Brain and Language **78,** 17–42 (2001)
doi:10.1006/brln.2000.2436, available online at http://www.idealibrary.com on IDEAL®

Assessing Working Memory and Language Comprehension in Alzheimer's Disease

Maryellen C. MacDonald, Amit Almor, Victor W. Henderson, Daniel Kempler, and Elaine S. Andersen

University of Southern California

Published online April 2, 2001

Studies of language impairments in patients with Alzheimer's disease have often assumed that impairments in linguistic working memory underlie comprehension deficits. Assessment of this hypothesis has been hindered both by vagueness of key terms such as ''working memory'' and by limitations of available working memory tasks, in that many such tasks either seem to have little relationship to language comprehension or are too confusing or difficult for Alzheimer's patients. Four experiments investigated the usefulness of *digit ordering,* a new task assessing linguistic working memory and/or language processing skill, in normal adults and patients with probable Alzheimer's disease. The digit ordering task was shown to be strongly correlated with the degree of dementia in Alzheimer's patients. The task correlated with measures of language processing on which patients and normal controls performed differently. The results are interpreted as indicating that linguistic representations, linguistic processing, and linguistic working memory are intertwined, such that a deficit of one (e.g., working memory) cannot be said to ''cause'' a deficit in the other. The implications of this approach are explored in terms of task demands in comprehension and memory measures, and interpretation of previous results in the literature.

Key Words: Alzheimer's disease; working memory; comprehension; sentence processing.

Some of the most dramatic cognitive changes in patients with Alzheimer's disease (AD) are changes in memory. Memory declines encompass spatial memory, such that patients may become disoriented only a few feet from their homes, semantic long-term memory, such that patients forget the names of familiar people and objects, and working memory, so that patients have difficulty following the thread of a conversation. A much more subtle impairment in these patients is an impairment of sentence comprehension, as evidenced by impaired performance in some sentence processing tasks (Grober & Bang, 1995; Rochon, Waters, & Caplan, 1994; Waters, Caplan, & Rochon, 1995). The subtle nature of this impairment, however, and in particular the fact that its detection requires careful testing, has sparked considerable debate regarding its origin. Because impairments in working memory are so dramatic (Baddeley,

This work was supported by a grant from the National Institutes of Health to Maryellen MacDonald (NIH RO1 AG11773), by a pilot grant to Maryellen MacDonald from the NIH-funded project, ''Multiple Disciplinary Approaches to Biogerontology,'' Caleb E. Finch, PI, and by the Alzheimer's Disease Research Center at the University of Southern California (NIH AG05142), Caleb E. Finch, PI. For additional referral of patients, we acknowledge Dr. Helena Chui, Co-Director, Alzheimer's Disease Diagnostic and Treatment Center, Rancho Los Amigos National Rehabilitation Center (DHS 94-20356).

Address correspondence and reprint requests to Maryellen C. MacDonald, Hedco Neuroscience Building, University of Southern California, Los Angeles, CA 90089-2520. Fax: 213 740–5687 E-mail: mcm@gizmo.usc.edu.

0093-934X/01 $35.00

Della Sala & Spinler, 1991), some researchers have suggested that the language impairments in AD stem from reduced working memory capacity in these patients (e.g., Rochon et al., 1994). Other researchers have suggested that AD patients have lost core linguistic knowledge in addition to having compromised working memory function (Grober & Bang, 1995). This debate concerning whether language processing impairments stem from loss of linguistic knowledge or from loss of computational capacity is echoed in aphasia research (Martin, Wetzel, Blossom-Stach, & Feher, 1987). In all cases, researchers have adopted a number of assumptions about what linguistic working memory is, including crucially the assumption that knowledge of language and working memory capacity can be independently damaged. We will explore an alternative position here, in which linguistic knowledge, processing operations, and processing capacity are inextricably intertwined. This position has enabled us to devise new ways of interpreting relationships between working memory tasks and comprehension tasks and consequently has led us to develop new ways of assessing working memory in AD patients. We begin with an analysis of how working memory tasks are designed and interpreted.

Working Memory Tasks

A number of measures of working memory capacity have been employed in the study of normal and impaired populations, including various word and digit spans, in which participants are presented with a string of items for immediate recall either in the order presented or in reverse order. One of the best known measures of linguistic working memory for the normal population is Daneman and Carpenter's (1980) *reading span* task, in which participants read sentences aloud and remember the sentence-final words for later recall. This task correlates well with a number of measures of language comprehension (Daneman & Carpenter, 1980; Daneman & Merikle, 1996; Just & Carpenter, 1992; but cf. Waters & Caplan, 1996). Daneman and Carpenter interpreted these correlations as indicating that the reading span task taps the linguistic working memory capacity that is used in on-line language comprehension. In their view, the key feature that distinguishes reading span from simple word or digit span tasks is that, unlike these simple span tasks, reading span has both a processing (reading aloud) and a storage (remembering words) component.

The reading span task has been successfully administered to aphasic patients (Miyake, Carpenter, & Just, 1994), though the interpretation of the results is controversial (Caplan & Waters, 1995; Martin, 1995; Miyake, Carpenter, & Just, 1995). However, this task and its auditory equivalent, listening span, both seem quite poorly suited to assessing the working memory capacity of Alzheimer's patients. Our own attempts to use these tasks with even mildly impaired AD patients have yielded very inconsistent results.

This apparent insensitivity of the otherwise useful reading/listening span tasks is quite unfortunate. AD patients are often described as having working memory impairments, and so a highly sensitive assessment of working memory could be quite illuminating concerning the cognitive deficits in AD. We therefore attempted to devise new tasks that would assess performance to a fine enough degree that we could not only distinguish impaired from unimpaired populations but could also distinguish milder from more moderately impaired AD patients.

The first step of our investigation was therefore to scrutinize the various tasks that AD patients can and cannot do in an attempt to understand why reading span scores are so poor in these patients. Our conclusion is that patients do poorly on this task simply because they have tremendous difficulty understanding and remembering instructions to do two things at once—reading out loud and remembering sentence-final words (Baddeley et al., 1991). The same can be said for the auditory equivalent of reading span, the listening span task (Daneman & Carpenter, 1980), in which participants hear sentences, judge whether they are true or false, and remember sentence-final words. Our experience using these tasks indicates that many patients forget one or another part of the task; for example, on some trials they forget to read aloud or that they are supposed to be attending to the sentence-final words. This analysis suggests that if we could retain the crucial features of reading/listening span while making the task less complicated, we might be able to develop a useful assessment for AD patients.

According to Daneman and Carpenter's original intuition, the crucial feature of their span tasks is that they consist of two different tasks, an actual sentence-processing task, and a memory task. This dual-task approach was inspired by Baddeley's formulation of the notion of working memory as having both a storage and processing component (Baddeley, 1986; Baddeley & Hitch, 1974). Of course this dual-task aspect is precisely what we suspect causes AD patients to fail to perform the tasks, and so it initially appears quite difficult to capture both storage and processing components in a single task that is doable by these patients.

An Alternative View

Fortunately, Daneman and Carpenter's intuition is not the only way to think of span tasks and why they correlate with other comprehension tasks. A different interpretation is offered by MacDonald and Christiansen (in press), who argued that the reading span task is simply a measure of language-processing skills, like measures of lexical decision latency, sentence–picture matching, and reading times (for similar points see Ericsson & Kintsch, 1995; Martin, 1995). MacDonald and Christiansen suggested that there are reliable correlations between performance on the reading span task and performance on other comprehension measures because all of these tasks, independent of whether they are called ''working memory tasks'' or ''comprehension tasks,'' rely on many of the same language comprehension processes.

This alternative interpretation leads to a rather different view of what it would mean to have a sensitive task of ''linguistic working memory'' for AD patients. On this view, some tasks termed ''working memory tasks'' and some tasks termed ''comprehension tasks'' might be good diagnostics of general language comprehension abilities, but only to the extent that the tasks capture crucial aspects of the comprehension process. Simple digit span tasks and lexical decision tasks, for example, are not likely to be informative about broad language comprehension abilities, because they each tap a very narrow subset of the processes that are involved in language comprehension. Indeed, there is good evidence that simple span tasks, which appear to rely largely on phonological storage mechanisms, are not well correlated with comprehension abilities in normal (Daneman & Carpenter, 1980) and impaired populations (Martin, 1995; Waters, Caplan, & Hildebrandt, 1991). The reading span task, however, requires a broad range of language comprehension skills involved in reading aloud, and is therefore likely to correlate well with general comprehension performance. Similar broad ''comprehension'' tasks, like reading and answering comprehension questions, would also be good predictors of general linguistic abilities, because they also draw on a broad range of comprehension skills. Thus working memory tasks have no privileged status in this account. Instead, they are just one kind of comprehension task which might prove useful in predicting individual differences in comprehension abilities. By this account, it should be possible to develop other tasks that assess comprehension ability as well as the reading and listening span tasks, possibly without complex instructions to do two or more things at once, which is too challenging for the AD patients.

