Journal of Memory and Language **46,** 199–226 (2002)
doi:10.1006/jmla.2001.2803, available online at http://www.academicpress.com on IDEAL®

The Word-Frequency Mirror Effect in Young, Old, and Early-Stage Alzheimer's Disease: Evidence for Two Processes in Episodic Recognition Performance

David A. Balota, Gregory C. Burgess, Michael J. Cortese, and David R. Adams

Washington University

Two experiments address the nature of the word-frequency mirror effect in episodic recognition performance and the underlying cognitive changes that occur in both healthy aging and in early-stage Dementia of the Alzheimer's Type (DAT). In Experiment 1, five groups of participants (young, healthy old, healthy old-old, very mildly demented individuals, and mildly demented individuals) studied lists of high- and low-frequency words and were given a yes/no episodic recognition test. The results indicated that there was a dramatic decrease in hit rate for low-frequency words across age and DAT, but no decrease for high-frequency words, thereby eliminating the low-frequency advantage typically found in recognition performance for the DAT individuals. For the distractor items, there was a clear advantage in rejecting low-frequency words compared to high-frequency words, and the size of this advantage was constant across groups of participants. This between-group pattern was replicated in a second experiment, in which only young adults were required to respond either under short or long response deadlines. The results are discussed with respect to an attentional control framework in which cognitively impaired groups of participants, and young adults at a short response deadline, rely more on baseline familiarity processes than on recollection-based processes. Discussion focuses on the nature of the recollection- and familiarity-based processes.

Key Words: aging; mirror effect; recognition memory; word frequency; two-process model.

Individuals with Dementia of the Alzheimers Type (DAT) produce breakdowns in a wide variety of memory tasks. Indeed, the major diagnostic criterion used by most clinicians is a deterioration in declarative/episodic memory performance (e.g., Nebes, 1992). However, the memory breakdowns are not limited to declarative/episodic memory tasks. There is also some evidence suggesting that there are breakdowns in semantic memory knowledge (e.g., Butters et al., 1987) along with impairment in procedural learning tasks in the mild form of dementia (see, e.g., Ferraro, Balota, & Conner, 1993).

Although there are a wide variety of memory changes in DAT, it is also clear that some processes appear to be relatively intact, at least early in the disease process. For example, these individuals appear to produce normal levels of repetition priming (see Balota & Duchek, 1991; Balota & Ferraro, 1996; Gabrieli et al., 1999). Interestingly, both Ferraro et al. and Gabrieli et al. have provided evidence that one will find breakdowns in early-stage DAT individuals when the implicit memory tasks place a relatively high demand on attentional systems. This attentional account is consistent with accumulating evidence that DAT individuals produce breakdowns in the attention demanding aspects of Stroop (e.g., Fisher, Freed, & Corkin, 1990; Spieler, Balota & Faust, 1996), reading with distraction (Duchek, Balota, & Thessing, 1998), spatial attention tasks (Simone & Baylis, 1997a, 1997b), and homograph disambiguation (Balota & Duchek, 1991; Faust et al., 1997). The role of attentional processes is also relevant to the changes that have been found in semantic memory tasks. For example, the large changes that are found in category fluency and Boston naming tasks in early-stage DAT individuals may be due to the attentional demands of the tasks instead of

This work was supported by Grant NIA PO1 AGO3991. Thanks are extended to the clinical core at the Washington University ADRC, who provided the well-specified population for this study; Martha Storandt, who provided the psychometric information; and Larry Jacoby, Steve Joordens, James Nairne, and Susan Sergent-Marshall, who provided useful comments on a previous version of this article.

Address correspondence and reprint requests to David A. Balota, Department of Psychology, One Brookings Hall, Washington University, St. Louis, MO 63130. E-mail: dbalota@artsci.wustl.edu.

0749-596X/01 $35.00

the integrity of the semantic memory structure *per se* (see, e.g., Balota, Watson, Duchek, & Ferraro, 1999; Nebes, Martin, & Horn, 1984; Ober & Shenaut, 1995a, 1995b). This is supported by evidence which suggests that when one minimizes the attentional demands of semantic tasks (e.g., via semantic priming procedures) very mild and mildly demented individuals produce relatively normal levels of semantic processing, even across varying levels of prime–target strengths (see Balota et al., 1999; Ober & Shenaut, 1995a, for a review).

Based on the above results, one might argue that the memory deficits observed in DAT individuals are accentuated by the attention demanding aspects of tasks, whereas more automatic operations appear to be relatively intact (see Balota & Faust, 2001; Jorm, 1986, for reviews). Similar arguments have been made regarding the cognitive changes in healthy older adults (e.g., Hasher & Zacks, 1988). It is in this light that the present experiments explore two interrelated issues: First, we attempt to take advantage of the memory profile exhibited in healthy aging and in DAT individuals to provide leverage on understanding one of the basic principles of human memory performance, referred to as the mirror effect in episodic recognition performance. Second, we use the mirror effect in episodic recognition performance to provide a better understanding of the underlying mechanisms that produce the observed episodic memory breakdowns in both healthy older adults and in individuals with early-stage DAT. We now turn to a brief discussion of the mirror effect and recent work regarding episodic memory performance in healthy older adults and in DAT individuals.

Mirror Effect in Episodic Recognition Performance

Glanzer and Adams (1985) coined the term "mirror effect" and argued that this pattern is so ubiquitous that it appears to reflect a basic principle of human episodic recognition performance. The mirror effect refers to the finding that "if there are two classes of stimuli, and one is more accurately recognized than the other, then the superior class is *both* more accurately recognized as old when old *and also* more accurately recognized as new when new" (Glanzer & Adams, 1990, p. 5). Consider, for example, the word-frequency effect in episodic recognition performance. Low-frequency words produce both a higher hit rate and a higher correct rejection rate than high-frequency words. The mirror pattern across hits and correct rejections is found across a number of quite distinct variables such as concreteness, list length, meaningfulness, pictures vs words, and other variables (see Glanzer & Adams, 1985, for a review).

Although there are a number of variables that produce the mirror effect in episodic recognition performance, we will primarily focus on word frequency in the present article. This finding is of interest for a number of distinct reasons: First, unlike many of the other variables that produce the mirror effect in recognition memory, the word-frequency effect is either eliminated (see Watkins, LeCompte, & Kim, 2000) or reversed in recall performance (e.g., Balota & Neely, 1980; Glanzer & Bowles, 1976). This dissociation between word-frequency effects as a function of retrieval task has provided a testbed for memory models. Second, although a number of models can accommodate the higher hit rate for low-frequency words compared to high-frequency words, it is more difficult to accommodate the lower false alarm rate to low-frequency words compared to high-frequency words (see Stretch & Wixted, 1998a, 1998b, for a discussion of this issue). Specifically, why should word frequency influence performance on items that were not presented during study?

Joordens and Hockley (2000) and Reder et al. (2000) have both recently developed two-process accounts of the word-frequency mirror effect. Although there are clear differences in their theoretical perspectives, there are also some important similarities. For example, both models suggest that for hit rates, there are opposing influences of a type of item-specific recollective process (which benefits low-frequency words) and a baseline familiarity process (which benefits high-frequency words). In the absence of any study episode, subjects rely primarily on baseline familiarity and so produce higher false alarms to high-frequency words compared to low-frequency words. On the other hand, because the increased recollection for low-frequency words typically outweighs any baseline advantage for high-frequency words, one finds higher hit rates for low-frequency words than high-frequency words. Both Joordens and Hockley and Reder et al. provide converging evidence for these two processes contributing to the mirror effect by using Tulving's (1985) remember/know distinction. The notion here is that remember responses are more likely to reflect recollective processes, whereas know responses are more likely to reflect familiarity based processes. Both Joordens and Hockley and Reder et al. found that low-frequency words provide more remember responses than high-frequency words, whereas high-frequency words provide more know responses than low-frequency words (also see Gardiner & Java, 1990; Strack & Forster, 1995, for a similar pattern). This was interpreted as support for a class of two-process models.

Given the importance of word-frequency and mirror effects in understanding episodic recognition performance, the present investigation explored this effect in healthy young, healthy older adults, and DAT individuals. Because of the specific cognitive changes that occur in both memory and attention in healthy older adults and in early stages of DAT, it is possible that one might find a qualitatively different pattern of mirror effects in these populations. Specifically, it is possible that older adults and early stage DAT individuals may be likely to rely more on baseline familiarity of the stimulus instead of recollective item specific information. This may produce more "yes" responses to both high-frequency targets *and* high-frequency distractors. Of course, based on the Joordens and Hockley (2000) and Reder et al. (2000) arguments regarding the relevance of remember and know judgments to the word-frequency mirror effect, one might also expect age (and DAT) to modulate the percentage of remember and know responses to high-and low-frequency words. Although we are unaware of any studies that have directly explored remember/know judgments in DAT individuals, it is interesting to note that there is evidence that in episodic recognition performance older adults are indeed more likely to produce "know" responses than "remember" responses compared to healthy younger adults (see Java, 1996; Mantyla, 1993; Parkin & Walter, 1992). Within the Joordens and Hockley and Reder et al. framework, this suggests that at least older adults may be more likely to rely on familiarity-based information to drive recognition judgments, thereby potentially modulating the hit rate advantage for low-frequency words. We now turn to a brief review of the role of familiarity in memory-impaired groups.

Familiarity Effects and Recognition Performance in Individuals with Episodic Memory Loss

Verfaellie and Cermak (1999) have noted that ". . . patients with amnesia often experience a sense of familiarity for recently encountered people or events. This feeling of familiarity is striking, because patients remain at a loss to describe where or when they encountered these people or experienced these events." In fact, there is some evidence that amnesics with focal lesions to the limbic system show a relative sparing of recognition memory (Aggleton & Shaw, 1996), suggesting that this familiarity may be used to support episodic recognition decisions (see, however, Haist, Shimamura, & Squire, 1992). Verfaellie and Cermak provided evidence which suggests that amnesics can use perceptual fluency to make episodic recognition judgements, and interestingly, they also provided evidence that these individuals rely more on perceptual fluency than healthy controls. Of course, it is likely that individuals with focal lesions to the limbic system may have intact cortical structures that provide useful access to perceptual fluency information that can also be used for making episodic recognition decisions. These may be subserved by nonlimbic cortical structures that engage initial processing of the stimulus, i.e., fluency operations (Jacoby, 1983a, 1983b). It is also possible that frontal cortical areas that may be more related to metacognitive reflection of the increase in familiarity may also be intact in these individuals (see, e.g., Moscovitch & Winocur, 1992). However, Parkin and Walter (1992) found that performance on the Wisconsin Card Sorting task (a task that has been viewed as reflecting primarily frontal operations) was negatively correlated with the relative incidence of "know" responses in episodic recognition performance in healthy older adults. In this light, it appears that decreases in frontal function in older adults may also contribute to an increase in the reliance on familiarity processes to drive recognition performance.

Turning to DAT individuals, the predictions would appear to be a bit more complex. In these individuals, there is rather widespread neuropathology in medial temporal, frontal, and parietal areas (see Kanne et al., 1998; Morris et al., 1996). Because of the widespread involvement of this disease, one might expect that DAT individuals will produce decreased perceptual priming that is critical for the buildup of familiarity information. However, as noted above, there is now accumulating evidence that when DAT individuals are presented implicit memory tasks that minimize attentional operations (simple repetition priming effects), there is relatively little breakdown in the implicit priming effects (e.g., Balota & Duchek, 1991; Gabrielli et al., 1999; Moscovitch, 1986). In one study particularly relevant to the present research, Balota and Ferraro (1996) investigated the repetition × word-frequency interaction in healthy young, older adults and in individuals with DAT in the lexical decision task—a task which has been shown to place a premium on familiarity information (e.g., Balota & Chumbley, 1985; Seidenberg & McClelland, 1989). The results of this study yielded a large long-term repetition effect for all groups of participants, which was larger for low-frequency words than for high-frequency words. Thus, even DAT individuals produced large repetition effects in a task that places a premium on familiarity-based information.

