Journal of Experimental Psychology: General
2001, Vol. 130, No. 2, 199–207

0096-3445/01/$5.00 DOI: 10.1037//0096-3445.130.2.199

Working Memory Span and the Role of Proactive Interference

Cindy Lustig
Duke University

Cynthia P. May
College of Charleston

Lynn Hasher
University of Toronto and The Rotman
Research Institute of Baycrest Centre

The authors investigated the possibility that working memory span tasks are influenced by interference and that interference contributes to the correlation between span and other measures. Younger and older adults received the span task either in the standard format or one designed to reduce the impact of interference with no impact on capacity demands. Participants then read and recalled a short prose passage. Reducing the amount of interference in the span task raised span scores, replicating previous results (C. P. May, L. Hasher, & M. J. Kane, 1999). The same interference-reducing manipulations that raised span substantially altered the relation between span and prose recall. These results suggest that span is influenced by interference, that age differences in span may be due to differences in the ability to overcome interference rather than to differences in capacity, and that interference plays an important role in the relation between span and other tasks.

Working memory is traditionally thought of as the capacity to handle current demands for storing and manipulating activated memory representations. In this guise, working memory plays a role of fundamental importance in major models of higher level cognitive functioning, including in models of language comprehension, general intelligence, memory retrieval, life span development, and frontal lobe functioning (e.g., Baddeley, 1996; Daneman & Carpenter, 1980; Just & Carpenter, 1992; Moscovitch, 1992; Shallice & Burgess, 1993). Working memory's status as a cognitive primitive in these models is supported by at least two decades of research showing that individual and group differences in capacity as measured by working memory span tasks predict individual and group differences on a variety of cognitive tasks.

Cindy Lustig, Department of Psychology, Duke University; Cynthia P. May, Department of Psychology, College of Charleston; Lynn Hasher, Department of Psychology, University of Toronto and The Rotman Research Institute of Baycrest Centre.

The research reported here was supported by National Institute on Aging Grants AG 2753 and AG 4306. Portions of the data presented here were previously reported at the 1998 meeting of the Psychonomics Society, Dallas, Texas. We thank Karen Z. H. Li, Jennifer Green, Stephanie Davis, Leigh Coughenour, Kelly Murray, Aaron Levy, Erin Elliot, Carolyn Yoon, Michelle Lee, and Leticia Carrasco, whose helpfulness on this project is greatly appreciated. We would also like to thank Timothy Salthouse for suggesting the reliability analyses, Malcolm Binns for verifying the power analyses, and Patrick Curran for statistical advice throughout the project.

Correspondence concerning this article should be addressed to Cindy Lustig, who is now at Washington University, Department of Psychology, Campus Box 1125, St. Louis, Missouri 63130-4899, or to Lynn Hasher, Department of Psychology, University of Toronto, 100 St. George Street, Toronto, Ontario, M5S 3GS, Canada. Electronic mail may be sent to cal9@acpub.duke.edu or hasher@psych.utoronto.ca.

However, not all investigators agree that capacity is the source of the span's predictive power (e.g., Baddeley, Logie, Nimmo-Smith, & Brereton, 1985; Martin, 1995; Tirre & Pena, 1992; Waters & Caplan, 1996; Whitney, Arnett, Driver, & Budd, 1999). Most working memory span tasks are quite complex, involving storage and retrieval of information across multiple test trials as well as various operations to process that information, and thus it is difficult to determine exactly what constructs they assess. In addition, an alternative interpretation of span tasks posits that they not only involve the ability to handle current information, but also require the successful management of previous information (e.g., Dempster & Corkill, 1999; May, Hasher, & Kane, 1999).

Consider, for example, the reading span task developed by Daneman and Carpenter (1980). In that task, participants read a series of unrelated sentences for comprehension while trying to remember the last word from each sentence. The sentences are presented in groups, or sets, that typically start with the smallest one (two sentences) and then increase systematically to the largest set (four or five sentences). Participants are given several trials (most often three to five) at each set size, and an individual's working memory span score is determined by the largest set size for which he or she completes the majority of the word recall trials successfully. In this task, participants may recall as many as 15 to 20 different lists of sentence-ending words. Because of the multiple memory tests inherent in the task, whatever the rehearsal buffer and operating demands may be within a single set of sentences (the aspect of processing that working memory span is intended to measure), there are major memory demands across sets (or lists) of items. To do well on multiple-list recall tasks, such as working memory span tasks, individuals must be successful at managing the memory component of the task so as to recall only those words that end the current set of sentences and no words that ended any previous sets of sentences (see May et al., 1999). This should be relatively easy on the first few trials, as set sizes are small (often only two items) and few other sets will have been learned and subsequently recalled before the current trial. Recalling only current-set items becomes increasingly difficult as the task proceeds because the set sizes become progressively larger and many more items will have been presented and recalled before the current set.

Performance on these largest, late-presented sets may be particularly important in ultimately determining span size. This is because individuals who are most successful at recalling these largest, latest sets and hence have the largest spans are likely to be those who can best eliminate words presented on previous trials from consideration as candidates for recall when the most recent set is the target. A classic and alternative way of describing this is to say that individuals with large spans may be those who are best able to handle the proactive interference (PI) that builds up across the series of multiple study and test trials that constitute the span task. Following this logic, working memory span tasks may well measure the ability to reduce the competition or interference from items presented on previous trials, so that on subsequent trials, an individual is retrieving from only the most recently presented set. Working memory span tasks may also measure their intended operation, the capacity to simultaneously store and process currently relevant information, but if so, this ability may be obscured by the presence of interference in the task.

