Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Memory and Language 51 (2004) 80–96
Journal of Memory and Language
www.elsevier.com/locate/jml

Working memory capacity and resistance to interference☆

Klaus Oberauer,[a,*] Elke Lange,[a] and Randall W. Engle[b]

[a] University of Potsdam, Allgemeine Psychologie I, P.O. Box 60 15 53, Potsdam 14415, Germany
[b] Georgia Institute of Technology, Atlanta, GA, USA

Received 7 October 2003; revision received 27 February 2004
Available online 6 May 2004

Abstract

Single-task and dual-task versions of verbal and spatial serial order memory tasks were administered to 120 students tested for working memory capacity with four previously validated measures. In the dual-task versions, similarity between the memory material and the material of the secondary processing task was varied. With verbal material, three additional words had to be read aloud in the retention interval, and their phonological and semantic similarity to memory list words was varied orthogonally. With spatial material, choice RT tasks in the retention interval used stimuli from either the same or a different kind as the memory stimuli. Similarity had little effect on dual-task costs. For correlational analyses, individual dual-task costs were measured in various ways, which varied as to their direction of correlation with working memory capacity. In general, these correlations were low. Dual-task costs, although measured reliably, did not correlate across verbal vs. spatial tasks. The results lend little support to theories identifying working memory capacity with the ability to resist interference, or the ability to coordinate two concurrent tasks.
© 2004 Elsevier Inc. All rights reserved.

Working memory capacity is often measured by tasks combining memory for serial order with a processing demand. For example, the reading span task designed by Daneman and Carpenter (1980) requires reading of a series of sentences, often followed by some judgment, together with memory for the last words of the sentences in their order of presentation. Likewise, operation span (Turner & Engle, 1989) and counting span (Case, Kurland, & Goldberg, 1982) combine arithmetic operations with retention of words or digits. These so-called "complex span tasks" can be described as dual tasks requiring simultaneous short-term retention of some information and processing of other, often unrelated, information, thereby matching the definition of working memory as a system for simultaneous storage and processing (Baddeley, 1986).

Several studies have shown that complex span tasks can be dissociated from simple span tasks (i.e., serial recall tasks without additional processing demand) by factor analysis (Cantor, Engle, & Hamilton, 1991; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Kail & Hall, 2001; Oberauer, Süß, Wilhelm, & Wittmann, 2003) and by neuroimaging (Fletcher & Henson, 2001; Postle, Berger, & D'Esposito, 1999; Smith, Geva, Jonides, Miller, Reuter-Lorenz, & Koeppe, 2001; Smith & Jonides, 1999). Moreover, complex spans are better predictors than simple spans of performance in complex cognitive tasks such as language comprehension (for a review see Daneman & Merikle, 1996) and reasoning (Conway et al., 2002; Engle et al., 1999; Kyllonen & Christal, 1990; Süß, Oberauer, Wittmann, Wilhelm,

☆ This work was supported by a TransCoop grant from the Alexander v. Humboldt Foundation to K. Oberauer and R.W. Engle, and by Grant OB 121/3-2 of Deutsche Forschungsgemeinschaft (DFG). We thank Petra Grüttner for her invaluable help by organizing the data collection. We also thank Katrin Göthe, Christina Körner, and Nadine Meyer for their help in scoring the tests.
* Corresponding author.
E-mail address: ko@rz.uni-potsdam.de (K. Oberauer).

0749-596X/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2004.03.003

& Schulze, 2002). It seems that "complex" or dual-task spans capture systematic variance not contained in "simple" spans, and this variance is strongly associated with complex cognition. Understanding what this variance is would constitute an important step toward understanding reasoning ability and, indeed, general fluid intelligence (Engle et al., 1999).

One hypothesis to account for the unique variance of complex span is that it reflects—more than simple span—the ability to resist interference. Complex span tasks require maintenance of a list of items in the face of other information that needs to be processed but is irrelevant to the memory task. Therefore, Engle et al. (1999) proposed that working memory capacity is the ability to temporarily maintain representations activated in the face of distraction. Their view can be summarized by the equation "complex span = simple span + controlled attention." Complex span tasks share with simple span tasks that list elements must be maintained in working memory, but in addition, complex span requires controlled (or executive) attention in order to guard these representations from distracting information from the secondary task. Starting from mathematical models of performance in a working memory task, Oberauer and Kliegl (2001) likewise came to the conclusion that a parameter capturing the amount of interference between elements to be held in working memory can account for individual and age differences in capacity.

The view that links working memory capacity to the ability to resist interference has been questioned by Jenkins, Myerson, Hale, and Fry (1999). They regressed participants' complex spans on their simple spans and consistently observed slopes smaller than one, implying that with increasing simple span dual-task costs, that is the difference between simple and complex spans, also increase. If simple span reflects working memory capacity, and if working memory capacity is the ability to resist interference, the opposite should be observed: People with high simple spans should suffer less, not more, dual-task interference. Oberauer and Süß (2000) argued that this analysis is misleading for two reasons: First, the same simple span score that is used to calculate individual dual-task costs is also used as a predictor of these costs, leading to an artificial lowering of the slopes due to regression to the mean. Second, simple span is not an adequate measure of working memory capacity, given the evidence that complex span measures something different from simple span (see Myerson, Jenkins, Hale, & Sliwinski, 2000, for a response to this critique). One obvious way to avoid these problems is to compute individual dual-task costs from measures of simple and complex spans and predict them by an independent measure of working memory capacity. This was done in the present study.

The term interference is often used in a broad, descriptive way encompassing every loss of performance in a primary task due to additional information processing. Here we want to investigate two theoretically specified concepts of interference. The first one, implicated by the view of Engle et al. (1999), is interference through distraction of attention. This means that a limited attentional resource is partially drawn away from the primary task by representations that are irrelevant to it but nonetheless demand resources, either because they require a response (as in dual tasks) or because they attract attention involuntarily (as in selective-attention tasks). The second concept, suggested by Oberauer and Kliegl (2001), is interference through partial overwriting of overlapping representations. If several distributed representations are held in working memory at the same time, they tend to overwrite each other to the degree that they share some of their features (see also Nairne, 1990). One important difference between these two concepts of interference regards the role of similarity. Overwriting implies that the amount of interference is a function of the representations' similarity (i.e., degree of overlap), whereas distraction of attention is not necessarily linked to the similarity of the representations involved. Therefore, in the present study we varied the degree of similarity between the potentially interfering representations.

Another hypothesis, closely related to but not identical with the view of Engle et al. (1999), is the assumption that the specific variance associated with complex spans reflects the efficiency of a central executive that is needed to coordinate the two partial tasks in dual-task settings (Baddeley, 1996; Baddeley & Della Sala, 1996). Since the central executive is thought of as a domain-general system, this hypothesis implies that the amount of dual-task costs, which reflects the efficiency of the central executive, be correlated across domains. Preliminary evidence for this contention has been obtained by Bayliss, Jarrold, Gunn, and Baddeley (2003) and by Oberauer et al. (2003). The present study provided another opportunity to test the generality of dual-task costs.

We constructed simple and complex span tasks with verbal and spatial content. The simple span tasks were forward serial recall tasks. The complex span tasks were the same serial recall tasks with an unrelated processing requirement added in the retention interval between presentation and recall of the list. For the spatial tasks, this was a speeded classification of abstract patterns as symmetrical or non-symmetrical. For the verbal tasks, the intervening processing requirement was simply to read three additional words aloud. The complex tasks were modeled closely after tasks that had high loadings on working memory factors in two large factor-analytic studies ("verbal span" in Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000; "word span (dual)" and "pattern span (dual)" in Oberauer et al., 2003).

Much research has been conducted to demonstrate that dual-task interference in working memory is larger

when the two tasks come from the same broad content domain (Baddeley, 1986; Cocchini, Logie, Della Sala, MacPherson, & Baddeley, 2002). This finding could be interpreted as evidence for similarity-based interference. Baddeley and his colleagues, however, interpret the dissociations between broad content domains as reflecting separate sub-systems. If this is correct, these findings would imply nothing with regard to whether representations within each system interfere with each other to the degree that they overlap. Therefore, we manipulated similarity within content domains, in ways similar to the manipulation of inter-item similarity in serial recall tasks (e.g., Baddeley & Ecob, 1970; Saint-Aubin & Poirier, 1999).