If indeed familiarity-based information is relatively intact in both healthy older adults and in early-stage DAT individuals (at least as reflected in the lexical decision task), but more attention demanding/recollective processes break down, what are the predictions regarding the word-frequency mirror effect in these individuals? It is quite possible that these individuals may be more likely to rely on baseline familiarity information and hence be more likely to respond old to high-frequency words than to low-frequency words for both targets and distractors. Interestingly, Wilson et al. (1983) provided a pattern that is consistent with this prediction in a group of DAT individuals. Specifically, they found that for old items there was a slight reversal of the word-frequency effect in DAT individuals, whereas for new items, there was the standard finding of more false alarms to high-frequency words than low-frequency words. Thus, DAT individuals appeared to be responding "old" based on baseline familiarity. In a second experiment, Wilson et al. tested only healthy control individuals either immediately or after a delay of 1 week. They reasoned that DAT individuals were possibly biased to respond old to high-frequency words simply because they were unable to retrieve list context information. If this were the case, then one might find a similar pattern in healthy older adults after such a delay, wherein list context information would be minimized. However, the results did not support this prediction. Specifically, healthy older adults showed a normal word-frequency effect even after a 1-week delay. Based on these results, Wilson et al. argued that it appears that DAT individuals do not engage in the appropriate encoding operations that produce a substantial boost for low-frequency words.

Although the Wilson et al. results are consistent with the notion that DAT individuals will not produce a word-frequency effect for old items, there are a number of issues that still need to be addressed. First, the interpretation of their second experiment that varied retention interval with healthy control individuals is constrained by the fact that the lure items did not show a word-frequency effect. Wilson et al. acknowledged the difficulty posed by this failure to find a word-frequency effect for the lure items. Second, and more importantly, because Wilson et al. included only a single group of healthy older ($N = 20$) individuals and a single group of DAT ($N = 17$) individuals, it is unclear whether this reduction of the word-frequency effect for old items will vary across different levels of healthy aging and different levels of dementia severity. Of course, it is possible that one might only find this pattern in relatively severe amnesic individuals (also see Huppert & Piercy, 1976, for a similar pattern in Korsakoff patients). In fact, the mean hit rate minus false alarm rate for the DAT individuals in the Wilson et al. study was very low, i.e., .20 compared to the healthy controls at .65.

In the first experiment, we included five groups of participants. This allowed us to independently investigate the influence of age and the influence of severity of DAT with three groups per comparison. Specifically, for the age comparison we compared healthy young adults (mean age of 19.4 years), healthy young-old adults (mean age of 71.4 years), and healthy old-old adults (mean age of 85.0 years). Turning to the comparison across DAT levels, we compared healthy age-matched control individuals with very mild DAT individuals and mild DAT individuals. In order to equate age across groups, we included both the healthy young old and the healthy old-old participants in the control group. Hence, in the DAT analysis, the healthy control group had a mean age of 76.4 years, the very mild DAT individuals had a mean age of 77.6 years, and the mild DAT individuals had a mean age of 77.6 years.

In addition to providing estimates of accuracy on the recognition test, we also included estimates of processing speed during the recognition test. The notion is that if across our groups of participants there is an increased reliance on more automatic familiarity based processes, then one might find response latency to also produce the predicted change in latencies for hit rates, but not for false alarm rates. Of course, without a measure of response latency, it is also possible that changes in accuracy may be accompanied by opposing influences on response latency across groups of participants. Thus, the response latency measures provide an important converging piece of evidence regarding differences across groups in episodic recognition performance.

EXPERIMENT 1

Methods

Participants. Four groups of older adults and one group of young adults participated in this study. The young group included 28 undergraduates at Washington University who were either paid or received course credit for their participation. The mean age of the young group was 19.4 years, with a range of 18 to 22 years. The older adults included both healthy older adults and older adults with DAT who were recruited from the participant pool at the Alzheimer's disease Research Center at Washington University. There were two groups of healthy older adults, a younger old group (mean age of 71.4 years, with a range of 63 to 79 years) and an older old group (mean age of 85.0 years, with a range of 80 to 92 years). These participants were screened for depression, severe hypertension, possible reversible dementias, and other disorders which could affect cognitive performance. There were 43 participants in the younger old group and 25 participants in the older old group.

There were two groups of older adults with DAT. The severity of the dementia was scaled according to the Clinical Dementia Rating (CDR) scale developed at Washington University (Berg, 1988; Hughes et al., 1982). The first was a group classified with very mild DAT (CDR = 0.5). There were 35 participants in this group. The mean age of this group was 77.7 years, ranging from 61 to 91 years. The second group of DAT individuals was classified with mild DAT (CDR = 1.0). This group included 17 participants, with a mean age of 77.6 years and a range from 56 to 90 years. DAT individuals were included or excluded based on criteria set by the National Institute of Neurological and Communicative Disorders and Stroke and Alzheimer's disease and Related Disorders Association (McKhann et al., 1984). The diagnosis of Alzheimer's disease by the clinical core at Washington University has been excellent (93% diagnosis accuracy confirmed at autopsy) and well documented (e.g., Berg et al., 1998).

Psychometric performance. All of the healthy older adults and the DAT individuals participated in an independent 2-h battery of psycho-

TABLE 1

Psychometric Means (*SD*) as a Function of Group

	CDR 0 (Young-Old)	CDR 0 (Old-Old)	CDR 0.5 (Very Mild DAT)	CDR 1.0 (Mild DAT)
WAIS Information	22.6 (3.8)	21.1 (3.7)	17.9 (5.4)	12.9 (5.3)
Boston Naming	56.7 (4.0)	52.9 (7.5)	51.5 (7.6)	40.8 (12.6)
Logical Memory	9.8 (2.7)	8.5 (3.1)	6.0 (3.7)	2.2 (1.5)
Associate Memory	14.4 (3.8)	13.3 (4.2)	11.0 (3.4)	7.3 (2.3)
Benton Copy	9.8 (0.6)	9.6 (0.8)	9.4 (1.0)	7.9 (2.5)
Trailmaking A[a]	37.7 (11.5)	46.9 (15.9)	45.5 (21.1)	83.4 (42.5)
Block Design	32.8 (8.0)	29.9 (6.7)	26.2 (8.1)	16.1 (10.0)
Digit Symbol	49.3 (9.6)	38.9 (8.8)	39.4 (12.4)	27.1 (12.1)
Digit Span	11.6 (2.2)	11.4 (2.3)	10.7 (2.1)	9.4 (2.1)
Word Fluency	29.3 (10.8)	28.1 (11.3)	26.9 (11.4)	18.1 (10.3)
Mental Control	7.5 (1.7)	7.6 (1.8)	7.5 (1.8)	5.8 (2.4)

[a] For this task, lower scores reflect better performance.

metric tests designed to assess cognitive functioning in a number of distinct domains. Memory was assessed with the Wechsler Memory Scale (WMS; Wechsler & Stone, 1973), Associates Recall and Recognition subscales (paired-associates learning), Logical Memory subscale (surface-level story memory), and forward and backward Digit Span from the WMS. Lexical retrieval processes were assessed by the Word Fluency test, which required participants to name as many words as possible beginning with a specified letter (e.g., P or S) in a 60-s time period (Thurstone & Thurstone, 1949), and the Boston Naming Test (Goodglass & Kaplan, 1983), in which participants named line drawings. Measures of general intelligence were the Information, Block Design, and Digit Symbol subtests of the WAIS-R (Wechsler, 1955). The Benton Copy Test and Trail Making Form A assessed visual perceptual motor performance. In the Benton Copy test, participants copy a geometric figure; in the Trail Making Form A test, participants connect numerically ordered dots that result in a specific pattern (Armitage, 1946). Participants also received the WMS Mental Control test, which evaluates the ability to produce quickly a well-rehearsed letter or digit sequence, such as the alphabet, in a specified amount of time. The results from the psychometric battery are displayed in Table 1. As expected, all measures produced main effects of group (all $ps < .01$). However, it should also be noted that the largest breakdown in cognitive performance in these tasks occurs in the Mildly Demented Groups. Although there was the expected pattern of psychometric performance among the remaining groups, when the Mildly Demented Group was excluded from the analyses, the Benton Copy, Digit Span, Word Fluency, and Mental Control tasks did not reach significance.

Materials. Forty-eight high- and 48 low-frequency words were selected for this study (see the Appendix for complete list). All items, with the exception of SEQUIN and SLEUTH, were listed in the Kučera and Francis (1967) norms. The mean frequency of the low-frequency words was 2.18, and the mean frequency of the high-frequency words was 77.38. The words ranged from three to seven letters in length and were matched in length across word frequency. All words were concrete nouns with the exception of NATION. The words from these two lists were used to create study and test lists. The length of these lists was varied across groups in order to minimize ceiling and floor effects. For young and healthy older adults all 96 words were used. For participants with very mild DAT, 72 words were used (36 high and 36 low frequency). For the participants with mild DAT, 48 words were used (24 high and 24 low frequency). For those participant groups with list lengths less than the maximum, the selection of words included in the experiment was counterbalanced, thereby ensuring that there were no systematic item differences across list length.

Half of the high-frequency and half of the low-frequency words for a given list length were chosen to form the study list, and the remaining half of the items within each frequency band were used as lure items on the recognition test. The assignment of words to the study list was counterbalanced across participants such that each word occurred both as a study item and a lure item.

Procedure. Participants were tested individually. All instructions were given to the participants by an experimenter who was present with the participant during the testing session. Before the experiment began, participants were instructed that this was a memory experiment and that they would see a list of words presented on the computer screen one at a time. They were further instructed to read each word aloud and to try to remember the word for a later test. Participants were allowed to ask questions before the experiment began.

During the first part of the experiment, words from the study list were presented one at a time centered on a computer screen. The words were presented in a random order, which was determined separately for each participant. Participants read each word aloud. Upon the detection of voice onset, via a Gerbrands voice-key, a 2-s stimulus duration was initiated, after which the next stimulus word was presented.

The second part of the experiment was designed as a filler and consisted of 10 trials of an arithmetic verification task. In this task, a simple addition or subtraction problem was presented on the screen, along with the solution. Participants were instructed to indicate if the answer was correct or incorrect by pressing the right and left mouse buttons respectively. A 1-s intertrial interval separated the trials in this part of the experiment.

The final part of the experiment was the recognition test. Participants were instructed that they would see a list of words one at a time on the computer and that these words would either be words from the first part of the experiment or new words. They were instructed to press the right mouse button if they thought the word was presented in the first part of the experiment and to press the left mouse button if the word had not been presented. Because response latency was being measured, participants were encouraged to respond as fast as possible during the recognition test, but not at the cost of accuracy, which was the primary dependent measure. Each of the words from the test list was presented one at a time in a random order at the center of the computer screen. The word remained on the screen until the participant responded. The response initiated a 1-s intertrial interval.

Results

Mixed-factor analyses of variance were performed on hits, false alarms, d', C (e.g., Snodgrass & Corwin, 1988), and response latencies to examine the effects of word frequency and group on recognition performance. For each dependent measure, there were two ANOVAs conducted. One ANOVA included Age (young, young-old, and old-old) as the grouping factor, and the second ANOVA included DAT (healthy old, very mild DAT, and mild DAT) as the grouping factor. Unless noted, all effects are significant at $p < .05$.

Accuracy. The mean percentage of hits (top panel) and false alarms (bottom panel) are presented in Fig. 1.[1] There are three points to note in Fig. 1. First, with the exception of the mildly demented individuals, all groups produced the standard mirror effect, i.e., higher hit rate and lower false alarm rate for low-frequency words compared to high-frequency words. Second, as shown in the hit rate there is a dramatic decrease (37%) across participant groups for the low-frequency words, whereas for the high-frequency words, there is virtually no change in performance. Third, the change in the frequency effect across groups for hits is in sharp contrast to the pattern produced for the false alarms, wherein all groups appear to produce an equivalent benefit for rejecting low-frequency words compared to high-frequency words.

[1] Although it is customary to graph between group conditions with bar graphs, we have purposefully graphed these data in a line graph to highlight the apparent continuous change across age groups in cognitive performance. It is at least possible that some of the changes observed in the healthy older groups may reflect undetected very early stage Alzheimer's disease.

