Domain-specific processing in short-term serial order memory

[RUNNING HEAD: Domain specificity in serial order memory]

Alexander Soemer[1], Satoru Saito[2]

[1]University of Potsdam

Department of Psychology

Potsdam, Germany

[2]Kyoto University

Graduate School of Education

Kyoto, Japan

Correspondence concerning this article should be addressed to:

Alexander Soemer

University of Potsdam

Department of Psychology

Karl-Liebknecht-Str. 24-25

14476 Potsdam, Germany

alexander@soemurai.de

Abstract

Recent studies of short-term serial order memory have suggested that the maintenance of order information does not involve domain-specific processes. We carried out two dual-task experiments aimed at resolving several ambiguities in those studies. In our experiments, encoding and response of one serial reconstruction task was embedded within encoding and response of a concurrent serial reconstruction task. Order demands in both tasks were independently varied so as to find revealing patterns of interference between the two tasks. In Experiment 1, participants were to maintain and reconstruct the order of a list of verbal materials, while maintaining a list of spatial materials or vice-versa. Increasing the order demands in the outer reconstruction task resulted in small or non-reliable performance decrements in the embedded reconstruction task. Experiment 2 sought to compare these results against two same-domain baseline conditions (two verbal lists or two spatial lists). In all conditions, increasing order demands in the outer task resulted in small or non-reliable performance decrements in the embedded task. However, performance in the embedded tasks was generally lower in the same-domain baseline conditions than in the cross-domain conditions. We argue that the main effect of domain in Experiment 2 indicates the contribution of domain-specific processes to short-term serial order maintenance. In addition, we interpret the failure to find consistent cross-list interference irrespective of domain as indicating the involvement of grouping mechanisms in concurrently performed serial order tasks.

Keywords: Serial order memory, verbal memory, spatial memory, domain specificity, working memory

Introduction

The ability to remember and reproduce the serial order of recent events is one of the fundamental activities underlying cognition in a variety of domains. For example, order information distinguishes words that are made up of the same combination of phonemes (e.g., *dog* vs. *god*) or different routes composed of certain landmarks within the same spatial environment. In both cases, elements of the sequence are drawn from a limited pool (phonemes or landmarks), while the to-be-retained sequence itself is new (a new word or a new route).

There has been a considerable debate about the nature of the cognitive processes involved in maintaining such verbal or spatial sequence information in short-term memory. One of the most contested issues in this area concerns the existence of processes that are functionally specialized for the type of to-be-ordered material (Depoorter & Vandierendonck, 2009; Jones et al., 1995; Logie et al., 2012; Meiser & Klauer, 1999; Morey & Mall, 2012; Saito et al., 2008; Vandierendonck, 2015). The present article reports the results of two dual-task studies investigating this issue with a concurrent serial reconstruction paradigm.

Functional Specialization in Serial Order Memory

Much early work on short-term serial order memory has been framed within the multiple-component working memory model of Baddeley and Hitch (1974). This model assumes that there are separate domain-specific stores and maintenance mechanisms for verbal and visuospatial materials. These verbal and visuospatial components are supervised by a domain-general central executive or attentional control system that coordinates and integrates information from the components. The dissociation between verbal and visuospatial components has been supported by a wealth of evidence from experiments that require the concurrent performance of short-term memory and rehearsal suppression tasks. In

a typical demonstration of this dissociation, verbal rehearsal suppression interferes more with verbal than with visuospatial serial recall, while visuospatial rehearsal suppression interferes more with visuospatial than with verbal serial recall.

A recent debate among working memory theorists concerns the question of whether such functional specialization translates to the short-term maintenance of order information. Supportive evidence for a functional specialization comes from a study by Saito et al. (2008) in which Japanese participants performed immediate written serial recall of Japanese Kanji characters. In one experiment, participants carried out a verbal rehearsal suppression task while retaining lists of phonologically and visually similar or dissimilar Kanji characters. This suppression task removed the phonological similarity effect, which is a typical result for verbal serial recall, while leaving the visual similarity effect intact. Saito et al. (2008) argued that this independence of similarity effects indicates the involvement of domain-specific verbal and visuospatial contributions to order retention (also see Logie et al., 2012).

There are, however, important arguments supporting an alternative domain-general view of order maintenance. One of these arguments concerns the similarity of behavioral phenomena across domains (Hurlstone, Baddeley, & Hitch, 2014). Most notably, primacy and recency effects, which are typically observed in the immediate recall and reconstruction of verbal lists, are also found with a variety of non-verbal materials (Avons, 1999; Guérard & Tremblay, 2008; Jones et al. 1995; Parmentier & Jones, 2000; Smyth & Scholey, 1996). Other characteristic phenomena that have been replicated across domains include the recall advantage for aurally presented lists over visually presented lists (Penney, 1989; Tremblay et al., 2006), grouping effects (Henson, 1998; Parmentier & Maybery, 2008), and transposition gradients (Henson, 1998; Parmentier & Jones, 2000).

One parsimonious interpretation of such cross-domain similarities is that they originate from the very same cognitive mechanisms. Experimental evidence in favor of this view

comes from a study of Jones et al. (1995) in which groups of participants carried out serial reconstruction with either letters or dot locations while concurrently performing verbal or spatial rehearsal suppression. Jones et al. (1995) demonstrated that changing-state versions of the suppression tasks (articulation of letter sequences and tapping of key sequences) disrupted both letter and dot reconstruction. Importantly, the detrimental effects of changing-state rehearsal suppression seemed to be of comparable size for both verbal and spatial reconstruction. On the basis of these results, Jones et al. (1995) argued for a model of short-term serial order memory in which order is retained in lists containing items from various domains. Memory loss for order information in these lists occurs whenever secondary tasks contain order cues that conflict with the order cues from the primary tasks, just like changing-state rehearsal suppression does. Importantly, this happens irrespective of the domain of the items.

This view was subsequently questioned by Meiser and Klauer (1999) who demonstrated that the domain-general effects of changing-state rehearsal suppression are restricted to the encoding phase. Referring to the multiple component model of working memory (Baddeley & Hitch, 1974), Meiser and Klauer argued that changing-state effects disrupt the central executive rather than the verbal and visuospatial components. In addition, in their experiments steady-state rehearsal suppression (repeating the same letter or the same key throughout) carried out during the retention interval seemed to be more domain-specific: Verbal serial reconstruction was reliably disrupted by verbal but not by spatial rehearsal suppression, while spatial serial reconstruction was reliably disrupted by both verbal and spatial rehearsal suppression.

