

# A Diffusion Model Analysis of Adult Age Differences in Episodic and Semantic Long-Term Memory Retrieval

Julia Spaniol and David J. Madden  
Duke University Medical Center

Andreas Voss  
Albert-Ludwigs-Universität Freiburg

Two experiments investigated adult age differences in episodic and semantic long-term memory tasks, as a test of the hypothesis of specific age-related decline in context memory. Older adults were slower and exhibited lower episodic accuracy than younger adults. Fits of the diffusion model (R. Ratcliff, 1978) revealed age-related increases in nondecisional reaction time for both episodic and semantic retrieval. In Experiment 2, an age difference in boundary separation also indicated an age-related increase in conservative criterion setting. For episodic old–new recognition (Experiment 1) and source memory (Experiment 2), there was an age-related decrease in the quality of decision-driving information (drift rate). As predicted by the context-memory deficit hypothesis, there was no corresponding age-related decline in semantic drift rate.

*Keywords:* aging, reaction time, context memory, recognition test

Older adults typically perform worse than younger adults on tests of episodic memory, such as recall, recognition, or source memory. By contrast, tests of semantic memory—such as lexical decision, semantic categorization, or semantic priming—often show small or nonexistent age differences (Balota, Dolan, & Duchek, 2000; Light, 2000b; Zacks, Hasher, & Li, 2000). One locus of these age differences is the encoding of items for later retrieval (e.g., Craik, 1986, 1994; Glisky, Rubin, & Davidson, 2001), but retrieval itself also appears to be vulnerable to age-related decline (e.g., Anderson, Craik, & Naveh-Benjamin, 1998; Burke & Light, 1981). Further, as we show below, previous research suggests that the degree to which successful retrieval depends on the processing of contextual information is a critical determinant of age differences in memory performance.

One challenge in directly comparing episodic and semantic retrieval, and age differences therein, however, is the fact that different empirical metrics have been used to characterize performance in the two domains (Verhaeghen, 2000; but see McKoon & Ratcliff, 1979; McKoon, Ratcliff, & Dell, 1986). Episodic task performance is typically measured on an accuracy scale, whereas semantic task performance is most often measured in terms of reaction time (RT), accuracy often being at ceiling. This problem cannot be solved by simply using a common dependent variable

across tasks, because RT and accuracy each have different properties depending on the value of the other variable (Pachella, 1974; Santee & Egeth, 1982).

In the present study we report new evidence for differential age effects on episodic and semantic retrieval components and for an age-related deficit in the retrieval of episodic context information. We use the diffusion model (Ratcliff, 1978) to address the “metric problem” by simultaneously modeling accuracy and RT data and deriving estimates of the cognitive processes underlying younger and older adults’ retrieval performance.

## Aging and Context Memory

Long-term memory tasks vary with respect to their reliance on contextual representations. Semantic memory tasks generally do not require memory for context information, though it has been suggested that semantic task performance may be influenced by contextual cues (e.g., see Muter, 1978, for a demonstration of context effects in semantic retrieval). Episodic recognition tasks rely explicitly on context memory, by requiring the discrimination between events experienced in the context of an experimental study phase and events experienced extraexperimentally. Source memory tasks, by definition, make heavy demands on context memory, as do serial-order memory tasks and recall. The hypothesis that aging leads to disproportionate declines in context processing was first formulated in the early 1980s (e.g., Burke & Light, 1981; Rabinowitz & Ackerman, 1982) and continues to be a central theme of research on cognitive aging (e.g., Braver et al., 2001). Age differences in memory tasks tend to increase as a function of the tasks’ reliance on memory for contextual detail (for empirical and theoretical reviews, see Light, 1996; 2000b; Spaniol & Bayen, 2004; Spencer & Raz, 1995; Verhaeghen, Marcoen, & Goossens, 1993; Zacks et al., 2000). Older adults’ poor performance on tasks with high context reliance has been attributed to age deficits in specific cognitive processes such as self-initiated processing (e.g., Craik, 1986, 1994), recollection (e.g., Jacoby, 1999), or associative encoding (e.g., Naveh-Benjamin, 2000).

---

Julia Spaniol and David J. Madden, Center for the Study of Aging and Human Development and Department of Psychiatry and Behavioral Sciences, Duke University Medical Center; Andreas Voss, Institut für Psychologie, Albert-Ludwigs-Universität Freiburg, Freiburg, Germany.

This research was supported by Grants R37 AG02163 and R01 AG11622 from the National Institute on Aging. We are grateful to Sara E. Moore, Susanne M. Harris, and Leslie Crandell Dawes for technical assistance.

Correspondence concerning this article should be addressed to Julia Spaniol, who is now at the Rotman Research Institute, Baycrest, 3560 Bathurst St., Toronto, Ontario M6A 2E1, Canada. Email: jsaniol@rotman-baycrest.on.ca

Formal models have been used to determine which aspects of context processing are affected by age-related deficits. Early applications of formal models of episodic memory (e.g., Estes', 1955, stimulus fluctuation model) to cued-recall data provided support for an age-related deficit in the integration of item and context information during encoding (Kliegl & Lindenberger, 1993) or in the amount of context information encoded and the degree of contextual fluctuation over time (Balota, Duchek, & Paullin, 1989). More recently, Howard and Kahana's (2002) temporal context model, a single-store distributed model of item-context processing in episodic recall, was used to account for free-recall data from younger and older adults (Howard, Kahana, & Wingfield, 2005; see also Kahana, Howard, Zaromb, & Wingfield, 2002). Results suggested an age-related deficit in the retrieval of context information, possibly due to interference from item-context associations that were formed during retrieval but not during encoding. Finally, on the basis of correlational analyses, age deficits in episodic memory (including tasks with high context-memory demands) have also been linked to age-related declines in general cognitive processing resources such as speed or working memory. Importantly, however, these variables appear not to account for all of the age-related variance in episodic memory measures (e.g., Hertzog, Dixon, Hultsch, & MacDonald, 2003; Verhaeghen & Salthouse, 1997).

In sum, there is strong empirical support for an age-related deficit in tasks requiring memory for context information. Although the source of this deficit has been a matter of some debate, a complete explanation seems to require the assumption of a specific cause above and beyond age-related declines in general processing resources (see also Light, 1996, 2000b; Spencer & Raz, 1995). We henceforth refer to this idea as the *context-memory deficit hypothesis*.

### Overcoming the Metric Problem: Previous Approaches

We now return to the issue of the "metric problem"—the fact that performance in episodic tasks tends to be measured in terms of accuracy, whereas performance in semantic tasks is measured in terms of RT with accuracy near ceiling. One strategy for solving the metric problem is a dual-task design that enables a comparison of semantic and episodic retrieval costs, in terms of their effects on performance in a concurrent, unrelated task. Using this strategy, Veiel and Storandt (2003) measured younger and older adults' performance on semantic word generation (Experiment 1) and episodic free-recall tasks (Experiments 2 and 3), under full-attention and divided-attention conditions. From hierarchical regression analyses, the authors determined that, when controlling for full-attention performance on the secondary task (visual target detection), retrieval costs on the secondary task during the divided-attention conditions were not significantly different for younger and older adults, for either episodic or semantic retrieval. Veiel and Storandt concluded that age differences in both episodic and semantic retrieval can be accounted for by a general slowing mechanism (e.g., Salthouse, 1996; Salthouse & Madden, in press).

A second approach to the metric problem is to integrate accuracy and RT measures of episodic and semantic retrieval in an analysis of speed-accuracy trade-off (SAT) functions for both retrieval types. The retrieval SAT methodology has been used extensively in the general literature (e.g., Doshier & Rosedale,

1991; Hintzman & Curran, 1997) but has rarely been applied to investigate age-related changes in memory (for an exception see Laver, 2000). To our knowledge, the only direct comparison of episodic and semantic tasks using this approach is that of Healy and Light (2004). These authors also used a response deadline procedure (with time points spanning 100–2,000 ms), allowing them to chart complete SAT functions for episodic retrieval (old-new recognition) and semantic retrieval (lexical decision). Analyses of theoretically derived parameters of the SAT functions indicated that the onset of the retrieval process (both episodic and semantic) occurred earlier for younger adults than for older adults but that the dynamic and asymptotic differences between episodic and semantic retrieval were similar for the two age groups. Thus, these findings, like those of Veiel and Storandt (2003), contradict the hypothesis of a specific age-related deficit in context memory.

As with every method, both the dual-task and SAT methods have limitations. Dual-task costs can be difficult to interpret as a result of potential individual (and age group) differences in task strategies and in the allocation of processing resources to the primary and secondary tasks (Salthouse, 1988). In addition, dual-task costs give only an indirect measure of the component processes of memory retrieval (i.e., the quality of the retrieved information vs. the response criteria against which this information is evaluated) and are uninformative as to the relative contributions of these processes. The SAT procedure with response signals, in contrast, does allow for a more fine-grained analysis of different aspects of retrieval (onset, rate, and overall quality), but it is not always clear whether the processes underlying performance in the SAT paradigm are equivalent to those underlying performance under standard (self-paced) conditions (e.g., Wickelgren, 1977; but see Ratcliff, 1988).

We propose that age differences in episodic and semantic retrieval are best examined in the context of a theoretically informed measurement model that can account for both RT and accuracy data when these are acquired under standard conditions (i.e., without SAT or dual-task manipulations). As we show in the following sections, the diffusion model (e.g., Ratcliff, 1978; Ratcliff, Van Zandt, & McKoon, 1999), a sequential-sampling model of two-choice decisions, fulfills this requirement because it has been shown to account for correct and error RT patterns across a variety of cognitive domains, for both younger and older adults.

### The Diffusion Model

The diffusion model (e.g., Ratcliff, 1978; Ratcliff et al., 1999) assumes that RTs in two-choice decisions can be decomposed into nondecisional (e.g., perceptual-motor) and decisional processes whose duration is determined by systematic as well as random influences. The nondecisional RT component is captured by model parameter  $t_0$ . The process underlying the decisional RT component is illustrated in Figure 1. The drift rate, model parameter  $\nu$ , is the systematic influence that moves the decision process from a starting point (parameter  $z$ ) toward one of two response boundaries. As soon as a boundary is reached, the decision process terminates, and a response is initiated. In our example, an arrow pointing toward the upper boundary illustrates a positive drift rate. The value of the drift rate depends on the quality of the information being recovered. For example, in an old-new recognition test, recently studied test words would evoke a higher *old* drift than words that were

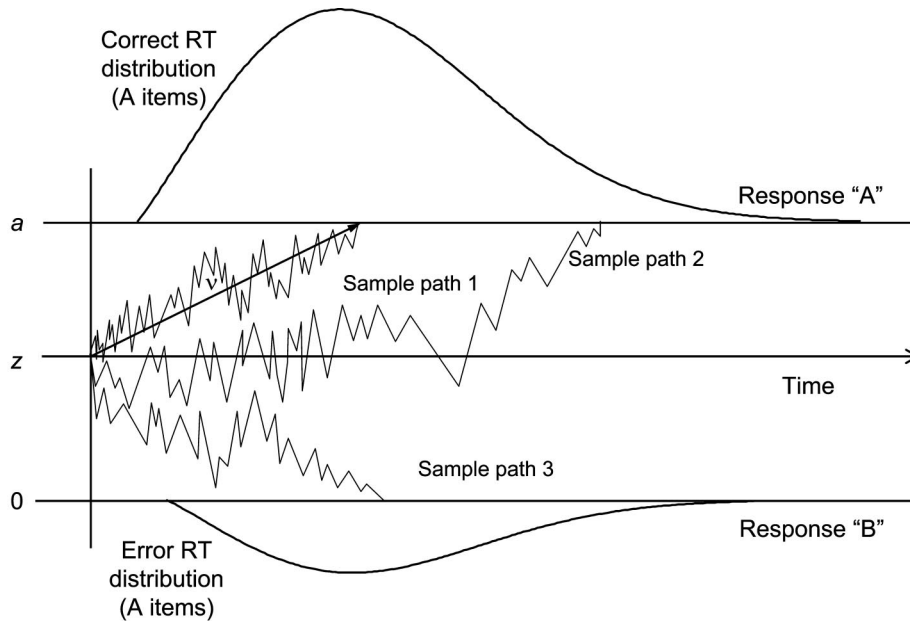


Figure 1. Illustration of the diffusion process for a category A item, in a task requiring the discrimination of items into category A or category B. The diffusion process begins at the starting point  $z$  and is driven toward the upper boundary  $a$  (“A” response) by a positive drift rate,  $\nu$ . A single drift rate is shown, although drift rates are assumed to be normally distributed across trials. Three sample paths illustrate that random variation affects the information accumulation process. Sample paths 1 and 2 result in the correct (“A”) response, whereas sample path 3 drifts toward the lower boundary 0 (“B”) response, resulting in an error. RT = reaction time.

studied less recently (see Ratcliff, Thapar, & McKoon, 2004, for a discussion of the similarity between drift rate and the familiarity signal in global memory models). Boundary separation parameter  $a$  captures the distance between the lower and upper boundaries; the value of  $a$  thus determines how much information is required on average before either response is initiated. Because of random noise in the information accumulation process, the time required by the decision process to reach one of the two boundaries and thereby initiate a response is variable, and occasionally the process terminates at the incorrect boundary. This “within-trial” variability is illustrated by the three sample paths in Figure 1; it is incorporated in the diffusion model as a scaling parameter that is fixed rather than estimated from the data.