Similarity in the spatial domain was varied by using two categories of materials: partially filled matrices ("patterns") and partially connected dot grids ("lines"). Examples for these stimuli are shown in Fig. 1. In the similar condition, the processing task used stimuli of the same category as the items of the memory list, whereas in the dissimilar condition stimuli from the other category were used. In the verbal domain we varied phonological and semantic similarity orthogonally. All items in the memory lists were nouns referring to animals or plants. In the semantically similar conditions, the words used for the processing task were also animal or plant nouns, whereas in the dissimilar condition they were inanimate nouns. Phonological similarity was realized by having the three words to be read aloud overlap phonologically with three of the words in the memory list.

We computed the amount of interference (i.e., dual-task costs) by comparing performance in the single and the corresponding complex tasks in several ways. Most dual-task studies use a single measure of dual-task costs without even discussing alternative ways of computing costs. As we will argue below, there are several ways to compute dual-task costs, each relying on its own rationale, and it is not clear a priori that they all reflect the same construct, and if they don't, which one is best suited as indicator of a theoretically interesting variable such as resistance to interference. Therefore, a further goal of this work is to provide an empirically informed comparison of different ways to compute dual-task costs.

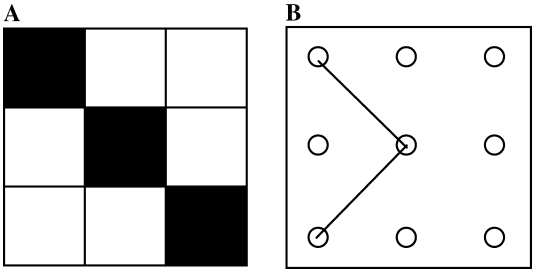

Fig. 1. Example stimuli for the spatial experimental tasks: Patterns (A) and lines (B).

To summarize, we pursue four goals: (1) We tested the hypothesis that working memory capacity reflects the ability to resist interference with the contents of working memory. To this end, we calculated individual dual-task costs in complex span tasks and predicted them by an independent measure of working memory capacity (called WMC here), consisting of one verbal, one numerical, and two spatial tasks validated as representative working memory tasks by Oberauer et al. (2000). If working memory capacity is identical with—or strongly related to—the ability to resist interference in complex span tasks, then there should be a negative correlation between WMC scores and dual-task costs. This prediction holds for both concepts of interference discussed above. (2) We investigated whether the amount of interference depends on the similarity between the memory material and the distracting material. The concept of interference through overlap predicts that dual-task costs will be larger in the similar than in the dissimilar conditions, whereas the concept of interference as distraction of attention makes no such prediction. (3) We tested the hypothesis that dual-task costs in complex span tasks are domain-general. If they are, then indicators of dual-task costs from purely verbal task combinations should correlate with those from purely spatial task combinations. (4) An additional methodological goal was to compare several reasonable procedures to calculate dual-task costs with regard to their psychometric quality, their correlations among each other, and their relationship to working memory capacity.

Method

Participants

Participants were 120 high-school students from Potsdam. Their age ranged from 17 to 19 years, and 54 of them were male. They received 30.00, DM (about $15.00) for participation in three 1-h sessions.

Design and materials

The experimental design was implemented by 11 blocks of memory tasks; five verbal and six spatial. Two of the spatial and one of the verbal blocks were single tasks, requiring serial recall of a list of items. The remaining blocks were dual tasks, combining serial recall with a processing requirement in the retention interval. Each block consisted of four consecutive sub-blocks. Within each sub-block, each memory list length was realized once. The list lengths were presented in a new pseudo-random order within each sub-block; these

orders were the same for all participants. The block structure is illustrated in Fig. 2. For practical reasons, the verbal single-task block was divided into two parallel blocks, each consisting of two sub-blocks, so that there was an equal number of verbal and spatial blocks. Table 1 gives an overview of the design.

The ranges of list lengths for each block were chosen based on previous experience with similar tasks and a small pilot study. We attempted to cover a large range of difficulty of the primary task (i.e., serial recall) in order to avoid restrictions of range. In addition, we intended to construct span scores from the data, mimicking the adaptive procedure by which spans are commonly determined. This requires that we obtain data from list lengths large enough to reach the maximum length that can be recalled without error by even the most able participants.

The blocks were administered in pairs, such that the two blocks in the same row of Table 1 always followed each other. Half of the participants started with the verbal block within each pair, the other half with the spatial block. The order of the six pairs of blocks was counterbalanced over participants: In each of two sessions, a participant either worked through the three block pairs with even numbers or the three pairs with odd numbers. All 12 orders of block pairs possible within this constraint were realized for an equal number of participants.

Materials for verbal tasks

All memory lists of the verbal tasks consisted of nouns designating animals or plants (animate nouns). In the dual tasks with high semantic similarity, the secondary task list also consisted of animate nouns,

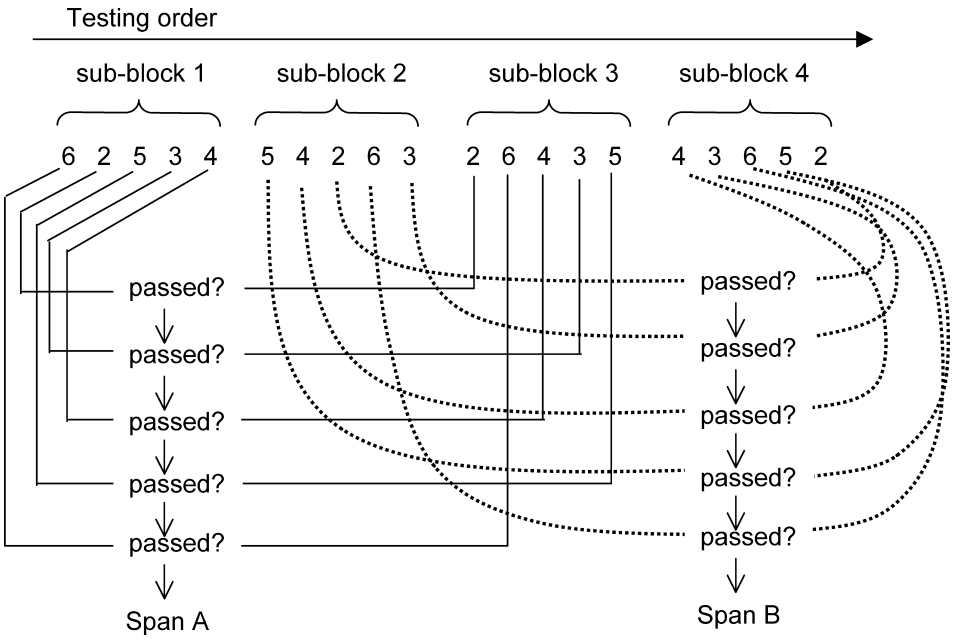

Fig. 2. Illustration of the sub-block structure of a task (digits in the top row represent list lengths of individual trials) and the procedure to emulate spans. Two parallel spans A and B are built from two pairs of sub-blocks by scoring trials as "passed" or "failed," starting with the smallest list length and proceeding until a participant failed both trials on a given list length level.

Table 1
Experimental design

Block number	Verbal	List lengths	Spatial	List lengths
1	Single (1)	3–9	Single patterns	2–6
2	Single (2)	3–9	Single lines	2–6
3	Dual, sem. similar, phon. similar	2–6	Dual patterns similar	1–4
4	Dual, sem. similar, phon. dissimilar	2–6	Dual patterns dissimilar	1–4
5	Dual, sem. dissimilar, phon. similar	2–6	Dual lines similar	1–4
6	Dual, sem. dissimilar, phon. dissimilar	2–6	Dual lines dissimilar	1–4

Note. The total number of trials in each block can be obtained by multiplying the range of list lengths by 4 (except the two verbal single blocks, which contained each list length only twice).

whereas in the low semantic similarity condition they consisted of inanimate nouns. In conditions with high phonological similarity each of the three words in the secondary task list was phonologically similar to one word in the memory list. The lists for all tasks were constructed by a computer program that sampled words at random without replacement from one of five sets described below. In this way, no word was repeated in a memory list within a block.

Set 1 consisted of 56 pairs of animate nouns between one and three syllables long. Pairs were constructed such that they overlapped phonologically, that is, they shared several phonemes at the same within-word positions. Twenty-one of these pairs were rhyming words. Eight nouns figured once in the first-noun role and once in the second-noun role; apart from this, no noun was repeated in this set. *Set 2* consisted of 56 pairs of one- to three syllable nouns; the first noun of each pair was identical to the first noun of Set 1. The second was an inanimate noun (i.e., a noun designated a non-living thing) and was selected such that it had high phonological overlap with the first noun. Twenty-five of these pairs were rhyming words. *Set 3* consisted of 24 additional animate nouns in the same length range, none of which was identical with one of the nouns in Sets 1 and 2. *Set 4* consisted of 60 nouns of the same category as Set 3; this set was reserved for the secondary tasks. *Set 5* comprised 60 one- to three syllable inanimate nouns, which were also used for the secondary tasks.

