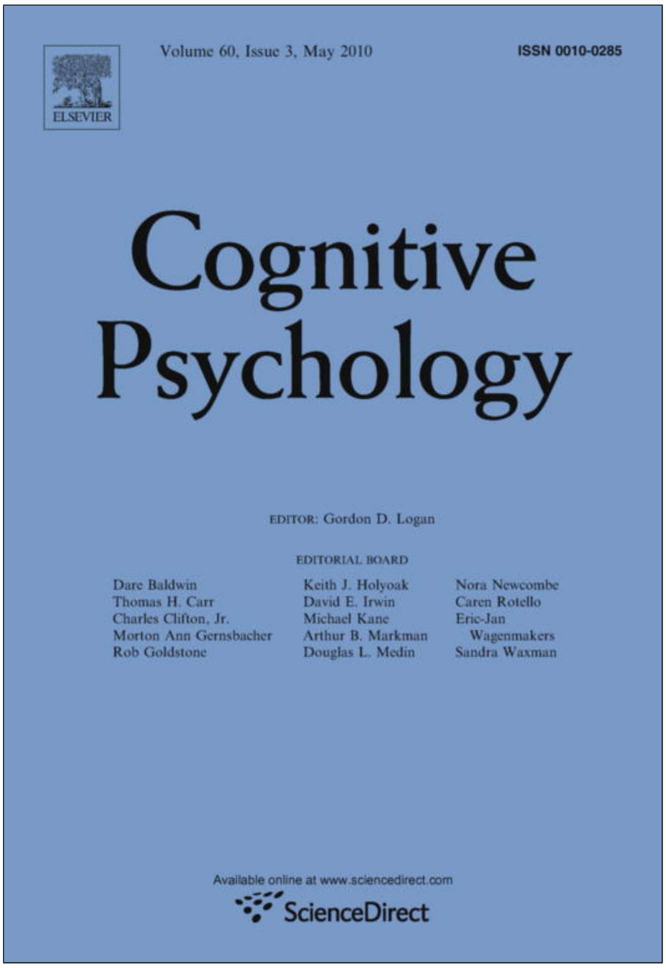
Volume 60, Issue 3, May 2010
ISSN 0010-0285
ELSEVIER
Cognitive Psychology
EDITOR: Gordon D. Logan
EDITORIAL BOARD
Dare Baldwin
Thomas H. Carr
Charles Clifton, Jr.
Morton Ann Gernsbacher
Rob Goldstone
Keith J. Holyoak
David E. Irwin
Michael Kane
Arthur B. Markman
Douglas L. Medin
Nora Newcombe
Caren Rotello
Eric-Jan Wagenmakers
Sandra Waxman
Available online at www.sciencedirect.com
ScienceDirect

Cognitive Psychology 60 (2010) 127–157

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Individual differences, aging, and IQ in two-choice tasks

Roger Ratcliff *, Anjali Thapar, Gail McKoon

The Ohio State University, Bryn Mawr College, United States

ARTICLE INFO

Article history:
Accepted 4 September 2009
Available online 4 December 2009

Keywords:
Reaction time
Aging
IQ
Diffusion model
Individual differences

ABSTRACT

The effects of aging and IQ on performance were examined in three two-choice tasks: numerosity discrimination, recognition memory, and lexical decision. The experimental data, accuracy, correct and error response times, and response time distributions, were well explained by Ratcliff's (1978) diffusion model. The components of processing identified by the model were compared across levels of IQ (ranging from 83 to 146) and age (college students, 60–74, and 75–90 year olds). Declines in performance with age were not significantly different for low compared to high IQ subjects. IQ but not age had large effects on the quality of the evidence that was obtained from a stimulus or memory, that is, the evidence upon which decisions were based. Applying the model to individual subjects, the components of processing identified by the model for individuals correlated across tasks. In addition, the model's predictions and the data were examined for the "worst performance rule", the finding that age and IQ have larger effects on slower responses than faster responses.

1. Introduction

A central finding in research on aging is that as people age, their response times (RTs) increase. In several simple two-choice tasks such as numerosity discrimination, recognition memory, and lexical decision, as well as in some perceptual tasks, the increase in RTs is coupled with little or no decrease in accuracy. This provides a puzzle: if RTs are used as the dependent measure, there appears to be a deficit in processing, but if accuracy is the dependent measure, there appears to be little or no deficit. Adding to the puzzle, RTs are sometimes, but not always, shorter for subjects with high IQs than subjects with low IQs, and accuracy is sometimes, but not always, greater (Jensen, 1987).

* Corresponding author. Address: Department of Psychology, The Ohio State University, Columbus, OH 43210, United States. Fax: +1 614 688 3984.
E-mail address: ratcliff.22@osu.edu (R. Ratcliff).

doi:10.1016/j.cogpsych.2009.09.001

Much of the work on ability and aging has employed global measures of ability. However, for any specific ability, there is no standard set of tests to measure it (see Bowles & Salthouse, 2008, for discussion of this issue). Moreover, sometimes a single task, for example, the WAIS digit/symbol coding task, is used as a measure of both speed of processing ability and reasoning ability. For any particular ability, it is often the case that different investigators choose different tests, and some of the tests have several variants. Because individual tests may have idiosyncrasies, the use of multiple subtests is recommended to attempt to average idiosyncrasies out (Bowles & Salthouse, 2008; Little, Lindenberger, & Nesselroade, 1999; Oberauer, Suβ, Wilhelm, & Wittmann, 2003; Tucker-Drob & Salthouse, 2008).

Our aim in this article was to conduct a study of aging, abilities, and individual differences that was at the same time both modest and comprehensive. The study was modest in that we selected only three tasks, three age groups, and two IQ measures. The study was comprehensive in that a processing model, Ratcliff's diffusion model (1978), was used to provide a complete analysis of the behavioral data, including accuracy and RTs for correct and error responses and the shapes of RT distributions. The study was also comprehensive in that we examined a wide range of IQs and a wide range of ages. We analyzed the data at the individual subject level, using individual subjects' data and components of processing (as identified by the model) in calculating correlations and in structural equation modeling.

Specifically, we tested three groups of subjects: college age, 60–74 year olds, and 75–90 year olds. Each subject participated in three tasks: a control task that makes little or no perceptual or memory demands (numerosity discrimination), a simple memory task (item recognition), and a task that assesses lexical knowledge (lexical decision). For the numerosity task, a number of asterisks (between 31 and 70) were displayed on a PC monitor and subjects were asked to judge whether the number was "large" or "small." For item recognition, subjects studied lists of single words and they were asked to judge whether test words had or had not occurred in a just-studied list. For lexical decision, subjects were asked to judge whether strings of letters were words. The subjects varied in IQ from 83 to 146, as measured by the WAIS-R matrix reasoning and vocabulary scales, a broader range than we have used in previous applications of the diffusion model to aging issues. We set 83 as the lower bound so that the subjects that were included in the study had low IQs but were still in the normal functioning range and would have no trouble with instructions. For these subjects and tasks, the aim was that the diffusion model provide a unified account of the effects of age and IQ.

In the study reported here, the memory task was item recognition, not other memory tasks such as associative memory or source memory. We used item recognition because the diffusion model has provided complete explanation of data in our earlier experiments (Thapar, Ratcliff, & McKoon, 2003; Ratcliff, Thapar, Gomez, & McKoon, 2004; Ratcliff, Thapar, & McKoon, 2001; Ratcliff, Thapar, & McKoon, 2003; Ratcliff, Thapar, & McKoon, 2004; Ratcliff, Thapar, & McKoon, 2006a; Ratcliff, Thapar, & McKoon, 2006b; Ratcliff, Thapar, & McKoon, 2007, henceforth referenced as RTM). For item recognition, other studies have found only small effects of age on memory (e.g., Balota, Dolan, & Duchek, 2000; Bowles & Poon, 1982; Craik, 1994; Craik & Jennings, 1992; Erber, 1974; Gordon & Clark, 1974; Kausler, 1994; Naveh-Benjamin, 2000; Neath, 1998, chap. 16; Old & Naveh-Benjamin, 2008; Rabinowitz, 1984; Schonfield & Robertson, 1966). However, most of these studies measured only accuracy, not RTs, even though older adults are typically slower than young adults. 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 and so the products of earlier operations are not fully available for later operations (e.g., Salthouse, 1996). In this context, findings that older adults show no deficits in accuracy for item recognition are surprising. In contrast to item recognition, age has a large effect on other memory tasks, such as associative recognition, cued and free recall, and source memory. Apart from associative recognition, these tasks are not amenable to diffusion model analyses and we focus on item recognition in this article (see Ratcliff, Thapar, & McKoon, in preparation, for an analysis of such memory paradigms).

Lexical decision has been a popular task with which to examine speed of processing. In contrast to many other tasks, it might be expected that lexical knowledge would improve with age. Over a lifetime of 60–70 years, the number of encounters for many words must greatly exceed the number of encounters in the first 20 years. Yet despite so many years of practice, lexical decision 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. However, while responses slow with age, accuracy does not change. Averaging over 22 lexical decision experiments, Myerson, Ferraro, Hale, and Lima (1992) found that error rates were about the same for old and young subjects. Thus, in lexical decision, like item recognition, older adults show a deficit in RTs but not accuracy.

Our third task, numerosity discrimination, has become a useful benchmark paradigm for examining experimental factors such as biases, optimality, and training, as well as the effects of such manipulations as sleep deprivation (e.g., Ratcliff & Van Dongen, 2009; Starns & Ratcliff, in press). We hypothesized that IQ effects would be smaller for this task because it does not require memory or vocabulary as item recognition and lexical decision do. The task shows the same patterns of results for young versus old subjects as item recognition memory and lexical decision, namely large changes in RT but small or non-existent changes in accuracy as a function of age. For all three tasks, the inclusion of the lower IQ subjects allows us to examine whether the lack of a deficit in accuracy as a function of age generalizes to lower IQ subjects.

In our previous work on aging and speed of processing (RTM papers), we have systematically examined the effects of aging on RTs and accuracy in a number of two-choice tasks, including the ones used here (see also Spaniol, Madden, & Voss, 2006). The data are well-described by Ratcliff's diffusion model for two-choice decisions (Ratcliff, 1978; Ratcliff, 1981; Ratcliff, 1985; Ratcliff, 1988; Ratcliff, 2002; Ratcliff, 2006; Ratcliff & McKoon, 2008; Ratcliff & Rouder, 1998; Ratcliff & Rouder, 2000; Ratcliff & Smith, 2004; Ratcliff, Van Zandt, & McKoon, 1999; Smith, 2000; Smith & Ratcliff, 2009; Smith, Ratcliff, & Wolfgang, 2004). From RTs and accuracy, the model abstracts estimates of the components of processing that underlie decisions: the quality of the information on which a decision is based, the criterial amounts of evidence that must be accumulated before a decision is made, and the time taken up by nondecision processes such as stimulus encoding, memory access, and response output. The model provides a framework in which to understand mappings from speed and accuracy measures of performance to components of processing. In addition to the tasks used in this study, the model has been successful in explaining performance for signal detection-like tasks, brightness discrimination with masked stimuli, recognition memory, lexical decision, and letter discrimination with masked stimuli (Gomez, Ratcliff, & Perea, 2007; Ratcliff, 2008a; Ratcliff, Gomez, & McKoon, 2004; Ratcliff et al., 1999; Smith et al., 2004; Voss, Rothermund, & Voss, 2004; Wagenmakers, Ratcliff, Gomez, & McKoon, 2008).

However, the diffusion model has not yet been used to evaluate effects of IQ on performance or interactions between IQ and aging (but see Schmiedek, Oberauer, Wilhelm, Suβ, & Wittmann, 2007, for some preliminary investigations). More generally, there is a need for studies that incorporate both a wide range of IQs and a wide range of ages (Deary, Der, & Ford; 2001; Roberts & Stankov, 1999).

RTM's experiments showed that older subjects (60–90 year olds) typically adopt more conservative decision criteria than college-age subjects, that is, they accumulate more evidence before making a response. In addition, the duration of the nondecision component is longer for them. However, most notably, the quality of the evidence upon which their decisions are based is often as good as that for college students. For numerosity discrimination, length discrimination, lexical decision, and recognition memory, RTM found, at most, slight decreases with age. There were significant declines only in two perceptual tasks, masked letter discrimination and masked brightness discrimination. 60–74 year olds showed a deficit for letter discrimination and 75–90 year olds showed deficits for both letter and brightness discrimination. These latter two findings are predictable from psychophysical research where it has been shown that deficits occur at earlier ages for high spatial frequency information like that used in masked letter discrimination than for low spatial frequency information like that used in brightness discrimination (Coyne, 1981; Fozard, 1990; Owsley, Sekuler, & Siemsen, 1983; Spear, 1993).

The diffusion model is especially useful in that it can be applied to data for individual subjects, separating out the components of processing for each one. Ratcliff et al. (2006a), for example, examined individual differences in performance in and across four tasks (numerosity discrimination, letter and brightness discrimination, and recognition memory) for 10 college students and two groups of 10 older subjects (60–74 and 75–90 year olds). Applying the model to the individual subjects' data, correlations were calculated across the four tasks for the three processing components (quality of evidence, criterial amounts of evidence, nondecision). The correlations were moderately high across the tasks for all three components. In other words, if a subject had a high value for one of the components in one task, he or she tended to have a high value for that component on the other tasks. In contrast, within tasks, RTM have routinely found little correlation between components of processing.

In RTM's studies, older subjects were matched to college students in terms of IQ, so the range of IQ values was relatively restricted (the lowest IQs were around 100) and most of the correlations with IQ were small. For the experiment reported here, the range of IQ values was larger (83–146, as mentioned above). We expected the model analyses to show larger individual differences among parameter values than in the RTM studies and thus to give a better understanding of which components of processing are consistent across tasks for what types of individuals. The broader range of IQ's and wider range of performance across individuals also provided stringent tests of the diffusion model.

Beyond individual differences, the experiment addressed two important issues. The first was the "worst performance" rule (Coyle, 2003). Often, it has been found that IQ correlates more strongly with longer RTs than shorter RTs. This has been labeled the worst performance rule because the longer RTs are assumed to come from more difficult test items. Below, we show how the diffusion model can predict this finding. We examined whether the rule applies across the large set of data provided by the experiment – the three tasks, the three age groups, and the wide range of IQs.

The second issue was age and IQ interactions. It could be that performance declines differentially at a slower rate for higher than lower IQ subjects as it might according to the cognitive reserve hypothesis (Satz, 1993; Stern, 2002). This hypothesis states that factors such as education and IQ may serve as a protective factor to mitigate the effects of aging on cognitive function. Alternatively, it could be that the decline is the same for all levels of IQ (cf. Salthouse, 2006). To anticipate the results of the experiment, we found that any decline is about the same for all subjects, both higher and lower IQ.