Other nonsystematic influences—not depicted in Figure 1—can be modeled explicitly, allowing the model to account for differences in correct and error RT distributions. Ratcliff and Rouder (1998) showed that large across-trial variability in drift rate (assumed to be normally distributed with  $SD = s_\nu$ ) is associated with slow error responses, whereas large variability in starting point (assumed to be uniformly distributed with range  $s_z$ ) is associated with fast error responses. If values of  $s_\nu$  and  $s_z$  are of moderate size, errors are slow if accuracy is in a moderate range, and errors are fast if accuracy is extremely low or extremely high. In addition, Ratcliff and Tuerlinckx (2002) showed that variability in the nondecisional RT component (assumed to be uniformly distributed with range  $s_r$ ) can account for fast responses. For high drift rates, variability in the nondecisional component shortens the leading edge of the RT distribution.

The validity of the diffusion parameters has been tested experimentally (e.g., Ratcliff, 1985; Ratcliff et al., 1999; Voss, Rother-

mund, & Voss, 2004). For example, Voss et al. (2004) showed that boundary separation increased following the introduction of accuracy rewards; drift rates decreased when stimuli were harder to discriminate; nondecisional RT increased when the motor demands of responding were higher; and the position of the starting point relative to the two decision boundaries varied as a function of response-specific payoffs. The speed–accuracy manipulations included in most of the studies on aging by Ratcliff and colleagues (see next paragraph) are another important example of successful validity tests of the model parameters. In these studies, speed–accuracy instructions selectively affected decision boundaries, but not other parameters, such as drift rate. A full review of the diffusion modeling literature is beyond the scope of this article; next we summarize findings from studies that have applied the diffusion model to adult age differences in cognition.

#### Age Differences in Diffusion Model Parameters

A series of recent studies have applied the diffusion model to investigate age differences across a range of two-choice decision tasks: visual signal detection, numerosity judgments, and distance judgments (Ratcliff, Thapar, & McKoon, 2001); brightness perception (Ratcliff, Thapar, & McKoon, 2003); masked letter discrimination (Thapar, Ratcliff, & McKoon, 2003); recognition memory (Ratcliff, Thapar, & McKoon, 2004); and lexical decision (Ratcliff, Thapar, Gomez, & McKoon, 2004). These studies have consistently found older adults to have longer nondecisional RT components compared with younger adults, with the degree of slowing varying as a function of cognitive domain. Another regularity reported for all tasks except brightness discrimination was

an age-related increase in boundary separation (parameter  $a$ ), indicating that older adults require more information before initiating a response than do younger adults. Surprisingly, age differences in drift rate were almost always nonsignificant, which suggests that age differences in RT distributions could not generally be attributed to an age-related decline in the quality of the information recovered during the decision process. The only exception was observed for masked letter discrimination (Thapar et al., 2003), a task that requires fast visual processing. Here drift rate was significantly lower for older adults, a result consistent with other reports of age-related visual-perceptual decrements (Schneider & Pichora-Fuller, 2000; Scialfa, 2002). Finally, for recognition memory, drift variability (parameter  $sv$ ) was slightly greater for older adults than for younger adults (Ratcliff, Thapar, & McKoon, 2004).

### The Current Study

As noted previously, the only direct comparison of episodic and semantic retrieval processes for younger and older adults has been made in the context of SAT methodology (Healy & Light, 2004). The primary goal of the current study was to perform this direct comparison, using the diffusion model to estimate the contributions of the component processes—nondecisional and decisional, systematic and random—that give rise to accuracy and RT patterns in each age group. Age-related deficits in any component of context processing (encoding, storage/forgetting, retrieval) should lead to declines in the drift parameter ( $\nu$ ), which represents the quality of the retrieved information. If age-related declines in this parameter are greater for episodic retrieval than for semantic retrieval, this would be particularly strong support for the context-memory deficit hypothesis (e.g., Light, 1996, 2000b; Spencer & Raz, 1995).

We used an experimental design that equated important task demands across episodic and semantic retrieval conditions (study-test task structure, stimuli, study without knowledge of the type of subsequent test, and response demands; see Doshier & Rosedale, 1991, for a similar strategy). Although many prior applications of the diffusion model have involved manipulations of speed-accuracy criteria using task instructions and feedback, as well as extensive practice spanning multiple experimental sessions (e.g., Ratcliff, Thapar, & McKoon, 2004), we did not take that approach. Instead we opted for conditions more representative of other studies that have not applied the diffusion model, although we recognized the potential costs inflicted by this strategy (greater variability in the data, a smaller range of accuracies for each retrieval type). We used old-new recognition (Experiment 1) and a two-choice source memory task (Experiment 2) as episodic retrieval tasks, and living-nonliving discrimination as a semantic retrieval task. Our rationale for choosing these tasks was that they shared a similar two-choice structure and differed only in their episodic-semantic character. In addition, comparing semantic memory (no contextual demands) to recognition (low contextual demands) and source memory (high contextual demands) allowed us to study the sensitivity of model parameters to different retrieval conditions and to test predictions derived from an age-related context-memory deficit hypothesis.

Two hypotheses guided our research. First, on the basis of previous findings by Ratcliff and colleagues (see brief review

above), we expected to replicate age-related increases in nondecisional RT ( $t_0$ ) and in conservative decision boundary separation ( $a$ ), for both episodic and semantic retrieval. Second, in line with an age-related context-memory deficit hypothesis, we hypothesized that age differences in drift rate ( $\nu$ ) would be observed for episodic retrieval but not for semantic retrieval. As stated above, if older adults' difficulties in episodic memory result from an age-related deficit in the retrieval of contextual information, then this deficit should be reflected in the drift rate parameter because the drift rate reflects the quality of the information driving the decision process. Thus, if older adults retrieve less context information than younger adults, older adults' drift rates should be lower than younger adults' when decisions depend on context retrieval, as is the case for recognition and, especially, source memory decisions, but not for semantic decisions. This prediction conflicts to some extent with the Ratcliff, Thapar, and McKoon (2004) finding of no age difference in recognition drift rates; however, these authors did report a trend in the direction of our prediction.

### General Method

#### Participants

The Institutional Review Board of the Duke University Medical Center approved the research procedures, and all participants gave written informed consent. Participant characteristics are shown in Table 1. In Experiment 1, there were 24 younger adults (12 women) between 18 and 22 years of age, and 24 older adults (12 women) between 66 and 84 years of age. Experiment 2 comprised 24 younger adults (12 women) between 18

Table 1  
*Participant Characteristics by Age Group*

Characteristic	<i>M</i>		<i>SD</i>	
	Younger adults	Older adults	Younger adults	Older adults
Experiment 1 ( $n = 24$ per group)				
Age (years)	19.17	71.08*	1.05	4.57
Education	12.75	16.33*	0.94	2.71
Acuity	15.21	27.08*	1.02	10.93
Vocabulary	64.75	64.25	4.26	3.66
Digit symbol accuracy	98.03	96.64*	1.92	2.42
Digit symbol RT	1,313.60	1,862.90*	168.25	323.60
Experiment 2 ( $n = 24$ per group)				
Age (years)	21.63	71.46*	3.23	6.30
Education	15.29	16.54	2.29	2.17
Acuity	15.63	29.38*	1.69	10.46
Color vision	13.96	13.25*	0.20	0.79
Vocabulary	64.50	64.29	3.79	3.42
Digit symbol accuracy	97.45	97.05	2.71	2.99
Digit symbol RT	1,296.70	1,979.00*	166.65	392.95

*Note.* Vocabulary = raw score (maximum of 70) on the Wechsler Adult Intelligence Scale—Revised (Wechsler, 1981); Acuity = denominator of the Snellen fraction for corrected near vision; Color vision = raw score (maximum of 14) on the Dvorine color plates (Dvorine, 1963); Digit symbol accuracy and Digit symbol RT = percentage correct and reaction time (ms), respectively, on a computer test of digit symbol coding (Salt-house, 1992).

\* Significant age group difference within each experiment,  $p < .05$ .

and 29 years of age and 24 older adults (13 women) between 63 and 87 years of age, none of whom had participated in Experiment 1. The younger adults were students who received course credit or were compensated for their participation. The older adults were community-dwelling individuals who were compensated for participating in the study. Participants completed a vocabulary measure (Wechsler, 1981) and a computer-administered digit symbol substitution test (Salthouse, 1992). All participants demonstrated corrected near visual acuity of at least 20/40. No color vision test was administered to participants in Experiment 1 because all stimuli in that experiment were black-and-white. Participants in Experiment 2 (which used color stimuli) demonstrated normal color vision of at least 12 points on the Dvorine (1963) pseudoisochromatic plates. All participants were screened for major health problems (e.g., stroke, Parkinson's disease) and the use of psychotropic medications.

### Stimuli and Apparatus

The stimuli used in both experiments consisted of 468 nouns, 4–9 letters long ( $M = 6$ ), with a Kučera-Francis (1967) word frequency of 10–849 ( $Md = 25$ ). In the judgment of the authors and two independent raters, half of these words could be unambiguously classified as living (e.g., “rabbit”) or nonliving (e.g., “mirror”), respectively.

The experimental task was created and run in MATLAB (Mathworks, Natick, MA), using the Psychophysics Toolbox extensions (Brainard, 1997; Pelli, 1997). Stimulus presentation was controlled by a 2.0 GHz processor, Pentium 4 microcomputer with a 19-in., flat panel LCD. Viewing distance was approximately 60 cm, although head movement was not restrained. Participants responded using the “Z” and “/” keys on the computer keyboard, by resting the index finger of each hand on the two response keys during testing.

### Design and Procedure

Following two short practice blocks, participants in both studies completed six study–test blocks. Each block involved a study phase, a retention interval with a perceptual-motor filler task, and either an episodic or a semantic test. Depending on the type of test, each study–test block could

thus be classified as an episodic (E) or a semantic (S) block. Two block orders, ESSESE and SEESES, were counterbalanced across participants within each age group. The assignment of left and right response keys (“Z” and “/”) to responses (e.g., “living” vs. “nonliving” during semantic tests) was also counterbalanced, so that each of the four combinations of block order and response-key assignment occurred equally often within each age group.

Each of the six study–test blocks used a different word list. These lists, each composed of an equal number of living and nonliving words, contained both the study words and the test words used in each block. The order of presentation of these lists was identical for all participants, but, as noted previously, the order of semantic and episodic study–test blocks was counterbalanced across participants. Thus, each word was used equally often in episodic and semantic study–test blocks within each age group. The order in which the words of each list were presented within the study and test phases of each block was individually randomized for each participant.

## Experiment 1

### Method

As described in the General Method section, participants completed two short practice blocks (one episodic and one semantic) and six study–test blocks (three episodic and three semantic). The sequence for each block included a study phase, a choice RT task, and a test phase. The task is illustrated in Figure 2.

During the study phase of each block, participants viewed 39 words (including one primacy and two recency buffer words), each with a duration of 3 s, presented without interstimulus intervals (ISIs). Across the six study–test blocks, half of the buffer words were living, half were nonliving. Within each block, half of the 36 nonbuffer words were living, half nonliving. Words appeared in white lowercase 56-point Arial font on a black background in the center of the screen.