Consistent with this view is substantial evidence that the ability to manage PI contributes significantly to span performance. Much of this evidence comes from the individual and group differences literature, in which several investigators have focused on the suggestion that the ability to resist interference is related to overall span performance (e.g., Chiappe, Hasher, & Siegel, 2000; Conway & Engle, 1994; Dempster, 1981; Dempster & Cooney, 1982; Jensen, 1964; May et al., 1999; Rosen & Engle, 1998; Rosner, 1972; Whitney et al., 1999). Further evidence comes from experimental manipulations of interference-relevant variables within the span task itself. A number of studies show that factors that act to decrease PI within the span task (e.g., reducing similarity among items) also act to boost span scores (e.g., Conrad & Hull, 1964; Shah & Miyake, 1996; Young & Supa, 1941). As well, recent evidence suggests that span measures, like other memory tasks (e.g., Underwood, 1957), can be influenced by previous laboratory experience (Lustig & Hasher, in press).

Thus, working memory span tasks may not be simple indices of the general mental capacity to store and manipulate information from the current trial. Instead, span performance is influenced by (perhaps among other things) the ability to control the buildup of PI across successive test trials. In particular, it is important to suppress or inhibit items from the no-longer-relevant, previous trials, so that they do not interfere with recall of the relevant items from the current trial (May et al., 1999). A major question then is whether span predicts performance on a wide range of high-level cognitive tasks because of its ability to measure *current* item processing, *past* item processing, or some combination of the two. That question is the focus of the present study.

To address the source of span's predictive power, we capitalized on a recently reported finding by May et al. (1999) in which the amount of PI within the span task was experimentally manipulated. Specifically, May et al. modified the administration of the Daneman and Carpenter (1980) reading span task to create versions of the task that, relative to the standard administration, reduce the impact of PI on the largest set sizes. As described above, the standard administration of span presents the largest set sizes last, after numerous other trials, and a substantial amount of PI may build up before participants receive the longest trials. In one alternative administration used by May et al., the sequence of trials was reversed so that the largest set sizes were presented first rather than last, thus reducing the impact of PI on the longest trials. In a second alternative administration, distinctive breaks were added between each span trial to enhance the temporal distinctiveness between trials and to decrease similarity between trials, two factors well known to reduce the extent of PI (e.g., Underwood, 1945, 1957; Wickens & Cammarata, 1986; Wickens & Gittis, 1974). Consistent with the suggestion that PI plays an important role in determining span performance, May et al. found that these PI-reducing manipulations led to increased span scores.

The present study extends these findings by examining how span's ability to predict performance on other cognitive tasks changes depending on whether the span task is administered in the standard, high-PI version versus in a version designed to reduce the impact of PI on working memory span. Specifically, we examined how correlations between span and recall of a short prose passage varied across situations in which PI was relatively high versus low within the span task itself. We chose prose recall as a predicted measure for several reasons. First, beginning with the classic study of Daneman and Carpenter (1980), memory for such passages has been widely used in the literature to examine the relation between span and language comprehension. Daneman and Carpenter asked participants to read a series of short passages and to answer questions about factual, gist, and inferential information after each passage. Of course, this is the precise content required by prose recall (see, e.g., Alba & Hasher, 1983). Following this lead, a number of studies investigating the relationship between span and language have used prose recall as their measure of language comprehension (e.g., Drevenstedt & Bellezza, 1993; Hess & Tate, 1992; Holland & Rabbitt, 1990; Jurden & Reese, 1992; Lee-Sammons & Whitney, 1991; Light & Anderson, 1985; Morrow, Leirer, & Altieri, 1992; Stine, Lachman, & Wingfield, 1993), and prose recall is now widely viewed as an appropriate measure of language comprehension (see, e.g., Daneman & Merikle, 1996). We note that correlations between prose recall tasks and span tasks are as high as correlations between span and many other tasks (see Daneman & Merikle, 1996).

We also chose prose recall because, like other language measures (for example, determining anaphoric reference), it is likely to be affected by PI. Evidence for PI's influence on language tasks accrues from several sources. To begin with, groups that are differentially vulnerable to interference (including learning disabled children, poor readers, and older adults) also show deficits on language recall and comprehension tasks (e.g., Gernsbacher, 1997; Gernsbacher & Faust, 1991; Light & Capps, 1986; Townsend, 1982; Zacks & Hasher, 1988; see Dempster & Corkill, 1999; Kemper, 1992, for reviews). Further, among healthy college students, measures of interference proneness (such as performance on the Brown-Peterson task) predict performance on standard reading comprehension tasks and memory tasks (Dempster, 1985; Dempster & Cooney, 1982). Finally, recall of short passages such as the one used here and throughout the language processing literature is influenced by variables known to affect PI, such as the number of prior ideas or the similarity between sentences within the passage (Kintsch & Young, 1984; Sehulster & Crouse, 1972). Because PI also affects span tasks, it seemed quite possible that a shared influence of PI on both span and prose recall might contribute to the correlation between them.

The general logic of the study was straightforward: If capacity is the basis on which span tasks predict other measures, reducing the amount of PI in the span task should have no impact on the correlation between span and prose recall. (The alternative administrations created by May et al., 1999, do not affect the capacity requirements of the span task; in all versions the same amount of information is to be processed and recalled.) If PI is the basis on which span tasks predict other measures, reducing the amount of PI in the span task should reduce the correlation between span and prose recall. We tested both younger and older adults on the standard and PI-reducing versions of the span task. Doing so allowed us to explore the pattern of correlations across a wider range of span scores than if we had tested only college students, thus increasing the generalizability of our findings. Further, older adults' differential sensitivity to interference (e.g., Hasher & Zacks, 1988; Kane & Hasher, 1995; McDowd & Filion, 1992; Winocur & Moscovitch, 1983) makes them particularly interesting in studies of interference on cognitive tasks. Our findings for both groups were the same. Operations that reduced interference for each group raised their span scores (in replication of May et al., 1999) and reduced the ability of span to predict prose recall.