2. Experiment

The numerosity judgment task was used to provide a baseline against which lexical decision and recognition memory could be compared because it requires little perceptual or cognitive resources. On each trial an array of asterisks was displayed on a PC monitor screen and subjects were asked to judge whether the number was larger or smaller than 50. The other two tasks, recognition memory and lexical decision, are tasks that engage more central cognitive processes – memory and knowledge of words. In all three tasks, the independent variables were manipulated such that accuracy ranged from high to moderately low. Sweeping out RTs over a wide range of accuracy values provides maximal constraints on fitting the diffusion model to data (Ratcliff & Tuerlinckx, 2002).

2.1. Method

2.1.1. Subjects

Forty-five college-age subjects, 43 subjects from 60–74 years old, and 42 subjects from 75–90 years old participated in the experiment. The college-age subjects were recruited at Bryn Mawr College and in surrounding areas. The older adults were community-dwelling volunteers from the Bryn Mawr, PA, and Columbus, OH, areas. All subjects were paid for their participation. All subjects met the following inclusion criteria: a score of 26 or above on the Mini-Mental State Examination (Folstein, Folstein, & McHugh, 1975) and no evidence of disturbances in consciousness, medical or neurological disease causing cognitive impairment, history of head injury with loss of consciousness, or current psychiatric disorder. The subjects in both older adult age groups completed the Alzheimer Disease Assessment Scale – Cognitive Portion (ADAS-Cog; Rosen, Mohs, & Davis, 1984) and all had scores below 15. They also completed the Center for Epidemiological Studies – Depression scale (Radloff, 1977) and the only significant difference was that the CES-D score was higher for the college age students than for the 60–74 and 75–90 year old subjects. For IQ, subjects completed the Vocabulary and Matrix Reasoning subtests of the Wechsler Adult Intelligence Scale – 3rd Edition (WAIS-III; Wechsler, 1997). There were no differences among scaled Matrix Reasoning scores, but there was a difference between the unscaled (raw) scores for the college age group and the 60–74 and 75–90 year old groups. There were no other significant differences on any of the measures. The means and standard deviations of all these measures are shown in Table 1.

Table 1
Subject characteristics.

Measure	Young 18–25		Old 60–74		Very old 75–90	
	M	SD	M	SD	M	SD
Mean age	20.8	1.7	68.6	4.1	81.5	3.9
Years education	13.8	1.2	15.0	2.6	14.6	2.9
MMSE	28.6	1.3	27.9	1.6	27.3	1.3
WAIS-III vocabulary	12.6	2.9	11.7	2.9	12.6	3.4
WAIS-III matrix reasoning (scaled)	11.6	3.0	11.7	3.2	13.0	3.8
WAIS-III matrix reasoning (raw)	18.7	4.4	13.8	5.6	12.7	6.1
WAIS-III IQ	112.1	14.2	109.7	14.4	115.8	17.8
CES-D	13.7	7.8	9.5	7.1	10.3	8.1
ADAS-Cog	N/A	N/A	5.3	2.7	5.9	2.4

Note. MMSE = Mini-Mental State Examination; WAIS-III = Wechsler Adult Intelligence Scale-3rd edition; CES-D = Center for Epidemiological Studies-Depression Scale.

Each subject participated in four sessions, one to collect demographic and IQ information (i.e., the above measures) and one on each of the three tasks, in counterbalanced order. For all three tasks, subjects were instructed to respond quickly but not at the expense of making avoidable errors.

2.1.2. Stimuli

For recognition memory and lexical decision, the stimuli were high, low, and, only for lexical decision, very low frequency words. There were 800 high frequency words with frequencies from 78 to 10,600 per million (mean = 325, SD = 645, Kucera & Francis, 1967); 800 low frequency words, with frequencies of 4 and 5 per million (mean = 4.41, SD = 0.19); and 741 very low frequency words, with frequencies of 1 per million or no occurrence in the Kucera and Francis' corpus (mean = 0.365; SD = 0.48). All of the very low frequency words occurred in the Merriam-Webster Ninth Collegiate Dictionary (1990), and they were screened by three Northwestern undergraduate students; any words that they did not know were eliminated. For all three tasks, stimuli were chosen randomly without replacement from these pools. The stimuli were presented on the screen of a PC and responses were collected on the PC's keyboard.

2.1.3. Numerosity discrimination

On each trial, between 31 and 70 asterisks were placed in random positions in a 10×10 array of blank characters. Subjects were asked to press the "/" key if the number of displayed asterisks was "large" and the "z" key if the number was "small." There were 30 blocks of 40 trials with all stimuli presented once in each block. For data analyses, the numbers of asterisks were grouped into eight experimental conditions such that the mean RTs and accuracy values for the numbers grouped together were similar. "Small" responses to 31–50 asterisks and "large" responses to 51–70 asterisks were counted as correct. Subjects were given examples of large and small numbers of asterisks at the beginning of the session. There was one practice block of trials for which there was error feedback to aid subjects in calibration of the large–small dimension but there was no feedback in later blocks.

2.1.4. Recognition memory

There were 26 study-test blocks. For each block, the study list consisted of eight high and eight low frequency words displayed for 1 s each. Four of the high and four of the low frequency words were presented once and four of each were presented twice. One additional filler word, a very low frequency word, was placed at the end of the study list to serve as a buffer item. The test list immediately followed the study list and consisted of the 16 studied words plus 16 new words, eight high and eight low frequency. The first two test words in the test list were fillers, either two new very low frequency words or one new very low frequency word and the last, filler, item of the study list. Subjects were asked to press the "/" if the test word had been presented in the immediately preceding study list and the "z" key if not. There was no error feedback.

2.1.5. Lexical decision

Words were selected from the high, low, and very low frequency pools and nonwords were selected from a pool of 2341 pseudowords that were generated from words by randomly replacing all the vowels with other vowels (except for "u" after "q"). There were 70 blocks of trials with each block containing 30 letter strings: five high frequency words, five low frequency words, five very low frequency words, and 15 pseudo words. Subjects were asked to press the "/" key if the letter string was a word and the 'z' key if it was not. There was no error feedback.

3. Diffusion model

The diffusion model is designed to explain the cognitive processes involved in making simple two-choice decisions. As described above, the model separates the quality of evidence entering a decision from the decision criteria and from nondecision processes. Decisions are made by a noisy process that accumulates information 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 the quality of match between the test item and memory in memory and lexical decision tasks. The mean of the distribution of times taken up by the nondecision component (the combination of encoding, response execution, and so on) is labeled 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).

The values of the components of processing vary from trial to trial, under the assumption that subjects cannot accurately set the same parameter values from one trial to another (e.g., Laming, 1968; Ratcliff, 1978). Across trial variability in drift rate is normally distributed with SD η, across trial variability in starting point is uniformly distributed with range s_z, and across trial variability in the nondecision component is uniformly distributed with range s_t. Also, there are "contaminant" responses – slow outlier response times as well as responses that are spurious in that they do not come from the decision process of interest (e.g., distraction, lack of attention). To accommodate these responses, we assume that, on some proportion of trials (p_o), a uniform distributed random delay between the minimum and maximum RT for the condition is added to the decision RT (see Ratcliff & Tuerlinckx, 2002). The assumption of a uniform distribution is not critical; recovery of diffusion model parameters is robust to the form of the distribution (Ratcliff, 2008b).

The values of all the parameters, including the variability parameters, are estimated simultaneously from data by fitting the model to all the data from all the conditions of an experiment. The model can successfully fit data from single subjects if there are around 400–1000 total observations per subject, which typically takes about 45 min for the kinds of tasks considered in this article. Variability in the parameter estimates is much less than differences in the parameters across subjects so that correlations are meaningful and not contaminated by noise to a high degree. The model can be understood as decomposing accuracy and RT data for correct and error responses into components of processing.

For the numerosity discrimination experiment, we initially assumed that drift rates were equal and opposite for "small" responses to small stimuli and "large" responses to large stimuli. For example, the drift rate for 31–35 asterisks would have the opposite sign but the same numerical value as the drift rate for 66–70 asterisks. However, for many of the subjects, the zero point of drift was biased. Some subjects put the zero point of drift toward 60 asterisks and some put it toward 40 asterisks. To accommodate these biases, we used a drift criterion, a value added or subtracted from the drift rate (like the criterion in signal detection theory; see Ratcliff & McKoon, 2008, for a detailed discussion). The addition of a drift criterion can, for example, make the drift rate for the condition with 31–35 asterisks larger numerically than the drift rate for the condition with 66–70 asterisks.

The diffusion model was fit to the data for each task and each subject by minimizing a chi-square value with a general SIMPLEX minimization routine that adjusts the parameters of the model until it finds the parameter estimates that give the minimum chi-square value (see Ratcliff & Tuerlinckx, 2002, for a full description of the method). The data entered into the minimization routine for each experimental condition were the 0.1, 0.3, 0.5, 0.7, 0.9 quantile RTs for correct and error responses and the corresponding accuracy values. The quantile RTs and the diffusion model were used to generate the predicted cumulative probability of a response by that quantile response time. Subtracting the cumulative probabilities for each successive quantile from the next higher quantile gives the proportion of responses between adjacent quantiles. For the chi-square computation, these are the expected values, to be compared to the observed proportions of responses between the quantiles (i.e., the proportions between 0, 0.1, 0.3, 0.5, 0.7, 0.9, and 1.0, which are 0.1, 0.2, 0.2, 0.2, 0.2, and 0.1) multiplied by the number of observations. Summing over $(\text{observed–expected})^2/\text{expected}$ for all conditions gives a single chi-square value to be minimized.

The diffusion model is tightly constrained. The most powerful constraint comes from the requirement that the model fit the right-skewed shape of RT distributions (Ratcliff, 1978; Ratcliff & McKoon, 2008; Ratcliff et al., 1999). In addition, changes in response probabilities, quantile RTs, and the relative speeds of correct and error responses across experimental conditions that vary in difficulty are all captured by changes in only one parameter of the model, drift rate. The other parameters cannot vary across levels of difficulty. For the response criteria, subjects could only set them as a function of difficulty if they already knew, before the accumulation process started, what the level of difficulty would be. For the nondecision component, we usually assume that the duration of stimulus encoding, matching against memory, response output, and other such nondecision processes do not vary with difficulty.

4. Results: RTs and accuracy

For the college students, correct RTs less than 300 ms and greater than 3000 ms were eliminated from analyses. This excluded 2.3%, 3.7% (one subject contributed 1.3% of the 3.7%) and 2.9% (the same subject contributed 1.2% of the 2.9%) of the data for the numerosity, recognition memory, and lexical decision tasks, respectively. For the 60–74 year old subjects, the cutoffs were 300 ms and 3500 ms, excluding 3.2%, 0.6%, and 1.4% of the data, and for the 75–90 year old subjects, the cutoffs were 300 ms and 4500 ms, excluding 4.1%, 0.6%, and 1.3% of the data from the numerosity, recognition memory, and lexical decision tasks, respectively.

Overall, the effects of age on performance replicated those obtained by RTM. Fig. 1 shows how accuracy and median RTs varied across experimental conditions as a function of age. Plots are shown for the two responses for each task (‘‘large” and ‘‘small” numbers of asterisks, ‘‘old” and ‘‘new” responses in recognition memory, and ‘‘word” and ‘‘nonword” in lexical decision).

For the numerosity task, the RT functions for ‘‘large” and ‘‘small” responses appear to be different from each other but (as mentioned above) this is because many of the subjects were biased, setting the average midpoint at around 55 asterisks rather than the correct value, 50. If the functions for ‘‘large” and ‘‘small” responses were flipped left to right and the midpoint set to around 55, the two functions would be similar, with six out of the eight points on each function aligning with each other. For recognition memory and lexical decision, positive responses could not mirror negative responses because there were different numbers of conditions for positive and negative responses and the stimuli were not inherently symmetrical.

Overall, there was little difference among the age groups in accuracy. For these analyses, all the conditions of each experiment were averaged. For numerosity discrimination, mean accuracy values were 0.80, 0.83, and 0.81, for the college-age, 60–74, and 75–90 year olds, respectively, $F(2, 127) = 3.72$, $p < 0.05$, for recognition memory, the means were 0.75, 0.76, and 0.75, for the college-age, 60–74, and 75–90 year olds, respectively, $F(2, 127) = 0.17$, $p > 0.05$, but for lexical decision, the means were 0.86, 0.93, and 0.93, for the college-age, 60–74, and 75–90 year olds, respectively, $F(2, 127) = 18.07$, $p < 0.05$.

There were large differences among the age groups in median RTs, with college students faster than 60–74 year olds and 60–74 year olds faster than 75–90 year olds (for numerosity discrimination, medians 605, 883, and 1028 ms, respectively, $F(2, 127) = 33.12$, $p < 0.05$, for recognition memory, medians of 668, 866, and 934 ms, respectively, $F(2, 127) = 41.34$, $p < 0.05$, and for lexical decision, medians of 659, 893, and 948 ms, respectively, $F(2, 127) = 433.22$, $p < 0.05$).