Complementing these investigations, other researchers have sought to study potential domain-specific interference between two concurrently maintained sets of items. In one study by Morey and Mall (2012), participants were required to retain interleaved lists of

verbal and spatial items, and to reconstruct one of the lists after a short delay. In one experimental condition, participants received a cue in advance that indicated to them which of the lists had to be reconstructed after the delay. Thus, on these trials, participants only needed to encode and maintain one of the lists presented to them. In a second condition, participants were not cued in advance, meaning that they had to encode and maintain both lists. Morey and Mall (2012) found that both verbal and spatial serial reconstruction performance was higher when participants were cued in advance. In fact, performance in that case was similar to performance in a single task baseline condition[1]. Because the two dual-task conditions differed only in the presence or absence of a cue, it seemed that order encoding and/or maintenance of two lists interfered with each other. Due to the interleaved presentation mode, however, it is difficult to separate encoding from maintenance interference. This is important to note inasmuch as proponents of the multiple-component view might raise a similar argument to that of Meiser and Klauer (1999), stating that encoding rather than maintenance involves domain-general central executive processes and thus might account for the observed interference.

In another study by Depoorter and Vandierendonck (2009), participants had to remember a list of visually presented verbal materials and a list of visually presented spatial locations. Importantly, encoding and response for one list was embedded between encoding and response for the second list (encoding outer list → encoding embedded list → response embedded list → response outer list). The two lists contained items either from the same domain (outer-verbal/embedded-verbal condition or outer-spatial/embedded-spatial

---

1   In addition, analyses of serial position curves showed that reconstruction performance for the last item in the auditory-verbal set was not reliably affected by cueing. Morey and Mall (2012) interpreted this result as evidence for an additional role of verbal short-term memory in order retention. Although not discussed in their paper, better memory for the last one or two items with auditory presentation compared to memory for the same items with visual presentation is a well-known effect in short-term memory called the *modality effect* (Penney, 1989). There are a number of explanations for this effect, most of which refer to acoustic rather than verbal factors. In addition, there are demonstrations of a modality effect in spatial short-term memory (Tremblay et al., 2006), leading one to suspect that the "verbal" recency advantage in Morey and Mall's study was, in fact, an artifact of having different presentation modes for verbal and spatial items.

condition) or from different domains (outer-verbal/embedded-spatial condition or outer-spatial/embedded-verbal condition). Another manipulation consisted of varying the task requirements such that both the outer task and the embedded task could either be item recognition tasks or order recognition tasks. Using this design, Depoorter and Vandierendonck (2009) found that performance in the outer serial order tasks was reliably lower when the embedded tasks were also serial order tasks, and this result was obtained both in the same-domain conditions and the cross-domain conditions. In contrast, the outer item tasks did not exhibit analogous patterns of performance decrements depending on the embedded task requirements.

A possible interpretation of this result is that processes involved in short-term serial order retention are shared between domains, while short-term item retention can be supported by additional domain-specific verbal and visuospatial processes. Logie et al. (2012), however, subsequently pointed out that the patterns of interference in the Depoorter and Vandierendonck (2009) study might be related to output rather than maintenance activities, because the paradigm required the same type of response (key presses) in both tasks. Vandierendonck (2015) has recently addressed this and other criticisms of the paradigm by replacing the order recognition tasks with auditory-verbal serial recall and visuospatial order reconstruction tasks, and by requiring different responses for the two tasks. Again, it was found that embedded order tasks interfered more with outer order tasks than did embedded item tasks.

It is important to note that the argument against domain specificity in these studies is based on performance decrements in the outer order tasks depending on whether or not the embedded tasks require the retention of order information. Despite efforts to make the two order tasks more dissimilar in presentation and response characteristics (Vandierendonck, 2015), decrements in the outer tasks are potentially related to interference from activities

underlying order encoding, order maintenance, or order output in the embedded tasks. In contrast, performance differences in the embedded tasks—which should theoretically only originate from order maintenance requirements for the outer tasks—were either not observed (Depoorter & Vandierendonck, 2009) or not investigated (there were no combinations of outer item tasks with embedded item and order tasks in the study of Vandierendonck, 2015). Thus, the possibility remains that there is a role for domain-specific processes in serial order maintenance.

Furthermore, there is yet another potential difficulty associated with distinguishing between domain-general and domain-specific processes in dual-task paradigms. While extreme domain-general accounts and extreme domain-specific accounts predict a clear presence or absence of cross-list maintenance interference, the picture becomes more complex under a view that allows for both domain-general and domain-specific contributions to order maintenance. In this case, different amounts of cross-list interference may be expected for situations in which the amount of to-be-retained order information is within as opposed to beyond the limits of what hypothetical domain-specific processes can handle. Under conditions of low load, large cross-domain interference effects might not be found at all, as domain-specific processes should theoretically be able to handle the to-be-remembered information. In contrast, cross-domain interference should be observed when domain-specific capacity limits are exceeded, as in this situation the share of domain-general processing in order maintenance increases. If that were the case, in studies that have found large cross-domain interference—supporting domain-general accounts—order maintenance requirements might have been too large to allow for the detection of domain-specific contributions.

To summarize, demonstrations of maintenance interference between two order representations in short-term memory (1) under conditions of comparable task requirements

and presentation modes, (2) with tasks that emphasize order information and minimize item information, (3) using a design which isolates order encoding from maintenance, and (4) controlling for the possibility that the pattern of results differ between low-load and high-load conditions are still missing in the literature. We sought to provide them with the current set of experiments.

The Present Study

Participants were to perform two serial order tasks either with lists of items from different domains (Experiments 1 and 2) or from the same domains (Experiment 2). Contributions from item memory were minimized by repeatedly using the same items throughout the experiment and by using serial reconstruction instead of the more widely applied serial recall task. Thus, we would argue that the load manipulation mainly varied order memory demands.

