almost always slower than correct responses.

There were large individual differences across supersubjects. For example, in Experiment 1, the mean RTs for correct responses for young supersubjects varied from 510 ms to 870 ms and their accuracy varied from .83 to .94. There was similar variability across the older subjects. Our strategy for fitting the diffusion model was to fit it both to the data averaged across the supersubjects and to the data for each supersubject, comparing the parameters for the fit to the average data to the average parameter values for the fits to the individual supersubjects.

There were differences in speed–accuracy relationships across young and older supersubjects: For the young supersubjects, the fastest were the least accurate and the slowest were the most accurate (e.g., for Experiment 1, average accuracy for the fastest was .87 and average accuracy for the slowest was .93). In contrast, for the older supersubjects, the fastest were no less accurate than the slowest, except for 1 supersubject (2 subjects) with lower accuracy (the accuracy values were .96, .96 accuracy, and .89 for these three sets of supersubjects, respectively). These results suggest that the young subjects were more likely to trade speed for accuracy than were the older subjects. This might be a characteristic of young subjects in general, or it might be a characteristic of young subjects participating in experiments for course credit in introductory psychology subject pools.

Brinley Plots

A procedure that has been standard for examining RTs in aging research is to plot older subjects' RTs for each experimental condition against young subjects' RTs for the same conditions to produce what is called a "Brinley plot" (Brinley, 1965). The typical result has been that the plot is a straight line with a slope greater than one. This has been taken as evidence for a decrease in the speed of processing with age. This view has received much support (e.g., Birren, 1965; Cerella, 1985, 1990; Cerella, Poon, & Williams, 1980; Fisk & Warr, 1996; Salthouse, 1985, 1996; Salthouse, Kausler, & Sauls, 1988) and much criticism and debate (Allen, Ashcraft, & Weber, 1992; Allen et al., 1993; Cerella, 1994; Fisk & Fisher, 1994; Fisher & Glaser, 1996; Hartley, 1992; Hertzog, 1992; Lima et al., 1991; Madden, 1989; Madden, Pierce, & Allen, 1992; Myerson et al., 1992; Myerson, Wagstaff, & Hale, 1994; Perfect, 1994).

Ratcliff, Spieler, and McKoon (2000) argued against this view from a theoretical perspective by showing that in the frameworks of explicit models, there are multiple ways that slowing in mean RT can be produced, such as changes in the rate of accumulation of evidence, decision criteria settings, or both. The fits of the diffusion model to the data from four experiments (Ratcliff, Thapar, et al., 2004; Ratcliff et al., 2001, 2003; Thapar et al., 2003) showed that sometimes decision criteria settings alone are responsible for the increase in RT, and sometimes there is also a decrement in the rate of accumulation of evidence. See Myerson, Adams, Hale, and Jenkins (2003) and Ratcliff, Spieler, and McKoon (in press) for further discussion of these issues.

Although in models like Ratcliff's diffusion model Brinley plots can be produced from any of several different mechanisms, and therefore are not theoretically constraining (Ratcliff et al., 2000), we present the plots here for the data from our experiments to show that our results provide linear functions consistent with the results from previous studies. Figure 2 shows data and fitted straight lines for Experiments 1 and 2. In each case, the mean RTs for correct responses for older subjects are plotted against the mean RTs for young subjects for words of each frequency class. Figure 2 also shows error bars representing two standard errors in the mean RTs. The straight lines lie within confidence regions of the data points, so they are reasonably described by linear functions (see Ratcliff, Spieler, & McKoon, in press). For Experiment 1, with pseudowords as the nonwords, the slope is 1.40; for Experiment 2, with random letter strings, it was significantly

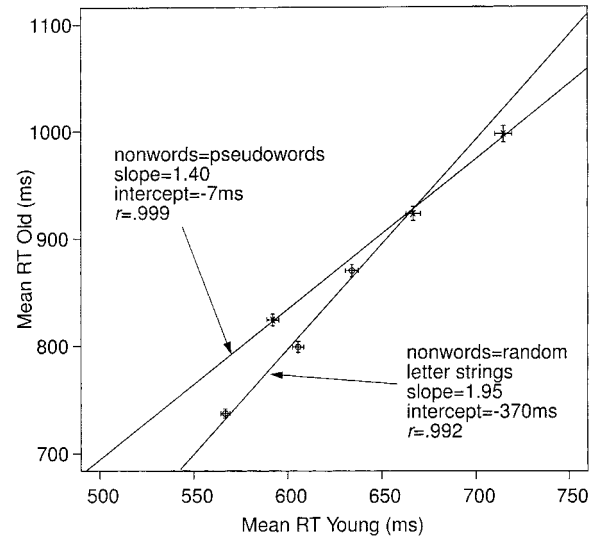


Figure 2. Brinley plots for mean correct response time (RT) for Experiments 1 and 2. The points on the graph represent the high-, low-, and very low-frequency word conditions for older and young subjects. Straight lines are fitted to the means for Experiments 1 and 2 separately, and error bars are 2 *SEs* in mean RT.

different, 1.95 (as can be seen from the two standard error bars presented on the function).

The fact that the slope varies according to whether the nonwords are pseudowords or random letter strings illustrates one of the problems with the slowing hypotheses derived from Brinley plot analyses: It would not be expected that the amount of cognitive slowing for older subjects relative to young subjects in lexical access differs as a function of the type of nonword used in the experiment.