Method

Participants

One hundred sixty-seven young adults (ages 18–24) and 117 older adults (ages 61–75) participated in this experiment. Young adults were undergraduates at Duke University, the University of Arizona, and the College of Charleston. Older adults were community-dwelling volunteers in Durham, North Carolina, and Tucson, Arizona. All participants were native English speakers. The data from 17 young and 17 older adults were excluded because of health problems (e.g., recent history of stroke, uncorrected vision), participation in prior similar experiments, or failure to follow instructions during the course of the study. The final set of data was based on the performance of 150 young and 100 older participants.

Materials

Reading span. The sentences used in the present study came from the May et al. (1999) study. They selected 45 sentences from those originally created by Daneman and Carpenter (1980) and shortened them to range from 7 to 13 words in length. For example, the original sentence, "It was shortly after this that an unusual pressure of business called me into town" was changed to, "Shortly after that, an unusual pressure of business called me into town." These 45 sentences were then randomly assigned into five sets of two, five sets of three, and five sets of four sentences. Within a given set size, sentences were rotated to create five orders of presentation that were then counterbalanced across the 50 participants within each of the presentation formats given to each age group.

Two different formats of presentation were used to assess working memory span for both young and older adults: (a) the "standard" format, in which the five trials for set size two were presented first, followed by the five trials for set size three and the five for set size four; and (b) the "descending" format, in which the five trials for set size four were presented first, followed by the five trials for set size three and the five for set size two. A third format, the "descending-with-breaks" condition, was used only with younger adult participants. In this condition, the descending sequence was followed but with distinctive, nonverbal tasks included after the recall of each successive set of sentences. Two familiar and distinctive sentences (i.e., "Jack and Jill went up the hill" and "Mary had a little lamb") served as examples given to all participants as part of the instructions.

Filler tasks. The break tasks consisted of 14 unique nonverbal tasks (for example, Hidden Patterns and Card Rotations) from the Kit of Factor-Referenced Tests (Educational Testing Service, 1976). The nonverbal filler used between reading and recall of the story was the Digit Symbol subtest of the Wechsler Adult Intelligence Scale—Revised (WAIS–R; Wechsler, 1981).

Story. The story used in this experiment was a 140-word paragraph with a Flesch–Kincaid grade level of 7.8 years. This story was selected to be similar in length and reading ease to those used by Daneman and Carpenter (1980). The story was divided into 24 idea units for scoring purposes, following a widely used procedure in prose memory (see, e.g., Bransford & Johnson, 1972; see Appendix for idea units).

Design and Procedure

Participants completed a working memory span task and then read a story, did a nonverbal filler task, and, finally, recalled the story. Older adults were randomly assigned to either the standard or descending span conditions ($n = 50$ in each condition); young adults were randomly assigned to the standard, descending, or descending-with-breaks span conditions ($n = 50$ in each condition) of May et al. (1999). Participants in the descending-with-breaks span condition received instructions for the break tasks prior to the span instructions and received a different break task for 45 s between each span trial.

The instructions for the reading span task were the same as those used by May et al. (1999) and closely followed those described by Daneman and Carpenter (1980). Participants were instructed to read and understand the sentence on each index card and try to remember the last word. As soon as the participant finished reading a card the next one was presented for reading and comprehension until a blank card appeared, indicating that participants should attempt to recall the last word for each of the sentences in that set. The words could be recalled in any order, with the exception that the last word of the set was not to be recalled first. Participants in the descending-with-breaks condition worked on one of the break tasks for 45 s before moving onto the next trial; all other participants began the next trial immediately after recall.

After the span task, participants were given the story and asked to read it aloud at a pace that would permit them to understand it and, subsequently, to answer some questions. They next received instructions for the nonverbal filler task (which was different from any of the tasks used as "breaks") and worked on this task for 90 s.

Following the filler task, participants wrote down as much of the story as they could remember. They were asked to use the original words and phrasing of the story when possible, but were told that it was more important to recall the main ideas and events of the story. Participants were given up to 5 min to recall the story and no one required more time.

When finished with the experimental tasks, participants completed Version 3 of the Extended Range Vocabulary Test (ERVT; Educational Testing Service, 1976) and other demographic and health forms. All older adults were tested in the morning (from 9 to 11 a.m.) and all young adults were tested in the afternoon (from 12 to 5 p.m.), following evidence suggesting that these times of testing are likely to maximize the performance of each age group (e.g., Intons-Peterson, Rocchi, West, McLellan, & Hackney, 1998; May & Hasher, 1998).

Results

Participants

The mean age, years of education, and ERVT (vocabulary) score for participants in each of the five conditions can be found in Table 1. The data for years of education and ERVT score were analyzed using a one-way, five-factor analysis of variance

Table 1
Means and Standard Deviations for Age, Years of Education, and Vocabulary Scores for Each Group of Participants

	Older adults		Younger adults		
Demographics	Ascending	Descending	Ascending	Descending	Descending-breaks
Age					
M	68.9	68.6	19.1	19.4	19.3
SD	3.8	4.5	1.2	1.5	1.3
Education					
M	16.4	16.4	13.0	13.1	12.9
SD	2.2	2.1	1.1	1.3	1.1
Vocabulary					
M	34.8	37.2	21.5	20.2	19.4
SD	7.7	7.4	7.9	8.6	7.1

(ANOVA) followed by Tukey post hoc tests with an alpha level of .05. Older adults had significantly more years of education and higher ERVT scores than did young adults. Within each age group, there were no differences in years of education (minimum significant difference = 0.91) or ERVT score (minimum significant differences = 4.26) as a function of span format.