In both experiments, a crucial manipulation consisted of an independent variation of list length. The rationale behind this manipulation was that if order maintenance processes are shared between domains and if these processes are capacity-limited, it should be possible to induce a trade-off situation between concurrent order maintenance activities. Under a domain-general view, one would then expect that both increases in same-domain and cross-domain order maintenance requirements (i.e., same-list load and cross-list load) reduce serial reconstruction performance in each task, because identical processes underlie order maintenance in both lists. Conditional on the assumption that there are no other ways of avoiding maintenance interference between the two lists, an extreme domain-general view would predict that the same-domain and cross-domain load effects are of equal size. In contrast, under an extreme domain-specific view, one might expect that increases in same-domain order requirements (i.e., same-list load) should reduce serial reconstruction performance, while cross-domain order requirements (i.e., cross-list load) should have negligible effects. Lastly, if one allows for both domain-general and domain-specific

contributions to order maintenance, there might be different amounts of domain-general involvement depending on whether the maintenance requirements approach or exceed the limits of what domain-specific processes can handle alone. According to this view, detrimental effects of cross-domain load might only be visible beyond a certain load threshold or, alternatively, they might generally be smaller than same-domain load effects, because between-domain competition would only concern those maintenance processes that are shared between the domains.

Note that such predictions are based on the assumption that interference between domains is solely determined by competition for capacity-limited order maintenance processes. Although this assumption seems to underlie much of the previous work on this issue, one must take into account other factors that might reduce cross-list interference in comparison to within-list interference. A potential concern here is that the domain-wise presentation *and response* mode allows for the grouping of items (e.g., Frankish, 1985; Ryan, 1965) which might reduce the likelihood of interference between the two lists irrespective of the items' domain. We will come back this issue in more detail later.

Experiment 1

In Experiment 1, participants carried out two serial reconstruction tasks, one task with a closed set of 4, 6, or 8 pseudowords,[2] and a second task with a closed set of 4, 6, or 8 spatial locations indicated by black squares. All combinations from 4-4 loads to 8-8 loads were tested. Encoding and reconstruction of one list was embedded between encoding and reconstruction of the second list (encoding task A → encoding task B → reconstruction task B → reconstruction task A). For half of the participants, the verbal task was embedded into

---

2   The term "pseudowords" is used to indicate non-existing words that are phonotactically possible in a language.

the spatial task (embedded-verbal group), and this order was reversed for the other half (embedded-spatial group).

Japanese two-mora pseudowords were selected by first generating all two-mora combinations that are phonotactically possible in the Japanese language and then removing real words from the list. For the remaining items, the probability of two morae being combined in that order was computed based on the Japanese written corpus by Tamaoka and Makioka (2004). A total of 318 items with medium bi-mora probability were selected and served as pool from which 9 items were selected for each participant at the beginning of the experiment. The selection of items was essentially random, but was constrained by the requirement that the same mora could only be used once in the set of 9 words to reduce item similarity. For each participant, an individual set of locations was generated by adding a random amount of spatial noise to a 3×3 matrix of 9 initially equally spaced locations ($\pm$ 50 pixels for x and y coordinates on a screen with 1280×1024 resolution).

Participants

Thirty-four male and twelve female undergraduate students between 18 and 25 years of age (median = 20) were recruited from the institute's participant pool, 23 participants being assigned to each one of the task order conditions based on their participant numbers. All participants reported normal or corrected-to-normal vision and were rewarded with a 1000JPY book coupon after the completion of the experiment.

(Insert Figure 1 about here)

Procedure

We will describe the trial procedure for the embedded-spatial task order condition here (see *Figure 1*). The procedure for the embedded-verbal condition was analogous.

In a trial, participants were first prompted with a fixation cross. After a short blank screen, 4, 6, or 8 pseudowords appeared centrally one after another painted in a black 22 point-sized font on a white background. After the disappearance of the last pseudoword, 4, 6, or 8 black 20×20 pixel squares were each presented at a different location followed one after another. Both pseudowords and squares were presented for 1250 milliseconds each, and they were separated by a 250 millisecond blank screen. After the last square disappeared, the screen remained blank for 2250 milliseconds (retention interval). After the retention interval, the unfilled black squares representing the locations appeared on the screen simultaneously. Participants indicated the order of square presentation via mouse clicks. When the mouse pointer entered one of the unfilled squares, this square was filled black so as to indicate that a mouse click would confirm selection of the square. The squares disappeared from the screen upon clicking, so that no double selections were possible. After spatial serial reconstruction was completed, the pseudowords were presented again simultaneously. Participants indicated order by clicking on the pseudowords one after another using the mouse. When the mouse pointer was inside an invisible boundary around a pseudoword, this pseudoword was underlined so as to indicate that a mouse click would confirm selection of the pseudoword. Pseudowords disappeared from the screen upon clicking, so that no double selections were possible. After verbal serial reconstruction was completed, the participants could either rest or proceed with the next trial by pressing a key.

At the beginning of an experimental session, participants were given the opportunity to practice one trial of the verbal and the spatial serial order tasks separately with 4, 6, and 8 items, and dual-task trials with combinations of 4-4, 6-6, and 8-8 items. In the main session, participants then performed 10 trials for each possible combination of verbal and spatial loads (3 verbal load conditions $\times$ 3 spatial load conditions $\times$ 10 trials = 90 trials).

Data analysis

Data in this paper were analyzed with mixed-effects logistic and linear regression models for accuracy and response time data, respectively. We first carried out a model selection procedure in which we compared predictive fit indexes of different regression models with various combinations of main effects and interactions. We chose the Deviance Information Criterion (DIC; Spiegelhalter et al., 2002) as our measure of model fit with a penalty for model complexity. As a general decision rule, the model with the lowest DIC value is to be preferred. Both the DICs and the parameters of the models were estimated using Markov chain Monte Carlo methods[3]. All regression models in the comparison process included random intercepts for subjects and random slopes for both embedded load and outer load with respect to subjects[4].

Statements about the presence of effects are based on the regression coefficients in the selected models. In the accuracy analyses, the expected detrimental effects of load on serial reconstruction performance are represented by regression coefficients $b < 0$, while coefficients equal to zero or with positive signs indicate no effect or effects in the opposite (and unexpected) direction. This logic is reversed for the response time analyses, as more items should lengthen (not shorten) response times.

