Psychology and Aging
2007, Vol. 22, No. 1, 56–66

0882-7974/07/$12.00 DOI: 10.1037/0882-7974.22.1.56

Application of the Diffusion Model to Two-Choice Tasks for Adults 75–90 Years Old

Roger Ratcliff
Ohio State University

Anjali Thapar
Bryn Mawr College

Gail McKoon
Ohio State University

The effects of aging on simple 2-choice decision making was investigated with the diffusion model (R. Ratcliff, 1978). Data for 75- to 90-year-olds were collected and compared with previous data from 60- to 75-year-olds and college students for 5 tasks: a signal detection–like task, letter and brightness discrimination with masking, recognition memory, and lexical decision. The model fit the data well and therefore allows components of processing to be examined as a function of age. Compared with decision-making processes in college students, decision criteria and nondecision components of processing increased with participants' age. However, the quality of the evidence on which decisions were based decreased with age only for letter and brightness discrimination.

Keywords: aging, reaction time, cognitive ability, diffusion model

A central finding about aging is that as people age, their response times (RTs) increase. Also, there is sometimes a decrease in accuracy. Recently, Ratcliff, Thapar, and McKoon (2001, 2003, 2004); Ratcliff, Thapar, Gomez, and McKoon (2004); and Thapar, Ratcliff, and McKoon (2003) (henceforth, RTM) examined the effects of aging in several two-choice tasks: two signal detection–like tasks, a brightness discrimination task with masked stimuli, a recognition memory task, a lexical decision task, and a letter discrimination task with masked stimuli. We separated out aging effects on component processes of the tasks by applying the diffusion model (Ratcliff, 1978, 1981, 1985, 1988, 2002; Ratcliff & Rouder, 1998, 2000; Ratcliff & Smith, 2004; Ratcliff, Van Zandt, & McKoon, 1999). We found that 60- to 75-year-old subjects adopted more conservative decision criteria than did college-age subjects and also had slower nondecision components of processing (encoding, response execution, memory access, lexical access). However, the quality of the stimulus evidence driving the decision process was not significantly worse for the older subjects than for the young subjects except in masked letter discrimination. The finding of lower quality stimulus information for masked letter discrimination but not for masked brightness discrimination is consistent with the psychophysical finding that deficits occur with age for high but not for low spatial frequency stimuli (Spear, 1993).

Roger Ratcliff and Gail McKoon, Department of Psychology, Ohio State University; Anjali Thapar, Department of Neural and Behavioral Sciences and Department of Psychology, Bryn Mawr College.

Preparation of this article was supported by National Institute of 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, Ohio State University, 1835 Neil Avenue, Columbus, OH 43210. E-mail: ratcliff.22@osu.edu

In the studies described in this article, application of the diffusion model was extended to 75- to 90-year-olds. The questions addressed were whether the diffusion model could fit the data from the 75- to 90-year-olds as well as it had fit the data from the younger subjects and whether components of processing differ between 75- to 90-year-olds and 60- to 75-year-olds. Earlier research has found substantial loss of cognitive abilities for 80- to 100-year-olds (Baltes, 1998; Baltes & Smith, 2003; Singer, Lindenberger, & Baltes, 2003; Singer, Verhaeghen, Ghisletta, Lindenberger, & Baltes, 2003). In our studies, the 75- to 90-year-old subjects were active and well functioning, as evidenced by their willingness and ability to take part in our experiments, and they were matched to college students in terms of IQ and education.

Method

In Experiments 1–6, we used the same test lists, procedures, and instructions as in the experiments by RTM, and full descriptions can be found in those articles. All of the subjects in the experiments reported here (different subjects in each experiment) and in RTM's experiments met the following inclusion criteria: 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, history of head injury with loss of consciousness, or current psychiatric disorder. All subjects completed either the Picture Completion and Vocabulary subtests or the Digit Symbol and Information subtests of the Wechsler Adult Intelligence Scale—3rd Edition (WAIS–III; Wechsler, 1997). All of these measures showed that our 75- to 90-year-old subjects matched the other two groups (characteristics are shown in Table 1). For the Experiments 1–6, there were 24, 35, 33, 35, 33, and 34 subjects, respectively. The same two sites used for earlier experiments (Bryn Mawr College and Northwestern University) provided subjects, and the same experimenters using the same apparatus tested subjects.

Table 1
Subject Characteristics

Measure	Signal detection		Discrimination: Letter		Discrimination: Brightness		Recognition memory		Lexical decision: Pseudowords		Lexical decision: Random letters	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Mean age	79.2	3.14	80.1	4.05	79.8	3.54	79.2	3.46	81.2	4.66	80.5	3.54
Years education	16.3	3.23	15.9	3.27	16.2	2.55	16.1	2.22	14.7	3.04	16.0	2.65
MMSE	28.8	1.13	28.3	1.31	28.8	1.27	28.1	1.23	28.1	1.64	28.4	1.50
WAIS–III Vocab/Info	14.3	2.39	14.2	2.54	14.7	1.90	14.9	2.11	15.2	2.44	14.7	2.04
WAIS–III PC/Dig-Sym	12.5	2.69	12.4	3.09	13.1	2.73	12.1	2.38	13.2	2.68	12.9	2.46
CES–D Total	8.4	5.12	9.5	7.09	7.4	4.99	9.4	5.68	9.7	5.14	7.9	6.51

Note. For Experiments 1–6, there were 24, 35, 33, 35, 33, and 34 subjects, respectively. MMSE = Mini-Mental State Examination; WAIS–III = Wechsler Adult Intelligence Scale—3rd edition (subjects were given either Vocabulary [Vocab] and Picture Completion [PC] subtests or Information [Info] and Digit-Symbol [Dig-Sym] subtests); CES–D = Center for Epidemiological Studies—Depression Scale.

The experimental tasks were chosen to span a range of possible limitations on cognitive performance. Experiment 1 used a signal detection task: Subjects were asked to judge whether the distance between two dots was large or small. The stimuli were clearly visible on a computer monitor and were displayed until the subject responded, so there were no perceptual or memory limits on performance. In Experiments 2 and 3, in which, respectively, letter discrimination and brightness discrimination were tested, perceptual information about the stimuli was limited through masking. Experiment 4 was a standard recognition memory task, and Experiments 5 and 6 were standard lexical decision tasks.