Reading Span

The reading span data were scored in two ways. The "traditional" method was developed by Daneman and Carpenter (1980) and measures a participant's span as the largest set size at which all words for the majority of trials, here three out of five, are correctly recalled. Partial credit (of .5) was given if all of the items for two (out of five) trials at that set size were correctly recalled. Scoring began with set size two for all participants, regardless of span format. Criteria (at least three out of five trials correct) for set size two had to be obtained before continuing to set size three, and criteria for set size three had to be met before continuing to set size four. Because this measure is somewhat arbitrary in the criteria it uses (e.g., for assigning partial credit) and, in particular, because it limits the range of scores and thus restricts the range for correlations (e.g., Daneman & Green, 1986), we also used a second, more sensitive, method. This "weighted items" method (see, e.g., Chiappe et al., 2000; May et al., 1999; Rosen & Engle, 1998) calculates span as the total number of words correctly recalled in fully correct trials (for example, three trials correct at set size two and one trial correct at set size three would give a span score of 9). We note that the two measures correlate highly ($r = .89$, $p < .001$), as has been reported elsewhere (see, e.g., Whitney et al., 1999). The mean scores are presented in Table 2.

The span manipulations used in the present study were designed to vary the extent of PI in the task. Because these manipulations were used by May et al. (1999), we expected to replicate their key findings, as follows: (a) Young adults would show their usual advantage compared with older adults when the standard span conditions were contrasted; (b) the interference-reducing, descending format would increase the span scores of older adults relative to the standard condition; and (c) the descending-with-breaks condition would increase the span scores of young adults over span scores in the other two conditions, which would not differ from each other, given young adults' relative ability to handle PI.

To test these specific predictions, the data were analyzed using planned contrasts done on both scoring methods.[1] Because the outcomes were similar across the two methods, we report only the data for the weighted item span score. As expected, and consistent with a large literature (e.g., Gick, Craik, & Morris, 1988; Salthouse & Babcock, 1991; Zacks & Hasher, 1988), young adults had higher span scores than older adults when span was administered in the standard, ascending condition, $t(98) = 4.00$, $p = .0001$. Older adults tested with the PI-reducing descending format had higher span scores than did older adults tested with the standard format, $t(98) = 2.27$, $p = .02$. Although the span scores for young and older adults differed when the standard format was used, this difference was no longer reliable when young and older adults were tested with the descending format, $t(98) = 1.29$, $p = .20$.[2] Finally, young adults tested in the descending-with-breaks condition had higher span scores than young adults tested in the other two conditions, $t(148) = 3.68$, $p = .0003$, whereas performance in ascending and descending conditions did not differ, $t(98) < 1$.

Thus, reducing the PI component of the span task by administering it so that the longest and most difficult trials occurred before previous items could disrupt recall (before PI could build up) improved the span scores of older adults but not young adults. Further reducing the impact of PI by introducing unique break tasks after each trial raised the scores of young adults.

The results of two subsidiary analyses support our claim that the experimental manipulations increased span performance for both young and older adults by reducing PI. First, we examined the rate of intrusions (words that were not the last words of sentences in the current set) as a function of item span for each group. (See Table 3.) The rate of intrusions was generally low (as is common

[1] A third set of analyses using the total number of words recalled as the dependent measure led to identical conclusions.

[2] We note that the effect size for the age difference in the ascending condition was of relatively large size and consistent across this study and May et al. (1999). The effect size for the age difference in the ascending condition was relatively large according to Cohen (1988), $d = .77$ in the current study, and $d = .72$ in May et al. The effect size for the age difference in the descending condition was small, $d = .29$, for both studies. Further, this difference changed direction from one study to the next, favoring older adults in May et al. and young adults in the current study.

Table 2
Means and Standard Deviations of Standard Span, Item Span, and Prose Recall Scores and Correlation Between Span and Prose Recall for Each Group of Participants

Span and prose recall measures	Older adults: Ascending	Older adults: Descending	Younger adults: Ascending	Younger adults: Descending	Descending-breaks
Standard score					
M	2.81	3.05	3.17	3.12	3.53
SD	0.60	0.53	0.67	0.59	0.61
Item score					
M	20.10	23.56	26.20	25.52	30.72
SD	8.07	6.05	8.86	7.21	7.61
Prose recall					
M	12.43	12.88	14.69	14.72	14.93
SD	3.26	3.56	4.32	3.65	3.94
Correlation	.29	.08	.27	.33	.02

in the PI literature; see, e.g., Dempster, 1985; Postman & Keppel, 1977). However, older adults tested in the ascending condition (whose span performance is most likely to be influenced by PI) had a higher rate of intrusions than any other group. Using the same planned contrasts design to analyze these data as for the span data, we found that older adults tested in the ascending condition made more intrusions than older adults tested in the descending condition, $t(98) = 2.96, p = .003$. The number of intrusions made by older and younger adults did not differ for participants tested in the descending condition ($t < 1$). The intrusion rates for young adults were in general very low, and although the pattern of means is consistent with our predictions (young adults in the descending-with-breaks condition have a slightly lower intrusion rate than young adults in the two no-breaks conditions), this difference is not reliable, $t(148) = 1.02$.

Second, we examined the number of correctly completed trials at set size four, once again using the planned contrasts design. (See Table 3.) If, as we claim, our experimental manipulations boosted span performance by reducing the impact of PI on the longest trials, participants tested in the PI-reducing conditions (descending for older adults, descending-with-breaks for young adults) should obtain higher scores on this set size relative to their counterparts in the other conditions. In general, all our participants found these large set size trials to be very difficult; even in the young descending-with-breaks group, only three participants successfully completed the recall of all five trials at this set size.