We accordingly developed such a task. This task, *digit ordering*, involved hearing a series of digits and immediately reporting them back in ascending numerical order (Cooper, Sagar, Harvey, & Sullivan, 1991, report a similar task). For example, a participant might hear the digits 7, 2, 4, 9, 1, for which the correct response was 1, 2, 4, 7, 9. The task seems on the surface to have little in common with language comprehension, and certainly it has no overt sentence processing component, in contrast to the reading and listening span tasks. This surface dissimilarity is deceptive, however, and the task does seem to share certain key features with language comprehension. In order to perform the task accurately, a participant must first take an auditory signal (the digits) and convert it to phonological and lexical semantic representations. These semantic representations are then used to formulate a message for response (the digits in ascending numerical order), and phonological information is likely to be used to keep track of some digits while others are being ordered. Finally, the resulting message must be assembled for speech production and articulated. MacDonald and Christiansen (in press) have suggested that this simultaneous activation of phonological and semantic information is an important step in sentence comprehension. Thus although the sequence of events in the digit ordering task is clearly not exactly what goes on during a conversation, it does draw on a range of language comprehension and production processes to a much greater degree than do simple span tasks. For example, the most similar commonly used span task is backward digit span, in which a participant hears a series of digits and reports them back in reverse order. It is likely that some semantic information, in addition to phonological information, is utilized in immediate recall tasks such as backward span (e.g., Martin & Lesch, 1996). However, the degree of semantic involvement is likely to be smaller in the backward span task than in the digit ordering task, because in the latter task, semantic information (the meaning of the individual digits) is not just a mnemonic aid, it is the key information that must be accessed in order to perform the task.

Crucially, the digit ordering task draws on these language processes not through a complex dual-task paradigm but with extremely simple instructions based on the readily understood concept of putting numbers ‘‘in order.’’ In Baddeley’s (1986) terms, the task has both storage (phonological activation of the digits) and processing (reordering the digits) components, but participants are not explicitly instructed to do two different things at once. The simple instructions appear to make the digit ordering task well suited to use with AD patients.

If it proves to be sensitive to AD patients’ language-processing abilities, the digit ordering task can be viewed in several ways. From a traditional working memory perspective, digit ordering has both storage and processing components (albeit not an explicit ‘‘sentence processing’’ component as with reading or listening span) and should therefore tap linguistic working memory, which ought to correlate with comprehension abilities in AD patients. By our preferred perspective, digit ordering offers no special insight into ‘‘working memory’’ but nonetheless contains a number of subcomponents of language comprehension and production, and thus should be a useful assessment of comprehension abilities, particularly the ability to activate and manipulate phonological and semantic information from auditory input.

Of course any interpretation of relationships between the digit ordering task and more traditional language comprehension and working memory tasks awaits a demonstration that the digit ordering task is actually sensitive to variations across individuals. The following experiments are designed to provide this demonstration. First, Experiment 1 compares the digit ordering task to the reading span task in young normal adults, the population where reading span has been most extensively used. Experiment 2 then investigates the use of digit ordering in AD and normal elderly populations. This study compares digit ordering and a number of common neuropsychological tests and investigates the extent to which various tasks correlate with overall measures of dementia. Two additional experiments are then presented that investigate the relationship between digit ordering performance and several different types of language-processing tasks in AD and normal elderly participants. Armed with evidence from these experiments about the behavior of the ordering task and its relationship to other tasks in several different populations, we return to the issues of interpretation in the General Discussion.

EXPERIMENT 1. DIGIT ORDERING VS. READING SPAN IN YOUNG HEALTHY ADULTS

By our analysis, the combination of semantic processing and phonological storage that is required by the digit ordering task resembles the language comprehension processes that are tapped much more explicitly by the reading span task (Daneman & Carpenter, 1980). Our first experiment was undertaken to test this claim by examining the relation between performance in the digit ordering task and the reading span task. To set a context against which the relation between the two tasks could be assessed, we administered the reading span task twice, on different dates and with different items, and compared the relation between digit ordering and reading span to the relation between reading span measured on two different occasions. We expected the two reading span assessments to correlate reasonably well with one another. For example, Waters and Caplan (1996) found that two administrations of the reading span task about 2 months apart correlated with one another at .41. Correlations between digit ordering and reading span should be of similar magnitude if in fact the two tasks pose similar processing demands.

Method

Participants. Thirty-six University of Southern California psychology undergraduate students volunteered for course credit or were paid for participation.

Materials. The original stimulus list from the Daneman and Carpenter (1980) reading span task were modified to allow testing on two separate occasions.[1] The materials were structured so that the task increased in difficulty through the testing session. At the easiest level, two sentences were presented before sentence-final words were recalled, and additional sentences were added in later blocks. The original materials contained five trials at difficulty levels of two through five sentences, and there were an additional three trials with six sentences each. Because participants almost never succeed in this task with six-sentence trials, we split the original set of materials into two lists that had three trials each at levels 2–5.

Digit ordering trials were developed using the digits 1–9. Difficulty ranged from two to six digits to order, with four trials at each level. The digits were sampled equally often across all levels of difficulty.

Procedure. Participants were tested in three separate sessions, each of which was separated from the other sessions by a minimum of one week. The reading span task was administered in the first and third sessions; assignment of reading span list to testing session was counterbalanced across participants. The digit ordering task was administered in the second session. All participants were tested on the same digit ordering list.

The reading span task was administered as described in Daneman and Carpenter (1980), with blocks of increasing difficulty as the task progressed. The major difference from their method was that there were only three trials at each difficulty level, and testing was stopped when participants made an error in more than one trial at a given level. Digit span testing proceeded through increasing levels of difficulty, similar to the reading span task. Digits in each trial were presented auditorily at the rate of about one

[1] We thank Patricia Carpenter for providing us with these items.

digit per second. Participants were instructed to report the digits in ascending numerical order. At each testing level, testing was stopped if the participant failed to sequence the digits in two or more trials correctly (out of a total of four trials at each level). To allow a straightforward comparison of performance in the two tasks, reading and digit ordering scores were analyzed using the percentage of correct trials at each testing session.

Results and Discussion

Table 1 summarizes performance in the three sessions. Participants had higher reading span scores in the last session than in the first, $t(34) = 2.83$, $p < .01$, presumably reflecting increased familiarity with the task. To assess the relation between performance in the three sessions, we calculated the correlation between participants' performance in those sessions. Performance of the reading span task in the two sessions was reliably correlated, $r = .54$, $F(1, 34) = 14.09$, $p < .01$. This result is generally consistent with the correlation of .41 that Waters and Caplan (1996) reported for two assessments of the reading span task on different occasions separated by several months. Correlation between our participants' performance on the digit ordering task and the reading span task (averaged across the two sessions) was also of similar magnitude, $r = .45$, $F(1, 34) = 8.39$, $p < .01$.

These results suggest that the digit ordering task and reading span task have similar abilities to assess individual differences in linguistic abilities or linguistic working memory in the young normal population. Of course we are not claiming that the task does not tap some other abilities beyond strictly linguistic ones. For example, the digit ordering task might have a visuo-spatial component, evidenced by some participants' reports that they attempted to visualize some numbers while performing the task. Other span tasks may similarly involve visualization; some reading span participants report that they try to visualize the to-be-remembered words while reading sentences aloud. We do not see the potential for spatial or other components in the digit ordering task as a disadvantage; our claim is that the task is a useful assessment of linguistic abilities, not that it is specifically unrelated to non-linguistic abilities. Moreover, normal conversational language processing draws on a variety of abilities and modalities, including spatial abilities to integrate visual and acoustic information, as demonstrated by many current head-mounted eyetracking studies in which participants listen to instructions to move objects on a table (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Thus if digit ordering does have spatial or other traditionally nonlinguistic components, we can speculate that some of these may also contribute to the task's sensitivity to language-processing ability. These issues await further investigation; at this point it is simply clear that the digit ordering task correlates well with reading span, a traditional test of linguistic working memory.

Despite these promising findings, our results did suggest that the digit ordering task could benefit from some modification for use with younger adults, in that it might not have been hard enough for our participants. Seven of the 36 participants

TABLE 1
Reading Span and Digit Ordering Performance (in Number of Correct Trials) in Experiment 1

	Mean	Range	*SD*	% Correct
Reading span session 1	4.61	2–9	2.06	38.4
Reading span session 2	5.53	3–12	2.05	46.1
Digit ordering	18.1	14–20	1.59	90.5

Note. Total possible correct trials for the reading span task was 12 trials; total possible correct for digit ordering was 20 trials.

scored the maximum possible correct (20 trials) on the digit ordering task. By contrast, only one participant scored the maximum number correct (12) in one session of the reading span task. Thus the addition of a block of digit ordering trials with seven digits each might improve sensitivity for young normal participants. Some pilot testing with seven digits suggested that this modification should be accompanied by the inclusion of extra digit names in the lists, such as ''ten,'' ''eleven,'' or possibly ''zero,'' because with only the digits 1–9 in presentation lists and seven digits to recall, a few participants developed a strategy of detecting which two digits were missing from the presentation list. No participants used this strategy when at least three potential digits were not included in a trial (e.g., six-digit trials sampled from the digits 1–9, seven-digit trials sampled from 1–10), and our exploration of this strategy suggests that it is extremely ineffective when three or more digits are not included in a trial. Thus use with young normal participants awaits further testing, but as will become clear below, the addition of harder trials is not necessary for testing elderly adults or AD patients, as these participants rarely approach ceiling on this task.

EXPERIMENT 2. DIGIT ORDERING TASK IN ELDERLY ADULTS AND AD PATIENTS

Given the relationship between digit ordering and reading span established in Experiment 1, our next goal was to investigate whether the digit ordering task was simple enough for AD patients to perform, and if so whether it provided a useful measure for assessing the general capabilities of these patients. To test the utility of the digit ordering task in this population, we added the task to a large battery of standardized tests administered to elderly healthy and demented participants. If AD patients are impaired in their language comprehension ability, then the extent of their impairment should be more strongly reflected in their digit ordering performance than in their performance of any other task that requires only storage of phonological information (e.g., simple digit span).