For Experiment 1, we manipulated stimulus difficulty by varying the distances between the two dots and, in Experiments 2 and 3, by varying the time between presentation of the stimulus and presentation of the mask. In Experiment 3, the stimuli were patches of pixels, and we manipulated difficulty using both stimulus duration and brightness, with brightness ranging from mostly white pixels to mostly dark pixels. For recognition memory, words on the study lists were high, low, or very low in frequency and were presented either once or three times. Test words that had not appeared in the study list were also high, low, or very low in frequency. The same high-, low-, and very-low-frequency words were used in the lexical decision experiments. In Experiment 5, the nonwords were pseudowords, and in Experiment 6, they were random letter strings.

For Experiments 1–4, on alternating blocks of trials, instructions stressed that responses be either as accurate as possible or as fast as possible. Subjects were given feedback appropriate to the instructions: In accuracy blocks, accuracy feedback was given on each trial, and in speed blocks, a "too slow" message was given after responses that exceeded 700 ms in the signal detection and brightness discrimination experiments, 650 ms in the letter discrimination experiments, and 800 ms in the recognition memory experiment. For Experiments 5 and 6, subjects were instructed to respond quickly and accurately.

Results

Response times shorter than 300 ms and longer than 4,000 ms were eliminated from the data (<1.8% of the data for each experiment). Figure 1 shows the RTs and accuracy values averaged across all the independent variables in each experiment for the 75- to 90-year-olds reported here and the college students and 60- to 75-year-olds reported in the RTM articles.

For the experiments that included the speed or accuracy instruction manipulation, college students were more willing to trade accuracy for speed than were older subjects, as shown by the larger differences for college students who received speed instructions

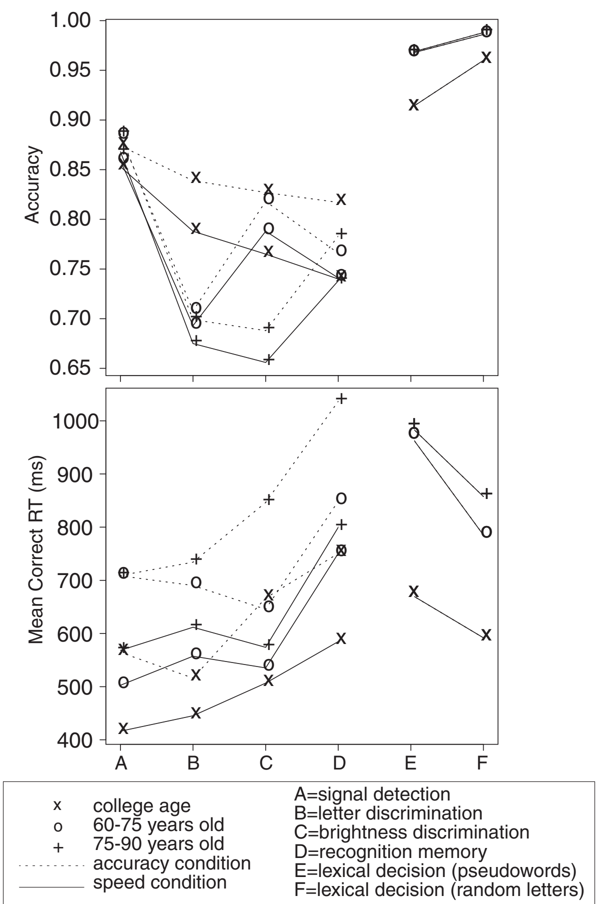

Figure 1. Plots of accuracy and mean response time (RT) averaged over all subjects and all conditions in the experiments. In each plot, there are two lines for each group of subjects for the first four experiments. In the accuracy plot, the top line corresponds to the accuracy condition, and the bottom line corresponds to the speed condition. For the RT plot, the top line is for the accuracy condition, and the bottom line for the speed condition.

compared with those who received accuracy instructions. With speed instructions, college students were always faster than 60- to 75-year-olds, who were always faster than 75- to 90-year-olds. With accuracy instructions, 75- to 90-year-olds had RTs that were longer than those of 60- to 75-year-olds except in the signal detection experiment; however, some of the differences were small and not reliable. For 75- to 90-year-olds, the effect on RTs of speed instructions versus accuracy instructions varied across experiments from as little as 100 ms to as much as 300 ms. The RT effects were smaller for college students and 60- to 75-year-olds.

The effects of age on accuracy varied across tasks. Accuracy was roughly equivalent for the three subject groups for the signal detection task; there were relatively small decrements as a function of age in recognition memory; and accuracy was higher for both groups of older subjects than for college subjects in lexical decision. In the letter discrimination task, in which the stimuli had high spatial frequency, college students were over 10% more accurate than either group of older subjects. In the brightness discrimination task, in which the stimuli had lower spatial frequency, college students and 60- to 75-year-olds showed equivalent levels of accuracy, whereas the 75- to 90-year-olds showed more than a 10% drop.

Quantile Probability Functions

Quantile probability functions provide a summary picture of the shapes of RT distributions, how they vary across conditions (i.e., levels of accuracy), and how correct RTs compare with error RTs. The functions for the 75- to 90-year-olds were constructed in the same way as they were for the younger subjects described in the RTM articles. The probability of a response determines position on the *x*-axis, and quantile RTs are plotted vertically on the *y*-axis. In Figures 2 and 3, the .1, .3, .5 (median), .7, and .9 quantiles are plotted. Correct responses fall on the right-hand sides of the functions, and error responses on the left (the probabilities of correct responses are usually greater than .5, and the probabilities of error responses are usually lower than .5). The quantiles for correct responses for the easiest (most accurate) stimulus conditions fall on the far right of the plots, and their error quantiles on the far left. For more difficult conditions, the quantiles are nearer the center. The quantiles are ordered so that for each vertical line (i.e., each stimulus difficulty condition), the lowest is the .1 quantile, the next lowest is the .3 quantile, and so on.