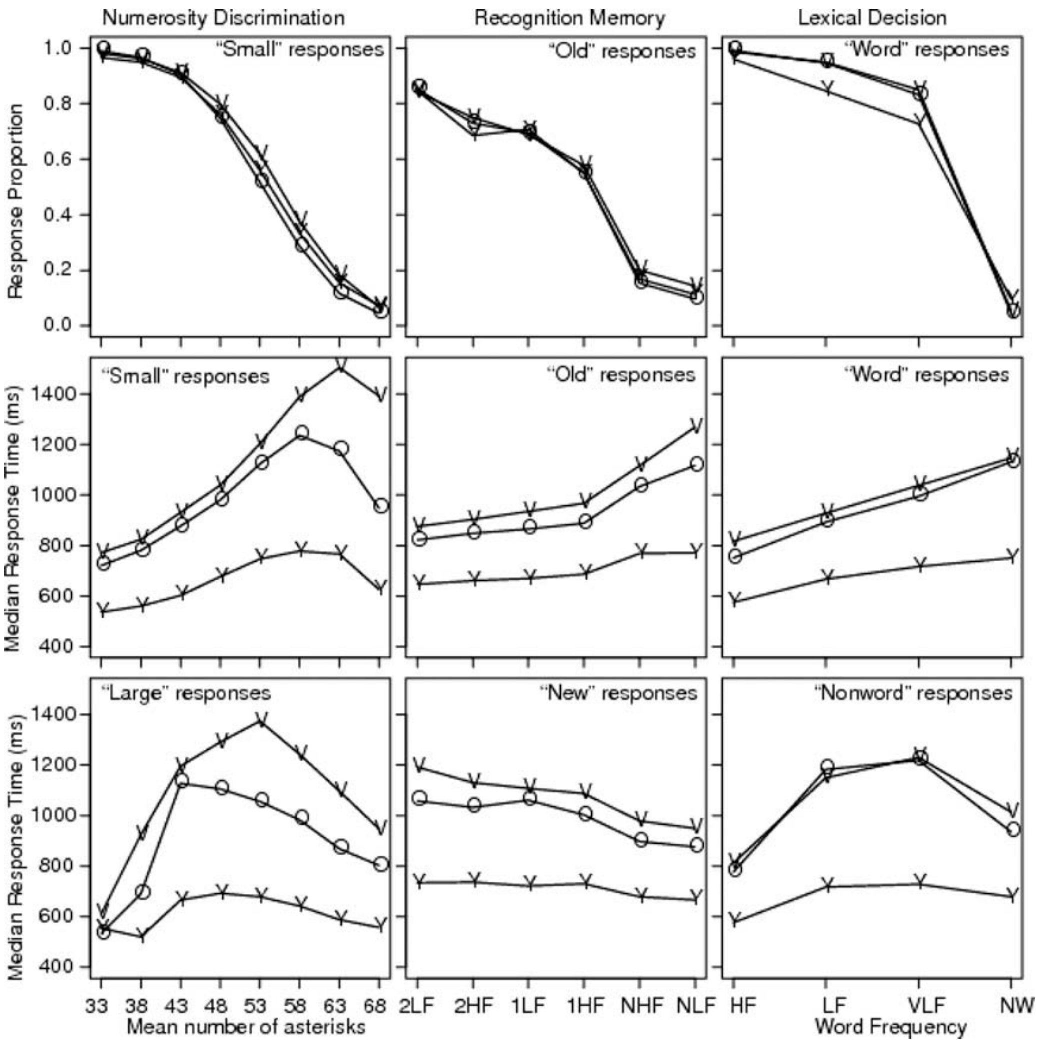

Fig. 1. Accuracy and median RT as a function of experimental condition for the three tasks (numerosity discrimination, recognition memory, and lexical decision) for the three subject groups, college age ("Y"), 60–74 year olds ("O"), and 75–90 year olds ("V"). HF = high frequency words, LF = low frequency words, VLF = very low frequency words, and NW = nonwords. For recognition memory, the digits 1 and 2 (in front of HF and LF) refer to the number of presentations and *N* refers to new words.

For all three groups of subjects, median RTs for correct responses increased as accuracy decreased, as would be expected. The pattern for error responses is more complicated: median error RTs decreased as error rate decreased in numerosity discrimination and in lexical decision for "nonword" responses. But they increased as error rate decreased in recognition memory and for "word" responses in lexical decision.

Fig. 2 shows how accuracy and median RTs varied with age, task, and IQ, averaged over the more accurate conditions for each task (i.e., excluding the conditions for which accuracy was near chance). The conditions were: the ranges of asterisks with means of 33, 38, 43, 58, 63, and 68 in numerosity discrimination; new items and twice presented old items in recognition memory, and all of the conditions for lexical decision.

The main findings for IQ are that accuracy increased with IQ for lexical decision and recognition, but there were at most only small changes in correct median RTs as a function of IQ. For the numerosity task, used as a baseline, IQ did not significantly affect accuracy or speed.

Figs. 1 and 2 summarize the data that the diffusion model must explain: performance on multiple conditions in each of three tasks for a wide range of ages and a wide range of IQs. In all cases, the model must explain accuracy, the shapes of the RT distributions for correct and error responses, and the relative speeds of correct and error responses.

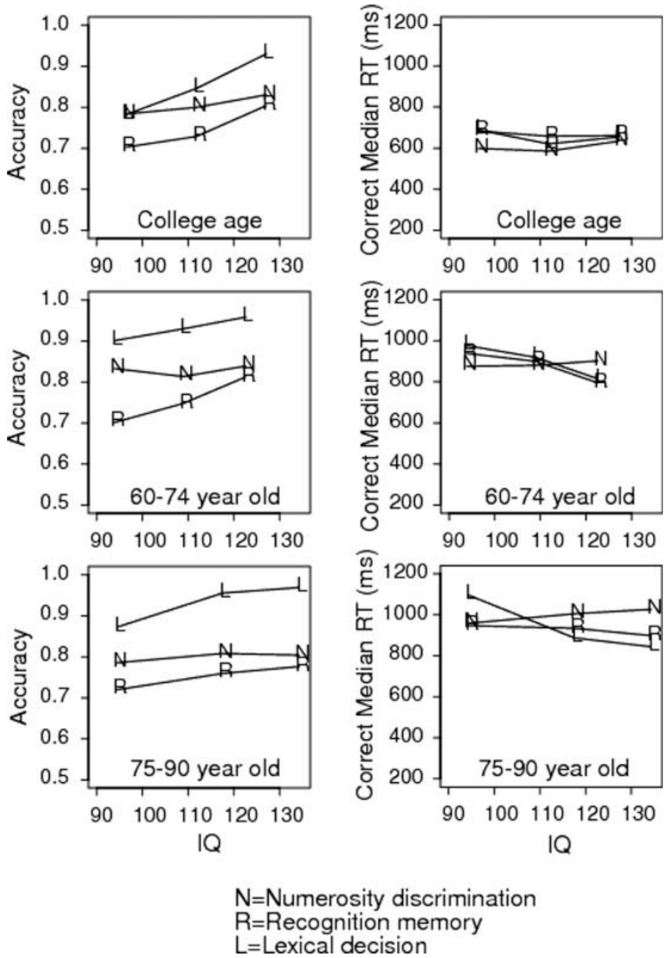

Fig. 2. Accuracy and median RT as a function of IQ. Subjects were divided into three groups by IQ, task, and age group. Conditions for which accuracy was near chance were excluded. The conditions used were: the ranges of asterisks with means of 33, 38, 43, 58, 63, and 68 for numerosity discrimination; new items and twice presented old items for recognition memory, and all of the conditions for lexical decision.

5. Diffusion model analyses

In this section, we show that the model fit the data well. In the sections following, interpretations of the data and the components of processing identified by the model for aging and IQ are discussed.

The model was applied to the data for each task for each subject individually. To show all of the empirical data and all of the model's predicted data for all of the conditions for all three tasks for all of the subjects would require far too many figures or tables. Instead, we illustrate that the model's predicted data match the empirical data in two ways. First, for each experimental condition for each task, we averaged the data from all the subjects. In most of the RTM papers, comparisons have been made between the average of model parameters across fits to individuals, and fits to group data. There have been almost no cases for individual parameters in which the means were significantly different. Second, we picked one condition with average accuracy lower than ceiling (to provide the widest range across subjects) for each task to compare the empirical data and predicted values in plots of accuracy values and each of the quantile RTs for each of the individual subjects.

The model fit the data well, and it did so conforming to the crucial assumption mentioned above that only drift rate, not the nondecision component or the criteria, was allowed to vary with the difficulty of experimental conditions. For instance, the slower and less accurate RTs for low compared to high frequency words in lexical decision should be explained only by a difference in their drift rates. It is assumed that boundary separation and the nondecision component do not vary with difficulty because it is assumed that they are not adjustable as a function of the amount of evidence accumulated early during the decision process. Also, because the decision process is highly stochastic, using early accumulated evidence would be a very inaccurate way of adjusting their values.

5.1. Averaged data: quantile probability functions

We use quantile probability functions to display the quality of the model's fits to data. For each condition in each task, the 0.1, 0.3, 0.5, 0.7, and 0.9 quantile RTs are plotted as a function of the proportions of correct and error responses for that condition. Fig. 3 shows the quantile probability functions averaged over subjects for each condition in each task with *x*'s for the data and *o*'s for the values predicted from the model's best-fitting parameter values. Each column of RTs (the 0.1, 0.3, 0.5, 0.7, and 0.9 quantiles) represents one experimental condition. The points to the right for each plot represent correct responses and those to the left, error responses. For example, in lexical decision the proportion of "word" responses for very low frequency words was about 0.8, on the right, and the proportion of "nonword" responses was about 0.2, on the left. Responses to high frequency words have the highest probability of correct responses and so their quantiles for correct responses are farthest right and their quantiles for error responses are farthest left.

For numerosity discrimination, all the conditions in the experiment are displayed in a single plot. There are 16 columns of RT quantiles. The eight on the right half are correct responses to "small" numbers and correct responses to "large" numbers. The eight on the left are errors, "small" responses for "large" numbers and "large" responses for "small" numbers. Plotting all the data together like this was possible because the starting point of the diffusion process (z) was about halfway between the two

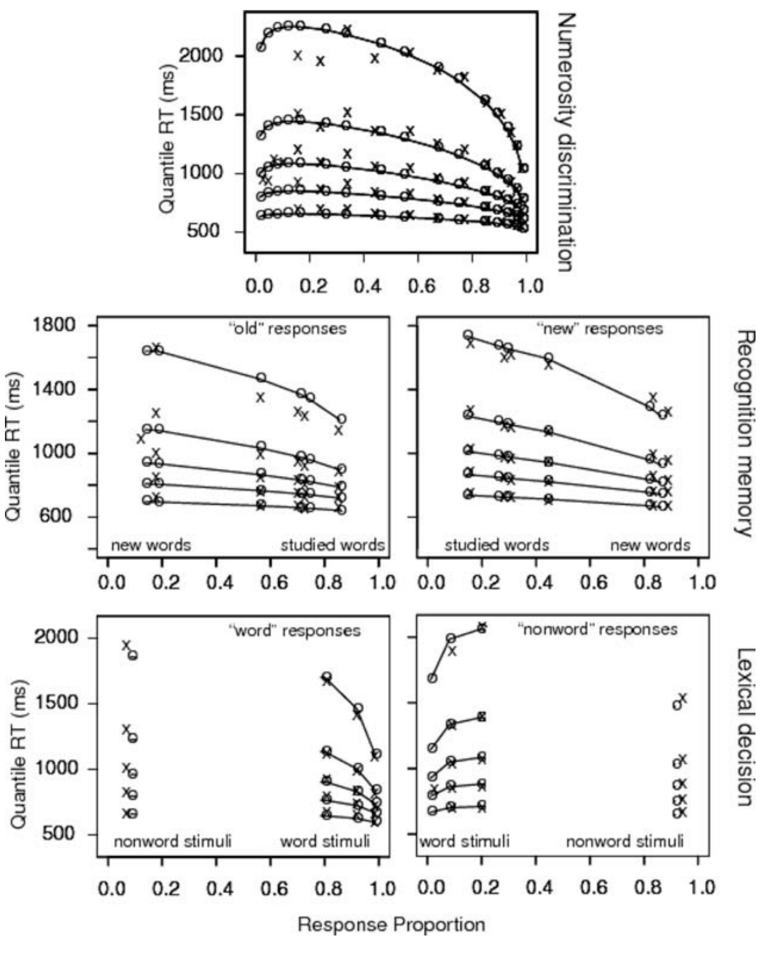

Fig. 3. Quantile probability plots for the three tasks for data averaged over all subject groups. The *x*'s are the data and the *o*'s are the predictions joined by the lines. The five lines stacked vertically above each other are the values predicted by the diffusion model for the 0.1, 0.3, 0.5, 0.7, and 0.9 quantile RTs as a function of response proportion for the conditions of the experiments.

boundaries, making the functions for "small" responses and "large" responses nearly symmetric (though shifted on the drift rate scale to have a zero point of drift about 55, averaged across subjects).

In contrast, the recognition memory and lexical decision functions are shown in two panels, one for "old" or "word" responses and one for "new" or "nonword" responses, because the estimate of the starting point was not halfway between the boundaries and so "old" and "new" responses and "word" and "nonword" responses have different vertical locations on the quantile probability plots, with the location on the *x*-axis for each of the experimental conditions determined by its difficulty.

The data predicted by the model, derived from the model parameter values that provided the best fit to the empirical data, match the empirical values well. Predicted response proportions are within a few percent of the empirical values and quantile RTs are within 10's of milliseconds of the empirical values. There were only two exceptions: in the numerosity task, the 0.9 quantiles for errors for three of the conditions missed by 50–200 ms, and in the recognition memory task, the 0.9 quantiles for correct "old" responses missed by about 50 ms. It is not surprising that the model's predictions miss in a few cases. The range of data combined over ages and IQ levels is wide, so the good fit of the model to the average data is impressive.

For some of the error conditions, some subjects had fewer than the five responses needed to compute five quantiles and so only the median is plotted in the figures. The subjects for whom there were fewer than five error responses were the slower, more accurate subjects, so leaving them out of the analyses would have led to a bias in parameter estimates toward the faster, less accurate subjects.

5.2. Individual subjects

For each subject, there were eight conditions for numerosity discrimination, six for recognition memory, and four for lexical decision. To illustrate the fits of the model to the data, for each task, we chose one condition for which average accuracy was not at ceiling (1.0) or floor (0.5). The conditions were 41–45 asterisks for numerosity discrimination, high frequency words presented twice for recognition memory, and very low frequency words for lexical decision. For all 130 subjects, Figs. 4 and 5 plot the empirical values of accuracy and RT quantiles against the values predicted from the best-fitting parameters of the diffusion model.

If the data and predictions matched exactly, the slopes would have a slope of 1.0. The functions are close to 1.0, which is especially impressive given that there was only one session of data per task, that there were 130 subjects, and that there were large individual differences among the subjects with response proportions varying from 0.5 to close to 1.0, 0.1 quantile RTs varying from 400 ms to 1000 ms, and 0.9 quantile RTs varying from 800 ms to 3000 ms. The match between the data and the diffusion model predictions was equally close for all the other conditions that are not plotted in the figure.

In general, there were more observations for correct responses, so deviations between the model and data for correct responses are more serious than deviations for error responses. There were only a few subjects for whom the model's miss was larger than 10% (Fig. 4): five subjects for numerosity discrimination, six for recognition memory, and none for lexical decision.

Overall, the lower RT quantiles were less variable than the higher quantiles so deviations between experimental and predicted 0.1 quantiles are more serious than for the higher quantiles. There were again only a few instances in which the model predictions seriously missed the data.

For correct responses for numerosity discrimination (Fig. 4), there were 5 subjects for whom there were serious misses in the 0.1 and 0.3 quantile RTs. These misses occurred for 60–74 and 75–90 year olds who appear to have adopted a strategy not consistent with the diffusion model: For the easy conditions, they responded relatively quickly, but for the more difficult conditions (e.g., 45–55 asterisks), their responses were delayed by 200–300 ms. The model had to average over these conditions, leading to predicted 0.1 and 0.3 quantile RTs larger than the empirical values. There were also three subjects with large misses in the 0.9 quantile RTs.

For correct responses for recognition memory (Fig. 4), there were no serious misses between predictions and data in the 0.1 and 0.3 quantile RTs and only 3 or 4 misses in the higher quantiles. For correct responses for lexical decision (Fig. 4), there are 4 misses in the 0.1 and 0.3 quantile RTs and 4 misses in the higher quantiles.