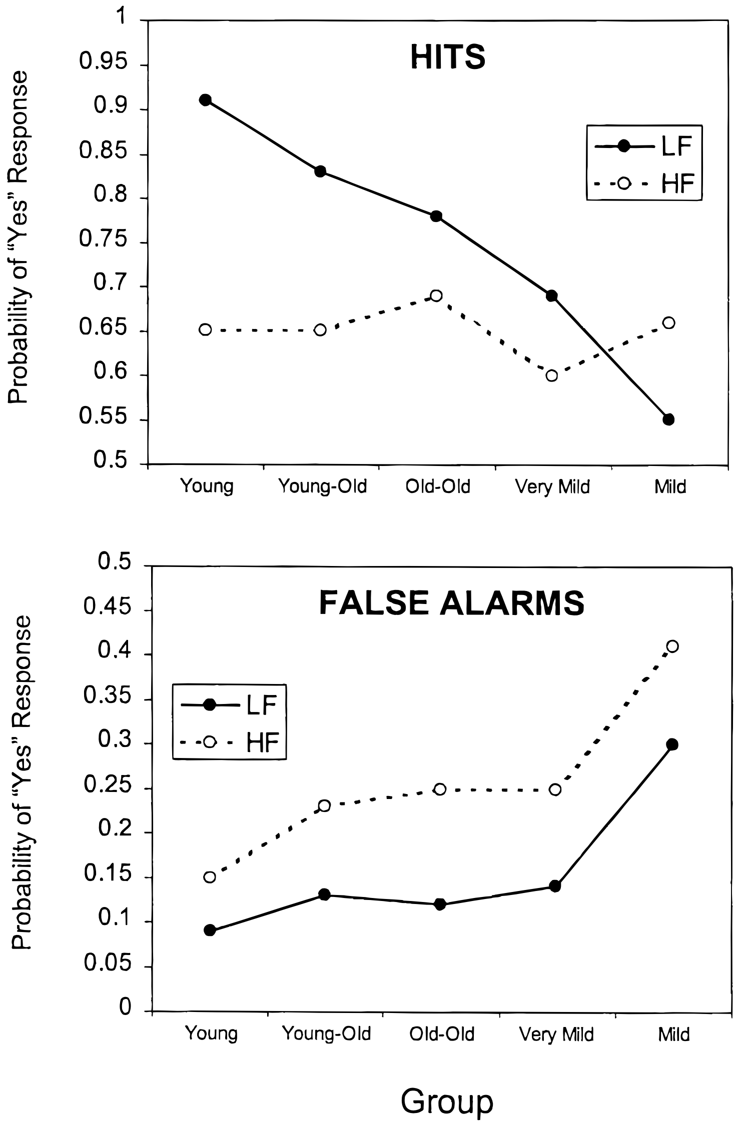

FIG. 1. Mean percentage hit and false alarm rate as a function of word frequency and group.

The above observations were supported by the ANOVAs. First, consider the changes in hit rates across groups. The hit rate reliably decreased across levels of DAT, $F(2,117) = 5.49$, $MSE = .34$, but not across age, $F < 1$. In addition there were more hits for low-frequency than high-frequency words in both the age analysis, $F(1,93) = 121.44$, $MSE = 1.39$, and the DAT analysis, $F(1,117) = 3.98$, $MSE = .07$. More importantly, the age × frequency interaction, $F(2,93) = 8.11$, $MSE = .09$, and the DAT × frequency interaction, $F(2,117) = 12.67$, $MSE = .22$, were both highly reliable in the analysis on hits. This reflects the fact that the frequency effect in hit rates decreased across age and dementia, with the eventual reversal of the frequency effect occurring for the mild DAT patients.

In the analyses of False Alarms, the false alarm rate increased across DAT, $F(2,117) =$

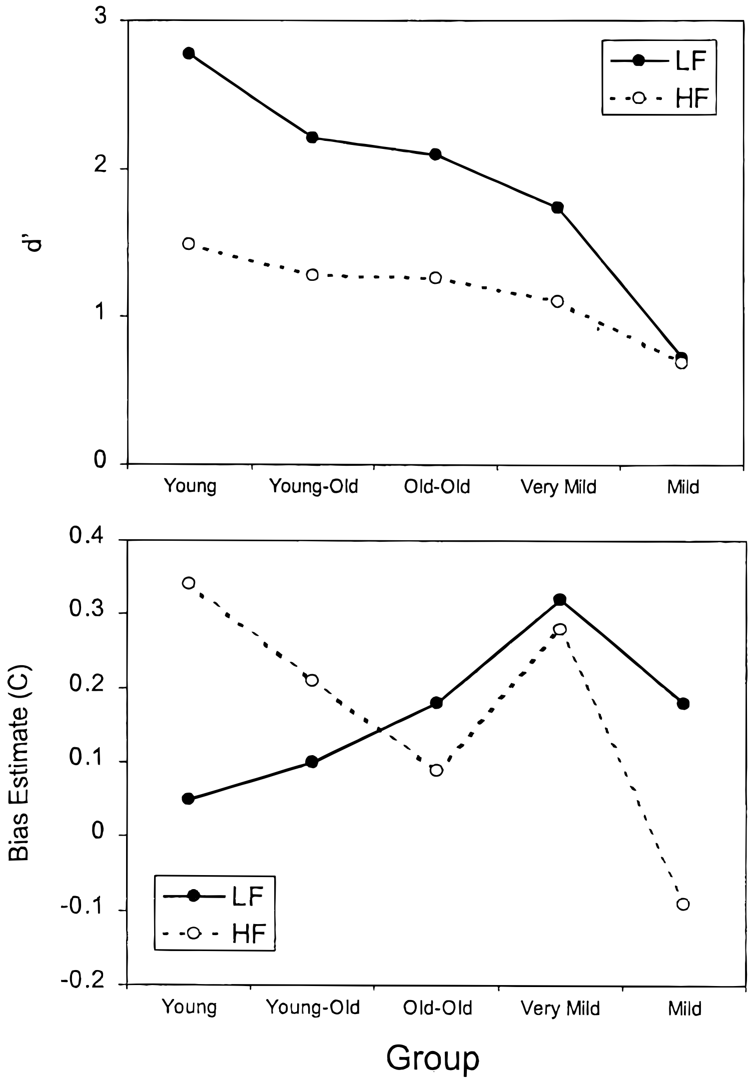

FIG. 2. Mean discriminability and bias estimates as a function of word frequency and group.

9.68, *MSE* = .43, and approached significance in the analysis with age, $F(2,93) = 2.39$, $MSE = .08$, $p < .10$. More false alarms were made to high-frequency words than low-frequency words in the analysis by age, $F(1,93) = 50.82$, $MSE = .45$, and DAT, $F(1,117) = 42.40$, $MSE = .53$. In contrast to the analyses on the hit rate, the frequency effect in false alarms was constant across groups as reflected by the lack of both an age × frequency, $F(2,93) = 1.62$, and a DAT × frequency, $F < 1$, interaction.

Signal detection analyses. A signal-detection analysis was also applied to the data to obtain measures of discriminability (d') and bias (C). As shown in Fig. 2, results from these analyses were generally consistent with the above analyses. Specifically, as shown in the top panel of Fig. 2, there is more of a decrease in discriminability for low-frequency words than high-frequency words across groups. Second, as shown in the bottom panel of Fig. 2, across groups, participants tend to be less conservative (lower

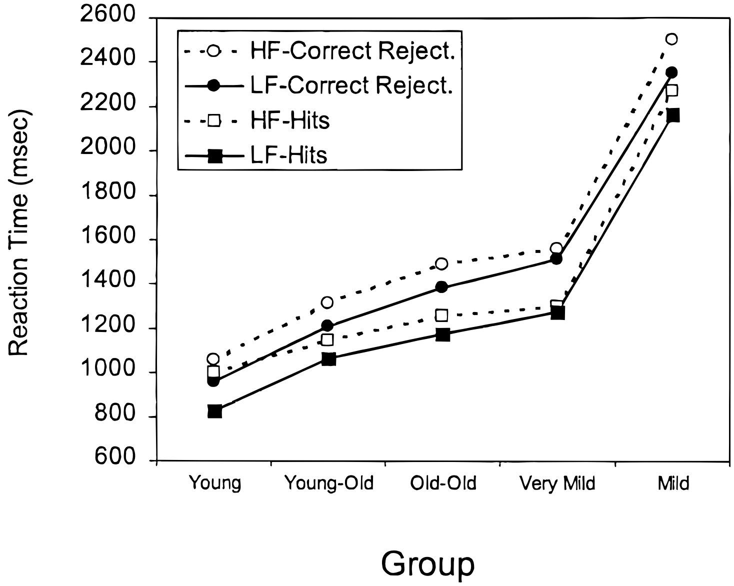

FIG. 3. Mean raw response latencies as a function of word frequency and group.

scores reflect less conservative) to high-frequency words as a function of age and DAT, whereas for low-frequency words there is a general tendency to become more conservative across groups.

Overall, d' decreased across age, $F(2,93) = 3.87$, $MSE = 3.42$, and DAT, $F(2,117) = 12.48$, $MSE = 1.15$. Discriminability was greater for low-frequency than high-frequency words in the analysis by age, $F(2,93) = 233.11$, $MSE = .20$, and by DAT, $F(2,117) = 48.84$, $MSE = .25$. Also, the interactions between age and frequency, $F(2,93) = 3.88$, $MSE = .20$, and DAT and frequency, $F(2,117) = 10.45$, $MSE = .25$, in the d' analyses were significant, indicating that the difference between low- and high-frequency words in discriminability decreased across age and DAT.

In the analyses of *C*, the bias estimate, the effect of Frequency was significant in the age analysis, $F(1,93) = 5.60$, $MSE = .08$. This was due to high-frequency words producing a less conservative bias than low-frequency words. The effect of Frequency also approached significance in the DAT analysis, $F(1,117) = 3.73$, $MSE = .35$, $p < .06$, which again reflected the less conservative bias for high-frequency words compared to low-frequency words. More importantly, the change in bias across groups for low- and high-frequency words was reflected in a significant age $\times$ frequency interaction, $F(2,93) = 5.92$, $MSE = .08$, which also approached significance in the DAT analyses, $F(2,117) = 2.98$, $MSE = .35$, $p < .06$.

Response latencies. Analyses were also performed on response latencies for correct responses during the recognition test. Extreme scores were eliminated first by removing any responses less than 250 ms or greater than 10,000 ms. Of the remaining observations, latencies that deviated by 2.5 standard deviations from each participant's overall mean were discarded. Overall, outliers accounted for 3% of the data. The data from one participant in the old-old group and two participants in the very mildly demented group were not included because there were no valid responses for at least one of the conditions.[2]

[2] To determine whether more responses were being discarded in certain conditions and across groups, we performed an analysis on the proportion of responses discarded. In this analysis, Subject Group included all five groups. The main effect of Group was significant $F(4,143) = 7.20$, $MSE = .04$. This was due to the higher outlier rate for the mildly de-

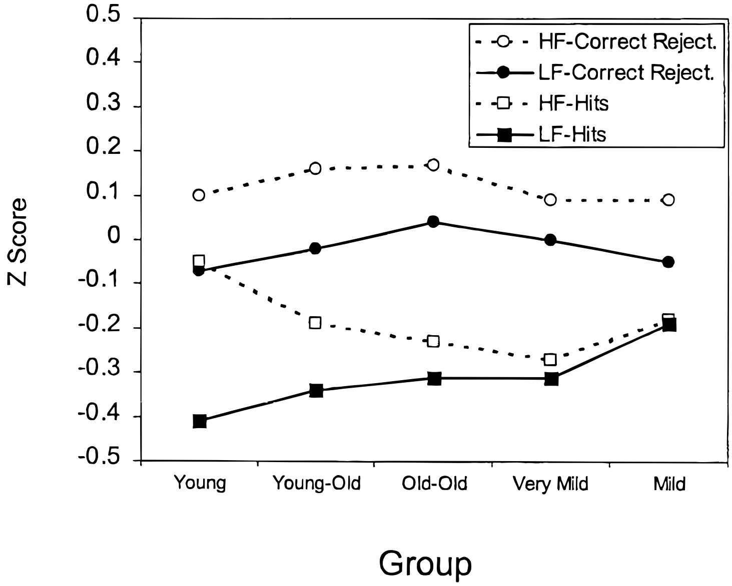

FIG. 4. Mean transformed *z* scores as a function of word frequency and group.

The mean response latency data are presented in Fig. 3. As shown here, there are three major points to note: First, there was a large increase in response latency for the mildly demented individuals. Second, high-frequency words produced slower response latencies compared to low-frequency words. Third, the distractor items were overall slower to reject than were the target items to accept. Of course, the last observation needs to be interpreted with some caution because the right button on the mouse was used to make the “old” response and the left key on the mouse was used to make the “new” response.