The quantile probability functions in the figures represent averages across subjects for each experimental condition. For Experiment 1, the signal detection experiment, there were 32 different

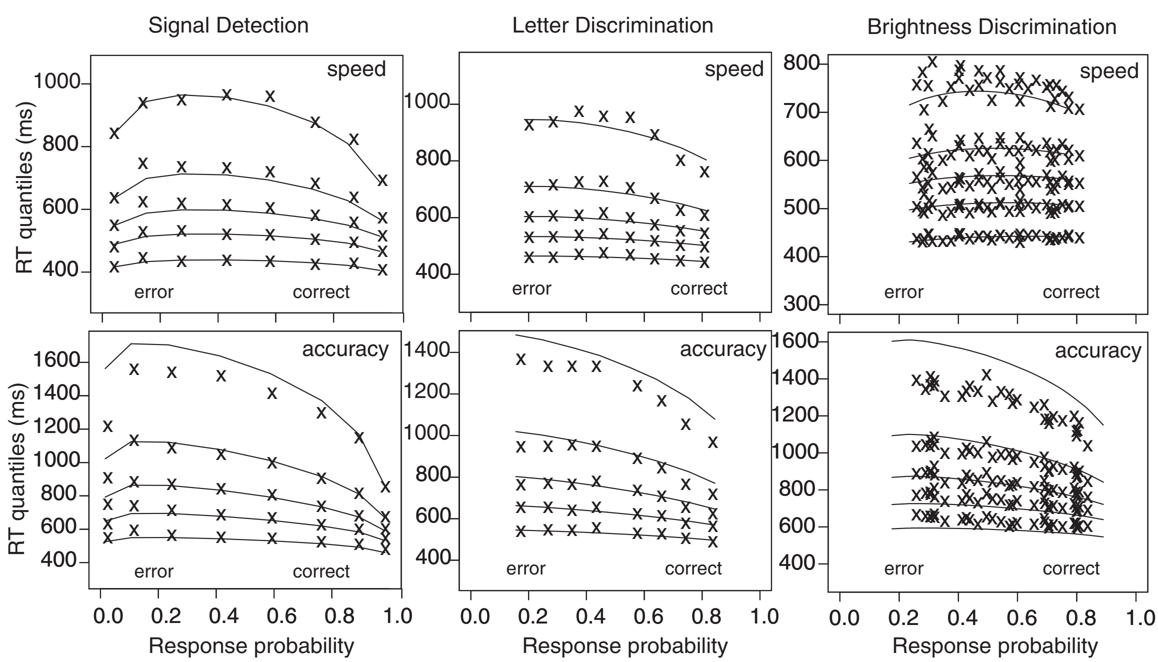

Figure 2. Quantile–probability plots for the signal detection, letter discrimination, and brightness discrimination experiments. The Xs represent the data averaged across subjects, and the lines represent the theoretical fits of the diffusion model. The quantile response times (RTs) in order from the bottom to the top are the .1, .3, .5, .7, and .9 quantiles, and in each vertical line, the quantiles have to have this order. For the brightness discrimination experiment in several conditions, a moderate proportion of subjects did not have enough responses to allow computation of quantiles, so no quantiles are presented for these extreme error conditions. For signal detection, the order of the conditions represents grouping presented in Ratcliff et al. (2006). For letter discrimination, the extreme points represent stimulus durations of 40, 30, 20, and 10 ms, respectively. For brightness discrimination, the extreme points represent more extreme stimuli with .65 and .35 proportion of white pixels and longer stimulus durations (150 ms), whereas the conditions in the center (accuracy near .5) represent more difficult conditions with .525 and .475 proportion of white pixels and shorter stimulus durations (50 ms).

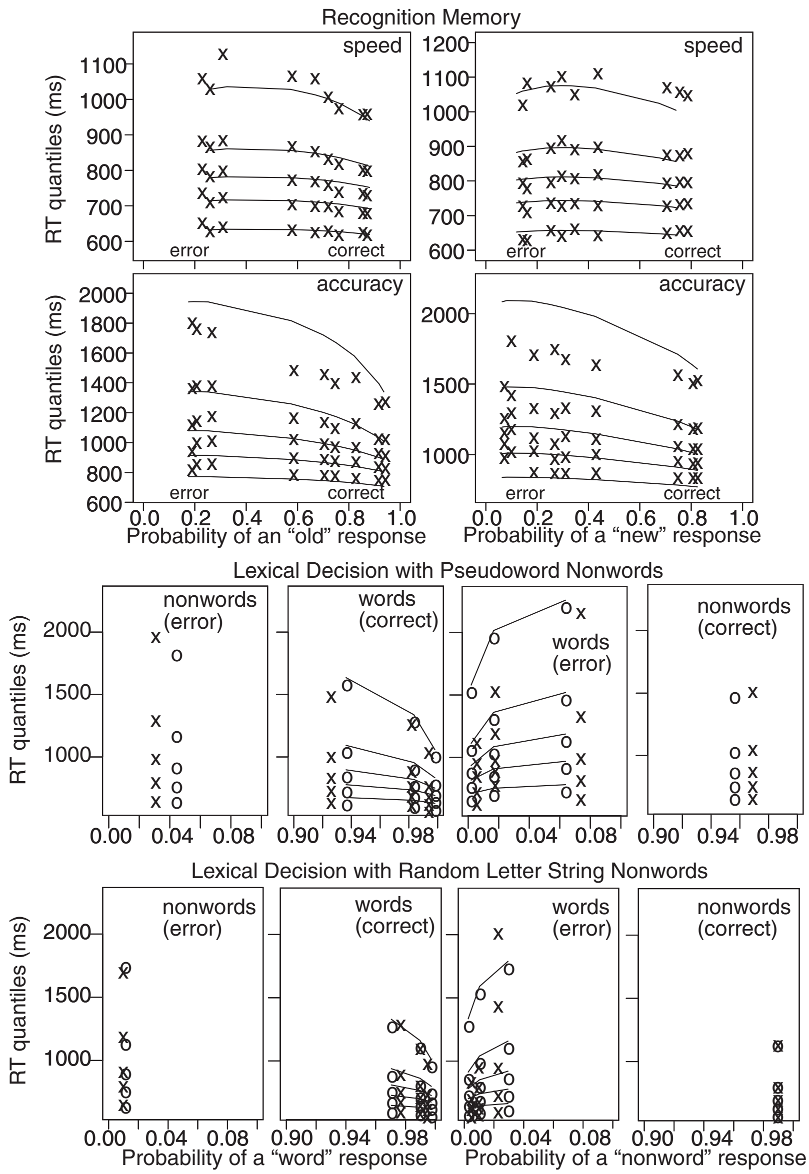

Figure 3. Quantile–probability plots for the recognition memory and lexical decision experiments. The Xs represent the data averaged across subjects, the Os represent theoretical predictions, and the lines represent the best fits of the model to the data. For recognition memory for the left-hand plots, the conditions from right to left represent 3V, 3L, 3H, 1V, 1L, 1H, NH, NL, and NV, where 3 = three presentations, 1 = one presentation, N = new words, V = very-low-frequency words, L = low-frequency words, and H = high-frequency words. For the right-hand panel, the conditions are in the reverse order. For the lexical decision experiments, word responses for the high-frequency words were more accurate than those for the low-frequency words, which were more accurate than those for the very-low-frequency words. RT = response time.

distances between the dots, grouped by similar RTs and accuracy (Ratcliff et al., 2001) into four conditions in the figures. For Experiment 2, there were four possible stimulus durations between presentation of a letter and presentation of the mask. For Experiment 3, there were 18 conditions, defined by crossing three stimulus durations with six different proportions of white versus black pixels. For these three experiments, responses for the two choices were approximately symmetric for both accuracy and RTs. In other words, in the signal detection task, for example, the same accuracy and correct and error RTs were found with responses of "large" to a large stimulus as with responses of "small" to the corresponding small stimulus. Therefore, the correct responses for the two choices were grouped together as were the error responses, giving a single quantile probability function for conditions with speed instructions and a single function for conditions with accuracy instructions. (For Experiment 3, there were too few observations to plot quantiles for errors in the most accurate conditions; therefore, for these conditions, no quantiles are displayed.)