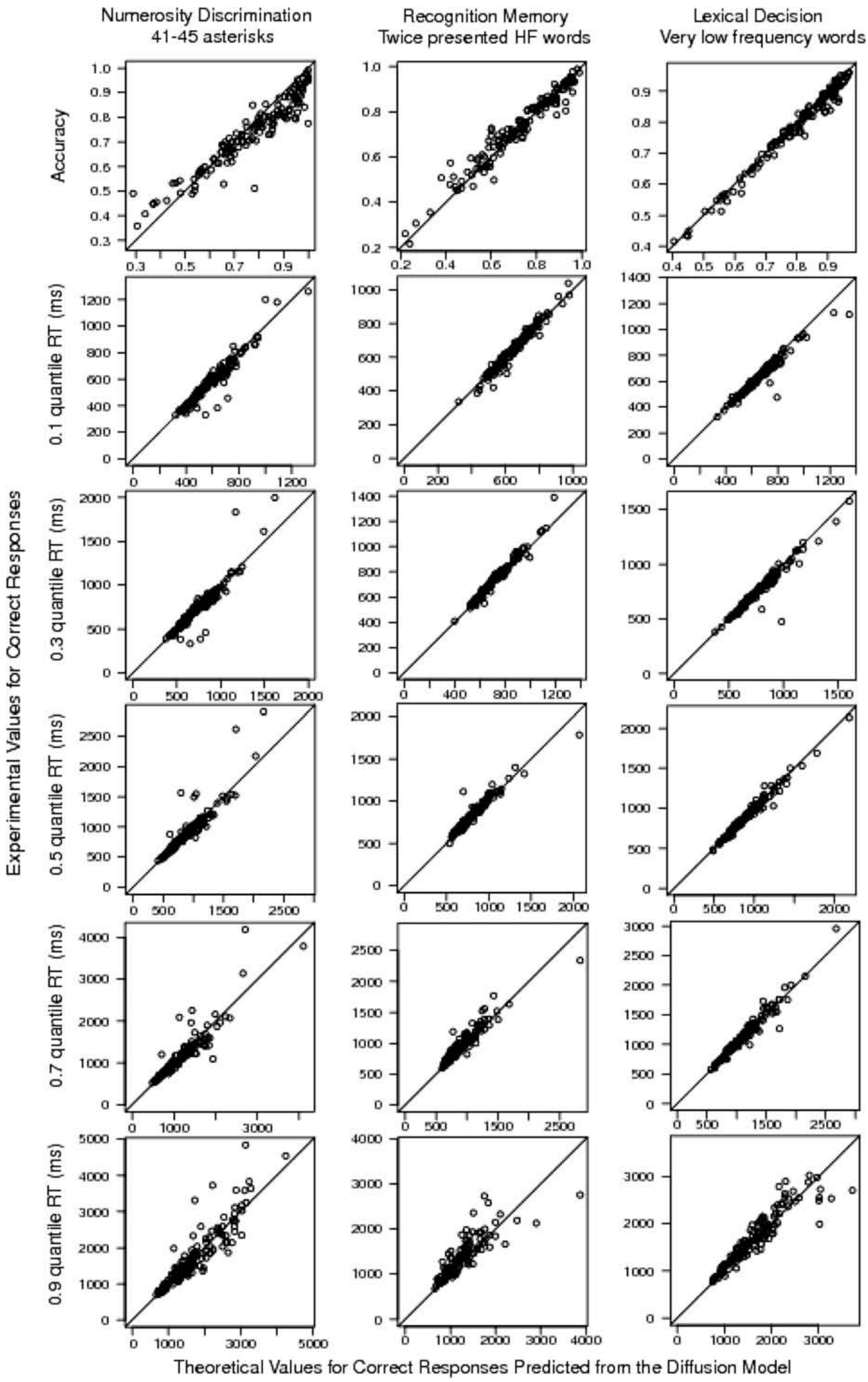

Fig. 4. Plots of accuracy and RT quantiles for data (y-axis) and predicted values from fits of the diffusion model (x-axis) for correct responses for a single condition for each of the three tasks for all subjects in the three age groups.

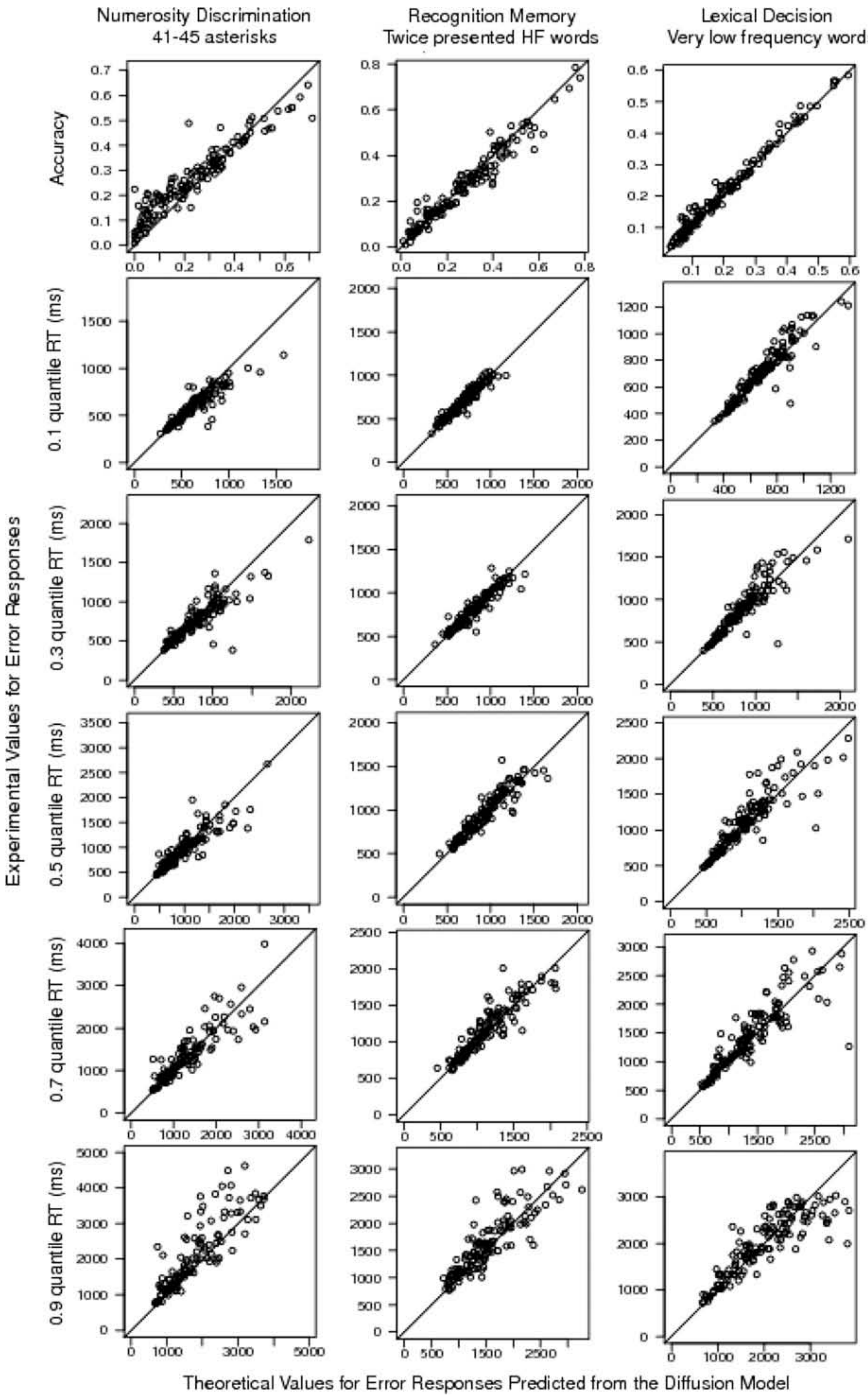

Fig. 5. Plots of accuracy and RT quantiles for data (*y*-axis) and predicted values from fits of the diffusion model (*x*-axis) for error responses for a single condition for each of the three tasks for all subjects in the three age groups.

The RT quantiles for errors, Fig. 5, show about the same quality of fits as for the correct RT quantiles but with much greater variability than for correct responses. For numerosity discrimination and recognition memory, some of the conditions had five or fewer error RTs and their RTs were excluded and so the variability in the data was much greater than for correct RT quantiles. This was less of a problem for lexical decision, for which there were at least 12 observations for each of the quantile RTs.

5.3. Chi-square goodness-of-fit

We calculated chi-square goodness-of-fit values for each task for each subject and the means and SDs of the chi-square values are shown in Table 2. The means are similar to values reported in previous studies (RTM). The degrees of freedom for the chi-square values were calculated as follows: for the five quantile RTs, there are six bins: two outside the 0.1 and 0.9 quantiles and four between the pairs of quantiles. This gives 12 degrees of freedom, −1 because the total probability adds to 1. Thus, for the numerosity discrimination task with eight conditions, the number of degrees of freedom is 76: 88 in the eight conditions −12 for the number of parameters for the model. For recognition memory with six conditions and 13 model parameters, there were 53 degrees of freedom, and for lexical decision with four conditions and 11 parameters, there were 33. The 0.95 critical value for the numerosity task is 97.4, for the recognition memory task it is 71.0, and for the lexical decision task it is 47.4. The average chi-square values (Table 2) range from the critical value up to two times the critical value.

The chi-square statistic has the property that as the number of observations increases, the power of the test increases so that even the smallest deviation can lead to significance. To illustrate this: 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 computations, the observed and expected proportions between two adjacent bins systematically miss by 0.1 (e.g., instead of the proportions being 0.2, one is 0.1 and the next is 0.3). Then the additional contribution from this miss to the chi-square is $N((0.1-0.2)^2/0.3 + N(0.3-0.2)^2/0.1)$, where N is the number of observations in the condition. For the numerosity discrimination task with about $N = 140$ per condition, the contribution to the chi-square from this systematic deviation would be 18.7. This means that a single systematic miss in one of the eight conditions' 10 quantile RTs (one miss out of 80) can add about 19% of the critical value of the chi-square (in Numerosity). The contributions for the recognition memory and lexical decision

Table 2
Means and SDs in parameter values for subject groups and tasks.

Parameter	Task	Subjects	a	z	T_{er}	η	s_z	p_o	s_t	χ^2
Mean	Numerosity	College	0.170	0.083	0.373	0.191	0.084	0.018	0.165	119
		60–74	0.224	0.106	0.484	0.164	0.072	0.018	0.168	146
		75–90	0.257	0.128	0.462	0.150	0.078	0.025	0.181	132
	Recognition	College	0.141	0.069	0.489	0.230	0.063	0.012	0.189	100
		60–74	0.170	0.074	0.632	0.237	0.037	0.002	0.194	92
		75–90	0.182	0.077	0.643	0.214	0.031	0.009	0.200	88
	Lexical	College	0.157	0.080	0.429	0.139	0.072	0.029	0.149	97
		60–74	0.204	0.095	0.539	0.113	0.028	0.025	0.154	79
		75–90	0.213	0.101	0.572	0.129	0.040	0.032	0.143	77
SD	Numerosity	College	0.057	0.027	0.034	0.064	0.043	0.030	0.050	37
		60–74	0.089	0.039	0.102	0.065	0.056	0.026	0.117	92
		75–90	0.070	0.038	0.114	0.069	0.060	0.041	0.128	54
	Recognition	College	0.040	0.019	0.052	0.072	0.042	0.021	0.082	44
		60–74	0.047	0.025	0.096	0.076	0.040	0.005	0.082	30
		75–90	0.049	0.028	0.087	0.081	0.036	0.020	0.102	26
	Lexical	College	0.047	0.024	0.043	0.090	0.049	0.034	0.075	37
		60–74	0.051	0.025	0.069	0.045	0.031	0.032	0.094	43
		75–90	0.048	0.026	0.085	0.069	0.041	0.037	0.087	47

Note: 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_t = range of the distribution of nondecision times, and χ^2 is the chi-square goodness-of-fit measure.

tasks would be greater because there are fewer conditions. This suggests that if the chi-square statistic is to be used to evaluate goodness-of-fit, then the size of any systematic deviations must be considered large enough that the model becomes unattractive. For the three tasks in this experiment, we believe that the chi-square misses are small enough (see the fits in Figs. 4 and 5) that we can assume the model fits well and proceed to examine aging and IQ effects on the parameters of the model, i.e., on the components of processing identified by the model.

5.4. Age, IQ, and components of processing

5.4.1. Components of processing as a function of age

Older subjects differed from college students in having higher criteria and longer nondecision times. Most importantly, they did not differ in drift rates for lexical decision or recognition. This means that the quality of the information entering the decision process was not significantly affected by age.

Tables 2 and 3 show the best-fitting parameter values for the three subject groups and the three tasks (averaged over IQ). The tables also show the standard deviations in the values, computed across subjects. Significance can be assessed by computing *z* statistics from the means and standard deviations in the tables.

For numerosity discrimination, there were changes in drift rate across the age groups. This does not replicate other studies with this task (Ratcliff, 2008a; Ratcliff et al., 2001; Ratcliff et al., 2007). However, the reason drift rates varied here is that variability in drift rate across trials was larger for the college-age subjects than the 60–74 year old subjects, and larger for the 60–74 year old subjects than the 75–90 year old subjects (this was not true for recognition memory and lexical decision). The analog to this in signal detection theory is that a larger difference in means and a larger SD can lead to no change in d'. In the earlier studies, SD in drift across trials did not vary systematically across age groups.

The findings for lexical decision and recognition broadly replicate those from the RTM articles. They extend application of the model to a significantly larger range of IQ's (83–146). For the RTM studies, older subjects' IQs were matched to college students', which limited the IQ range.

Table 3
Means and SDs in drift rates for subject groups and tasks.

Parameter	Task	Subjects	Drift rates						Drift criterion
Mean	Numerosity	College	0.451	0.342	0.206	0.069	–	–	0.078
		60–74	0.430	0.324	0.197	0.064	–	–	0.075
		75–90	0.372	0.280	0.169	0.056	–	–	0.095
	Recognition	College	0.159	0.334	0.052	0.168	−0.266	−0.328	–
		60–74	0.196	0.297	0.040	0.138	−0.291	−0.352	–
		75–90	0.192	0.271	0.044	0.113	−0.249	−0.317	–
	Lexical	College	0.457	0.227	0.127	−0.240	–	–	–
		60–74	0.412	0.238	0.141	−0.253	–	–	–
		75–90	0.437	0.280	0.169	−0.249	–	–	–
SD	Numerosity	College	0.141	0.105	0.062	0.035	–	–	0.070
		60–74	0.136	0.100	0.060	0.030	–	–	0.086
		75–90	0.164	0.123	0.072	0.033	–	–	0.097
	Recognition	College	0.115	0.170	0.087	0.103	0.131	0.143	–
		60–74	0.172	0.159	0.110	0.114	0.125	0.138	–
		75–90	0.143	0.160	0.107	0.099	0.144	0.174	–
	Lexical	College	0.130	0.083	0.078	0.081	–	–	–
		60–74	0.136	0.102	0.077	0.086	–	–	–
		75–90	0.201	0.145	0.102	0.123	–	–	–

The drift rate conditions are: for numerosity discrimination, 31–35 and 66–70 asterisks, 36–40 and 61–65 asterisks, 41–45 and 56–60 asterisks, 46–50 and 51–55 asterisks. For recognition memory, 1*P* low frequency, 2*P* low frequency, 1*P* high frequency, 2*P* high frequency, new high frequency, new low frequency (where *P* represents the number of presentations). For lexical decision, high frequency, low frequency, very low frequency, nonwords. The drift criterion is added to the drift rate for low numbers of asterisks and taken away from the drift rate for high number of asterisks.

