Psychology and Aging
2004, Vol. 19, No. 1, 157–170

0882-7974/04/$12.00 DOI: 10.1037/0882-7974.19.1.157

Eye-Fixation Patterns of High- and Low-Span Young and Older Adults: Down the Garden Path and Back Again

Susan Kemper and Angela Crow
University of Kansas

Karen Kemtes
University of Nevada, Las Vegas

Young and older adults' eye fixations were monitored as they read sentences with temporary ambiguities such as "The experienced soldiers warned about the dangers conducted the midnight raid." Their fixation patterns were similar except that older adults made many regressions. In a 2nd experiment, high- and low-span older adults were compared with high- and low-span young adults. First-pass fixations were similar, except low-span readers made many regressions and their total fixation times were longer. High-span readers also used the focus operator "only" (e.g., "Only experienced soldiers warned about the dangers . . .") to immediately resolve the temporary ambiguities. No age group differences were observed. These results are discussed with reference to theories of the role of working memory in sentence processing.

Two experiments were conducted to provide new data relevant to the ongoing debate over the role of working memory capacity limitations in sentence processing and the possibility of age differences in the processing of complex syntactic constructions (Just & Carpenter, 1992). This debate has focused on the processing of complex sentences containing temporary syntactic ambiguities, such as "The experienced soldiers warned about the dangers conducted the midnight raid." At issue is whether or not both young and older adults or high- and low-span individuals experience garden-path effects: an increase in processing time reflecting the initial misinterpretation of the first verb "warned" as the main verb (MV) of the sentence rather than as the verb of a reduced relative clause (RRC). Existing studies have been severely criticized on a number of methodological and procedural grounds (see Caplan & Waters, 1999, and subsequent responses).

Just et al. (Just & Carpenter, 1992; Just & Varma, 2002; King & Just, 1991; MacDonald, Just, & Carpenter, 1992) have claimed that working memory capacity constrains the interpretation of temporary syntactic ambiguities, limiting the ability of low-span readers to make and sustain multiple interpretations of the ambiguous phrases. Older adults were claimed to resemble low-span readers and, thus, have difficulty making and sustaining multiple interpretations of ambiguous phrases. In contrast, Caplan and Waters, in another series of reviews and publications (Caplan & Waters, 1999, 2002; Waters & Caplan, 1996a, 1996b, 2001, 2002), have argued that syntactic parsing and other interpretive processes, including lexical access, assignment of thematic roles, and the determination of topic, focus, and coreference, rely on a specialized processing system with a separate sentence-interpretation resource (SSIR) unrelated to traditional span measures of working memory.

According to the Just and Carpenter (1992) capacity-constrained (CC) theory (see also the 3CAPS model of Just & Varma, 2002), low-span readers should have difficulty processing the temporary syntactic ambiguities and should exhibit garden-path effects. Initially, the RRC constructions are misinterpreted as MVs based on the greater frequency of occurrence of the MV construction; once disambiguating information is encountered, the phrase must be reinterpreted. High-span readers should be able to avoid garden-path effects by constructing multiple syntactic interpretations of the ambiguous phrases and retaining these interpretations until disambiguating information is encountered. Caplan and Waters's SSIR theory (1999) predicts similar patterns of online processing for all readers because interpretive processes are buffered from working memory limitations. All readers should show increased processing delays at points of maximal syntactic complexity.

Because it is well established that older adults typically show impairments of working memory on span tests, research findings from studies comparing sentence processing by young and older adults have been critical to this debate. For example, Kemtes and Kemper (1997) have attempted to replicate the finding by MacDonald et al. (1992) that high- and low-span readers would allocate processing time differently to sentences containing temporary syntactic ambiguities. Kemtes and Kemper compared young and older adults differing in working memory span. They hypothesized that older adults as a group should resemble low-span young adults and exhibit garden-path effects. Kemtes and Kemper failed to replicate the results of MacDonald et al.: They found that all readers, regardless of age or span group, showed garden-path

Susan Kemper and Angela Crow, Gerontology Center, University of Kansas; Karen Kemtes, Department of Psychology, University of Nevada, Las Vegas.

Additional materials for this article are available on the Web at dx.doi.org/10.1037/0882-7974.19.1.157.supp

This research was supported by National Institute on Aging Grant RO1 AG009952 and the Roy A. Roberts endowment at the University of Kansas. We thank Tracy Mitzner and Heather Humphrey for their help with eye tracking and Chris Lankford for his advice.

Correspondence concerning this article should be addressed to Susan Kemper, Gerontology Center, 3090 Dole Building, 1000 Sunnyside, University of Kansas, Lawrence, KS 66045. E-mail: skemper@ku.edu

effects for sentences with temporary syntactic ambiguities. They also reported that older low-span readers had difficulty answering probe questions about the ambiguous sentences. Caplan and Waters (1999) take these results as support for their theory that interpretive processes are buffered from working memory limitations, whereas postinterpretive processes, such as question answering, are not.

Kemtes and Kemper (1999) suggested that subtle differences between their procedures and those of MacDonald et al. (1992) may have contributed to the discrepant findings. To resolve this issue and address the criticisms raised by Caplan and Waters (1999), this experiment was repeated using an alternative methodology: eye tracking. Eye tracking was chosen over other methods because it is natural and unobtrusive. It is widely used to develop and test theories of sentence processing and has superceded other reading methods, such as word-by-word reading, cross-modal priming, and lexical decision (see Ferreira & Anes, 1994). It permits participants to use parafoveal information to preview upcoming words and phrases and to make regressive fixations to review previous words and phrases. These may be important components of skilled reading. Eye tracking has been shown to be sensitive to a variety of linguistic and discourse manipulations that affect the time course of sentence processing. It has the advantage over the auditory moving-window paradigm used by Waters and Caplan (2001, 2002) in that the reading task imposes few demands on the participant: (a) The eye-tracking system used in the current study does not require the use of a bite bar or other device to constrain head movements; (b) stimuli are presented naturally, and no artificial segments are imposed on the flow of information; and (c) sentences are read for content without additional processing demands such as the acceptability judgment task used by Waters and Caplan. Although Waters and Caplan claim that the auditory moving-windows paradigm presents stimuli in the "primary modality of language reception" (p. 130), it is arguable whether syntactically complex sentences, such as those used in the current experiment, commonly occur in oral language; indeed, they may be rarely used in oral language just because they are difficult to produce and understand (Beaman, 1984; Chafe, 1982; Hildyard & Olson, 1982). Hence, monitoring eye-fixation patterns while reading may be a more sensitive and naturalistic method for examining the immediate processing of the complex syntactic constructions than word-by-word reading and auditory moving windows procedures.

Experiment 1

Eye-fixation patterns of young and older adults were compared as they read sentences containing temporary syntactic ambiguities. Three measures were obtained from the eye-fixation records: first-pass fixation times, leftward regressions to previous parts of the sentence, and total fixation times reflecting first-pass fixations plus subsequent fixations arising from regressions and rereading the sentence in whole or in part. First-pass fixations were assumed to reflect immediate syntactic processing and semantic analysis, corresponding to Waters and Caplan's interpretive processes. They define interpretive processes as "largely unconscious, obligatory, on-line, first-pass comprehension processes devoted to assignment of the literal, preferred, discourse-congruent meaning . . . [including] acoustic-phonetic conversion; lexical access; recognition of intonation contours; assignment of propositional values, such as thematic roles, attribution of modification, and scope of quantification; and determination of discourse-level semantic values, such as topic, focus, co-reference, and others" (2001, p. 128). First-pass fixations were expected to vary with the time course of syntactic processing, peaking at the most complex region of the sentence. The Just and Carpenter (1992) model implies that older readers should exhibit a peak in first-pass fixations late in ambiguous RRC sentences when they encounter the second verb. Having been led down the garden path, older readers must attempt to reanalyze the first verb phrase as an RRC. Young readers should exhibit a peak in first-pass fixations early in ambiguous MV and RRC sentences when they encounter the ambiguous first verb if they compute both the MV and RRC interpretations. They may also exhibit a peak late in the sentence when they encounter disambiguating information.

Regressions were expected to reflect breakdowns of immediate processing, peaking whenever readers were unable to assign a syntactic interpretation (Ehrlich & Rayner, 1983; Ferreira & Clifton, 1986; Frazier & Rayner, 1982; Kennedy, 1983). According to the Just and Carpenter (1992) model, regressions to previous parts of the sentence were expected to peak for older readers when they encounter the second verb of ambiguous RRC sentences and reread previous parts of the sentence in order to reanalyze the first verb phrase. The target "landing site" of the regression was expected to indicate whether they resorted to rereading the entire sentence by regressing to the subject noun phrase or to rereading just the critical first verb phrase. Young readers may also make regressions to previous parts of the sentence when they encounter disambiguating information in MV and RRC sentences. Total fixation times were assumed to reflect the time course of immediate processing as well as any reprocessing of the sentence in whole or in part after regressions. Older adults were expected to spend more time rereading the initial parts of ambiguous RRC sentences to correctly reinterpret the first verb phrase as an RRC.

Method

Participants

Ten older adults and 10 young adults participated. All young participants were college students recruited by means of postings on campus bulletin boards and class announcements. All older participants were community-dwelling adults who were recruited from a registry of prior research participants. All participants were monolingual speakers of English. All were paid a modest honorarium for their participation. The participants are described more fully in Table 1. On the basis of a one-way analysis of variance (ANOVA) comparing age groups, the participants did not differ in educational level, although they did differ in working memory span as measured by the Digits Forward and Digits Backward tests from the Wechsler Adult Intelligence Scale-Revised (Wechsler, 1958) and the Daneman and Carpenter Reading Span Test (Daneman & Carpenter, 1980). An alpha level of .05 was set for this and all subsequent t and F tests.