Quantile Probability Functions

One of the main arguments for the use of quantitative models is that providing a complete explanation of processing in a task requires accounting for all aspects of the experimental data. A model that deals with only correct mean RTs is incomplete because it does not deal with accuracy, error RTs, or RT distributions. Such a model—if extended to make predictions about accuracy, error RTs, and RT distributions—is almost guaranteed to make incorrect predictions. To fully test the diffusion model, it is simultaneously fit to all aspects of the data. Plotting all of these aspects of the data separately would make their relative behaviors difficult to grasp, so we display the data in quantile probability functions (Figures 3 and 4).

In quantile probability functions, response probabilities are plotted against quantile RTs. The probability of a response for a particular stimulus type determines the position of a point on the *x*-axis, and the quantile RTs for that stimulus type determine position on the *y*-axis. Specifically, in Figures 3 and 4, the .1, .3, .5 (median), .7, and .9 quantiles are plotted for each of the four experimental conditions: three values of word frequency and nonwords. There are separate plots for word responses and nonword responses for each experiment. Because all of the response prob-

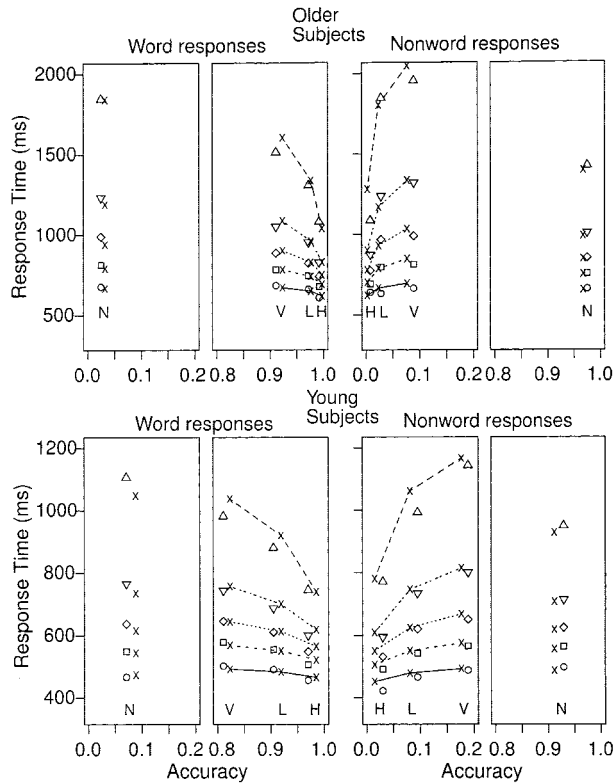


Figure 3. Quantile probability plots for Experiment 1. The lines and x symbols represent the theoretical fits of the diffusion model, and the triangles, diamonds, squares, and circles represent the empirical quantile response times (RTs). The lines in order from the bottom to the top are for the .1, .3, .5, .7, and .9 quantile RTs. The symbols H, L, V, and N below the .1 quantile RTs identify the high-, low-, and very low-frequency word conditions and the nonword condition, respectively.

abilities are above .8 (for correct responses) and below .2 (for error responses), we have only plotted the parts of the functions from 0 to .2 and from .8 to 1.0. The symbols other than x s represent the data points and the lines, and x s are the best fitting functions from the diffusion model, which is discussed later.

Quantile probability functions are used to provide a summary picture of the shapes of the RT distributions. The .1 quantile RTs for correct responses change by less than 50 ms across the various levels of accuracy for both young and older subjects. In contrast, the .9 quantile RTs change by up to several hundred milliseconds. Across all conditions, the change in mean error RT is mainly reflected in the RT distribution spreading rather than shifting.

The results show the expected patterns for the effects of word frequency and type of nonword (cf. Ratcliff, Gomez, & McKoon, 2004) on RT distributions and accuracy values. The question for the diffusion model is what components of processing are responsible for the effects of age on RT and accuracy.

Fits of the Diffusion Model to the Data

The diffusion model was fit to the experimental data by minimizing a chi-square value with a general SIMPLEX minimization routine that adjusts the parameters of the model to find the param-

eters that give the minimum chi-square value (see Ratcliff & Tuerlinckx, 2002, for a full description of the methods). The data entered into the minimization routine for each experimental condition were the RTs for each of the five quantiles for correct and error responses as well as the accuracy values. The quantile RTs and the diffusion model were used to generate the predicted cumulative probability of a response by that quantile RT. 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 computation, these are the expected values to be compared to the observed proportions of responses between the quantiles (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) that are 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.

Research on fitting the diffusion model to data (Ratcliff & Tuerlinckx, 2002) has found that when long or short outlier RTs are added to simulated data, the chi-square method cannot accurately recover the parameter values that were used to generate the