The results from the ANOVAs yielded main effects of age, $F(2,92) = 10.62$, $MSE = 3570188.1$, and DAT, $F(2,115) = 36.67$, $MSE = 879476.77$. In addition, the main effect of frequency was reliable in the ANOVA with age, $F(1,92) = 36.91$, $MSE = 26095.76$, and DAT, $F(1,115) = 12.25$, $MSE = 47650.49$. Participants responded in less time to old items than new items in the analysis by Age, $F(1,92) = 59.86$, $MSE = 37329.85$, and DAT, $F(1,115) = 33.28$, $MSE = 125414.96$. No other effects were significant.

Of course, the above analyses on raw response latencies may be a bit misleading because there are also considerable increases in variance across the groups of participants. In order to rescale the data so effect sizes are equivalent at different points of the RT scale, Faust, Balota, Spieler, and Ferraro (1999) have provided evidence that a *z* transformation of the data is useful. In this analysis, we converted each RT to a *z* score based on an individual’s overall grand mean and standard deviation. The *z* transformed data are presented in Fig. 4. These data are quite clear, and nicely converge on the accuracy data. First, as one can see the response latencies to targets were relatively faster than to

mented group (.07). More responses to old items were discarded than for new items, $F(1,143) = 20.33$, $MSE = .02$. Also, the interaction between Old/New and Frequency was significant, $F(1,143) = 14.40$, $MSE = .00$. This was due to a difference in outlier rates between low-frequency old and new items (.04 vs .02), whereas no difference was found between high-frequency old and new items (.03 vs .03). However, the mildly demented group had a higher rate for high-frequency old compared to new items (.08 vs .05), whereas the outlier rate was the same for low-frequency old and new items (.07 vs .07). This pattern produced a reliable Group × Frequency × Old/New interaction, $F(4,143) = 5.73$, $MSE = .00$. Of course, because the outlier rate was so low in these cells, with very little variance in some of the cells, one must be cautious not to over interpret the importance of these outlier differences.

lures. Second, for the lure items, as in the accuracy data, the response latencies to high-frequency new items were relatively slower than to low-frequency new items and this difference did not change across groups. On the other hand, for the old items, the response latencies were relatively faster for low-frequency words than for high-frequency words, and this word-frequency effect decreases across groups.

The above observations were supported by separate ANOVAs on the *z* scores, as a function of age and DAT. There was a main effect of old/new both in the analysis by Age, $F(1,92) = 62.43$, $MSE = .14$, and DAT, $F(1,115) = 47.44$, $MSE = .16$. Relatively faster responses were made to low-frequency words than to high-frequency words in the analysis by age, $F(1,92) = 64.67$, $MSE = 2.78$, and DAT, $F(1,115) = 13.95$, $MSE = .80$. The interaction between group and frequency was significant in the analysis by age, $F(2,92) = 4.24$, $MSE = .18$, due to the frequency effect decreasing across age groups, but this interaction was not significant in the DAT analysis, $F(2,115) = 1.08$. The group $\times$ frequency $\times$ old/new interaction was marginally significant in the analysis by age, $F(2,92) = 2.82$, $MSE = .04$, $p < .07$, but not in the analysis by DAT, $F < 1$. This marginally reliable interaction across age groups was due to a gradual reduction in the frequency effect for old items, whereas the frequency effect for new items remained constant.

Discussion

The results of Experiment 1 are quite clear. Specifically, as shown in Fig. 1, the effect of both age and DAT on the hit rate and correct rejection rate changed quite dramatically as a function of word frequency. In particular, there was virtually no change across groups in the word-frequency effect for the distractor items. This was reflected in both the accuracy data and the response latency data. On the other hand, there was a large change across groups in performance for the old items. In particular, the hit rate for the low-frequency words decreased dramatically across groups, whereas for the high-frequency words there was virtually no change in performance. A very similar pattern was found in the response latency data, which is based on *z* score transformed data to adjust for overall differences in speed and variance across groups.

These results are consistent with the Wilson et al. (1983) study described above and also extend this pattern to different age groups, different levels of dementia severity, and to response latency measures. It is very clear that low-frequency hits are disproportionately influenced by both Age and DAT compared to high-frequency hits. Moreover, there is no influence of either Age or DAT on the influence of word frequency on the distractor items.

Before turning to the theoretical interpretation of these results, we present the results from a second experiment. This experiment is based on the notion that one can mimic performance in healthy older adults in a group of healthy young adults by imposing a response deadline. An intriguing recent example of this approach is a study reported by Jacoby (1999). Jacoby found that putting young adults under a temporal deadline increased their reliance on familiarity based information, compared to recollection-based information, as reflected by an increased susceptibility to repetitions of distractor information. Moreover, the younger adult pattern under time pressure was quite similar to that of the older adults who were not under time pressure. Thus, Jacoby argued that by pushing young adults to respond quickly, there was increased reliance on familiarity-based information, just as found in older adults without time pressure.

In Experiment 2, we extend the Jacoby logic and use a relatively novel tempo response deadline approach to investigate the word-frequency mirror effect (also see Hintzman, Caulton, & Curran, 1994; Joordens & Hockley, 2000). The tempo response deadline approach is useful because it produces strong constraints on when in time the participant must respond to the cue based on both auditory and visual temporal cues (see Balota & Shields, 1988; Kello & Plaut, 1998). If the pattern of results from Experiment 1 is due to an increased reliance on baseline familiarity information to drive the response, and word frequency can modulate this familiarity, then one might actually find a similar dissociation in a group of young adults under short and long response deadlines. In particular, one should expect a normal word frequency mirror effect at the slow deadline, but an elimination of the word frequency mirror effect at the fast deadline.

EXPERIMENT 2

Method

Participants. Forty-eight undergraduates from Washington University in St. Louis participated in this experiment as partial fulfillment of an introductory psychology course requirement.

Materials. Four lists, each containing 48 words, were created for this experiment. Each list contained 24 high-frequency words (mean frequency of 237.8 occurrences per million) and 24 low-frequency words (mean frequency of 7.5 occurrences per million). The words were concrete nouns from four to seven letters in length. Word length was not reliably different across word frequency, $t(190) = 1.43$, $p = 0.15$. These words were counterbalanced across subjects so that each item appeared an equal number of times in each condition created by crossing tempo (fast vs slow), presentation (studied vs distractor), and block order (fast tempo block first vs slow tempo block first). Ninety-six medium-frequency words were selected for practice items.

Procedure. The experiment was composed of two blocks—one block with a slower 1000-ms response signal rhythm (or *tempo*) and one block with a faster 500-ms tempo. Each block consisted of a study phase, an arithmetic distractor task, and a yes/no recognition test. These tasks were preceded by three practice sets to give the subjects an opportunity to practice the tempo procedure. Each practice set was identical to the test phase, except that only 8 medium-frequency words were studied and 16 medium-frequency words were tested. In total, the participants received 48 practice trials at each tempo prior to the administration of the experimental trials for a given tempo block.

Subjects were instructed that words shown during the study phase would be used later in a yes/no recognition test. Forty-eight words were displayed individually in uppercase letters at the center of a computer screen. During the study phase, the following events occurred during each trial: (a) the study word was presented at the center of the screen for 2000 ms; (b) the stimulus was removed from the screen; and (c) a 500 ms blank intertrial interval was initiated.

In order to minimize any recency effects, subjects were given a short arithmetic task immediately following the study list. The arithmetic task began when the participant pressed a key on the keyboard. Five simple addition and subtraction equations (with an answer) were shown individually. The subject determined if the equation presented on the screen was correct or incorrect. Participants responded by pressing the "/" key on the keyboard if the equation was correct or the "Z" key if the equation was incorrect. Each equation remained at the center of the screen for 2000 ms or until a response was made. There was a 500-ms blank screen between trials. After the arithmetic task, the participant pressed a key to continue to the recognition test.

The procedure for the recognition test was based, in part, on tempo naming tasks used by Balota and Shields (1988) and Kello and Plaut (1998). In this procedure, subjects are given a series of cues in rhythm. The cues (both visual and auditory) rhythmically indicate the point in time when the response should be initiated. This rhythmic cueing procedure is very similar to a response deadline procedure. The primary difference is that subjects are given clear temporal cues during each trial (as reflected by the rhythmic patterning) that indicate when to initiate the response. As described below, subjects are remarkably successful in producing responses at the targeted rhythmic delay within a narrow temporal window.

The timing of recognition decisions was guided through the use of a rhythmic sequence of tones and visual cues, which flanked the location of the recognition test stimulus. As illustrated in Fig. 5, the following sequence of events occurred on each trial: (a) a 50-ms 750-Hz tone was presented simultaneously with the onset of a pair of ampersands, separated by the number of spaces corresponding to the recognition target plus 10 spaces, which served as vi-

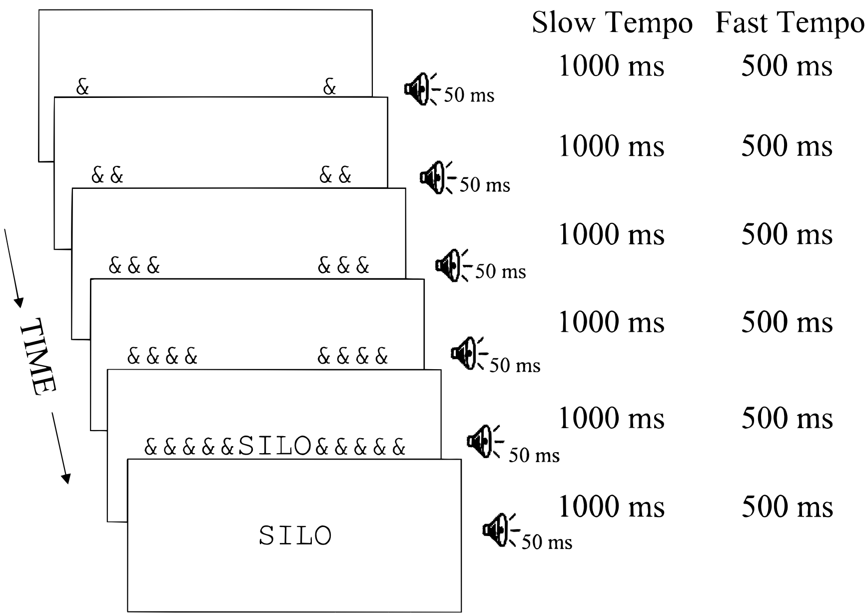

FIG. 5. Sequence of events in response deadline paradigm.

sual cues indicating where and when the target stimulus would be presented; (b) the flankers remained on the screen for the duration of one tempo (either 500 or 1000 ms); (c) a second 50-ms 750-Hz tone was presented with an additional pair of flankers presented at the immediate inside of the original pair; (d) the presentation of the tones and the flankers was repeated three more times in rhythm; (e) on the fifth presentation of visual/auditory cues, the recognition test stimulus was presented within the flankers; (f) after time equivalent to one tempo passed, the flankers were removed from the screen and the target stimulus alone remained at the center of the screen; and (g) participants pressed either the "/" key for old or "z" key for new. After each response, the participants were also given feedback to help them stay on tempo. If their keypress was correct and within a 100-ms window of the correct tempo (i.e., ±50 ms), an "O" would appear at the center of the monitor. For every 50 ms that the response was early, a "−" would appear to the left of the "O," whereas for every 50 ms of a late response a "+" sign was presented to the right of the "O." After an incorrect response, the word "WRONG" was shown and a short tone was presented.

After the recognition test was completed for Block 1, the participants were instructed to press a key to continue to the practice trials for the second phase of the experiment. The second phase was conducted in an identical fashion to the first except that different study and test words were presented and the duration of the tempo changed based on the counterbalancing order.

Results

Prior to analysis, the data were screened for outliers. Responses which were ±1000 ms from the tempo were omitted from all calculations. The means and SDs for each subject were computed from all remaining observations. Those trials which were 2.5 SDs from the mean for that subject were omitted from further analyses. This screening procedure eliminated 2.8% of the data.