Materials

The MV/RRC sentences were originally developed by MacDonald et al. (1992) and supplemented by Kemtes and Kemper (1997). The critical contrast is between Example 1a and 1b: In Example 1a, the sentence ends in a prepositional phrase (PP), whereas that in Example 1b ends with a second verb predicate (2V). Hence, in Example 1a, the first verb is initially and correctly interpreted as an MV; in Example 1b, that initial MV

Table 1
Comparison of Young and Older Adults in Experiment 1

	Young adults		Older adults			
Variable	*M*	*SD*	*M*	*SD*	*F*(1, 19)	*p*
Age	19.7	1.9	75.2	3.0		
Education	13.5	0.6	15.2	2.6	3.94	.06
Vocabulary	30.3	3.2	33.2	2.3	5.34	.03
WAIS–R—Digits Forward	7.1	1.2	4.9	1.6	12.17	.01
WAIS–R—Digits Backward	5.6	1.3	3.7	0.9	13.26	.01
Reading Span	3.9	0.8	2.3	0.5	25.60	<.01

Note. WAIS–R = Wechsler Adult Intelligence Scale—Revised.

interpretation of the first verb must be reanalyzed as an RRC once the disambiguating second verb is encountered. Examples 1c and 1d serve as controls. Example 1c is an unambiguous MV sentence; the verb cannot be used in an RRC. Example 1d is an unambiguous unreduced relative clause sentence; the relative pronoun "who" serves to immediately disambiguate its structure. Each sentence was divided into three critical regions: the subject noun phrase (subject NP), the first verb phrase (1VP), and the second verb phrase or prepositional phrase (2VP/PP). These critical regions were used in the analysis of the eye-movement data.

Subject NP | 1VP | 2V/PP (1)

The experienced solidiers | warned about the dangers | before the midnight raid. (1a)

The experienced soldiers | warned about the dangers | conducted the midnight raid. (1b)

The experienced soldiers | spoke about the dangers | before the midnight raid. (1c)

The experienced soldiers | who were told about the dangers | conducted the midnight raid. (1d)

There were 40 sets of MV/RRC sentences. They were assigned to four stimulus lists such that each list contained 10 examples of each type (Examples 1a–d) of MV/RRC sentence but only 1 sentence from each set. In addition to the experimental sentences, each list contained 60 filler sentences of various syntactic forms. Unlike the studies of Waters and Caplan (2001, 2002), only grammatically well-formed, semantically acceptable sentences were used in the reading study. The lists were randomized, and each was broken into two blocks of 50 sentences.

Procedure and apparatus

Participants were first acquainted with the equipment and then given a block of 20 practice sentences to read. The participants were asked to read each sentence carefully for meaning and advised they would be tested on the sentences at the conclusion of the experiment. No comprehension test was administered. Each participant was assigned randomly to one of four stimulus lists. Then two blocks of sentences were presented; order of the blocks was counterbalanced across participants.

An Applied Sciences Laboratories eye tracker (Model 504) with a magnetic head tracker was used to record eye movements. Eye movements were sampled 60 times/s with an accuracy of 0.5 degrees visual angle at 16 in. The head tracker noted displacements, relative to a base unit, of a sensor attached to a visor worn by the reader and corrected the record of eye movements for head movements. Head movements were sampled 100 times/s with an accuracy of 0.03 degrees. Stimuli were presented using GazeTracker software (Lankford, 2001), which also analyzed the eye-movement data. The eye tracker was calibrated at the start of each session and between blocks for each participant. One microcomputer controlled the eye tracker; it was interfaced with a second computer running the GazeTracker software for presentation and analysis.

Each trial consisted of a fixation point centered on a blank screen for 500 ms followed automatically by the presentation of a sentence. The participants controlled presentation by pressing the mouse when they had completed reading the sentence. Participants sat in an adjustable chair with a headrest. They wore reading glasses if they normally did so. The chair could be raised or lowered to accommodate to bi- or trifocals. The participants also wore a visor with a small magnetic sensor for the head tracker. The sentences were presented on a Dell 17-in. (278.6 cm) flat-panel computer screen at a viewing distance of 16 in. (262.2 cm). The fixation point and stimulus items were presented in white (125.5 lux) on a black background (0.03 lux) to maximize pupil size. Text was presented in Arial typeface with a mean size for individual letters of 0.57 degrees. The participants held a computer mouse in their preferred hand, which was used to control sentence presentation.

Sentences were segmented into critical regions as in Example 1. Two reading time measures were computed for each critical region: the duration of the first-pass fixations to the region and the total duration of all fixations to a region. First-pass fixation duration is the sum of all fixations to a region beginning with the initial fixation inside a region and ending with the first saccade leftward to a prior region or rightward to a successive region. Total fixation duration included all first-pass fixations as well as any fixations resulting from regressions to the region or subsequent refixation after a leftward or rightward saccade to another region. In addition, first-pass regressions from one region leftward to a previous region were also identified. The mean number of first-pass regressions to the subject NP from subsequent regions was determined; regressions to other regions were infrequent (<5% of all regressions) and were not analyzed further, although they did contribute to the calculation of total fixation durations. The average duration of first-pass regressions to the subject NP was also computed. Fixations were defined as a minimum of two successive sampled eye positions occurring within a diameter of 30 pixels, approximately 0.5 degrees of visual angle.

Because the critical regions differed in length, the fixation and regression durations were analyzed following the procedure recommended by Trueswell, Tanenhaus, and Garnsey (1994). First-pass fixation durations, first-pass regression durations, and total durations for each participant were first regressed on the number of characters (letters and spaces) in each region, and then the residuals, reflecting length-corrected reading times, were used in the actual analysis. Raw, unadjusted first-pass fixation and regression durations and total fixation durations may be found at dx.doi.org/10.1037/0882-7974.19.1.157.supp.

Data from 2% of the experimental trials for young adults and 4% of the trials for older adults were lost as a result of eyeblinks, large head movements, or other eye tracking failures.

Results

The analysis of the MV/RRC sentences involved four measures: first-pass fixation times, mean number of regressions to the subject NP, duration of these regressions to the subject NP, and total fixation times. An omnibus ANOVA with age group, critical region, sentence type, and ambiguity is reported followed by a region-by-region decomposition of significant effects and interactions. Lower order main effects and interactions subsumed by higher order interactions are not reported.

First-Pass Fixation Time

A 2 (age group) × 3 (region) × 2 (sentence type) × 2 (ambiguity) ANOVA compared first-pass fixation times. The results are summarized in Figure 1. The four-way interaction was significant, $F(1, 18) = 12.80$, $p = .01$, $\eta^2 = .41$. At Regions 1 and 2, subject NP and 1VP, there were no significant effects or interactions: First-pass fixation times did not vary with age group, ambiguity, or sentence type. At Region 3, 2V/PP, the two-way Sentence Type × Ambiguity interaction, $F(1, 18) = 28.76$, $p < .01$, $\eta^2 = .63$, was significant but did not interact with age. First-pass fixations to Region 3 by young and older adults were longer for ambiguous RRC sentences than for the other three types of sentences.

Regressions to the Subject NP

A 2 (age group) × 2 (region) × 2 (sentence type) × 2 (ambiguity) ANOVA compared the mean number of regressions to the subject NP from the other two critical regions: 1VP and 2V/PP. The results are summarized in Figure 2. The two-way interaction of age group and ambiguity was significant, $F(1, 18) = 10.77$, $p = .01$, $\eta^2 = .37$. Young adults made few regressions from ambiguous ($M = 0.25$, $SE = 0.11$) or unambiguous ($M = 0.35$, $SE = 0.14$) sentences regardless of type. Older adults made few regressions from unambiguous sentences ($M = 0.30$, $SE = 0.13$) but made more regressions back to the subject NP for the ambiguous sentences ($M = 1.28$, $SE = 0.23$). The average duration of these regressions was compared with a similar ANOVA; only the main effect of age group was significant, $F(1, 18) = 15.04$, $p < .01$, $\eta^2 = .29$. Young adults made longer regressions ($M_{Duration} = 0.14$, $SE = 0.04$) than older adults ($M_{Duration} = 0.10$, $SE = 0.05$).

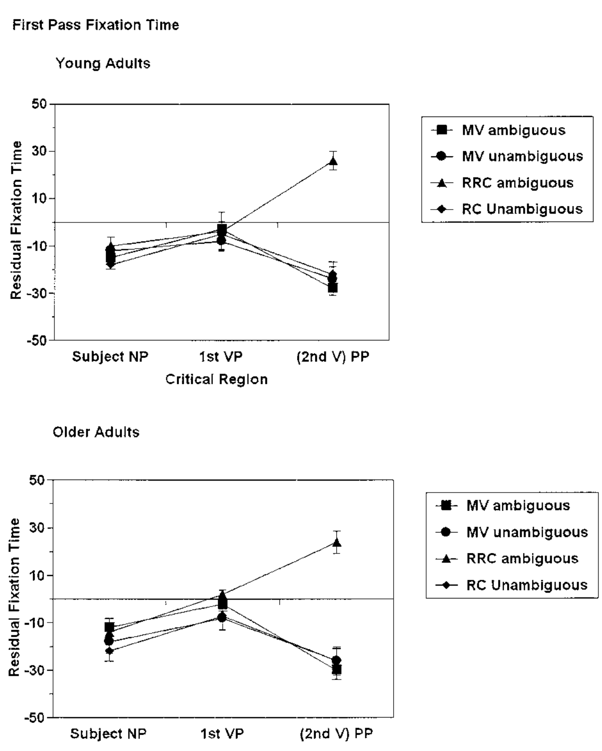

Figure 1. Residual first-pass fixation time (and *SE*s) for young and older adults for ambiguous main verb (MV) and reduced relative clause (RRC) sentences and their unambiguous controls. NP = noun phrase; VP = verb phrase; V = verb; PP = prepositional phrase; RC = relative clause.