For the dual tasks, the computer program first determined three serial positions in the memory list at random and assigned to them the first noun of one of the pairs selected at random from Set 1 (for the semantically similar condition) or Set 2 (for the semantically dissimilar condition). For trials with list length 2, one of the three nouns was dropped. If the trial had a list length larger than three, the remaining serial positions of the memory list were filled at random from Set 3. The serial positions selected for the nouns taken from the pairs were the same in matching trials of the four dual-task blocks.

In the condition with high phonological similarity, the second noun from each pair was assigned at random to one of the three serial positions in the secondary-task list. This is the list of words that had to be read aloud in the retention interval. In this way, each word in the secondary-task list was phonologically similar to one word of the memory list. In the condition with low phonological similarity, the secondary-task words were selected at random from Set 4 (for high semantic similarity) or Set 5 (for low semantic similarity). Thus, the secondary-task words in this condition were not phonologically related to words in the memory list beyond chance level.

The single-task memory lists within the range of list lengths also covered by dual tasks (3–6) were constructed by the same procedure as the dual-task memory lists. This ensured that within the comparable range of list lengths the words used by single-tasks and dual-tasks were sampled from the same pool. Thus, any difference between single-task and dual-task performance cannot be attributed to specific characteristics of the words used. The words used to construct single-task lists of higher length were selected at random from Set 4.

Materials for spatial tasks

There were two categories of stimuli used for the spatial tasks (see Fig. 1). The patterns were partially filled 3×3 matrices. The lines were constructed from a 3×3 dot arrangement in which a subset of dots was connected by straight lines, forming a single connected line drawing. For the memory lists, we constructed sets of relatively simple patterns and lines, because pilot testing suggested that with complex patterns, spans would hardly go off the floor for many participants. The set of patterns, for instance, included all matrices with two adjacent cells filled, all matrices with three filled cells in a row, all L-formed and T-formed patterns, etc. The set of lines consisted only of single lines, two connected lines, and Z-formed drawings. The sets of stimuli for the secondary task were more complex. Since the secondary task was classifying stimuli according to their symmetry along a vertical axis, half of the stimuli constructed for the secondary task were symmetrical and half were not. None of them was identical with a stimulus used for the memory task.

All tasks were constructed by a computer program that selected stimuli at random without replacement from the appropriate sets. There were two blocks of single tasks, one consisting of lists of patterns, the other consisting of lists of lines. The four dual-task blocks were formed by combining each category of stimuli for the memory task with each category for the secondary task, thus generating two blocks where the same category was used for both the memory and processing task (i.e., high similarity), and two blocks where the stimuli for the two tasks were from different categories (i.e., low similarity).

Working memory capacity test

The test used to obtain an independent measure of working memory capacity (WMC) consisted of four tasks previously used by Oberauer et al. (2000, 2003): (1) Reading span: Participants read lists of sentences, each presented for 4 s, and memorized the last word of each sentence. After a varying number of sentences, the last words had to be recalled in order. (2) Memory updating numerical: One digit is presented in each of several frames on the screen, which must be memorized. This is followed by a sequence of arithmetic operations presented in individual frames, by which the

content of that frame must be updated. (3) Memory updating spatial (STM): In each of a number of frames, a dot is presented in one of 9 possible locations. After sequential presentation of the dots, participants must recall dot positions for selected frames. (4) Spatial coordination: A varying number of dots is sequentially presented in cells of a 10×10 grid. The pattern which the dots would form if presented simultaneously must be reproduced in an empty grid. Each task consisted of 11–15 trials, administered by computer. Readers are referred to the original references for a detailed description.

Procedure

We tested participants individually in quiet rooms. The WMC test was administered in a single session about six months before the experimental tasks in the context of another study. Participants worked through six blocks of experimental tasks in each of two sessions. They received a booklet containing a written instruction for each block, followed by answer sheets for the tasks to follow. Each block of test trials was preceded by two practice trials.

Participants started each trial by pressing the space bar. This triggered the display of the word "Achtung" (alert) for 500 ms, followed by a 500 ms blank, after which the first stimulus of the memory list was presented. All memory stimuli were presented sequentially in the center of the screen. Words were displayed in red ink within a white rectangle on a black background; spatial stimuli were presented in white. Words were presented for 1 s, spatial stimuli for 1.6 s. Each stimulus was followed by a 200 ms inter-stimulus interval (ISI). After the last memory item and its ISI, the word "Ende" (end) was displayed centrally in white ink for 500 ms, followed by a 500 ms blank. When the task was a single task, the line "please write down words" or "please draw patterns" was displayed at this point, and participants were required to recall the memory list in the correct order on the answer sheet. When they finished that, participants started the next trial by pressing the space bar.

In case of a dual-task block, the display of "Ende" was followed by presentation of the first stimulus of the secondary-task list. For the verbal tasks this was a word printed in yellow within a white frame in the center of the screen, slightly shifted to the bottom. Participants were required to read this word aloud as quickly as possible and press the space bar when ready, after which the current word was deleted and the next word was displayed with an ISI of 200 ms. There were three words to be read aloud on each trial. Participants were aware that their responses were tape-recorded. For the spatial task, the stimuli were presented centrally, slightly removed to the bottom relative to the memory stimuli, and flanked by the words "symmetrical" and "not symmetrical" to the left and the right, respectively, corresponding to the arrow keys to which each response was mapped in the instructions. Participants were required to press the correct arrow key as quickly as possible. A false response elicited a 300 ms warning tone of 300 Hz. In both tasks, the line "too slow!" was displayed for 500 ms as feedback to each reaction time exceeding 5 s. After each key press, the next stimulus was presented with a response–stimulus interval of 200 ms. Following the reaction to the third stimulus, the instruction to write down the words or to draw the patterns was displayed as in the single tasks.

For the verbal tasks, the answer sheets consisted of a number of columns with empty lines. There was one column for every trial, and they were numbered by the trial number. Each column had a number of rows equal to the maximum list length of the task (i.e., 9 for single tasks and 6 for dual tasks), so that participants could not anticipate the list length of each trial from the answer sheet. They were required to write down the words beginning with the top row, leaving a blank line for words they could not remember, and leaving blank the remaining lines exceeding the actual list length. For the spatial tasks, the answer sheets contained rows of empty matrices (for patterns) or dot grids (for lines). Answers were given by marking the filled cells in the matrices (e.g., by a cross) or by drawing the appropriate lines connecting dots, respectively. Again, the number of empty patterns or grids was equal to the maximum list length.

Data treatment

We computed two measures of memory performance in the experimental tasks. One is the percentage of items recalled correctly in the correct serial position, averaged over all trials in a block. The second performance measure is a pseudo-span score computed by emulating a span testing procedure. The procedure is illustrated in Fig. 2. We first paired sub-block 1 with sub-block 3 and sub-block 2 with sub-block 4 within each block, thus obtaining two independent sets of trials containing each list length twice. From this we determined the pseudo-span separately for each pair of sub-blocks in the following way: First, each participant's span was set to the smallest list length of the block minus one. Starting with the smallest list length, the span was incremented to that length if the participant passed both trials in a pair of sub-blocks, and at that length minus 0.5 if he or she passed one of the two trials. A trial was scored as passed if the complete list was recalled correctly, and as failed otherwise. If at least one trial in a pair of sub-blocks was passed, the procedure moved on to the next list length, otherwise

it stopped, and all higher list lengths were disregarded for the computation of the span.

Results

Summary statistics of the two performance measures for the experimental tasks are given in Table 2. We obtained a reliability estimate for the percent-correct score by determining percentage correct separately for each of the four sub-blocks and computing Cronbach's α over these four independent indicators. The reliability of the span scores in each block was estimated by computing Cronbach's α on the two parallel span estimates from the two pairs of sub-blocks.