For Experiments 4, 5, and 6 (recognition memory and lexical decision tasks), accuracy and RTs were not symmetric for the two choices, so there were two quantile probability functions for each experiment: for recognition memory, one for "old" responses and one for "new" responses, and for lexical decision, one for "word" and one for "nonword" responses. For Experiment 4, there were six conditions for studied words (presented once or three times crossed with three levels of frequency) and three conditions for words that had not been studied (three levels of frequency). For Experiments 5 (pseudowords as nonwords) and 6 (random letter strings as nonwords), there were three conditions for words, defined by three levels of word frequency.

Overall, with both speed and accuracy instructions, the .1 quantile RTs for correct responses changed by less than 100 ms across levels of accuracy, whereas the .9 quantile RTs changed by as much as several hundred ms. In general, error responses were slower than correct responses, and, as with correct responses, changes in median error RTs across conditions were mainly reflected in the RT distributions spreading rather than shifting (changing by as much as 1,000 ms in the .9 quantiles in lexical decision). The patterns for the 75- to 90-year-olds qualitatively match those for the two younger groups described in the RTM articles. Comparing the 75- to 90-year-olds reported here with the 60- to 75-year-olds in the earlier experiments, we found that accuracy was lower for 75- to 90-year-olds only in brightness discrimination, whereas RTs were longer in all conditions except with accuracy instructions in Experiment 1. The question for the diffusion model is what components of processing are responsible for these effects.

Interpreting the Data Through the Diffusion Model

Our goal with the diffusion model is to explain the cognitive processes involved in making simple two-choice decisions. The model separates the quality of evidence entering a decision from the decision criteria and from other, nondecision processes such as stimulus encoding and response execution. Decisions are made by a process in which information accumulates over time from a starting point z toward one of two response criteria, or boundaries, a and 0. When a boundary is reached, a response is initiated. The rate of accumulation of information is called the drift rate (v), and it is determined by the quality of the information extracted from the stimulus in perceptual tasks and by the quality of the match between the test item and memory in memory and lexical decision tasks. The nondecision components of processing such as encoding and response execution are combined into one component with mean T_{er}. Within-trial variability (noise) in the accumulation of information from the starting point toward the boundaries results in processes with the same mean drift rate terminating at different times (producing RT distributions) and sometimes at the wrong boundary (producing errors). It is assumed that components of processing vary from trial to trial. Across-trial variability in drift rate (normally distributed with SD η) and starting point (uniformly distributed with range s_z), in conjunction with boundary positions and drift rates, determines the relative speed of correct responses versus error responses. It is also assumed that the nondecision component varies across trials, uniformly distributed with range s_t. For further details of the model, see the RTM articles and Ratcliff and Tuerlinckx (2002).

The main manipulations in the experiments were stimulus difficulty and, in Experiments 1–4, speed instructions versus accuracy instructions. We manipulated difficulty with distance between the dots in the signal detection task, stimulus duration in the letter discrimination task, brightness and stimulus duration in the brightness discrimination task, number of repetitions and word frequency in recognition memory, and word frequency in lexical decision. Differences in difficulty are modeled by differences in drift rate. Speed–accuracy tradeoffs are modeled by changes in the distance between the boundaries in the decision process—wider boundaries require more information before a decision can be made, and this leads to more accurate and slower responses. The assumptions that only drift rate can change with difficulty and that only boundary separation can change between speed and accuracy instructions produce a highly constrained model, one that would be falsified by many possible deviations of the data from predicted values (Ratcliff, 2002).

Because of the results of the RTM studies, we expected response boundaries and T_{er} to increase with age in all six experiments. For the signal detection task, there were no perceptual or memory limits on the information available to the subjects; therefore, according to Ratcliff et al. (2001), drift rates should not vary with age. For the other tasks, the RTM studies found that drift rates were lower for 60- to 75-year-olds than for college students only in masked letter discrimination, but for 75- to 90-year-olds, lower drift rates might also be expected in masked brightness discrimination and recognition memory.

We fit the diffusion model to the data using a standard minimization routine (Ratcliff & Tuerlinckx, 2002). Each subject's data were fit individually, and the resulting parameter values were averaged across subjects (Tables 2 and 3). Standard errors in the parameter values can be found by dividing the standard deviations by the square root of the number of subjects for each experiment. To perform significance tests for differences in parameter values between the 75- to 90-year-old subjects tested here and the 60- to 75-year-old subjects from the RTM articles, we combined the standard errors from the two groups to produce a pooled standard error; this pooled standard error doubled was used as the critical value.

The model was also fitted independently to the data averaged across subjects: Accuracy values and each quantile RTs were averaged across subjects for each condition. These fits provided the lines in Figures 2 and 3. Group data have often been used in fitting models, and the assumption (usually implicit) is that the parameter values for fits to the group data will be the same as averages from fits for the individual subjects. This was true for our experiments. Parameter values obtained from the fits to the group data and average parameter values across individuals were within 2 standard errors of each other for all parameters with only two exceptions: For brightness discrimination, the differences were the result of a relatively low number of errors in high accuracy conditions, and for the lexical decision experiment with random letter strings, the differences occurred because some conditions had very high accuracy and consequently very low numbers of errors.