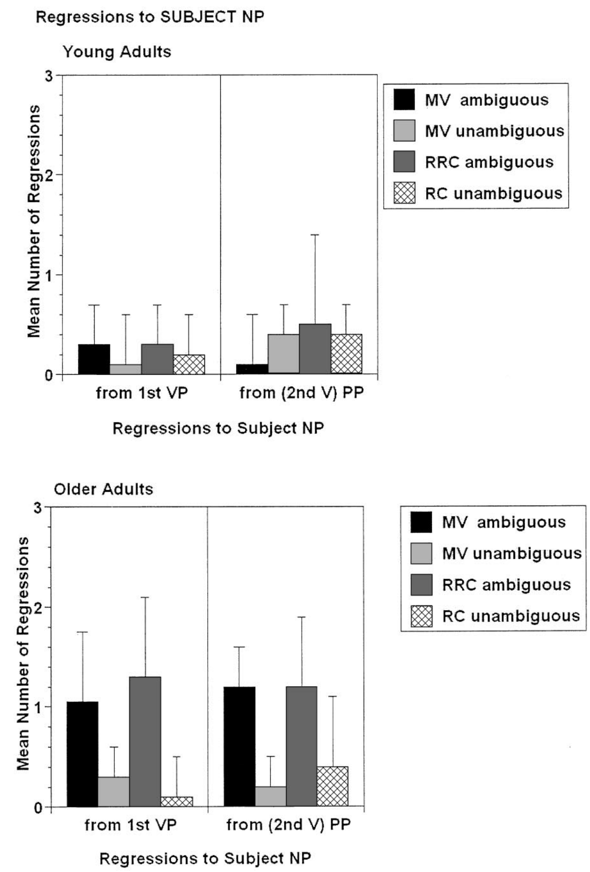

Figure 2. Regressions (and *SE*s) to the subject noun phrase (NP) for young and older adults for ambiguous main verb (MV) and reduced relative clause (RRC) sentences and their unambiguous controls. VP = verb phrase; V = verb; PP = prepositional phrase; RC = relative clause.

Total Fixation Time

A 2 (age group) × 3 (region) × 2 (sentence type) × 2 (ambiguity) ANOVA compared total fixation times for the three critical regions: subject NP, 1VP, and 2V/PP. The results are summarized in Figure 3. The three-way interaction of region, sentence type, and ambiguity was significant, $F(2, 16) = 22.71$, $p < .01$, $\eta^2 = .56$. At Region 1, subject NP, and region 2, 1VP, there were no significant effects or interactions. At Region 3, 2V/PP, the Type × Ambiguity interaction was significant, $F(1, 18) = 28.73$, $p = .01$, $\eta^2 = .62$, but did not interact with age group. Total fixation times were longer for ambiguous RRC sentences than for the other types of sentences.

Discussion

First-pass fixation times to the subject NP or 1VP did not vary by age group, sentence type, or ambiguity. Only at Region 3,

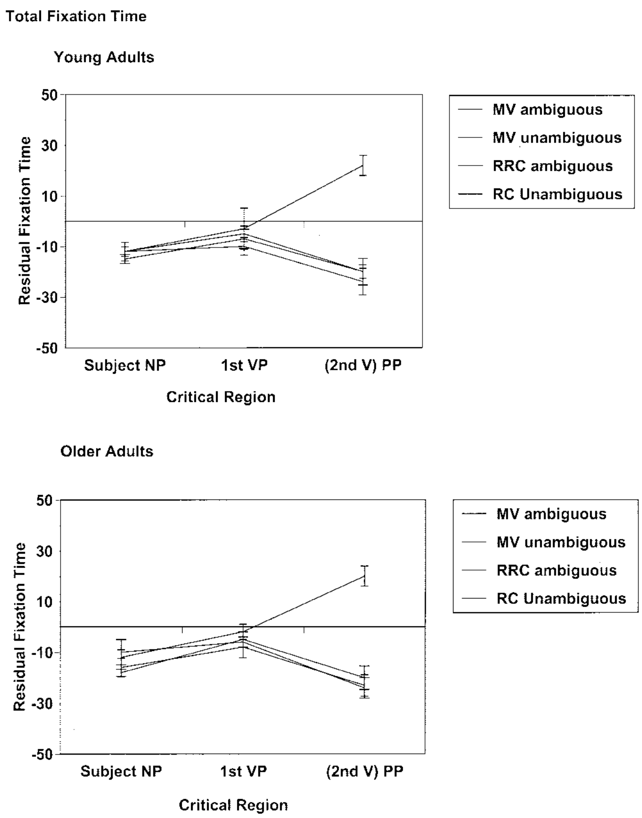

Figure 3. Residual total fixation time (and *SE*s) for young and older adults for ambiguous main verb (MV) and reduced relative clause (RRC) sentences and their unambiguous controls. VP = verb phrase; V = verb; PP = prepositional phrase; RC = relative clause.

2V/PP, did first-pass fixation times diverge as a function of sentence type and ambiguity. Region 3 corresponds to the 2VP of the RRC sentences or the PP of MV sentences; it thus disambiguates the syntactic role of the 1VP in Region 2. At Region 3, first-pass fixation times were longer for RRC sentences with temporary syntactic ambiguities than for MV sentences or unambiguous control sentences. Young adults made few regressions back to the subject NP. Older adults made more regressions to the subject NP from Regions 2, 1VP, and 3, 2V/PP, in order to interpret the ambiguous sentences. These regressions resulted in short fixations, which did not affect differentially their total fixation times for the subject NP, suggesting the older adults were quickly rereading the subject NP. Total fixation times for both young and older adults were longer for the ambiguous RRC sentences at Region 3, corresponding to the disambiguating information.

These results are not fully consistent with either the CC model of Just and Carpenter (1992; also the 3CAPS model of Just & Varma, 2002) or the SSIR model of Waters and Caplan (1996a, 1996b). According to the CC model, fixation patterns by young and older adults should differ, reflecting the effects of working memory limitations on sentence processing. The similarity of the first-pass fixation times and total fixation times for young and older adults for MV and RRC sentences does not support the CC model; both young and older adults experienced a garden-path effect for the RRC sentences at Region 3. The regression data support the Just and Carpenter model because young adults were able to resolve the temporary syntactic ambiguity without rereading the subject NP, presumably by drawing on information in working memory, whereas the older adults could do so only after rereading the subject NP. However, these regressions resulted in very short refixations of the subject NP that did not significantly increase total fixation times, suggesting the older adults did not engage in substantial reanalysis of the subject NP.

The SSIR model (Caplan & Waters, 1999, 2002; Waters & Caplan, 1996a, 1996b, 2001, 2002) holds that there is a specialized working memory system for interpretive processing that is at least partially distinct from a more general working memory system involved in postinterpretive processing. The similarity of first-pass fixation times for young and older readers supports the Waters and Caplan model. The regression data do not. Waters and Caplan have never discussed the role of regressive eye fixations with regard to interpretive versus postinterpretive processes because they rely on the auditory moving-windows paradigm, which does not permit regressions or distinguish first-pass from total listening times. If regressive eye movements reflect interpretive processes, the regressions to the subject NP from Regions 2 and 3 in ambiguous sentences by older adults would be inconsistent with the SSIR model. However, regressions and total fixation times are generally assumed to reflect initial parsing problems and reanalysis and thus would be postinterpretive. Although there were significant age group differences in regressions from Regions 2 and 3 to the subject NP, total fixation times to the subject NP were similar for both groups. This pattern implies that the older adults' leftward regressions, although more numerous than those of the young adults, did not result in extensive reanalysis of the subject NP but served as a postinterpretive confirmation of their analysis of the sentence. Under this view, immediate processing of the ambiguous sentences is similar for young and older adults, although postinterpretive processing is affected by age group differences, presumably reflecting older adults' reduced working memory capacity.

Experiment 2

This experiment was designed to examine both age group differences and span group differences in eye-fixation patterns to ambiguous sentences. To clearly establish the role of working memory in sentence processing, an "excluded middle" design was used to select a group of high-span older adults and a group of low-span older adults, excluding older adults with intermediate memory spans. Then high- and low-span young adults were carefully selected to match the span scores of the two groups of older adults. If working memory capacity affects sentence processing, high- and low-span groups should differ in fixation patterns. If working memory capacity does not affect sentence processing, high- and low-span groups should exhibit similar fixation patterns, according to the SSIR model. At issue was whether or not age group differences would be observed in addition to span group differences.

In addition to the MV/RRC sentences used in the first experiment, a second type of RRC sentence was used to assess working memory span differences in fixation patterns. Ni, Crain, and Shankweiler (1996; Experiment 2) reported different first-pass fixation patterns for two types of RRC constructions: those marked by the determiner "the" versus those marked by the focus operator "only." The focus operator sets up an expectation of contrasting set, as in "Only businessmen loaned money at low interest rates." Ni et al. (1996) claim that this contrast is semantic or referential and biases the initial interpretation of the verb as an RRC. Ni et al. reported that readers did not experience a garden-path effect for "only" sentences but did so for "the" sentences. In a related experiment (Experiment 4), Ni et al. reported different fixation patterns for high- and low-span readers for "only" sentences: rapid first-pass processing of "only" sentences by low-span readers coupled with many regressions versus slower first-pass processing of "only" sentence by high-span readers coupled with few regressions. Liversedge, Paterson, and Pickering (1998) and Clifton, Bock, and Radó (2000) have challenged these findings, reporting that "only" does not bias first-pass fixation times. Neither study compared high- and low-span readers. Kemtes (1998) failed to find significant age group differences for these types of sentences using an immediate grammatical judgment task. Experiment 2 was designed to examine the issues raised by these discrepant findings.

Experiment 1 did not investigate whether young and older readers differed in their comprehension of MV and RRC sentences. In Experiment 2, probe questions were inserted after 50% of the sentences to monitor the participants' comprehension. The probe questions assessed whether the readers were able to avoid the incorrect, garden-path interpretation of the ambiguous RRC sentences by correctly interpreting the subject of the first verb. They also assessed whether the readers were able to correctly interpret the subject of both verbs in "the" and "only" sentences.