Latencies from the verbal and the spatial secondary tasks were discarded as outliers if they exceeded an individual's mean in the respective condition by three standard deviations. This resulted in a loss of 1.5% of the data. Latencies from the spatial tasks were also discarded when they were associated with a wrong decision. We first report results of the experimental manipulations of similarity, and then turn to the relationship of working memory capacity with dual-task costs. For all analyses, the α level was set to .05. The effect size is given as partial eta^2 (η_p^2), which reflects the proportion of variance accounted for by the effect in question, relative to the total of the effect variance and its error variance.

Similarity and dual-task interference

Dual-task costs can be estimated by comparing single-task and dual-task spans, or by comparing the percentage correct values computed over those list lengths that are shared by single and dual-task conditions (PC* in Table 2). For the verbal tasks, mean accuracy for the shared list lengths 3–6 was 81% in the single-task condition and 59% on average in the four dual-task conditions, a significant difference, $t(119) = 19.5$. For the spatial tasks, mean accuracy for list lengths 2–4 was 73% for the two single tasks and 60% for the four dual tasks, also a significant difference, $t(119) = 11.9$. As can be seen in Table 2, the dual-task effect was also apparent as substantial reductions in spans.

We investigated effects of the similarity manipulations, together with effects of list length, on dual-task costs, secondary task latency, and secondary task accuracy (for the spatial tasks). Dual-task costs (DTC) were computed separately for each similarity condition by subtracting the dual-task score from the corresponding single-task score, for both percentage correct (over shared list lengths) and span. The DTC computed from percentage correct are shown in Fig. 3. The statistical analyses revealed a number of higher-order interactions, which did not show any interpretable pattern and are most likely due to idiosyncratic features of the materials used in each design cell; for the sake of brevity we report only the theoretically interesting effects.

Verbal tasks

The percent-correct DTC from the verbal tasks were submitted to an ANOVA with phonological similarity (2), semantic similarity (2), and list length (4) as factors. Dual-task costs tended to be largest for the intermediate list lengths (4 and 5), $F(1, 119) = 43.72$, $\eta_p^2 = .269$ for the quadratic contrast. There was no effect of phonological similarity, $F < 1$, but there was a significant effect of semantic similarity, $F(1, 119) = 8.35$, $\eta_p^2 = .066$. As can be seen in Fig. 3, DTC were slightly larger in conditions with semantically similar secondary task words overall. Semantic similarity interacted with the linear

Table 2
Means, standard deviations, and reliabilities of single and dual-task performance measures

	Mean PC (*SD*)	Mean PC* (*SD*)	Reliability of PC	Mean span (*SD*)	Reliability of span
Verbal single	63 (10)	81 (11)	.81	4.42 (.84)	.76
Verbal phon+/sem+	65 (15)	59 (17)	.87	2.77 (.91)	.66
Verbal phon−/sem+	64 (16)	58 (17)	.85	2.83 (.94)	.64
Verbal phon+/sem−	66 (15)	60 (16)	.85	3.04 (.90)	.70
Verbal phon−/sem−	67 (14)	61 (16)	.84	3.0 (.91)	.75
Spatial single pattern	64 (17)	76 (17)	.82	2.83 (1.16)	.82
Spatial single lines	61 (20)	70 (18)	.78	2.39 (1.05)	.67
Spatial pattern similar	67 (18)	62 (20)	.79	1.92 (1.06)	.65
Spatial pattern dissimilar	66 (19)	61 (20)	.83	1.80 (.94)	.67
Spatial lines similar	63 (22)	59 (23)	.84	1.85 (1.04)	.77
Spatial lines dissimilar	63 (23)	58 (24)	.89	1.80 (1.14)	.79

PC, percent items recalled correctly, averaged over all list lengths administered; PC*, percent items recalled correctly, averaged over only the list lengths shared by single and dual-task conditions; phon+, phonologically similar; phon−, phonologically dissimilar; sem+, semantically similar; and sem−, semantically dissimilar. Reliabilities are Cronbach's α over 4 variables built from the 4 sub-blocks within each task for PC, and over the two parallel span measures, respectively (see text for details).

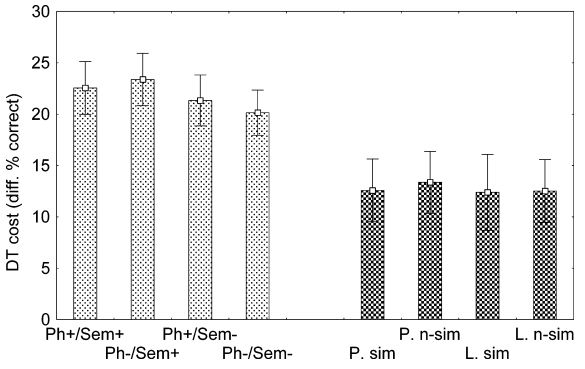

Fig. 3. Mean dual-task costs, calculated as difference in percent of items recalled correctly, for the similarity conditions in verbal tasks (left side) and spatial tasks (right side). Ph+, high phonological similarity; Ph−, low phonological similarity; Sem+, high semantic similarity; Sem−, low semantic similarity; P, patterns; and L, lines. Error bars represent two standard errors.

contrast of list length, $F(1,119) = 43.02$, $\eta_p^2 = .266$. Fig. 4A shows that the increase in DTC with high semantic similarity was confined to small list lengths and absent or even reversed for large list lengths.

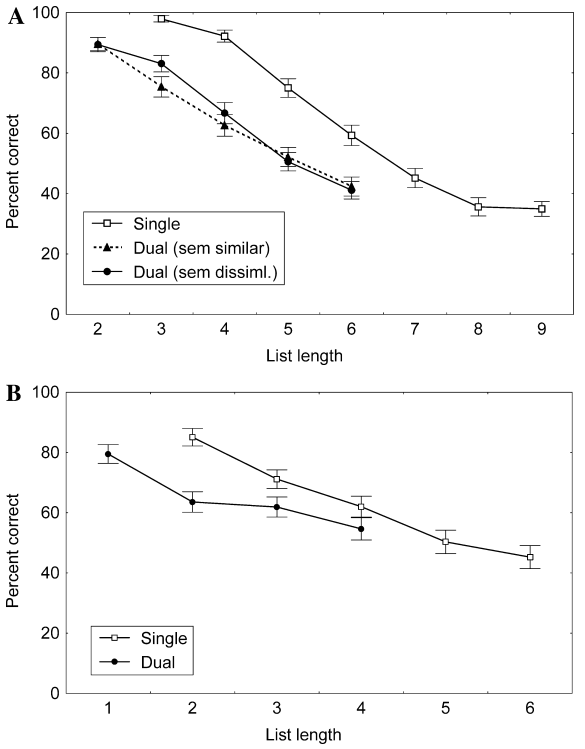

Fig. 4. Percent correct for single tasks and dual tasks by list length. (A) verbal tasks; (B) spatial tasks. Dual-task conditions are distinguished only where statistically reliable.

The analysis of span DTC by an ANOVA with semantic similarity (2) and phonological similarity (2) revealed a similar pattern: There was a main effect of semantic similarity, $F(1,119) = 15.13$, $\eta_p^2 = .113$, but no effect of phonological similarity ($F < 1$) and no interaction ($F = 1.1$).

The latencies of the secondary task (reading words aloud) were analyzed by an ANOVA with semantic similarity (2), phonological similarity (2), memory list length (4), and serial position in the secondary task list (3) as factors. There was an effect of serial position, $F(2,236) = 73.03$, $\eta_p^2 = .382$, reflecting the fact that participants took an average of 1267 ms until pressing the space bar after reading the first word, but only 688 and 697 ms for the next two words. Latencies increased slightly with the length of the memory list, $F(3,354) = 3.14$, $\eta_p^2 = .026$, but this effect was limited to the first of the three words read in each trial, as reflected by the interaction of list length and serial position, $F(6,708) = 5.11$, $\eta_p^2 = .048$. There was no main effect of semantic similarity, $F < 1$. The main effect of phonological similarity was significant, $F(1,119) = 5.84$, $\eta_p^2 = .047$. Words that were phonologically similar to one word in the memory list were read 60 ms slower than dissimilar words. This effect did not interact with memory list length ($F < 1$) and just marginally with the quadratic contrast of serial position, $F(1,119) = 4.08$, $p = .046$, $\eta_p^2 = .033$.

Spatial tasks

Spatial DTC obtained from percent correct were submitted to an ANOVA with material (patterns vs. lines), similarity (high vs. low), and list length (2–4) as factors. There was a trend for DTC to become smaller with increasing list length, $F(1,119) = 63.88$, $\eta_p^2 = .349$, for the linear trend. Fig. 4B shows that this is probably due to the singular low performance in the dual-task conditions at list-length two. The main effect of similarity was not significant, nor was the effect of material (both $F < 1$). An ANOVA on DTC computed from the spans with material (2) and similarity (2) revealed that dual-task costs were larger for patterns than for lines, $F(1,119) = 13.06$, $\eta_p^2 = .099$, but there was again no effect of similarity ($F = 1.37$).

