Aging and Verbal Memory Span: A Meta-Analysis

Kara L. Bopp and Paul Verhaeghen

Department of Psychology and Center for Health and Behavior, Syracuse University, Syracuse, New York.

Using Brinley plots, this meta-analysis provides a quantitative examination of age differences in eight verbal span tasks. The main conclusions are these: (a) there are age differences in all verbal span tasks; (b) the data support the conclusion that working memory span is more age sensitive than short-term memory span; and (c) there is a linear relationship between span of younger adults and span of older adults. A linear model indicates the presence of three distinct functions, in increasing order of size of age effects: simple storage span; backward digit span; and working memory span.

In this article, we provide a meta-analysis of the effects of adult age on verbal short-term memory and verbal working memory span tasks, using graphical analysis on the original span metric.

Memory Span

Span tasks are a popular measure in cognitive psychology and neuropsychological testing. Two types of tasks can be distinguished: short-term memory span and working memory span.

Measures of short-term memory include forward digit span (Wechsler, 1955; 1981), letter span (Kinsbourne, 1974; Taub, 1975), and word span (Baddeley, 1986; Talland, 1965). In each of these a series of stimuli (digits, letters, or words, respectively) is presented, visually or auditorily, typically at the rate of one stimulus per second. The task of the participant is to repeat the stimuli in the order they were presented. Testing usually begins with a series length of two stimuli and increases by one after two trials. Testing ends when a participant is incorrect on two trials of the same length. Span can be calculated in a number of ways, including the total number of words recalled, proportion of words per set, number of words in correct trials, or as the longest series correctly recalled. The current study includes both the final scoring method, because this is the method most frequently used overall, and the Wechsler Adult Intelligence Scale and the Revised Wechsler Adult Intelligence Scale (WAIS–WAIS-R) method for scoring, which is frequently used overall, and the Wechsler Adult Intelligence Scale (WAIS) method for scoring, which is frequently used in digit span measures. (Note that, in the Results section, we explicitly test for differences between these methods to ascertain that method effects do not influence the results.)

Beginning in the 1980s, working memory span measures were introduced into the literature. Working memory refers to the dynamic relationship between passive storage and active manipulation or transformation of information held in memory (Baddeley & Hitch, 1974; Kane, Bleckley, Conway, & Engle, 2001; Miyake & Shaw, 1999; Miyake et al., 2000; Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000). An important feature of the model is the system’s limited capacity, which forces the resources to be shared between storage and processing. Working memory span measures such as reading span (Daneman & Carpenter, 1980), listening span (Bubacco & Salthouse, 1990; Salthouse & Bubacco, 1991), sentence span (an average of reading and listening span; Morrow, Menard, Stine-Morrow, Teller, & Bryant, 2001), and computation span (Bubacco & Salthouse; Salthouse & Bubacco) capture this interaction between concurrent storage and online processing. Reading span requires participants to read sets of sentences while remembering the last word of each sentence. Participants are sometimes asked to answer a question about the sentences in true–false or multiple-choice format. Listening span tasks require participants to listen to sentences and answer a simple question about the sentence, while remembering the last word of each sentence. Scores on listening span and reading span are sometimes averaged into a measure of sentence span. In computation span, a series of arithmetic problems is presented to participants to solve while the participants must also remember the last digit from each problem. Typically, in these working memory span tasks, three sets of trials are given for each span length. As with the short-term memory span tasks, the current study only examines span as measured by the longest series that is correctly recalled.

In cognitive aging research, a further distinction is often made within short-term span tasks (Craik, 1986; Greigore & Van der Linden, 1997; Myerson, Emery, White, & Hale, 2003). Mere storage tasks, such as forward digit span, are distinguished from tasks requiring short-term memory reordering, such as backward digit span and alphabet span tasks. In backward digit span, a series of digits is presented auditorily, and the participant’s task is to repeat the digits in reverse order (Wechsler, 1955, 1981). Reordering span tasks differ from short-term memory tasks proper in that the item manipulation presents a task requirement that goes beyond mere storage. Although both working memory span tasks and reordering span tasks require processing and storage, the type of processing differs. It is likely that reordering tasks differ from working memory span tasks in that item processing occurs after all items have been stored given that the presentation time is too fast for it to occur online. In the current study we examine only backward digit span because of the lack of studies utilizing other reordering span tasks. Note that the tradition in the psychometric literature (most notably Wechsler, 1955) is to add scores for forward and backward digit spans to obtain a single span for short-term memory. We used only studies that provided separate scores in the current analysis.