Accuracy. The mean percentage hit rate and false alarm rate are displayed in Fig. 6. As shown in this figure, the results are quite clear and very consistent with the predictions based

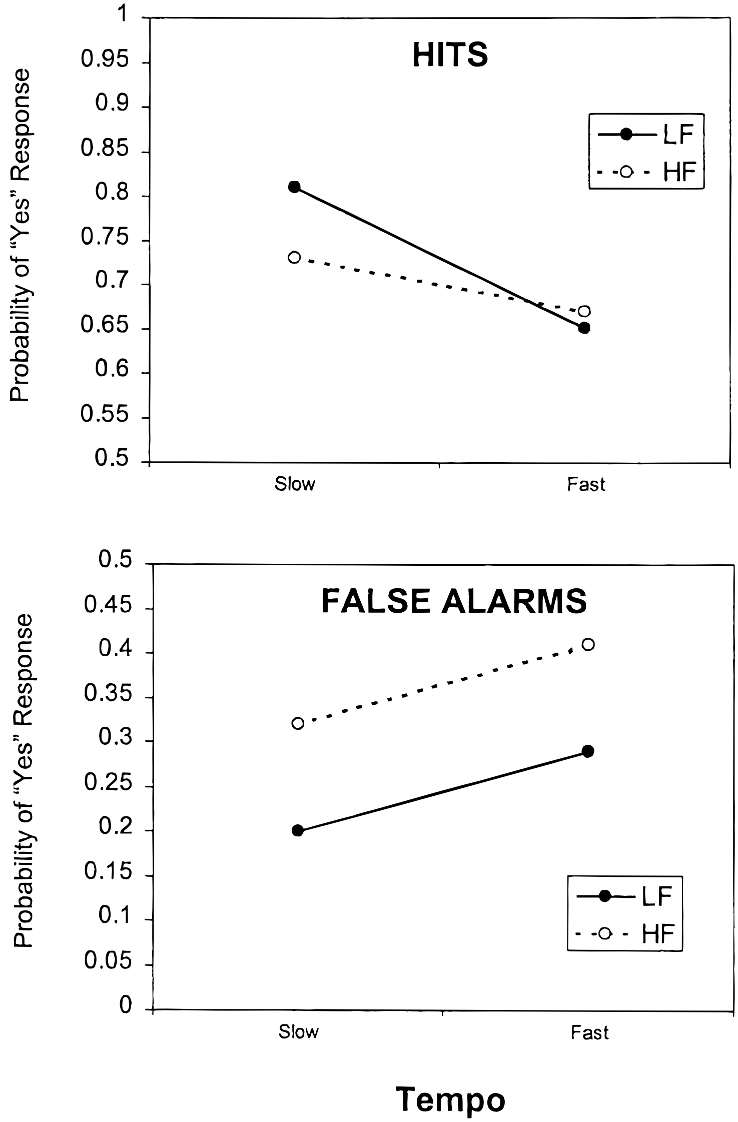

FIG. 6. Mean hit and false alarm rate as a function of word frequency and tempo.

on the results from Experiment 1. Specifically, at the slow tempo, there is evidence of the mirror effect in which low-frequency words produce a higher hit rate and lower false alarm rate compared to high-frequency words. However, at the fast tempo the results are quite different. In particular, there is a slight reversal of the word frequency effect with the high-frequency words producing a higher hit rate than the low-frequency words. However, there was virtually no modulation of the frequency effect by tempo for the distractor items.

The above observations were supported by a 2 (high vs low frequency) × 2 (fast vs slow tempo) × 2 (old vs new) mixed-factor ANOVA. This analysis yielded a significant three-way interaction between tempo, old/new, and word frequency, $F(1,46) = 6.48$, $MSE = .008$, $p = .014$. In order to further investigate this interaction, separate 2 (high vs low frequency) × 2 (fast vs

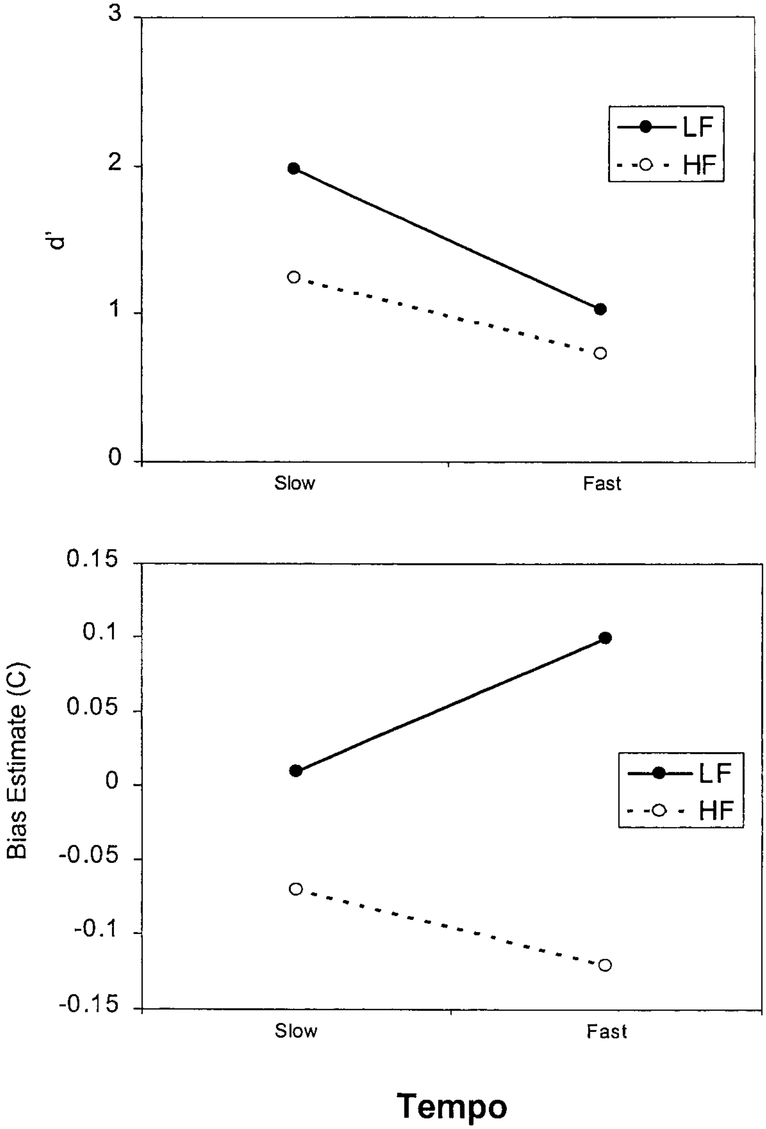

FIG. 7. Mean discriminability and bias estimates as a function of word frequency and tempo.

slow tempo) ANOVAs were conducted on the targets and the distractors. The results from these analyses were also clear. Specifically, for the targets, there were main effects of tempo, $F(1,47) = 26.89$, $MSE = .02$, and Frequency, $F(1,47) = 4.55$, $MSE = .008$. More importantly, there was a highly reliable tempo $\times$ frequency interaction, $F(1,47) = 11.69$, $MSE = .0092$, which reflected the fact that the frequency effect was reliable at the slow tempo, $F(1,47) = 20.75$, $MSE = .007$, and slightly reversed, but not reliable, at the fast tempo ($F < 1.00$). Turning to the false alarm results, there was again a highly reliable main effect of tempo, $F(1,47) = 24.89$, $MSE = .017$, and frequency, $F(1,47) = 86.75$, $MSE = .0089$. However, in contrast to the hit rate, there was no frequency $\times$ tempo interaction, $F < 1.00$.

Signal detection analyses. As in Experiment 1, we also submitted the present results to a sig-

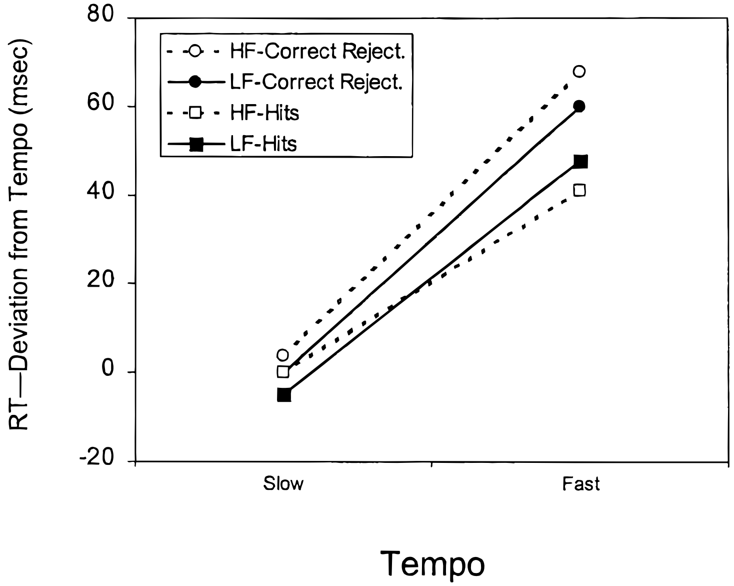

FIG. 8. Mean response latency deviation from the correct response tempo as a function of word-frequency and tempo.

nal detection analysis. The estimates of bias and sensitivity as a function of tempo and word frequency are displayed in Fig. 7. Overall, these results are consistent with the results from the signal detection analyses from Experiment 1. First, as shown in the top panel, the difference in d' between high- and low-frequency words decreases at the fast tempo compared to the slow tempo. Second, as shown in the bottom panel, the changes in the Bias estimates across tempos go in the opposite direction for high- and low-frequency words.

The above observations were supported by separate ANOVAs on the sensitivity and criterion estimates. For the sensitivity measures, there was a highly reliable effect of tempo, $F(1,47) = 46.03$, $MSE = .55$, frequency, $F(1,47) = 13.07$, $MSE = .22$, and a reliable interaction between tempo and frequency, $F(1,47) = 9.09$, $MSE = .25$, which indicates that sensitivity to low-frequency words changes more across tempo than for high-frequency words. Turning to the measure of bias, the main effect of Frequency was reliable, $F(1,47) = 24.64$, $MSE = .019$, with a reliable interaction between frequency and tempo, $F(1,47) = 4.83$, $MSE = .02$, reflecting the pattern that there was an increased bias to high-frequency words at the fast tempo compared to the slow tempo, whereas for low-frequency words there was a decreased bias across tempos.

Response latency. Deviations in response latency from the targeted response tempo were also calculated for both hits and correct rejections. (Because there is no between-participants comparison, as in Experiment 1, the z score transformation of the data is not warranted.) The mean deviations from the targeted tempo are displayed in Fig. 8. There are three points to note in this figure. First, participants were quite good at responding in line with the tempos, with an average deviation of 28 ms from the targeted response deadline. Second, responses deviated from the target tempo more at the fast deadline compared to the slow deadline. Third, and most importantly, these data also conform to the accuracy data. Specifically, for the hit rates, at the slow tempo, response latencies were faster to low-frequency words than to high-frequency words, whereas at the short tempo, there is some hint of a reversal of the word-frequency effect. However, turning to the distractor items, there only appears to be a main effect of frequency with the high-frequency words producing shorter response latencies than the low-frequency words at both the short and the long tempos.

The above observations were, in general, supported by an ANOVA. The overall ANOVA yielded a main effect of tempo, $F(1,47) = 46.22$, $MSE = 6132$, but the frequency effect did not reach significance, $F(1,47) = 1.26$, $MSE = 484.85$. Although the overall ANOVA did not yield a reliable three-way interaction among tempo, frequency, and old/new, $F(1,47) = 2.29$, $MSE = 668$, $p = .14$, the separate two-way ANOVAs did support the above interpretation. Specifically, for the hit rate, there was a reliable two-way interaction between tempo and frequency, $F(1,47) = 4.49$, $MSE = 376$, indicating that the word frequency effect was in the predicted direction at the slow tempo, but actually reversed at the fast tempo.[3] However, for the correct rejection rate, there was no hint of an interaction between frequency and tempo, $F < 1.00$. Thus, the response latency data are in accord with the accuracy data.