Method

Participants

A group of 24 older adults were given a battery of working memory tests (described later). Based on their performance on these tests, 8 high-span and 8 low-span older adults were selected for inclusion in the reading study. The 8 older adults with the highest span scores, on at least three of four measures, and the 8 older adults with the lowest span scores, on at least three of four measures, were selected. From a group of 32 young adults who were also given the battery of working memory tests, 8 high-span and 8 low-span young adults were selected to match the distributions of the groups of high- and low-span older adults on the working memory battery. Ten very high-span young adults, whose working memory scores were above the range of scores by the older adults on at least three of the four span measures, were excluded. In addition, six young adults were excluded because their scores were intermediate between those of the high- and low-span older adults. All older participants were community-dwelling adults who were recruited from a registry of prior research participants. All young participants were college students recruited by means of postings on campus bulletin boards and class announcements. All participants were monolingual speakers of English. All were paid a modest honorarium for their participation. More detailed participant data are given in Table 2. On the basis of a 2 × 2 ANOVA comparing age groups and span groups, the participants did not differ in educational level.

The participants were also given a survey of reading habits, eliciting information on how many hours per week they spent reading and what types of materials they read. The age groups and span groups did not differ in how much reading they did per week ($M = 28$ hr/week, $SD = 8.30$). The age groups did differ in how they distributed these hours across different types of reading materials; not surprisingly, the young adults reported reading college textbooks as well as readings assigned by their classes, including literary works such as novels and plays as well as books on contemporary social issues. The older adults reported more newspaper and magazine reading and reading of historical nonfiction. The older adults ($M = 35.60$ of 40 correct, $SD = 3.80$) did score higher on the vocabulary test than the young adults ($M = 30.60$, $SD = 2.80$), $F(1, 28) = 15.10$, $p = .01$, $\eta^2 = .96$; the span group main effect and the Age Group × Span Group interaction were not significant.

Working Memory Measures

The battery of working memory tests included the Digits Forward and Digits Backward tests from the WAIS-R (Wechsler, 1958) as well as the Daneman and Carpenter Reading Test (Daneman & Carpenter, 1980). In addition, a grammaticality judgment test modeled after that of Waters and Caplan (1996a, 1996b) was administered. Acceptable sentences (e.g., It was the mother who carried the baby) and unacceptable sentences (e.g., It was the baby who carried the mother) were presented one at a time on a computer screen. Participants were timed as they decided whether or not the sentence made sense. Response time and accuracy were recorded. The sentences were also divided into sets of increasing length, as in the Daneman and Carpenter test. After the final sentence in each set had been presented, participants were asked to recall the final word of each sentence in the set. Thus, this test yields three measures: a span measure, referred to

Table 2
Comparison of Young and Older Adults Classified as High- and Low-Span Participants in Experiment 2

	Young adults				Older adults			
	High span		Low span		High span		Low span	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age	20.0	2.1	19.5	1.6	75.6	3.2	74.4	5.1
Education	13.5	0.5	13.6	0.5	14.8	2.2	14.5	3.0
Vocabulary	30.4	1.4	30.8	3.5	36.1	2.6	35.0	4.6
WAIS–R—Digits Forward	7.5	0.9	5.5	0.5	6.8	0.8	4.2	0.9
WAIS–R—Digits Backward	6.1	0.8	3.4	0.5	5.9	0.7	3.4	0.5
Reading Span	4.3	0.5	2.8	0.5	4.2	0.4	2.1	0.4
WC span	3.8	0.3	2.3	0.4	3.5	0.3	2.1	0.5
WC RT (s)	3.9	0.5	5.8	0.4	4.3	0.6	6.1	0.6
WC accuracy (%)	98.5	1.4	96.8	2.4	97.6	1.1	97.0	1.3

Note. WAIS–R = Wechsler Adult Intelligence Scale—Revised; WC = Waters-Caplan; RT = response time.

as the Waters-Caplan (WC span), a reaction time measure (WC RT), and an accuracy score (WC accuracy). Waters and Caplan also computed a composite measure, comp *Z*, by converting each score to a *z* score and averaging the three *z* scores; this composite was not used in the current study. The digit span and reading span measures were highly correlated, and these measures also correlated with the WC RT measure, and, with the exception of Digits Forward, the WC span measure. The WC accuracy measure was highly correlated with the participants' educational level and vocabulary scores (Table 3). These correlations are based on the entire sample of 24 older adults and 36 young adults who were screened for participation in the reading study.

A series of 2 × 2 ANOVAs was conducted to compare the participants based on age group and span group. High- and low-span groups did differ on the four span measures (Digits Forward, Digits Backward, reading span, and WC span), all $Fs(1, 28) \geq 11.53$, $p < .01$, $\eta^2 \geq .71$. They also differed in WC RT, $F(1, 28) = 89.28$, $p < .01$, $\eta^2 = .78$, but not on the WC accuracy score, $F(1, 28) = 2.91$, $p = .10$, $\eta^2 = .02$. The age groups did not differ on any of the span measures, the WC accuracy measure, or WC RT. None of the Age Group × Span Group interactions were significant.

Materials

Two types of sentences were used in this study. In addition to the MV/RRC sentences used in the first experiment, a second type of sentence was also used. The second type also contained temporary syntactic ambiguities distinguished by the use of "the" or "only" in the initial subject NP ("the/only" sentences). Originally developed by Ni et al. (1996), they differ somewhat in their internal phrase structure and decomposition into critical regions from the MacDonald et al. sentences. The critical contrast is between Examples 2a and 2b; each contains an RRC followed by an MV. The quantifier "only" in Example 2b, according to Ni et al., serves to block the ambiguity by focusing the interpretation of the subject NP on a subset of the subject NP; that subset is specified by the following expression, the first verb (1V). Hence, the 1V is immediately interpreted as an RRC modifying the subject NP, not as an MV. Examples 2c and 2d are controls; both contain unambiguous RRC sentences (the main verb interpretation of the 1V is blocked by its morphological form, e.g., "stolen" cannot be used as an MV except as a passive "were stolen").

Subject NP | 1V | 2V | 2VP (2)

The businessmen | loaned money at low interest | were told | to record their expenses. (2a)

Only businessmen | loaned money at low interest | were told | to record their expenses. (2b)

The vans | stolen from the parking lot | were found | in a back alley. (2c)

Only vans | stolen from the parking lot | were found | in a back alley. (2d)

There were 40 sets of MV/RRC sentences and 48 sets of "the/only" sentences. They were assigned to four stimulus lists such that each list contained 10 examples of each type of MV/RRC sentence and 12 examples of each type of "the/only" sentences but only 1 sentence from each set. In addition to the experimental sentences, each list contained 112 filler sentences of various syntactic forms. Unlike the studies of Waters and Caplan (2001, 2002), only grammatically well-formed, semantically acceptable sentences were used in the reading study. The lists were randomized, and each was broken into four blocks of 50 sentences.

Probe questions were inserted after 50% of the sentences, randomly selected. The probe questions for the fillers and control sentences tested for comprehension of the MV or sentence predicate. "Don't know" was the correct answer to 10% of the probe questions about filler sentences. The probe questions for the experimental MV/RRC sentences tested for the correct interpretation of the 1VP. Thus, for Examples 1a and 1b, the probe was "Who warned about the dangers?" For Example 1a, it is correctly answered by "the soldiers," indicating an MV interpretation; for Example 1b, answering "the soldiers" indicates an incorrect MV interpretation. Correct answers to Example 1b were "don't know" or "someone else." There were two types of questions for "the/only" sentences. Half probed for correct interpretation of the first VP. For Examples 2a and 2b, the probe was "Who loaned the money?" Answering "the businessmen" indicates an incorrect MV interpretation in either case. Correct answers for Examples 2a or 2b were "the bank," "don't know," or "someone else." The second type of question for "the/only" sentences probed for correct interpretation of the 2VP. For Examples 2a and 2b, the probe was "Who recorded their expenses?" The correct answer was "the businessmen."

Procedure

Testing took place over 2 days. On the first day, participants were given the battery of working memory tests. Those selected for inclusion in the high- or low-span groups were then invited back a second day, when the reading study was administered. Participants were first acquainted with the

Table 3
Correlations Between Participants' Scores on the Working Memory Measures: Education and Vocabulary

Variable	1	2	3	4	5	6	7	8
1. Education	—	.228	−.105	−.113	−.186	−.125	.013	.457*
2. Vocabulary	.361	—	.231	.187	.142	.213	.011	.428*
3. WAIS–R—Digits Forward	−.164	.224	—	.746**	.753**	.336	.659**	−.212
4. WAIS–R—Digits Backward	.084	.071	.731**	—	.652**	.475*	.631**	−.174
5. Reading Span	−.202	.246	.690**	.694**	—	.465*	.642**	−.215
6. WC span	−.076	.126	.289	.407*	.417*	—	.625**	−.127
7. WC RT	.058	−.048	−.700**	−.775**	−.730**	−.650**	—	−.179
8. WC accuracy	.478*	.435*	−.312	−.310	−.255	−.215	−.118	—

Note. Correlations for 36 young adults screened for participation are presented in the upper half; those for the 24 older adults screened for participation are presented in the lower half.
WAIS–R = Wechsler Adult Intelligence Scale-Revised; WC = Waters and Caplan; RT = response time.
* $p < .05$. ** $p < .01$.

equipment and then given a block of 20 practice sentences to read; probe questions accompanied 50%. The experimenter recorded the participants' responses to the probe questions. Each participant was assigned randomly to one of four stimulus lists. Then four blocks of 50 sentences were presented; blocks were counterbalanced across participants. Other details of the presentation and analysis of eye fixation patterns were the same as those followed in Experiment 1. As in Experiment 1, first pass and total fixation times were analyzed using length-adjusted residual reading times. Unadjusted, raw first-pass and total fixation times may be found at dx.doi.org/10.1037/0882-7974.19.1.157.supp.

Results

The MV/RRC sentences and "the/only" sentences were analyzed separately. Each analysis involved four measures: first-pass fixation times, the mean number of regressions to the subject NP, total fixation times, and accuracy in answering the probe questions. An omnibus ANOVA with age group, critical region, sentence type, and ambiguity is reported followed by a region-by-region decomposition of significant effects and interactions. Lower order main effects and interactions subsumed by higher order interactions are not reported. Note that none of the main effects for age group or interactions involving this factor were significant.