Memory Span and Aging: Developmental Theory

Age differences in memory span measures are expected by most extant theories of cognitive aging. In fact, the concepts of age-related deficits in short-term memory or working memory, more notably in attentional–executive working memory control (Baddeley, 1986, Engle, Kane, & Tuholski, 1999;
Kane et al., 2001; Miyake et al., 2000), have been invoked as a mechanism to explain age-related declines in a wide variety of tasks of fluid cognition, either as a direct causal mechanism (e.g., Hasher & Zacks, 1988; Mayr & Kliegl, 1993) or as a mediator between a rather general decrease in processing resources and higher-order cognition (e.g., Salthouse, 1996).

Many aging theories also predict larger age-related effects for working memory span than for short-term memory span. First, processing resource theories (e.g., Belleville, Rouleau, & Caza, 1998; Craik, Morris, & Gick, 1990; Dobbs & Rule, 1989; Foos, 1989; Light, Zelinski, & Moore, 1982; Salthouse, Babcock, & Shaw, 1991) state that aging depletes the cognitive resources available for processing. The concomitant expectation is that working memory tasks, with their demands on both processing and storage, will yield larger age effects than short-term memory tasks, which require mere storage.

Second, theories focusing on coordinative task requirements (e.g., Mayr & Kliegl, 1993), predict larger age differences in working memory span than in short-term memory span, because working memory tasks require the coordination of concurrent storage and processing demands, which is a demand absent in short-term memory tasks. Coordination theory furthermore predicts that span tasks can be separated into distinct categories, namely, tasks that do and do not require coordination, with step functions in the observed age differences.

Third, the speed theory of cognitive aging (Salthouse, 1996) predicts that age-related slowing of computational processes will cause a decrease in short-term memory—working memory span, either because items are not encoded sufficiently well (Salthouse’s “limited-time mechanism”), or—in the case of working memory tasks—because items that are stored away in a temporary visual or verbal buffer might have decayed from that buffer by the time they are needed for subsequent processing (Salthouse’s “simultaneity mechanism”). Statistically controlling for speed has indeed been found to reduce the magnitude of age-related differences in short-term memory and working memory performance (Salthouse, 1991; Verhaeghen & Salthouse, 1997). The speed theory would predict age differences in mere storage tasks (where the limited-time mechanism operates), and larger age differences still in tasks requiring concurrent processing and storage (governed by both the limited-time and simultaneity mechanism). The existence of a simultaneity mechanism would furthermore lead to the prediction that tasks requiring reordering (such as backward digit span) yield larger age differences than simple storage spans.

Fourth, the inhibition theory of cognitive aging (Hasher & Zacks, 1988) states that an age-related breakdown in inhibitory control leads to an increase in the number of simultaneously active messages, thereby effectively shrinking the size of short-term memory or working memory. With regard to span measures, Hasher and colleagues have suggested that working memory tasks require successful management of previous information; that is, that in order to be successful on a span task, it is necessary to effectively inhibit stimuli from previous sets (Chiappe, Hasher, & Siegel, 2000; Lustig, May, & Hasher, 2001; May, Hasher, & Kane, 1999). Therefore, the inhibition theory would predict that span tasks with repeated information should yield larger age differences. Consequently, forward digit span and letter span should be more age sensitive than word span. In addition, tasks that use similar stimuli for storage and processing (i.e., reading span, listening span, sentence span, and computation span tasks) would be expected to yield larger age differences than tasks that use different stimuli for storage and processing (i.e., operation span tasks).

It should be noted that the status of expected age differences in span measures requiring reordering (such as backward digit span) is somewhat unclear. It seems that most developmental theories (with the exception of the coordination theory) would expect that age differences in these tasks fall somewhere in between the smaller effect on short-term memory span and the larger effect on working memory span. Whereas some would place backward digit span closer to working memory span because of the requirement of effortful processing (Lezak, 1995), others consider it closer to short-term memory tasks because of its similarity to forward digit span (Wechsler, 1955).