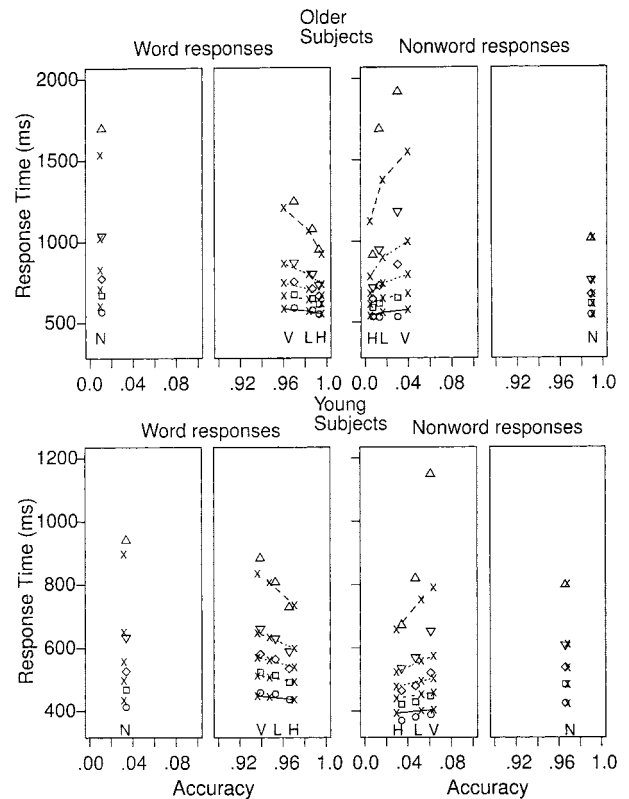


Figure 4. Quantile probability plots for Experiment 2. The lines and x symbols represent the theoretical fits of the diffusion model, and the triangles, diamonds, squares, and circles represent the empirical quantile response times (RTs). The lines in order from the bottom to the top are for the .1, .3, .5, .7, and .9 quantile RTs. The symbols H, L, V, and N below the .1 quantile RTs identify the high-, low-, and very low-frequency word conditions and the nonword condition, respectively.

data. To address this problem, as noted earlier, short outliers (RTs shorter than 300 ms for young subjects and shorter than 350 ms for older subjects) were trimmed out by examining the time at which accuracy began to rise above chance (e.g., Swenson, 1972) and long outliers (responses longer than 3,000 ms for young subjects and 4,000 ms for older subjects here) were also eliminated from analyses. Ratcliff and Tuerlinckx showed that any remaining long contaminant RTs can be explicitly modeled. A parameter (p_o) is added to represent the probability of a contaminant in each condition of the experiment. The contaminant is assumed to come from a uniform distribution that has maximum and minimum values corresponding to the maximum and minimum RTs in the condition. The psychological assumption behind this choice is that subjects are delayed by a random amount of time on some small proportion of trials (e.g., a momentary lapse of attention). For the data reported here, the value of the probability parameter, p_o , was not significantly different for older and young subjects for the two experiments.

For the fits of the model presented here, seven parameters were held constant across the four stimulus conditions (three values of word frequency and nonwords): boundary separation (a), the starting point (z), the mean duration of the nondecision component of processing (T_{er}), the range of the variability in the nondecision component of processing across trials (s_r), the range of the variability in the starting point across trials (s_z), the standard deviation in across-trial variability in drift rate (η), and the probability of contaminant RTs (p_o). Holding these seven parameters constant reflects the assumption that the quality of the information from the stimulus does not differentially affect any of these components of the decision process. Only drift rate can vary to model changes in performance across stimulus conditions. Changes in drift rate move points along the quantile probability function. With the seven constant parameters plus a different value of drift rate for each condition, the model must account for accuracy values, the relative speeds of correct and error responses, and the shapes of the RT distributions for correct and error responses (all the data shown by the quantile probability functions in Figures 3 and 4). With only drift rate changing, the model must account for the changes in

accuracy and distribution shape for both error and correct response conditions as a function of word frequency and nonword.

We fit the diffusion model to the data in two ways. First, the data from each supersubject were fit individually and the parameter values were averaged across supersubjects, and second, the data were averaged over supersubjects and then the model was fit to the averaged data. We could not fit data from individual subjects because there would be too few error RTs in some conditions for many subjects to provide quantiles.

The means for each of the parameter values across supersubjects are shown in Tables 3 and 4 along with their standard deviations. Standard errors in the parameter values (the basis for significance tests additional to those presented next) can be found by dividing the standard deviations by the square root of the number of supersubjects. The parameters from the fits to the averaged data are also shown in Tables 3 and 4. These latter fits were used as the basis for the predictions displayed in Figures 3 and 4 (the lines). Group data have often been used in fitting models, and the assumption (usually implicit) is that the fits and parameter values to the group data will turn out to be the same as the averages from the fits for the individual subjects (other fits of the diffusion model show this to apply for the diffusion model; see Ratcliff, Thapar, et al., 2001, 2003, in press; Thapar et al., 2003). We provide both for comparison. The parameter values obtained from the group data and the average parameter values across the supersubjects were either within two standard errors or close to within two standard errors of each other (see Tables 3 and 4). Also, the parameter values are in the range of parameter values from applications of the diffusion model (Ratcliff, 2002; Ratcliff, Gomez, & McKoon, 2004; Ratcliff & Rouder, 1998, 2000; Ratcliff et al., 1999, 2001).

In general, the model captures the changes in RT distributions for both correct and error responses as well as accuracy values as a function of stimulus condition, with only drift rate changing across conditions. The only systematic misses are in the .9 quantile RTs for errors in some of the conditions.