Goodness of fit. Figures 2 and 3 show that the model does a good job of capturing changes in correct and error RT distributions and accuracy values, with only drift rate changing across conditions of stimulus difficulty. For the experiments in which two

Table 2
Means and Standard Deviations in Parameter Values Across Subjects for Fits of the Diffusion Model to the Experiments

Experiment	a_s	a_a	T_{er}	η	s_z	s_t	p_o	z_s	z_a	χ^2	χ^2 60–75 yr	χ^2 college	*df*
							Mean						
Signal detection	.118	.183	.395	.170	.056	.183	.000			141	135	115	77
Letter discrimination	.103	.156	.408	.216	.011	.151	.004			88	119	84	77
Brightness discrimination	.089	.193	.432	.186	.063	.220	.036			1175	951	680	377
Recognition memory	.100	.184	.616	.184	.028	.257	.001	.045	.081	427	431	368	180
Lexical decision (psuedo)		.199	.543	.100	.041	.129	.047		.091	91	203[a]	304[a]	77
Lexical decision (random)		.204	.533	.150	.071	.099	.050		.108	75	71[a]	172[a]	77
Signal detection (60–75 yr)	.102	.190	.368	.192	.050	.122	.001						
Signal detection (college)	.084	.155	.312	.161	.029	.169	.004						
							Standard deviation						
Signal detection	.026	.044	.051	.052	.040	.058	.001			44			
Letter discrimination	.035	.052	.051	.052	.040	.049	.011			34			
Brightness discrimination	.020	.074	.044	.056	.022	.039	.002			643			
Recognition memory	.032	.067	.050	.099	.021	.064	.001	.018	.033	376			
Lexical decision (pseudo)		.053	.076	.054	.041	.083	.040		.021	59			
Lexical decision (random)		.046	.056	.069	.046	.064	.041		.028	117			

Note. The last two columns show the average chi-square values from the college student and 60- to 75-year-old groups from the Ratcliff, Thapar, and McKoon (2001, 2003, 2004); Ratcliff, Thapar, Gomez, and McKoon (2004); and Thapar, Ratcliff and McKoon (2003) studies. The last two columns of means (for the signal detection experiment with college and 60- to 75-year-old subjects) are fits to the data sets from Ratcliff, Thapar, and McKoon (2001) in which more up-to-date programs were used in fitting all the other experiments and in which a parameter representing variability in T_{er} (s_t) and a parameter representing the proportion of contaminant RTs (p_o) were used. a_s = boundary separation for speed condition; a_a = boundary separation for accuracy condition; T_{er} = nondecision component of response time; η = standard deviation in drift across trials; s_z = range of the distribution of starting point (z); s_t = range of the distribution of nondecision times; p_o = proportion of contaminants; z_s = starting point for speed condition; z_a = starting point for accuracy condition.
[a]Values are from data collapsed into supersubjects so that the chi-square values are inflated relative to the values for the 75- to 90-year-old subjects.

variables were manipulated (presentation duration and brightness in brightness discrimination; word frequency and repetitions in recognition memory), the RT quantiles lie on the same functions. This is consistent with the assumption that both variables affect a single common component of processing, drift rate.

The model also captures the effects of speed and accuracy instructions, with only boundary separation changing. The only systematic misses are in the .9 quantile RTs with accuracy instructions in brightness discrimination and recognition memory (with smaller misses in signal detection and letter discrimination) and in the .1 quantile RTs for brightness discrimination and recognition memory. First, the .9 quantile RTs may miss because subjects did not allow processes to run to completion (RTs are long, e.g., 1.5 s), thereby reducing their .9 quantile RTs relative to predictions. Second, it may be that accuracy instructions slow processing for components other than the decision process, leading to a larger value of T_{er}. A modest increase in T_{er} (e.g., 20–30 ms, see Rinkenauer, Osman, Ulrich, Müller-Gethmann, & Mattes, 2004) would produce a slight increase in the predicted .1 quantile RTs (e.g., 20–30 ms), which would allow the model to better match the data in the brightness discrimination and recognition memory experiments. This would require a smaller value of boundary separation (a) to produce the best fits, which would reduce the predicted .9 quantile RTs to better match the data. The change in T_{er} and a would not significantly affect the values of the other parameters of the model and therefore would not change any conclusions.

The chi-square values averaged over individual subjects are shown in Table 2. Across all the experiments, the chi-square values for the 75- to 90-year-olds are similar to those for the college students and 60- to 75-year-olds in RTM's experiments. The similarity of the chi-square values shows that the model fits the data well across all three age groups (see Ratcliff, Thapar, Gomez, & McKoon, 2004, p. 285, for discussion of the power of the chi-square test).

Differences in parameter values with age. Table 4 summarizes z tests for parameter values for the 75- to 90-year-olds compared with those of RTM's 60- to 75-year-olds and RTM's 60- to 75-year-olds compared with college students with pooled standard deviations as noted above (significance level = .05).

The three parameters identified in the RTM articles as most likely to be involved in slowing of older relative to younger adults are boundary separation, the nondecision component of processing, and drift rate. As Figure 4 and Table 4 show, boundary

Table 3
Means and Standard Deviations in Drift Rates and Drift Criteria for Fits of the Diffusion Model to the Experiments

Experiment	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8	v_9	v_{cr13}	v_{cr46}	v_{cr79}
						Mean						
Signal detection	.449	.250	.140	.041								
Letter discrimination	.310	.219	.116	.043								
Brightness discrimination	.181	.101	.036	.226	.130	.048	.245	.158	.053	−.036	.011	.056
Recognition memory	.334	.285	.185	.113	.082	.027	−.243	−.229	−.168			
Lexical decision (psuedo)	.435	.303	.207	−.276								
Lexical decision (random)	.565	.454	.363	−.425								
Signal detection (60–75)	.470	.256	.141	.074						.036		
Signal detection (college)	.453	.251	.133	.054						.077		
						Standard deviation						
Signal detection	.111	.099	.079	.073								
Letter discrimination	.199	.162	.105	.069								
Brightness discrimination	.152	.086	.039	.169	.114	.041	.182	.114	.046	.061	.046	.055
Recognition memory	.177	.165	.115	.078	.070	.063	.155	.146	.127			
Lexical decision (psuedo)	.124	.099	.074	.096								
Lexical decision (random)	.164	.142	.117	.130								

Note. For signal detection, drift rates (v) represent grouping presented in the text. For letter discrimination, v_1–v_4 represent stimulus durations of 40, 30, 20, and 10 ms respectively. For brightness discrimination, the first three, second three, and third three drift rates are for 50-, 100-, and 150-ms stimulus durations, respectively. Within each group of three drift rates, the first has .35 and .65 pixel conditions combined, the second, .425 and .575 combined, and the third, .475 and .525 combined. v_{crij} represents the drift criterion for conditions i and j. For recognition memory, the first three drift rates are for items presented three times, the next three for items presented once, and the last three for new items. Within each group of three, the first drift rate is for very-low-frequency words, the second for low-frequency words, and the third for high-frequency words. For lexical decision, the first drift rate is for high-frequency words, the second for low-frequency words, the third for very-low-frequency words, and the fourth for nonwords. The last two columns of means (for the signal detection experiment with college and 60- to 75-year-old subjects) are fits to the data sets from Ratcliff, Thapar, and McKoon (2001) in which more up-to-date programs were used in fitting all the other experiments and in which a parameter representing variability in the nondecision component of response time (range of the distribution of nondecision times) and a parameter representing the proportion of contaminant response times were used.