All four theories, therefore, agree that age differences should be larger for working memory span than for short-term memory span. The coordination theory is the only theory that does not assume that age differences should emerge in simple storage tasks, although it is not incompatible with that result. A few predictions, however, are theory specific: The inhibition theory is the only theory that makes a clear distinction between the types of items used for recall; the coordination theory has an all-or-none character claimed by none of the other theories.

### Meta-Analysis

In the present meta-analysis we extend the findings of the two previously published meta-analyses (Verhaeghen, Marcoen, & Goossens, 1993; Verhaeghen & Salthouse, 1997). Those meta-analyses focused on effect sizes (i.e., mean standardized differences between younger and older adults or age-by-span correlations) for broad classes of short-term memory and working memory tasks. Our analysis involves a more detailed examination of most verbal span measures. The emphasis is on graphical analysis (see Cerella, Poon, & Williams, 1980, for an early example); that is, we compiled and used span data to construct a Brinley plot, which displays performance of older adults on a given span task as a function of performance of younger adults on the same task. This allows for the examination of span scores in their original metric. The current graphical analysis allows for an examination of variability and consistency across studies utilizing span tasks. In the context of aging research, Brinley plots have been used to test (and account) for a mathematical relation between performance of older and younger adults. Very often, it is found that average performance of a group of older adults can be predicted reliably from the average performance of a group of younger adults across tasks that measure similar types of processing (e.g., Cerella, 1990; Cerella et al.; Myerson, Hale, Wagstaff, & Poon, 1990; Verhaeghen, 2000; Verhaeghen & Cerella, 2002; Verhaeghen & Marcoen, 1993; Verhaeghen et al., 2002). The number of mathematical functions needed to adequately explain the data are taken to be an indication of the number of age-related mechanisms involved (Dunn & Kirsner, 1988).

It should be noted that, typically, Brinley plots are derived from a wide variety of tasks or of experimental manipulations (e.g., Cerella, 1994; Verhaeghen, 2000), leading to a wide range of values on the x axis. This is not the case for the Brinley analysis that we perform here: The difficulty range on the
MEMORY SPAN AND AGING META-ANALYSIS

Sample of Studies

We collected studies by consulting the PsycINFO electronic database, using keywords such as memory span and working memory. In addition, because many studies on cognition and aging report span measures as part of the descriptive information about their samples, we manually examined a sizeable proportion of articles contained in the main journals in the field, namely Psychology and Aging, volumes 1(1) through 17(3), Journal of Gerontology: Psychological Sciences, volumes 36(1) through 57B(3), The Gerontologist, volumes 26(6) through 42(5), and Aging, Neuropsychology, and Cognition, volumes 1(1) through 9(2). The reason we did not manually search older volumes was that memory span was the relative lack of prevalence of articles examining span in these journals. We used references cited in selected studies and in review papers to identify additional studies. We included all studies that (a) compared a younger (age, M < 30 years) versus an older (age, M > 60 years) adult sample, both groups free of any cognitive problems; and (b) assessed participants on at least one memory span task. We concluded our data collection in November 2002.

The resulting database includes 123 studies, contained in 104 articles of memory span and aging. Several studies contained more than one span task. The data set includes 64 studies of forward digit span, 5 studies of letter span, 28 studies of reading span, 5 studies of word span, 57 studies of backward digit span, 28 studies of reading span, 12 studies of listening span, 10 studies of sentence span (reading and listening span data combined), and 14 studies of computation span. The mean age reported of the younger adults in the articles ranged from 16 to 29 years (M = 21.2, SD = 2.4) and ranged from 60.7 to 77.8 years for older adults (M = 70.1, SD = 2.84). The age difference between the younger and older adults averaged across all studies had M = 48.8, SD = 3.9. All articles are listed in the reference section at the end of this article, marked with an asterisk. Given the large sample of studies, we decided not to include the full tabulation of spans by study in the article. Interested readers can obtain this tabulation under the form of an Excel spreadsheet from either author.