The reliability of the effects is determined by 95% Bayesian credible intervals (CRI) around the regression coefficients computed from the MCMC output. These intervals can be interpreted as encompassing the 95% most likely coefficient values, given the collected data and certain prior assumptions about likely coefficient values[5]. Similar to a decision rule with confidence intervals (see Kruschke, 2011), an effect can be considered reliable if the credible

---

3   MCMC samples were obtained by running three chains with 200000 iterations in the Gibbs sampler JAGS version 3.4.0 (website: mcmc-jags.sourceforge.net). Uninformative prior distributions were chosen for all parameters. The output from JAGS was then fed into the statistical software R via the package R2jags for further computations.

4   We also checked whether this was actually necessary by comparing the DICs of models with different random effect structures. In all analyses, retaining all random parameters was indicated.

5   In the present case, prior assumptions practically played no role, as we used uninformative prior distributions on all parameters for the MCMC simulations.

interval around the corresponding regression coefficient is entirely below or above 0, with the interval itself giving the direction of the effect and an estimation of effect size.

In addition, we provide Bayes Factors (BF) as a measure of the strength of evidence for the default hypothesis, stating that there are detrimental effects of load, against the alternative hypothesis, stating that there are no detrimental effects or even unexpected facilitative effects of load. BFs > 1 support the default hypothesis, while BFs < 1 support the alternative hypothesis. For example, a BF of 100 means that the obtained data are about 100 times more likely to have occurred under the default hypothesis, while a BF of 0.01 means that the obtained data are about 100 times more likely to have occurred under the alternative hypothesis. Note that due to the limits of numerical approximation we sometimes report BFs approaching zero or positive infinity, which are theoretically impossible results. In these cases, we used an arrow symbol to indicate that the "real" BFs were very small or very large, respectively. The Bayes factors were computed following the method described in Hoijtink (2012) for inequality constrained hypotheses.

Lastly, when analyzing data from a dual-task interference paradigm, one usually needs to ensure that the concurrent task is performed with sufficient accuracy in order to conclude that there was potential interference on the primary task. In the present case, we had to ensure that serial order is maintained in the concurrent task and, thus, a maintenance load is incurred. However, adopting a strict accuracy criterion for the concurrent task led to a considerable loss of trials, especially in the 8-8 condition. We therefore report two types of analyses, one adopting a 100% accuracy criterion for the concurrent task, and a second type including all trials. In the following, these analyses will be labeled *strict* analyses and *raw* analyses, respectively. Although the theoretical weight lies on observations in the embedded tasks, for completeness, we also report the analyses of the outer tasks. Descriptive statistics can be found in *Supplements A1* and *A2*.

Accuracy. In both the strict and raw analyses, the number of correct responses divided by load (i.e., list length) and irrespective of serial position served as the dependent variable. We used this measure to model performance with regard to the effects of embedded load, outer load, and interactions between these two factors. The results of the model selection procedure can be seen in *Table 1*. A plot of the data is given in *Figure 2.*

(Insert Figure 2, Table 1 about here)

Strict analyses. Verbal serial reconstruction performance was calculated for trials with perfect spatial serial reconstruction performance, and spatial serial reconstruction performance was calculated for trials with perfect verbal serial reconstruction performance. The selected model for the embedded verbal task contained a reliable effect of verbal load in the expected direction (b = -0.49, CRI = [-0.68; -0.31], BF = 2983), while the selected model for the outer verbal task contained a reliable effect of verbal load in the expected direction (b = -0.56, CRI = [-0.74; -0.39], BF ∞) and a non-reliable effect of spatial load (b = -0.07, CRI = [-0.24; 0.10], BF = 4).

The selected model for the embedded spatial task contained a reliable effect of spatial load in the expected direction (b = -0.89, CRI = [-1.06; -0.70], BF ∞) and a non-reliable effect of verbal load (b = -0.08, CRI = [-0.26; 0.09], BF = 5). The selected model for the outer spatial task contained a reliable effect of spatial load in the expected direction (b = -0.60, CRI = [-0.81; -0.40], BF = 2971) and a reliable effect of verbal load (b = -0.28, CRI = [-0.52; -0.02], BF = 57). This result was qualified by a reliable spatial versus verbal load interaction with a positive coefficient (b = 0.03, CRI = [0.01; 0.05], BF = 0.03). Note that the BF for the interaction indicates that there is around 33 times more support for the alternative hypothesis compared to the default hypothesis.

Raw analyses. The selected model for the embedded verbal task contained a reliable effect of verbal load with an expected negative sign (b = -0.49, CRI = [-0.66; -0.32], BF = 3159). The selected model for the outer verbal task contained reliable effects of both verbal load (b = -0.71, CRI = [-0.95; -0.48], BF = 1525) and spatial load (b = -0.36, CRI = [-0.60; -0.12], BF = 625) in the expected direction. The interaction between the two factors was reliable but with an unexpected positive coefficient (b = 0.03, CRI = [0.01; 0.06], BF = 0.01). Note that the BF for this interaction indicates that there is 100 times more support for the alternative hypothesis compared to the default hypothesis.

In the analysis of the embedded spatial task, the selected model contained a reliable effect of spatial load (b = -1.27, CRI = [-1.52; -1.03], BF = 3019) and a reliable effect of verbal load (b = -0.67, CRI = [-0.92; -0.41], BF = 948), both effects having the expected negative signs. The interaction between these two factors was reliable, with the coefficient being positive (b = 0.08, CRI = [0.05; 0.11], BF   0). The BF for this interaction indicates practically unlimited support for the alternative hypothesis. The selected model for the outer spatial task contained reliable effects of spatial load (b = -0.64, CRI = [-0.83; -0.46], BF = 2889) and verbal load (b = -0.38, CRI = [-0.58; -0.18], BF = 1503) with both effects being in the expected direction. The interaction between these two factors was reliable, however, with the coefficient having an unexpected positive sign (b = 0.03, CRI = [0.02; 0.05], BF   0). Again, the BF indicates practically unlimited support for the alternative hypothesis.

(Insert Figure 3, Table 2 about here)

Response times. Because response time data showed a marked right skew, we applied a log10 transformation and scaled the data by a factor of ten. The results of the model selection procedure can be seen in *Table 2*. A plot of the data is given in *Figure 3*.