MV/RRC Sentences

A preliminary analysis compared residual first-pass fixation time, regressions to the subject NP, and residual total fixation time for the 1st versus 4th block of trials. The effect of block was nonsignificant and there were no significant interactions with the block factor. A follow-up analysis examined fixation times and regression data only for those sentences corresponding to probe questions the participants answered correctly. The overall pattern of results was similar to that reported below.

First-Pass Fixation Time. A 2 (span group) × 2 (age group) × 3 (region) × 2 (sentence type) × 2 (ambiguity) ANOVA compared first-pass fixation times for the three critical regions. The results are summarized in Figure 4. The three-way Region × Sentence Type × Ambiguity interaction was the only significant finding, $F(2, 27) = 84.69$, $p < .01$, $\eta^2 = .86$. At Region 1, the subject NP, there were no significant effects or interactions: First-pass fixation times did not vary with age group, span group, sentence type, or ambiguity. This was also true for Region 2, 1VP. At Region 3, the 2V/PP, first-pass fixation times for unambiguous control sentences and ambiguous MV sentences were faster than those for ambiguous RRC sentences; there were no age group or span group differences in first-pass fixation times at Region 3.

Regressions to the subject NP. A 2 (span group) × 2 (age group) × 2 (region) × 2 (sentence type) × 2 (ambiguity) ANOVA compared the mean number of regressions to the subject NP from the other two critical regions, 1VP and 2V/PP. The results are summarized in Figure 5. Only the Span Group × Ambiguity interaction was significant, $F(1, 28) = 47.69$, $p = .01$, $\eta^2 = .61$. High-span participants made few regressions from the 1VP or 2V/PP region ($M = 0.21$, $SE = 0.11$) to the subject NP. Low-span readers made few regressions for unambiguous sentences from the 1VP or 2V/PP ($M = 0.27$, $SE = 0.12$) to the subject NP but made more regressions for ambiguous sentences ($M = 0.89$, $SE = 0.10$). The average duration of first-pass regressions was analyzed with a similar ANOVA comparing span groups, age groups, region, sentence type, and ambiguity. The two-way Span Group × Ambiguity interaction was significant, $F(2, 27) = 41.21$, $p < .01$, $\eta^2 = .76$. High-span readers' regressions did not differ with condition ($M_{Duration} = 0.08$, $SE = 0.04$); low-span readers made longer regressions than high-span readers ($M_{Duration} = 0.12$, $SE = 0.07$), especially for ambiguous sentences ($M_{Duration} = 0.24$, $SE = 0.05$).

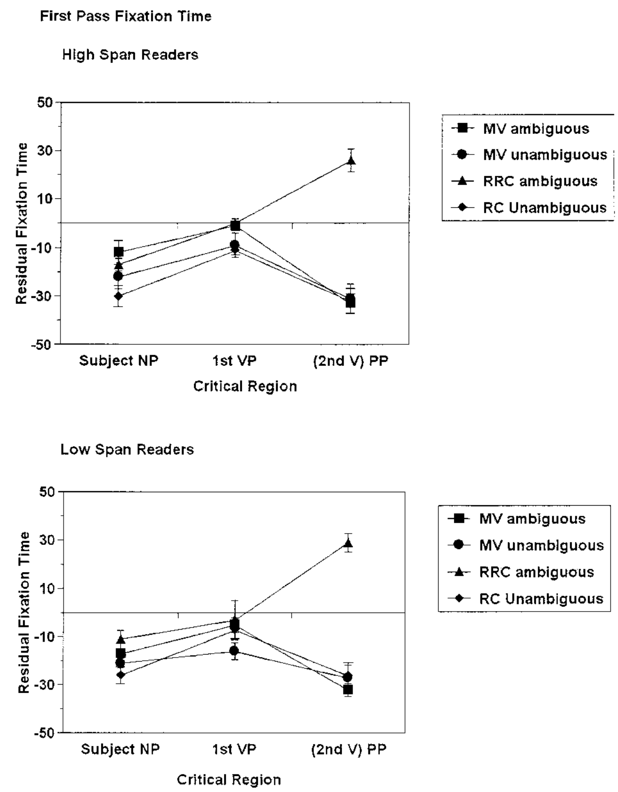

Figure 4. Residual first-pass fixation time (and *SE*s) for high- and low-span readers for ambiguous main verb (MV) and reduced relative clause (RRC) sentences and their unambiguous controls. NP = noun phrase; VP = verb phrase; V = verb; PP = prepositional phrase; RC = relative clause.

Total fixation time. A 2 (age group) × 2 (span group) × 3 (region) × 2 (sentence type) × 2 (ambiguity) ANOVA compared total fixation times for the three critical regions: subject NP, 1VP, and 2V/PP. The results are summarized in Figure 6. The four-way interaction of span group, region, sentence type, and ambiguity was marginally significant, $F(2, 27) = 3.26$, $p = .05$, $\eta^2 = .20$. At Region 1, subject NP, the Ambiguity × Span Group interaction was significant, $F(1, 28) = 7.86$, $p = .01$, $\eta^2 = .18$. Total fixation times for high-span readers did not vary with sentence type or ambiguity. Total fixation times for low-span readers for ambiguous sentences were inflated by regressions to this region from Regions 2 and 3. As a result, total fixation times for low-span readers to ambiguous sentences were longer than those for unambiguous sentences.

In the analysis of total fixation times for Region 2, 1VP, the Sentence Type × Ambiguity × Span Group interaction was significant, $F(1, 28) = 12.13$, $p = .01$, $\eta^2 = .25$. Total fixation times for high-span readers did not vary with sentence type or ambiguity. Total fixation times for low-span readers for ambiguous sentences

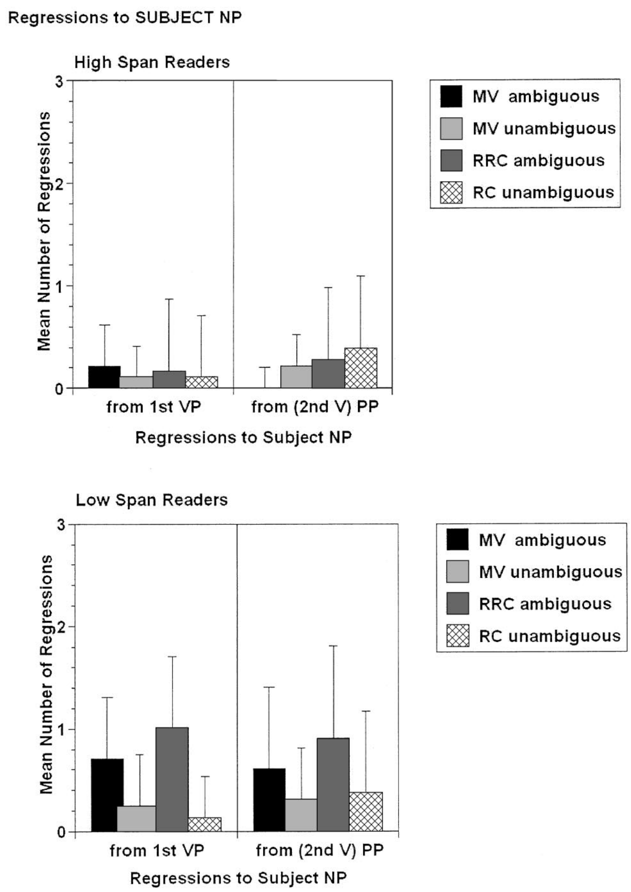

Figure 5. Regressions (and *SE*s) to the subject noun phrase (NP) for high- and low-span readers for ambiguous main verb (MV) and reduced relative clause (RRC) sentences and their unambiguous controls. RC = relative clause; VP = verb phrase; V = verb; PP = prepositional phrase.

were inflated by refixations to this region after regressions to the subject NP. As a result, total fixation times for low-span readers were longer for ambiguous sentences than for unambiguous sentences.

The Sentence Type × Ambiguity interaction was significant, $F(1, 28) = 24.58$, $p < .01$, $\eta^2 = .41$, in the analysis of total fixation times for Region 3, 2V/PP. For both high- and low-span readers, total fixation times to the disambiguating 2V in ambiguous RRC sentences were longer than those to the disambiguating PP in ambiguous MV sentences or the 2V or PP in unambiguous sentences.

Probe questions. A 2 (age group) × 2 (span group) × 2 (sentence type) × 2 (ambiguity) ANOVA was used to analyze the percentage of probe questions answered correctly. The results are reported in Table 4. The Type × Ambiguity × Span interaction was significant, $F(1, 28) = 10.97$, $p < .01$, $\eta^2 = .28$. High-span readers answered 86% of the questions correctly regardless of sentence type or ambiguity. Low-span readers answered 76% correctly for questions about ambiguous and unambiguous MV sentences and unambiguous RRC sentences. Low-span readers answered correctly only 55% of the probe questions about the ambiguous RRC sentences. Note that the correct answer was "Don't know." High-span readers answered 98% of the questions correctly about the filler sentences; low-span readers answered 84% correctly, including the 10% of the filler probe questions that were correctly answered by "Don't know." This suggests that the low-span readers were able to respond "Don't know" when they were uncertain of the answer to the probe question and that their incorrect answers to the questions about ambiguous sentences represent actual misinterpretations, rather than a reluctance to respond "Don't know." Low-span readers incorrectly answered questions about the RRC sentences with the subject NP 36% of the time and gave other incorrect answers 9% of the time.