Brinley Analysis

In a Brinley analysis, performance of older adults is plotted as a function of performance of younger adults (Brinley, 1965; Cerella et al., 1980). In Brinley space, the diagonal represents the same score for younger and older adults, or no age difference. In the reaction-time literature, several explicit models have been advanced to explain the shape and parameters of young–old functions (for an overview, see Cerella, 1990). The relevant interpretable shapes include a linear function, a power function through the origin (the information-loss model), and a quadratic function through the origin with a linear coefficient fixed at unity (the overhead model). In the present analysis, we decided to fit only linear models, mainly because the rationales behind the nonlinear models are specific to reaction times and do not generalize easily to accuracy. We conducted the Brinley analyses by using the weighted least squares algorithm, weighting for sample size.

We found initial Brinley model runs to be suspect (viz., small R²s and unusual parameter values); therefore, we conducted an outlier analysis by using leverage values and Cook’s d. For this analysis, we regressed span of older adults on span of younger adults for each of the three task types (simple storage, reordering, and working memory) separately. We identified and removed three data points as outliers from all further analyses. Cook’s d values were 1.89 for forward digit span (younger, 9.9; older, 7.8) in Kemper and Sumner (2001), 1.45 for backward digit span (younger, 8.2; older, 5.4) in Kemper and Sumner, and 0.01 for backward digit span (younger, 10.6; older, 8.3) in Fisk, Cooper, Hertzog, Anderson-Garlach, and Lee (1995). The three studies leverage values were 0.15, 0.08, and 0.39, respectively. No obvious aspects of methodology or subject sampling differentiated these studies from the studies that remained in the analysis, but we note that the span scores reported for younger adults in each of these studies is unusually large.

Results

Figure 1 shows the plot of mean span scores of older adults as a function of mean span scores of younger adults (k = 203). Descriptive statistics on these means are provided in Table 1. Prior to the full Brinley analysis, we conducted one preliminary analysis to check for method effects. In this analysis, we compared data from the same types of span tasks, forward digit span and backward digit span, using different methods, namely (a) the WAIS versions prior to the 1981 WAIS-R (k = 17 and 14, for forward digit span and backward digit span, respectively); (b) the 1981 WAIS revision (WAIS-R; k = 15 and 13, respectively); and (c) other methods not disclosed in the article (k = 30 and 30, respectively). A Brinley plot of the data is provided in Figure 2. As we can see, the different methods yield overlapping clouds of data points. One single line sufficed for forward digit span, with R² = .72; when we entered separate intercepts and slopes for each of the three methods, ∆R² = .013; incremental F(4, 56) = 0.69, ns. A single line fit the data quite well for backward digit span, with R² = .78. There were, however, significant differences between methods; when we entered separate intercepts and slopes for each of the three methods, ∆R² = .061; incremental F(4, 53) = 4.94, p < .05. The WAIS version of backward digit span separated from the other two methods. For the WAIS version, the intercept was −0.02 (SE = 0.71) and the slope was 0.88 (SE = 0.16); the common line for the two other backward digit span methods had an intercept of 1.89 (SE = 0.31) and a slope of 0.59 (SE = 0.05). The suggestion is that there are no age differences on backward
digit span when measured by the WAIS (the line is statistically indistinguishable from the diagonal); age differences do emerge when the WAIS-R and other methods are being used. The number of data points obtained from studies using the WAIS is quite small; therefore, we decided not to split the backward digit span measures into subgroups. Rather, as a control analysis, we repeated all Brinley regression analyses subsequently reported here, omitting the WAIS data from the backward digit span pool. This did not change the pattern of the results.

We did model fitting incrementally; that is, we started off with a baseline comparison model (Model 1), assuming that a single regression line suffices to capture the age differences. Subsequent models (Models 2 and 3) refined this analysis in a nested fashion, by allowing different intercepts and slopes for different types of span tasks.

The baseline model (Model 1) assesses the fit of a single linear regression line.

\[
\text{span}_{\text{older}} = a + b \text{ span}_{\text{younger}}. \tag{1}
\]

This model fits the data quite well ($R^2 = .90$). The estimated slope was 1.02 ($SE = 0.02$) and the estimated intercept was −0.87 ($SE = 0.14$). The intercept was significantly different from 0; the slope was not significantly different from 1. This line, then, is parallel to the diagonal, suggesting that, regardless of the nature of the span task, older adults retain about one item fewer than younger adults. The mean span score (weighted for sample size) is 5.61 for younger adults and 4.87 for older adults.