Analysis of the parameter estimates using t tests and analyses of variance (ANOVAs; .05 significance level, with all t tests using the Welch correction for degrees of freedom) showed that the older

Table 3
Parameter Values and Standard Deviations From Fits of the Diffusion Model

Experiment and group	a	z	T_{er}	η	s_z	p_o	s_r	$M \chi^2$
1. Young (average parameters over supersubjects)	0.127	0.063	0.440	0.101	0.053	0.023	0.164	303.9
1. Young (fit to average data)	0.126	0.062	0.447	0.120	0.069	0.033	0.154	2295.1
1. Young SDs (over supersubjects)	0.030	0.014	0.029	0.050	0.031	0.024	0.042	116.6
1. Older (average parameters over supersubjects)	0.183	0.090	0.539	0.100	0.048	0.057	0.174	202.6
1. Older (fit to average data)	0.184	0.091	0.548	0.097	0.064	0.066	0.177	909.9
1. Older SDs (over supersubjects)	0.043	0.020	0.062	0.042	0.033	0.049	0.097	119.7
2. Young (average parameters over supersubjects)	0.130	0.072	0.399	0.097	0.072	0.038	0.149	172.4
2. Young (fit to average data)	0.124	0.069	0.398	0.084	0.065	0.044	0.141	1334.8
2. Young SDs (over supersubjects)	0.029	0.016	0.024	0.023	0.014	0.028	0.032	64.5
2. Older (average parameters over supersubjects)	0.175	0.094	0.476	0.091	0.040	0.065	0.145	71.3
2. Older (fit to average data)	0.166	0.090	0.480	0.111	0.042	0.063	0.150	381.2
2. Older SDs (over supersubjects)	0.036	0.020	0.034	0.029	0.024	0.027	0.030	38.5

Note. The χ^2 values are the means across subjects, except for the fit to average data. a = boundary separation; z = starting point; T_{er} = nondecision component of response time; η = standard deviation in drift across trials; s_z = range of the distribution of starting point (z); p_o = proportion of contaminants; s_r = range of the distribution of nondecision times.

Table 4
Parameter Values and Standard Errors From Fits of the Diffusion Model for Drift Rates

Experiment and group	v_{HF}	v_{LF}	v_{VLF}	v_N
1. Young (average parameters over supersubjects)	0.491	0.283	0.182	-0.279
1. Young (fit to average data)	0.481	0.281	0.180	-0.279
1. Young <i>SDs</i> (over supersubjects)	0.086	0.050	0.040	0.054
1. Older (average parameters over supersubjects)	0.470	0.330	0.216	-0.289
1. Older (fit to average data)	0.435	0.287	0.200	-0.268
1. Older <i>SDs</i> (over supersubjects)	0.128	0.088	0.073	0.091
2. Young (average parameters over supersubjects)	0.487	0.383	0.331	-0.355
2. Young (fit to average data)	0.437	0.356	0.300	-0.321
2. Young <i>SDs</i> (over supersubjects)	0.058	0.050	0.034	0.043
2. Older (average parameters over supersubjects)	0.495	0.391	0.316	-0.390
2. Older (fit to average data)	0.465	0.367	0.296	-0.363
2. Older <i>SDs</i> (over supersubjects)	0.068	0.054	0.045	0.056

Note. The subscripts of the drift rates (v) are N = new words, H = high-frequency words, L = low-frequency words, and V = very low-frequency words.

subjects differed from the young subjects in several ways. First, the value of the nondecision component of RT was larger for older subjects than for young subjects by about 100 ms in Experiment 1, $t(17.84) = 3.63$, and about 80 ms in Experiment 2, $t(17.25) = 3.11$, approximately replicating Ratcliff, Thapar, et al. (2004), Ratcliff et al. (2001, 2003), and Thapar et al. (2003). Second, drift rates were slightly higher for older subjects than for young subjects, by about 4.7%, but this difference was significant in neither Experiment 1, $F(1, 21) = 0.31$, $MSE = 0.0226$, nor Experiment 2, $F(1, 18) = 0.17$, $MSE = 0.0095$. Drift rates differed among stimulus conditions for both Experiment 1, $F(3, 63) = 320.89$, $MSE = 0.00101$, and Experiment 2, $F(3, 54) = 241.82$, $MSE = 0.000408$. (Note that drift rates for nonwords have the opposite sign from drift rates for words, so they were converted to the same sign for the ANOVAs because better performance means more positive drift rates for words and more negative drift rates for nonwords.) Third, both boundary separation and starting point were greater for older subjects than for young subjects in both Experiment 1, $t(17.84) = 3.63$; $t(17.14) = 3.67$, and Experiment 2, $t(17.24) = 3.10$; $t(17.41) = 2.72$. Fourth, neither the standard deviation in drift rate nor the range of the distribution of nondecision components of processing was different for older and young subjects for either Experiment 1, $t(20.90) = -0.057$; $t(13.39) = 0.31$, or Experiment 2, $t(17.36) = -0.48$; $t(17.90) = -0.26$. There was no difference in the range of the distribution of starting points in Experiment 1, $t(20.67) = -0.36$, but young subjects had a larger range in Experiment 2, $t(14.32) = -3.65$. Fifth, the older subjects had a larger proportion of contaminant RTs in Experiment 1, $t(14.17) = 2.14$, and in Experiment 2, $t(17.97) = 2.21$.

In summary, the main differences between the older and young subjects were longer nondecision components of processing and more conservative decision criteria for the older subjects. We had thought that the older subjects might have a greater rate of accumulation of evidence from the stimuli than the young subjects: Their average rate turned out to be slightly (4.7%), although not

significantly, larger. The differences among other parameter values were small or nonsignificant.