It might be thought that older subjects would show more variability across trials in their drift rates, nondecision times, and starting point, but there were differences in variability only for starting point in recognition and lexical decision, and the differences were such that there was more variability for the college students than the older subjects.

5.4.2. Correlations between IQ and accuracy and between IQ and median RTs

The correlations in the following subsections were computed for each age group for each task and then, because the correlations were similar across the age groups, they were averaged.

Table 4 shows the correlations between IQ and accuracy and between IQ and median RT. For IQ and accuracy, the correlations are positive and quite large for lexical decision and recognition memory. For IQ and median RTs, the correlations were negative, small for numerosity discrimination and a little larger for recognition memory and lexical decision. So, overall the higher IQ subjects tended to respond more accurately and more quickly than the lower IQ subjects, but not strongly so for RT.

To assess the significance level of the correlations, a correlation of 0.29 with 42 degrees of freedom would be significant at the 0.05 level for a two-tailed test (there were 45, 43, and 42 subjects in our three groups). As a rule of thumb, we considered only correlations larger than this as meaningful.

5.4.3. Correlations between components of processing and accuracy and median RTs

In the diffusion model, larger values of boundary separation and the nondecision component produce longer median RTs. Also, larger values of drift rate produce higher accuracy and shorter median RTs.

For all three tasks, the data followed this pattern. Drift rate was strongly positively correlated with accuracy and negatively correlated with median RTs (Table 4). Boundary separation and the nondecision component were positively and strongly correlated with median RTs but not accuracy.

These findings replicate RTM's findings except that the correlations of model parameters and data were larger than in the RTM studies because of the wider range of abilities and performance. For example, in recognition memory in Ratcliff et al. (2004), SDs across subjects were almost half the size of those reported in Tables 2 and 3. The correlation between the nondecision component and median RT was strong here but not in the earlier studies, and correlations between drift rate and median RTs were strong here but not in the earlier studies.

5.4.4. Correlations between IQ and components of processing

In this section, we report correlations between IQ and components of processing and in later sections, we show plots of the components as a function of IQ. The correlations were computed for each age group for each task and then averaged.

As might be expected, IQ correlated with drift rate for the lexical decision and recognition memory tasks (Table 5), indicating that higher IQ subjects are better at remembering words and better at discriminating words from nonwords. This is not surprising because vocabulary was one of the subtests of IQ we used. In contrast, for the numerosity discrimination task, the correlation between IQ and drift rate was weaker, indicating that the ability to judge numerosity is not as highly related to either subtest of our IQ measure, matrix reasoning or vocabulary.

IQ did not correlate significantly with the nondecision component or with boundary separation in lexical decision or recognition. For numerosity discrimination, the IQ-boundary correlation was

Table 4
Correlations of IQ and model parameters with accuracy and median RT.

Experiment	Measure	IQ	a	T_{er}	v	Median RT
Numerosity discrimination	Accuracy	0.242	0.243	0.164	0.441	0.055
	Median RT	0.110	0.620	0.303	−0.483	
Recognition memory	Accuracy	0.502	0.105	0.137	0.861	−0.250
	Median RT	−0.208	0.468	0.558	−0.430	
Lexical decision	Accuracy	0.721	0.082	−0.116	0.576	−0.247
	Median RT	−0.319	0.626	0.446	−0.561	

Table 5
Correlations between IQ and components of processing in each task and between the components of processing across the tasks.

	IQ/numerosity	IQ/recognition	IQ/lexical decision	Numerosity/ recognition	Numerosity/ lexical decision	Recognition/ lexical decision
Boundary separation	0.33	0.12	−0.08	0.42	0.33	0.46
Nondecision component	−0.22	−0.03	−0.08	0.43	0.47	0.56
Drift rate	0.24	0.55	0.53	0.47	0.47	0.63

significant, but much of this was due to a few older, higher IQ subjects who adopted extremely conservative decision criteria to avoid errors (in the range of 0.4–0.6 in boundary separation where the mean was around 0.22).

5.4.5. *Correlations between components of processing across tasks*

For each age group, correlations were computed between drift rate, boundary separation, and the nondecision component for each pair of tasks across subjects. The correlations were then averaged across the age groups. These averages are reported in Table 5.

Perhaps surprisingly, the correlations were positive and significant for all the pairs of tasks, demonstrating that if an individual has a high value for one of the parameters on one task, they likely have a high value on the other tasks. This pattern replicates that obtained by Ratcliff et al. (2006a), but with a considerably wider range of IQ values and almost four times as many subjects.

5.4.6. *Correlations among the different components of processing*

The correlations between drift rate and boundary separation, drift rate and the nondecision component, and the nondecision component and boundary separation are shown in Table 6. The correlations were computed across subjects in each age group and then averaged over the three age groups. Out of the 27 correlations for the separate age groups (three pairs of parameters by three tasks by three age groups), only one was significant. This shows that the model parameters are determined by relatively independent aspects of the data, similar to the findings in the RTM studies.

5.5. *The Effects of Age and IQ on components of processing*

Fig. 6 displays plots of drift rate, boundary separation, and nondecision time as a function of age and IQ for the three tasks. *Y*'s designate college-age subjects, *O*'s 60–74 year olds, and *V*'s 75–90 year olds. In each panel, the thick lines are linear regression lines for the nine data points.

There are dramatic differences in drift rates as a function of IQ for lexical decision and recognition memory. Drift rates decrease by almost a factor of 2 from high to low IQ as IQs change from an average of 125 to an average of 95. For numerosity discrimination, the decrease is much smaller, only about 1/3.

For all three tasks, the boundary separation and nondecision time parameters show at most small changes as a function of IQ. There is enough power to detect decreases if they were there because, as described above, these parameters show consistent changes with age (Table 2).

Fig. 7 shows the same information as Fig. 6 but with all the conditions for all the tasks. The decrease in drift rate with IQ occurs for most of the conditions, which means that the decrease shown in Fig. 6 is not the result of averaging a few conditions with a large decrease and other conditions without a

Table 6
Correlations among model parameters.

Task	a–T_{er}	a–v	T_{er}–v
Numerosity	−0.264	−0.001	−0.100
Recognition	−0.136	0.047	0.042
Lexical	−0.053	−0.124	−0.058

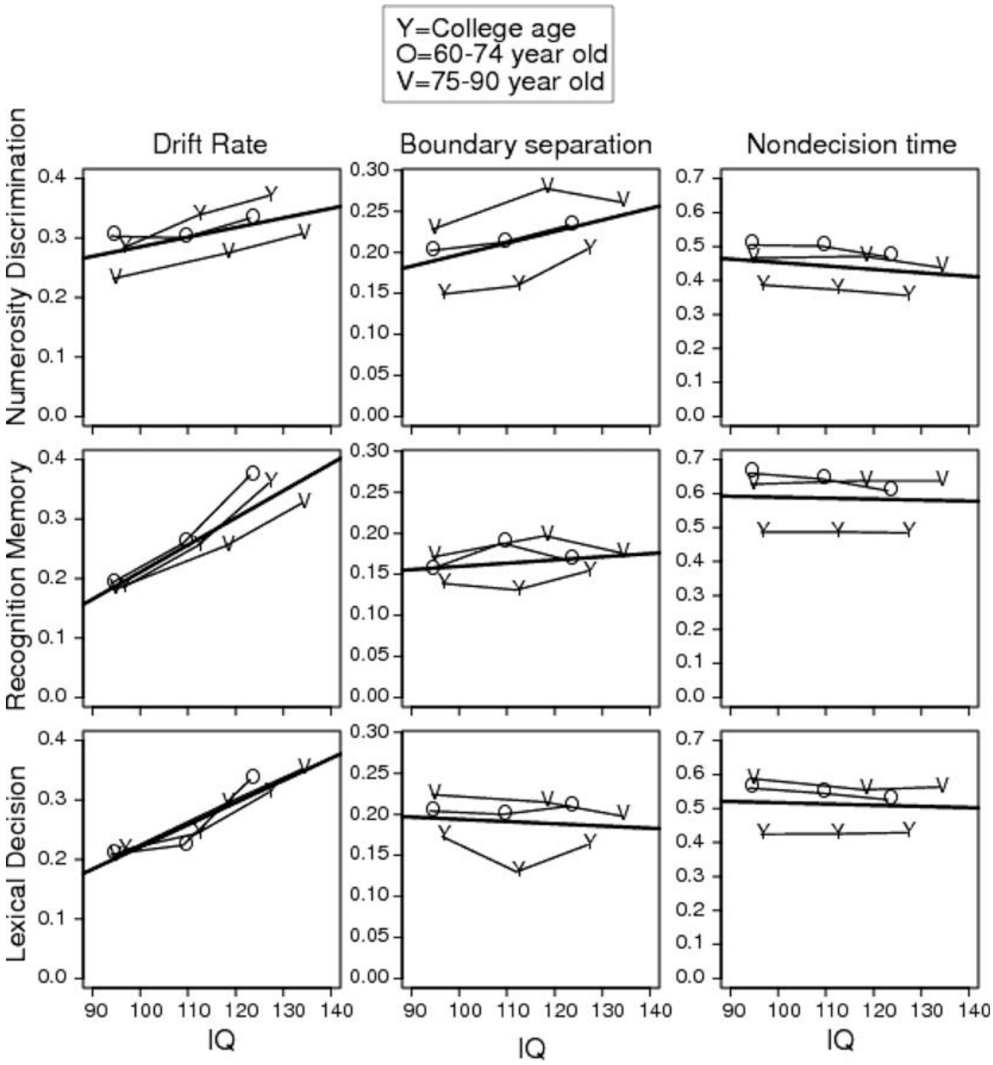

Fig. 6. Plots of average drift rate, boundary separation, and nondecision time as a function of IQ for the three age groups and the three tasks. The subjects were divided into three groups as a function of IQ.

decrease. We also found that the across trial variability parameters (η, s_z, and s_t did not change as a function of IQ. Estimates of these variability parameters are more variable than a, T_{er}, and v, but even so, there were no trends in the values as a function of IQ.

5.6. The joint effects of IQ and age on drift rates

Fig. 8 shows one of the most important results of this research: there is no differential decrease in drift rate as a function of age for low versus high IQ subjects. The left hand side of the figure shows two possibilities. In the top panel, there is a larger decrease in drift rate with age for low IQ subjects than high IQ subjects. In the bottom panel, the decreases are the same. The right hand panels show the results from Fig. 6 re-plotted to show drift rate as a function of age and IQ for the three tasks. The differences in drift rate between high and low IQ subjects are not larger for the 75–90 year olds than for the college students. (Note that there is no problem with floor effects because there is little decrease in drift rate across age groups.)

6. Structural equation modeling

Structural equation modeling fits a regression model simultaneously to the nondecision component, boundary separation and drift rate parameters and two IQ measures to investigate whether a

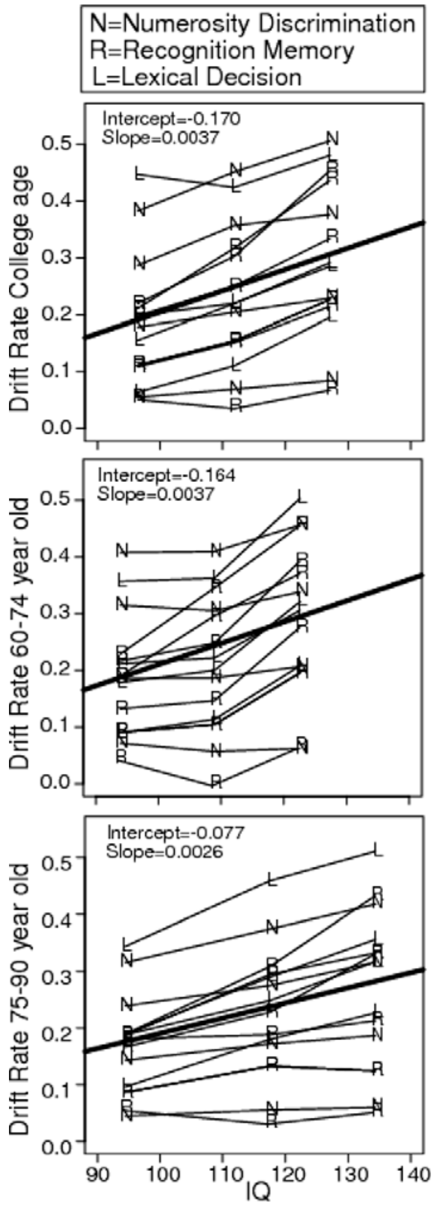

Fig. 7. Plots of the drift rates for all the individual conditions of each task as a function of IQ. Conditions with negative drift rates (e.g., nonwords in lexical decision) had the sign changed to positive. The thick line is the average linear regression line to all the data.

common factor can account for relationships among tasks and how these common factors are associated with the two IQ measures. For example, if task A is correlated with task B and task B with task C, then there may be a common factor underlying the three tasks. Structural equation modeling also explicitly takes into account measurement error.