Reaction times for the decisions in the secondary task were submitted to an ANOVA with material (2), similarity (2), memory list length (3), and serial position of reaction (3) as factors.[1] The theoretical interesting effects were: A significant effect of serial position, $F(2,222) = 235.71$, $\eta_p^2 = .68$, due to a much longer time for the first of the three decisions (1138 ms) than for the following two (793 and 762 ms, respectively); a significant main effect of

[1] Eight participants dropped out of this analysis because of empty design cells due to errors.

Table 3
Reliabilities and correlations of WMC tasks

	Reading span	MU numerical	MU spatial	Spatial STM
Reading span	.80	.49	.28	.19
MU numerical		.71	.58	.51
MU spatial			.70	.47
Spatial STM				.67

Note. Entries in the main diagonal are Cronbach's α, based on 11–15 trials as test items.

memory list length, indicating a *decrease* of reaction times with longer memory lists, $F(1, 111) = 63.99$, $\eta_p^2 = .366$ for the linear contrast; and a conspicuous absence of main effects of similarity and of material (both $F < 1$).

Overall accuracy in the secondary task was 94%. In an ANOVA analogous to that performed on the reaction times, there was no main effect of similarity, nor of material (both $F < 1$). Several other effects became significant but were small (hardly exceeding 2 percentage points) and non-systematic.

To summarize, there were substantial dual-task costs in both verbal and spatial tasks, although the effect size was clearly larger in case of the verbal tasks (see Fig. 3). Similarity, however, had little effect on the size of this interference. The only statistically reliable effects were a slight increase of dual-task costs with high semantic similarity, and a slight increase of word reading latencies with high phonological similarity.

Dual-task interference and working memory capacity

We tested whether working memory capacity is related to the amount of dual-task interference by correlating a measure of capacity with various measures of dual-task costs (DTC), explained below. The reliabilities of the four standard WMC tasks and their intercorrelations are displayed in Table 3. As a composite measure of WMC we used the unweighted average of *z*-scores obtained from the four standard working memory tasks. A composite score was used for most analyses in order to reduce the contribution of task-specific variance and emphasize the common variance among several indicators of working memory capacity (cf. Oberauer, in press).

The correlations between the composite WMC variable and performance in the experimental tasks (mean percent correct over all list lengths, and spans) are summarized in Table 4. These correlations show that WMC is strongly related to both single- and dual-task performance in the verbal domain, whereas the relationship was conspicuously low for the spatial domain. The latter observation is surprising since tasks very similar to the experimental spatial tasks used here were strongly correlated with other WMC measures in previous studies (Kane, Hambrick, Tuholski, Wilhelm, Payne, & Engle, in press; Oberauer et al., 2003).[2] One potentially relevant difference to these previous studies is that here we deliberately chose relatively simple spatial configurations as items. This could have made it easier to encode each item as a single chunk, instead of having to integrate its pieces actively (see Oberauer et al., 2003, for evidence that the integration of elements into structures is an important aspect of working memory capacity).

There are several different ways to calculate DTC, each of which follows its own rationale, consisting of assumptions about how an underlying psychological variable translates into performance measured in a particular metric. The underlying variable could be the amount of available activation or strength of an item's memory representation or any other theoretical construct held responsible for better or worse memory performance—we will refer to it summarily as *resource* here (without commitment to assumptions from resource theories), because the assumptions about the relationship between this variable and observable performance have become known as *performance–resource functions* (Norman & Bobrow, 1975).

We will show that the correlation of DTC with an external variable such as WMC depends on the index chosen to represent DTC. In the literature on dual-task interference usually only one of these DTC indices is used, and the underlying rationale is often left implicit. In order to give a more complete picture of the data, and to provide a direct comparison of the various possible ways to obtain individual estimates of DTC, we computed different DTC indices and correlated them with the composite WMC score. For these analyses, we eliminated data from one participant who was an outlier on most DTC indices for the verbal tasks.[3]

[2] The low correlations found here do not reflect a dissociation of verbal and spatial working memory—if anything, the correlation of the spatial experimental task was higher with the two verbal-numerical tasks among the standard WMC measures (reading span and numerical memory updating) than with the two spatial standard tasks.

[3] This participant had 0% correct on two out of four subblocks of the verbal single task and therefore received negative DTC values, which deviated from the means by more than 3 SD for most DTC indices.

Table 4
Correlations between percent correct and span measures from experimental tasks and working memory capacity

	WMC	Single-PC (V)	Dual-PC (V)	Single span (V)	Dual span (V)	Single PC (S)	Dual PC (S)	Single span (S)
Single-PC (V)	.50							
Dual-PC (V)	.49	.71						
Single span (V)	.52	.77	.66					
Dual span (V)	.55	.65	.87	.74				
Single PC (S)	.22	.22	.27	.04	.10			
Dual PC (S)	.26	.27	.30	.08	.13	.83		
Single span (S)	.09	.08	.22	0	.07	.75	.67	
Dual span (S)	.20	.22	.24	.07	.12	.84	.89	.70

Note. $N = 119$.

The most straightforward way to compute DTC is to subtract the performance in single tasks from that in comparable dual tasks. We did this for percentage correct (*PC difference*) and for span (*span difference*). The implicit rationale of this procedure is that equal amounts of performance decrements reflect equal amounts of resource reductions. This assumption seems quite reasonable for a span score if one assumes that span reflects the number of free "slots" in a working memory that is limited to a specific number of chunks to be held (e.g., Cowan, 2001), and the secondary task takes away a certain number of "slots." We see no obvious rationale, however, for using the difference measure with percentage correct. Such an index would imply that equal decrements in percentage correct reflect equal reductions in an underlying resource for performance levels close to the floor of the measurement scale (e.g., a reduction from 30 to 10% correct), intermediate (from 60 to 40%), or close to ceiling (from 90 to 70%). This is at least a questionable assumption.

Another popular measure of DTC is the proportional drop in performance under dual-task conditions relative to single task performance. The rationale of this measure was expressed by Myerson et al. (2000) by the metaphor of stocks and losses: Those who have more have more to lose. If interference increases the likelihood to lose one item from memory by a constant amount, then people who can hold a larger number of items in memory under single-task conditions (i.e., people with a larger single-task span) will lose a larger number of items from interference. The reduction in span can be expected to be proportional to the single-task span. To capture this rationale, we computed the *proportional decrease in span* index.

Proportional loss is also often used to measure dual-task on a percentage correct scale (*proportional decrease in PC*). With this measure of DTC, a decrease close to ceiling (from 90 to 70% = .22) would be regarded as smaller than the same absolute decrease lower on the scale (e.g., from 60 to 40% = .33, or from 30 to 10% = .67). In other words, an equivalent reduction in resources would be expected to result in larger losses on the percent correct scale the closer the overall level is to ceiling. Contrary to this assumption, one could argue that when performance is close to 100% in the single-task condition, it might drop relatively little in the dual-task condition because of "spare resources": Even a substantial reduction in resources due to the dual-task load could leave enough to maintain performance on a high level, resulting in a smaller loss than when the same dual-task interference hit a performance level far from ceiling. The logic of "spare resources" underlies the common hypothesis that capacity and task complexity should interact, such that individuals or groups with lower capacity should be particularly vulnerable to an increase in task complexity or a secondary task load (e.g., King & Just, 1991; Salthouse, 1992).

The rationale of the capacity × complexity interaction on percentage correct is more compatible with a measure of DTC in terms of the *proportional increase of errors*. This index is the mirror image of the one based on the proportional decrease of percentage correct. Applying the proportional increase of errors measure to the examples from above shows that it does the opposite of what the proportional decrease in PC index does: It inflates DTC close to ceiling (from 10 to 30% errors = 2.0) and shrinks those close to the floor (from 70 to 90% errors = .29).