separations were larger for 75- to 90-year-olds than for 60- to 75-year-olds with speed instructions in the signal detection[1] and brightness discrimination experiments and with accuracy instructions in the brightness discrimination and recognition memory experiments. In RTM's experiments, 60- to 75-year-olds had larger boundary separations than did college students for all the tasks except brightness discrimination. Boundary separation is assumed to be under subjects' control, so it might be expected that differences with age would not be consistent across experiments; instead, boundary settings would depend on how subjects differentially interpreted instructions. However, this was not the finding: The separations were consistently smaller for college students relative to 75- to 90-year-olds in all the experiments, indicating a general trend of increasing boundary separation with age.

The nondecision component of processing was slower for 75- to 90-year-olds than for 60- to 75-year-olds only in signal detection, recognition memory, and lexical decision with random letter strings. In RTM's experiments, the nondecision component was slower for 60- to 75-year-olds than for college students in all the tasks.

The most striking result is that for three of the tasks—signal detection, recognition memory, and lexical decision—drift rates did not significantly decrease with age. This indicates that the quality of information entering the decision process did not decline with age. For letter discrimination, drift rates decreased from college students to 60- to 75-year-olds but no further for 75- and 90-year-olds. For brightness discrimination, drift rates were the same for college students and 60- and 75-year-olds and decreased from 60- and 75-year-olds to 75- and 90-year-olds. Brightness discrimination was the only task for which drift rates were lower for 75- and 90-year-olds than for 60- and 75-year-olds. These last two results were replicated in Ratcliff, Thapar, and McKoon (in press) with small groups of subjects.

The estimates of variability parameters have proportionally larger standard deviations than do the estimates of other parameters (Ratcliff & Tuerlinckx, 2002), so differences among them must be much larger to be significant. Overall, in Experiments 1–6 and in the RTM studies, there was a tendency for larger variabilities in drift rate, starting point, and the nondecision component of processing for the 75- and 90-year-olds than for the younger

[1] The original fits of the signal detection experiment (Ratcliff et al., 2001) did not use more recent fitting programs that include variability in the nondecision component of processing (s_t) and the possibility of contaminant (e.g., outlier) response times (p_o). In addition, the data presented here had slightly different biases in the zero point of drift across subjects, so a drift criterion parameter that represents this difference was added. These changes increased the number of degrees of freedom from 78 to 159 (because "large" responses to large stimuli and "small" responses to small stimuli were fit separately, but the data and fits were displayed combined as in Ratcliff et al., 2001). The only major effect on parameter values was an increase in the value of T_{er} because the inclusion of variability allowed shorter values of the quantiles to be accommodated (the new minimum was $T_{er} - s_t/2$).

Table 4
Effects of Subject Groups on Parameters of the Diffusion Model

	Parameter difference					
	60 to 75-year-old vs. college-age subjects			75 to 90-year-old vs. 60 to 75-year-old subjects		
Experiment	a_s	T_{er}	v	a_s	T_{er}	v
Signal detection	higher	longer	*ns*	higher	longer	*ns*
Letter discrimination (masked)	higher	longer	lower	*ns*	*ns*	*ns*
Brightness discrimination (masked)	*ns*	longer	*ns*	higher	*ns*	lower
Recognition memory	higher	longer	*ns*	*ns*	longer	*ns*
Lexical decision pseudowords	higher	longer	*ns*	*ns*	*ns*	*ns*
Lexical decision random letter strings	higher	longer	*ns*	*ns*	longer	*ns*

Note. For the lexical decision experiment, there is one value for *a,* and for the other four experiments, the value of *a* for the speed (s) condition is used. a = boundary separation; T_{er} = nondecision component of response time; v = drift rate.

groups (Table 2). However, a few differences were significant: first, the differences in η between college students and 60- to 75-year-olds for letter discrimination and recognition memory, and between 60- to 75-year-olds and 75- to 90-year-olds for brightness discrimination; second, the differences in s_z between college students and 60- to 75-year-olds for lexical decision with random letter strings and between 60- to 75-year-olds and 75- to 90-year-olds for brightness discrimination; and third, the differences in s_t between college students and 60- to 75-year-olds for signal detection and between 60- to 75-year-olds and 75- to 90-year-olds for brightness discrimination and recognition memory.

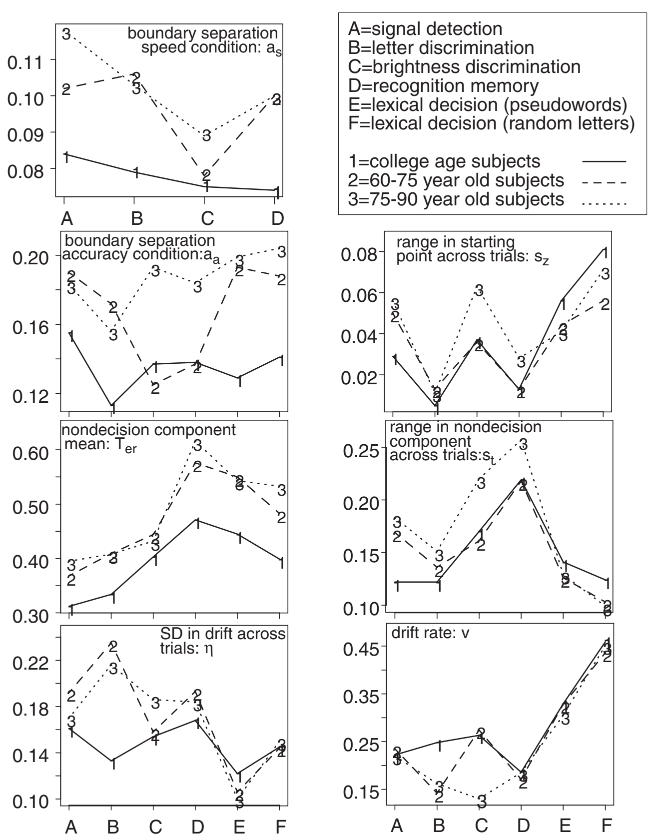

Figure 4. Plots of the parameter values from the diffusion model as a function of experiments (A–F) and of subject groups. The results from the college-age subjects and the 60- to 75-year-old subjects are from Ratcliff, Thapar, and McKoon (2001, 2003, 2004); Ratcliff, Thapar, Gomez, and McKoon (2004); and Thapar, Ratcliff and McKoon (2003). The drift rate is the average over all conditions in the experiments.