Schmiedek et al. (2007) applied structural equation modeling to the data from 8 two-choice RT tasks with 80 observations per task for 135 subjects (from Oberauer et al., 2003). They fit the exGaussian distribution (Hohle, 1965; Ratcliff, 1979; Ratcliff & Murdock, 1976) and a simplified version of the diffusion model to the data (using the EZ method, Wagenmakers, Van Der Maas, & Grassman, 2007; but see Ratcliff, 2008b). Oberauer et al. had collected six different measures of working memory along with RT data. Schmiedek et al. applied structural equation modeling to the exGaussian parameters obtained from the RT data and working memory measures and also to the diffusion model parameters from the RT data and working memory measures. For the diffusion model parameters, there was a strong relationship between the working memory measures and drift rate. They also found that standardized regression parameters for the latent diffusion model parameters (i.e., single parameters for

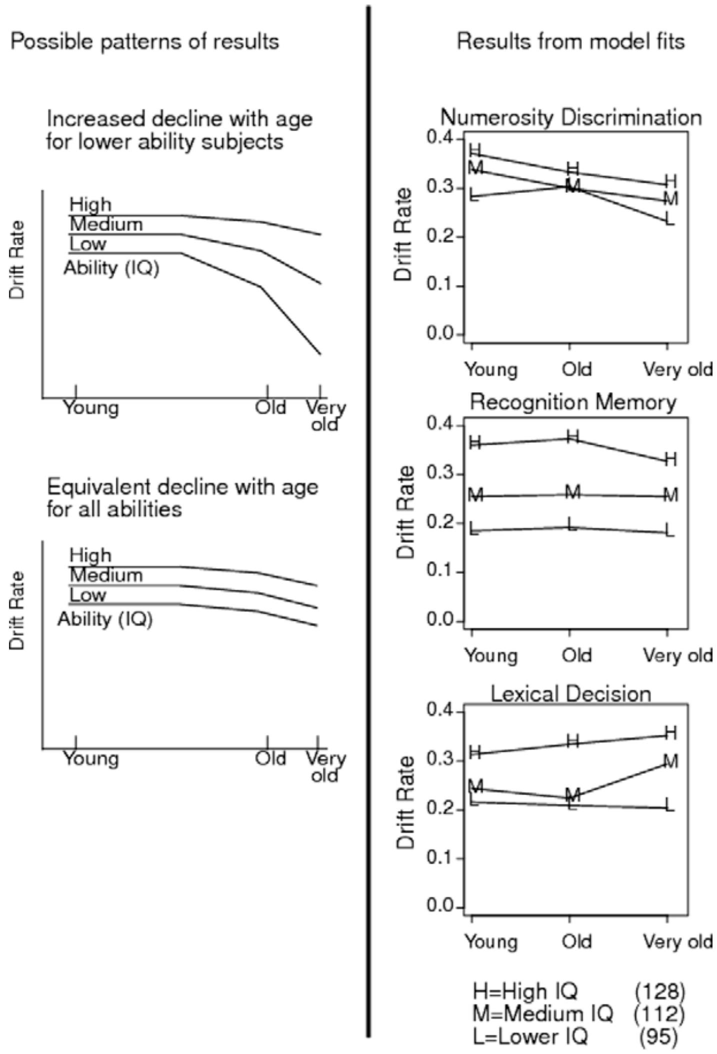

Fig. 8. Plots of drift rate for high, medium, and low IQ groups as a function of age.

nondecision component, boundary separation and drift rate that represent those parameters for all the tasks, c.f., the ellipses in Fig. 9) and the diffusion model parameters for the individual tasks were positive. This means that common diffusion model parameters were able to account for individual differences common to all the tasks.

We applied structural equation modeling (using LISREL 8.8, Joreskog & Sorbom, 1993) to the three main diffusion model parameters for the three tasks and the two components of IQ that we measured (matrix reasoning and vocabulary). This provides an analysis to show whether common factors account for individual differences in model parameters and to what extent the common factors can be identified with the matrix reasoning (fluid intelligence) and vocabulary (crystallized intelligence) components of IQ. Because the model parameters correlate across tasks, we expect that common factors will represent model parameters across tasks in the structural equation models. Due to model identification issues, the IQ variables were included in the model using the pseudo-variable approach described in Bollen (1989, p.173). There were three latent common factors representing boundary separation (a), the nondecision component (T_{er}), and drift rate (v). Fig. 9 shows the model structure. The observed variables are shown in the square boxes at the bottom (with the letters N, R, and L representing the numerosity, recognition, and lexical decision tasks). For each link and for the residuals, there are three numbers; these are the values for college age, 60–74 year old, and 75–90 year old subjects, respectively.

The sample sizes used here are on the extreme lower end of acceptability (e.g., Schumaker & Lomax, 1996), however, it is likely that the clear dissociations between the model parameters and IQ allow the model to fit acceptably. The fits of the model for the three subject groups were

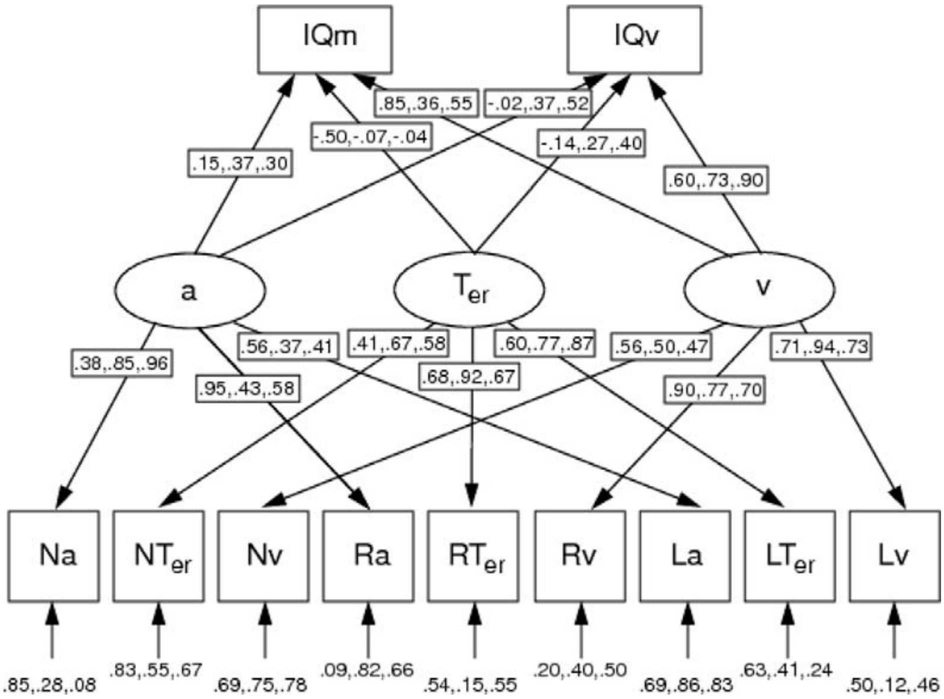

Fig. 9. A graphical presentation of a structural equation model for diffusion model parameters and two IQ components. Numbers in squares are the normalized regression coefficients (Gamma coefficients for the numbers just below the IQ measure boxes and Lambda-*x* coefficients for the numbers below the latent diffusion model parameters). The three numbers represent in order: college age, 60–74 year old, and 75–90 year old subjects. The numbers at the bottom are residuals. In the boxes at the bottom, *N* refers to numerosity discrimination, *R* to recognition memory, and *L* to lexical decision, and *a* = boundary separation, T_{er} = nondecision component, *v* = drift rate diffusion model parameters.

χ^2 = 60.2 (RMSEA = 0.12, CFI = 0.76, SRMR = 0.15), 49.5 (RMSEA = 0.09, CFI = 0.91, SRMR = 0.10), and 40.1 (RMSEA = 0.05, CFI = 0.89, SRMR = 0.11) for the college age, 60–74 year old, and 75–90 year old subjects, respectively, with 37 degrees of freedom. The chi-square value for young subjects is significant, probably because the range of values across subjects is smaller, but the chi-square values for the other two groups are not significant. Evaluation of goodness-of-fit for structural equation modeling is not as straightforward as other methods. The LISREL program produces around 20 goodness-of-fit indices (several closely related) and most of the measures have recommended ranges rather than statistical tests for significance. Apart from the fit for young subjects, the CFI value is around the recommended value for a good fit, 0.9, (but this is qualified of course by the low sample size).

As reflected in the correlation analyses, the standardized regression coefficients or factor loadings between each latent diffusion model parameter and the corresponding parameter for each task are positive and vary from 0.37 to 0.96 with a mean of 0.66 and all are significant at the 0.05 level. The coefficients for the numerosity discrimination drift rates are lower than for the other two tasks which, along with the lower correlations between numerosity drift rate and both IQ and drift rates for the other tasks, suggests that numerosity discrimination performance is not as strongly related to IQ and the other cognitive tasks. The results show that common model parameters can account for much of the variability across subjects in the separate tasks. This provides the same interpretation that is suggested by the simple correlations shown in Table 6 but fits the whole data set rather than just pairwise correlations.

The standardized gamma coefficients between the latent diffusion model parameters and the two IQ measures are consistently above zero for the six values of drift rate, that is, for the two IQ measures by three subject groups, and all are significant at the 0.05 level. However, the standardized gamma coefficients for boundary separation and the nondecision component and the two IQ measures are not consistently of the same sign; only six of the twelve are significant and one has the opposite sign from the others. This result again supports the interpretation from Table 6 that drift rate and not the other parameters correlates reliably with overall IQ.

We examined two other structural equation models. First, we ran a model with all the subjects combined and added age as a factor at the same level as the two IQ measures. This resulted in a model that loaded drift rate onto IQ (standardized gamma coefficients of 0.46 and 0.66 for matrix and vocabulary, respectively) and that loaded boundary separation and the nondecision component onto age (standardized gamma coefficients of 0.42 and 0.61, respectively). The fit of the model was not as good in chi-square terms as the analyses above, $\chi^2 = 105.2$ (RMSEA = 0.10, CFI = 0.92, SRMR = 0.08), with 45 degrees of freedom. But the CFI value was acceptable. In the second analysis, we added age as a factor at the same level as the IQ measures to produce three separate analyses as in Fig. 9. Drift rate loaded onto IQ as in Fig. 9, but nothing loaded onto age (standardized gamma coefficients were in the range −0.27 to 0.26 across the three analyses and goodness-of-fit and other model parameters were within a few percent of those shown in Fig. 9). Lindenberger and Baltes (1997) found correlations in the range of −0.4 to −0.6 of age with five different abilities (perceptual speed, reasoning, memory, knowledge, and fluency). Their study had individuals that ranged from age 70 to age over 100 and had over 500 subjects in the sample. With our smaller samples, we did not have enough power to obtain such results.

These analyses show that there were no differences in model parameters as a function of age within each group and that age effects were only obtained for the large difference between the groups. The IQ and common factor results for both of these analyses were similar to those in Fig. 9.

7. The worst performance rule

A puzzling finding in IQ research has been that IQ is sometimes more highly correlated with slow responses than fast ones, even though slow responses have more variability than fast responses. Evidence for this finding, which occurs in many, but not all studies, has been summarized by Coyle (2003), also Baumeister and Kellas (1968), Diascro and Brody (1993), Jensen (1982), Kranzler (1992), Larson and Alderton (1990); but see Salthouse (1998).

With simulations, Ratcliff, Schmiedek, and McKoon (2008; also Schmiedek et al., 2007) demonstrated the conditions under which the diffusion model predicts the worst performance rule. They assumed (given earlier data as well as the data described above) that IQ corresponds to drift rate. Accuracy and RT data were simulated by generating model predictions from parameter values that are typical of those found empirically. RT quantile and accuracy predictions were generated with 200 observations for each experimental condition and correlations were calculated between the five RT quantiles (0.1, 0.3, 0.5, 0.7, and 0.9) and drift rate, boundary separation, and the nondecision component. These were then averaged over experimental conditions in each experiment.

First, consider the drift rate correlations. When there is no across trial variability in any of the components of processing, and no across subject variability in boundary separation or the nondecision component, the model predicts a shallow U-shaped function. The correlation is negative for the 0.1 and 0.3 quantiles, more negative for the 0.5 and 0.7 quantiles, and less negative for the 0.9 quantile. The model predicts the U-shaped function because of a complicated interaction between differences in the RT quantiles as a function of drift rate and variability in the RT quantiles (see Ratcliff et al., 2008).

The assumptions that there is no across trial variability and that there is no across subject variability in boundary separation or the nondecision component are not realistic. When across trial variability in the model parameters is added, again using values typical of fits to earlier data (values of η, s_z, and s_t), there is little change in the pattern of correlations. However, when across subject variability in boundary separation and the nondecision component are added, again using values typical of earlier studies, the pattern of correlations between drift rate and the RT quantiles changes from the U-shaped function to the worst performance rule. The correlation between drift rate (IQ) and quantiles increases from the 0.1 to the 0.9 quantile.

This change from a shallow U-shaped function to an almost linear function is straightforward to explain. Suppose that there are across subject differences only in the nondecision component. Differences in RT quantiles as a function of drift rate are smaller for the faster quantiles than the slower quantiles. If the nondecision component varies randomly across subjects, then all of the quantile RTs are perturbed by the same amount. As a result, the small differences in the 0.1 quantiles as a function of drift rate become obscured more than the higher quantiles. Thus the correlation between faster responses (lower quantile RTs) and drift rate is reduced relative to the correlation between slower responses (higher quantiles). This same result also holds when there is variability across subjects in boundary separation.

For the correlation between boundary separation and quantiles, the pattern is similar to the pattern for drift rates but with the opposite sign. Without across subject variability in the nondecision component, the function decreases slightly from a high value of correlation to slightly lower value. When across subject variability in the nondecision component and drift rate are added, the correlations between RT quantiles and boundary separation increase from the lower to the higher quantiles.

The pattern for the nondecision component is different. The correlations between the RT quantiles and the nondecision component are positive, but they decrease from the lower to the higher quantiles (Ratcliff et al., 2008). Differences in T_{er} across subjects induce shifts in the whole RT distribution. Because the 0.1 quantile RTs are less variable than the 0.9 quantile RTs, differences in the 0.1 quantiles are more reliable than differences in the 0.9 quantiles and so the correlations decrease over quantiles.

In sum, the model can explain both U-shaped and linear functions and it can do this for drift rate, boundary separation, and the nondecision component. Whether the functions for real data are more U-shaped or more like the worst performance rule depends on how much variability there is across subjects in boundary separation and the nondecision component relative to variability in drift rate. This can be answered by generating predictions using the model parameters derived from fits to data.

7.1. Experimental data

Given that the model fits the data well, then the best-fitting parameter values can be used to generate quantile RTs for individual subjects which then can be used to produce predicted shapes for the functions relating correlations to RT quantiles. Fig. 10 shows the worst performance rule functions from the data and Fig. 11 shows the model's predictions for them.

Fig. 10 shows the functions for the three tasks and the three age groups. Correlations were computed for each condition and then averaged over the conditions. Just as for Fig. 2, only the conditions for which performance was above chance were included (six conditions for numerosity discrimination and four each for recognition memory and lexical decision).

For all the tasks and all the age groups, the correlations with boundary separation were positive and increasing with RT quantiles, and the correlations with the nondecision component were positive and decreasing with RT quantiles.

In the data (Table 5), IQ and drift rate were positively correlated for lexical decision and recognition memory. Consequently, their correlation-RT quantile functions should track each other, and Fig. 10 shows that they did. For the numerosity task, IQ and drift rate were not strongly correlated and Fig. 10 reflects this: the correlation-RT quantile functions for IQ and drift rate did not always show the same shapes and the correlations of IQ with RT quantiles were close to zero.

For neither drift rate nor IQ did the data for recognition and lexical decision unequivocally show the worst performance rule. Instead, the correlations with RT quantiles were negative and either they decreased with increasing quantiles or they were shallowly U-shaped (although for IQ, many of the correlations were close to zero).