The second model (Model 2) separates short-term memory tasks (i.e., forward digit span, backward digit span, word span, and letter span; $k = 139$) from working memory tasks (i.e., computation span, listening span, reading span, and sentence span; $k = 64$), by allowing distinct intercepts and slopes for these two sets of tasks. We did not find the intercepts to be significant, so we ran the model without intercepts. Compared with a single regression line without a constant ($R^2 = .985$), the second model separating short-term memory and working memory tasks fit significantly better: $R^2 = .991$; $\Delta R^2 = .006$; incremental $F(1, 201) = 131.65$, $p < .05$. We found the slope to equal 0.92 ($SE = 0.01$) for the short-term memory tasks and 0.74 ($SE = .02$) for the working memory tasks. For both regression lines, the slopes were reliably smaller than 1. The mean span scores for older adults on the short-term memory span tasks were 6.03 and 3.01 on the working memory span tasks. For younger adults, these numbers were 6.59 and 4.03, respectively.

In our third model (Model 3), we separated backward span scores ($k = 57$) from the short-term memory measures requiring simple storage ($k = 82$), while still keeping working memory separate. We still did not find the intercepts for short-term memory and working memory to be significant, so we ran the model with only an intercept for backward span. This third model fitted the data significantly better than Model 2: $R^2 = .991$; $\Delta R^2 = .000$; incremental $F(2, 199) = 3.15$, $p < .05$. For the simple storage tasks, the slope equals 0.92 ($SE = 0.01$). For backward digit span, the intercept equals 0.76 ($SE = 0.35$) and the slope equals 0.78 ($SE = 0.07$). For the working memory tasks, the slope equals 0.74 ($SE = 0.02$). All slopes, with the exception of the slope for simple storage tasks, are significantly smaller than 1. The intercept for backward digit span is

<table>
<thead>
<tr>
<th>Task</th>
<th>$k$</th>
<th>$M$</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Age Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forward digit span</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Younger</td>
<td>64</td>
<td>7.59</td>
<td>0.99</td>
<td>5.91</td>
<td>10.50</td>
<td>0.53</td>
</tr>
<tr>
<td>Older</td>
<td>7.06</td>
<td>1.02</td>
<td>4.19</td>
<td>9.80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Word span</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Younger</td>
<td>13</td>
<td>5.17</td>
<td>0.47</td>
<td>4.29</td>
<td>6.02</td>
<td>0.65</td>
</tr>
<tr>
<td>Older</td>
<td>4.52</td>
<td>0.37</td>
<td>3.90</td>
<td>5.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Letter span</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Younger</td>
<td>5</td>
<td>5.86</td>
<td>0.32</td>
<td>5.09</td>
<td>6.43</td>
<td>0.97</td>
</tr>
<tr>
<td>Older</td>
<td>4.89</td>
<td>0.30</td>
<td>4.47</td>
<td>5.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Backward digit span</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Younger</td>
<td>57</td>
<td>5.88</td>
<td>1.10</td>
<td>4.40</td>
<td>8.93</td>
<td>0.54</td>
</tr>
<tr>
<td>Older</td>
<td>3.34</td>
<td>0.96</td>
<td>4.04</td>
<td>8.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Computation span</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Younger</td>
<td>14</td>
<td>4.58</td>
<td>1.14</td>
<td>3.28</td>
<td>6.70</td>
<td>1.54</td>
</tr>
<tr>
<td>Older</td>
<td>3.04</td>
<td>1.05</td>
<td>1.83</td>
<td>5.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Listening span</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Younger</td>
<td>12</td>
<td>3.74</td>
<td>0.58</td>
<td>2.85</td>
<td>5.40</td>
<td>1.27</td>
</tr>
<tr>
<td>Older</td>
<td>2.47</td>
<td>0.47</td>
<td>1.66</td>
<td>4.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reading span</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Younger</td>
<td>28</td>
<td>3.49</td>
<td>0.57</td>
<td>2.40</td>
<td>5.00</td>
<td>0.63</td>
</tr>
<tr>
<td>Older</td>
<td>2.86</td>
<td>0.49</td>
<td>2.01</td>
<td>4.30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sentence span</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Younger</td>
<td>10</td>
<td>4.81</td>
<td>0.65</td>
<td>2.70</td>
<td>5.30</td>
<td>1.01</td>
</tr>
<tr>
<td>Older</td>
<td>3.80</td>
<td>0.53</td>
<td>2.32</td>
<td>4.50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Tables values are weighted for sample size; $k =$ number of studies.
significantly larger than 0. We considered additional splits of the data, but they did not yield significant improvements in model fit. This third model then provides the best empirical account of the data. It is shown in Figure 1, overlaid on the data.