Method

Participants. One hundred five elderly adults were drawn from a pool of participants in a longitudinal study of aging and dementia at the Alzheimer's Disease Research Center at USC. All had undergone standardized diagnostic clinical and neuropsychological evaluations. Participants were assigned to groups based on their Clinical Dementia Rating (CDR) (Berg, 1989). Sixty-four participants were classified as healthy elderly (CDR $=$ 0). Sixteen participants presented evidence of mild cognitive impairment and were classified as having questionable dementia (CDR $=$ .5), 17 were mildly demented (CDR $=$ 1), and 8 were moderately demented (CDR $=$ 2). Participants in these latter two groups met the clinical criteria for probable Alzheimer's disease (McKhann, Drachman, Folstein, Katzman, Price, & Stadlan, 1984). As can be seen in Table 2, normal participants were slightly younger and had slightly more years of education than participants in the other groups. The age differences were not reliable across the groups, $F < 1$. However, years of education did differ across groups, $F(3, 104) = 3.58$, $p < .05$. Subsequent analyses indicated that only the healthy elderly group and moderate dementia group differed in this measure, a result that should be interpreted cautiously, given the small sample ($N = 8$) in the moderate AD group.

Table 2 also indicates general measures of cognitive function and mood. The Mini-Mental State Exam (MMSE) (Folstein, Folstein, & McHugh, 1975) contains a variety of questions assessing memory, language, spatial ability, and factual knowledge. Despite the variety of questions, the performance relies heavily on language abilities, both because of the number of language-related questions and also because adequate language processing is necessary to understand and answer all questions on the test. As expected, MMSE scores were highest in the healthy elderly group and worst in the moderate AD group, $F(3, 104) = 91.46$, $p < .0001$. Pairwise tests indicated that all four participant groups were significantly different from one another on the MMSE, all F's > 4.0, p's $< .05$. Mood, as measured by the Geriatric

TABLE 2
Participant Characteristics, Means and (Standard Deviations), Experiment 2

Clinical Dementia Rating (CDR)	*N*	Age, years	Education, years	Mood (GDS)	Cognitive function (MMSE)
CDR = 0 (healthy elderly)	64	77.44 (7.71)	15.75 (2.27)	3.74 (4.16)	28.31 (1.40)
CDR = .5 (questionable dementia)	16	78.63 (8.21)	15.38 (2.68)	8.38 (5.94)	27.00 (2.19)
CDR = 1 (mild AD)	17	78.89 (6.95)	14.72 (2.91)	5.89 (5.23)	20.50 (3.52)
CDR = 2 (moderate AD)	8	79.63 (5.26)	13.00 (2.07)	6.00 (3.85)	17.50 (4.21)

Note. The Geriatric Depression Scale (GDS) is a 30-point scale with higher values indicating greater depression. The Mini-Mental State Exam (MMSE) is a 30-point scale with lower values indicating increased dementia.

Depression Scale (Yesavage et al., 1983), varied with dementia category, $F(3, 104) = 4.61$, $p < .005$. Pairwise tests indicated that only the healthy and questionable dementia groups differed reliably from one another in mood, $F(1, 78) = 4.19$, $p < .01$.

Procedure. As part of their standardized evaluation in the longitudinal project, participants were periodically tested by a trained research associate on a neuropsychological battery designed to assess cognitive domains commonly affected by aging or dementia. The scores presented here were from the first testing session completed by a participant that included the digit ordering task. In this battery, the digit ordering task consisted of a total of 10 trials (2 at each level) rather than 20 trials as in Experiment 1. Other tests in the battery included the forward and backward digit span from the revised Wechsler Adult Intelligence scale (Wechsler, 1987); tests of short- and long-term memory (total score on three immediate recall trials of at 10-item supraspan word list used by the Consortium to Establish a Registry for Alzheimer's Disease (Morris, Mohs, Rogers, Fillenbaum, & Heyman, 1988), and the delayed recall score of words from the same list); picture naming (Boston Naming Test—Goodglass, Kaplan, & Weintraub, 1982); phonological fluency (a timed test in which participants report as many words as possible that begin with a given phoneme); semantic fluency (reporting as many animal names as possible in a timed interval; Morris et al., 1988); and a spatial task, judgment of line orientation (Benton, Hannay, & Varney, 1975). The mean performance of each dementia rating group is shown in Table 3. Because percentage correct cannot be calculated for the fluency measures, all data are reported as raw scores, not percentages.

Results and Discussion

We performed several different analyses to determine how the digit ordering task compared to the more established neuropsychological tests. First, we computed the correlation matrix between the eight tasks shown in Table 3, plus age, education, and mood. This matrix is presented in Table 4. The table shows that age had a small

TABLE 3
Group Means and (*SD*s) on Neuropsychological Tests Administered in Experiment 2

	Clinical dementia rating			
Test (total possible correct)	Healthy elderly	Questionable dementia	Mild dementia	Moderate dementia
Digit ordering (10)	8.31 (1.30)	7.44 (1.55)	4.22 (2.86)	3.00 (2.87)
Forward digit span (14)	6.55 (1.28)	6.25 (1.43)	5.67 (1.18)	6.00 (1.07)
Backward digit span (14)	7.13 (2.37)	6.88 (2.94)	4.61 (2.06)	3.87 (2.80)
Immediate recall (30)	20.33 (4.09)	17.63 (5.24)	9.61 (4.88)	8.63 (3.78)
Delayed recall (10)	6.89 (1.75)	5.00 (3.14)	0.44 (0.70)	0 (0)
Boston Naming Test (60)	56.69 (4.47)	52.94 (7.71)	36.50 (9.56)	31.25 (13.57)
Phonological fluency (no limit)	42.03 (11.19)	39.00 (12.26)	28.71 (11.66)	22.0 (13.36)
Semantic fluency (no limit)	19.38 (4.24)	15.00 (5.33)	8.33 (3.27)	57.75 (2.38)
Judgment of line orientation (30)	24.55 (5.14)	23.06 (5.31)	17.28 (7.70)	13.88 (10.48)

Note. Higher scores reflect better performance on all measures.

TABLE 4
Correlation Matrix for Age, Education, Mood, and Neuropsychological Variables in Experiment 2

Measure	Age	Education	Mood (GDS)	MMSE	Digit ordering	Forward digits	Backward digits	Immediate recall	Delayed recall	Boston Naming	Phonological fluency	Semantic fluency	Judgment of line orientation
Age	1												
Years of education	−.11	1											
Mood (GDS)	.125	−.154	1										
Mini-Mental State (MMSE)	−.08	.276	−.135	1									
Digit ordering	−.178	.188	−.128	.8	1								
Forward digit span	−.13	.21	−.166	.28	.315	1							
Backward digit span	−.21	.227	−.062	.506	.469	.488	1						
Immediate recall	−.14	.223	−.194	.769	.626	.381	.541	1					
Delayed recall	−.187	.291	−.224	.788	.681	.294	.503	.859	1				
Boston Naming Test	−.154	.404	−.216	.772	.633	.204	.354	.673	.743	1			
Phonological fluency	−.057	.222	−.076	.579	.554	.335	.522	.636	.603	.459	1		
Semantic fluency	−.138	.226	−.206	.687	.585	.252	.408	.696	.741	.711	.605	1	
Judgment of line orientation	−.2	.402	−.068	.512	.499	.182	.407	.412	.446	.466	.321	.429	1

Note. GDS, Geriatric Depression Scale; MMSE, Mini-Mental State Exam.

negative relationship to the measures in this sample, years of education was moderately positively correlated with performance on the test battery, and geriatric depression scores (in which higher scores indicate greater depression) were mildly negatively correlated with performance. Most of the cognitive tests were positively correlated with one another, a common result for these standard measures when used across a very broad range of cognitive abilities. In particular, scores on the Mini-Mental State Exam exhibited fairly strong positive correlations with most measures, as would be expected from this general test of cognitive performance.

To further explore the relationships between these measures, we next performed a hierarchical regression with an *F* value to enter of 4.0, in which nine cognitive tests shown in Table 3 predicted MMSE, the general measure of dementia severity. Because the results in Table 4 indicate that age, years of education, and mood (GDS) were mildly associated with MMSE and with some other tasks, all three of these factors were initially forced into the regression equation. These three factors combined to account for 8.7% of the variance in MMSE, a reliable amount $F(3, 101) = 3.20$, $p < .05$. Only three of the nonforced factors entered the regression. The first was digit ordering, which raised the total variance in MMSE accounted for to 66.2%; the second was immediate recall, which accounted for an additional 11% of the variance; and the third was the Boston Naming Test, which accounted for an additional 3.8% of variance, such that the total variance accounted for by these three factors plus age, mood, and years of education was 81%. An identical pattern was obtained when age, mood, and years of education were not forced into the regression, in that the same three factors entered the regression in the same order with virtually the same amount of variance accounted for, and no other factors entered the regression.

These results show that digit ordering scores were strongly correlated with the degree of dementia in this sample. We suspect that the immediate recall task entered the regression because the MMSE contains several similar tasks, including sentence and word repetition. The Boston Naming Test, a picture-naming measure of patients' semantic and lexical representations, is not a test of what is conventionally considered working memory and presumably reflects semantic abilities in this sample.

Because Alzheimer's dementia is often not viewed as a continuum of normal aging (Berg, 1985), we also conducted a second hierarchical regression that excluded normal control participants. This analysis investigated whether the digit ordering task can predict MMSE within the questionable, mild, and moderate dementia groups. As before, control variables were first forced into the regression analysis before other neuropsychological scores were allowed to enter. Age, education, and mood accounted for 7.8% of the variance in this group, which, given the smaller sample size, was not a reliable amount, $F(3, 37) = 1.05$, ns. The same three factors as in the first regression entered this regression in the same order as before. Digit ordering accounted for an additional 55.1% of the variance in MMSE, immediate recall contributed an additional 13.1%, and the Boston Naming Test contributed 3.7%, bringing the total variance accounted for to 80.3% in this sample. This result shows that digit ordering not only is useful in distinguishing healthy elderly adults from those with dementia, but also is very sensitive to the degree of dementia within the range of questionable to moderate dementia.