Strict analyses. Response times were calculated and analyzed for trials with perfect performance in both tasks. For the embedded verbal task, this resulted in two models having an identically low DIC (1946). Both models contained reliable effects of verbal load with an expected positive sign (both $b = 1.15$, CRI = [0.99; 1.31], BF ∞), while one of the models contained an additional but non-reliable effect of spatial load ($b = -0.01$, CRI = [-0.15; 0.15], BF = 0.85). The selected model for the outer verbal task contained a reliable effect of verbal load in the expected direction ($b = 1.14$, CRI = [0.93; 1.35], BF ∞), a non-reliable effect of spatial load ($b = -0.16$, CRI = [-0.37; 0.04], BF = 0.06), and a non-reliable interaction between verbal and spatial load ($b = 0.02$, CRI = [0; 0.05], BF = 22).

The selected model for the embedded spatial task contained a reliable effect of spatial load in the expected direction ($b = 0.92$, CRI = [0.76; 1.09], BF ∞), a non-reliable effect of verbal load ($b = 0$, CRI = [-0.17; 0.18], BF = 1), and a non-reliable interaction between the two factors ($b = -0.02$, CRI = [-0.04; 0], BF = 0.04). In the analyses of the outer spatial task, the lowest DIC was observed in two models (2535). Both models contained a reliable effect of spatial load in the expected direction with slightly different estimates (outer load only model: $b = 0.97$, CRI = [0.81; 1.13], BF ∞; full model: $b = 0.79$, CRI = [0.51; 1.07], BF ∞), while the latter contained an additional but non-reliable effect of verbal load ($b = -0.20$, CRI = [-0.46; 0.07], BF = 0.08), and a non-reliable interaction between the two effects ($b = 0.03$, CRI = [-0.01; 0.08], BF = 14).

Raw analyses. The selected model for the embedded verbal task only included a reliable effect of verbal load ($b = 1.18$, CRI = [1.02; 1.33], BF ∞) in the expected direction. In the analyses of the outer verbal task, three models had an identical DIC (7014). In all three models, however, only the effect of verbal load was reliable, with the coefficient having the expected positive sign (model including only the outer load effect: $b = 1.27$, CRI = [1.12; 1.42], BF ∞; model including both load effects: $b = 1.27$, CRI = [1.11; 1.42], BF ∞;

model including both load effects and the load interaction: $b = 1.36$, CRI = [1.17; 1.55], BF $\approx \infty$).

The selected model for the embedded spatial task contained a reliable effect of spatial load ($b = 1.03$, CRI = [0.86; 1.19], BF $\approx \infty$) in the expected direction, a non-reliable effect of verbal load ($b = -0.01$, CRI = [-0.16; 0.15], BF = 0.79), and a non-reliable interaction between these two factors ($b = -0.02$, CRI = [-0.03; 0], BF = 0.03). The selected model for the outer spatial task contained a reliable effect of spatial load ($b = 1.26$, CRI = [1.04; 1.47], BF $\approx \infty$) and a reliable effect of verbal load ($b = 0.26$, CRI = [0.05; 0.47], BF = 195) with both effects being in the expected direction. In addition, the interaction between the two loads was reliable in an unexpected direction ($b = -0.04$, CRI = [-0.07; -0.02], BF ≈ 0).

Summary and Discussion

Adopting a strict accuracy criterion for the concurrent task, it was found that increases in verbal but not spatial load reduced verbal serial reconstruction performance. Conversely, only increases in spatial but not verbal load reduced spatial reconstruction performance. The response time data unsurprisingly showed increases in reconstruction times the more items had to be reconstructed. In contrast, the effect of cross-domain load on response times was non-reliable. One important difference emerged between the strict and the raw analyses in that embedded spatial reconstruction performance was affected by both same-domain and cross-domain load (see *Table 1*)[6].

In sum, serial reconstruction performance remained surprisingly resistant to increases in cross-domain load, and this pattern was observed in both the accuracy and the response time data. At first glance (and with the exception of the asymmetric cross-domain interference in the raw analyses), this result seems to be difficult to explain under a purely domain-general

---

6   Additional analyses showed that cross-domain load effects appeared in trials with low accuracy in the concurrent task.

view. Based on the idea that cross-domain load increases should tax domain-general order maintenance processes, one would have expected a clear pattern of cross-domain load effects for both serial reconstruction tasks, ideally similar in size to same-list load effects. The results of Experiment 1, thus, rather corroborate the viewpoint that domain-specific processes contribute to serial order memory (Logie et al., 2012; Saito et al., 2008).

Such a conclusion, however, does not take into account alternative factors that might reduce cross-list interference. As outlined in the introduction, a potential concern in the present case is that the domain-wise presentation and response modes might induce a grouping of the materials according to their domain. In a typical study investigating the effect that grouping has on serial order performance, participants are presented with a single list of (usually verbal) items and are instructed to recall the list in correct serial order after a short delay. At some point during list presentation, a temporal boundary (e.g., an extended pause) or a non-temporal boundary (e.g., a change in presentation location) is inserted, which effectively causes participants to perceive two sublists. This manipulation results in the appearance of primacy and recency effects for each sublist and an overall increase in recall accuracy for the whole list (Farrell, 2012; Frankish, 1985; Henson, 1998; Hurlstone & Hitch, 2015; Parmentier & Maybery, 2008; Ryan, 1969).

Most theories of serial order memory assume that the effects of grouping are related to the hierarchical coding of positional information in grouped lists (Burgess & Hitch, 1999; Farrell, 2012; Henson, 1998). According to these theories, associations are formed between the list's items and contextual cues that serve as positional markers. At recall, these cues are reinstated one after another and, thereby, the associated items are retrieved in correct order. In grouped lists, each item is associated with two kinds of positional markers, one representing positional information relative to other items in a sublist, and another marker representing an item's sublist membership. An alternative but older idea is that grouping acts

as some sort of chunking (Wickelgren, 1964). According to this account, sublists can be compressed into a single unit or chunk, which results in a reduction of the quantity of information that must be retained at a time. This compression is reversed at recall in order to access within-sublist order information.

Both positional marking and chunking can be regarded as ways of separating two lists' memory representations from each other. If this separation can be achieved to a sufficient degree, the level of cross-list interference should not depend on the numbers of items in the concurrent list. Thus, grouping may explain the lack of cross-list interference in Experiment 1 without the need to refer to any domain-specific processes. However, if grouping mechanisms could entirely account for this observation, we should find exactly the same pattern of results with concurrent serial reconstruction of two verbal lists or two spatial lists. We tested this prediction in Experiment 2.