Fig. 11 shows the same plots as Fig. 10 but for the values predicted by the model from the parameters that best fit the data. The model predictions match the data closely, which means that the failure to find strong or consistent evidence for the worst performance rule for IQ is a property of both the model (based on parameters from fits to data) and the data.

The results from this study show that the worst performance rule is by no means universal. The results that most nearly conformed to the worst performance rule were obtained in the recognition memory and lexical decision tasks where IQ effects were large, RTs were long, and accuracy was relatively high (especially for 60–74 and 75–90 year olds). Despite this, there was no significant consistent evidence to support the rule.

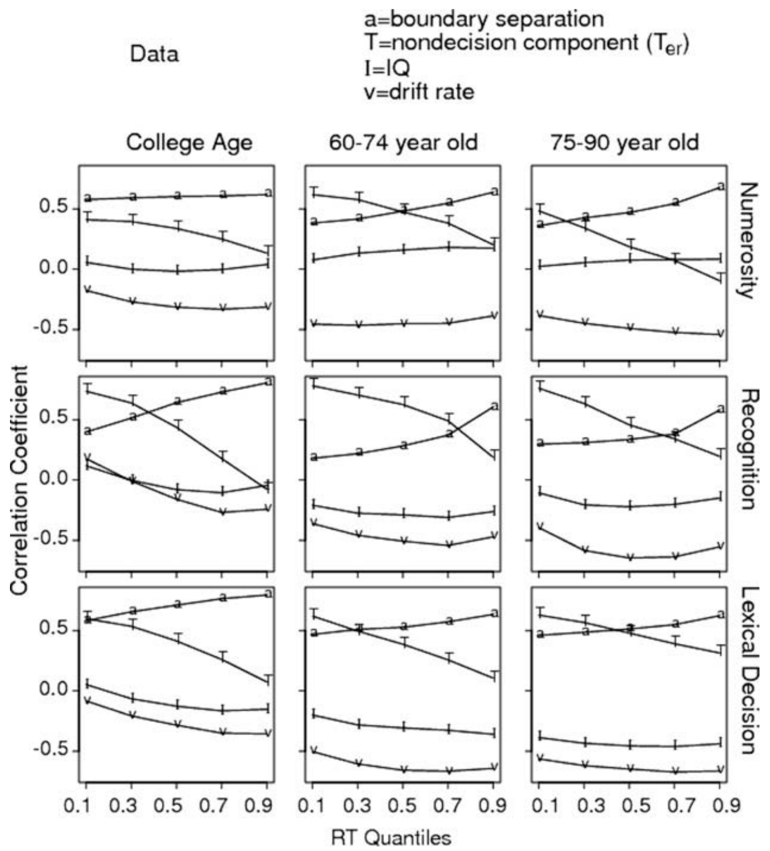

Fig. 10. Correlations between the empirical quantile RTs and diffusion model parameters and empirical quantile RTs and IQ for the three subject groups and three tasks. a = boundary separation, T = nondecision component, v = drift rate, and I = IQ.

8. General discussion

In this study, we examined the effects of age and IQ on performance in three two-choice tasks. Numerosity discrimination, which requires little perceptual or memory load, was used as a baseline. The other two tasks were lexical decision and recognition memory. In previous studies (RTM), the range of IQs was restricted because older subjects' IQs were matched to college-age subjects'. Here, IQ ranged from 83 to 146. Also, the numbers of subjects in the age groups (45 college students, forty-three 60–74 year olds, and forty-two 75–90 year olds) were reasonably large. In the next paragraphs, we first review the findings in this study and then relate the results to the wider literature.

The data were interpreted in terms of Ratcliff's diffusion model. The model explained all of the data: accuracy, correct and error RT quantiles (i.e., RT distributions), and the relative speeds of correct and error responses, for each task, experimental condition, age group, and level of IQ individually for the 130 subjects. For all three tasks, there were several levels of difficulty (e.g., number of asterisks for numerosity discrimination, word frequency for recognition memory and lexical decision). As would be expected, increases in difficulty led to decreases in accuracy and increases in RTs. In the model, drift rate represents the quality of the evidence on which a decision is based and so drift rate should vary with difficulty. Across levels of difficulty, the model fit the data, i.e., accuracy and RT distributions, well, with only the single parameter, drift rate, varying. There were only a few systematic deviations between theory and data (Figs. 3–5).

We note that the diffusion model explained the data well with only a single session of data per task (45 min of performance). This is impressive because it is rare for computational models in cognitive psychology to provide reliable estimates of individual differences from only a single session on a single task.

As a function of age, RTs slowed considerably but the decline in accuracy was minimal for all three tasks. The model explains the patterns of data for age as follows: older subjects slowed because they

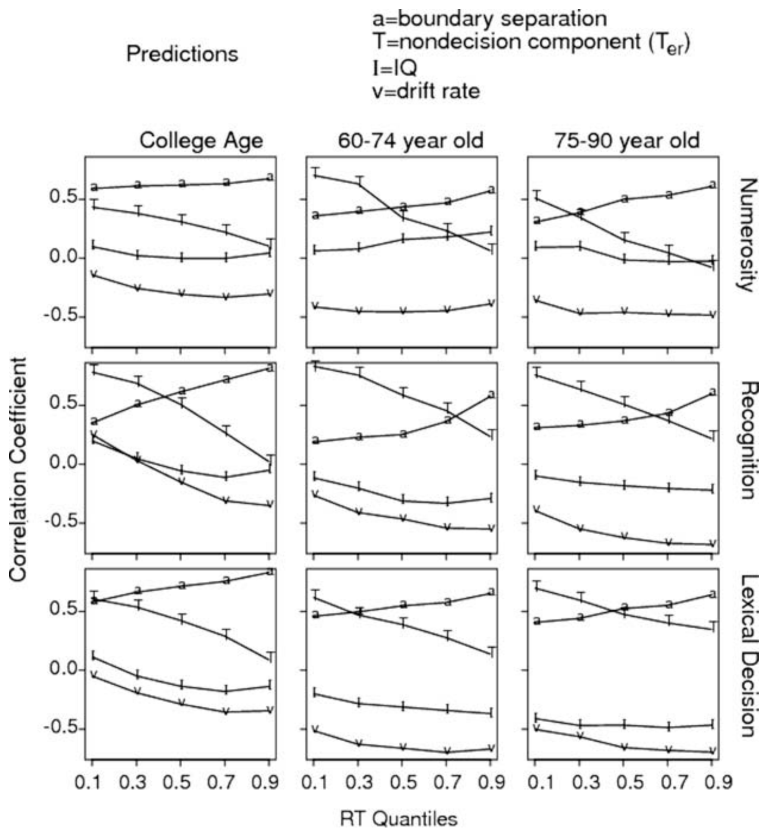

Fig. 11. Correlations between the predicted quantile RTs and diffusion model parameters and predicted quantile RTs and IQ for the three subject groups and three tasks. *a* = boundary separation, *T* = nondecision component, *v* = drift rate, and *I* = IQ.

set wider boundaries and their nondecision component was longer. Accuracy declined only minimally because drift rate declined only minimally. The finding that drift rate does not noticeably decline with age is notable and replicates previous findings (RTM). In particular, the results for the 75–90 year old group replicated results from the Ratcliff et al. (2007) experiments for this age group.

For lower compared to higher IQ subjects, RTs were generally longer and accuracy was lower for the lexical decision and recognition memory tasks, but less so for the baseline, numerosity judgment task. In the diffusion model analyses, lower IQ subjects were slower and less accurate than higher IQ subjects because the quality of evidence on which their decisions were based, drift rate, was poorer. The magnitude of the difference in drift rates between lower and higher IQ subjects was large, especially for lexical decision and recognition. In contrast, there was no significant effect of IQ on boundary separation or the nondecision component.

It might be thought that components of processing would show more variability across trials for older than younger subjects, or lower than higher IQ subjects. For instance, older subjects might have been more variable in drift rates or starting point from trial to trial. However, this was not the case. None of the components of processing varied across trials more for the older groups than the young group or more for the lower than the higher IQ groups. The only two significant variability effects were first, variability in drift rate across trials decreased (not increased) with age for numerosity discrimination, and second, the range of starting points across trials was larger, not smaller, for the college-age subjects compared to the 60–74 year olds and 75–90 year olds for recognition memory and lexical decision.

A central claim of the diffusion model is that the three main components of processing tap different aspects of decision making. This suggests that, within a task, it is possible for drift rate, boundary separation, and the nondecision component not to correlate with each other (in the absence of other factors that might affect more than one component), and this is the result we obtained. The result means that one of the components cannot be predicted from another. For example, a subject's boundary settings do not predict the duration of the nondecision component or the quality of the evidence that enters the decision process.

In the model, drift rates are the main determinant of accuracy. Accordingly, we found that, within each task, drift rate and accuracy were positively correlated. Likewise, boundary separation and the nondecision component were positively correlated with RT. Furthermore, for each of the three components, the values were correlated across tasks. For example, if a subject had a high drift rate on one task, then she or he also had a high drift rate on the other tasks. Structural equation modeling supported this pattern of relationships across tasks. Common factors among the model parameters (i.e., a single value of boundary separation, a single value for the nondecision component, and a single value for drift rate for a subject across all three tasks) account for a large proportion of the individual differences within each age group. The results also show that the common drift rate factor explains a large proportion of the variance across subjects in the two IQ measures, matrix reasoning and vocabulary. These results are consistent with the results of Schmiedek et al. (2007) who showed that common factors from several working memory tasks were related to drift rates in eight choice RT tasks.

The worst performance rule (Coyle, 2003) is the finding that IQ is more strongly correlated with slow than fast responses. We obtained this result for lexical decision and recognition memory, but not the numerosity task, and only for the two older groups. This suggests that the worst performance rule is less universal than might be deduced from previous literature (see Salthouse, 1998).

Interactions between IQ and speed of processing have a long history of inquiry. For example, Woodworth (1938) argued that the correlation between IQ and the speed of performing a task is low (varying from 0 to 0.35). The majority of studies has found correlations in the same range as Woodworth, low to moderately negative between IQ and RT (e.g., Deary, Der, & Ford, 2001; Detterman, 1987; Sheppard & Vernon, 2008). Our results also show moderately weak relationships between IQ and RTs.

With the diffusion model analysis, we can explain the weak relationships. Within a task, drift rate varies across levels of difficulty and, as difficulty increases, responses become slower. Thus, averaging over subjects, RT varies with drift rate as in Fig. 3. We would then expect that, because drift rate correlates with IQ, IQ and speed would be correlated at least moderately strongly. However, this is a correlation across subjects between the average drift rate for a subject and the IQ of the subject. The correlation is low because boundary separation and the duration of the nondecision component vary across subjects, which produces variability in RTs that is unrelated to drift rate (see also Ratcliff et al., 2008 for a similar explanation of the worst performance rule).

There have been two different approaches to measuring speed of processing in the aging literature. One is to examine performance on one or a small number of fairly standard two-choice tasks (e.g., Cerella, 1994; Fisk & Fisher, 1994; Myerson, Wagstaff, & Hale, 1994; Myerson et al., 1992; Perfect, 1994). The other is to use several different kinds of tasks in order to average out the idiosyncrasies that might be tied to any one task. The measures of speed from each task are then taken together as representative of an overall measure of speed (e.g., Ball et al., 2002; Bowles & Salthouse, 2008; Lindenberger & Baltes, 1997; Little et al., 1999; Oberauer et al., 2003; Salthouse, 1996; Tucker-Drob & Salthouse, 2008).

For the first approach to speed of processing, the RTM experiments and the experiment in this article examine the components of processing that are affected by age in tasks that require fast, two-choice decisions. Speed is measured in the same way for all of the experiments, specifically, the RT for a decision. The main result has been that the slow responses of older subjects are due to the more conservative decision criteria that they adopt and their longer nondecision times.

In the second approach, when researchers use a variety of tasks to measure speed, the standard two-choice task might be one of them but others can be quite different. For example, in the digit-symbol task, a list of mappings between symbols and digits is displayed on a PC monitor throughout a series of tests. For each test, another digit-symbol pair is displayed and subjects respond according to whether it is the same or different than one of the pairs in the list. In this task, and others like it, speed is measured by the number of test items a subject can do in a certain amount of time (e.g., 40 s). Another, quite different, measure is the minimum stimulus duration at which subjects can attain 75% correct performance under various levels of cognitive demand, with demand implemented by adding distractor stimuli or a concurrent task (e.g., Ball et al., 2002; Salthouse, 1996).

The digit-symbol and related tasks are likely amenable to theoretical analyses by models such as the diffusion model because accuracy and RT can be measured for "same" and "different" stimuli, as in other two-choice tasks. In contrast, the stimulus duration that provides enough evidence to produce, say, 75% accuracy, is a function of drift rate, much more than boundary separation. This is logically separate from decision time in two-choice tasks, which is a function of the other two components of processing, boundary separation and the nondecision component, as well as drift rate (see Ratcliff et al., 2003; Thapar et al., 2003). In future research, it will be important to determine whether these different measures are strongly correlated across subjects when analyzed in a modeling framework.

Often in past studies, variability in RTs across subjects has been greater for older individuals than younger individuals (Hale, Myerson, Smith, & Poon, 1988; Hultsch, MacDonald, & Dixon, 2002; Myerson & Hale, 1993; Myerson, Robertson, & Hale, 2007; Robertson, Myerson, & Hale, 2006; Williams, Hultsch, Strauss, Hunter, & Tannock, 2005). For example, in the numerosity discrimination task described here, the SD's across subjects in mean correct RT are 145 ms, 264 ms, and 319 ms for college age, 60–74 year olds, and 75–90 year olds, respectively. The diffusion model explains this in the following way (Ratcliff, Spieler, & McKoon, 2000, Fig. 6): in general, boundary separation is smaller and less variable for younger than older subjects. The larger separation for older subjects leads to a larger range of mean RTs across subjects. Although it has been argued that the increased variability for older subjects' RTs is a statistical artifact (because SDs increase with mean RTs), the diffusion model provides an explanation for the finding. Of course, whether the explanation is correct for a particular data set requires fitting the model to the data and evaluating its goodness-of-fit.

The results in the RTM experiments and the experiment reported here show that older subjects adopt more conservative decision criteria than young subjects. Starns and Ratcliff (in press) investigated the aspects of performance that the younger compared to older subjects might be trying to optimize with their decision criteria settings. Starns and Ratcliff defined optimality in terms of "reward rate" (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006), where reward rate (the term comes from the animal literature) is the number of responses that are correct per unit time. It is clear from the earlier RTM studies that subjects adjust their boundary settings to conform to speed instructions or accuracy instructions; they can trade accuracy for speed or speed for accuracy. Starns and Ratcliff found that young subjects often set their boundaries at values that give close to optimal performance, aiming for the most correct responses per unit time. In contrast, older subjects set their boundaries to obtain close to the maximum accuracy that could be obtained if the boundaries were set very wide apart. The actual boundary values were set at a point that made accuracy within a percent or two of this maximum. In other words, they do not set their boundaries to maximize the number correct per unit time, rather, they set them wide enough so as not to make errors that they could avoid (if they adopted narrower settings). This provides a further part of the explanation for slowing with age.