An important question in any hierarchical regression when the potential factors are positively correlated concerns whether the order of factors entering the regression reflects luck as much as true differences between the factors. That is, if two factors have very similar correlations with the dependent variable, small differences in participants' performance or measurement error could have changed which factor entered the regression. As is clear from Table 4, other factors besides digit ordering are strongly correlated with MMSE. Our goal in this experiment was to evaluate the effectiveness of digit ordering. Having found it to be an easily administered, sensitive measure in this population, we advocate its wider usage in test batteries, but we are certainly not suggesting that other tasks should now be abandoned. Indeed, given our claim that language processing comprises a number of component skills, we (like others) administer a wide variety of standardized tests.

A related question concerns why the digit ordering task works as well as it does in this sample. We have suggested that one reason is that digit ordering is sensitive to a number of subcomponents of language processing, and that the MMSE performance is heavily influenced by language abilities. A second reason is the task instructions are easy to understand, yielding good compliance, and that the degree of difficulty is well-scaled for the elderly and AD populations (unlike for young adults in Experiment 1, for which we probably made the task too easy). The absence of ceiling effects in the healthy elderly and the absence of floor effects in the mild–moderate range clearly contribute to the task's sensitivity. The task is not completely immune to floor effects, however. About half of the moderately demented patients scored only one trial correct (although none got zero trials correct), suggesting that the task may have limited usefulness in the moderate–severe dementia range. As this degree of dementia often leaves patients too impaired to perform many language tasks, it is not likely that floor effects in digit ordering will pose a significant problem in studies of language processing in AD.

To this point we have shown that the digit ordering task compares well to a standard measure of linguistic working memory (reading span) in young adults, and that it is a sensitive measure of dementia severity in a sample of elderly adults and AD patients. These results show the promise of the task, but a key component is missing, namely an investigation of the relationship between digit ordering performance and language-processing abilities. The next two experiments address this issue in samples of elderly normal and demented adults. Some of this research has been reported elsewhere with a more abbreviated presentation of working memory tasks (Kempler, Almor, Tyler, Andersen, & MacDonald, 1998; Almor, Kempler, MacDonald, Andersen & Tyler, 1999). These studies were designed to provide a careful analysis of various language comprehension tasks and language-processing abilities in AD patients. Here our focus is somewhat different, in that we investigate the extent to which the digit ordering task provides a useful predictor of performance on language-processing tasks. To foreshadow our results somewhat, we found that digit ordering correlated well with performance on some, but not all, language comprehension tasks in AD and normal elderly participants. Our analysis of this variation in correlations offers some insights into the effect of task demands on performance and the nature of working memory. For ease of exposition, we present the research as Experiments A and B; the use of letter names serves as a reminder that some of these data have been presented elsewhere and are reviewed and augmented here to emphasize the role of digit ordering in understanding AD patients' linguistic performance.

EXPERIMENT A. ORDERING TASKS AND ON VS. OFF-LINE LANGUAGE PROCESSING

Our first study (Kempler et al., 1998) compared performance of AD and normal elderly participants on sentence comprehension in two different tasks. One task was grammaticality judgment—participants heard a sentence and said whether they thought it was ''good English'' or not. The second task was an on-line cross-modal naming task in which participants heard an auditory sentence fragment and then named a visual target word that appeared on a computer screen at the offset of the auditory stimulus. The latency to name this target is related to the extent to which it constitutes a grammatical and sensible continuation to the auditory sentence fragment (Tyler, 1992). Good comprehension of the sentence structures under investigation should yield large differences in naming times between conditions in which the target is a good continuation and conditions in which the target is semantically or syntactically anomalous. In this way, we can measure processing difficulty as the comprehension process is ongoing. Moreover, several researchers have argued that these on-line tasks have lower working memory demands than tasks such as grammaticality judgment or sentence–picture matching, because the comprehender does not have to hold the sentence in memory while making some evaluative judgment about it (see Tyler, 1992, for review). The question for the present article concerns whether the digit ordering task is related to performance on both on- and off-line language-processing tasks, and what any relationships signify for theories of working memory and language comprehension in these patients.

Method

Participants. Eleven AD patients and nine healthy elderly controls participated. See Table 5 for participant information. AD patients were recruited through the USC Alzheimer's Disease Research Consortium or through the affiliated Alzheimer's Disease Diagnostic and Treatment Center at Rancho Los Amigos Medical Center. All patients met criteria for probable AD as described in Experiment 2. There were no reliable differences in age ($t(18) = 1.38$, $p > .10$) or education ($t(18) = -1.60$, $p > .10$) across the two groups.

Materials and procedure. For the cross-modal naming study, 80 grammatical/ungrammatical sentence pairs were constructed, such that the ungrammatical versions differed from the grammatical versions by the addition, deletion, or substitution of a single word. Two types of grammatical violations were tested: (1) subject verb agreement, in which a singular or plural subject was followed by either <u>was</u> or <u>were</u>, e.g. . . . *the students *<u>was/were</u>* (words that were presented as visual targets are underlined in these examples) and (2) verb transitivity violations, in which transitive verbs either were or were not followed by a direct object (*Bill poured/*poured with <u>cereal</u>*), and intransitive verbs either were or were not followed by a required preposition (*Susan went to/*went <u>meetings</u>*). Thus for the agreement items, a change of target created the grammatical or ungrammatical continuation, whereas the grammaticality manipulation in the transitive items was contained in the auditory fragment, just before the appearance of the target word.

The stimuli were digitally recorded. For the sentence fragment, entire sentences were recorded in order to create natural intonation, and extra material was digitally excised in order to create a fragment.

Participants were tested in a quiet room in their home or at USC, sitting in front of a Macintosh computer. In each trial, participants heard a neutral context sentence and the edited sentence fragment. The visual target appeared on the computer screen in large font at the offset of the auditory fragment. Participants were instructed to name the target as quickly and accurately as possible. To encourage participants to attend to the meaning of the sentence and integrate the visual target with the auditory fragment, a secondary task was administered following the naming of the visual target. Healthy control subjects pressed a key to indicate whether the visual target was a good continuation of the auditory fragment. Pilot testing indicated that the AD patients could not perform both this task and the primary naming task, so AD patients were asked comprehension questions about the auditory fragment on 20%

TABLE 5
Participant Information for Experiments A–B

	MMSE (*SD*)	Age, years (*SD*)	Education, years (*SD*)
Experiment A participants			
AD ($n = 11$)	21.0 (2.0)	80.7 (6.8)	15.0 (2.3)
Controls ($n = 9$)	29.1 (1.1)	77.0 (4.4)	16.8 (2.8)
Experiment B participants			
AD ($n = 10$)	21.0 (3.8)	82.0 (4.3)	15.0 (2.5)
Controls ($n = 10$)	29.4 (1.1)	78.0 (4.2)	17.0 (1.9)

TABLE 6
Percentage Correct (and *SE*) on Working Memory Tasks and Normalized and Raw Cross-modal Naming Times (and *SE*), for Agreement and Transitivity Stimuli, Experiment A

	AD patients		Healthy elderly	
Working memory measures				
Backward digit span	42.4 (5.1)		60.2 (7.6)	
1–40 by 3's	75.3 (12.0)		95.2 (3.1)	
Digit ordering	42.3 (8.0)		85.6 (5.0)	
	Normalized data	Raw times (ms)	Normalized data	Raw times (ms)
Subject–verb agreement				
Compatible target	−0.25 (.04)	1110 (101)	−0.26 (.07)	789 (64)
Incompatible target	0.27 (.04)	1355 (137)	0.28 (.07)	879 (69)
Transitivity				
Compatible target	−0.31 (.02)	1060 (79)	−0.31 (.05)	781 (53)
Incompatible target	0.37 (.02)	1285 (96)	0.34 (.07)	960 (95)

of the trials. In neither case was the secondary task scored but only served to direct participants' attention to the meaning of the auditory stimuli.

Participants were tested four times on four different counterbalanced lists, with at least 1 week between sessions for the AD patients, and 3 weeks for control participants. A grammaticality judgment task was administered in a fifth testing session. Forty sentences were adapted from the cross-modal stimuli, half from verb agreement items and half from transitivity items. Whereas the visual target that terminated the cross-modal naming stimuli did not always finish a sentence (e.g., *The students were*), the grammaticality judgment items were all complete sentences (e.g., *The students were very tired*). Half of the items of each type were grammatical, and half ungrammatical. Participants listened to a spoken sentence and were asked to judge whether it was ''good English'' or not.

Two standard measures of working memory were administered at various points during the five testing sessions. These were backward digit span and counting from 1 to 40 by threes (both from the revised Wechsler Memory Scale, Wechsler, 1987). We also administered the MMSE and the digit ordering task with 20 trials, as in Experiment 1.

Results

Working memory tasks.[2] Mean performance of the two groups on the working memory measures is shown in Table 6. Control participants performed much better than AD patients on digit ordering, $t(18) = 4.31$, $p < .001$, and performance was strongly correlated with the general cognitive measure MMSE, $r(18) = .74$, $p < .001$. By contrast, the two standardized tasks were only moderately successful at distinguishing AD and control participants. Controls were better than AD patients at backward digit span ($t(18) = 1.97$, $p < .05$) and did not differ in the 1–40 task ($t(18) = 1.46$, $p > .15$). The two tasks did not correlate with each other reliably $r(18) = .30$, $p > .20$. Correlations with MMSE were somewhat higher; for the 1–40 measure, $r(18) = .42$, $.05 < p < .10$, and for backward digit span, $r(18) = .48$, $p < .05$.