Summary. First-pass fixations to the subject NP or 1VP did not vary by sentence type, ambiguity, or span group. Only at Region 3 did first-pass fixation times begin to diverge as a function of sentence type and ambiguity. Region 3 corresponds to the 2VP of the RRC sentences or the PP of MV sentences; it thus disambiguates the syntactic role of 1VP in Region 2. At Region 3, first-pass fixation times by both high- and low-span readers to RRC sentences with temporary syntactic ambiguities were longer than those to MV sentences or unambiguous control sentences. Low-span readers encountered difficulty interpreting the ambiguous MV and RRC sentences and made many regressions to the subject NP in order to interpret 1V and 2V/PP regions. In contrast to Experiment 1, these regressions inflated their total fixation times for Regions 1 and 2 of the ambiguous sentences, suggesting the

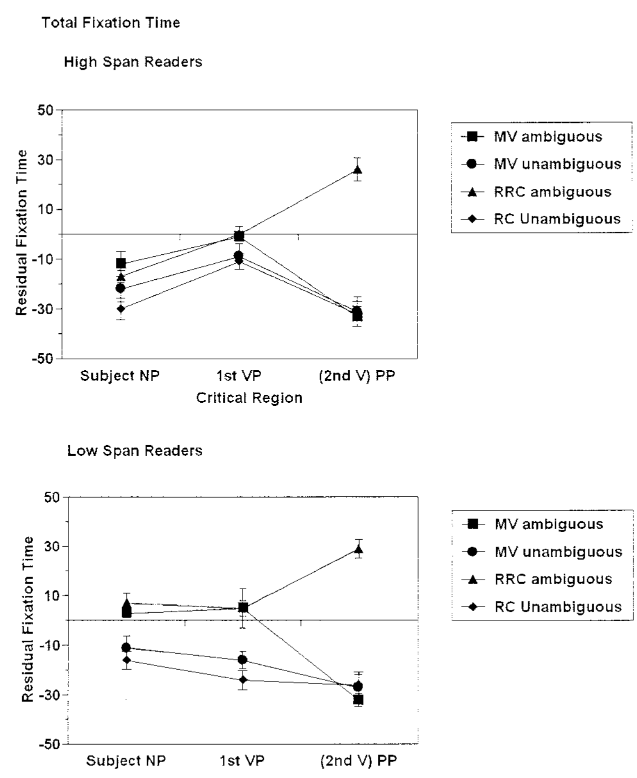

Figure 6. Residual total fixation time (and *SE*s) for high- and low-span readers for ambiguous main verb (MV) and reduced relative clause (RRC) sentences and their unambiguous controls. RC = relative clause; NP = noun phrase; VP = verb phrase; V = verb; PP = prepositional phrase.

Table 4
Percentage of Questions Answered Correctly by Young and Older Adults Classified as High- and Low-Span Participants for MV and RRC Sentences

	Young adults				Older adults			
	High span		Low span		High span		Low span	
Sentence type	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
MV ambiguous (%)	88	3.4	75	3.4	91	3.4	73	3.4
RRC ambiguous (%)	84	3.2	54	3.2	86	3.3	56	3.2
MV control (%)	84	3.2	77	3.3	82	3.4	76	3.3
RRC control (%)	85	2.9	76	2.9	86	3.3	78	2.9

Note. MV = main verb; RRC = reduced relative clause.

low-span readers were attempting to carefully resolve the ambiguity in order to answer the probe questions. Total fixation times for all readers were longer for Region 3 for ambiguous RRC sentences, corresponding with the disambiguating information. Despite their attempts to resolve the ambiguity by making regressions to the subject NP, low-span readers often did not correctly interpret the RRC sentences as indicated by their responses to the probe questions. Although high- and low-span readers produced different patterns of eye fixations while reading MV/RRC sentence, those of young and older adults were similar when matched for memory span.

"The/Only" Sentences

A preliminary analysis compared residual first-pass fixation time, regressions to the Subject NP, and residual total fixation time for the first versus fourth block of trials. The effect of block was nonsignificant and there were no significant interactions with the block factor. A second analysis examined fixation times and regression data only for those sentences corresponding with probe questions the participants answered correctly. The overall pattern of results was similar to that reported below.

First-pass fixation time. A 2 (span group) × 2 (age group) × 4 (region) × 2 (sentence type) × 2 (ambiguity) ANOVA compared first-pass fixation times for the four critical regions, the subject NP, 1V, 2V, and 2VP. The results are summarized in Figure 7. The four-way interaction of span group, region, sentence type, and ambiguity was significant, $F(3, 26) = 16.64$, $p < .01$, $\eta^2 = .66$. At Region 1, the subject NP, the Type × Span Group interaction was significant, $F(1, 28) = 13.87$, $p < .01$, $\eta^2 = .33$. First-pass fixation times for low-span readers did not vary with sentence type. First-pass fixation times for high-span readers varied with sentence type; their first-pass fixation times to "only" sentences were longer than their first-pass fixations to "the" sentences. For Region 2, 1VP, there were no significant effects or interactions. First-pass fixation times did not vary with sentence type or ambiguity for either span group.

For Region 3, 2V, the Type × Ambiguity × Span interaction was significant, $F(1, 28) = 29.22$, $p < .01$, $\eta^2 = .51$. Low-span readers' first-pass fixation times differed with ambiguity but not with sentence type: First-pass fixation times to ambiguous sentences were longer than those to unambiguous sentences for Region 3. For high-span readers, first-pass fixation times for unambiguous "the" and "only" sentences were similar for Region 3 and similar to those for ambiguous "only" sentences whereas first-pass fixation times to ambiguous "the" sentences were slower for Region 3. For Region 4, 2VP, there were no significant effects or interactions. First-pass fixation time did not vary with sentence type or ambiguity for either span group.

Regressions to the subject NP. A 2 (span group) × 2 (age group) × 3 (region) × 2 (sentence type) × 2 (ambiguity) ANOVA compared the mean number of regressions to the subject NP from the other three critical regions: 1VP, 2V, and 2VP. The results are summarized in Figure 8. The three-way Span Group × Region × Ambiguity interaction was significant. High-span participants made few regressions from the 1VP or 2V to the subject NP for either ambiguous or unambiguous sentences ($M = 0.17$, $SE = 0.08$). Low-span readers made more regressions than high-span readers and made more regressions from Region 3, 2V ($M = 1.37$, $SE = 0.08$) than from Region 2, 1VP ($M = 1.08$, $SE = 0.11$) or Region 4 ($M = 0.48$, $SE = 0.11$) for ambiguous sentences. Low-span readers made few regressions from any region of unambiguous sentences ($M = 0.21$, $SE = 0.11$). The average dura-

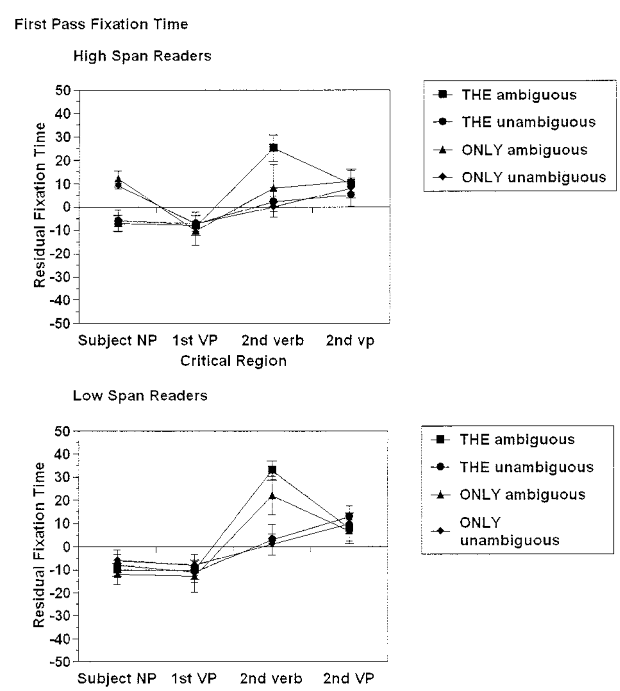

Figure 7. Residual first-pass fixation time (and *SE*s) for high- and low-span readers for ambiguous "the" and "only" sentences and their unambiguous controls. NP = noun phrase; VP = verb phrase.

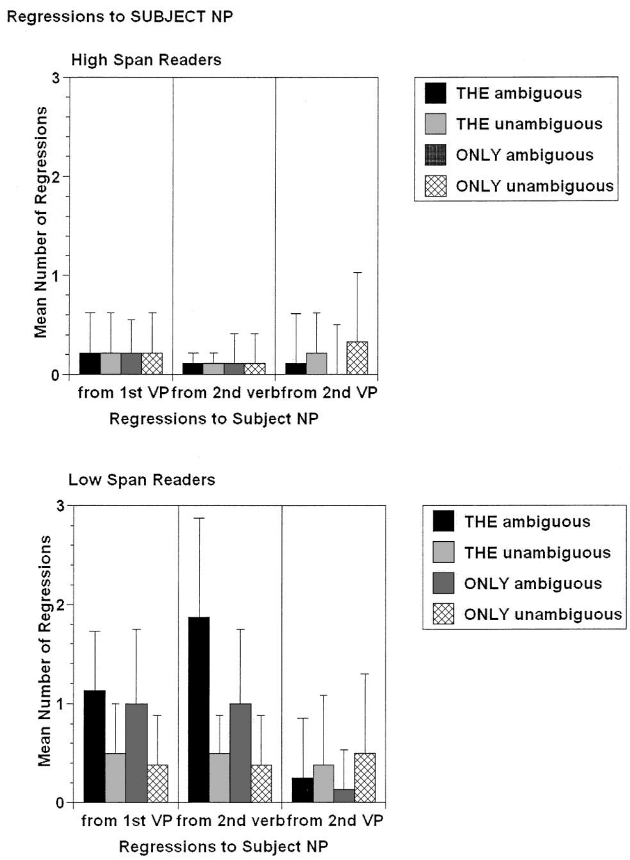

Figure 8. Regressions (and *SE*s) to the subject noun phrase (NP) for high- and low-span readers for ambiguous "the" and "only" sentences and their unambiguous controls. VP = verb phrase.

tion of first-pass regressions to the subject NP was analyzed with a similar ANOVA comparing span groups, age groups, region, sentence type, and ambiguity. The two-way Span Group × Ambiguity interaction was significant, $F(2, 27) = 14.52$, $p < .01$, $\eta^2 = .403$. High-span readers' regressions did not differ with condition ($M_{\text{Duration}} = 0.04$, $SE = 0.07$); low-span readers made longer regressions than high-span readers ($M_{\text{Duration}} = 0.12$, $SE = 0.08$), especially for ambiguous sentences ($M_{\text{Duration}} = 0.18$, $SE = 0.05$).