To minimize variability in encoding strategies, we asked participants to provide pleasantness judgments during the study phase. While each word was presented, participants judged whether it was pleasant or unpleasant

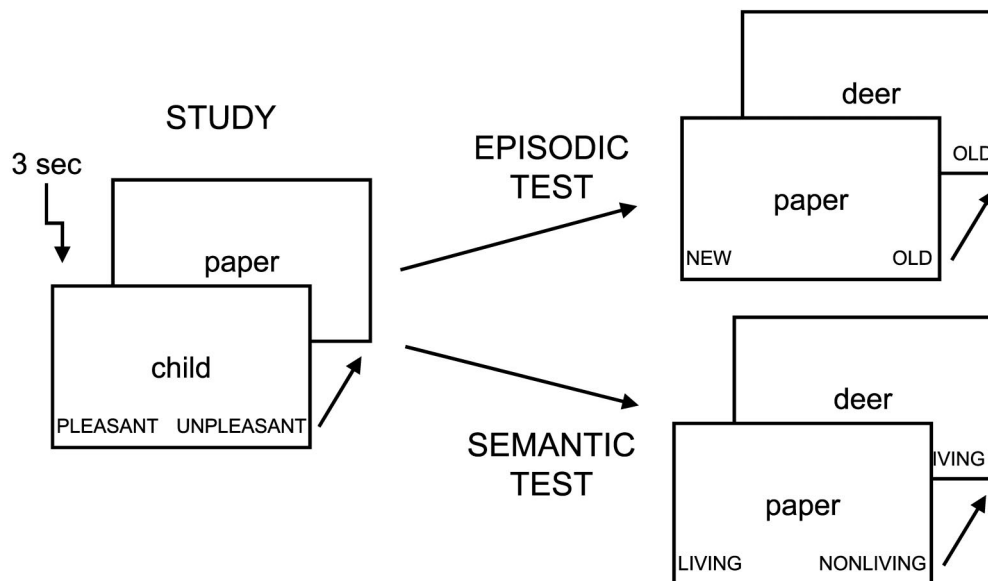


Figure 2. Illustration of study–test blocks in Experiment 1. In each block, participants received either an episodic or a semantic test. Study and test phases were separated by a brief retention interval featuring an independent perceptual-motor task (not shown).

and indicated their response by pressing the appropriate response key. The words *pleasant* and *unpleasant* appeared in white uppercase 36-pt. Arial font in bottom-left and bottom-right positions on the screen to remind participants of the response-key assignment. We recorded pleasantness judgments and RTs but did not use them in any of the analyses.

During a retention interval that lasted at least 1 min, participants completed a choice RT task. They pressed the left response key if the word *left* appeared or the right response key if the word *right* appeared. The words *left* and *right* each occurred 11 times in random intermixed order. Each word stayed on the screen until a response was made. After a 1-s blank screen, the next word appeared.

Prior to the beginning of the test phase, we informed participants about the nature of the upcoming test (i.e., living–nonliving judgments or old–new judgments) and instructed them to respond as quickly and accurately as possible. Seventy-eight test words were presented in white lowercase 56-pt. Arial font on a black background in the center of the screen. We excluded the first 6 words of the test list from analysis. These included the 3 primacy and recency buffer words from the study phase plus 3 other words. The remaining 72 words included the 36 nonbuffer words from the study phase and 36 new words (half living, half nonliving). The words *old* and *new* (episodic blocks) or *living* and *nonliving* (semantic blocks) appeared in white uppercase 36-pt. Arial font in bottom-left and bottom-right positions on the screen to remind participants of the response-key assignment. The test word remained on the screen until a response occurred. As soon as the participant pressed a response key, the test word disappeared, and the next test word was presented after a 1,000-ms delay. No performance feedback was given.

The 72 nonbuffer test words included two sets of 36 words (divided randomly), and the assignment of which of these were the 36 items from the study list (old words) and which were the 36 new test words was counterbalanced across participants in each age group. The use of words in each set as old or new was determined by the response order counterbalancing variable.

## Results

All statistical tests reported in this article used an alpha significance level of  $p < .05$  unless specified otherwise. RTs were excluded from the analyses if they were either less than 300 ms or greater than 3,000 ms. Less than 1% of responses in episodic and semantic tests in each age group in Experiment 1 were excluded for this reason (see Ratcliff & Tuerlinckx, 2002, for a discussion of the effects of outlier RTs on the estimation of diffusion models). We first present analyses performed on raw accuracy and RT data, to provide the reader with a detailed picture of the data that entered the modeling process. We then report results from fits of the diffusion model.

### Accuracy

We calculated the sensitivity index  $d'$  for old–new recognition and used an analogous measure for living–nonliving decisions, arbitrarily classifying correct “living” responses as hits and incorrect “living” responses as false alarms. The resulting data are presented in Table 2. We performed a  $2 \times 2$  split-plot analysis of variance (ANOVA) on  $d'$  with age (younger adults vs. older adults) as a between-subjects variable and retrieval type (semantic vs. episodic) as a within-subjects variable. We obtained significant effects of age,  $F(1, 46) = 3.96$ ,  $MSE = .51$ ,  $\eta^2 = .08$ , and retrieval type,  $F(1, 46) = 90.02$ ,  $MSE = .24$ ,  $\eta^2 = .53$ , as well as a significant Age  $\times$  Retrieval Type interaction,  $F(1, 46) = 33.31$ ,  $MSE = .24$ ,  $\eta^2 = .20$ . Additional planned comparisons indicated

Table 2  
*Hit Rates, False Alarm Rates, and  $d'$  in Experiments 1 and 2*

	<i>M</i>		<i>SD</i>	
	Younger adults	Older adults	Younger adults	Older adults
Experiment 1				
Episodic				
HR	0.94	0.92	0.09	0.06
FAR	0.05	0.13	0.07	0.07
$d'$	3.50	2.63	0.82	0.45
Semantic				
HR	0.95	0.97	0.06	0.02
FAR	0.02	0.01	0.02	0.01
$d'$	3.86	4.14	0.57	0.55
Experiment 2				
Episodic				
HR	0.90	0.79	0.09	0.17
FAR	0.11	0.22	0.09	0.13
$d'$	2.68	1.74	0.92	1.05
Semantic				
HR	0.93	0.97	0.09	0.04
FAR	0.01	0.01	0.02	0.02
$d'$	3.62	3.87	0.64	0.49

Note. HR = hit rate; FAR = false alarm rate.

that older adults had lower  $d'$  values than younger adults in the episodic task,  $t(46) = 4.51$ ,  $SE = .19$ ,  $\eta^2 = .31$ , but there was no age difference in the semantic task,  $t(46) = 1.75$ ,  $SE = .16$ ,  $\eta^2 = .06$ . Accuracy was significantly higher in the semantic task than in the episodic task, for both younger adults,  $t(23) = 2.16$ ,  $SE = .17$ ,  $\eta^2 = .17$ , and older adults,  $t(23) = 15.03$ ,  $SE = .10$ ,  $\eta^2 = .91$ .

### RT

RT data for correct and error responses are presented in Table 3. We performed a  $2 \times 2$  split-plot ANOVA on median RT with age (younger adults vs. older adults) as a between-subjects variable and retrieval type (semantic vs. episodic) as a within-subjects variable, separately for correct and error response RTs. In the analysis of correct RTs, the main effects of age,  $F(1, 46) = 18.05$ ,  $MSE = .03$ ,  $\eta^2 = .28$ , and retrieval type,  $F(1, 46) = 71.83$ ,  $MSE = .01$ ,  $\eta^2 = .59$ , were statistically significant, as was the Age  $\times$  Retrieval Type interaction,  $F(1, 45) = 5.30$ ,  $MSE = .01$ ,  $\eta^2 = .05$ . Additional planned comparisons indicated that older adults were significantly slower than younger adults for both episodic retrieval,  $t(46) = 4.22$ ,  $SE = .04$ ,  $\eta^2 = .28$ , and semantic retrieval,  $t(46) = 3.84$ ,  $SE = .03$ ,  $\eta^2 = .24$ . Semantic retrieval was faster than episodic retrieval for both younger adults,  $t(23) \geq 5.99$ ,  $SE = .01$ ,  $\eta^2 = .61$ , and older adults,  $t(23) \geq 6.29$ ,  $SE = .02$ ,  $\eta^2 = .63$ . The significant interaction reflected the fact that, for correct responses, the RT cost of increasing task difficulty, from semantic retrieval to episodic retrieval, was greater for older adults than for younger adults. In the analysis of RT for error responses, the effects of age,  $F(1, 45) = 6.00$ ,  $MSE = .19$ ,  $\eta^2 = .12$ , and retrieval type,  $F(1, 45) = 22.81$ ,  $MSE = .07$ ,  $\eta^2 = .34$ , were significant, indicating that error RTs were longer for older adults than for younger adults and that error RTs were longer for episodic retrieval than for

Table 3  
Median Reaction Time (in ms) for Correct and Error Responses  
in Experiments 1 and 2

Response type	<i>M</i>		<i>SD</i>	
	Younger adults	Older adults	Younger adults	Older adults
Experiment 1				
Correct responses				
Episodic	814	985	96	173
Semantic	742	858	88	119
Error responses				
Episodic	1,125	1,330	381	320
Semantic	1,085	861	415	294
Experiment 2				
Correct responses				
Episodic	1,024	1,442	154	333
Semantic	713	892	122	144
Error responses				
Episodic	1,582	1,770	671	479
Semantic	978	1,349	718	756

semantic retrieval. The Age  $\times$  Retrieval Type interaction was not significant,  $F(1, 45) = .01$ ,  $MSE = 0.07$ ,  $\eta^2 < .01$ .

### Diffusion Models

We used a method introduced by Voss et al. (2004) to estimate the parameters of the diffusion model. This method uses the Kolmogorov–Smirnov (KS) test statistic  $T$  (Kolmogorov, 1941) as the optimization criterion in an iterative search for the best-fitting model solution. The value of  $T$  reflects the maximum vertical distance between empirical cumulative RT distributions and the theoretical cumulative RT distributions for a given set of model parameters. A Simplex search (Nelder & Mead, 1965) identifies the parameter combination associated with the smallest  $T$  statistic (i.e., the best model fit). A significant  $T$  statistic (e.g.,  $p < .05$ ) indicates that the discrepancy between empirical and theoretical RT distributions is unlikely to have come about by chance, thus signaling model misfit. Voss et al. pointed out that, unlike chi-square and weighted least-squares methods, which require RT distribution quantiles as input, the KS test has the advantage of using raw RTs, thereby preventing some of the information loss that may result from averaging RTs within each quantile.<sup>1</sup> Although Voss et al. proposed that the KS test may be more robust against the influence of outlier RTs than the maximum-likelihood method (see Ratcliff & Tuerlinckx, 2002; Voss et al., 2004), this claim has not been tested formally.

Applying the KS test to estimate parameters of the diffusion model requires that two theoretical cumulative RT distributions be fitted simultaneously to two empirical cumulative RT distributions—one associated with the lower response boundary (e.g., the *new* boundary in an old–new recognition test), the other associated with the upper boundary (e.g., the *old* boundary in an old–new recognition test). A practical approach proposed by Voss et al. (2004) is to assign a negative sign to one of the cumulative RT distributions (e.g., the lower boundary RT distribution) and con-

catenating the two distributions so they form a single distribution with negative and positive RT ranges. A single KS test can then be used to assess model fit.

We estimated independent models of episodic and semantic retrieval for each participant using each participant’s RT distributions for old and new words (episodic retrieval) and for living and nonliving words (semantic retrieval). For the episodic model, the upper boundary was associated with “old” responses, and the lower boundary was associated with “new” responses. Following the logic described above, we simultaneously estimated two episodic submodels in which all parameters except the drift rate were set equal across the old and new word types. Thus, we estimated the following six episodic parameters for each participant in a single modeling step:  $t_0$ ,  $a$ ,  $z$ ,  $\nu_{Old}$ ,  $\nu_{New}$ , and  $s_\nu$ . We report  $z/a$  instead of  $z$  because it can be interpreted as a measure of response bias (see Voss et al., 2004). Values of  $z/a$  greater than 0.5 indicate a bias toward the response associated with the upper boundary, whereas values less than 0.5 indicate a bias toward the response associated with the lower boundary.

For the semantic model, the upper boundary was associated with “living” responses, the lower boundary was associated with “non-living” responses, and separate drift rates were estimated for living and nonliving words. The semantic model thus also contained two submodels, estimated simultaneously, that differed only in drift rates. We estimated the following six semantic parameters for each participant:  $t_0$ ,  $a$ ,  $z$ ,  $\nu_{Living}$ ,  $\nu_{Nonliving}$ , and  $s_\nu$ .

In all, we estimated 196 submodels (48 participants  $\times$  4 submodels). Each submodel was based on 108 responses minus the number of excluded responses, if any. Group means of the model parameters are presented in Figure 3. We submitted the model parameters to a series of analyses. In a first analysis we tested for the presence of response bias in each age group, separately for episodic and semantic models. In a second step we examined effects of age and word type (old–new, living–nonliving) on drift rates, separately for episodic and semantic models. We were not able to examine effects of word type on other parameters because, to increase statistical power, those parameters were estimated for both word types simultaneously. Finally, we analyzed effects of age and retrieval type (episodic vs. semantic) on each parameter to test the research hypotheses described in the Introduction.

**Response bias.** Significant bias to respond “old” in the episodic task (i.e.,  $z/a > 0.5$ ) was present in both younger adults,  $t(23) = 2.67$ ,  $SE = .02$ ,  $\eta^2 = .24$ , and older adults,  $t(23) = 5.41$ ,  $SE = .01$ ,  $\eta^2 = .56$ . Similarly, bias to respond “living” in the semantic task (i.e.,  $z/a > 0.5$ ) was present in both younger adults,  $t(23) = 5.78$ ,  $SE = .02$ ,  $\eta^2 = .59$ , and older adults,  $t(23) = 5.49$ ,  $SE = .02$ ,  $\eta^2 = .57$ . These biases may reflect an anchoring on the response alternative that was named first in the task instructions (where the word *old* always preceded *new* and *living* always preceded *nonliving*).