Grammaticality judgments. AD patients performed significantly worse on grammaticality judgments (mean 79.5% correct, $SD = .11$) than the healthy controls, who were near ceiling (95.6% correct, $SD = .04$), $t(18) = 4.22$, $p < .001$. All of the working memory tasks were reliably correlated with grammaticality judgment performance: for digit ordering, $r(18) = .60$, $p < .01$; for backward digit span, $r(18) = .46$, $p < .05$; and for 1–40, $r(18) = .46$, $p < .05$. A scatterplot of individual partici-

[2] Most of the data concerning digit ordering were not reported in the original Kempler et al. (1998) work.

pant scores on digit ordering and grammaticality judgment can be found in Appendix 1.

Cross-modal naming. Because the performance of the AD patients can vary substantially from one testing session to another, averaging raw response times from different sessions can be greatly misleading. This problem becomes especially severe after the deletion of error trials which could result in an artificial change in condition means (for example, error trial removal from a session in which a patient had long latencies would result in the incorrect shortening of the mean RT of the conditions to which the error trials belonged). To address this problem and to facilitate comparisons across the two participant groups, naming times were normalized using a z-transformation, and all analyses were performed on normalized data. The z-transformation was based on each participant's mean and standard deviation for trials from this experiment, computed separately for each session. Filler trials were not included in calculation of the mean and standard deviation for the transformation. The raw and normalized means of each group's performance in the different conditions are listed in Table 6.[3] An analysis of variance yielded a reliable main effect of grammaticality, such that naming times were shorter to visual targets that formed a grammatical continuation than to those that formed an ungrammatical continuation. This result was equally true for both patients and control participants, and held for both stimulus types (agreement and transitivity), and there was no interaction between grammaticality and these other factors. In other words, both control participants' and AD patients' naming times suggested good sensitivity to the grammatical violations that were tested in this study. Given this lack of difference between participant groups, it is not surprising that scores on the grammaticality judgment task, digit ordering, and the standardized measures were not reliably correlated with on-line sensitivity to grammatical violations (a difference score between ungrammatical and grammatical naming times), all p's $> .10$. A scatterplot of sensitivity scores (the difference of normalized RTs to grammatical and ungrammatical trials) and digit ordering scores for all participants is contained in Appendix 1.

Discussion

The results of Experiment A showed that in a language comprehension task in which healthy elderly adults and patients performed very differently (grammaticality judgment), the accuracy of performance was correlated with digit ordering performance. In the on-line task, however, there were no differences in the normal and patient groups, and thus there was less variance that digit ordering could account for. These results, beyond confirming the usefulness of the digit ordering task in distinguishing AD and healthy elderly participants, have implications for the role of working memory in language comprehension abilities, and the role of task demands in performance. The major question raised by the data in Experiment A is why AD patients and healthy elderly adults performed so similarly in the on-line task, when the other memory and comprehension tasks showed them to be so different in their abilities. Two alternative interpretations are possible, reflecting the alternative views of working memory that were discussed in the introduction. First, patients' success with the on-line cross-modal naming task may indicate that AD patients have intact syntactic processes but impaired working memory (Rochon et al., 1994). On this view, reduced working memory resources would naturally lead to impaired performance on working memory tasks, and it would also impair performance on off-line

[3] Note that the z-scores do not add to zero, as the deletion of error and outlier trials may lead to an unequal number of observations in each cell.

tasks requiring an explicit judgment, as such tasks are thought to burden working memory resources. The Experiment A data may thus reflect the superiority of on-line tasks to off-line ones (Tyler, 1992), in that on-line tasks can reveal patients' true language-processing capacities, unhindered by unnatural working memory demands. We will term this the ''intact on-line abilities'' hypothesis. A related perspective on the relationship between on-line and working memory tasks is provided by Caplan and Waters (1999). They have suggested that most on-line language comprehension tasks tap ''interpretive'' (early automatic parsing) processes, while off-line tasks and working memory tasks, presumably including digit ordering, tap ''post-interpretive'' processes that occur after language comprehension and are involved in judgment and decision-making. Thus off-line and working memory tasks correlate well because they both tap ''post-interpretive'' processes, in contrast to the on-line tasks, which do not. This approach does not necessarily predict that on-line tasks will reveal AD patients' ''true'' abilities, but like the ''intact abilities'' approach, it does predict that on- and off-line tasks assess qualitatively different stages of language comprehension. The approach is controversial, and several researchers have argued that the interpretive/post-interpretive dichotomy is inconsistent with language comprehension data (Just, Carpenter & Keller; 1996; MacDonald & Christiansen, in press).

An alternative interpretation to these positions arises from a close analysis of the task demands in the cross-modal naming task. For the stimuli that were used in Experiment A, very local associative information could influence naming times for the visual target. With the visual target *cereal*, for example, the grammatical auditory fragment ended with the word *pours,* while the ungrammatical ended with *with.* Even the most minimal sensitivity to cooccurrence information would be expected to yield shorter naming times to *cereal* in the grammatical than in the ungrammatical case; naming times therefore need not reflect full syntactic analysis of the sentences, or even processing of anything other than the last few words. In the grammaticality judgment task, however, the locus of any ungrammaticality was unpredictable, because complete sentences were presented rather than sentence fragments as in the cross-modal naming task. Participants therefore had to process the entire sentence to detect whether a grammatical violation was present. In this interpretation of the data, the cross-modal naming task is not a superior measure that reveals patients' true processing abilities; on the contrary, it is a task that allows participants to perform well with only shallow processing of the input and is therefore of limited interest, at least with the particular stimulus materials in used in Experiment A. We will term this view the ''shallow processing'' hypothesis.

We can extend this approach further to account for the relationship between grammaticality judgment and the digit ordering task. Exactly what information is used to perform grammaticality judgment is a matter of debate and may vary with syntactic structure and with the participant (see Allen & Seidenberg, 1999, for discussion). However, the task does demand at least some sentence-processing abilities, in contrast to the on-line task, for which very shallow processing of the input could yield normal-looking patterns of naming times. On this view, the grammaticality judgment and digit ordering tasks correlate well with one another because they share some language comprehension processes, such as converting an acoustic signal into a semantic representation, and assessing the serial order of words in the input.

These two alternative interpretations of the cross-modal naming task are tightly linked to particular views of working memory. According to the ''intact on-line abilities'' account, the cross-modal task reflects patients' true skills, while the grammaticality judgment task is contaminated by extraneous working memory demands. In the ''shallow on-line processing'' account, however, good on-line performance can be achieved without actually interpreting the stimulus sentences, such that the task could overestimate abilities. The more rigorous off-line task is a more accurate assessment of patients' sentence processing skills on this view. Moreover, on the analysis of digit ordering offered here, good tests of comprehension abilities (such as grammaticality judgment) should correlate with digit ordering not because the comprehension task ''uses'' the working memory being assessed with digit ordering, but rather because both tasks involve similar language comprehension processes.

The data from Experiment A do not distinguish between these two radically different interpretations of the on-line task, but Experiment B is designed to shed light on this question. Whereas the cross-modal naming stimuli in Experiment A relied on very simple, local relationships of adjacent or nearly adjacent words (subject–verb agreement and verb transitivity), the materials used in Experiment B involved long distance dependencies between pronouns and their antecedents. According to the ''intact on-line abilities'' hypothesis, in which on-line tasks reveal patients' true processing abilities unfettered by extraneous working memory demands, patients in Experiment B should again show good sensitivity to fragment–target compatibility in the cross-modal naming task. Moreover, we should find that once again performance on working memory tasks does not correlate well with on-line comprehension performance, on the view that on-line tasks are free from such extraneous memory demands. By contrast, the ''shallow processing'' hypothesis predicts that once we prevent patients from responding accurately via only shallow processing of adjacent words, patients' performance will be impaired compared to controls. Thus the pronominal reference materials, which cannot be integrated except by thorough processing of the discourse, should pose difficulties for AD patients. Furthermore, because we contend that the ordering tasks are a good assessment of language comprehension abilities, this hypothesis predicts that on-line performance with pronominal reference materials should correlate well with the digit ordering task.

EXPERIMENT B. DIGIT ORDERING AND PROCESSING REFERENCE

Almor et al. (1999) examined the on-line comprehension of pronouns and referentially dependent noun phrases in the cross-modal naming task, with two different types of stimuli. In the first set, pronouns were the visual targets following multisentence discourses, and in the second set, the visual targets were adjectives, so that integration of the adjective into the auditory sentence was dependent on having successfully interpreted the discourse, including referentially dependent expressions.

Method

Participants. Ten AD patients and 10 elderly controls were tested; participant information is contained in Table 5. Compared to the AD patients, the healthy controls were slightly younger, $t(18) = 2.00$, $p < .07$, and had slightly more years of education, $t(18) = 1.97$, $p < .07$. Because of these marginal differences, age and education were entered as covariates in all analyses.

Materials. Stimuli for the two cross-modal naming tasks consisted of short paragraphs containing several discourse participants. Examples of the pronoun target items are shown at the top of Table 7, and examples of the adjective target items are shown at the bottom of this table. Twenty paragraphs of each type were constructed.

The manipulation in the pronoun–target items affected whether the pronoun target formed a sensible continuation of the discourse, that is, whether the pronoun could sensibly refer to one of the entities introduced the first sentence of a three-sentence discourse. Crucially, information in the final sentence was not sufficient to distinguish compatible and incompatible visual targets. In the discourse in Table 7, for example, the auditory fragment *During the performance, the clown threw candy to . . .* does not indicate whether *him* or *them* is an appropriate pronoun; it is only through the introduction of *the children* in the first sentence that *them* becomes a good continuation and *him* becomes anomalous—the only grammatical interpretation of *him* is that it refers to some person not yet mentioned in the discourse.