Note that there is some heterogeneity in methods to estimate span. To assess the effect of these discrepancies, we reestimated the models, including only those studies that utilized exactly identical procedures within each type of span task (k = 126). The included procedures used two or three trials per length in serial order. The computation span procedure required math to be generated orally and memory for the second digit in the equation (Salthouse & Babcock, 1991). The listening span procedure required auditory presentation and memory for the last word (Salthouse & Babcock). The reading span procedure had visual presentation and memory for the last word without any additional task beyond reading (Hartley, 1986). The sentence span procedure was the same as that described by Stine and Wingfield (1990), combining previously mentioned reading and listening span. We could not include letter span because of heterogeneity in procedures. The word span procedure required auditory presentation and oral participant response. Finally, we included WAIS and WAIS-R procedures for forward digit span and backward digit span (Wechsler, 1955; 1981). The model fits were not significantly better compared with previous analysis (without intercepts); $R^2 = .988$, $R^2 = .992$, and $R^2 = .993$ for Models 1, 2, and 3, respectively, and the parameter values did not differ substantially from those obtained on the full sample of studies. The analysis suggests that procedural differences are not responsible for the observed variability among span scores.

**DISCUSSION**

The present meta-analysis supports four main conclusions: (a) there are age differences in all aspects of memory span; (b) there is a mathematical relationship between the span of younger adults and the span of older adults; (c) working memory span is more age sensitive than short-term memory span; (d) backward digit span is more age sensitive than short-term memory span but less than working memory span.

**Reliable Age Differences in all Span Tasks**

In our Brinley analysis, we found evidence for three groupings of age effects on the temporary memory buffer (see the subsequent paragraphs). Age effects were significant in all three groupings, as indicated by intercepts smaller than 0, slopes smaller than 1, or both. Moreover, only 16 out of the 203 studies included in the Brinley analysis showed a positive age effect (i.e., a higher mean for older than younger adults): 9 of these were in forward digit span, 6 in backward digit span, and 1 from the reading span task. The evidence points overwhelmingly to the existence of negative age differences in verbal span measures.

We note one potential method effect: Data gathered by use of the WAIS showed a smaller age deficit in backward digit span than data gathered by use of other methods, including the later WAIS-R version. It is unclear what accounts for this effect. It is possible that sampling differences or cohort effects play a role. It should be noted that approximately 50% of the studies did not report the methodology used to obtain forward and backward digit span scores, and therefore we could not test true methodological differences in the current study.

**The Mathematical Relation Between the Spans of Older and Younger Adults**

The Brinley analysis demonstrates that there is a mathematical relation between the span of older adults and the span of younger adults. It can be captured well by a linear model. About 93% of the variance in older adults’ mean span length can be explained by knowing (a) younger adults’ mean span length;
and (b) the nature of the task (i.e., whether it concerns a simple storage short-term memory task, a backward digit span task, or a working memory task). Young adults' mean span alone (i.e., without the task type specification) explained 90% of the variance. Such strong mathematical relations are typically taken as evidence that age differences are quantitative rather than qualitative (e.g., Cerella, 1994; Cerella et al., 1980; Verhaeghen, 2000). That is, these relations suggest that, within each of the three groupings, the span of older adults is tightly tied to the span of younger adults, and that whatever variations in procedure and stimuli drive span differences in younger adults will result in similar effects in older adults.