Total fixation time. A 2 (age group) × 2 (span group) × 4 (region) × 2 (sentence type) × 2 (ambiguity) ANOVA compared total fixation time for the four critical regions: subject NP, 1VP, 2V, and 2VP. The results are summarized in Figure 9. The four-way Span Group × Region × Sentence Type × Ambiguity interaction was significant, $F(3, 26) = 12.74$, $p < .01$, $\eta^2 = .56$. For Region 1, subject NP, the Sentence Type × Span Group interaction, $F(1, 28) = 43.83$, $p < .01$, $\eta^2 = .62$, and Ambiguity × Span Group interaction were significant, $F(1, 28) = 35.74$, $p < .01$, $\eta^2 = .56$. Total fixation times for high-span readers for "the" sentences were shorter than their total fixation times to "only" sentences. In contrast, for low-span readers, total fixation times for "the" and "only" sentences did not differ. However, total fixation times for ambiguous and unambiguous sentences for high-span readers did not differ, whereas those for low-span readers were longer for ambiguous sentences than for unambiguous sentences, reflecting their more numerous and longer regressions to the subject NP for the ambiguous sentences.

For total fixation time for Region 2, 1VP, the Ambiguity × Span Group interaction was marginally significant, $F(1, 28) = 3.24$, $p < .08$, $\eta^2 = .10$. Total fixation times for high-span readers were similar for ambiguous and unambiguous sentences, whereas among low-span readers they were longer for ambiguous sentences than for unambiguous sentences.

In the analysis of total fixation times for Region 3, 2VP, the Sentence Type × Ambiguity × Span Group interaction was significant, $F(1, 28) = 14.53$, $p = .01$, $\eta^2 = .34$. For high-span readers, total fixation times to ambiguous "the" sentences were longer than those to the other sentences. For low-span readers, total fixation times for both types of ambiguous sentences were longer than that for unambiguous sentences. At Region 4, 2V, there were no significant main effects or interactions. Total fixation time did not differ with sentence type or ambiguity.

Probe questions. The two types of probe questions were analyzed separately. A 2 (age group) × 2 (span group) × 2 (sentence type) × 2 (ambiguity) ANOVA was used to analyze the percentage of probe questions about the first verb that were answered correctly. For the first probe question about the 1VP, the Ambiguity × Span interaction was significant, $F(1, 28) = 57.87$, $p < .01$, $\eta^2 = .87$. High-span readers correctly answered 94% of the questions about unambiguous sentences and ambiguous "the/only" sentences. Low-span readers answered correctly 90% of the questions about unambiguous sentences but only 44% of the questions about ambiguous sentences. Low-span readers answered 39% of

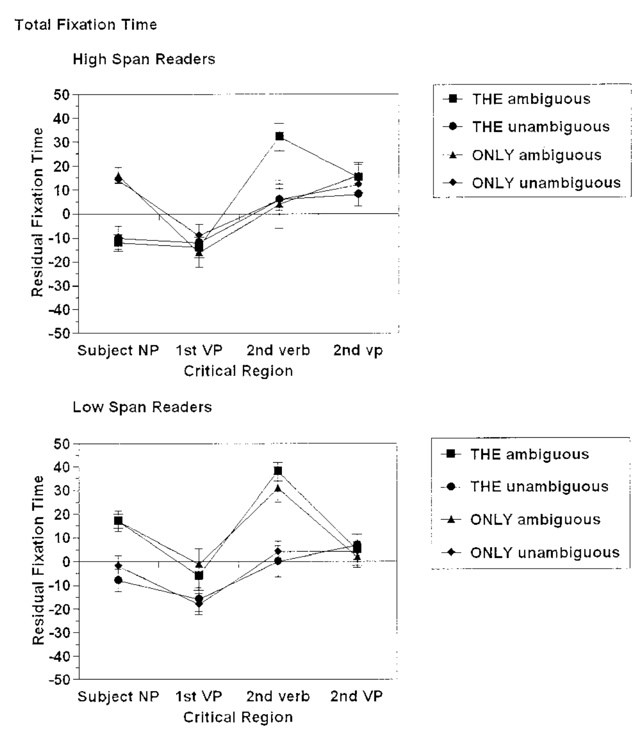

Figure 9. Residual total fixation time (and *SE*s) for high- and low-span readers for ambiguous "the" and "only" sentences and their unambiguous controls. VP = verb phrase.

the questions about ambiguous “the/only” incorrectly with the subject NP and gave other incorrect responses to 16% of the probe questions. Note that the correct answer to the first probe question was “Don’t know.” The results are reported in Table 5.

For the second probe question about the 2VP, only the main effect of sentence type was significant, $F(1, 28) = 8.80$, $p = .01$, $\eta^2 = .24$. Readers answered correctly 94% of the questions about “the” sentences but only 87% of the questions about “only” sentences.

Summary. When the quantifier “only” is first encountered, first-pass fixation times by high-span readers tended to be longer than those by the low-span readers. This suggests that high-span readers required extra time to assess the contrast implied by the focus operator and low-span readers were insensitive to this contrast. At Region 3, first-pass fixation times and total fixation times by both high- and low-span readers to ambiguous “the” sentences were longer than those to unambiguous “the” sentences, indicating that readers were using the disambiguating 2V to resolve the ambiguity. High-span readers were able to avoid this ambiguity effect at Region 3 for “only” sentences because they had allocated more time to Region 1 to interpreting the contrastive focus implied by “only.” Low-span readers were insensitive to the focus operator; consequently, their first-pass fixation times as well as total fixation times for Region 3 of ambiguous “only” and ambiguous “the” sentences were similar. Low-span readers encountered difficulty interpreting the ambiguous sentences and made more and longer regressions to the subject NP to interpret the 1V and 2VP regions. As a result, their total fixation times were inflated for Regions 1 and 2 of the ambiguous sentences. High-span readers often incorrectly interpreted the ambiguous “the” sentences but were able to use the “only” quantifier to focus on the correct interpretation of ambiguous “only” sentences. Despite their attempts to resolve the ambiguity by making regressions to the subject NP, low-span readers often did not correctly interpret either ambiguous “the” or “only” sentences. Although high- and low-span readers produced different patterns of eye fixations while reading “the/only” sentence, those of young and older adults were similar when matched for memory span. Whereas the results from “the” sentences support the SSIR model, those from “only” sentences reveal an effect of working memory span on first-pass interpretive processing. This study also confirms the original report by Ni et al. (1996) of different fixation patterns for high- and low-span readers for “the” and “only” sentences and suggests that Liversedge et al. (1998) and Clifton et al. (2000) failed to find different fixation patterns for “the” and “only” sentences because they did not compare fixation patterns of high- and low-span readers.

General Discussion

First-pass fixation times for ambiguous MV and RRC sentences show similar garden-path effects: All readers, regardless of age or working memory span, experience garden-path effects for ambiguous RRC sentences, with first-pass fixations peaking in Region 3. However, young and older adults in Experiment 1 and high- and low-span readers in Experiment 2 did produce different patterns of eye fixations when regressions and total fixation time were analyzed. Older adults in Experiment 1 and low-span readers in Experiment 2 made more regressions from Regions 2 and 3 to the subject NP for ambiguous MV/RRC sentences in order to resolve the ambiguity. In contrast, young adults in Experiment 1 and high-span readers in Experiment 2 were able to resolve the syntactic ambiguities without recourse to leftward regressions by relying on information in working memory.

Experiment 2 revealed three differences between high- and low-span readers that were not assessed in Experiment 1. First, introducing probe questions in Experiment 2 resulted in an increase in the duration of first-pass regressions by low-span readers to ambiguous sentences that was not observed in Experiment 1 for older adults. Whereas the older adults in Experiment 1 appear to have quickly reread the subject NP of ambiguous sentences, the low-span readers in Experiment 2 appear to have reanalyzed the status of the subject NP in response to the probe questions. The probe questions required them to identify the subject of either the first or second verb in order to answer the questions correctly. Second, high-span readers were able to answer probe questions about all types of sentences correctly with 80% or better accuracy. Low-span readers were able to answer the probe questions about unambiguous sentences as well as those about MV sentences with a similar level of accuracy. However, they often misinterpreted ambiguous RRC sentences, incorrectly assigning an MV interpretation. Thus, the low-span readers’ more numerous and longer regressions to the subject NP did not necessarily result in a correct interpretation of the ambiguous RRC sentences.

“The/only” sentences in Experiment 2 revealed another difference in first-pass fixation times as a function of working memory span: First-pass fixation time for high- and low-span readers diverged when the focus operator “only” was used to restrict the

Table 5
Percentage of Questions About the First Verb Phrase Answered Correctly by Young and Older Adults Classified as High- and Low-Span Participants for “The” and “Only” Sentences

	Young adults				Older adults			
	High span		Low span		High span		Low span	
Sentence type	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
“The” ambiguous (%)	93	0.9	44	3.9	94	2.4	45	1.4
“Only” ambiguous (%)	93	0.9	42	3.9	94	1.1	45	2.9
“The” unambiguous (%)	94	0.7	90	0.6	94	0.7	90	0.8
“Only” unambiguous (%)	93	1.4	94	0.9	92	3.5	88	3.5

interpretation of the subject NP. High-span readers allocated additional fixation time to the subject NP marked by the "only" operator presumably to assess its function; low-span readers did not appear to be sensitive to the contrast implied by "only." First-pass fixation time by all readers revealed a garden-path effect for the ambiguous "the" sentences; however, high-span readers, unlike low-span readers, were able to avoid the garden-path effect for the "only" sentences because they had allocated extra first-pass fixation time to subject NPs marked by the "only" operator.