We did not examine the consequences of grouping subjects into supersubjects differently than was done here. Larger or smaller numbers of subjects per group might have provided more power for the statistical tests, but it is clear that the difference in drift rates would not have reached significance even if different groupings were used.

Goodness of fit. The average chi-squares across supersubjects were 203 for Experiment 1 and 304 for older and young subjects, respectively; for Experiment 2, they were 71 and 168 for older and young subjects, respectively. The number of degrees of freedom for each experiment was 77 and the critical chi-square value was 98.5. Thus, for supersubjects, we see systematic statistical deviations between theory and data, but for individual subjects, few of the deviations would be significant because the number of observations would be reduced, which would lead to the chi-square values being reduced by approximately dividing by the number of subjects in each supersubject (between 2 and 6).

The chi-square statistic has the property that as the number of observations increases, the power of the test increases, so even the smallest deviation will lead to a significant chi-square if the number of observations increases enough. To illustrate, the chi-square value is the sum over all frequency classes of $(O - E)^2/E$, where O and E are the observed and expected frequencies. Suppose in our computation, the observed and expected probabilities for two adjacent quantiles systematically miss by .1 (e.g., one is .1 and the next is .3 instead of them both being .2). Then the systematic contribution from this miss to the chi-square is $N(.1)^2/.3 + N(.1)^2/.1$, where N is the number of observations in the condition. For this experiment with $df = 77$ and a critical chi-square value of 98.5, if $N = 100$ per condition, the contribution to chi square from this systematic deviation is 13.3, and if $N = 1,750$ per condition (the number of observations for the word conditions for a supersubject with 4 subjects), the contribution to chi-square is 233, a value over two times greater than the critical value. This suggests that if the chi-square statistic is to be used to evaluate goodness of fit, two questions must be asked: First, what size systematic deviations would be considered large enough to make the model unattractive, and second, what would be the contribution of such deviations to the chi-square for different sample sizes? In these experiments, for the average supersubject with 1,750 observations for each word condition, if there were deviations of .03 in each of the expected frequencies between quantile RTs, then summing over the three word conditions, 118 would be added to the chi-square value. If there were similar deviations for the nonword condition (with three times the number of observations), 236 would be added to the chi-square values. Thus, the significant values of chi-square are consistent with maximum systematic deviations of about .03 in cumulative probabilities.

Differences in accuracy between older and young subjects. Error rates for young subjects were about twice as large as error rates for older subjects in both experiments. This difference cannot be attributed to differences in drift rates because the differences in drift rates were small and not significant. Instead, the difference in accuracy comes from differences in criteria settings; older subjects adopt more conservative decision criteria (i.e., larger values of boundary separation a) than do young subjects.

We know that criteria can be adjusted by experimental manipulations. In Ratcliff, Thapar, et al. (2004), Ratcliff et al. (2001, 2003), and Thapar et al. (2003), explicit speed–accuracy manipulations were carried out. Both older and young subjects were able to make changes in RT of several hundred milliseconds with changes in accuracy of only a few percent. Furthermore, Wagenmakers et al. (2004, Experiment 1) performed an experiment with young subjects in which the same speed–accuracy manipulations were performed. They found that mean RT increased from 486, 534, 550, and 535 ms to 624, 709, 761, and 746 ms for high-, low-, and very low-frequency words and pseudowords, respectively. Corresponding changes in accuracy were from .92, .82, .70, and .88 to .98, .94, .84, and .96 for high-, low-, and very low-frequency words and pseudowords, respectively. These values show that young subjects can trade speed for accuracy and that the sizes of the accuracy differences are consistent with the argument that the accuracy differences between older and young subjects in the experiments presented here are the result of criteria settings. Comparisons of the fits to those presented in Wagenmakers et al. (see Ratcliff & Smith, 2004, Table 2) suggest that young subjects in Experiment 1 are operating nearer their settings for accuracy instructions than their settings for speed instructions.

Another way to show how criteria settings affect accuracy is to examine predictions from the model when the settings are varied. We took all of the parameter values for young subjects for Experiment 1 (Tables 3 and 4), except boundary separation, and generated predicted accuracy and RT values by using the boundary separations from the older subjects (the fourth row in Table 3). The accuracy values for high-, low-, very low-frequency words, and nonwords were .999, .975, .902, and .971, respectively. These values are within 1% of those for the older subjects for Experiment 1 (Table 2). On the basis of these results, we conclude that differences in speed–accuracy criterion settings for older versus young subjects are responsible for the lower accuracy for young subjects relative to older subjects in our experiments.

Correlations among parameters, accuracy values, and mean RT. Ratcliff, Thapar, et al. (2004) analyzed the correlations between model parameters and data. We examined the patterns of correlations for young and older subjects for Experiments 1 and 2 in the same way as in Ratcliff, Thapar, et al. The differences in accuracy among supersubjects for Experiment 2 were too small to provide any variability, so we present the analyses for Experiment 1 only. We averaged the mean RTs and accuracy values for correct responses and the absolute values of the drift rates over all conditions (three frequency classes of words and nonwords). We then computed the correlations for young and older subjects separately and averaged these together to produce the mean correlations that are presented in Table 5. The results are similar to those in Ratcliff, Thapar, et al.: accuracy and mean RT are uncorrelated; boundary separation and drift rate are moderately negatively correlated; boundary separation and mean RT are strongly correlated; drift rate and mean RT are moderately negatively correlated; accuracy and drift rate are moderately correlated (this effect is smaller than in Ratcliff, Thapar, et al., possibly partly because the supersubject grouping was performed on RT and not accuracy); and the parameter for the nondecision components of processing, T_{er} , is correlated only with mean RT. These results indicate that, averaged across all of the conditions of the experiment, RT was determined mainly by the boundary separation that subjects adopted. More