TABLE 7
Pronoun and Adjective Target Stimuli, Cross-Modal Task, Experiment B

Pronoun Target Stimulus
The children loved the silly clown at the party. The show was very funny. During the performance, the clown threw candy to
Compatible visual target: him
Incompatible visual target: them
Adjective Target Stimulus
Full Noun Phrase Condition:
The housewife watched the clumsy plumber working under the sink. The housewife showed the plumber where the leak was. The housewife could not believe that the plumber was so
Pronoun Condition:
The housewife watched the clumsy plumber working under the sink. She showed him where the leak was. She could not believe that he was so
Visual target, both conditions: clumsy

Use of a pronoun to introduce a new discourse participant is not felicitous, and thus good comprehension of the entire discourse should yield shorter naming times to the coreferential target *them* compared to *him*.

The manipulation in the adjective target items occurred not at the visual target but within the discourse, which either made extensive use of pronouns in the second and third sentences after introducing two characters in the first sentence, or repeated the full noun phrases (e.g., *housewife* and *plumber* in the example in Table 7) in the second and third sentences. In contrast to the pronoun-target study, the visual target was the same word (adjective) in both versions of a stimulus, and all adjectives formed sensible continuations of the discourse. This manipulation is more subtle than in the previous studies, but naming times should reflect the ease of developing this discourse in the pronoun and repeated name conditions, and the ease of integrating the visual target with the context. English speakers typically find the pronoun versions more natural for this type of discourse, and normal young adult comprehenders have shorter processing times when pronouns are used for subsequent reference instead of repeated full noun phrases (Gordon, Grosz, & Gilliom, 1993). Thus intact language comprehension abilities should yield shorter naming times for adjective targets in the pronoun version than in the full noun phrase version.

Procedure. The cross-modal naming procedure was identical to that used in Experiment A. The participants were tested in four sessions, which included tests of digit ordering and MMSE.[4]

Results

Cross-modal naming. Normalized and raw naming times for AD and control participants are shown in Table 8. Analyses of these data found impairments in the AD patients relative to controls. With respect to the pronoun targets, although both the patients and the healthy controls were sensitive to the context appropriateness of those targets, the patients were much less sensitive than controls, yielding an interaction between target compatibility and participant group, $F(1, 18) = 29.78$, $p < .001$. Thus, in the same on-line cross-modal naming task used in Experiment A, patients processed pronouns less effectively than did healthy elderly adults.

With respect to the adjective targets, there were no main effects of participant group or target compatibility, but again there was a significant interaction between these factors, $F(1, 18) = 6.72$, $p < .05$. The nature of this interaction was that patients' on-line comprehension was better in the full noun phrase condition than in

[4] The digit ordering data were not reported in the original Almor et al. (1999) work. Instead, a similar ''month ordering'' task was reported, in which subjects heard lists of months of the year and had to report them back in ascending calendar order. We have found that the two ordering tasks correlate reliably, and that scores on month ordering are often somewhat lower than on digit ordering. This difference probably reflects the fact that the order of months is slightly less familiar than the order of digits, and that months are articulatorily more complex, with an average of 2.4 syllables each, compared to 1.1 syllables for the digits 1–9.

TABLE 8
Digit Ordering Scores, Normalized and Raw (ms) Cross-Modal Naming Times, and (Standard Errors), Experiment B

	AD patients		Healthy elderly	
Digit ordering (20 possible):	9.3 (1.48) (46.5%)		17.90 (.59) (89.5%)	
	Normalized data	Raw RT	Normalized data	Raw RT
Cross-Modal Naming				
Pronoun targets				
Compatible target	−.14 (.22)	858 (78)	−.40 (.13)	730 (45)
Incompatible target	.09 (.20)	1042 (160)	.54 (.23)	965 (69)
Adjective targets				
Full noun phrase discourse	−.08 (.19)	856 (46)	.07 (.15)	728 (27)
Pronoun discourse	.06 (.14)	873 (44)	.01 (.26)	705 (29)

the pronoun condition, whereas normal elderly adults showed better performance in the pronoun condition than in the repeated noun phrase condition, consistent with other work with young normal adults (Gordon et al., 1993).

Correlations with digit ordering. Digit ordering means are shown in Table 8. The efficiency of processing pronouns (gauged by the difference between each participant's normalized naming latency for inappropriate pronouns and his/her normalized latency for appropriate ones) correlated with digit ordering scores, $r(16) = .51$, $p < .02$, one tail. Furthermore, the extent to which a participant benefited from replacing pronouns with full noun phrases also correlated with the digit ordering score, $r(16) = .44$, $p < .04$. Thus in contrast to Experiment A, digit ordering was reliably correlated with cross-modal naming performance for both types of visual targets.

Discussion

In Experiment B, AD patients exhibited impaired performance in the on-line cross-modal naming task, in that they showed reduced differences between grammatical and felicitous conditions compared to ungrammatical and infelicitous ones, compared to the healthy elderly controls. This result clearly favors the ‘‘shallow processing’’ hypothesis over the ‘‘intact on-line abilities’’ hypothesis because it links patients' impaired performance not to the kind of task (off-line vs. on-line) but to situations in which deeper processing is required, as when maintaining long distance referential information in discourse. The fact that, in contrast to Experiment A, performance in this task with both the pronoun and the adjective targets was correlated with performance of the digit ordering task further supports our contention that the digit ordering task taps the processes operating in the course of regular language comprehension that is not artificially shallow. These results do not indicate that on-line tasks should always be viewed with suspicion in studies of patient populations, but rather that the demands imposed by all tasks must be carefully scrutinized.

One aspect of task demands in the on-line task deserves further attention. Recall that a secondary task was included in the cross-modal naming studies to encourage attention to the auditory stimulus and integration of the visual target and auditory sentence fragment. Because the healthy elderly were asked to judge sentence–target compatibility on every trial while AD patients received a comprehension question on 20% of the trials, it is possible that the different demands of these secondary tasks could have affected the results. For example, perhaps healthy adults performed better than AD patients with pronoun targets in Experiment B because the secondary task directed their attention to sentence–target compatibility. Although it is tempting to point to the secondary task as an explanation of any individual result, the entire pattern of results across Experiments A–B does not point to the secondary task as a primary factor. The subject–verb agreement items (Experiment A), the verb transitivity items (Experiment A), and the pronoun items (Experiment B) all had equal numbers of compatible and incompatible targets. Of these, targets were repeated in the verb agreement study (*was/were* targets) and pronoun study (*him/them* targets). It is difficult to see how the secondary tasks could differentially direct attention such that group differences were observed only in the pronoun study and not in the other two. Moreover, group differences were also found in the adjective target study in Experiment B, in which the visual target did not vary across pronoun and repeated noun condition and was always a felicitous continuation. Thus while we continue to seek ways to hold the secondary task constant across groups in cross-modal naming studies, we do not believe the variation in the secondary task here can explain the results obtained.

Several other issues related to the difference between the items in Experiments A and B need to be addressed before the shallow processing hypothesis can be accepted. First, the items in Experiment A hinged on grammatical relations (subject–verb number agreement and verb argument structure) while the items in Experiment B hinged on discourse relations (pronoun antecedent number agreement and pronoun vs. full NP appropriateness). Thus one interpretation of these results is that AD patients simply have a discourse impairment but not a grammatical impairment. Second, the items in Experiment A differed from the items in Experiment B in the number of intervening words between the target and the part of the context which determined whether the target was a good or bad continuation. In Experiment B, pronoun targets were not presented immediately after the antecedent but only after many words, but in Experiment A, targets followed closely (within one or two words) after the crucial linguistic context. It may be that patients are impaired only when there are many words separating the visual targets from the context element which determines the target's appropriateness. These two issues require further investigation, some of which we have already begun. In Almor et al. (1999) we showed that AD patients' pronoun problems do not necessarily implicate a discourse impairment but instead could be seen as the outcome of normal discourse processing in the context of reduced computational ability. In Almor, MacDonald, Kempler, Anderson, & Tyler (2001), we reported two experiments that tested the effect of the number of intervening words between the target and the element in the context it is related to AD patients' cross-modal naming performance with both grammatical items (as in Experiment A here) and discourse items (as in Experiment B). The results of these experiments indicated that the number of intervening words was not a crucial factor underlying patients' performance with these stimuli. While a complete discussion of these issues is beyond the scope of the present article, we note here that in all and only those cases in which AD patients presented impaired comprehension performance relative to controls, performance correlated with digit ordering scores. Thus, the digit ordering task provides a strong predictor of AD patients' success or failure in language processing tasks requiring thorough processing of the input. The exact reasons why processing may be more or less difficult for some stimuli or tasks cannot be revealed by the digit ordering task but should instead be investigated by careful consideration of the language-processing stimuli and tasks.

Finally, the distinctions between Experiment A–B are related to Caplan and Waters' (1995) claims about ''interpretive'' vs. ''post-interpretive'' processes. It is not clear which of these stages are tapped by the on-line tasks in Experiment B. Caplan and Waters (1999) suggested that pronoun comprehension might best be construed as a postinterpretive process, and if digit ordering also taps postinterpretive processes (as they claim for reading span), then their framework would predict that, at least for the pronoun targets in Experiment B, digit ordering could correlate positively with cross-modal naming times. For the adjective target items, however, the cross-modal naming times reflect integration of an adjective into a sentence (e.g., *The housewife could not believe the plumber was so clumsy*), which perfectly fits Caplan and Waters' definitions of interpretive processing. If so, cross-modal naming would not be expected to correlate with digit ordering in this case. Thus the distinction between interpretive and postinterpretive processing does not seem to make the correct predictions about the range of results in Experiments A–B.