Our results speak to the cognitive reserve hypothesis. This is the hypothesis that factors such as educational attainment, intellectual ability, and/or socioeconomic status play a protective role to mitigate the effects of aging on cognitive function (Satz, 1993; Stern, 2002; Stern, Albert, Tang, & Tsai, 1999). A number of studies have demonstrated that these factors are associated with slower rates of cognitive decline in normal aging (e.g., Albert et al., 1995; Butler, Ashford, & Snowdon, 1996; Evans, Beckett, Albert, et al., 1993). However, we did not find this trend. In the diffusion model, variations in cognitive reserve would be implemented as variations in drift rate. Subjects with more cognitive reserve (higher IQ in our study) should have larger values of drift rate; our results show that they do. But the differences in drift rates between higher and lower IQ subjects were not affected by age. Drift rates declined little with age, but about as much for the higher IQ subjects as for the lower IQ subjects. This lack of differential decline in drift rates is consistent with work by Lindenberger and Baltes (1997), Rabbitt, Chetwynd, and McInnes (2003), and Singer, Verhaeghen, Ghisletta, Lindenberger, and Baltes (2003) who found little differential decline as a function of ability (but see Deary, MacLeannan, & Starr, 1998, who did find evidence for differential age-related declines as a function of ability). Our results are also consistent with research investigating the effects of education on age-related cognitive decline where decreases in cognitive abilities are as large for highly educated subjects as low educated subjects (Christensen et al., 2007; Tucker-Drob, Johnson, & Jones, 2009; Van Dijk, Van Gerven, Van Boxtel, Van der Elst, & Jolles, 2008; Van Gerven, Meijer, & Jolles, 2007).

In sum, the diffusion model analysis provides a somewhat different view than some earlier views of the effects of age and IQ on speed of processing. It separates components of processing from each other and it specifies dissociations between the components, age, and IQ for the three tasks in this experiment. IQ affected drift rate whereas age affected boundary separation and the nondecision component. In other work (Ratcliff et al. in press), we have applied the diffusion model to understand age and IQ effects on associative recognition and in item recognition (the task used here) and compared the analyses to the effects of age and IQ on cued and free recall performance. IQ affects both recognition tasks in the same way as here. As this modeling approach begins to address the wider domain that explores general abilities, we will begin to see alternative interpretations begin to emerge as well as confirmation of existing views.

Acknowledgments

Preparation of this article was supported by NIA grant R01-AG17083 and NIMH grant R37-MH44640. We would like to thank Mike Edwards for a great deal of help on the structural equation modeling.

References

Albert, M. S., Jones, K., Savage, C. R., Berkman, L., Seeman, T., Blazer, D., et al (1995). Predictors of cognitive change in older persons: MacArthur studies of successful aging. *Psychology and Aging, 10*, 578–589.

Allen, P. A., Madden, D. J., & Crozier, L. C. (1991). Adult age differences in word-level and letter-level processors during word recognition. *Psychology and Aging, 6*, 261–271.

Ball, K., Birch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, et al (2002). Effects of cognitive training interventions with older adults. *The Journal of the American Medical Association, 288*, 2271–2281.

Balota, D. A., Dolan, P. O., & Duchek, J. M. (2000). Memory changes in healthy young and older adults. In E. Tulving & F. I. M. Craik (Eds.), *Handbook of memory* (pp. 395–410). Oxford University Press.

Baumeister, A. A., & Kellas, G. (1968). Distributions of reaction times of retardates and normals. *American Journal of Mental Deficiency, 72*, 715–718.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced choice tasks. *Psychological Review, 113*, 700–765.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

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

Bowles, R. P., & Salthouse, T. A. (2008). Vocabulary test format and differential relations to age. *Psychology and Aging, 23*, 366–376.

Butler, S. A., Ashford, J. W., & Snowdon, D. A. (1996). Age, education and changes in the Mini-Mental State Exam scores of older women: Findings from the Nun Study. *Journal of the American Geriatrics Society, 44*, 675–681.

Cerella, J. (1994). Generalized slowing in Brinley plots. *Journal of Gerontology: Psychological Sciences, 49*, 65–71.

Christensen, H., Anstey, K. J., Parslow, R. A., Maller, J., Mackinnon, A., & Sachdev, P. (2007). The brain reserve hypothesis, brain atrophy and aging. *Gerontology, 53*, 82–95.

Coyle, T. R. (2003). IQ, the worst performance rule, and Spearman's law: A reanalysis and extension. *Intelligence, 31*, 567–587.

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

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, N.J.: Lawrence Erlbaum Associates.

Deary, I. J., Der, G., & Ford, G. (2001). Reaction times and intelligence differences: A population- based cohort study. *Intelligence, 29*, 389–399.

Deary, I. J., MacLeannan, W. J., & Starr, J. M. (1998). Is age kinder to the initially more able?: Differential ageing of a verbal ability in the Healthy Old People in Edinburgh Study. *Intelligence, 26*, 357–375.

Detterman, D. K. (1987). What does reaction time tell us about intelligence? In P. A. Vernon (Ed.), *Speed of information processing and intelligence* (pp. 177–199). Norwood, NJ: Ablex.

Diascro, M. N., & Brody, N. (1993). Serial versus parallel processing in visual search tasks and IQ. *Personality and Individual Differences, 14*, 243–245.

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

Evans, D. A., Beckett, L. A., Albert, M. S., et al (1993). Level of education and change in cognitive function in a community population of older persons. *Annals of Epidemiology, 3*, 71–77.

Fisk, A. D., & Fisher, D. L. (1994). Brinley plots and theories of aging: The explicit, muddled, and implicit debates. *Journal of Gerontology: Psychological Sciences, 49*, 81–89.

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

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, US: Academic Press, Inc..

Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the go/no-go task. *Journal of Experimental Psychology: General, 136*, 389–413.

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.

Hale, S., Myerson, J., Smith, G. A., & Poon, L. W. (1988). Age, variability, and speed: Between- subjects diversity. *Psychology and Aging, 3*, 407–410.

Hohle, R. H. (1965). Inferred components of reaction times as a function of fore period duration. *Journal of Experimental Psychology, 69*, 382–386.

Hultsch, D. F., MacDonald, S. W. S., & Dixon, R. A. (2002). Variability in reaction time performance of younger and older adults. *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 57*, 101–115.

Jensen, A. R. (1982). Reaction time and psychometric g. In H. J. Eysenck (Ed.), *A model for intelligence* (pp. 93–132). New York: Springer.

Jensen, A. R. (1987). Individual differences in the Hick paradigm. In P. A. Vernon (Ed.), *Speed of information processing and intelligence* (pp. 101–175). Norwood, NJ: Ablex.

Joreskog, K. G., & Sorbom, D. (1993). *Lisrel 8: Structural equation modeling with the SIMPLIS command language*. Hillsdale, NJ: Erlbaum.

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

Kranzler, J. H. (1992). A test of Larson and Alderton's (1990) worst performance rule of reaction time variability. *Personality and Individual Differences, 13*, 255–261.

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

Laming, D. R. J. (1968). *Information theory of choice reaction time*. New York: Wiley.

Larson, G. E., & Alderton, D. L. (1990). Reaction time variability and intelligence: A "worst performance" analysis of individual differences. *Intelligence, 14*, 309–325.

Lindenberger, U., & Baltes, P. B. (1997). Intellectual functioning in old and very old age: Cross-sectional results from the Berlin Aging Study. *Psychology & Aging, 12*, 410–432.

Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods, 4*, 192–211.

Merriam-Webster (1990). *Merriam-Webster's ninth new collegiate dictionary* (9th ed.). Springfield, MA: Author.

Myerson, J., Ferraro, F. R., Hale, S., & Lima, S. D. (1992). General slowing in semantic priming and word recognition. *Psychology and Aging, 7*, 257–270.

Myerson, J., & Hale, S. (1993). General slowing and age invariance in cognitive processing: The other side of the coin. In J. Cerella, J. Rybash, W. Hoyer, & M. L. Commons (Eds.), *Adult information processing: Limits on loss* (pp. 115–141). San Diego, CA: Academic Press.

Myerson, J., Robertson, S., & Hale, S. (2007). Aging and intraindividual variability in performance: Analysis of response time distributions. *Journal of Experimental Analysis of Behavior, 88*, 319–337.

Myerson, J., Wagstaff, D., & Hale, S. (1994). Brinley plots, explained variance, and the analysis of age differences in response latencies. *Journal of Gerontology: Psychological Sciences, 49*, 72–80.

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). *Human memory: An introduction to research, data, and theory*. Pacific Grove, CA: Brooks/Cole.

Oberauer, K., Suβ, H.-M., Wilhelm, O., & Wittmann, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision, and coordination. *Intelligence, 31*, 167–193.

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of aging on item and associative measures of memory: A meta-analysis. *Psychology and Aging, 23*, 104–118.

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? *Journal of Gerontology: Psychological Sciences, 49*, 60–64.

Rabbitt, P., Chetwynd, A., & McInnes, L. (2003). Do clever brains age more slowly? Further exploration of a nun result. *British Journal of Psychology, 94*, 63–71.

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

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

Ratcliff, R. (1985). Theoretical interpretations of speed and@osu.edu 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. (2006). Modeling response signal and response time data. *Cognitive Psychology, 53*, 195–237.

Ratcliff, R. (2008a). Modeling aging effects on two-choice tasks: Response signal and response time data. *Psychology and Aging, 23*, 900–916.

Ratcliff, R. (2008b). The EZ diffusion method: Too EZ? *Psychonomic Bulletin and Review, 15*, 1218–1228.

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

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation, 20*, 873–922.

Ratcliff, R., & Murdock, B. B. Jr., (1976). Retrieval processes in recognition memory. *Psychological Review, 83*, 190–214.

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 letter identification. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 127–140.

Ratcliff, R., Schmiedek, F., & McKoon, G. (2008). A diffusion model explanation of the worst performance rule for reaction time and IQ. *Intelligence, 36*, 10–17.

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., 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. (2006a). Aging and individual differences in rapid two-choice decisions. *Psychonomic Bulletin and Review, 13*, 626–635.

Ratcliff, R., Thapar, A., & McKoon, G. (2006b). Aging, practice, and perceptual tasks: A diffusion model analysis. *Psychology and Aging, 21*, 353–371.

Ratcliff, R., Thapar, A., & McKoon, G. (2007). Application of the diffusion model to two-choice tasks for adults 75–90 years old. *Psychology and Aging, 22*, 56–66.

Ratcliff, R., Thapar, A., & McKoon, G. (in preparation). Individual differences, aging, and IQ in four memory tasks.

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

Ratcliff, R., & VanDongen, H. P. A. (2009). Sleep deprivation affects multiple distinct cognitive processes. *Psychonomic Bulletin and Review, 16*, 742–751.

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

Roberts, R. D., & Stankov, L. (1999). Individual differences in speed of mental processing and human cognitive abilities: Toward a taxonomic model. *Learning and Individual Differences, 11*, 1–120.

Rosen, W. G., Mohs, R. C., & Davis, K. L. (1984). A new rating scale for Alzheimer's disease. *American Journal of Psychiatry, 141*, 1356–1364.

Robertson, S., Myerson, J., & Hale, S. (2006). Are there age differences in intraindividual variability in working memory performance? *Journals of Gerontology: Series B: Psychological Sciences and Social Sciences, 61*, 18–24.

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

Salthouse, T. A. (1998). Relations of successive percentiles of reaction time distributions to cognitive variables and adult age. *Intelligence, 26*, 153–166.

Salthouse, T. A. (2006). Mental exercise and mental aging. *Perspectives on Psychological Science, 1*, 68–87.

Satz, P. (1993). Brain reserve capacity on symptom onset after brain injury: A formulation and review of evidence for threshold theory. *Neuropsychology, 7*, 273–295.

Schmiedek, F., Oberauer, K., Wilhelm, O., Suβ, H.-M., & Wittmann, W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General, 136*, 414–429.

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

Schumaker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Erlbaum.

Sheppard, L. D., & Vernon, P. A. (2008). Intelligence and speed of information-processing: A review of 50 years of research. *Personality and Individual Differences, 44*, 535–551.

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

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

Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review, 116*, 283–317.

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

Spaniol, J., Madden, D. J., & Voss, A. (2006). A diffusion model analysis of adult age differences in episodic and semantic long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 101–117.

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

Starns, J. J. & Ratcliff, R. (in press). The effects of aging on the speed–accuracy compromise: Boundary optimality in the diffusion model. *Psychology and Aging*.

Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society, 8*, 448–460.

Stern, Y., Albert, S., Tang, M. X., & Tsai, W. Y. (1999). Rate of memory decline in AD is related to education and occupation: Cognitive reserve? *Neurology, 53*, 1942–1947.

Tucker-Drob, E. M., Johnson, K. E., & Jones, R. N. (2009). The cognitive reserve hypothesis: A longitudinal examination of age-associated declines in reasoning and processing speed. *Developmental Psychology, 45*, 431–446.

Tucker-Drob, E. M., & Salthouse, T. A. (2008). Adult age trends in the relations among cognitive abilities. *Psychology and Aging, 23*, 453–460.

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.

Van Dijk, K. R. A., Van Gerven, P. W. M., Van Boxtel, M. P. J., Van der Elst, W., & Jolles, J. (2008). No protective effects of education during normal cognitive aging: Results from the 6-year follow-up of the Maastricht Aging Study. *Psychology and Aging, 23*, 119–130.

Van Gerven, P. W. M., Meijer, W. A., & Jolles, J. (2007). Education does not protect against age-related decline of switching focal attention in working memory. *Brain and Cognition, 64*, 158–163.

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory and Cognition, 32*, 1206–1220.
Wagenmakers, E.-J., Ratcliff, R., Gomez, P., & McKoon, G. (2008). A diffusion model account of criterion shifts in the lexical decision task. *Journal of Memory and Language, 58*, 140–159.
Wagenmakers, E.-J., Van Der Maas, H. L. J., & Grassman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin and Review, 14*, 3–22.
Wechsler, D. (1997). *The Wechsler adult intelligence scale III*. San Antonio, CA: Psychological Corporation, Harcourt Brace.
Williams, B. R., Hultsch, D. F., Strauss, E. H., Hunter, M. A., & Tannock, R. (2005). Inconsistency in reaction time across the life span. *Neuropsychology, 19*, 88–96.
Woodworth, R. S. (1938). *Experimental Psychology*. New York: Henry Holt & Company.