The majority of older adults tested in the ascending condition were unable to recall *any* trials at set size four, but the majority of older adults tested in the descending condition were able to successfully produce at least one trial at this set size. Older adults in the descending condition tended to recall more trials at set size four than did older adults tested in the ascending condition, $t(98) = 1.60, p = .11$. In addition, the age difference in set size four recall was much larger in the standard, ascending condition than in the PI-reducing descending condition. Young adults had much higher scores than older adults when tested in the ascending condition, $t(98) = 3.87, p = .00001$. In contrast, the age difference in the descending condition was marginal, $t(98) = 1.51, p = .13$. Finally, young adults tested in the descending-with-breaks condition scored higher at set size four than did young adults tested in the two no-breaks conditions, $t(148) = 4.51, p = .0001$. Thus, these data are consistent with the conclusion that PI plays a substantial role in determining performance.

In summary, reversing the format of span administration so that the longest trials are presented first, a manipulation intended to reduce competition from previous items, improves the performance of older adults relative to their performance on the standard, ascending version of the task. Because young adults are relatively skilled at reducing or able to reduce PI (relative to older adults), a more extreme manipulation (the addition of distinctive breaks) was required to boost their performance. The same manipulations that boost overall span performance (descending for older adults,

Table 3
Means and Standard Deviations of Intrusion Rates and Number of Sets Recalled at Set Size Four for Each Group of Participants

Additional span interference measures	Older adults: Ascending	Older adults: Descending	Younger adults: Ascending	Younger adults: Descending	Descending-breaks
Intrusion rates					
M	.17	.09	.10	.09	.07
SD	.16	.13	.11	.10	.10
Set size four recall					
M	0.76	1.14	1.68	1.50	2.52
SD	1.04	1.03	1.35	1.18	1.31

descending-with-breaks for young adults) also reduce intrusion rates and improve performance on the longest trials, supporting the idea that the beneficial effects of these manipulations were indeed the result of a reduction in PI.

Prose Recall

Two raters scored each participant's recall of the prose passage for gist recall of the idea units. Following training on scoring, the interrater reliability, calculated on the number of idea units for each participant, was extremely high ($r = .97$). Discrepancies of 1 point were resolved by taking the average of the two scores; discrepancies greater than 1 point were resolved by discussion. As is commonly found in the prose memory literature, older adults had lower story recall scores ($M = 12.7$) than did young adults ($M = 14.8$), $F(1, 248) = 19.34$, $p = .001$, but within each age group story recall score did not differ as a function of span condition ($F < 1$).

The critical question for this investigation was whether reducing the PI component in the span task would reduce its ability to predict recall of the story. To answer this question, we calculated the correlations between the item span score and the story recall score for each group. On the basis of a considerable literature, our expectation was that the standard format span task would correlate with recall. As can be seen from Table 2, this correlation for both young and older adults was close to .30. Correlations in this range are widely reported (see Daneman & Merikle, 1996).[3] The descending span task also showed a reliable correlation for young adults.

As can be seen, the same manipulations that increased span scores decreased the correlation between span and story recall. For older adults, only the standard ascending span predicted story recall; when PI was reduced by using the descending administration, span performance no longer predicted story recall performance. For young adults, reducing the PI component of the task by using the descending-with-breaks format both raised span scores and eliminated the relation between span and story recall.[4]

We note that the absence of reliable correlations between the PI-reducing manipulations and recall cannot be attributed to measurement problems. First, within each age group, the range of span scores on the critical, PI-reducing conditions is as great as it is for the standard conditions. Second, the range of prose recall scores is consistent across all the span conditions. Further, the split half reliability (Spearman–Brown) scores for all the span measures are in a range that is considered good to strong (.6–.8) and so do not limit the ability to see correlations (or their absence) between span and story recall.[5]

Finally, as a check on the validity of the correlations, we compared the story recall performance of the 12 participants with the lowest and highest item span scores within each group. (See Table 4.) Separate ANOVAs for each age group's story recall scores followed by contrasts of the story recall scores of the low- versus high-span subgroups were consistent with the correlation results. The story recall performance of low-span participants was substantially lower than that of high-span participants for older adults tested using the ascending, high-PI version of the span task, $t(22) = 2.03$, $p = .049$, but not for older adults tested using the descending, reduced-PI version of the span task ($t < 1$). Similarly, the story recall performance of low- and high-span young adults differed for young adults tested in the two no-breaks conditions, $t(46) = 2.59$, $p = .01$, but not for young adults tested in the descending-with-breaks condition ($t < 1$). The results from this subsidiary analysis are consistent with the correlation results: In the PI-reducing conditions (descending for older adults, descending-with-breaks for young adults), span fails to predict prose recall.

Table 4
Means and Standard Deviations of Prose Recall Scores of Low- and High-Span Quartile Participants in Each Group

Span quartile	Older adults		Younger adults		Descending-breaks
	Ascending	Descending	Ascending	Descending	
Low-span					
M	10.63	13.08	12.96	13.29	13.83
SD	3.08	3.43	4.62	3.52	3.92
High-span					
M	13.25	13.17	15.38	16.83	14.67
SD	2.69	3.42	3.81	3.41	4.64

Discussion

The normative view in cognition is that working memory span tasks measure the capacity available for individuals to maintain and operate on current information. However, as many have noted, the tasks that assess working memory are actually quite complex

[3] We have been asked why groups that differ in interference proneness (young and older adults) would show the same correlation between span (on which mean performance will differ) and story recall (on which mean performance will also differ). In response, we would say that unless the relationship between span and prose recall performance is curvilinear and steeper at one end of the function than at another, similar correlations would be expected. The correlations should also be similar, if flat, if there is one factor (say interference) that is causally related to both measures and that can be eliminated in one task by experimental manipulation.