Experiment 2

We introduced two new same-domain conditions to the paradigm. In the verbal-verbal condition, one task was performed with two-digit numbers, while the second task was performed with pseudowords. In the spatial-spatial condition, there were two different sets of locations. If the very same domain-general processes were recruited for verbal and spatial order maintenance and if the lack of cross-list interference in Experiment 1 could be explained by grouping rather than by the presence of domain-specific order maintenance processes, then the pattern of results should not differ between the newly introduced same-domain and the cross-domain conditions. In particular, we should neither see effects of cross-list load in any of the conditions nor a decrease in embedded task performance when the domains of embedded and outer tasks match. In contrast, one can think of at least two possible outcomes for the case of having domain-specific contributions to order maintenance. First, without grouping, one can expect to find cross-list load effects in the same-domain

conditions but not in the cross-domain conditions. However, if grouping effects are taken into account, one may expect a lack of cross-list load effects in all conditions but general performance decrements in the same-domain conditions compared to the cross-domain conditions.

Materials

In the embedded-verbal condition, the embedded task was the previously used pseudoword reconstruction task, while the outer task was either the previously used location reconstruction task or a second verbal serial reconstruction task with a list of two-digit numbers. In the embedded-spatial condition, the embedded task was the previously used location reconstruction task, while the outer task was either the previously used pseudoword reconstruction task or a second location reconstruction task with a different set of locations.

Pseudowords were sampled using the method described in Experiment 1. For the number reconstruction task, we sampled for each participant a set of 9 numbers from a pool containing two-digit numbers that always differed in the first and the second digit. Thus, there were 81 possible items from which the numbers could be sampled. In contrast to Experiment 1, the sets of locations for the square and circle reconstruction tasks were sampled from a pool of 81 possible locations defined by a 9×9 matrix of equally spaced regions on the screen. The location sampling procedure involved two steps. First, a subset of 18 locations was chosen from the 9×9 matrix. In order to select a spatially balanced subset of locations, the first location was constrained to be near the center of the screen, and all subsequent locations had to be within a certain minimum and maximum distance from each other. In a second step, the selected 18 locations were alternately assigned to the square and circle reconstruction tasks, resulting in two spatially interleaved sets of items.

In the embedded-verbal condition, the locations in the outer spatial task were indicated by filled squares, as in Experiment 1. In the embedded-spatial condition, the locations in the embedded spatial tasks were indicated by filled black squares, while the locations in the outer spatial task were indicated by filled black circles.

Participants

Forty male and twenty female undergraduate students between 18 and 36 years of age (median = 21) were recruited from the institute's participant pool, 30 participants being assigned to each one of the task order conditions based on their participant numbers. All participants reported normal or corrected-to-normal vision and were rewarded with a 1000JPY book coupon after the completion of the experiment.

Procedure

The procedure for the replication conditions was identical to Experiment 1. For the novel same-domain conditions, we used the same presentation and response parameters; in other words, the only difference between pseudoword and number reconstruction were the items themselves, while the only differences between the two location reconstruction tasks were the shape of the items indicating the locations and the locations sets themselves.

At the beginning of an experimental session, participants were given the opportunity to practice single task trials (embedded-verbal group: pseudowords, numbers, square locations; embedded-spatial group: square locations, circle locations, pseudowords) with 4, 6, and 8 items, and dual-task trials (embedded-verbal group: pseudowords-numbers, pseudowords-squares; embedded-spatial group: squares-circles, squares-pseudowords) with 4-4, 6-4, and 8-4 items.

Results

Data of one participant were lost due to a recording error. A second participant withdrew his participation after the beginning of the experiment due to time constraints. Thus, the analyses were conducted with the data of 58 participants. The modeling procedure was largely the same as in Experiment 1. However, a random slope parameter for the within-participant factor *outer task domain* with respect to subjects was added to the random effects structure in the analyses[7]. Because of the complexity of the design, we report only the results of the analyses for the embedded tasks, as they are the main focus of the present research. Additional statistics for the outer tasks can be found in the supplement section (*Supplements A5* and *A6*).

Accuracy. As in Experiment 1, the number of correct responses divided by load served as the dependent variable. In the strict analyses, serial reconstruction performance in the embedded tasks was calculated for trials with perfect outer task performance. In the raw analyses, all trials were included irrespective of the performance in the outer task. Model selection was carried out with models containing effects of embedded load, outer load, and outer task domain. The latter effect was coded such that the cross-domain conditions served as the baseline conditions. Thus, a negative regression coefficient for the factor outer task domain indicates a performance decrement for the same-domain conditions in comparison to the cross-domain conditions, which may be taken as an indicator of domain-specific contributions to serial order maintenance. The embedded-verbal and embedded-spatial conditions were analyzed separately. The results of the model selection procedure can be seen in *Table 3*. A simplified plot of the data is given in *Figure 4* (see *Supplement A3* for descriptive statistics).

(Insert Figure 4, Table 3 about here)

---

7   Again, it was evaluated whether these parameters were actually necessary by comparing the DICs of different regression models that included all fixed effect parameters but different random effect parameters. Retainment of all slope parameters was indicated.

Strict analyses. For the embedded verbal task, the model with the lowest DIC contained reliable main effects of verbal load (b = -0.65, CRI = [-0.85; -0.47], BF = 3101), outer load (b = -0.25, CRI = [-0.43; -0.06], BF = 144), and outer task domain (b = -1.55, CRI = [-2.34; -0.81], BF = 1497) in the expected direction. In addition, there was a reliable interaction between outer load and outer task domain with a positive sign (b = 0.25, CRI = [0.12; 0.38], BF   0). The BF for this latter interaction indicates practically unlimited support for the alternative hypothesis. Additional models were fitted separately for each outer task domain condition in order to examine potential differences of the effect of outer load in these conditions. In both conditions, however, the selected models only contained verbal load effects in the expected direction (verbal-spatial: b = -0.64, CRI = [-0.87; -0.42], BF = 3068; verbal-verbal b = -0.72, CRI = [-0.96; -0.49], BF = 1495).