**Drift rates: Effects of age and word type.** We submitted the absolute values of episodic drift rates to a  $2 \times 2$  split-plot ANOVA with age (younger adults vs. older adults) as a between-subjects

<sup>1</sup> However, a possible disadvantage of the KS statistic is that it captures only the largest difference between two cumulative RT distributions, rather than fitting the probability mass between quantiles, as in the methods discussed by Ratcliff and Tuerlinckx (2002).

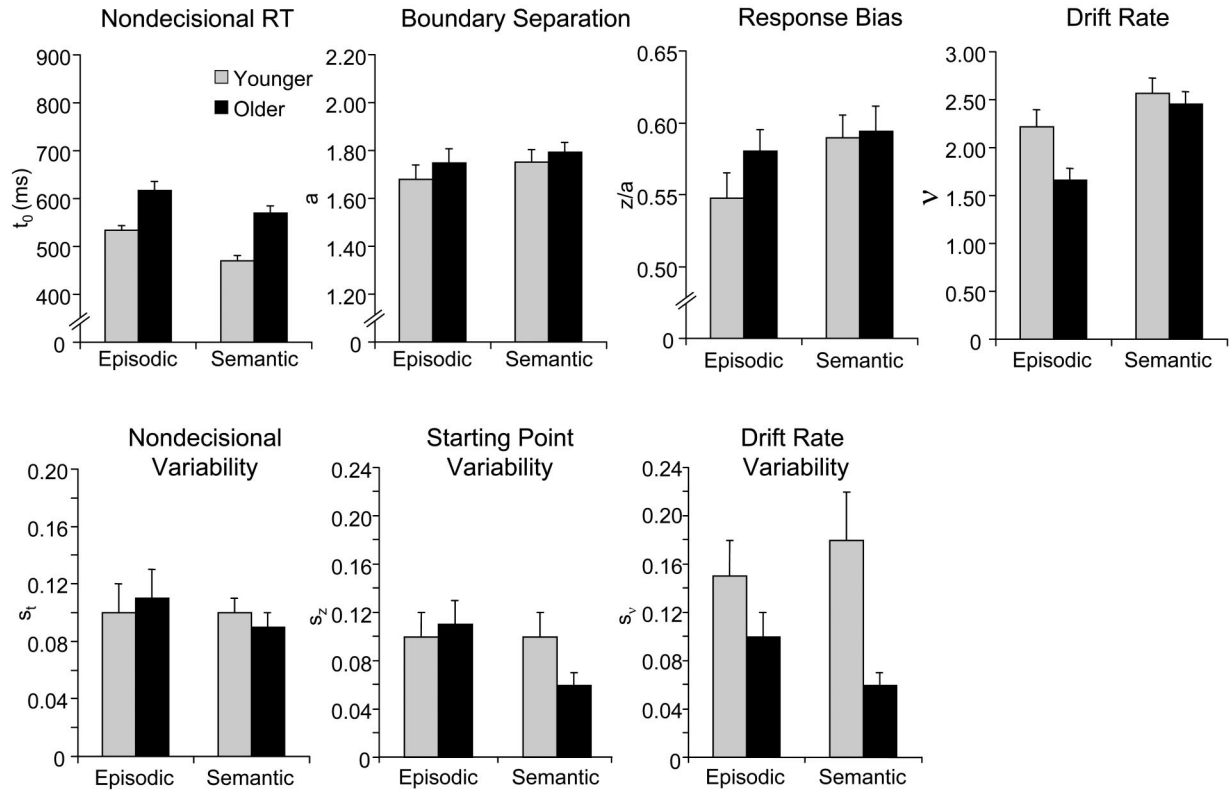


Figure 3. Mean estimates of diffusion model parameters for younger and older adults in Experiment 1. For drift rates, absolute values are shown, averaged across old and new words (episodic test) and over living and nonliving words (semantic test). For  $z/a$ , departures from 0.5 indicate a bias in favor of the response associated with the upper boundary ( $z/a > 0.5$ ) or the lower boundary ( $z/a < .05$ ). Error bars represent standard errors. RT = reaction time.

variable and word type (old vs. new) as a within-subjects variable. Older adults' drift rates were significantly lower than those of the younger adults,  $F(1, 46) = 7.10$ ,  $MSE = 1.07$ ,  $\eta^2 = .13$ . There was no significant effect of word type on drift rates,  $F(1, 46) = 1.00$ ,  $MSE = .53$ ,  $\eta^2 = .02$ , nor an Age  $\times$  Word Type interaction,  $F(1, 46) = 0.41$ ,  $MSE = .53$ ,  $\eta^2 = .01$ . By contrast, a similar ANOVA on semantic drift rates yielded no significant effect of age,  $F(1, 46) = 0.32$ ,  $MSE = 1.04$ ,  $\eta^2 = .01$ , but a significant effect of word type,  $F(1, 46) = 12.50$ ,  $MSE = .45$ ,  $\eta^2 = .21$ . The Age  $\times$  Word Type interaction was not significant,  $F(1, 46) = 0.20$ ,  $MSE = .45$ ,  $\eta^2 < .01$ . Drift rates were higher for responses to nonliving words than for responses to living words, suggesting that nonliving words possessed a more distinctive long-term memory representation than living words.

*All parameters: Effects of age and retrieval type.* We performed  $2 \times 2$  split-plot ANOVAs, separately for each parameter ( $t_0$ ,  $a$ ,  $z/a$ ,  $v$ ,  $s_v$ ,  $s_z$ , and  $s_v$ ), with age (younger adults vs. older adults) as a between-subjects variable and retrieval type (episodic vs. semantic) as a within-subjects variable.

The nondecisional RT component (parameter  $t_0$ ) was significantly longer for older adults than for younger adults,  $F(1, 46) = 23.99$ ,  $MSE = .01$ ,  $\eta^2 = .34$ . In addition,  $t_0$  was significantly longer for episodic retrieval than for semantic retrieval,  $F(1, 46) = 34.24$ ,  $MSE = .01$ ,  $\eta^2 = .44$ . There was no significant Age  $\times$  Retrieval Type interaction,  $F(1, 46) = 0.81$ ,  $MSE = .01$ ,  $\eta^2 < .01$ .

Boundary separation (parameter  $a$ ) showed no significant effects of age,  $F(1, 46) = 0.93$ ,  $MSE = .07$ ,  $\eta^2 = .02$ ; retrieval type,  $F(1, 46) = 1.25$ ,  $MSE = .06$ ,  $\eta^2 = .03$ , or their interaction,  $F(1, 46) = 0.06$ ,  $MSE = .06$ ,  $\eta^2 < .01$ .

The analysis of response bias (parameter  $z/a$ ) yielded no significant effects of age,  $F(1, 46) = 1.05$ ,  $MSE = .01$ ,  $\eta^2 = .03$ ; retrieval type,  $F(1, 46) = 3.82$ ,  $MSE = .01$ ,  $\eta^2 = .08$ ; or their interaction,  $F(1, 46) = 0.96$ ,  $MSE = .01$ ,  $\eta^2 < .01$ .

To analyze drift rates, we calculated mean drift rates for each retrieval type, averaging the absolute values of  $v_{\text{Old}}$  and  $v_{\text{New}}$  to obtain  $v_{\text{Episodic}}$ , and averaging the absolute values of  $v_{\text{Living}}$  and  $v_{\text{Nonliving}}$  to obtain  $v_{\text{Semantic}}$ , for each participant. The effect of retrieval type on mean drift rates was significant,  $F(1, 46) = 34.12$ ,  $MSE = .46$ ,  $\eta^2 = .40$ , as was the Age  $\times$  Retrieval Type interaction,  $F(1, 46) = 5.22$ ,  $MSE = .46$ ,  $\eta^2 = .06$ . However, there was no significant main effect of age,  $F(1, 46) = 3.35$ ,  $MSE = .83$ ,  $\eta^2 = .07$ . Additional tests indicated that older adults had lower drift rates than younger adults in the episodic task,  $t(46) = 2.67$ ,  $SE = .21$ ,  $\eta^2 = .13$ , but not in the semantic task,  $t(46) = 0.56$ ,  $SE = .21$ ,  $\eta^2 = .01$ . Furthermore, episodic drift rates were lower than semantic drift rates for both younger adults,  $t(23) = 2.62$ ,  $SE = .13$ ,  $\eta^2 = .23$ , and older adults,  $t(23) = 5.54$ ,  $SE = .14$ ,  $\eta^2 = .57$ .

There were no significant effects of age, retrieval type, or their interaction on the variability of the nondecisional RT component

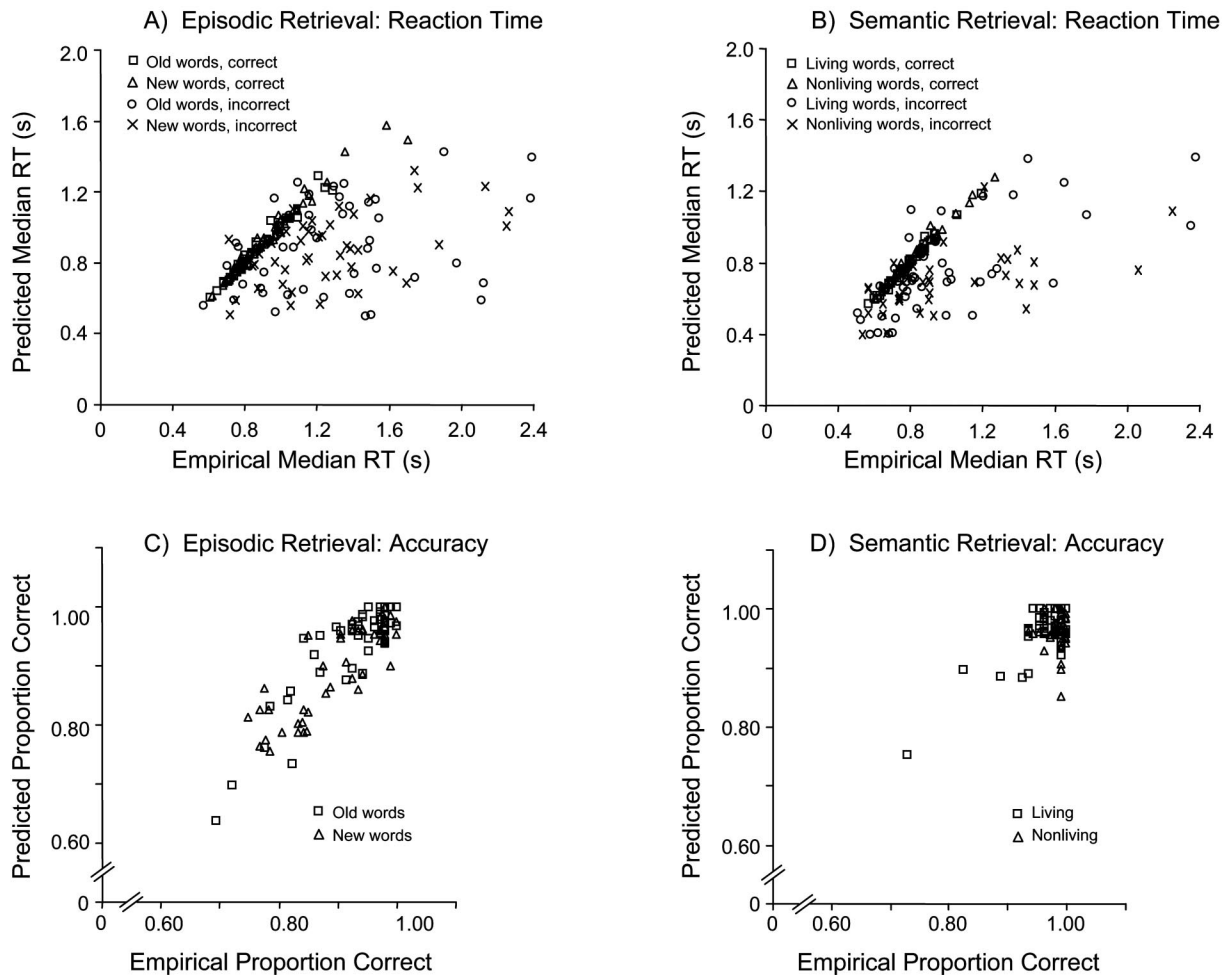
(parameter  $s_1$ ),  $F(1, 46) \leq 0.91$ ,  $MSE = .01$ ,  $\eta^2 < .01$ , or on starting point variability (parameter  $s_2$ ),  $F(1, 46) \leq 1.82$ ,  $MSE = .01$ ,  $\eta^2 \leq .03$ . Drift variability (parameter  $s\nu$ ) was lower in older adults than in younger adults,  $F(1, 46) = 8.30$ ,  $MSE = .02$ ,  $\eta^2 = .16$ , but no effects involving retrieval type were significant,  $F(1, 46) \leq 2.53$ ,  $MSE = .01$ ,  $\eta^2 \leq .06$ .