Correlations. Relationships among the dependent variables (accuracy, correct RTs, and error RTs) and the main parameters of the model (drift rate, boundary separation, and the nondecision component) averaged over all the experiments are shown in Table 5. For those experiments with speed or accuracy instructions, we calculated correlations only for conditions with speed instructions because the data were more stable and the boundary separation parameters more consistent across subjects than for conditions with accuracy instructions. For the dependent variables, means were calculated across all the conditions of an experiment.

The correlations were remarkably similar across the six experiments reported here and across RTM's experiments. For example, the correlations between correct mean RTs and *a* were .85, .89, .75, .86, .89, and .57, and the correlations between accuracy and drift rate were .67, .79, .94, .80, .57, and .35 for Experiments 1–6,

Table 5
Average Correlations Across Experiments for the Main Features of the Data and the Parameter Values

Variable	ERT	Pr	CRT	a_s	T_{er}	η
Pr	−.11					
CRT	.80[a]	−.03				
a_s	.74[a]	.03	.80[a]			
T_{er}	.28	.17	.25	−.06		
η	−.03	.21	.07	.21	.28[a]	
drift	−.30[a]	.69[a]	−.20	−.08	.21	.58[a]

Note. The critical value of the correlation coefficient for 21 correlations averaged over six experiments is .13. For the recognition memory and lexical decision experiments, all the drift rates for new items and nonwords were negative, so their absolute values were used. ERT = error RT; Pr = probability of the response (accuracy); CRT = correct RT; a_s = boundary separation for the speed condition; T_{er} = nondecision component of processing; η = SD in drift across trials. Response times are averaged over conditions.
[a] Values had the same sign for each experiment.

respectively. (The lower correlations for lexical decision with random letters as nonwords in Experiment 6 might be due to extreme values of drift rate that result in a lower range of drift rates across subjects.) Given the general similarity, the table reports the correlations averaged across experiments.

We computed a significance value, .13, from the 21 correlations in Table 5. The value was obtained under the assumption that both the data and the parameter values come from normal distributions for each experiment. The .13 value was obtained from repeated comparisons carried out in Monte Carlo simulations for each parameter and data statistic. There are seven data statistics and parameter values (see Table 5), and each was used in six comparisons. The significance value was the value of the 500th largest of 10,000 simulations (the 5% point). Differences moderately larger than .13 are likely significant ("likely" because the data and distributions of parameter values might deviate from normality). Correlations with absolute values greater than .13 and the same sign for each experiment (asterisked values in Table 5) are certainly significant.

The main results (Table 5) are that mean RTs for both correct and error responses strongly correlate with boundary separation and that accuracy strongly correlates with drift rate. The other results are as follows: The standard deviation in drift rate across trials is correlated with drift rate. Correct and error RTs are correlated with each other and are weakly negatively correlated with drift rate. Neither accuracy and mean RT nor boundary separation and drift rate are correlated. The nondecision component of processing is not strongly correlated with any of the other quantities.

The results broadly replicate those of RTM. The overall level of RT (both correct and error RTs) is determined by the boundary separation that subjects adopted. More conservative subjects responded more slowly, less conservative subjects more quickly. RT is at most weakly determined by drift rate: The quality of the information on which decisions were based was only slightly better for faster subjects than for slower subjects. Overall accuracy is mainly determined by drift rate: Subjects with higher drift rates performed more accurately than subjects with lower drift rates. Accuracy is not a function of boundary separation: More conservative subjects were not more accurate than less conservative subjects.

It is important to stress that these correlations concern individual differences in overall levels of performance averaged across all the conditions with speed instructions in each experiment. Even though overall accuracy and RT are not correlated across subjects, it is the case that within a subject, changes in drift rate, for example, have strong and reliable effects on both accuracy and RT (Figures 2 and 3).

General Discussion

For two-choice decision tasks, the diffusion model allows components of processing to be extracted from RT and accuracy data. In this article, application of the model was extended from the college age and 60- to 75-year-old subjects of earlier experiments (RTM) to 75- to 90-year-olds.

The model fit the data well, apart from some modest misses in the .1 and .9 quantiles of RT distributions in some conditions for some experiments. The model is highly constrained, especially in the behavior of RT distributions. This was proven by Ratcliff (2002). Several fake data sets that were plausible but never observed empirically were generated; for example, for one set, RT distributions had normal distributions instead of the right-skewed distributions that are observed empirically. For another set, RT distributions shifted as task difficulty increased instead of spreading. The model was fit to all the fake data sets, and in each case, the model failed to fit significantly.

The most salient result of the experiments reported here is that the quality of the information entering the decision process, drift rate, was as high for the 75- to 90-year-olds as for the 60- to 75-year-olds in five out of the six experiments and as high as for the college students in four out of the six. Drift rates differed between the 75- to 90-year-olds and the 60- to 75-year-olds only for brightness discrimination. Drift rates differed between both the 60- to 75-year old and 75- to 90-year-old groups and the college students for letter discrimination.

The signal detection task offers a useful control for the other experiments. It shows that drift rates do not decline as participants age in a task with little cognitive or memory load and with no limit on the availability of perceptual information. In masked letter discrimination, in which the stimuli have high spatial frequencies, drift rates decreased between the college students and the 60- to 75-year-olds (RTM) but not between the 60- to 75-year-olds and the 75- to 90-year-olds. For masked brightness discrimination, in which the stimuli have low spatial frequencies, drift rates did not significantly decrease between the college students and the 60- to 75-year-olds, but they did decrease between the 60- to 75-year-olds and the 75- to 90-year-olds.

Drift rates for recognition memory did not significantly decline across the three groups of subjects (RTM's studies and Experiment 4). Previously, the conclusion in the literature has been that aging has little effect on recognition memory (Balota, Dolan, & Duchek, 2000; Bowles & Poon, 1982; Craik, 1994; Craik & Jennings, 1992; Craik & McDowd, 1987; Erber, 1974; Gordon & Clark, 1974; Kausler, 1994; Neath, 1998; Naveh-Benjamin, 2000; Rabinowitz, 1984; Schonfield & Robertson, 1966). However, this conclusion has been based only on accuracy measures. In Experiment 4 and in RTM's experiments, older adults were much slower than college students. This presents a puzzle: Slowing for older adults has often been interpreted as a deficit such that, for example, cognitive operations are not fully completed in the available time or the products of earlier operations are not fully available for later operations (e.g., Salthouse, 1996). The diffusion model reconciles the RT and accuracy data: Older adults are as accurate as college students because the quality of their information from memory is as good as that of the students. Older adults are slower because they set their response boundaries more conservatively.