In the embedded spatial task analysis, the lowest DIC was observed in three models (4045) that differed in the presence/absence of a non-reliable effect of outer load or a non-reliable interaction between spatial load and outer load. Because all models lead to the same conclusions, we report the results for the model with the fewest parameters. This model contained a reliable main effect of spatial load (b = -0.85, CRI = [-0.99; -0.71], BF ∞) and a reliable main effect of outer task domain (b = -1.87, CRI = [-2.70; -1.02], BF ∞) in the expected direction. In addition, there was a reliable interaction between spatial load and outer task domain with a positive sign (b = 0.15, CRI = [0.04; 0.27], BF = 0.01). The BF for this interaction indicates 100 times more support for the alternative hypothesis compared to the default hypothesis.

Raw analyses. In the embedded verbal task analysis, the model with the lowest DIC contained reliable main effects of verbal load (b = -0.53, CRI = [-0.65; -0.40], BF = 1477), outer load (b = -0.16, CRI = [-0.28; -0.04], BF = 142), and outer task domain (b = -0.55, CRI = [-0.89; -0.20], BF = 953) in the expected direction. The three-way interaction between

these factors contained in the selected model was not reliable ($b = 0$, CRI = [0; 0.01], BF = 0.05).

In the embedded spatial task analysis, two models had an identically low DIC (10974). The model with the fewer parameters contained reliable effects of spatial load ($b = -0.83$, CRI = [-0.99; -0.67], BF = 1501), outer load ($b = -0.24$, CRI = [-0.40; -0.09], BF = 432), and outer task domain ($b = -1.3$, CRI = [-1.62; -0.99], BF = 1485). In addition, there was a non-reliable interaction between spatial load and outer load ($b = 0.02$, CRI = [0; 0.03], BF = 0.02) and a reliable three-way interaction with a positive sign ($b = 0.02$, CRI = [0.01; 0.02], BF   0). The BF of this latter interaction indicates practically unlimited support for the alternative hypothesis.

Response times. As in Experiment 1, data were log10 transformed and scaled by a factor of ten prior to the statistical analyses. In the strict analyses, response times were calculated for trials with perfect performance in both the embedded and outer tasks. In the raw analyses, all trials were included irrespective of performance in the two tasks. The models contained effects of embedded load, outer load, and outer task domain. The latter effect was coded such that the cross-domain conditions served as the baseline conditions. Thus, a positive regression coefficient for the outer task domain factor indicated an increase in response times from the cross-domain conditions to the same-domain conditions, which can be taken as an indicator of domain-specific contributions to serial order maintenance. The embedded-verbal and embedded-spatial conditions were analyzed separately. The results of the model selection procedure can be seen in *Table 4*. A simplified plot of the data is given in *Figure 5*. See *Supplement A4* for descriptive statistics.

(Insert Figure 5, Table 4 about here)

Strict analyses. In the analysis of the embedded verbal task, the selected model contained reliable effects of verbal load ($b = 0.83$, CRI = [0.63; 1.04], BF $\approx \infty$) and outer task domain ($b = 0.57$, CRI = [0.14; 0.99], BF = 167) as well as a reliable interaction between verbal load and outer load ($b = 0.05$, [0.02; 0.09], BF = 1542) with all effects being in the expected direction. In addition, there was a reliable three-way interaction between verbal load, outer load, and outer task domain with a negative sign ($b = -0.02$, CRI = [-0.03; -0.01], BF ≈ 0). The BF for this latter interaction indicates practically unlimited support for the alternative hypothesis. The effect of outer load contained in the selected model failed to reach the reliability criterion ($b = -0.18$, CRI = [-0.39; 0.05], BF = 0.07).

The selected model for the embedded spatial task contained a reliable effect of spatial load ($b = 0.86$; [0.73; 0.98], BF $\approx \infty$) only.

Raw analyses. In the analysis of the embedded verbal task, the selected model contained a reliable effect of verbal load in the expected direction ($b = 1.12$; [1.00; 1.24], BF $\approx \infty$). In addition, there were reliable interactions between verbal load and outer task domain ($b = 0.08$; [0.01; 0.16], BF = 42) and a three-way interaction between verbal load, outer load, and outer task domain ($b = -0.02$; [-0.03; -0.01], BF ≈ 0). The BF for the latter interaction indicates practically unlimited support for the alternative hypothesis. The main effects of outer load ($b = 0.06$; [-0.06; 0.17], BF = 5) and outer task domain ($b = 0.24$; [-0.21; 0.69], BF = 6) were non-reliable.

In the analysis of the embedded spatial task, the selected model contained a reliable effect of spatial load in the expected direction ($b = 0.98$; [0.84; 1.11], BF $\approx \infty$), a reliable interaction between spatial and outer load ($b = -0.02$; [-0.03; -0.01], BF ≈ 0), and a reliable interaction between outer load and outer task domain ($b = 0.05$; [0.01; 0.09], BF = 87). The BF for the former indicates practically unlimited support for the alternative hypothesis. The

main effects of outer load (b = 0.05; [-0.08; 0.18], BF = 4) and outer task domain (b = -0.10; [-0.44; 0.24], BF = 0.36) were non-reliable. To understand the interaction between outer load and outer task domain, additional models were fitted separately for each outer task domain condition. In the spatial/verbal condition, there was a reliable effect of spatial load in the expected direction (b = 0.97; [0.81; 1.12], BF ∞) only. The effect of verbal load (b = 0.04; [-0.12; 0.19], BF = 2) and the interaction between spatial and verbal load (b = -0.02; [-0.03; 0], BF = 0.05) did not reach the reliability criterion. Likewise, in the spatial/spatial condition, there was a reliable effect of embedded spatial load in the expected direction (b = 0.99; [0.83; 1.15], BF ∞) only, and a non-reliable effect of outer spatial load (b = 0.12; [-0.04; 0.27], BF = 13). However, the interaction between the two loads was reliable in this model (b = -0.02; [-0.04; -0.01], BF = 0.02), with the BF indicating 50 times more support for the alternative hypothesis compared to the default hypothesis.

Summary and Discussion

In Experiment 2, performance in the embedded tasks was generally worse when the domain of the embedded and outer tasks matched, albeit the drop in performance was small (~-3.7% for verbal reconstruction and ~-6.1% for spatial reconstruction, respectively). As in Experiment 1, increases in same-list load unsurprisingly reduced reconstruction performance, while cross-list load effects were small to non-reliable. The analyses of response times revealed consistent effects of same-list load but non-reliable effects of cross-list load. The effects of domain were non-reliable in the response time analyses with the exception of the embedded-verbal condition in the strict analyses.