**Model fit.** We assessed goodness-of-fit for a total of 96 models (48 participants  $\times$  2 models, one each for episodic and semantic retrieval) using KS tests (for details, see Voss et al., 2004). Significant outcomes at the  $p < .05$  level indicated model misfit. Only one older adults' episodic model test was significant. Out of 96 tests, however, 4–5 tests would be expected to be significant by chance alone. Thus, using a quantitative criterion (the  $T$  statistic), we found that there was no evidence to suggest that model fit was poor.

Figure 4 provides a graphical illustration of model fit. For each participant and experimental condition, median RT for correct and error responses as predicted by the diffusion model was plotted against the empirical median RT (top panel), and the predicted

proportion of correct responses was plotted against the empirical proportion of correct responses (bottom panel). We present younger and older adults' data within the same plots because inspection of each age group's scatterplots revealed no systematic differences.

The RT plots for both episodic and semantic retrieval show that the model was successful in accounting for median correct RTs, which are clustered tightly on and around the diagonal. Median incorrect RTs, in contrast, were systematically underpredicted. The underprediction of median error RTs is likely due to the fact that accuracy was very high overall, with most participants contributing only a small number of misses (episodic retrieval:  $M = 7.94$ ,  $SD = 8.0$ ; semantic retrieval:  $M = 4.54$ ,  $SD = 4.98$ ) and false alarms (episodic retrieval:  $M = 10.15$ ,  $SD = 8.29$ ; semantic retrieval:  $M = 1.56$ ,  $SD = 1.57$ ; see also Table 2). As a result, the parameter estimates largely reflected each participant's roughly 200 correct RTs. Some of the bad fits represent a single observation. Indeed, there was a significant negative correlation between model misfit (the absolute deviation between empirical and pre-



**Figure 4.** Empirical and predicted median reaction time (RT) and accuracy in Experiment 1. Predicted values were based on the diffusion model as described in the Model Fit section of the Results section for Experiment 1. A: RT values for episodic retrieval; B: RT values for semantic retrieval; C: accuracy values for episodic retrieval; D: accuracy values for semantic retrieval.

dicted median RT) and the number of observations underlying the empirical median RT, for both episodic retrieval,  $r = -.64$ ,  $p < .01$ , and semantic retrieval,  $r = -.52$ ,  $p < .01$ . Voss et al. (2004) reported a similar finding in their Experiment 1. Voss et al.'s Experiment 2 showed that, when error rates were higher, their KS model estimation method did produce excellent model fits for both correct responses and errors.

### Discussion

The analyses of the diffusion model parameters in Experiment 1 yielded a significant main effect of retrieval type, without an accompanying Age  $\times$  Retrieval Type interaction, for the nondecisional RT component  $t_0$ . This result indicates relatively slower encoding and response processes for episodic retrieval as compared with semantic retrieval. The effect of retrieval type on the nondecisional RT component may represent an adaptive response to changes in perceived task difficulty, similar to a finding by Voss et al. (2004, Experiment 1). This adjustment likely affects response preparation or execution stages, rather than the encoding of the test stimuli, because episodic and semantic stimuli did not differ perceptually.

Older adults in Experiment 1 showed reduced accuracy and RT performance compared with younger adults, particularly for episodic retrieval. Fits of the diffusion model allowed us to pinpoint the cognitive processes involved in these age differences. As predicted, the results of Experiment 1 replicated previous findings by Ratcliff et al. (see brief review in the Introduction section) of age-related increases in nondecisional RT (parameter  $t_0$ ). Nondecisional processing exhibited age-related slowing of approximately 90 ms. However, contrary to our prediction and to most previous reports by Ratcliff et al. (Ratcliff, Thapar, Gomez, & McKoon, 2004; Ratcliff et al., 2001; Ratcliff, Thapar, & McKoon, 2004; Thapar et al., 2003), there was no age-related increase in conservative decision boundaries. This finding is not unique; Ratcliff et al. (2003) also reported a nonsignificant age difference in boundary separation in a brightness discrimination task. Our result suggests that, in this experiment at least, older adults' longer RTs for both retrieval types could not be accounted for by an age-related increase in conservatism but were largely caused by nondecisional slowing.

In line with our hypothesis, there was significant age-related decline in the drift rate parameter for episodic retrieval but not for semantic retrieval (see Figure 3).<sup>2</sup> Episodic drift rate was lower than semantic drift rate for both age groups. Note that, given the model architecture, lower drift rates lead to increases in RT and decreases in accuracy (e.g., Thapar et al., 2003). Our result suggests that the information driving the decision process during episodic retrieval was weaker for older adults than for younger adults, whereas the information driving semantic decisions was comparable in both age groups. The present result for episodic drift differs from that reported by Ratcliff, Thapar, and McKoon (2004). This discrepancy may reflect several procedural differences between our experiment and Ratcliff, Thapar, and McKoon's study. First, our participants received less task practice than Ratcliff, Thapar, and McKoon's participants. Therefore, older adults in our experiment may have had less opportunity to adopt optimal task settings in the episodic task. Second, unlike Ratcliff, Thapar, and McKoon, we included a 60-s retention interval, which may have

led to greater forgetting of episodic context information for older adults than for younger adults. Third, we used pleasantness judgments as an explicit encoding instruction and provided a 3-s study time per item, whereas Ratcliff, Thapar, and McKoon simply asked participants to study items for a later test, with only 1 s of study time per item. It is possible that younger adults benefited more than older adults from the encoding instructions and from the relatively long study times in our experiment.

The context-deficit hypothesis predicts that the age difference in episodic drift should be particularly pronounced in tasks that strongly rely on context memory. To replicate and extend our findings from Experiment 1, we therefore conducted a second experiment in which we replaced old–new recognition with a more context-dependent task: two-choice source memory judgments.

### Experiment 2

Experiment 2 differed from Experiment 1 with respect to the nature of its episodic retrieval task. Instead of old–new recognition, Experiment 2 included a context-memory task that tested participants for their memory of the screen location in which each test word was presented during the study phase.

### Method

The following aspects of the method in Experiment 2 differed from those in Experiment 1. During the study phase of each block, 28 words (half living, half nonliving; including two primacy buffers and two recency buffers) were presented in white lowercase 56-pt. Arial font, against a black background. Half of the words appeared inside a red color patch, located at the top of the screen, and the other half appeared inside a blue color patch, located at the bottom of the screen. The Experiment 2 tasks are illustrated in Figure 5. Among the 24 nonbuffer words, each combination of location (top or bottom) and semantic status (living or nonliving) occurred six times. We counterbalanced the assignment of the 24 nonbuffer test words to the two screen positions across participants in each age group. The 24 nonbuffer test words included a random sequence of two sets of 12 words (six living, six nonliving). The top–bottom assignment of each set was determined by the response order counterbalancing variable.

At the beginning of each study phase, we instructed participants to try to remember the spatial location of each word by associating the word with its location or with the color of the patch in which the word was shown. The duration of each word was 5 s; there were no ISIs. We instructed participants to refrain from responding while they studied the words.

A possible strategy for remembering the location of each word would be to attend to only one location (e.g., by ignoring all words presented in the bottom position). To discourage participants from using this strategy, we ensured that a brief recognition test followed each study phase. Three studied words (at least one from each location) and three new words were shown in random order, and participants provided old–new recognition judgments. All participants performed above chance on the recognition trials, and informal inquiries indicated that none of the participants used the spatially selective encoding strategy.

The retention interval comprised two perceptual-motor speed tasks that together lasted about 1.5 min: the left–right choice RT task that was also used in Experiment 1 and a simple RT task. Neither of these tasks was used in any of the analyses reported here.

<sup>2</sup> Accuracy in the semantic retrieval condition is near ceiling. Therefore it cannot be ruled out that a more difficult semantic task would produce group differences in semantic drift rates.

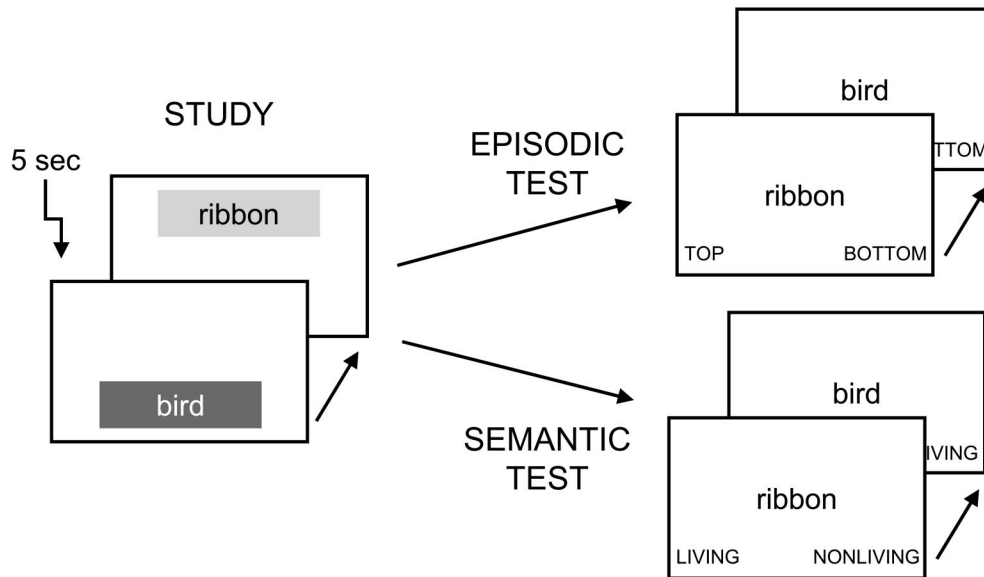


Figure 5. Illustration of study–test blocks in Experiment 2. In each block, participants received either an episodic or a semantic test. Study and test phases were separated by a brief retention interval featuring independent perceptual-motor tasks (not shown).

During the test phase, the words from the study phase were presented again, this time in the center of the screen, in white lowercase 56-pt. Arial font against a black background. Duration of the test word was again response-limited. As soon as a response was made, the test word disappeared, and the next test word appeared after a 1,000-ms delay. The four buffer words were presented first and were not included in any of the analyses. The 24 nonbuffer words were shown in a new random order. The words *top* and *bottom* (episodic blocks) or *living* and *nonliving* (semantic blocks) appeared in white uppercase 36-pt. Arial font in bottom-left and bottom-right positions on the screen to remind participants of the response-key assignment.

### Results

RTs that were either less than 300 ms or greater than 4,000 ms were excluded from the analyses. We used a higher upper RT cutoff than in Experiment 1 because RTs in the top–bottom judgment task were longer than RTs in the old–new recognition task used in Experiment 1. Less than 1% of semantic test responses were excluded in each age group. Less than 1% of younger adults' and 3.35% of older adults' episodic test responses were excluded.

### Accuracy

Accuracy measures were calculated as in Experiment 1. For the episodic task, we arbitrarily classified correct “top” responses as hits and incorrect “top” responses as false alarms. Hit rates, false-alarm rates, and  $d'$  values are shown in Table 2. A  $2 \times 2$  split-plot ANOVA on  $d'$  values with age group (younger adults vs. older adults) as a between-subjects variable and retrieval type (semantic vs. episodic) as a within-subjects variable yielded significant main effects of age,  $F(1, 46) = 4.00$ ,  $MSE = .70$ ,  $\eta^2 = .08$ , and retrieval type,  $F(1, 46) = 93.18$ ,  $MSE = .61$ ,  $\eta^2 = .09$ , as well as a significant Age  $\times$  Retrieval Type interaction,  $F(1, 46) = 14.09$ ,

$MSE = .61$ ,  $\eta^2 = .61$ . Additional tests indicated that older adults had lower  $d'$  values than younger adults for episodic retrieval,  $t(46) = 3.29$ ,  $SE = .28$ ,  $\eta^2 = .19$ , whereas there was no significant group difference in  $d'$  for semantic retrieval,  $t(46) = 1.55$ ,  $SE = .17$ ,  $\eta^2 = .05$ . Accuracy was significantly higher in the semantic task than in the episodic task for both younger adults,  $t(23) = 4.29$ ,  $SE = .22$ ,  $\eta^2 = .44$ , and older adults,  $t(23) = 9.23$ ,  $SE = .23$ ,  $\eta^2 = .79$ .