GENERAL DISCUSSION

In this article we have pursued both a theoretical point concerning the interpretation of working memory tasks and their role in language comprehension, and an empirical investigation concerning a new task for assessing language processing and/or working memory abilities in patients with AD. Our analysis of the reading span task (Daneman & Carpenter, 1980) led us to develop the simpler digit ordering task, which proved to be a sensitive measure of AD patients' cognitive and linguistic performance. We argued that digit ordering poses demands similar to those of reading span, evidenced by the correlation between performance of young healthy adults in the reading span and digit ordering tasks, with simple enough instructions that AD patients could perform the task. The results of Experiments A–B on the comprehension of various grammatical, semantic, and discourse relations further showed that the digit ordering task provides a valuable predictor of which comprehension tasks would lead to impaired performance by AD patients.

As indicated in the introduction, this relationship between digit ordering and language comprehension can be viewed from two distinct perspectives. From a traditional working memory perspective, our results could be interpreted as indicating that digit ordering is a useful new test of linguistic working memory, and it correlates with comprehension tasks that draw heavily on linguistic working memory. An alternative perspective (MacDonald & Christiansen, in press) holds that linguistic working memory is not the ''resource'' that enables language processing. Instead, individuals have varying degrees of linguistic skill, owing to biological and experiential factors, and they bring these skills to bear on both ''comprehension'' and ''working memory'' tasks. On this view, correlations between two tasks are consistent with the claim that they have common task demands, independent of whether they are called working memory tasks (reading span and digit ordering, for example) or comprehension tasks. More specifically, AD patients are impaired in their ability to convert acoustic input into a rich syntactic, semantic, and discourse representation, as evidenced by their poor performance compared to controls on (a) digit ordering, (b) off-line grammaticality judgments, and (c) on-line cross-modal naming tasks in which the stimuli require thorough processing, as in Experiment B. By contrast, when the on-line task does not require thorough processing of the sentence but allows participants to respond on the basis of shallow processing of a few nearby words, as in Experiment A, AD patients appear much like normals and their performance is uncorrelated with performance on ordering tasks or other tasks that make more substantial processing demands.

Connectionist Language Processing and ''Working Memory''

The perspective we offer is related to a broader account advocated by MacDonald and Christiansen (in press), in which linguistic working memory is not separable from linguistic knowledge and processing (see also Martin & Saffran, 1997). A full explication of that account is beyond the scope of this article, but some aspects of the approach are particularly relevant to studies of impaired populations. MacDonald and Christiansen suggested that a great deal of recent language comprehension work is consistent with a connectionist approach to language processing, and that connectionist models have implications for the role of working memory in processing. Within the connectionist framework, processing of input is achieved through the passing of activation between interconnected units in a network. The network's capacity to process information varies as a function of the input (e.g., whether the material is complex or simple), the properties of the network (how activation is passed through connections, etc.), and the interaction of these properties—how much a network with a particular architecture has experienced similar input before. With the entire network contributing to processing, and with processing capacity tied to the efficiency of passing of activation across large numbers of units, where is working memory? The most common answer within these systems is one in which processing capacity (or working memory) is a property of the network itself and not as a separate entity that can vary independently of the architecture and experience that governs the network's processing efficiency.

A number of researchers have investigated capacity variation and the nature of individual differences and impairments within connectionist architectures. These differences are instantiated in networks in various ways, including variations in the number of computational units that the network has available to learn and process information (Harm & Seidenberg, 1999; Patterson, Seidenberg, & McClelland, 1989), variation in the amount of training the network receives (MacDonald & Christiansen, in press; Munakata, McClelland, Johnson, & Siegler, 1997), variation in the efficiency with which the network is able to pass information among units (Dell, Schwartz, Martin, Saffran, & Gagnon, 1997), and variation in the amount of ''noise'' in the input signal presented to the network (St. John & Gernsbacher, 1998). Thus the connectionist approach is fully compatible with the notion that individuals and networks may vary in processing capacity, by virtue of initial architecture, linguistic experience, and/or changes in the architecture after training (simulating damage in impaired populations). In all cases, however, these manipulations affect the behavior of the whole network, both its processing and its representation.

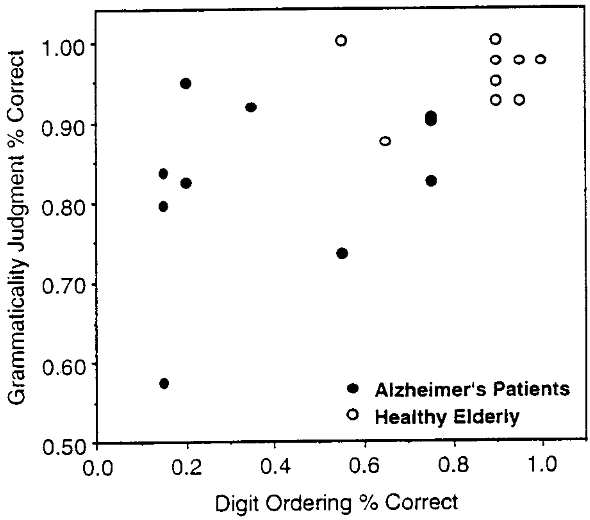
Implications for impairments. This last point has important consequences for accounts of language impairments after brain damage. As noted in the introduction, there is substantial debate in the literature concerning whether AD patients' and aphasics' linguistic impairments stem from reduced computational capacity or loss of linguistic knowledge (Grober & Bang, 1995; Martin et al., 1987; Rochon et al., 1994). This debate is effectively eliminated within the connectionist perspective, in that damage to a network affects both its representations and its processing capacity. Clearly this approach does not answer all questions concerning impaired processing, and there is still ample room for debates concerning the nature of impairments and damage. For example, a patient making some semantic errors in visual word recognition tasks might have damage to semantic systems, orthographic systems, or possibly phonological systems, and networks simulating humans' visual word recognition typically assume these interacting levels of representation (e.g., Plaut & Sallice, 1993). Whatever the claim about the nature of the damage, however, this damage will impair both the processing capacity of the network and the representations embodied in the network.

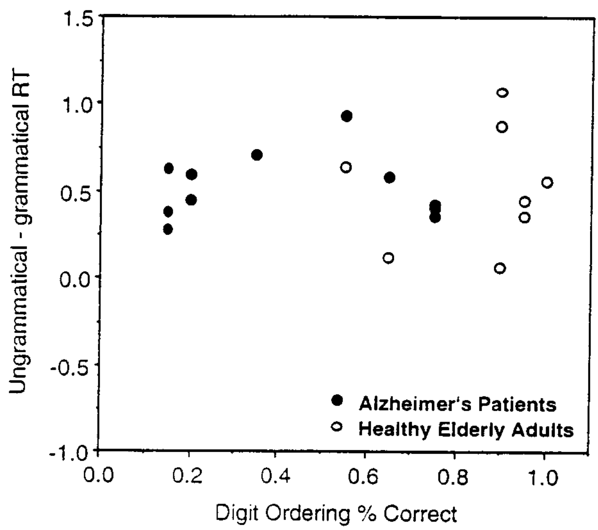
Imaging studies. Similarly, this account offers an alternative perspective on imaging data, which are becoming increasingly important in studies of normal and impaired language processing (e.g., Posner & Pavese, 1998). For example, suppose a particular language-processing task (or subtraction of images between two tasks) yielded activation in Broca's area. On the traditional view in which linguistic knowledge and working memory are separate entities, and processing involves placing knowledge in working memory, researchers could debate whether this activation reflects the locus of linguistic knowledge or (more typically) the locus of the relevant working memory (e.g., Just et al., 1996). We think that this dichotomy is quite awkward—if the activation in Broca's area reflects the working memory rather than the knowledge, then where is linguistic knowledge? Or if the activation reflects knowledge, then where is the computation space where language processing occurs? Our view avoids this dichotomy: Processing involves passing activation through a distributed network, and activation patterns in an imaging study indicate the operation of that (human) network. As there is no one part of the network that can be identified as ‘‘working memory’’ or ‘‘knowledge,’’ activation data do not indicate working memory or knowledge but instead localize the processing required to perform the task. Thus imaging data, in our view, reflect network activities in response to particular task demands, a view that is shared by imaging researchers. For example, Postle and D'Esposito (1999), using tasks requiring memory for shape and location information, found no evidence of a separate working memory system. They argued that their fMRI data support ‘‘a model of visual working memory function in which material-specific mnemonic representations are maintained in the same neuronal networks that subserve sensory analysis of these stimuli’’ (p. 590). We think that avoiding the awkward ‘‘knowledge or working memory’’ debate makes imaging results more coherent and more interesting than within the traditional perspective.

These claims concerning the inseparability of capacity and knowledge clearly need additional testing. MacDonald and Christiansen (in press) investigated their implications for individual differences largely within the normal young adult population, and we have just begun to consider the implications for impairments in AD here. More concretely, the digit ordering task itself needs additional validation, information on test–retest reliability, the establishment of norms, etc. What these preliminary data have suggested, however, is that the digit ordering holds promise for studies of impaired populations, and that a detailed analysis of task demands can lead to a novel perspective on the relationship between what have traditionally been termed ‘‘working memory’’ and ‘‘comprehension’’ tasks, and between the notions of ‘‘working memory’’ and ‘‘language comprehension’’ more generally.