[4] A series of effect size and power analyses are consistent with these conclusions, although a Fisher's z test for differences between correlation coefficients yields only marginal results ($z = 1.07$ for old ascending vs. old descending, $z = 1.48$ for young descending-with-breaks vs. the two no-breaks conditions). With respect to the effect size analyses, we first note that the effect sizes for the correlations are larger in the standard conditions than in the PI-reducing conditions. The correlations between span and prose recall in the standard conditions are of intermediate size ($r = .30$; Cohen, 1988) and consistent with the larger literature. In contrast, the correlations between span and prose recall in the two PI-reducing conditions fall slightly below even the recommended standard for small effect sizes ($r = .10$). Further, the effect sizes for the differences between the correlations ($q = .22$ for older adults, $q = .32$ for younger adults) are also in the intermediate range ($q = .30$). However, because our sample size was rather small for an individual differences study, we had low power to detect these differences statistically using the Fisher's z test.

[5] We thank Timothy Salthouse for these suggestions. We note that although young adults tested in the descending condition have a numerically lower reliability than other groups (.6), they have the highest correlation between span and prose recall of any group. It is therefore unlikely that the low reliabilities in the PI-reducing conditions are the cause of the low correlations between span and prose recall.

and performance on them is multiply determined. In particular, May et al. (1999) observed that most assessments of working memory span have multiple study and test trials. As a result, working memory span tasks also require that participants manage the past such that items from antecedent test trials do not interfere at the time of recall with items from the current set of to-be-recalled materials. If the past is not well managed, presumably by being suppressed when no longer relevant (Hasher, Zacks, & May, 1999; Zacks & Hasher, 1994), then the consideration set at the time of recall will consist of currently relevant items along with others that were previously relevant, setting the stage for poorer recall than if only the currently relevant items were active and thereby lowering span scores (e.g., Anderson, 1974, 1976; Watkins & Watkins, 1975).

Working memory span scores are inherently vulnerable to the negative effects of prior learning because the effects are known to build up rapidly across successive trials within a task. As a result, trials farther from the beginning of a task will be more impacted by interference than trials closer to the beginning (e.g., Keppel & Underwood, 1962; Wickens, 1970). As it happens, most span tasks have their longest trials late in the series, and it is success on these late trials that is important for obtaining a high span score. As a result, the assessment of span is, in many instances, confounded by the buildup of PI inherent in successive tests of memory.

In replication of May et al. (1999), we confirmed that working memory span scores can be affected by manipulations that reduce the impact of interference on these longest trials. Reducing the impact of PI on these longest trials by presenting them first in the sequence, rather than last, improves the span scores of older adults. Further reducing the impact of PI by also adding brief nonverbal filler tasks between each set of sentences (thus increasing the distinctiveness of each trial) improves the span scores of young adults. Because the amount of information to be stored and processed remains the same across the different experimental manipulations, these findings, like those of May et al., are not easily explained by a capacity view of span. As well, they confirm that interference can indeed confound measures of working memory span. What they do not tell us is whether this confound plays a role in the predictive power that measures of working memory span have shown (see Daneman & Merikle, 1996).

To address this question, we compared the correlations between working memory span and prose recall when span was measured with more (the standard task) versus less (the experimental manipulations) interference operating on the largest set sizes. Our findings were generally straightforward. For both older and young adults, the correlations were reliable (and around .30) when the standard administration of span was given, and for both groups the correlations fell short of reliability (and were less than .10) when interference in the span task was reduced. The correlation pattern was confirmed by examining the prose recall performance of the people with the highest and lowest span scores within each group. For people tested in the standard condition (and for whom span predicted prose recall), subgroups with high span scores had higher prose recall scores than did subgroups with low span scores. For people tested in the PI-reducing conditions, subgroups with high and low span scores had equivalent prose recall scores. Thus, the confound in the span measurement appears to carry over into the predictive realm, suggesting that the ability to control the buildup of interference in standard administrations of span tasks is at least part of the reason individual differences in span predict individual differences in prose recall.

This finding is not actually so surprising because story recall, like many language-based tasks that span predicts, has a memory component inherent in it, and all memory-based tasks, whether dealing with arbitrary nonsense syllables, pairs of words, lists of words, stories, life events, or the linkages that need to be established for language comprehension to occur are to some degree vulnerable to the same interference-based memory effects (e.g., Alba & Hasher, 1983; see Dempster & Corkill, 1999, for a discussion of the role of interference in on-line reading comprehension). Thus, the present data can be seen as powerful in their implications for a mechanism that underlies group and individual differences in the ability to do a wide range of tasks. However, the present data are also merely suggestive. Of course, we used just one set of materials to assess span and just one task to look at individual differences. Our sample sizes, although large for an experimental study, were more moderate in scope for a correlational one. We unfortunately did not control the prior laboratory experience of our participants (i.e., the number of previous studies people had participated in), and we have recently learned that prior experience has a negative effect on at least one measure of working memory span (see Lustig & Hasher, in press). Such a failure of control may or may not account for a puzzling aspect of the present findings. Specifically, although interference susceptibility can account for age differences in span (a large literature suggests this; see Kane & Hasher, 1995) and for individual differences in the predictive utility of span, it does not fare so well in accounting for both simultaneously. That is because there are three groups in the present study with similar spans, old ascending, young ascending, and young descending, with different correlational patterns. We have not been able to generate a single factor account of both individual and group differences, and that raises the possibility that another, as yet unspecified factor is also operating. We note, however, that general capacity views of span cannot account for this pattern either (and, further, they cannot account for the changes in span scores and in patterns of correlations when interference is high versus low within an age group).