### RT

The RT data are presented in Table 3. We performed a  $2 \times 2$  split-plot ANOVA on median RT with age (younger adults vs. older adults) as a between-subjects variable and retrieval type (semantic vs. episodic) as a within-subjects variable, separately for correct and error RTs. In the analysis of correct RTs, the main effects of age,  $F(1, 46) = 40.69$ ,  $MSE = .05$ ,  $\eta^2 = .47$ , and retrieval type,  $F(1, 46) = 136.74$ ,  $MSE = .03$ ,  $\eta^2 = .71$ , were statistically significant, as was the Age  $\times$  Retrieval Type interaction,  $F(1, 46) = 10.59$ ,  $MSE = .03$ ,  $\eta^2 = .05$ . Follow-up tests indicated that older adults were significantly slower than younger adults in the episodic task,  $t(46) = 5.59$ ,  $SE = .07$ ,  $\eta^2 = .40$ , as well as in the semantic task,  $t(46) = 4.63$ ,  $SE = .04$ ,  $\eta^2 = .32$ . Additionally, correct responses were significantly slower in the episodic task than in the semantic task for both younger adults,  $t(23) = 12.30$ ,  $SE = .03$ ,  $\eta^2 = .87$ , and older adults,  $t(23) = 7.96$ ,  $SE = .07$ ,  $\eta^2 = .73$ . The significant interaction reflects the fact that RT costs of an increase in task difficulty (from semantic retrieval to episodic retrieval) were greater for older adults than for younger adults. In the analysis of error RTs, the main effect of retrieval type was significant,  $F(1, 31) = 14.48$ ,  $MSE = .31$ ,  $\eta^2 = .32$ , indicating that error responses were slower in the episodic task than in the

semantic task. There were no significant effects of age,  $F(1, 31) = 2.11$ ,  $MSE = .67$ ,  $\eta^2 = .06$ , or of the Age  $\times$  Retrieval Type interaction,  $F(1, 31) = 0.32$ ,  $MSE = .31$ ,  $\eta^2 = .01$ .

*Diffusion Models*

We fit diffusion models as described for Experiment 1, the only difference being the definition of the parameters to reflect the change in the episodic task. For episodic models, the upper boundary was now associated with “top” responses, and the lower boundary was associated with “bottom” responses. We estimated separate drift rates for top ( $\nu_{top}$ ) and bottom words ( $\nu_{bottom}$ ).

Each of the 196 submodels (48 participants  $\times$  4 submodels) was based on 72 responses minus the number of excluded responses, if any. Group means of the model parameters are presented in Figure 6. As in Experiment 1, we submitted the model parameters to a series of analyses in three stages: (a) testing for the presence of response bias in each age group, separately for episodic and semantic models; (b) examining effects of age and word type (old–new, living–nonliving) on drift rates, separately for episodic and semantic models; and (c) analyzing the effects of age and retrieval type (episodic vs. semantic) on each parameter.

*Response bias.* Younger adults showed no significant response bias in the episodic task,  $t(23) = 0.87$ ,  $SE = .02$ ,  $\eta^2 = .03$ ;

neither did older adults,  $t(23) = 1.94$ ,  $SE = .02$ ,  $\eta^2 = .14$ . In contrast, a significant bias to respond “living” in the semantic task (i.e.,  $z/a > 0.5$ ) was present in both younger adults,  $t(23) = 5.22$ ,  $SE = .02$ ,  $\eta^2 = .54$ , and older adults,  $t(23) = 2.76$ ,  $SE = .02$ ,  $\eta^2 = .25$ . This effect, similar to those seen in Experiment 1, may reflect anchoring on the response option named first in the instructions.

*Drift rates.* We submitted the absolute values of episodic drift rates to a  $2 \times 2$  split-plot ANOVA with age (younger adults vs. older adults) as between-subjects variable and word type (top vs. bottom) as within-subjects variable. Older adults’ episodic drift rates were significantly lower than younger adults’,  $F(1, 46) = 14.66$ ,  $MSE = .67$ ,  $\eta^2 = .24$ . Neither the effect of word type,  $F(1, 46) = 0.39$ ,  $MSE = .36$ ,  $\eta^2 = .01$ , nor the Age  $\times$  Word Type interaction,  $F(1, 46) < 0.01$ ,  $MSE = .36$ ,  $\eta^2 < .01$ , was significant. A  $2 \times 2$  split-plot ANOVA on semantic drift rates with age (younger adults vs. older adults) as a between-subjects variable and word type (living vs. nonliving) as a within-subjects variable yielded a significant effect of word type,  $F(1, 46) = 9.30$ ,  $MSE = .65$ ,  $\eta^2 = .16$ , indicating that drift rates for nonliving words were higher than drift rates for living words, similar to our finding in Experiment 1. Neither the age group effect,  $F(1, 46) = 2.30$ ,  $MSE = 1.12$ ,  $\eta^2 = .05$ , nor the Age  $\times$  Word Type interaction,  $F(1, 46) = 2.28$ ,  $MSE = .65$ ,  $\eta^2 = .04$ , was significant.

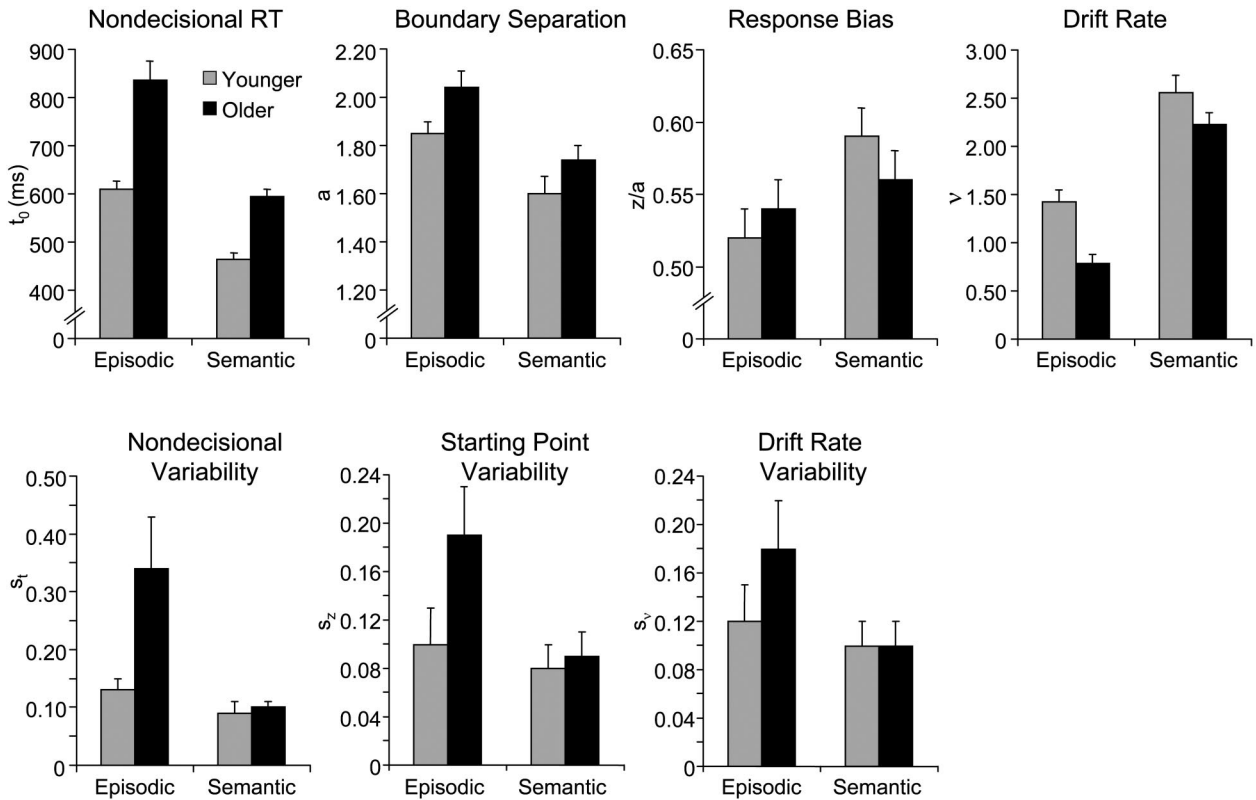


Figure 6. Mean estimates of diffusion model parameters for younger and older adults in Experiment 2. For drift rates, absolute values are shown, averaged across words presented in top and bottom locations (episodic test) and across living and nonliving words (semantic test). For  $z/a$ , departures from 0.5 indicate a bias in favor of the response associated with the upper boundary ( $z/a > 0.5$ ) or the lower boundary ( $z/a < .05$ ). Error bars represent standard errors. RT = reaction time.

*All parameters: Effects of age and retrieval type.* As in Experiment 1, we performed ANOVAs with age (younger adults vs. older adults) as a between-subjects variable and retrieval type (episodic vs. semantic) as a within-subjects variable, separately for each parameter ( $t_0$ ,  $a$ ,  $z/a$ ,  $\nu$ ,  $s_e$ ,  $s_z$ , and  $s_\nu$ ). For the analysis of drift rates we used mean episodic drift rates (calculated by averaging the absolute values of  $\nu_{\text{Top}}$  and  $\nu_{\text{Bottom}}$ ) and mean semantic drift rates (calculated by averaging the absolute values of  $\nu_{\text{Living}}$  and  $\nu_{\text{Nonliving}}$ ). These parameters are presented in Figure 6.

As in Experiment 1, there were significant effects of age,  $F(1, 46) = 48.33$ ,  $MSE = .02$ ,  $\eta^2 = .51$ , and retrieval type,  $F(1, 46) = 68.35$ ,  $MSE = .01$ ,  $\eta^2 = .58$ , on nondecisional RT (parameter  $t_0$ ), as well as a significant Age  $\times$  Retrieval Type interaction,  $F(1, 46) = 4.21$ ,  $MSE = .01$ ,  $\eta^2 = .03$ . Additional tests indicated that  $t_0$  was greater for episodic retrieval than for semantic retrieval for both younger adults,  $t(1, 23) = 9.84$ ,  $SE = .01$ ,  $\eta^2 = .81$ , and older adults,  $t(1, 23) = 5.44$ ,  $SE = .04$ ,  $\eta^2 = .56$ . The  $t_0$  parameter was greater for older adults than for younger adults for both episodic retrieval,  $t(1, 46) = 4.99$ ,  $SE = .05$ ,  $\eta^2 = .35$ , and semantic retrieval,  $t(1, 46) = 6.97$ ,  $SE = .02$ ,  $\eta^2 = .51$ .

Boundary separation (parameter  $a$ ) was significantly greater for older compared with younger adults,  $F(1, 46) = 6.31$ ,  $MSE = .10$ ,  $\eta^2 = .12$ , and for episodic compared with semantic retrieval,  $F(1, 46) = 20.73$ ,  $MSE < .09$ ,  $\eta^2 = .31$ . There was no significant Age  $\times$  Retrieval Type interaction on  $a$ ,  $F(1, 46) = 0.12$ ,  $MSE = .09$ ,  $\eta^2 < .01$ .

Response bias (parameter  $z/a$ ) was present only in the semantic task (see above), and neither the main effect of age,  $F(1, 46) = 0.00$ ,  $MSE = .01$ ,  $\eta^2 < .01$ , nor the Age  $\times$  Retrieval Type interaction,  $F(1, 46) = 1.91$ ,  $MSE = .01$ ,  $\eta^2 = .04$ , was significant.

We again used mean episodic drift rates (calculated by averaging the absolute values of  $\nu_{\text{Top}}$  and  $\nu_{\text{Bottom}}$  for each participant) and mean semantic drift rates (calculated by averaging the absolute values of  $\nu_{\text{Living}}$  and  $\nu_{\text{Nonliving}}$  for each participant) to facilitate a comparison across retrieval types. The effects of age,  $F(1, 46) = 9.96$ ,  $MSE = .57$ ,  $\eta^2 = .18$ , and retrieval type,  $F(1, 46) = 121.32$ ,  $MSE = .33$ ,  $\eta^2 = .72$ , on drift were significant, indicating higher drift for semantic as compared with episodic retrieval, and for younger as compared with older adults. Unlike in Experiment 1, the Age  $\times$  Retrieval Type interaction was nonsignificant,  $F(1, 46) = 1.80$ ,  $MSE = .33$ ,  $\eta^2 = .01$ . However, separate planned comparisons for each retrieval type, motivated by our findings in Experiment 1, showed a significant age difference in episodic drift,  $t(1, 46) = 3.83$ ,  $SE = .17$ ,  $\eta^2 = .24$ , but not in semantic drift,  $t(1, 46) = 1.52$ ,  $SE = .26$ ,  $\eta^2 = .05$ .