Drift rates for lexical decision also showed no significant differences among the three age groups, suggesting that vocabulary does not change with age. Less accurate performance by college students results from their less conservative decision criteria (Ratcliff, Thapar, Gomez, & McKoon, 2004, and Experiments 5 and 6).

In most tasks, the 60- to 75-year-olds set more conservative decision criteria than college students (RTM) and the 75- to 90-year-olds set more conservative criteria in all six experiments reported here. However, the 75- to 90-year-olds were more conservative than the 60- to 75-year-olds only for some of the experiments. Our interpretation of these results is that older subjects tend to adopt more conservative criteria, but that this is variable across individuals and their understanding of speed and accuracy instructions. Such variability is to be expected if criteria settings are under the control of subjects (as they must be because speed–accuracy instructions have large effects on RT). In the experiments reported here and in the RTM articles, with fully functional older adults matched on relevant characteristics, it is likely that the 75- to 90-year-olds and the 60- to 75-year-olds were similar enough that differences in criteria sometimes occurred and sometimes did not. The same interpretation applies to the nondecision component of processing. Whereas the 60- to 75-year-olds were almost always slower in this component than the college students, the 75- to 90-year-olds were only sometimes slower than the 60- to 75-year-olds. Again, there may be less difference between fully functioning and matched 75- to 90-year-olds and 60- to 75-year-olds than the age difference might suggest.

In all of the studies in this article and in the RTM articles, if the RT data were considered in isolation from the accuracy data, the suggestion would be that aging has a relatively large effect on cognitive processes. On the other hand, if the accuracy data were considered alone, the suggestion would be that aging has a relatively small effect. In the diffusion model framework, the RT and accuracy data are jointly interpreted: The large differences in RTs arise from differences in nondecision components of processing and criteria settings. Accuracy is similar across the age groups (in all but the letter and brightness discrimination experiments) because drift rates are similar.

Previous research with fully functioning 80- to 100-year-olds has indicated a substantial decline in cognitive abilities relative to 60- to 75-year-olds. Baltes and colleagues (Baltes, 1998; Baltes & Smith, 2003; Singer, Lindenberger, & Baltes, 2003; Singer, Verhaeghen, et al., 2003) reported significant declines in memory, language fluency, general knowledge, and especially perceptual speed. Our results are consistent with the findings on perceptual speed: 75- to 90-year-olds were always slower than 60- to 75-year–olds, who were always slower than college students (with one minor exception in one condition). However, we found high levels of accuracy for both our older groups in recognition memory, lexical decision, and signal detection. One reason performance was better for our oldest subjects than for Baltes and colleagues' subjects might be that ours were younger, with a mean age of about 80 and an upper limit of 90. Another reason might be that the components of processing in our simple two-choice tasks are relatively preserved for 75- to 90-year-olds. We believe that processes like those in our tasks are representative of the building blocks that make up higher level processes, and we hypothesize that they are the last cognitive processes to show decrements with advanced age.

As the theoretical analyses of the diffusion model are brought to various tasks, the quality of the information extracted from stimuli is decoupled from criterion effects and from nondecision components of processing. Instead of a monolithic account of processing speed in terms of only mean correct RTs, we have instead an account based on all aspects of the data. The quality of information extracted from stimuli can be separated from subject-adjustable decision criteria. The data and analyses from the studies reported here add to a growing body of support for the diffusion model in particular and quantitative modeling approaches in general.

References

Balota, D. A., Dolan, P. O., & Duchek, J. M. (2000). Memory changes in healthy older adults. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 395–409). New York: Oxford University Press.

Baltes, M. M. (1998). The psychology of the oldest-old: The fourth age. *Current Opinion in Psychiatry, 11,* 411–415.

Baltes, P. B., & Smith, J. (2003). New frontiers in the future of aging: From successful aging of the young old to the dilemmas of the fourth age. *Gerontology, 49,* 123–135.

Bowles, N. L., & Poon, L. W. (1982). An analysis of the effect of aging on recognition memory. *Journal of Gerontology, 37,* 212–219.

Craik, F. I. M. (1994). Memory changes in normal aging. *Current Directions in Psychological Science, 3,* 155–158.

Craik, F. I. M., & Jennings, J. (1992). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 51–110). Hillsdale, NJ: Erlbaum.

Craik, F. I. M., & McDowd, J. M. (1987). Recall, recognition, and aging. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 474–479.

Erber, J. T. (1974). Age differences in recognition memory. *Journal of Gerontology, 29,* 177–181.

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.

Gordon, S. K., & Clark, W. C. (1974). Adult age differences in word and nonsense syllable recognition memory and response criterion. *Journal of Gerontology, 29,* 659–665.

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

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1170–1187.

Neath, I. (1998). Developmental changes in memory. In *Human memory: An introduction to research, data, and theory* (pp. 350–375). Pacific Grove, CA: Brooks/Cole.

Rabinowitz, J. C. (1984). Aging and recognition failure. *Journal of Gerontology, 39,* 65–71.

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

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

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

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

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

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

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

Ratcliff, R., Thapar, A., & McKoon, G. (in press). Aging and individual differences in rapid two-choice decisions. *Psychonomic Bulletin and Review.*

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.

Rinkenauer, G., Osman, A., Ulrich, R., Müller-Gethmann, H., & Mattes, S. (2004). On the locus of speed-accuracy tradeoff in reaction time: Inferences from the lateralized readiness potential. *Journal of Experimental Psychology: General, 133,* 261–282.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103,* 403–428.

Schonfield, A. E. D., & Robertson, B. A. (1966). Memory storage and aging. *Canadian Journal of Psychology, 20,* 228–236.

Singer, T., Lindenberger, U., & Baltes, P. B. (2003). Plasticity of memory for new learning in very old age: A story of major loss? *Psychology and Aging, 18,* 306–317.

Singer, T., Verhaeghen, P., Ghisletta, P., Lindenberger, U., & Baltes, P. B. (2003). The fate of cognition in very old age: Six-year longitudinal findings in the Berlin Aging Study (BASE). *Psychology and Aging, 18,* 318–331.

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

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

Wechsler, D. (1997). *The Wechsler Adult Intelligence Scale* (3rd ed.). San Antonio, TX: Psychological Corporation.

Received January 13, 2006
Revision received June 8, 2006
Accepted June 27, 2006 ■