Discussion

The results from Experiment 2 were consistent with the interpretation from Experiment 1. Specifically, we argued that across groups in Experiment 1, participants appeared to rely more on baseline familiarity than recollection to drive a response. This was reflected in the data by a dramatic decrease in the frequency effect for the hit rate, whereas the frequency effect for the false alarm rate did not change across groups. If our interpretation of Experiment 1 was correct, then we expected to replicate this pattern in Experiment 2, where young adult subjects were put under time pressures to respond relatively quickly. In fact, this is precisely the pattern of results found in Experiment 2. Specifically, for hit rates, there was evidence of a crossover interaction in both accuracy and response latencies, whereas for false alarm rates, there was no hint of an interaction.

GENERAL DISCUSSION

The results of the present experiments converge on the notion that there are at least two types of information that, under some conditions, compete to drive an episodic recognition decision (see Jacoby, 1999). One component entails a more recollective or controlled process that involves a slower, more attention demanding analysis and contributes to the advantage in hit rates for low-frequency words compared to high-frequency words. The second component entails a baseline familiarity component and appears to be faster acting and is likely to be more automatic in nature. As Joordens and Hockley (2000) and Reder et al. (2000) have recently argued these two components may act in opposition when considering the hit rate for high- and low-frequency words. Specifically, low-frequency words produce higher recollection, as reflected by more “remember” responses, whereas high-frequency words produce higher familiarity, as reflected by more “know” responses. For healthy young adults, the low-frequency advantage in episodic recognition performance reflects the recollection-based process outweighing the familiarity-based process.

Within this framework, the results from Experiment 1 could be viewed as indicating that healthy older adults and individuals with early-stage DAT produce breakdowns in the attention demanding recollective process, but not in the more automatic familiarity-based process. This is reflected by the considerable breakdown in the hit rates for low-frequency words across groups, reflecting changes in the recollection process, with relatively little change in hit rate for high-frequency words, reflecting the more automatic, familiarity-based process. The additive effect of frequency across groups on the false alarm rates is also consistent with this notion because high-frequency words should produce more false alarms than low-frequency words in the absence of recollection for all

[3] We also included an ANOVA, which included Block order (fast tempo first vs slow tempo first) as a factor. As expected, the predicted interaction was somewhat larger when participants received the slow response deadline first and then received the fast response deadline. This pattern was expected because of a contrast effect of receiving a fast tempo after receiving a slow tempo.

groups, i.e., subjects are more likely to respond on baseline familiarity for nonpresented items.

The results from Experiment 2 are also consistent with this interpretation. Specifically, we were able to mimic the pattern of word-frequency effects in hit and false alarm rates found across groups in Experiment 1 in a single group of young adults who responded either at a relatively slow pace, which allowed recollective processes to be engaged, or at a relatively fast pace, which encouraged responses that were based on familiarity processes. This experiment was motivated by recent work by Jacoby (1999), suggesting that one can modulate the reliance on familiarity and recollection by manipulating response deadlines (also see Dolan & Balota, 1999; Joordens & Hockley, 2000).

Although we believe the present pattern, along with other extent data, is quite consistent with this general interpretation, there clearly are alternative accounts that need to be addressed. We address each of these in turn.

Simple Memory Differences vs Recollection/Attentional Differences across Groups and Deadlines

The present account places emphasis on the distinct roles of recollection and familiarity in the recognition performance. It is possible that the large decrease in performance on low-frequency words may be *totally* due to differences in the quality of encoding these items. In this light, single-process models that rely on the quality of encoding may be a more parsimonious account of the present results. Because high-frequency words are already poorly encoded, there is no change in performance on these items across groups. Although it is very clear that memory encoding differences are likely to contribute to the present results, we do not believe that this is the full story. There are two aspects of our data that are inconsistent with this account.

First, if the present results are totally due to different levels of memory encoding, then one might be able to eliminate the group $\times$ frequency interaction by partialing out overall memory performance on a different declarative memory test for the same group of individuals. One hundred twenty-three participants who participated in Experiment 1 also participated in a different experiment within the same 2-h testing session that included estimates of declarative memory as reflected by free recall performance (see Balota et al., 1999). In order to examine whether the pattern of accuracy data in the present study was attributable to varying influences from overall declarative memory performance, we partialed out the participants' recall performance via analyses of covariance. If differences in overall memory performance were responsible for the pattern of results obtained in Experiment 1, recognition performance would be similar across groups once differences in free recall were partialled out. The results of the analyses indicated that the frequency $\times$ group interaction remained significant in the analysis by age, $F(2,78) = 5.38$, $MSE = .01$, and by DAT, $F(2,93) = 12.10$, $MSE = .02$, even after the participants episodic recall performance was partialed out. In addition, there was still no frequency $\times$ group interaction for the false alarms in either the analysis by age or by DAT. Thus, when one covaries out baseline recall performance, the initial pattern of results does not substantially change. Furthermore, the same pattern of reliable effects occurs if one partials out Logical Memory performance (a standard psychometric declarative memory task). Specifically, the interaction in hits still remains highly reliable, $F(3,115) = 9.98$. (It is noteworthy that we could only conduct the analysis with two older adult groups and two DAT groups because we do not have available data from the Logical Memory test for the young adults.) Thus, simple differences in declarative memory across groups is not sufficient to account for the changes in the mirror effect across groups observed in Experiment 1.

In addition to partialing out memory performance, the results from Experiment 2 also suggest that memory encoding differences cannot totally accommodate the present results. Specifically, there was no difference in encoding across the subject groups who received the fast and slow tempos. However, simple temporal deadlines yielded a pattern of performance that was identical to the between-group performance found in Experiment 1. Joordens and Hockley (2000) have recently reported similar results from a deadline experiment. Thus, it is unlikely that differences in memory trace strength can account for these results. Rather, it appears that subjects rely on different types of information across groups and at different temporal deadlines.

Finally, it is also worth noting here that any account of the word-frequency mirror effect that totally relies on differences in memory trace strength for high- and low-frequency words must also accommodate the elimination or reversal of this effect in recall performance (e.g., Balota & Neely, 1980; Glanzer & Bowles, 1976). Although factors such as biases to report high-frequency words and increased associative connections for high-frequency words have been proposed to account for the reversal of the word frequency effect in recall performance, we are unaware of any clear tests of the degree to which these factors modulate the word frequency effect in recall performance. As discussed below, we believe that the attentional demands at retrieval, as a function of task constraints, are more likely candidates for the dissociation in word-frequency effects in recall and recognition performance.

Semantic/Lexical Changes across Groups

In addition to baseline differences in memory performance across groups, it is also possible that the present results are due to group differences in their vocabulary knowledge of the low-frequency words, i.e., semantic memory differences. Possibly, the semantic memory representations in the DAT individuals are degraded and low-frequency words, because of their infrequent use, are more susceptible to this degradation compared to high-frequency words. In fact, this is a component of the account that Wilson et al. (1983) proffered in their initial word-frequency episodic recognition study. Specifically, low-frequency words may be more degraded in DAT individuals and hence are less likely to be appropriately encoded and retrieved. Pushing this view to an extreme, one might expect poor performance on low-frequency words because these items are functionally being encoded as nonwords. However, there are four reasons that make this possibility untenable as the sole cause of the observed pattern of data: First, in a lexical decision study by Balota and Ferraro (1996) in which participants were required to discriminate words from nonwords, DAT individuals at the same cognitive level as the present DAT individuals, were quite accurate in recognizing low-frequency words and did not reliably differ from healthy control individuals (5.3% vs 6.3% error rates for the healthy control and DAT individuals, respectively). Second, the present word frequency $\times$ group interaction in hit rates also occurred in healthy aging, in which there is often an increase in vocabulary performance from healthy young to healthy older adults (see Balota & Ferraro, 1996, for a discussion of this issue). In fact, one might argue that this should work against observing the interaction. Third, the results from the tempo experiment converged on the same pattern of results, even though tempo was a within-participants manipulation. Finally, we partialed out differences in semantic/lexical performance from measures available from the psychometric battery (Boston Naming and WAIS information subtests), and again the group $\times$ word frequency interaction in hit rate was still reliable, $F(3,114) = 4.22$, $MSE = .10$, $F(3,114) = 4.90$, $MSE = .10$, respectively, and the interaction in false alarm rate did not approach significance (both Fs < 1.00). Based on these observations, it does not appear that the present results can be accounted for by differences between groups in vocabulary knowledge of low-frequency words.

Other Studies That Have Addressed the Role of Attention and Retrieval Speed in the Word Frequency Mirror Effect

Hintzman, Caulton, and Curran (1994) also investigated the role of attention in producing the word-frequency mirror effect in episodic recognition. In their study, they used both divided attention tasks and a response signal procedure. Because they found little change in the word-frequency mirror effect under single vs divided attention manipulations and also under short and long response deadlines, they concluded that the mirror effect does not reflect conscious, postretrieval judgments of the memorability of the items. The observation that attentional manipulations do not modulate the word-frequency mirror effect would appear to be inconsistent with the present results and interpretation. However, there are two aspects of the Hintzman et al. deadline experiments that warrant further attention. First, Hintzman et al. found relatively small word-frequency effects in the hit rates. In fact, the results from their first experiment did not produce a reliable word-frequency effect in the hit rate. Second, there was a clear tendency in their data to produce a pattern quite similar to the results from the present second experiment. For example, the results of their second experiment (with three response signal deadlines) indicated that there appears to be about a 7 to 8% advantage for low-frequency words at the long deadline, but virtually no difference exists between low- and high-frequency words at the short deadline. This change in the frequency effect did not occur for lure items and actually showed some tendency for the opposite pattern. In fact, in Hintzman et al.'s Experiment 6, they used shorter response signal lags because there appeared to be a differential bias to call high-frequency words old at the shortest (175 ms) lag in Experiment 3. The results of their Experiment 6 yielded a similar pattern. At this level, the results from Hintzman et al. appear quite consistent with the present results.

In addition, as noted above, Joordens and Hockley (2000) recently presented evidence from a deadline experiment with young adults, which is quite consistent with the results from the present Experiment 2. They found that speeding encoding and/or retrieval operations did not influence the word-frequency effect on the false alarm rates, but did influence, and in some cases reverse, the word-frequency effect for the hit rates. Interestingly, Joordens and Hockley also argued that Hintzman et al.'s. (1994) failure to modulate the word-frequency mirror effect via the presence of a secondary attention demanding task may have been due to the fact that retrieval processes appear to dominate performance of a secondary task (see, for example, Anderson, Craik, & Naveh-Bejamin, 1998). In this light, it may be more likely to modulate the word-frequency mirror effect by forcing a response via deadline procedures than by dividing attention via the presence of a secondary task at retrieval. Joordens and Hockley also viewed their results as most consistent with a two-process model of recognition performance, wherein speeded encoding or retrieval demands primarily modulate the influence of recollection, leaving the familiarity of the stimulus as the primary influence on performance. As noted above, this two-process framework was also consistent with Joordens and Hockley's observation that *know* responses are higher for high-frequency words than low-frequency words, whereas, *remember* responses show the reverse pattern (also see Reder et al., 2000).

Implications for the "Memory" Deficit in Healthy Aging and Alzheimers Disease

One simple implication of the present results is that older adults, and to a much larger extent DAT individuals, appear to have particular breakdowns in remembering relatively rare events. We have argued that this deficit may in part be related to attention demanding aspects of recollection in determining if a given stimulus item was seen earlier in the experimental context. Word-frequency may be a particularly tricky variable in the present experimental context, because familiarity (a strong correlate of frequency) is typically a useful piece of information in making recognition memory judgments, but can be misleading in this context.

Of course, there are other ways to increase the familiarity of a stimulus word via the context that it is embedded. For example, one might also manipulate the degree to which context is related to a particular item, thereby making the item more or less familiar due to the supporting context. In this light, the results from some recent studies exploring false memory are noteworthy. Specifically, Balota et al. (1999); Norman and Schacter (1997); Tun, Wingfield, Rosen, and Blanchard (1998); and Watson, Balota, and Sergent-Marshall (2001) have all reported evidence of increased, relative to veridical memory, susceptibility to false memories in healthy aging in the Deese/Roediger and McDermott paradigm. In this paradigm, participants are presented words (e.g., THREAD, PIN, EYE, SEWING, SHARP,

POINT, PRICK, THIMBLE, HAYSTACK, PAIN, HURT, and INJECTION) that are related to a critical nonpresented word (e.g., NEEDLE). Although veridical memory decreases in these subjects, false memory rates either remain the same or increase across groups. Balota et al. (1999) have extended this work to individuals with Alzheimer's disease. The results from their recall test are displayed in Fig. 9. As shown here, these data have an intriguing similarity to the hit rates displayed in Fig. 3. Specifically, for the items that have a high degree of support from the highly related list (e.g., the critical nonpresented word NEEDLE), there are relatively similar levels of false recall. On the other hand, for items that have relatively less support within the context of items, i.e., the items that were actually presented, there is a considerable breakdown in recall performance. Interestingly, as in the present study, Balota et al. demonstrated that this interaction was not simply due to differences in overall memory performance because this pattern persisted even after partialing out veridical memory performance. Thus, it is possible that familiarity, as defined by the context in which an item is embedded, will more likely drive memory performance in both healthy older adults and in DAT individuals. Balota et al. argued that this pattern could be accommodated within an attentional control framework in which healthy older adults and DAT individuals are more likely to select overall familiarity to drive a response instead of source-specific familiarity.