The variability of the nondecisional RT component (parameter  $s_r$ ) showed significant effects of age,  $F(1, 46) = 5.50$ ,  $MSE = .05$ ,  $\eta^2 = .11$ , and retrieval type,  $F(1, 46) = 9.17$ ,  $MSE = .12$ ,  $\eta^2 = .15$ , and an Age  $\times$  Retrieval Type interaction,  $F(1, 46) = 4.74$ ,  $MSE = .12$ ,  $\eta^2 = .08$ . Additional tests indicated that  $s_r$  was significantly greater for episodic retrieval than for semantic retrieval in older adults,  $t(1, 23) = 2.75$ ,  $SE = .09$ ,  $\eta^2 = .25$ , whereas younger adults'  $s_r$  did not vary as a function of retrieval type,  $t(1, 23) = 1.32$ ,  $SE = .03$ ,  $\eta^2 = .07$ . Older adults'  $s_r$  exceeded younger adults'  $s_r$  for episodic retrieval,  $t(1, 46) = 2.35$ ,  $SE = .09$ ,  $\eta^2 = .11$ , but not for semantic retrieval,  $t(1, 46) = 0.28$ ,  $SE = .03$ ,  $\eta^2 < .01$ .

Starting point variability (parameter  $s_z$ ) was greater for episodic retrieval than for semantic retrieval,  $F(1, 46) = 7.02$ ,  $MSE = .01$ ,

$\eta^2 = .13$ , but no effects involving age were significant,  $F(1, 46) \leq 3.18$ ,  $MSE \geq .01$ ,  $\eta^2 \leq .07$ . No effects on drift variability (parameter  $s_\nu$ ) were significant,  $F(1, 46) \leq 3.28$ ,  $MSE = .02$ ,  $\eta^2 \leq .07$ .

*Model fit.* We assessed goodness of fit for a total of 96 models (48 participants  $\times$  2 models, one each for episodic and semantic retrieval) using KS tests. None of the model tests were significant, suggesting that model fit was good overall.

Figure 7 provides a graphical illustration of model fit. For each participant and experimental condition, median RT for correct and error responses as predicted by the diffusion model was plotted against the empirical median RT (top panel), and the predicted proportion of correct responses was plotted against the empirical proportion of correct responses (bottom panel). We present younger and older adults' data within the same plots because inspection of each age group's scatterplots revealed no systematic differences.

The RT plots for both episodic and semantic retrieval show that the model was successful in accounting for median correct RTs. As in Experiment 1, however, the medians of the error RT distributions were systematically underpredicted. This again appears to be due to the fact that accuracy was very high overall, with most participants contributing only a small number of misses (episodic retrieval:  $M = 5.48$ ,  $SD = 5.01$ ; semantic retrieval:  $M = 1.81$ ,  $SD = 2.65$ ) and false alarms (episodic retrieval:  $M = 5.75$ ,  $SD = 4.39$ ; semantic retrieval:  $M = .42$ ,  $SD = .68$ ; see also Table 2). As a result, the parameter estimates largely reflected each participant's roughly 70 correct RTs. As in Experiment 1, some of the bad fits represent a single observation. Indeed, there was a significant negative correlation between model misfit (the absolute deviation between empirical and predicted median RT) and the number of observations underlying the empirical median RT, for both episodic retrieval,  $r = -.62$ ,  $p < .01$ , and semantic retrieval,  $r = -.50$ ,  $p < .01$ .

## Discussion

The results of Experiment 2 replicated several of the findings of Experiment 1. First, episodic retrieval was again associated with longer nondecisional RTs than semantic retrieval. Although this effect was present in both age groups, it was more pronounced for older adults. Also as in Experiment 1, episodic drift rates were lower than semantic drift rates in both age groups. Although the Age  $\times$  Retrieval Type interaction on drift rate was not significant, additional planned comparisons did replicate the Experiment 1 finding of a specific age-related decline in episodic drift rate but not semantic drift rate, lending support to the context-memory hypothesis (e.g., Light, 1996; 2000b).

Other results deviated from those of Experiment 1. First, there were effects of age and retrieval type on decision boundaries. Specifically, older adults set more conservative boundaries than younger adults, similar to previous reports by Ratcliff et al. (Ratcliff, Thapar, Gomez, & McKoon, 2004; Ratcliff et al., 2001; Ratcliff, Thapar, & McKoon, 2004; Thapar et al., 2003). Second, boundary settings were more conservative for episodic retrieval than for semantic retrieval in both age groups, especially in older adults. These between-groups and within-group differences in boundary settings may reflect adaptive responses to differences in task difficulty (see also Spaniol & Bayen, 2005; Thapar et al., 2003).

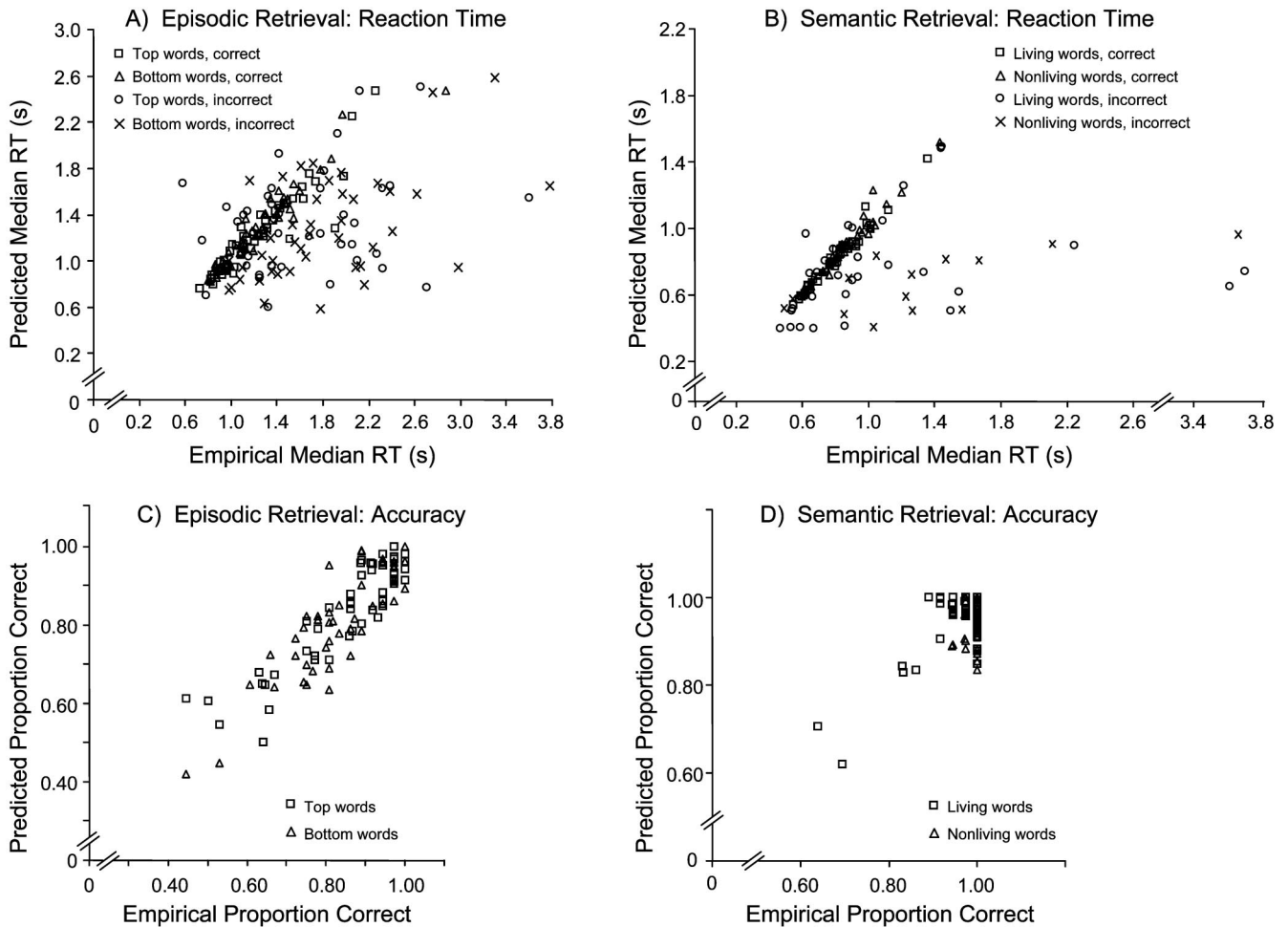


Figure 7. Empirical and predicted median reaction time (RT) and accuracy in Experiment 2. Predicted values were based on the diffusion model as described in the Model Fit section of the Results section for Experiment 2. A: RT values for episodic retrieval; B: RT values for semantic retrieval; C: accuracy values for episodic retrieval; D: accuracy values for semantic retrieval.

Age differences in the variability of the nondecisional RT component and the starting point of the diffusion process (parameters  $s_1$  and  $s_2$ ) indicate that these processes were noisier in older adults, especially for episodic retrieval. It is interesting to note that we did not replicate the age-related increase in drift rate variability (parameter  $s_3$ ) observed in Experiment 1.

### General Discussion

A fundamental problem in comparing the properties of episodic and semantic memory retrieval is that these two types of retrieval are typically assessed with different metrics: accuracy in the case of episodic retrieval and RT in the case of semantic retrieval (but see McKoon & Ratcliff, 1979; McKoon, Ratcliff, & Dell, 1986). This lack of a common metric has made it difficult to establish definitively that age-related decline in memory performance is specific to the context-dependent (episodic) domain. We addressed these issues in two experiments by using parameter estimates derived from the diffusion model of RT and accuracy (Ratcliff,

1978), which provided a common metric for a direct comparison of age-related changes in episodic and semantic memory retrieval. Using an estimation method introduced by Voss et al. (2004), we obtained good quantitative model fits. It is consequently possible to apply the diffusion model to data acquired in these more representative testing conditions, which do not feature extended task practice or speed-accuracy trade-off manipulations. This result has important methodological implications, demonstrating the feasibility and benefit of an explicit modeling approach to study population differences in cognitive mechanisms underlying two-choice decisions. However, it must be noted that, given the relatively high accuracy levels in both experiments, error RTs were systematically underestimated. If precise quantitative fitting of error RT distributions is a priority, it may be necessary to use task conditions that produce higher error rates.

On the basis of previous reports by Ratcliff et al. (Ratcliff, Thapar, Gomez, & McKoon, 2004; Ratcliff et al., 2001, 2003; Ratcliff, Thapar, & McKoon, 2004; Thapar et al., 2003), we

predicted that older adults would show longer nondecisional processing times and demonstrate greater conservatism than younger adults. In line with the context-memory deficit hypothesis, we also predicted that older adults would show lower drift rates than younger adults in episodic retrieval tasks (old–new recognition and source memory) but not in semantic retrieval tasks. Our data lent support to each of these predictions.

#### *Age-Related Increase in Nondecisional RT and Conservatism*

In Experiment 1, the age-related increase in the parameter estimating the nondecisional RT component (i.e., stimulus encoding, response preparation and execution) accounted for approximately 90 ms of the age difference in overall RT, whereas in Experiment 2, it accounted for approximately 225 ms of the age difference in RT for episodic retrieval and 130 ms of the age difference in semantic retrieval. Similar age differences on the  $t_0$  parameter have been reported elsewhere, although  $t_0$  values were lower than in the present study (e.g., Ratcliff, Thapar, & McKoon, 2004). It is interesting to note that in both current experiments, parameter  $t_0$  was also affected by retrieval type in both age groups. On average, the nondecisional RT component for episodic judgments exceeded that for semantic judgments by 56 ms in Experiment 1, by 145 ms for younger adults in Experiment 2, and by 240 ms for older adults in Experiment 2. The sensitivity of the nondecisional RT component to retrieval type may represent an adaptive response to perceived changes in task difficulty (i.e., a slowing of motor operations during the more difficult episodic tasks). Two subtly different interpretations are possible. According to the first, the presence of retrieval type effects on  $t_0$  indicates that this parameter fails to isolate purely nondecisional RT components and that the model formulation should be changed accordingly. However, this conclusion would contradict findings by Ratcliff et al. (see Introduction for a brief review) from experiments with speed–accuracy manipulations that strongly suggest that  $t_0$  is a nondecisional RT component. Alternatively, our finding may suggest that  $t_0$  is a valid measure of nondecisional RT but that  $t_0$  can be regulated by decisional processes, such as drift rate and boundary settings, which reflect the primary impact of changes in task difficulty.