A rationale that takes both floor and ceiling effects of the percent-correct scale into account can be taken from probabilistic test theory (Rasch, 1980). The basic assumption is that a continuous theoretical variable such as a resource is translated into percent correct by a sigmoid (e.g., logistic) function. One way to capture this logic in a measure of dual-task costs is to submit the percent correct variables to a probit transformation, which translates the probability of a correct answer into the corresponding z-score of a standard normal distribution. We calculated a measure called *Probit PC difference* as an indicator of DTC that represents the logic of probabilistic test theory. The Probit PC difference treats relatively large absolute differences between

single- and dual-task performance in the middle of the percent-correct scale as equivalent to smaller differences near both ends of the scale. Applied to our examples, a drop from 90% correct (1.28) to 70% (.52) would be equivalent to one from 30% (−.52) to 10% (−1.28), but larger than one from 60% (.25) to 40% (−.25).

From an individual-differences perspective, the most obvious method to isolate the variance due to secondary-task interference is the computation of residuals from regression equations. One rationale for doing so would be to assume that performance variance in the dual-task condition is a mixture of variance under single-task conditions and variance due to dual-task interference. To isolate the latter, one would have to predict dual-task performance from single-task performance and use the residual as an estimate of an individual's susceptibility to dual-task interference, or DTC. This method (applied to percent correct as performance measure) yielded what we called the *residual dual-task* index. Note that whereas all other indices discussed so far express higher DTC as higher values, a high residual dual-task score reflects dual-task performance higher than expected from a person's single-task performance, and hence relatively small DTC.

Another rationale could be to assume that single-task variance is a mixture of variance in dual-task performance and some specific source of variance for the single-task condition. Although counterintuitive at first sight, this assumption is defensible when it comes to working-memory tasks combining storage and processing. It has been argued since the seminal work of Daneman and Carpenter (1980) that dual-tasks combining storage and processing are purer measures of WMC than single-task measures of serial order memory (e.g., digit span or word span). One reason for this could be that single-task performance is a composite of WMC and the efficiency of an additional mechanism (such as the phonological loop) specialized to the maintenance of items in forward serial order. If this is true, the difference between single-task and dual-task performance in serial-order memory would reflect the additional benefit due to the specialized system, and one way to obtain a pure measure of its efficiency is to predict single-task performance from dual-task performance and take the residual as a measure of single-task benefit. This yields our *residual single task* index.

The DTC were computed individually for each similarity condition within the verbal and the spatial domain and then aggregated over similarity conditions because similarity hardly affected dual-task costs. Table 5 summarizes means and standard deviations of the various indices of DTC (or single-task benefits), together with estimates of reliabilities. The reliability estimates are Cronbach's α computed over four independent DTC-variables formed from the four sub-blocks within each block (for indices derived from percentage correct) or the two parallel span estimates (for indices derived from span scores). Also shown in Table 5 are the correlations of the DTC indices with the WMC scores. The DTC indices varied considerably in their reliability. Their relationships with WMC were generally low, and they differed systematically in direction depending on the index chosen. For example, the proportional decrease in PC, which inflates DTC in the low-performance range relative to an absolute difference and shrinks DTC in the high-performance range, correlated negatively with WMC at least for the verbal domain. The proportional increase in errors, in contrast, which inflates DTC in the high-performance range and shrinks them in the low-performance range, shows a positive

Table 5
Reliabilities and correlations with working memory capacity for various measures of dual-task costs

	Verbal			Spatial		
	Mean (*SD*)	Reliability	Correlation with WMC	Mean (*SD*)	Reliability	Correlation with WMC
PC difference	.23 (.11)	.74	−.16	.13 (.11)	.65	.01
Proportional decrease in PC	.28 (.15)	.81	−.26*	.13 (.20)	.57	−.02
Proportional increase in errors	1.52 (.92)	.42	.18*	.71 (.59)	.27	.22*
Probit (PC) difference	.77 (.37)	.57	.05	.42 (.43)	.60	.02
Span difference	1.51 (.58)	.34	.03	.76 (.63)	.51	−.16
Proportional decrease in span	.33 (.12)	.45	−.20*	.25 (.25)	.50	−.22*
Residual dual task	0 (.11)	.86	.32*	0 (.12)	.77	.10
Residual single task	0 (.07)	.58	.33*	0 (.09)	.64	.21*

Reliabilities are Cronbach's α over 4 variables built from the 4 sub-blocks within each task, except for spans, where they represent Cronbach's α over the two parallel span measures (see text for details). $N = 119$ for verbal tasks and 120 for spatial tasks. Correlations with * are significant.

Table 6
Regressions of dual-task cost indicators on four measures of working-memory capacity

	Verbal					Spatial				
	R^2	RS	MUN	MUS	SSTM	R^2	RS	MUN	MUS	SSTM
PC difference	.061	−.25	.05	.03	−.09	.057	−.21	.05	.01	.15
Proportional decrease in PC	.108*	−.29	−.03	−.01	−.07	.06	−.20	.12	−.13	.16
Proportional increase in errors	.06	−.01	.19	.15	−.12	.112*	−.04	−.13	.34	.09
Probit (PC) difference	.019	−.06	.12	.08	−.12	.06	−.20	.07	−.05	.18
Span difference	.004	.01	.04	.01	−.06	.062	−.24	.06	−.09	.02
Proportional decrease in span	.052	−.13	−.08	0	−.09	.099*	−.29	.10	−.16	.03
Residual dual task	.15*	.33	.03	.04	.08	.057	.20	.06	.02	−.13
Residual single task	.136*	.05	.33	.04	−.02	.091*	−.15	.23	.02	.14

Entries are unadjusted R^2 of the regression equation (* indicates a significant F ratio), followed by beta weights for reading span (RS), memory updating numerical (MUN), memory updating spatial (MUS), and spatial short-term memory (SSTM).

correlation to WMC. The PC difference index showed no reliable correlation with WMC. Thus, it seems that most of the correlations between DTC and WMC in Table 5 are artifacts from non-linear transformations on the differences between single- and dual-task performance.

How much variance in the DTC variables can at best be accounted for by independent measures of WMC? To explore this, we computed multiple regressions using the four standard WMC measures as predictors for each DTC index. The results are summarized in Table 6. This analysis allows task-specific variance of the standard WMC tasks, in addition to their shared variance, to contribute to the association with DTC variables. Nonetheless, this association was still found to be weak and unsystematic. The distribution of β-weights also shows that there was no consistent pattern of verbal WMC tasks predicting verbal DTC and spatial WMC tasks predicting spatial DTC.

The various measures of DTC are not equivalent, and some of them are not even strongly related. The correlations between different DTC measures calculated from the same single-task and dual-task data are displayed in the upper right triangle of Table 7 for verbal tasks and in the lower left triangle for spatial tasks. The entries in the main diagonal represent correlations between corresponding DTC indices across the two domains. These correlations were negligible throughout, defying the notion of a domain-general construct of resistance to dual-task interference or of dual-task coordination.

Another way to test for individual or group differences in the amount of interference is to equate individuals on single-task performance and look for differences in dual-task performance under the same conditions. We emulated this procedure by selecting for each participant within each domain the list length at which she or he came closest to a criterion. The criteria that allowed the best approximation to equal single-task performance were 85% correct for verbal and 70% correct for spatial tasks. Next we obtained for each participant the performance from the corresponding dual

Table 7
Correlations between various measures of dual-task costs

	Diff. PC	Prop. PC	Prop. error	Diff Probit	Diff span	Prop. span	Resid. single	Resid. dual
Diff. PC	.09	.97	.64	.86	.53	.66	.55	−.91
Prop. PC	.94	.04	.49	.75	.44	.66	.37	−.95
Prop. error	.66	.54	.06	.76	.47	.35	.74	−.45
Diff. probit	.96	.88	.60	.09	.69	.60	.78	−.67
Diff. span	.50	.48	.13	.56	−.03	.87	.61	−.43
Prop. span	.70	.71	.30	.68	.87	.11	.37	−.68
Resid. single	.71	.66	.50	.75	.50	.39	.02	−.34
Resid. dual	−.93	−.87	−.60	−.86	−.39	−.71	−.40	.20

Note. The upper right triangle contains correlations within the verbal tasks, the lower left triangle contains correlations within the spatial tasks, and the main diagonal contains correlations between corresponding cost measures from verbal and spatial tasks. Correlations >.19 are significant with $\alpha = .05$. $N = 119$ for verbal and 120 for spatial tasks.

task at the same list length. Participants were split at the median of the WMC score into a group with high and a group with low capacity. We conducted two ANOVAs (one for the verbal, one for the spatial tasks) with capacity group (high vs. low) and task condition (single vs. dual) as factors and performance on the selected list length as dependent variable. In both domains the critical interaction went in the expected direction (i.e., the groups did not differ in single-task performance, but the low capacity group showed worse dual-task performance), but failed the level of statistical significance, $F = 2.0$ for the verbal tasks and 1.6 for the spatial tasks.