Why might there be a shift to a reliance on familiarity-based information in healthy aging? One simple possibility is that the additional 50 years of experience in healthy older adults, compared to healthy young adults, has increased their reliance on familiarity-related information because this information is typically predictive of correct responses, while minimizing the demands on precious limited attentional resources. Unfortunately, this simple account would not accommodate the increased susceptibility to baseline familiarity in age-matched DAT individuals. A second related possibility is that changes in underlying neural systems produce breakdowns in attentional selection processes,

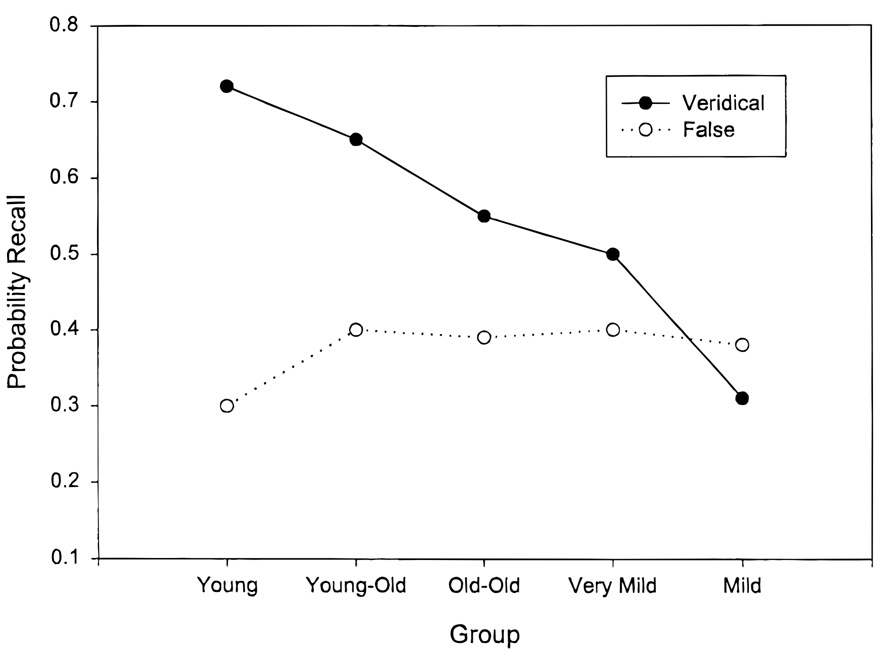

FIG. 9. Mean proportion correct recall and mean proportion false critical recall in the DRM paradigm as a function of Group (taken from Balota et al., 1999).

and this may increase the influence of prepotent relatively stronger sources of information, akin to the influence of the word dimension on color naming in the Stroop task. In this light, it is interesting to note that there is accumulating evidence that both healthy aging and early-stage Alzheimer's Disease produce changes in frontal cortical systems that have been viewed as being critical to some aspects of attentional control systems (e.g., Balota & Faust, 2001; Morris et al., 1996; West, 1996). We believe these changes in attentional control systems produce breakdowns both in the richness of the encoded memory trace and the ability to control different sources of activated information at retrieval. At this level, we believe it is important to reemphasize the contribution of attentional control systems in the episodic memory loss exhibited in both healthy older adults and in early stage Alzheimer's disease.

Of course, one must also be cautious not to overemphasize a single system underlying the cognitive changes that occur in healthy aging and in early stage Alzheimer's disease. There is clear evidence of heterogeneity in the neuropathological and cognitive profiles of healthy older adults and in DAT individuals. Recently, researchers have been able to use the cognitive and/or neuropathological profiles of individuals to predict performance on targeted tasks and/or components of tasks (e.g., Glisky, Polster, & Routhieaux, 1995; Kanne et al., 1998; Henke, Johnson, & De Leonardis, 1999). At this level, the present work might be viewed as an attempt to move the pendulum away from the view that Alzheimer's Disease is a declarative memory/medial temporal disease. There are additional neural systems (including frontal attentional systems) that may underlie the observed breakdowns in memory performance in early stage DAT.

The Nature of Recognition: Are Two Processes Enough?

So far in this article, we have been relatively vague about the nature of the recollection and familiarity components in recognition. One common interpretation of recollection is the retrieval of the source of the earlier occurrence of the stimulus. Specifically, the person remembers some episodic details of the encoding event. Of course, this is quite consistent with the observation that low-frequency words produce more "remember" responses than do high-frequency words (e.g., Gardiner & Java, 1990). On the other hand, there are aspects of the task demands of episodic recognition performance that suggest that additional processes may also play a role. One piece of information that has been shown to modulate episodic recognition performance is the processing fluency of the target. For example, Jacoby and Whitehouse (1989) have shown that increasing the speed of processing of a recognition test item, via the presentation of a masked identity prime, can lead to increases in false alarms and hit rates compared to an unrelated prime condition. This work suggests that there is an important role of fluency in making episodic recognition decisions, i.e., if the stimulus is processed unusually fast, then it is likely that is was recently presented. This is particularly relevant to the word-frequency mirror effect because low-frequency words produce larger fluency boosts due to repetition than do high-frequency words (e.g., Jacoby & Dallas, 1981). This relatively greater increase in fluency for low-frequency words, compared to high-frequency words, due to a study episode could be used to make an attribution that the stimulus is old.

The potential role of relatively larger changes in familiarity for low-frequency words compared to high-frequency words is consistent with a recent study by Dolan and Balota (1999). In this study, participants either read or heard a list of high- and low-frequency words in an opposition paradigm (see Jacoby, 1999). Subjects were instructed to respond "old" only to earlier read words. At a long deadline (1500 ms), individuals were better at rejecting low-frequency "heard" words, compared to high-frequency words, presumably due to low-frequency items producing greater recollection. More intriguing is the pattern at the short deadline (800 ms). If subjects relied on only baseline familiarity at the short deadline, then one would expect subjects to be relatively poor at rejecting high-frequency heard words compared to low-frequency words.

However, this was not the pattern. In fact, subjects were actually poorer at rejecting low-frequency heard words at the short deadline, yielding a crossover interaction between frequency and delay. This is consistent with the notion that subjects were using relative change in familiarity, instead of baseline familiarity, to make their episodic recognition decisions. Of course, based on the present results one might expect that if one shortened the deadline to 500 ms that one would find that baseline familiarity would modulate performance, and hence high-frequency heard words would be more poorly rejected.

The importance of relative familiarity as an account of the word-frequency mirror effect was originally espoused by Brown, Lewis, and Monk (1977), and similar views have more recently been endorsed by Benjamin, Bjork, and Hirshman (1998, also see Joordens & Hockley, 2000). The notion here is that participants have available an estimate of the baseline familiarity of the stimulus (or the average baseline familiarity of the stimuli being presented) along with the relative change in familiarity, as indexed by changes in the fluency of processing the stimulus from an earlier encoding event. It is possible that the difference between these two estimates provides a cue for recognition.

If participants can use the relative change in familiarity as a useful piece of information to drive episodic recognition decisions, then one might expect participants to consider low-frequency words more memorable than high-frequency words. This prediction was not supported in early work by Greene and Thapar (1994) and Wixted (1992), who found that subjects actually reported high-frequency words as more memorable than low-frequency words. However, it is interesting to note here that more recent work by Guttentag and Carroll (1998) and Benjamin (2001) has demonstrated that if subjects are asked to make their memorability judgments during a recognition test that indeed participants rate low-frequency words more memorable than high-frequency words, even for items that were considered new. Hence, it appears that exposure to a recognition test does increase the sensitivity to the memorability of low-frequency words. It is possible that subjects develop the metacognitive information available regarding the utility of a relative boost in familiarity to drive episodic recognition decisions.

The role of metacognitive/inferential processes in episodic recognition performance and the changes that we observed in the present study across age groups is intriguing in light of the work by Parkin and Walter (1992). They found that older adult's "know" judgments were negatively correlated with performance on Wisconsin Card Sorting task, which presumably reflects frontal lobe function. There is also evidence that metamemory performance is particularly disrupted in individuals with frontal lobe damage (e.g., Janowsky, Shimamura, & Squire, 1990). Finally, there is a relatively high incidence of senile plaque involvement (a neuropathological marker for Alzheimer's Disease) in frontal lobes of early-stage DAT individuals (e.g., Kanne et al., 1997; Morris et al., 1996). In this light, it is possible that present results may in part be due to changes that occur in older adults and in individuals with early-stage DAT in insights regarding the appropriate cues to use to make recognition judgments. Of course, such a possibility is also consistent with the rather large attentional breakdowns that have been observed in these groups (see Balota & Faust, 2001, for a review).

SUMMARY

The present experiments have shown that one can eliminate and partially reverse the low-frequency hit rate advantage in episodic recognition performance, while leaving the low-frequency false alarm advantage unaltered, i.e., one can eliminate the word-frequency mirror effect in episodic recognition performance. This was accomplished by comparing different groups of participants who exhibit both memory and attentional breakdowns and also by forcing young adults to respond within a narrow temporal deadline. We have viewed these results as most consistent with the view that there are multiple sources of information that are available during a recognition test to drive an "old" response. These include baseline familiarity, recollection of specific episodes, and possible attribution processes regarding the relative change in fluency of a stimulus. Because of the widespread changes in attentional processes in both healthy older adults and in early-stage DAT individuals, we believe that there may be increased likelihood of these individuals relying on inappropriate sources of familiarity in driving an "old" response.

APPENDIX

High-frequency words					
CAR	ROAD	WORLD	RADIO	STREET	COLLEGE
ROSE	HOME	PLANE	BEACH	BOTTLE	LIBRARY
KEY	BALL	FLOOR	CLOUD	JACKET	PICTURE
COW	TOWN	BOOK	BREAD	NATION	PALACE
CAT	KING	FIELD	DRESS	GARDEN	KITCHEN
SUN	BABY	HOTEL	MOUTH	STREAM	FATHER
CUP	POOL	TRUCK	SNAKE	MARKET	VILLAGE
TREE	BEAR	WHEEL	VALLEY	ENGINE	CHICKEN
Low-frequency words					
URN	FLEA	COBRA	WHARF	TABLET	CREVICE
BOAR	SILO	OLIVE	YACHT	TRIPOD	DUNGEON
OWL	LILY	OTTER	DWARF	GALAXY	PARASOL
SPA	HARP	ISLE	TUNIC	WIZARD	BEGGAR
GEM	LASS	VALET	ALTAR	CAVERN	MONSOON
KEG	VASE	GOURD	FLASK	AFGHAN	SLEUTH
EEL	PLUM	PECAN	ANVIL	JAGUAR	TROLLEY
VINE	LOFT	BANJO	SEQUIN	BONNET	VULTURE

REFERENCES

Aggleton, J. P., & Shaw, C. (1996). Amnesia and recognition memory: A re-analysis of psychometric data. *Neuropsychologia*, **34,** 51–62.

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.

Armitage, S. G. (1946). An analysis of certain psychological tests used for the evaluation of brain injury. *Psychological Monographs*, **60,** Whole No. 277.

Balota, D. A., Boland, J. E., & Shields, L. W. (1989). Priming in pronunciation: Beyond pattern recognition and onset latency. *Journal of Memory & Language*, **28,** 14–36.

Balota, D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language*, **24,** 89–106.

Balota, D. A., Cortese, M. J., Duchek, J. M. Adams, D., Roediger, H. L., III, McDermott, K. B., & Yerys, B. E. (1999). Veridical and false memories in healthy older adults and in dementia of the Alzheimer's Type. *Cognitive Neuropsychology*, **16,** 361–384.