The boundary separation parameter  $a$  showed no effects of age or retrieval type in Experiment 1. This was somewhat unexpected, given Ratcliff, Thapar, and McKoon's (2004) finding of an age-related increase in conservatism for old–new recognition. However, because of the procedural differences between our experiment and that of Ratcliff, Thapar, and McKoon, the discrepancy is difficult to interpret. In Experiment 2, boundary separation was greater for older adults than for younger adults, indicating an age-related increase in conservatism similar to that described previously (e.g., Ratcliff, Thapar, & McKoon, 2004). If this age-related increase in conservatism had strictly resulted from an adaptive response by older adults to relatively low task performance (e.g., Spaniol & Bayen, 2005; Thapar et al., 2003), then it should have been more pronounced for episodic retrieval than for semantic retrieval. This was not the case, although there was a main effect of retrieval type on boundary separation, which indicated that both age groups set more conservative decision boundaries during episodic retrieval than during semantic retrieval. To-

gether, these findings indicate that (a) older adults exhibit an increase in conservatism during two-choice memory decisions, at least when these decisions depend critically on the retrieval of context information, and (b) the effects of age and task difficulty on conservatism appear to be additive. This age-related change may represent a response on the part of older adults to decline in relatively early perceptual processes. Although the duration of the test words was well above perceptual threshold, age-related decline exists in the efficiency of visual feature extraction even under these resource-limited conditions (Schneider & Pichora-Fuller, 2000; Scialfa, 2002). Older adults may differentially emphasize certain aspects of task performance, in this case the decision criterion, to compensate for this decline in bottom-up visual processing (Madden, Whiting, Spaniol, & Bucur, 2005). This type of compensatory mechanism could explain why an age-related increase in conservatism has been observed even in tasks that make minimal or no demands on memory retrieval (see Ratcliff et al., 2001, 2003; Ratcliff, Thapar, Gomez, & McKoon, 2004; Thapar et al., 2003).

#### *Age Differences in Episodic Drift Parameters*

As predicted, episodic drift rates were lower for older adults than for younger adults in both experiments, whereas semantic drift rates did not differ as a function of age. To our knowledge, this finding is the first demonstration of age-related differences in episodic drift rates. Ratcliff, Thapar, and McKoon (2004) reported no significant age difference in drift for old–new recognition, despite a statistical trend in the same direction. As we suggested earlier, it is possible that Ratcliff, Thapar, and McKoon's methodology, which featured an instructional manipulation (speed vs. accuracy) and multiple experimental sessions (including at least one session devoted entirely to task practice), helped minimize age differences in drift. Ratcliff et al.'s study also featured no retention interval, shorter per-item study times, and different encoding instructions (studying for a later test, rather than making pleasantness judgments). Ratcliff, Thapar, and McKoon's approach is ideal for obtaining stable RT performance across different accuracy levels, thus creating optimal conditions for model estimation and for testing the viability of the diffusion model in the context of long-term memory judgments across different participant populations. Our experimental design, however, may have been more representative of other published studies in the literature on memory and aging, and the parametric age differences we have identified may therefore generalize more easily to other findings in the existing literature. That said, we believe that additional work is needed to identify the conditions that determine the presence of age differences in episodic drift rates.

Our finding is consistent with the hypothesis of an age-related context-memory deficit (e.g., Light, 1996, 2000b; Spencer & Raz, 1995). According to this hypothesis, the ability to process contextual detail—crucial for source memory and old–new recognition, but not for semantic judgments—is selectively impaired in old age. Several caveats are in order. First, neither of our episodic retrieval tasks provided “pure” measures of context memory. Both old–new recognition and our two-choice context-memory task also involve item memory, even though we minimized demands on item memory in Experiment 2 by presenting only old items. Nevertheless, it

is possible that the age differences in episodic drift rates may partly reflect an age-related decline in memory for items.

Second, although our results indicate an age-related decline in the quality of the contextual information driving the decision process during retrieval, it is not clear whether this effect originated from an age-related deficit at encoding, storage/forgetting, retrieval, or from some combination of these. Glisky et al. (2001), for example, presented data favoring an encoding account of source memory deficits in older adults with low scores on neuropsychological tests of frontal lobe function. These authors, however, acknowledged that retrieval effects may also have contributed to low source memory performance in their "low-frontal" older adult group. One strategy to investigate this question may be to manipulate encoding quality within subjects (e.g., using a levels-of-processing approach) and to assess effects on episodic drift rates in younger and older adults. However, the issue may be difficult to resolve conclusively, because any encoding manipulations may also have direct effects on retrieval.

Finally, given the current interest in dual-process (i.e., recollection vs. familiarity) accounts of episodic memory (e.g., Jacoby, 1999; Yonelinas, 2001) and in effects of aging on these different types of retrieval processes (e.g., Light, 2000a), it is important to point out that our conclusions do not hinge on either single- or dual-process assumptions. The information that drives the decision process can be either recollective or familiarity-based. Our data may well reflect a mixture of these processes, especially in Experiment 2 (source memory). Separating recollection and familiarity with experimental methods and studying their retrieval properties with the diffusion model is an interesting challenge for future research.

## References

- Anderson, N. D., Craik, F. I. M., & Naveh-Benjamin, M. (1998). The attentional demands of encoding and retrieval in younger and older adults: I. Evidence from divided attention costs. *Psychology and Aging, 13*, 405–423.
- Balota, D. A., Dolan, P. O., & Duchek, J. M. (2000). Memory changes in healthy older adults. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 395–408). Oxford, United Kingdom: Oxford University Press.
- Balota, D. A., Duchek, J. M., & Paullin, R. (1989). Age-related differences in the impact of spacing, lag, and retention interval. *Psychology and Aging, 4*, 3–9.
- Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436.
- Braver, T. S., Barch, D. M., Keys, B. A., Carter, C. S., Cohen, J. D., Kaye, J. A., et al. (2001). Context processing in older adults: Evidence for a theory relating cognitive control to neurobiology in healthy aging. *Journal of Experimental Psychology: General, 130*, 746–763.
- Burke, D. M., & Light, L. L. (1981). Memory and aging: The role of retrieval processes. *Psychological Bulletin, 90*, 513–546.
- Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive capabilities: Mechanisms and performances* (pp. 409–422). Amsterdam: Elsevier.
- Craik, F. I. M. (1994). Memory changes in normal aging. *Current Directions in Psychological Science, 3*, 155–158.
- Dosher, B. A., & Rosedale, G. (1991). Judgments of semantic and episodic relatedness: Common time-course and failure of segregation. *Journal of Memory and Language, 30*, 125–160.
- Dvorine, I. (1963). *Dvorine pseudo-isochromatic plates* (2nd ed.). New York: Harcourt.
- Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review, 62*, 145–154.
- Glisky, E. L., Rubin, S. R., & Davidson, P. S. R. (2001). Source memory in older adults: An encoding or retrieval problem? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 1131–1146.
- Healy, M. J. R., & Light, L. L. (2004, April). *Recognition and lexical decision retrieval dynamics in young and older adults*. Paper presented at the 10th Cognitive Aging Conference, Atlanta, GA.
- Hertzog, C., Dixon, R. A., Hultsch, D. F., & MacDonald, S. W. (2003). Latent change models of adult cognition: Are changes in processing speed and working memory associated with changes in episodic memory? *Psychology and Aging, 18*, 755–769.
- Hintzman, D. L., & Curran, T. (1997). Comparing retrieval dynamics in recognition memory and lexical decision. *Journal of Experimental Psychology: General, 126*, 228–247.
- Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46*, 269–299.
- Howard, M. W., Kahana, M. J., & Wingfield, A. (in press). Aging and contextual binding: Modeling recency and lag-recency effects with the temporal context model. *Psychonomic Bulletin and Review*.
- Jacoby, L. L. (1999). Ironic effects of repetition: Measuring age-related differences in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 3–22.
- Kahana, M. J., Howard, M. W., Zaromb, F., & Wingfield, A. (2002). Age dissociates recency and lag recency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 530–540.
- Kliegl, R., & Lindenberger, U. (1993). Modeling intrusions and correct recall in episodic memory: Adult age differences in encoding of list context. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*, 617–637.
- Kolmogorov, A. (1941). Confidence limits for an unknown distribution function. *Annals of Mathematical Statistics, 12*, 461–463.
- Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- Laver, G. D. (2000). A speed-accuracy analysis of word recognition in young and older adults. *Psychology and Aging, 15*, 705–709.
- Light, L. L. (1996). Memory and aging. In E. L. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 443–490). San Diego, CA: Academic Press.
- Light, L. L. (2000a). Dual-process theories of memory in old age. In T. J. Perfect & E. A. Maylor (Eds.), *Models of cognitive aging* (pp. 238–300). Oxford, United Kingdom: Oxford University Press.
- Light, L. L. (2000b). Memory changes in adulthood. In S. H. Qualls & N. Abeles (Eds.), *Psychology and the aging revolution: How we adapt to a longer life* (pp. 73–97). Washington, DC: American Psychological Association.
- Madden, D. J., Whiting, W. L., Spaniol, J., & Bucur, B. (2005). Adult age differences in the implicit and explicit components of top-down attentional guidance during visual search. *Psychology and Aging, 20*, 317–329.
- McKoon, G., & Ratcliff, R. (1979). Priming in episodic and semantic memory. *Journal of Verbal Learning and Verbal Behavior, 18*, 463–480.
- McKoon, G., Ratcliff, R., & Dell, G. (1986). A critical evaluation of the semantic/episodic distinction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 295–306.
- Muter, P. (1978). Recognition failure of recallable words in semantic memory. *Memory & Cognition, 6*, 9–12.
- 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.

- Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, 7, 308–313.
- Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41–82). Hillsdale, NJ: Erlbaum.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, 10, 437–442.
- Rabinowitz, J. C., & Ackerman, B. P. (1982). General encoding of episodic events by elderly adults. In F. I. M. Craik & S. E. Trehub (Eds.), *Aging and cognitive processes* (pp. 145–154). New York: Plenum Press.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59–108.
- Ratcliff, R. (1985). Theoretical interpretations of speed and accuracy of positive and negative responses. *Psychological Review*, 92, 212–225.
- Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling the accumulation of partial information. *Psychological Review*, 95, 238–255.
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9, 347–356.
- 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 & 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., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, 9, 438–481.
- Ratcliff, R., Van Zandt, T., & McKoon, G. (1999). Connectionist and diffusion models of reaction time. *Psychological Review*, 106, 261–300.
- Salthouse, T. A. (1988). Resource-reduction interpretations of cognitive aging. *Developmental Review*, 8, 238–272.
- Salthouse, T. A. (1992). What do adult age differences in the digit symbol substitution test reflect? *Journal of Gerontology: Psychological Sciences*, 47, 121–128.
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103, 403–428.
- Salthouse, T. A., & Madden, D. J. (in press). Information processing speed and aging. In J. DeLuca & J. Kalmar (Eds.), *Information processing speed in clinical populations*. New York: Psychology Press.
- Santee, J. L., & Egeth, H. E. (1982). Do reaction time and accuracy measure the same aspects of letter recognition? *Journal of Experimental Psychology: Human Perception and Performance*, 8, 489–501.
- Schneider, B. A., & Pichora-Fuller, M. K. (2000). Implication of perceptual deterioration for cognitive aging research. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 155–219). Mahwah, NJ: Erlbaum.
- Scialfa, C. T. (2002). The role of sensory factors in cognitive aging research. *Canadian Journal of Experimental Psychology*, 56, 153–163.
- Spaniol, J., & Bayen, U. J. (2004). Formal modeling in research on episodic memory and aging. *Psychology Science*, 46, 477–513.
- Spaniol, J., & Bayen, U. J. (2005). Aging and conditional probability judgments: A global matching approach. *Psychology and Aging*, 20, 165–181.
- Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: A meta-analysis. *Psychology and Aging*, 10, 527–539.
- 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.
- Veiel, L. L., & Storandt, M. (2003). Processing costs of semantic and episodic retrieval in younger and older adults. *Aging, Neuropsychology, and Cognition*, 10, 61–73.
- Verhaeghen, P. (2000). The parallels in beauty's brow: Time-accuracy functions and their implications for cognitive aging theories. In T. J. Perfect & E. A. Maylor (Eds.), *Models of cognitive aging* (pp. 50–86). Oxford, United Kingdom: Oxford University Press.
- Verhaeghen, P., Marcoen, A., & Goossens, L. (1993). Facts and fiction about memory aging: A quantitative integration of research findings. *Journal of Gerontology*, 48, P157–171.
- Verhaeghen, P., & Salthouse, T. A. (1997). Meta-analyses of age-cognition relations in adulthood: Estimates of linear and nonlinear age effects and structural models. *Psychological Bulletin*, 122, 231–249.
- Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition*, 32, 1206–1220.
- Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised*. New York: Psychological Corporation.
- Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, 41, 67–85.
- Yonelinas, A. P. (2001). Consciousness, control, and confidence: The 3 Cs of recognition memory. *Journal of Experimental Psychology: General*, 130, 361–379.
- Zacks, R. T., Hasher, L., & Li, K. Z. H. (2000). Human memory. In T. A. Salthouse & F. I. M. Craik (Eds.), *The handbook of aging and cognition* (2nd ed., pp. 293–357). Mahwah, NJ: Erlbaum.

Received March 8, 2005

Revision received September 28, 2005

Accepted October 6, 2005 ■