The linear function is characterized by an intercept that is 0 (for simple storage span and working memory span) or positive (for backward digit span) and a slope smaller than 1. At first blush, this model seems quite strange: The implication is that larger age differences are observed for conditions that yield higher spans for younger adults, that is, the conditions that are presumably easier. This implication goes against the complexity hypothesis (e.g., Cerella et al., 1980; Salthouse, 1991), which states that the more complex or more difficult tasks will yield the larger age differences.

This apparent discrepancy with the literature becomes understandable if we interpret our findings as follows: For simple storage span and for working memory span, older adults simply lose a constant proportion of items, compared with younger adults. That proportion is indicated by the slope of the linear function; that is, for simple storage span, older adults retain about 90% of the elements in working memory that younger adults can retain; for working memory span, this proportion goes down to about 75%. The line for backward digit span can be explained as deriving from a bipartite structure. That is, one interpretation of our findings is to consider the line for backward span as bifurcating from the line for forward span. An inspection of Figure 1 suggests that the point of bifurcation is close to the point where younger adults retain about four items. After this point, age differences increase proportionally with the span of the young (to be exact, forward digit span and backward digit span regression lines cross at span younger = 3.42). This interpretation raises the intriguing possibility that the age deficits in backward digit span may be related to the size of the focus of attention, which is often believed to be about four items (e.g., Cowan, 2001). The focus of attention is defined as the number of items held in working memory that is immediately available and accessible for processing. The interpretation is then that the size of the focus of attention changes little or not at all with aging; an additional assumption is that reordering items within the focus of attention occurs at little cost. Specifically, in older adults, the access processes within the focus of attention appear to be about 95% efficacious compared with access processes in younger adults; larger age differences (75% relative efficacy) only emerge when it becomes necessary to shunt items in and out of the focus of attention for reordering.

The focus-of-attention interpretation might also explain why age differences are slight in simple storage tasks (items can be retrieved without shunting), and why they are large in working memory tasks (in such tasks the focus of attention may shrink to include one and only one item; see, e.g., Garavan, 1998; McElree, 2001; Oberauer, 2002), which would necessitate shunting operations for all items. If this explanation is correct, then age differences in all span tasks might be explained by a single mechanism, namely the mechanism that drives the 75% information loss for items stored outside the focus of attention. This could be the simultaneity mechanism of cognitive slowing (Salthouse, 1996), or mechanisms related to age differences in volume and activation of the prefrontal cortex (Park, Polk, Mikels, Taylor, & Marshuetz, 2001; Raz, 2000; Reuter-Lorenz, et al., 2000; Rypma & D'Esposito, 2000). Circumstantial evidence for this model would include demonstrations of the absence of age differences in retrievability of items within the focus of attention coupled with an age deficit for availability of items outside the focus of attention. At least two studies have obtained such results (Verhaeghen & Basak, in press; Verhaeghen & Hoyer, in press).

Another unusual aspect of the model is that the positive intercept for backward digit span implies that, when the tasks are very difficult and very small span scores are obtained, the line will lie above the diagonal; that is, older adults will actually perform better than younger adults. One should keep in mind, however, that these are empirical functions and that they are only defined in the range from which they have been derived. Figure 1 shows that, for each of the task families, only a few data points actually fall in the span range for which the curves would lie above the diagonal.

**Working Memory Span is More Age Sensitive Than Short-Term Memory Span**

The Brinley analysis shows confirmation of prior findings of an age-related dissociation between simple storage span and working memory span. The former yields 95% relative retention of items in older adults as compared with younger adults; the latter yields 73%. It should be noted that, although the mean spans for short-term memory and working memory are quite different, possibly complicating our Brinley analysis, there is a region on the abscissa, between the span values of 4 and 6, where data are available for both short-term memory and working memory measures. An inspection of Figure 1 shows that for that region, there is little overlap between the data clouds for short-term memory and those for working memory measures. This vertical spatial separation in a region of horizontal overlap can be taken as additional strong evidence for the existence of distinct age effects on the two task types (Perfect, 1994).