Discussion

Working memory capacity and resistance to interference

The first goal of this study was to investigate relationship between working memory capacity and individual's ability to resist interference, as measured by the amount of dual-task costs in complex span tasks. Although we were able to measure individual participants' dual-task costs with reasonable reliability, they were only weakly if at all correlated with an independent measure of working memory capacity. Therefore, our data provide little support for the hypothesis that working memory capacity reflects the ability to resist interference in dual-task combinations of storage and processing (Engle et al., 1999; Oberauer & Kliegl, 2001). Our data also provide little support, however, for the contention of Jenkins et al. (1999) that dual-task interference *increases* with working memory capacity. Contrary to the "stocks and losses" metaphor endorsed by Myerson et al. (2000), dual-task costs were not a constant proportion of single-task span—in that case, there should have been a positive correlation between WMC and the absolute *span difference*, whereas the *proportional decrease in span* should have been uncorrelated with capacity. The opposite pattern was observed here: The *span difference*, as well as the *PC difference* indices, were largely uncorrelated with WMC, whereas the proportional decrease indices tended to be negatively correlated with WMC. This indicates that dual-task interference had an additive (or better: subtractive) effect on performance, whether measured as span or percentage correct, independent of a person's working memory capacity.

Similarity between tasks

Our second goal was to investigate whether dual-task interference in complex span tasks depends on the similarity between memory materials and stimuli to be used in the processing task. Similarity was manipulated in three different ways across two content domains. Overall, it was found to have little effect. The two exceptions were: (1) a small impact of semantic similarity on verbal memory performance, which was not observed consistently for all list lengths and therefore might be spurious, and (2) a small, but consistent effect of phonological similarity on reading times. Could a lack of power be responsible for our failure to detect effects of phonological similarity on recall, and of spatial similarity on all dependent variables? We think this is highly unlikely, given that our sample size was unusually large for a within-subjects experiment, our measures were reasonably reliable, and even small effects (such as those of semantic similarity) were detected.

The lack of more substantial similarity effects is surprising in light of previous studies that found larger interference with serial order memory from distractor tasks with similar, compared to less similar material (Lange & Oberauer, in press; Li, 1999; Turner & Engle, 1989; Wickelgren, 1965, 1966). On the other hand, Glanzer, Koppenaal, and Nelson (1972) found no effect of similarity on the reduction of recency in a free recall task by a distractor in the retention interval. The interfering effect of irrelevant sounds on serial order memory also are independent of the similarity between the memory items and the interfering material (for a review see Jones & Tremblay, 2000). We feel that our experimental tasks were more similar to those that did show an effect of similarity, since they involved memory for serial order, and the distracting material had to be processed in the retention interval (Lange & Oberauer, in press; Wickelgren, 1965) or between successive items (Li, 1999; Turner & Engle, 1989). In contrast, those studies that failed to find an effect of similarity differed from our paradigm in that they either used free recall (Glanzer et al., 1972) or displayed distracting material that was to be ignored (i.e., irrelevant sounds). The existing data therefore reveal no systematic pattern as to when interference between short-term retention and a processing task is increased by similarity between the two tasks' materials. Apparently, the many different ways to operationalize similarity in working-memory tasks have different effects. Therefore, our data cannot be generalized as implying that similarity-based interference never happens in working memory. Instead, they point to the need for theories of interference to indicate more precisely under which conditions similarity affects working-memory performance.

Domain-generality of dual-task costs

The third goal was to test whether dual-task costs are domain specific or general. For none of our DTC indices was there a significant correlation across domains. Thus, our data provide no support for the assumption that dual-task costs reflect a general executive attention (Engle et al., 1999) or the efficiency of a domain-general

central executive (Baddeley, 1996). This result might be limited to the particular kind of dual-task cost investigated here, that is the impairment in immediate memory by a secondary processing task from the same content domain. It could well be that dual-task combinations of tasks from different domains provide a better chance to tap a general executive function, because this function would be required only to coordinate processes from different domain-specific subsystems. There is substantial evidence that such cross-domain dual-task costs reflect a cognitive function that is specifically impaired in patients with Alzheimer's disease (Baddeley & Della Sala, 1996; Logie, Cocchini, Della Sala, & Baddeley, in press). We were interested in the particular kind of dual-task costs we measured because complex span tasks have been shown to be better predictors of complex cognition than simple span tasks, and we wanted to isolate the component of variance that sets them apart. Our results strongly suggest that the difference between complex and simple span tasks cannot be interpreted as measuring the added contribution of a general executive device. Therefore, the unique predictive power of complex span tasks is not easily attributed to general executive attention.

A potential objection against our dual-task cost variables is that they might be regarded incomplete. Research on dual-task performance usually measures DTC on both tasks. In our research, we had no measure of the processing component as a single task, and hence no estimate of dual-task cost on processing efficiency. In this regard, we followed the common practice of researchers using complex span tasks to assess working-memory capacity (Daneman & Carpenter, 1980). Confining DTC to the memory component receives justification from the finding that DTC for the memory component and DTC for the processing component of complex span tasks are uncorrelated (Oberauer et al., 2003), suggesting that they do not reflect the same construct.

Further support for keeping DTC costs from the component tasks separately comes from the finding in the present study that there is little trade-off in efficiency between memory and processing in complex span tasks. Contrary to what one should expect from the assumption of a common resource shared between retention of material in working memory and processing of other material (Anderson, Reder, & Lebiere, 1996; Just & Carpenter, 1992), there was hardly any effect of the memory demand (i.e., list length) on speed and accuracy of the processing task. For the verbal tasks, only the first of three words was read slower when the memory load was larger. For the spatial tasks, processing speed even increased with larger memory load, and accuracy remained constantly high throughout all list lengths. This is consistent with previous results (e.g., Oberauer, 2002) showing that memory list length does not affect the speed of concurrent processing operations as long as they do not require access to information in working memory, which was not the case here.

How to measure dual-task costs?

A fourth, methodological goal regarded the evaluation of various kinds of dual-task cost indicators. In our study, some DTC indicators had respectable reliability, whereas others did not. This does not mean that in other studies the same kinds of DTC will turn out to be reliable or less reliable, respectively. Rather, we wish to urge researchers to estimate the reliability of their DTC variables, and interpret correlations or group differences in its light. Another important result is that different DTC do not measure the same. Some of them are only moderately correlated. Moreover, the choice of a DTC can bias its correlation with other variables in systematic ways. In particular, DTC indices based on proportional loss of performance tend to inflate costs for people with overall low performance and shrink costs for those with high overall performance—a bias that is favorable to the hypothesis that high ability is associated with small dual-task costs. Conversely, DTC indices calculated from proportional increase of errors or deviation from a standard induce the opposite bias, thereby working against the above hypothesis. The effects of these biases can clearly be observed in Table 5. We want to make clear that our analysis is not intended to be a recipe for picking the DTC that serves one's hypothesis best. Each DTC is associated with assumptions providing its rationale, and future research should devote more attention to investigating which of these assumptions are warranted for particular kinds of tasks. Our personal preference is to use the probit PC difference index, because it is neutral with regard to the biases discussed above, and is based on a plausible rationale.

One further observation in the present data is relevant for the measurement of working memory capacity. The evaluation of the same performance data in terms of percentage of correctly recalled items and in terms of an emulated span measure allows a direct comparison of the two ways of scoring. Table 2 shows that for every single task block (with the exception of one tie), the reliability estimate for the percentage correct score was higher than that of the span scores. The reason for the lower reliability of the span scores is probably that the span procedure throws away information: First, each trial is only coded as passed or failed, disregarding the degree of failure to recall a list. Second, performance on list lengths higher than the first length at which a participant fails both trials is discounted. One must bear in mind, however, that our comparison is between two ways of scoring the same performance obtained with the same testing procedure, whereas

usually spans are obtained by a different testing procedure (i.e., adaptively increasing list lengths), which could result in better (or in worse) reliability than our pseudo-span.

Conclusions

The main conclusions from the present results are largely negative. Our data provide little support for the hypothesis of similarity-based interference between storage and processing in working memory. They are also difficult to reconcile with one idea suggesting similarity-independent interference, that is, the hypothesis that memory and concurrent processing must share a limited resource. The hypothesis that dual-task costs reflects a general capacity to resist interference was questioned by the lack of correlations between dual-task costs in verbal and spatial tasks, and by the small correlations of these costs with an independent assessment of working memory capacity. Thus, several promising current ideas about the nature of so-called complex span tasks might have to be rethought.