To return to the main focus of the study, we note that capacity accounts of working memory hold that span tasks measure the resources available for storing and manipulating activated memory representations and that this capacity is a critical determinant of performance in a range of cognitive domains. With respect to capacity, it is important to note that the different methods used here to administer the span task were equivalent in their demand for storage and manipulation both within each trial and summed across all trials. Thus, if span primarily measures capacity, neither span scores nor the ability of those scores to predict another task should have varied as a function of the way in which the span task itself was administered. Because both measures changed, capacity, as currently construed in the literature, cannot be the sole determinant of performance in span tasks or in span tasks' ability to predict other tasks (see also Martin, 1995; Wingfield, Waters, & Tun, 1998). (For comparisons of the inhibitory view of working memory function with views that emphasize constructs such as general executive control or self-initiated processing, please see Hasher, Tonev, Lustig, and Zacks, in press; Lustig, Hasher, & Tonev, in press.)

The present data clearly implicate processes that underlie interference as being critical determinants of span scores and of their predictive value. Elsewhere Hasher et al. (1999; Zacks & Hasher, 1994) have speculated that the ability to overcome PI depends largely on the efficiency with which one can inhibit the no-longer-relevant past as new information is presented. The benefit of efficient inhibition of the past is seen whenever a current task requires the retrieval of information from the most recent time period. This is because the set of items under consideration for retrieval will be narrowly focused when inhibition is effective, and will be broader, including more of the irrelevant past, when it is not. Individuals who differ in inhibitory efficiency should show differences on tasks that require rapid and accurate retrieval. As well, circumstances that alter inhibitory efficiency should also impact on tasks requiring accurate retrieval. Span and prose memory are two such tasks. Thus, an important mechanism underlying the effects seen here and perhaps others in the working memory span literature may well be inhibitory in nature.

References

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin, 93,* 203–231.

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology, 6,* 451–474.

Anderson, J. R. (1976). *Language, memory, and thought.* Hillsdale, NJ: Erlbaum.

Baddeley, A. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49A,* 5–28.

Baddeley, A., Logie, R., Nimmo-Smith, I., & Brereton, N. (1985). Components of fluent reading. *Journal of Memory and Language, 24,* 119–131.

Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11,* 717–726.

Chiappe, P., Hasher, L., & Siegel, L. S. (2000). Working memory, inhibitory control, and reading disability. *Memory & Cognition, 28,* 8–17.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion, and memory span. *British Journal of Psychology, 55,* 429–432.

Conway, A. R. A., & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General, 123,* 354–373.

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

Daneman, M., & Green, I. (1986). Individual differences in comprehending and producing words in context. *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 and Review, 3,* 422–433.

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin, 89,* 63–100.

Dempster, F. N. (1985). Proactive interference in sentence recall: Topic-similarity effects and individual differences. *Memory & Cognition, 13,* 81–89.

Dempster, F. N., & Cooney, J. B. (1982). Individual differences in digit span, susceptibility to proactive interference, and aptitude/achievement test scores. *Intelligence, 6,* 399–416.

Dempster, F. N., & Corkill, A. J. (1999). Interference and inhibition in cognition and behavior: Unifying themes for educational psychology. *Educational Psychology Review, 11,* 1–88.

Drevenstedt, J., & Bellezza, F. (1993). Memory for self-generated narration in the elderly. *Psychology and Aging, 8,* 187–196.

Educational Testing Service. (1976). *Kit of factor-referenced tests.* Princeton, NJ: Author.

Gernsbacher, M. A. (1997). Group differences in suppression skill. *Aging, Neuropsychology, and Cognition, 4,* 175–184.

Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 245–262.

Gick, M. L., Craik, F. I. M., & Morris, R. G. (1988). Task complexity and age differences in working memory. *Memory & Cognition, 16,* 353–361.

Hasher, L., Tonev, S. T., Lustig, C., & Zacks, R. (in press). Inhibitory control, environmental support, and self-initiated processing in aging. In M. Naveh-Benjamin, M. Moscovitch, & R. L. Roediger III (Eds.), *Perspectives on human memory and cognitive aging: Essays in honour of Fergus Craik.* Philadelphia: Psychology Press.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193–225). San Diego, CA: Academic Press.

Hasher, L., Zacks, R., & May, C. P. (1999). Inhibitory control, circadian arousal, and age. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII* (pp. 653–675). Cambridge, MA: MIT Press.

Hess, T. M., & Tate, C. S. (1992). Direct and indirect assessments of memory for script-based narratives in young and older adults. *Cognitive Development, 7,* 467–484.

Holland, C. A., & Rabbitt, P. M. A. (1990). Autobiographical and text recall in the elderly: An investigation of a processing resource deficit. *Quarterly Journal of Experimental Psychology, 42A,* 441–470.

Intons-Peterson, M. J., Rocchi, P., West, T., McLellan, K., & Hackney, A. (1998). Aging, optimal testing times, and negative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 362–376.

Jensen, A. R. (1964). *Individual differences in learning: Interference factor* (Final Report, Cooperative Research Project No. 1897). Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.

Jurden, F. H., & Reese, H. W. (1992). Educational context differences in prose recall in adulthood. *Journal of Genetic Psychology, 153,* 275–291.

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

Kane, M. J., & Hasher, L. (1995). Interference. In G. Maddox (Ed.), *Encyclopedia of aging* (2nd ed., pp. 514–516). New York: Springer-Verlag.

Kemper, S. (1992). Language and aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 213–270). Hillsdale, NJ: Erlbaum.

Keppel, G., & Underwood, B. J. (1962). Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior, 1,* 153–161.