Table 5
Correlations Between Mean Response Time (RT), Accuracy, and Model Parameters Averaged Over Young and Older Supersubjects for Experiment 1

Variable	1	2	3	4	5
1. a	—				
2. T_{er}	0.35	—			
3. v	−0.41	0.13	—		
4. Accuracy	0.30	0.25	0.37	—	
5. M RT	0.91	0.54	−0.56	0.11	—

Note. The correlations were computed for young and older supersubjects separately, and they were averaged. a = boundary separation; T_{er} = mean of the nondecision component of RT; v = drift rate.

conservative subjects produced longer RTs, and less conservative subjects produced shorter RTs. RT was not strongly determined by drift rate; that is, faster subjects did not have much better lexical information than slower subjects. However, accuracy was determined strongly by drift rate, that is, by whether lexical information was good or bad. Accuracy was more weakly correlated with boundary separation.

General Discussion

From the joint accuracy and RT data for both correct and error responses, the fits of the diffusion model extract values of the parameters for the several components of processing. The results show, first, that the nondecision component of RT was 80–100 ms longer for the older subjects than for the young subjects and, second, that the older subjects adopted higher, more conservative decision criteria settings. Ratcliff, Thapar, et al. (2004), Ratcliff et al. (2001, 2003), and Thapar et al. (2003) have shown that although older subjects' criteria settings are generally more conservative, it is also the case that they can vary their decision criteria considerably. We speculate that older subjects adopt more conservative criteria in most tasks because in some tasks, the information they extract from stimuli is relatively impoverished (Thapar et al., 2003). Whether older subjects can set their criteria as low as young subjects when asked to respond as quickly as possible is an open question.

Drift rates for the older and young subjects were about the same (the average difference, 0.017, which is about 4.7%, was nonsignificant). Drift rates measure the quality of the match between a test string of letters and lexical memory, and the finding that there is little difference between older and young subjects suggests that the older subjects, despite their years of practice and large differences in accuracy (Table 1), differed little in the quality of match from young subjects.

The results obtained in the experiments reported here for older versus young subjects are similar to those obtained in a signal detection task (Ratcliff et al., 2001), a recognition memory task (Ratcliff, Thapar, et al., 2004), a brightness discrimination task (Ratcliff et al., 2003), and a letter discrimination task with masking (Thapar et al., 2003). In all of these studies, older subjects were slower than young subjects by about 40 to 100 ms in the nondecision components of RT. In most conditions of most of the experiments, older subjects adopted more conservative decision

criteria than did young subjects. In all of the experiments, except masked letter discrimination, older subjects had about the same drift rates as young subjects. The difference between the two perceptual tasks—letter discrimination and brightness discrimination—is consistent with the literature on letter identification when using accuracy and threshold measures (Spear, 1993). In that literature, there is a deficit as a function of age for high spatial frequency stimuli (e.g., letters), but no deficit with low spatial frequency stimuli (e.g., brightness patches).

The same general differences that were found here between older and young subjects were found by Moreno et al. (2002) between aphasic patients and control subjects. The aphasic patients' lexical decisions were slower than the control subjects' and about equally accurate. Fitting the diffusion model to three groups of supersubjects (Ratcliff, Perea, et al., in press) showed that the aphasic patients were slower than the control subjects in the nondecision components of processing with values 577, 594, and 815 ms compared with 470, 459, and 481 ms for the control subjects (compared with a mean of 539 ms for the older subjects here). Also, the patients set their decision criteria much farther apart than did the control subjects, with values of a at 0.283, 0.233, and 0.393 compared with 0.147, 0.148, and 0.196 for the control subjects (compared with a mean of 0.183 for the older subjects here). However, although the drift rates were somewhat lower for the patients compared with the controls, drift rates for individual subjects overlapped between the two groups.

Semantic Priming Effects

One of the topics that has been studied a great deal in the aging literature is semantic priming in lexical decision. The decrease in RT that occurs when a target test word is preceded by a related word relative to an unrelated word is larger for older subjects than for young subjects. Summarizing over 22 experiments, Myerson et al. (1992) found that the average ratio of the priming effect for older to young subjects was 1.45 based on averaging mean RTs over experiments and 1.64 based on averaging the ratios for each experiment. There was also high variability in the ratio across experiments. Laver and Burke (1993) found that the semantic priming effect was constant as a function of baseline, but later analyses that separated lexical decision from other tasks confirmed the Myerson et al. findings (Hale & Myerson, 1995; Myerson, Hale, Chen, & Lawrence, 1997).

In the diffusion model, the simplest assumption about priming effects is that they result from a difference in drift rate, with the drift rate greater for primed than for unprimed test words. The preference of some of the authors of this article is to interpret priming effects in the framework of compound cue models (McKoon & Ratcliff, 1992; McNamara, 1992, 1994; Ratcliff & McKoon, 1988, 1994). The compound cue that determines drift rate would be the sum of the weighted products of the familiarity of the target word and the familiarity of the prime to all items in memory (see examples in Ratcliff & McKoon, 1988). In the unrelated condition, the prime and target would not share associates, so the sum of the products would be low; in the related condition, the prime and target would share associates, so the products for these shared associates would be large and their sum would be larger than that for the unrelated condition.