Balota, D. A., & Duchek, J. M. (1989). Spreading activation in episodic memory: Further evidence for age independence. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, **41,** 849–876.

Balota, D. A., & Duchek, J. M. (1991). Semantic priming effects, lexical repetition effects, and contextual disambiguation effects in healthy aged individuals and individuals with senile dementia of the Alzheimer type. *Brain and Language,* **40,** 181–201.

Balota, D. A., & Faust, M. E. (2001). Attention in Dementia of the Alzheimer's Type. In F. Boller & S. Cappa (Eds.), *Handbook of neuropsychology (2nd ed.)*. Amsterdam, The Netherlands: Elsevier.

Balota, D. A., & Ferraro, F. R. (1996). Lexical, sublexical, and implicit memory processes in healthy young and healthy older adults and in individuals with dementia of the Alzheimer type. *Neuropsychology*, **10,** 82–95.

Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory*, **6,** 576–587.

Balota, D. A., & Shields, L. W. (1988). *The locus of word-frequency effects in pronunciation*. Paper presented at the 1988 meeting of the Psychonomics Society, Chicago, IL.

Balota, D. A., Watson, J. M., Duchek, J. M., & Ferraro, F. R. (1999). Cross-modal semantic and homograph priming in healthy young, healthy old, and in Alzheimer's disease individuals. *Journal of the International Neuropsychological Society*, **5,** 626–640.

Benjamin, A. S. (2001). What you think you know isn't necessarily what you thought you would have known: Using word frequency effects as a model for training metacognition. Manuscript in preparation.

Benjamin, A. S., Bjork, R. A., & Hirshman, E. (1998). Predicting the future and reconstructing the past: A Bayesian characterization of the utility of subjective fluency. *Acta Psychologica*, **98,** 267–290.

Berg, L., McKeel, D. W., Jr., Miller, J. P., Storandt, M., Rubin, E. H., Morris, J. C., Baty, J., Coats, M., Norton, J., Goate, A. M., Price, J. L., Gearing, M., Mirra, S. S., & Saunders, A. M. (1998). Clinicopathological studies in cognitively healthy aging and Alzheimer's disease: Relation of histologic markers to dementia severity. *Archives of Neurology*, **55,** 326–335.

Brown, J., Lewis, V. J., & Monk, A. F. (1977). Memorability, word frequency and negative recognition. *Quarterly Journal of Experimental Psychology*, **29,** 461–473.

Burke, W. J., Miller, J. P.; Rubin, E. H., Morris, J. C., Coben, L. A., Duchek, J, Wittels, I. G., & Berg, L. (1988). Reliability of the Washington University Clinical Dementia Rating. *Archives of Neurology*, **45,** 31–32.

Butters, N., Granholm, E., Salmon, D. P., Grant, I., & Wolfe, J. (1987). Episodic and semantic memory: A comparison of amnesic and demented patients. *Journal of Clinical and Experimental Neuropsychology*, **9,** 479–497.

Dolan, P. O., & Balota, D. A. (1999, November). *Two effects of word frequency: Recollection, aging, and the Mirror Effect*. Poster presented at the 40th Annual meeting of the Psychonomic Society, Los Angeles, CA.

Faust, M. E., Balota, D. A., Spieler, D. H., & Ferraro, F. R. (1999). Individual differences in information-processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin*, **125,** 777–799.

Ferraro, F. R., Balota, D. A., & Connor, L. T. (1993). Implicit memory and the formation of new associations in non-demented Parkinson's disease individuals and individuals with senile dementia of the Alzheimer type: A serial reaction time investigation. *Brain and Cognition*, **21,** 163–180.

Gabrieli, J. D. E., Vaidya, C. J., Stone, M., Francis, W. S., Thompson-Schill, S. L., Fleischman, D. A., Tinklenberg, J. R., Yesavage, J. A., & Wilson, R. S. (1999). Convergent behavioral and neuropsychological evidence for a distinction between identification and production forms of repetition priming. *Journal of Experimental Psychology: General*, **128,** 479–498.

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition*, **18,** 23–30.

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **16,** 5–16.

Glanzer, M., & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, **2,** 21–31.

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, **13,** 8–20.

Glinsky, E. L., Polster, M. R., & Routhieaux, B. C. (1995). Double dissociation between item and source memory. *Psychology and Aging*, **9,** 229–235.

Goodglass, H., & Kaplan, E. (1983). *Boston Naming Test*. Philadelphia, PA: Lea & Febiger.

Greene, R. L., & Thapar, A. (1994). Mirror effect in frequency discrimination. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **20,** 946–952.

Guttentag, R., & Carroll, D. (1998). Memorability judgements for high- and low-frequency words. *Memory & Cognition*, **26,** 951–958.

Haist, F., Shimamura, A. P., & Squire, L. R. (1992). On the relationship between recall and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18,** 691–702.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 193–225). San Diego, CA: Academic Press.

Henkel, L. A., Johnson, M. K., & DeLeonardix, D. M. (1998). Aging and source monitoring: Cognitive processes and neuropsychological correlates. *Journal of Experimental Psychology: General*, **127,** 251–268.

Hintzman, D. L., Caulton, D. A., & Curran, T. (1994). Retrieval constraints and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20,** 275–289.

Hughes, C. P., Berg, L., Danziger, W. L., Coben, L. A., & Martin, R. L. (1982). A new clinical scale for the staging of dementia. *British Journal of Psychiatry*, **140,** 566–572.

Huppert, F. A., & Piercy, M. (1976). Recognition memory in amnesic patients: Effects of temporal context and familiarity of material. *Cortex*, **12,** 3–20.

Jacoby, L. L. (1983a). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **9,** 21–38.

Jacoby, L. L. (1983b). Remembering the data: Analyzing interactive processes in reading. *Journal of Verbal Learning and Verbal Behavior*, **22,** 485–508.

Jacoby, L. L. (1999). Deceiving the elderly: Effects of accessibility bias in cued-recall performance. *Cognitive Neuropsychology*, **16,** 417–436.

Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, **118,** 126–135.

Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1990). Memory and metamemory: Comparisons between patients with frontal lobe lesions and amnesic patients. *Psychobiology*, **17,** 3–11.

Java, R. (1996). Effects of age on state of awareness following implicit and explicit word-association tasks. *Psychology and Aging*, **11,** 108–111.

Joordens, S., & Hockley, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. *Journal of Experimental Psychology, Learning, Memory, and Cognition*, **26,** 1534–1555.

Jorm, A. F. (1986). Controlled and automatic information processing in senile dementia: A review. *Psychological Medicine*, **16,** 77–88.

Kanne, S. M., Balota, D. A., Storandt, M., McKeel, Jr., & Morris, J. C. (1998). Relating anatomy to function in Alzheimer's disease: Neuropsychological profiles predict regional neuropathology 5 years later. *Neurology*, **50,** 979–985.

Kello, C. T., & Plaut, D. C. (1998). *Modulation of effects in word reading without strategic control of pathways: Evidence from tempo naming*. Paper presented at the 39th annual meeting of the Psychonomics Society, Dallas, TX.

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

Mantyla, T. (1993). Knowing but not remembering: Adult age differences in recollective experience. *Memory & Cognition*, **21,** 379–388.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of the Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, **34,** 939–944.

Morris, J. C., Storandt, M., McKeel, D. W., Jr., Rubin, E .H., Price, J .L., Grant, E. A., & Berg, L. (1996). Cerebral amyloid deposition and diffuse plaques in "normal" aging: Evidence for presymptomatic and very mild Alzheimer's disease. *Neurology*, **46,** 707–719

Moscovitch, M., & Winocur, G. (1992). The neuropsychology of memory and aging. In F. I. M., Craik, & T. A. Salthouse, (Eds.), *The handbook of aging and cognition* (pp. 315–372). Hillsdale, NJ: Erlbaum.

Moscovitch, M. (1986). Afferent and efferent models of visual perceptual asymmetries: theoretical and empirical implications. *Neuropsychologia*, **24,** 91–114.

Nebes, R. D. (1992). Cognitive dysfunction in Alzheimer's Disease. In F. I. M Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition*, (pp. 373–446). Hillsdale, NJ: Erlbaum.

Nebes, R. D. (1994). Contextual facilitation of lexical processing in Alzheimer's disease: intralexical priming or sentence-level priming? *Journal of Clinical and Experimental Neuropsychology*, **16,** 489–497.

Nebes, R. D. (1991). Speed of thought differentiates dementia from depression. *Experimental Aging Research*, **17,** 86–87.

Nebes, R. D., Martin, D. C., & Horn, L. C. (1984). Sparing of semantic memory in Alzheimer's disease. *Journal of Abnormal Psychology*, **93,** 321–330.

Norman, K. A., & Schacter, D. L. (1997). False recognition in younger and older adults: Exploring the characteristics of illusory memories. *Memory & Cognition*, **25,** 838–848.

Ober, B. A., & Shenaut, G. K. (1995a). Semantic priming in Alzheimer's disease: Meta-analysis and theoretical evaluation. In P. A., Allen, & T. R. Bashore, (Eds.), *Age differences in word and language processing* (pp. 247–271). Amsterdam, The Netherlands: North-Holland/Elsevier.

Ober, B. A., & Shenaut, G. K. (1995b). Theoretical and practical issues in semantic priming research with Alzheimer's disease subjects. In L. F., DiLalla, & S. M. C. Dollinger, (Eds.), *Assessment of biological mechanisms across the life span* (pp. 121–150). Hillsdale, NJ: Erlbaum.

Parkin, A. J., & Walter, B. M. (1992). Recollective experience, normal aging, and frontal dysfunction. *Psychology and Aging*, **7,** 290–298.

Reder, L. M., Nhouyvansivong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember/know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26,** 294–320.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, **96,** 523–568.

Seidenberg, M. S., & McClelland, J. L. (1989). Visual word recognition and pronunciation: A computational model of acquisition, skilled performance, and dyslexia. In A. M. Galaburda, (Ed.), *From reading to neurons* (pp. 255–303). Cambridge, MA: MIT Press.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, **117,** 34–50.

Stretch, V., & Wixted, J. T. (1998a). Decision rules for recognition memory confidence judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24,** 1397–1410.

Stretch, V., & Wixted, J. T. (1998b). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **24,** 1379–1396.

Thurstone, L. E., & Thurstone, T. G. (1949). *Examiner manual for the SRT Primary Mental Abilities*. Chicago, IL: Science Research Associates.

Tun, P. A., Wingfield, A., Rosen, M. J., & Blanchard, L. (1998). Response latencies for false memories: Gist-based processes in normal aging. *Psychology and Aging*, **13,** 230–241.

Verfaellie, M., & Cermak, L. S. (1999). Perceptual fluency as a cue for recognition judgments in amnesia. *Neuropsychology*, **13,** 198–205.

Watkins, M. J., LeCompte, D. C., & Kim, K. (2000). Role of study strategy in recall of mixed lists of common and rare words. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **26,** 239–245.

Watson, J. M., Balota, D. A., & Sergent-Marshall, S. D. (2001). Semantic, phonological, and hybrid veridical and false memories in healthy older adults and in dementia of the Alzheimer's Type. *Neuropsychology*, **15,** 254–267.

Wechsler, D. (1955). *WAIS Manual*. New York: Psychological Corporation.

Wechsler, D., & Stone, C. P. (1973). *Manual: Wechsler Memory Scale*. New York: Psychological Corporation.

West, R. L. (1996). An application of prefrontal cortex function theory to cognitive aging. *Psychological Bulletin*, **120,** 272–292.

Whittlesea, B. W. A., & Williams, L. D. (1998). Why do strangers feel familiar, but friends don't? The unexpected basis for feelings of familiarity. *Acta Psychologica*, **98,** 141–166.

Whittlesea, B. W. A., & Williams, L. D. (2000). The source of feelings of familiarity: The discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **26,** 547–565.

Wilson, R. S., Bacon, L. D., Kramer, R. L., Fox, J. H., & Kaszniak, A. W. (1983). Word frequency effect and recognition memory in dementia of the Alzheimer type. *Journal of Clinical Neuropsychology*, **5,** 97–104.

Wixted, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **18,** 681–690.

(Received November 14, 2000)
(Published online October 3, 2001)