These results are not fully consistent with the CC model of Just and Carpenter (1992) (also the 3CAPS model of Just & Varma, 2002). The CC model argues that the capacity of working memory constrains language processing as well as other cognitive abilities. Not only should individuals with working memory limitations have increased difficulty processing syntactically complex sentences, but also the time course of sentence processing for simple versus complex sentences should vary with working memory. Low-span readers should allocate additional processing time at points of processing difficulty and be unable to retain multiple, alternative syntactic representations of ambiguous phrases. In contrast, high-span readers should be able to avoid garden-path effects in that they can compute and retain multiple representations of ambiguous phrases until disambiguating information is available. The CC model is supported by the regression patterns: Low-span readers had more difficulty processing the ambiguous sentences, particularly Regions 2 and 3, and they were unable to rely on information in working memory in order to do so. The finding that low-span individuals had difficulty correctly answering the probe questions about the ambiguous RRC and "the" sentences also supports this model. However, the similarity of first-pass fixation times for high- and low-span individuals for MV/RRC sentences in Experiment 2 does not support the CC model. Further, both high- and low-span readers showed a marked garden-path effect at Region 3 for the RRC sentences. The differences in first-pass fixations for high- and low-span readers for "the/only" sentences are consistent with the CC model: High-span readers allocated additional time to interpreting the subject NP in "only" sentences and were able to avoid the garden path for Region 3. However, high- and low-span readers produced similar first-pass fixations patterns for "the" sentences as they did for the MV/RRC sentences, and both groups experienced marked garden-path effects for these sentences, contrary to the CC model's predictions.

The Just and Carpenter (1992) model could be revised to accommodate these results while preserving a single capacity-limited working memory system involved in sentence processing and other cognitive abilities. In the revised model, both high- and low-span individuals experience garden-path effects because of temporary syntactic ambiguities. However, immediate sentence processing is affected by working memory limitations in two ways. First, low-span individuals are unable to quickly resolve temporary syntactic ambiguities; as a consequence, low-span individuals engage in regressive eye movements in order to recover from misanalyses, inflating total fixation times. High-span individuals can access relevant information in working memory to quickly resolve temporary syntactic ambiguities once they are detected. Second, low-span individuals are either not sensitive to the contrast in focus between "the" and "only" or they are unable to apply the "only" operator to restrict the interpretation of the subsequent NP. High-span readers can avoid garden-path effects by applying the focus operator to restrict their initial interpretation. Hence, a larger working memory enables readers to retain linguistic information and access it quickly to recover from initial misanalyses and to apply linguistic constraints to avoid such misanalyses; a smaller working memory limits readers' access to low-frequency linguistic information and forces them to compensate for initial misanalyses by rereading and reanalyzing critical linguistic information.

These results are also inconsistent with a more recent model put forth by MacDonald and Christiansen (2002). This model concedes an overall advantage to high-span individuals in terms of speed and accuracy of sentence processing but predicts similar patterns of online processing. This model also equates span differences in working memory with differences in linguistic expertise, affecting vocabulary knowledge as well as language processing. Hence, the finding that span scores were not highly correlated with vocabulary scores is inconsistent with this model. High-span readers do appear to have an advantage over low-span readers with regard to at least one type of linguistic knowledge: High-span readers appreciated the significance of the low-frequency "only" operator and were able to avoid garden-path effects for this type of sentence. High-span readers have presumably learned that "only" co-occurs with an NP modifier—in this case, the RRC—thereby blocking the MV interpretation. The MacDonald and Christiansen model is also unable to account for the marked differences in regressions for high- and low-span readers; it was not simply that low-span readers made more regressions, but they made more regressions for ambiguous sentences than for unambiguous sentences, an interaction that is inconsistent with the model.

The SSIR model of Waters and Caplan (1996a; Caplan & Waters, 1999) holds that there is a specialized working memory system for interpretive processing that is at least partially distinct from a more general working memory system involved in postinterpretive processing. Individual or group differences as measured by general working memory tests are linked to differences in postinterpretive processing affecting, for example, question–answering and sentence recall. Interpretive processing, involved in lexical access, syntactic analysis, focus, and thematic role assignment, should not be affected by such working memory limitations. Hence, the time course of immediate, interpretive processing should be invariant across groups and individuals. This model is supported by several findings in the current studies. First, both young and older adults experienced a garden-path effect for first-pass fixation times with ambiguous RRC sentences in Experiment 1, and both high- and low-span readers experienced similar garden-path effects for ambiguous RRC sentences in Experiment 2.

However, one aspect of the results is not consistent with the SSIR model. Although first-pass fixations by high- and low-span readers were similar for the MV, RRC, and "the" sentences in Experiment 2, they differed markedly for "only" sentences. High-span readers allocated additional first-pass time to processing subject NPs with the "only" operator and thereby avoided the garden-path effect at Region 3, whereas first-pass fixations for low-span readers did not differ for "the/only" sentences. First-pass fixations are generally considered to reflect immediate, online syntactic analyses and thus correspond to Waters and Caplan's interpretive processes. Thus, finding an effect of working memory span on first-pass processing is troubling for the SSIR theory. It may be that Waters and Caplan's (2001) inventory of "largely unconscious, obligatory, on-line, first-pass" (p. 128) interpretive processes must be redefined to exclude discourse-level semantic processes such as the application of the focus operator "only."

Because regressions and total fixation times are generally considered to reflect initial parsing problems and reanalysis, the results of Experiment 1 and 2 are generally supportive of the SSIR model. In Experiment 1, older readers' regressions did not affect total fixation times for the ambiguous sentences, suggesting the older readers were merely rereading the subject NP. In Experiment 2, low-span readers' regressions to the subject NP for ambiguous sentences resulted in inflated total fixation times, suggesting they were verifying their resolution of the ambiguity carefully in anticipation of answering a probe question. Thus, the observed age differences in regressions in Experiment 1 and span differences in regressions and total fixation times in Experiment 2 result from postinterpretive processes.

In general, the fixation patterns of older adults in Experiment 1 resembled those of low-span readers in Experiment 2. These parallels between the reading patterns of older adults and low-span individuals suggest that other differences between young and older adults, such as overall differences in vocabulary, processing speed, or efficiency of inhibitory processes do not substantially contribute to older adults' sentence-processing problems. Working memory limitations, affecting the resolution of syntactic ambiguities, contribute to older adults' reading problems by leading to more regressive eye fixations, longer total fixation times, and poor comprehension of garden-path sentences.

References

Beaman, K. (1984). Coordination and subordinate revisited: Syntactic complexity in spoken and written narrative discourse. In D. Tannen (Ed.), *Coherence in spoken and written discourse* (pp. 45–80). Norwood, NJ: Ablex.

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

Caplan, D., & G. S. Waters (2002). Working memory and connectionist models of memory: A reply to MacDonald and Christiansen (2002). *Psychology Review, 109,* 66–74.

Chafe, W. L. (1982). Integration and involvement in speaking, writing, and oral literature. In D. Tannen (Ed.), *Spoken and written language* (pp. 35–54). Norwood, NJ: Ablex.

Clifton, C., Jr., Bock, J., & Radó, J. (2000). Effects of the focus particle only and intrinsic contrast on comprehension of reduced relative clauses. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 591–619). Oxford, England: Elsevier.

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

Ehrlich, K., & Rayner, K. (1983). Pronoun assignment and semantic integration during reading: Eye movements and immediacy of processing. *Journal of Verbal Learning and Verbal Behavior, 22,* 75–87.

Ferreira, F., & Anes, M. D. (1994). Why study spoken language processes? In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 33–56). San Diego: Academic Press.

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language, 25,* 348–368.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology, 14,* 178–210.

Hildyard, A., & Olson, D. R. (1982). On the comprehension and memory of oral vs. written language. In D. Tannen (Ed.), *Spoken and written language: Exploring orality and literacy* (pp. 19–34). Norwood, NJ: Ablex.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149.

Just, M. A., & Varma, S. (2002). A hybrid architecture for working memory: A reply to MacDonald & Christiansen (2002). *Psychology Review, 109,* 55–65.

Kemper, S., & Kemtes, K. A. (1999). Limitations on syntactic processing. In S. Kemper & R. Kliegl (Eds.), *Constraints on language: Aging, memory, and grammar* (pp. 79–106). New York: Kluwer Academic.

Kemtes, K. (1998). *Decomposing adults' sentence comprehension: The roles of age, working memory, inhibitory functioning, and perceptual speed.* Doctoral dissertation, University of Kansas, Lawrence, KS.

Kemtes, K. A., & Kemper, S. (1997). Younger and older adults' on-line processing of syntactic ambiguities. *Psychology and Aging, 12,* 362–371.

Kennedy, A. (1983). On looking into space: In K. Rayner (Ed.), *Eye movements in reading: Perceptual and language processes* (pp. 237–251). New York: Academic Press.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30,* 580–602.

Lankford, C. (2001). *GazeTracker.* Charlottesville, VA: ERICA, Inc.

Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 55–75). Oxford, England: Elsevier.

MacDonald, M., & Christiansen, M. H. (2002). Reassessing working memory: A reply to Just and Carpenter and Waters & Caplan. *Psychology Review, 109,* 35–54.

MacDonald, M., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology, 24,* 56–98.

Ni, W., Crain, S., & Shankweiler, D. (1996). Sidestepping garden paths: Assessing the contributions of syntax, semantics, and plausibility in resolving ambiguities. *Language and Cognitive Processes, 11,* 283–334.

Trueswell, J. C., Tanenhaus, M. K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language, 33,* 285–318.

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

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

Waters, G. S., & Caplan, D. (2001). Age, working memory, and on-line syntactic processing in sentence comprehension. *Psychology and Aging, 16,* 128–144.

Waters, G. S., & Caplan, D. (2002). Working memory and online syntactic processing in Alzheimer's disease: Studies with auditory moving window presentation. *Journal of Gerontology: Psychological Sciences, 57B,* P298–P311.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence.* Baltimore: Williams & Wilkins.

Received August 17, 2002
Revision received July 25, 2003
Accepted July 25, 2003 ■