The raw data show the same effect (Table 1): The average age difference in simple storage spans combined (weighting for sample size) is 0.59 items; the average age difference in working memory span is 1.02 items. Anomalies exist, however, at the level of individual tasks: reading span, a working memory task, shows a small age deficit (0.63 items), whereas letter span, a simple storage task, shows a large effect (0.97 items). The larger age deficit in letter span may be due to increased susceptibility to interference with advancing age—many letters sound alike, and the code of short-term memory for letters is primarily auditory. Reading span also yields the lowest overall score, and the smaller age deficit in this measure might be caused simply by a floor effect.

**Age Differences in Backward Digit Span**

The finding of a distinction between forward and backward digit spans in the Brinley plot should be appealing to those researchers that identify short-term memory and working mem-
ory by task requirements: mere storage (forward digit span) versus reordering (backward digit span). Previous findings comparing forward and backward digit spans have yielded mixed results. In agreement with the current findings, Slivinski and Busche (1999), in a longitudinal study, found significant age-related declines on the backward digit span test, but they failed to find age effects on the forward digit span. Babcock and Salthouse (1990), in a meta-analysis (14 studies forward digit span–backward digit span), also found slightly larger age differences for backward digit span compared with forward digit span. The current result, however, is in line neither with prior meta-analytic reports of equivalent effect sizes for forward and backward digit spans (Verhaeghen et al., 1993), nor with three large-sample studies that likewise failed to find an increase in age difference between the forward and backward digit spans over the adult life span (Gregoire & Van der Linden, 1997; Myerson et al., 2003; Park et al., 2001).

One of the reviewers mentioned the possibility that the aggregation of digit forward span with word span and letter span might account for the discrepancy between our result and the large-scale individual studies already mentioned. To test for this possibility, we ran a regression analysis allowing separate slopes within simple storage span (forward digit span vs word and letter span), in addition to the functions for backward span and working memory span already included in our final model. The model fit was significantly better than that of the final model, $R^2 = .992$; $\Delta R^2 = .000$; incremental $F(1, 198) = 8.10, p < .01$. For forward span, the slope equaled 0.93 ($SE = 0.01$), and the slope for word span and letter span combined equaled 0.86 ($SE = 0.03$). Therefore, the relationship between simple storage span and backward span was not altered by the separation of forward digit span from word and letter span.

Figure 1 suggests a possible source of this discrepancy in findings. The age differences between mere-storage spans and reordering spans are slight and are only observed when the reordering span task yields scores of 6 or higher for younger adults. Under typical conditions, in which reordering span scores of younger adults are close to 6 (see Table 1), our model predicts that age differences will not be observed. Therefore, the age differences in backward digit span (0.54) and forward digit span (0.53) are not significantly different by examining averaged data.

Conclusions

In summary, we found that adult age differences are present in all span measures, that is, in tasks requiring mere storage (short-term memory) as well as in tasks requiring storage and processing (working memory). Simple storage tasks yield smaller age effects than working memory span tasks (this replicates the results of earlier meta-analyses; see Verhaeghen et al., 1993; Verhaeghen & Salthouse, 1997). The age effects of reordering tasks (only backward digit span in the current study) fall in between. These results are in line with most of the aging theories described in the introduction. Inhibition theory, however, predicts larger age differences in forward and backward digit span tasks than in other mere storage tasks, and this was not the case (Table 1). Coordination theory does not predict the observed dissociation between digit span backward and digit span forward.

Acknowledgments

This research was supported in part by Grant AG-16201 from the National Institute on Aging. The research was conducted in partial fulfillment of the requirements for K. L. Bopp’s doctoral degree. We thank John Cerella and Martin Slivinski for helpful comments on earlier versions of this article.

Address correspondence to Kara Bopp, who is now at Wofford College, 429 North Church St., Wofford College, Spartanburg, SC 29303; or Paul Verhaeghen, Department of Psychology, 430 Huntington Hall, Syracuse University, Syracuse, NY 13244-2340. E-mail: BoppKL@wofford.edu or pverhaeg@psych.syr.edu

References

Articles used in the meta-analysis are marked with an asterisk; the type of span is indicated by acronym (FS = forward digit, LET = letter, WS = word, BS = backward digit, AS = alphabet, CS = computation, LS = listening, RS = reading, and SS = sentence spans, respectively).


Received June 14, 2004
Accepted March 3, 2005
Decision Editor: Thomas M. Hess, PhD