One way to downplay the consequences of our results for the hypotheses discussed above would be to argue that our results, for some reason, are untypical. Contrary to previous studies (e.g., Oberauer et al., 2003), dual tasks in our study did not correlate more with the WMC score than the corresponding single tasks (see Table 4).[4] Another deviation from previous data is the fact that the spatial experimental tasks were only weakly correlated with WMC. Finally, as noted above, the failure to find an effect of similarity on the amount of interference could also be regarded as an anomaly. One could conclude that we were not successful to generate *similarity-based* interference with our dual-tasks, and hence did not obtain a sufficient amount of the variance that, in other studies, was responsible for the high correlation between complex span tasks across domains, and their high correlation with complex cognition. This deflationary interpretation is not easily dismissed, but it leaves open one important question: Given that there were substantial and reliable dual-task costs in our tasks, what, if not the susceptibility to interference, or the efficiency of task coordination by a central executive, do they reflect? Our data should at least motivate proponents of the interference account of working memory capacity (including ourselves) and proponents of the central executive account to specify more precisely under which conditions the amount of dual-task interference should reflect working memory capacity (or the capacity of a central executive).

A change of perspective that might be helpful to understand our results is to see the performance difference between single- and dual-task versions not as dual-task costs, but as single-task benefits. This makes sense if we see dual-task performance as a relatively pure manifestation of working memory capacity, whereas single-task performance would reflect a mixture of working-memory capacity and the recruitment of additional mechanisms or strategies. Such a view would be suggested by the theory of Cowan (2001), who assumed that working memory has the capacity to hold about 4 chunks, and forward serial recall exceeds the "magical number four" because of special strategies (e.g., ad hoc chunking of items). Adding a secondary task could be seen as interfering with these strategies, resulting in a purer measure of working memory capacity. Likewise, starting from the theory of Baddeley (1986) one could assume that single-task serial recall relies to a large degree on "slave systems" such as the phonological loop and the visuo-spatial sketch pad, which might have little relationship to each other and to complex cognition in general. Adding a secondary task that uses representations of the same domain would largely disable the slave systems, such that dual-task performance reflects mainly the capacity of more central mechanisms such as the episodic buffer (Baddeley, 2000). Individual differences in dual-task costs would then reflect quantitative or qualitative differences between people in their ability (or willingness) to use efficient specialized strategies for serial recall (in Cowan's framework), or differences in the efficiency of the slave systems utilized for serial recall in the single task condition (in Baddeley's framework). Under such a view, it would not be surprising that the single-task benefits added by specialized sub-systems, or by task-specific strategies, are not correlated across domains, and are not particularly strongly related to independent measures of working memory.

Hence, we might regard "simple" span tasks as the more complex measures, in that they reflect more of a mixture of different sources of variance. Whereas the starting point of our work was the equation "complex span = simple span + controlled attention" (Engle et al., 1999), it might be more fruitful to turn things around: simple span = complex span + specialized mechanisms or strategies.

[4] Since directly comparable results were not reported in Oberauer et al. (2003), we reanalyzed their data for correlations of the four tasks most similar to our experimental tasks with a composite of the four standard WMC tasks used here. These correlations were .48 for *single word span*, .68 for *dual word span*, .46 for *single pattern span*, and .53 for *dual pattern span*.

References

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limits on retrieval. *Cognitive Psychology, 30*, 221–256.

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

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

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Science, 4*, 417–423.

Baddeley, A. D., & Della Sala, S. (1996). Working memory and executive control. *Philosophical Transactions of the Royal Society of London, 351*, 1397–1404.

Baddeley, A. D., & Ecob, J. R. (1970). Simultaneous acoustic and semantic coding in short-term memory. *Nature, 227*, 288–289.

Bayliss, D. M., Jarrold, C., Gunn, D. M., & Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General, 132*, 71–92.

Cantor, J., Engle, R. W., & Hamilton, G. (1991). Short-term memory, working memory, and verbal abilities: How do they relate? *Intelligence, 15*, 229–246.

Case, R., Kurland, M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*, 386–404.

Cocchini, G., Logie, R. H., Della Sala, S., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition, 30*, 1086–1095.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence, 30*, 163–183.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–185.

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

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General, 128*, 309–331.

Fletcher, P. C., & Henson, R. N. A. (2001). Frontal lobes and human memory. Insights from functional neuroimaging. *Brain, 124*, 849–881.

Glanzer, M., Koppenaal, L., & Nelson, R. (1972). Effects of relations between words on short-term storage and long-term storage. *Journal of Verbal Learning and Verbal Behavior, 11*, 403–416.

Jenkins, L., Myerson, J., Hale, S., & Fry, A. (1999). Individual and developmental differences in working memory across the life span. *Psychonomic Bulletin & Review, 6*, 28–40.

Jones, D. M., & Tremblay, S. (2000). Interference in memory by process or content? A reply to Neath (2000). *Psychonomic Bulletin & Review, 7*, 550–558.

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

Kail, R., & Hall, L. K. (2001). Distinguishing short-term memory from working memory. *Memory & Cognition, 29*, 1–9.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. E. (in press). The generality of working-memory capacity: A latent-variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General*.

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

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?!. *Intelligence, 14*, 389–433.

Lange, E., & Oberauer, K. (in press). Overwriting of phonemic features in serial recall. *Memory*.

Li, K. Z. H. (1999). Selection from working memory: On the relationship between processing and storage components. *Aging, Neuropsychology, and Cognition, 6*, 99–116.

Logie, R., Cocchini, G., Della Sala, S., & Baddeley, A. D. (in press). Is there a specific executive capacity for dual task coordination? Evidence from Alzheimer's disease. *Neuropsychology*.

Myerson, J., Jenkins, L., Hale, S., & Sliwinski, M. (2000). Stocks and losses, items and interference: A reply to Oberauer and Süß (2000). *Psychonomic Bulletin & Review, 7*, 734–740.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18*, 251–269.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7*, 44–64.

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 411–421.

Oberauer, K. (in press). The measurement of working memory capacity. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence*. Thousand Oaks: Sage.

Oberauer, K., & Kliegl, R. (2001). Beyond resources: Formal models of complexity effects and age differences in working memory. *European Journal of Cognitive Psychology, 13*, 187–215.

Oberauer, K., & Süß, H. M. (2000). Working memory and interference. A comment on Jenkins, Myerson, Hale, and Fry (1999). *Psychonomic Bulletin & Review, 7*, 727–733.

Oberauer, K., Süß, H. M., Schulze, R., Wilhelm, O., & Wittmann, W. W. (2000). Working memory capacity—facets of a cognitive ability construct. *Personality and Individual Differences, 29*, 1017–1045.

Oberauer, K., Süß, H. M., Wilhelm, O., & Wittmann, W. W. (2003). The multiple faces of working memory—storage, processing, supervision, and coordination. *Intelligence, 31*, 167–193.

Postle, B. R., Berger, J. S., & D'Esposito, M. (1999). Functional neuroanatomical double dissociation of mnemonic and executive control processes contributing to working memory performance. *Proceedings of the National Academy of Science, 96*, 12959–12964.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests (expanded edition)*. Chicago: University of Chicago Press.

Saint-Aubin, J., & Poirier, M. (1999). Semantic similarity and immediate serial recall: Is there a detrimental effect on order information? *Quarterly Journal of Experimental Psychology A, 52*, 367–394.

Salthouse, T. A. (1992). Why do adult age differences increase with task complexity? *Developmental Psychology, 28*, 905–918.

Smith, E. E., Geva, A., Jonides, J., Miller, A., Reuter-Lorenz, P., & Koeppe, R. A. (2001). The neural basis of task-switching in working memory: Effects of performance and aging. *Proceedings of the National Academy of Science, 98*, 2095–2100.

Smith, E. E., & Jonides, J. (1999). Storage and executive processes in the frontal lobes. *Science, 283*, 1657–1661.

Süß, H. M., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working memory capacity explains reasoning ability—and a little bit more. *Intelligence, 30*, 261–288.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127–154.

Wickelgren, W. A. (1965). Acoustic similarity and retroactive interference in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 4*, 53–61.

Wickelgren, W. A. (1966). Phonemic similarity and interference in short-term memory for single letters. *Journal of Experimental Psychology, 71*, 396–404.