APPENDIX 1

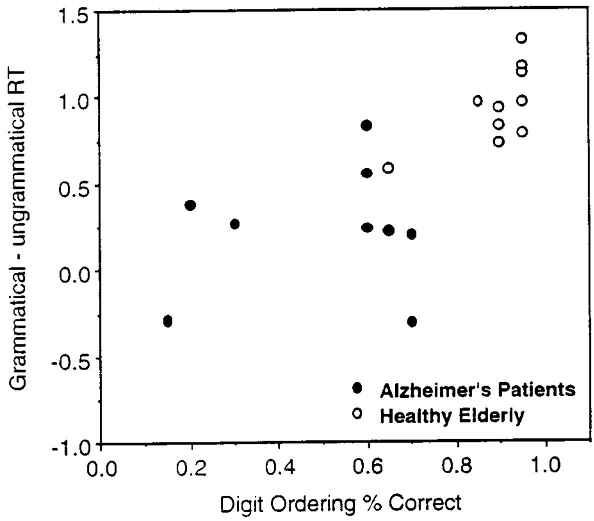
Scatterplots of AD and healthy elderly participants in Experiment A. Graphs show the relationship between digit ordering and grammaticality judgment (top graph) and digit ordering and grammaticality effect in cross-modal naming (bottom graph).

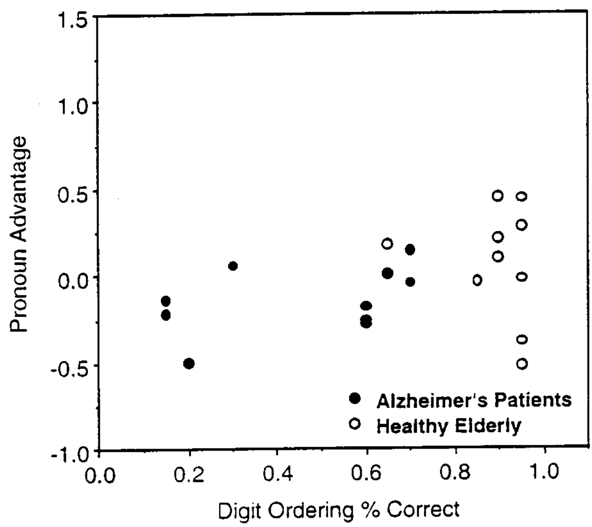




APPENDIX 2

Scatterplots of AD and healthy elderly participants in Experiment B. Graphs show the relationship between digit ordering and pronoun grammaticality effect in cross-modal naming (top graph), and between digit ordering and pronoun vs. repeated noun advantage in cross-modal naming (bottom graph).





REFERENCES

Allen, J., & Seidenberg, M. S. (1999). The emergence of grammaticality in connectionist networks. In B. MacWhinney (Ed.), *The emergence of language.* Mahwah, NJ: Erlbaum.

Almor, A., Kempler, D., MacDonald, M. C., Andersen, E. S., and Tyler, L. K. T. (1999). Why do Alzheimer's patients have difficulty with pronouns? Working memory, semantics, and pronouns in Alzheimer's disease. *Brain and Language,* **67,** 202–227.

Almor, A., MacDonald, M. C., Kempler, D., Andersen, E. S., and Tyler, L. K. T. (2001). Long distance number agreement in Alzheimer's Disease. *Language and Cognitive Processes,* **16,** 35–63.

Baddeley, A. (1986). *Working memory.* Oxford: Oxford Univ. Press.

Baddeley, A. D., Della Sala, S., & Spinnler, H. (1991). The two-component hypothesis of memory deficit in Alzheimer's disease. *Journal of Clinical and Experimental Neuropsychology,* **13,** 372–380.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation.* New York: Academic Press.

Benton, A. L., Hannay, H. L., & Varney, N. R. (1975). Visual perception of line direction in patients with unilateral brain disease. *Neurology,* **25,** 907–910.

Berg, L. (1985). Does Alzheimer's disease represent an exaggeration of normal aging? *Archives of Neurology,* **42,** 737–739.

Berg, L. (1989). Clinical dementia rating. *Psychoparmacology Bulletin,* **24,** 637–639.

Caplan, D., & Waters, G. S. (1995). Aphasic disorders of syntactic comprehension and working memory capacity. *Cognitive Neuropsychology,* **12,** 638–649.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral & Brain Sciences,* **22,** 77–126.

Cooper, J. A., Sagar, H. J., Jordan, N., Harvey, N. S., & Sullivan, E. V. (1991). Cognitive impairment in early, untreated Parkinsons Disease and its relationship to motor disability. *Brain,* **114,** 2095–2122.

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior,* **19,** 450–466.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review,* **3,** 422–433.

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical access in aphasic and nonaphasic speakers. *Psychological Review,* **104,** 801–838.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review,* **102,** 211–245.

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

Goodglass, H., Kaplan, E., & Weintraub, S. (1982). *Boston Naming Test.* Philadelphia: Lea & Febiger.

Gordon, P. C., Grosz, B. J., & Gilliom, L. (1993). Pronoun, names, and the centering of attention in discourse. *Cognitive Science,* **17,** 311–348.

Grober, E., & Bang, S. (1995). Sentence comprehension in Alzheimer's disease. *Developmental Neuropsychology,* **11,** 95–107.

Harm, M., & Seidenberg, M. S. (1999). Phonology, reading, and dyslexia: A connectionist modeling perspective. *Psychological Review,* **106,** 491–528.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review,* **98,** 122–149.

Just, M. A., Carpenter, P. A., & Keller, T. A. (1996). The capacity theory of comprehension: New frontiers of evidence and arguments. *Psychological Review,* **103,** 773–780.

Kempler, D., Almor, A., Tyler, L. K., Andersen, E. S., & MacDonald, M. C. (1998). Sentence comprehension in Alzheimer's Disease: A comparison of off-line vs. on-line sentence processing. *Brain and Language,* **64,** 297–316.

MacDonald, M. C., & Christiansen, M. H. (in press). Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review.*

Martin, N., & Saffran, E. M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology,* **14,** 641–682.

Martin, R. C. (1995). Working memory doesn't work: A critique of Miyake et al.'s capacity theory of aphasic comprehension deficits. *Cognitive Neuropsychology,* **12,** 623–636.

Martin, R. C. , Wetzel, F., Blossom-Stach, C., & Feher, E. (1987). An assessment of two theories of agrammatism and syntactic comprehension deficits. *Cognition,* **32,** 157–191.

Martin, R. C., & Lesch, M. (1996). Associations and dissociations between languge impairment and list recall: Implications for models of STM. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 149–178). East Sussex, UK: Psychology Press.

Martin, R. C., Wetzel, F., Blossom-Stach, C., & Feher, E. (1987). An assessment of two theories of agrammatism and syntactic comprehension deficits. *Cognition,* **32,** 157–191.

McKhann, G., Drachman, D., Folstein, M. F., 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 Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology,* **34,** 939–944.

Miyake, A., Carpenter, P., & Just, M. (1994). A capacity approach to syntactic comprehension disorder: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology,* **11,** 671–717.

Miyake, A., Carpenter, P., & Just, M. (1995). Reduced resources and specific impairments in normal and aphasic sentence comprehension. *Cognitive Neuropsychology,* **12,** 651–679.

Morris, J. C., Mohs, R. C., Rogers, H., Fillenbaum, G., & Heyman, A. (1988). Consortium to establish a registry for Alzheimer's disease (CERAD) clinical and neuropsychological assessment of Alzheimer's disease. *Psychopharmacology Bulletin,* **24,** 641–652.

Munakata, Y., McClelland, J. L., Johnson, M. H., & Siegler, R. S. (1997). Rethinking infant knowledge: Toward an adaptive process account of successes and failures in object permanence tasks. *Psychological Review,* **104,** 686–713.

Patterson, K., Seidenberg, M., & McClelland, J. (1989). Dyslexia in a distributed, developmental model of word recognition. In R. Morris (Ed.), *Parallel distributed processing.* Oxford: Oxford Univ. Press.

Posner, M. I., & Pavese, A. (1998). Anatomy of word and sentence meaning. *Proceedings of the National Academy of Sciences,* **95,** 899–905.

Postle, B. R. & D'Esposito, M. (1999). ''What''—then—''Where'' in visual working memory: An event-related fMRI study. *Journal of Cognitive Neuroscience,* **11,** 585–597.

Rochon, E., Waters, G., & Caplan, D. (1994). Sentence comprehension in patients with Alzheimer's Disease. *Brain and Language,* **46,** 329–349.

St. John, M., & Gernsbacher, M.A. (1998). Learning and losing syntax: Practice makes perfect and frequency builds fortitude. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Foreign language learning: Psycholinguistic experiments on training and retention* (pp. 231–255). Mawah, NJ: Erlbaum.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science,* **268,** 1632–1634.

Tyler, L. K. (1992). *Spoken language comprehension: An experimental approach to normal and disordered processing.* Cambridge, MA: MIT Press.

Waters, G., & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology A,* **49,** 51–79.

Waters, G., Caplan, D., & Hildebrandt, N. (1991). On the structure of verbal short-term memory and its functional role in sentence comprehension: Evidence from neuropsychology. *Cognitive Neuropsychology,* **8,** 81–126.

Waters, G. Caplan, D., & Rochon, E. (1995). Processing capacity and sentence comprehension in patients with Alzheimer's disease. *Cognitive Neuropsychology,* **12,** 1–30.

Waters, G. S., & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review,* **103,** 761–772.

Wechsler, D. (1987). *Wechsler Memory Scale—Revised.* New York: Psychological Corp.

Yesavage, J. A. Brink, T. L. Rose, T. L., Lum, O., Huang, V. Adey, M., & Leirer, V.O. (1983). Development and validation of a geriatric depression screening scale: A preliminary report. *Journal of Psychiatric Research,* **17,** 37–49.