Kintsch, W., & Young, S. R. (1984). Selective recall of decision-relevant information from texts. *Memory & Cognition, 12,* 112–117.

Lee-Sammons, W. H., & Whitney, P. (1991). Reading perspectives and memory for text: An individual differences analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 1074–1081.

Light, L. L., & Anderson, P. A. (1985). Working-memory capacity, age, and memory for discourse. *Journal of Gerontology, 40,* 737–747.

Light, L. L., & Capps, J. L. (1986). Comprehension of pronouns in young and older adults. *Developmental Psychology, 22,* 580–585.

Lustig, C., & Hasher, L. (in press). Working memory span: The effect of prior learning. *American Journal of Psychology.*

Lustig, C., Hasher, L., & Tonev, S. T. (in press). Inhibitory control over the present and the past. *European Journal of Cognitive Psychology.*

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

May, C. P., & Hasher, L. (1998). Synchrony effects in inhibitory control over thought and action. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 363–379.

May, C. P., Hasher, L., & Kane, M. J. (1999). The role of interference in memory span. *Memory & Cognition, 27,* 759–767.

McDowd, J. M., & Filion, D. L. (1992). Aging, selective attention, and inhibitory processes: A psychophysiological approach. *Psychology and Aging, 7,* 65–71.

Morrow, D. G., Leirer, V. O., & Altieri, P. A. (1992). Aging, expertise, and narrative processing. *Psychology and Aging, 7,* 376–388.

Moscovitch, M. (1992). Memory and working-with-memory: A component process model based on modules and central systems. *Journal of Cognitive Neuroscience, 4,* 257–267.

Postman, L., & Keppel, G. (1977). Conditions of cumulative proactive inhibition. *Journal of Experimental Psychology: General, 106,* 376–403.

Rosen, V., & Engle, R. W. (1998). Working memory capacity and suppression. *Journal of Memory and Language, 39,* 418–436.

Rosner, S. R. (1972). Primacy in preschooler's short-term memory: The effects of repeated tests and shift-trials. *Journal of Experimental Child Psychology, 13,* 220–230.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27,* 763–776.

Sehulster, J. R., & Crouse, J. H. (1972). Storage and retrieval of prose material. *Psychological Reports, 30,* 435–439.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125,* 4–27.

Shallice, T., & Burgess, P. (1993). Supervisory control of action and thought selection. In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness, and control. A tribute to Donald Broadbent* (pp. 171–187). Oxford, England: Clarendon.

Stine, E. A. L., Lachman, M. E., & Wingfield, A. (1993). The roles of perceived and actual control in memory for spoken language. *Educational Gerontology, 19,* 331–349.

Tirre, W. C., & Pena, C. M. (1992). Investigation of functional working memory in the reading span test. *Journal of Educational Psychology, 84,* 462–472.

Townsend, M. A. R. (1982). Intrusions of prior knowledge in prose recall as a function of age and exhortations to remember. *Psychological Reports, 51,* 291–294.

Underwood, B. J. (1945). The effect of successive interpolations on retroactive and proactive inhibition. *Psychological Monographs, 59*(No. 3).

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review, 64,* 49–60.

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

Watkins, O. C., & Watkins, M. C. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 442–452.

Wechsler, D. (1981). *The Wechsler Adult Intelligence Scale—Revised.* New York: Psychological Corporation.

Whitney, P., Arnett, P. A., Driver, A., & Budd, D. (1999). *Measuring central executive functioning: What's in a reading span?* Manuscript submitted for publication.

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review, 77,* 1–15.

Wickens, D. D., & Cammarata, S. A. (1986). Response class interference in STM. *Bulletin of the Psychonomic Society, 24,* 266–268.

Wickens, D. D., & Gittis, M. M. (1974). The temporal course of recovery from interference and degree of learning in the Brown-Peterson paradigm. *Journal of Experimental Psychology, 102,* 1021–1026.

Wingfield, A., Waters, G. S., & Tun, P. A. (1998). Does working memory work in language comprehension? Evidence from behavioral neuroscience. In N. Raz (Ed.), *The other side of the error term: Aging and development as model systems in cognitive neuroscience* (pp. 319–343). Amsterdam: Elsevier.

Winocur, G., & Moscovitch, M. (1983). Paired-associate learning in institutionalized and noninstitutionalized old people: An analysis of interference and context effects. *Journal of Gerontology, 38,* 455–464.

Young, C. W., & Supa, M. (1941). Mnemic inhibition as a factor in the limitation of the memory span. *American Journal of Psychology, 54,* 546–552.

Zacks, R., & Hasher, L. (1988). Capacity theory and the processing of inferences. In L. L. Light & D. M. Burke (Eds.), *Language, memory, and aging* (pp. 155–169). New York: Cambridge University Press.

Zacks, R., & Hasher, L. (1994). Directed ignoring: Inhibitory regulation of working memory. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory mechanisms in attention, memory, and language* (pp. 241–264). San Diego, CA: Academic Press.

Appendix

Prose Passage (Separated Into Idea Units)

Sharon Brown's car / had fallen apart / to the point of being a total write-off. / She needed to get around town, / so she applied to get a monthly discount bus pass. / She didn't relish the idea / of riding the bus, / but the passes were not expensive, they fit well into her budget. / The transit service also allowed her to do all of the activities / she needed her car for, / such as shopping, / visiting friends, and / going to the zoo. / The zoo trips / were especially nice / since she liked to go there at least once a month. / With the bus / she would not have to worry about paying / the five dollar / parking fee / which she had always perceived as unfair. / The more she thought about it, the more that Sharon realized / that the idea of taking the bus / was a good one.

Received May 17, 1999
Revision received December 16, 1999
Accepted January 10, 2000 ■