Assuming that priming effects are produced by increased drift rates, we used the parameter values for the diffusion model for Experiments 1 and 2 to examine the ratio of the sizes of priming effects for older relative to young subjects. We assumed a drift rate of 0.4 for primed test words and a drift rate of 0.3 for unprimed test words. We used the same two values for both older and young subjects because there were no significant differences in their drift rates in Experiments 1 and 2. All of the other parameter values were set to the values obtained in the experiments. RTs were generated from the model with these parameter values, and priming effects were then computed. The ratio of the priming effects for older to young subjects was 1.88 for Experiment 1 and 1.68 for Experiment 2. Both ratios fall in the range of those reported by Myerson et al. (1992). These results show that holding drift rates constant between the older and young subjects while the older subjects have more conservative decision criteria produces a ratio of priming effects in the typical range that has been obtained experimentally. The more conservative decision criteria are all that is needed to produce the right scaling behavior to give the larger priming effects for older subjects.

In summary, for lexical decision, like a variety of other two-choice tasks, the diffusion model provides a method for decoupling the rate of extraction of information from stimuli from criterion effects and from the nondecision components of RT. This allows hypotheses about the information extracted from stimuli to be separated from hypotheses about subject-adjustable decision criteria, which is in contrast to monolithic accounts of processing speed in terms of only mean correct RT.

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., & Crozier, L. C. (1991). Adult age differences in letter-level and word-level processing. *Psychology and Aging, 6*, 261–271.
- 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.
- Allen, P. A., Sliwinski, M., & Bowie, T. (2002). Differential age effects in semantic and episodic memory: Part II. Slope and intercept analyses. *Experimental Aging Research, 28*, 111–142.
- Balota, D. A., & Spieler, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General, 128*, 32–55.
- Birren, J. E. (1965). Age-changes in speed of behavior: Its central nature and physiological correlates. In A. T. Welford & J. E. Birren (Eds.), *Behavior, aging, and the nervous system* (pp. 191–216). Springfield, IL: Thomas.
- Brinley, J. F. (1965). Cognitive sets, speed and accuracy of performance in the elderly. In A. T. Welford & J. E. Birren (Eds.), *Behavior, aging and the nervous system* (pp. 114–149). Springfield, IL: Thomas.
- Buchanan, L., McEwen, S., Westbury, C., & Libben, G. (2003). Semantics and semantic errors: Implicit access to semantic information from words and nonwords in deep dyslexia. *Brain and Language, 84*, 65–83.
- Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review, 100*, 432–459.
- 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. (1994). Generalized slowing in Brinley plots. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 49*, 65–71.
- Cerella, J., Poon, L. W., & Williams, D. M. (1980). Age and the complexity hypothesis. In L. W. Poon (Ed.), *Aging in the 1980s: Psychological issues* (pp. 332–340). Washington, DC: American Psychological Association.
- Coyne, A. C. (1981). Age difference and practice in forward visual masking. *Journal of Gerontology, 36*, 730–732.
- Fisher, D. L., & Glaser, R. A. (1996). Molar and latent models of cognitive slowing: Implications for aging, dementia, depression, development, and intelligence. *Psychonomic Bulletin and Review, 3*, 458–480.
- Fisk, A. D., & Fisher, D. L. (1994). Brinley plots and theories of aging: The explicit, muddled, and implicit debates. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 49*, 81–89.
- Fisk, J. E., & Warr, P. (1996). Age and working memory: The role of perceptual speed, the central executive, and the phonological loop. *Psychology and Aging, 11*, 316–323.
- 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.
- Fozard, J. L. (1990). Vision and hearing in aging. In J. E. Birren, & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 150–170). San Diego, CA: Academic Press.
- Hale, S., & Myerson, J. (1995). Fifty years older, fifty percent slower? Meta-analytic regression models and semantic context effects. *Aging and Cognition, 2*, 132–145.
- 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.
- Kucera, H., & Francis, W. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Laver, G. D., & Burke, D. M. (1993). Why do semantic priming effects increase in old age? A meta-analysis. *Psychology and Aging, 8*, 34–43.
- 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: Wiley.
- 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.
- McKoon, G., & Ratcliff, R. (1992). Spreading activation versus compound cue accounts of priming: Mediated priming revisited. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1155–1172.
- McNamara, T. P. (1992). Priming and constraints it places on theories of memory and retrieval. *Psychological Review, 99*, 650–662.
- McNamara, T. P. (1994). Theories of priming: II. Types of primes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 507–520.
- Merriam-Webster. (1990). *Merriam-Webster's ninth new collegiate dictionary* (9th ed.). Springfield, MA: Author.
- Moreno, M. A., Buchanan, L., & Van Orden, G. C. (2002). Variability in aphasic patients' response times. *Brain and Cognition, 48*, 469–474.
- Myerson, J., Adams, D. R., Hale, S., & Jenkins, L. (2003). Analysis of group differences in processing speed: Brinley plots, Q-Q plots, and other conspiracies. *Psychonomic Bulletin and Review, 10*, 224–237.
- 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., Chen, J., & Lawrence, B. (1997). General lexical slowing and the semantic priming effect: The roles of age and ability. *Acta Psychologica, 96*, 83–101.
- Myerson, J., Wagstaff, D., & Hale, S. (1994). Brinley plots, explained variance, and the analysis of age differences in response latencies. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 49*, 72–80.
- Owsley, C., Sekuler, R., & Siemsen, D. (1983). Contrast sensitivity through adulthood. *Vision Research, 23*, 689–699.
- Perfect, T. J. (1994). What can Brinley plots tell us about cognitive aging? *Journals of Gerontology, Series B: Psychological Sciences and Social 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. (1981). A theory of order relations in perceptual matching. *Psychological Review, 88*, 552–572.
- Ratcliff, R. (1985). Theoretical interpretations of speed and accuracy of positive and negative responses. *Psychological Review, 92*, 212–225.
- Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling the accumulation of partial information. *Psychological Review, 95*, 238–255.
- Ratcliff, R. (2002). A diffusion model account of reaction time and accuracy in a two choice brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin and Review, 9*, 278–291.
- Ratcliff, R., Gomez, P., & McKoon, G. (2004). Diffusion model account of lexical decision. *Psychological Review, 111*, 159–182.
- Ratcliff, R., & McKoon, G. (1988). A retrieval theory of priming in memory. *Psychological Review, 95*, 385–408.
- Ratcliff, R., & McKoon, G. (1994). Retrieving information from memory: Spreading activation theories versus compound cue theories. *Psychological Review, 101*, 177–184.
- Ratcliff, R., Perea, M., Coleangelo, A., & Buchanan, L. (in press). A diffusion model account of normal and impaired readers. *Brain & Cognition*.
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9*, 347–356.
- Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 127–140.
- Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review, 111*, 333–367.
- Ratcliff, R., Spieler, D., & McKoon, G. (2000). Explicitly modeling the effects of aging on response time. *Psychonomic Bulletin and Review, 7*, 1–25.
- Ratcliff, R., Spieler, D., & McKoon, G. (in press). Analysis of group differences in processing speed: Where are the models of processing? *Psychonomic Bulletin and Review*.
- Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging on recognition memory. *Journal of Memory and Language, 50*, 408–424.
- Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging, 16*, 323–341.
- Ratcliff, R., Thapar, A., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging, 18*, 523–535.
- Ratcliff, R., & Tuerlinckx, F. (2002). Estimation of the parameters of the

- diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin and Review*, 9, 438–481.
- Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, 106, 261–300.
- Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multi-alternative decision field theory: A dynamic artificial neural network model of decision-making. *Psychological Review*, 108, 370–392.
- Salthouse, T. A. (1985). *A theory of cognitive aging*. Amsterdam: North-Holland.
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103, 403–428.
- Salthouse, T. A., Kausler, D. H., & Saults, J. S. (1988). Utilization of path-analytic procedures to investigate the role of processing resources in cognitive aging. *Psychology and Aging*, 3, 158–166.
- Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review*, 102, 567–591.
- Spear, P. D. (1993). Neural bases of visual deficits during aging. *Vision Research*, 33, 2589–2609.
- Swenson, R. G. (1972). The elusive tradeoff: Speed versus accuracy in visual discrimination tasks. *Perception and Psychophysics*, 12, 16–32.
- Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging*, 18, 415–429.
- Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (2004). *A quantitative account of strategic effects in lexical decision*. Manuscript in preparation.
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third Edition*. San Antonio, TX: Psychological Corporation.

Received June 3, 2003

Revision received November 25, 2003

Accepted December 8, 2003 ■



AMERICAN PSYCHOLOGICAL ASSOCIATION SUBSCRIPTION CLAIMS INFORMATION

Today's Date: _____

We provide this form to assist members, institutions, and nonmember individuals with any subscription problems. With the appropriate information we can begin a resolution. If you use the services of an agent, please do **NOT** duplicate claims through them and directly to us. **PLEASE PRINT CLEARLY AND IN INK IF POSSIBLE.**

PRINT FULL NAME OR KEY NAME OF INSTITUTION _____		MEMBER OR CUSTOMER NUMBER (MAY BE FOUND ON ANY PAST ISSUE LABEL) _____
ADDRESS _____		DATE YOUR ORDER WAS MAILED (OR PHONED) _____
CITY _____ STATE/COUNTRY _____ ZIP _____		<input type="checkbox"/> PREPAID <input type="checkbox"/> CHECK <input type="checkbox"/> CHARGE CHECK/CARD CLEARED DATE: _____
YOUR NAME AND PHONE NUMBER _____		(If possible, send a copy, front and back, of your cancelled check to help us in our research of your claim.) ISSUES: <input type="checkbox"/> MISSING <input type="checkbox"/> DAMAGED
TITLE _____	VOLUME OR YEAR _____	NUMBER OR MONTH _____
_____	_____	_____
_____	_____	_____

Thank you. Once a claim is received and resolved, delivery of replacement issues routinely takes 4–6 weeks.

(TO BE FILLED OUT BY APA STAFF)

DATE RECEIVED: _____	DATE OF ACTION: _____
ACTION TAKEN: _____	INV. NO. & DATE: _____
STAFF NAME: _____	LABEL NO. & DATE: _____

Send this form to APA Subscription Claims, 750 First Street, NE, Washington, DC 20002-4242

PLEASE DO NOT REMOVE. A PHOTOCOPY MAY BE USED.