The fact that embedded task performance was remarkably resistant to cross-list load increases even in the same-domain conditions is consistent with the idea that cross-list interference is reduced by grouping mechanisms such as multilevel positional marking or chunking. On the other hand, the main effect of domain found in the accuracy analyses is

difficult to reconcile with a purely domain-general view of serial order maintenance and calls for an explanation that includes domain-specific factors.

General Discussion

We conducted two experiments in which participants had to concurrently maintain and reconstruct the order of verbal and spatial lists. In both experiments, encoding and serial reconstruction of one list was embedded between encoding and serial reconstruction of a second list. Furthermore, we independently manipulated order maintenance requirements in the two lists by varying list lengths between 4 and 8 items. Experiment 1 required serial reconstruction of one verbal and one spatial list (in verbal-spatial or spatial-verbal order). Experiment 2 replicated this design with two additional same-domain conditions (verbal-verbal and spatial-spatial).

In both experiments, serial reconstruction performance remained largely unaffected by increases in cross-list load but was sharply reduced by increases in same-list load. In Experiment 1, there was some indication of a detrimental effect of verbal load on spatial serial reconstruction only when the accuracy criterion for the concurrent task was relaxed. In Experiment 2, on the other hand, verbal serial reconstruction seemed to suffer slightly from increases in spatial load in both analyses. In addition, when the accuracy criterion for the concurrent task was relaxed, verbal load seemed to reduce spatial serial reconstruction performance. Generally, however, the detrimental effects of the concurrent tasks were relatively weak and unstable across experiments, which indicate that cross-list interference can be avoided by ways unrelated to domain-specific or domain-general order maintenance processes. As suggested, we think that grouping mechanisms, such as hierarchical positional marking or chunking, help to separate the two lists' memory representations from each other and, because of that, the level of cross-list interference is only weakly related to the numbers of items in the concurrent list.

More important with regard to the present research question is the observation that in Experiment 2 serial reconstruction performance in both the embedded verbal and the embedded spatial tasks was generally lower when the domains of the lists in the embedded and outer tasks matched. A straightforward interpretation of this finding is that it indicates contributions of domain-specific processes to order maintenance. In line with previous research (Logie et al., 2012; Saito et al., 2008), these contributions could take the form of functionally specialized verbal and spatial rehearsal processes. Related to this, there is broad agreement among researchers that verbal item information can be rehearsed by a specialized articulation-based mechanism (Baddeley, 2012; Camos, Lagner, & Barrouillet, 2009; Logie et al., 2012), and it is possible that an identical mechanism supports rehearsal of verbal order information as well (Burgess & Hitch, 1999). In addition, some researchers have argued for the existence of a corresponding functionally specialized rehearsal mechanism in the spatial domain (e.g., Tremblay, Saint-Aubin, & Jalbert, 2006), although this idea seems to be more controversial (Klauer & Stegmaier, 1997; Morey et al., 2013).

An alternative way of explaining the main effect of domain in Experiment 2 is to assume that it is relatively more difficult to encode or retrieve (rather than maintain) the order of a list of items from one domain while maintaining the order of a second list of items from the same domain. Regarding this issue, most computational models of serial order memory incorporate mechanisms by which the successful retrieval of an item depends on that item's similarity to other items in the list (reviewed in Hurlstone et al., 2014). If one makes the additional assumption that items drawn from a single domain are more similar to each other than items drawn from different domains, the retrieval of items should be relatively more impaired by a concurrently held list containing items from the same domain.

Alternatively, item similarity may directly impair the representation of order information. Related to this, there is evidence that the same type of ordering mechanisms may operate in

both the verbal and the spatial domains (Hurlstone, Hitch, & Baddeley, 2014; Hurlstone & Hitch, 2015). Among these mechanisms are primacy gradients, which are the relative activation levels of items in memory resulting from their presentation order. Some models of serial order memory assume that the steepness of a primacy gradient, and thereby the strength of the order representation, is directly influenced by item similarity (e.g., Farrell & Lewandowsky, 2002).

Finally, one has to consider the possibility that the main effect of domain does not reflect domain-specific contributions to order memory. It is possible for example that the grouping cues that separate the two lists in the present paradigm were simply more salient in the cross-domain conditions as compared to the same-domain conditions. Recall that in the cross-domain conditions, participants could group the lists by spatial location (the verbal spatial items were always presented in the screen center). In the same-domain conditions, in contrast, the lists differed only in the identity of the items (squares versus circles in the spatial conditions, and pseudowords versus numbers in the verbal conditions). This explanation, however, does not seem convincing given that the participants knew in advance how many items would belong to each list and given that reconstruction was performed list-wise. Moreover, if grouping was weaker in the same-domain conditions than in the cross-domain conditions, how could we explain the lack of cross-list load effects? Thus, it seems unlikely that grouping was less effective in the same-domain conditions as compared to the cross-domain conditions. In sum, we think that the present data are more consistent with explanations referring to domain-specific contributions to serial order maintenance.

Conclusion

We have presented the results of two experiments demonstrating patterns of maintenance interference between two order representations (1) under conditions of comparable task requirements and presentation modes, (2) with tasks that emphasize order information and

minimize item information, (3) using a design which isolates order encoding from maintenance, and (4) controlling for the possibility that the pattern of results differ between low-load and high-load conditions. The two most important results concern the main effect of domain in Experiment 2, which is consistent with the involvement of domain-specific processes in order maintenance, and the surprisingly weak cross-list interference across conditions in both experiments, which seems to be related to grouping mechanisms operating in concurrently performed serial reconstruction tasks.

As we have outlined in the discussion section, domain-specific contributions to order maintenance may take different forms. One possibility is that there are independent maintenance processes in the verbal and spatial domains, respectively. Alternatively, concurrent encoding and/or retrieval of two lists with items from the same domain might be relatively more difficult due to confusion in memory related to either the item or the order representation. We suggest that future experimental research should seek to address these different possibilities in more detail.

Acknowledgements

References

Avons, S. E. (1999). Effects of visual similarity on serial report and item recognition. *The Quarterly Journal of Experimental Psychology*, *52A*, 217-240.

Baddeley, A. D. (2012). Working Memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1-29.

Baddeley, A. D., & Hitch, G. J. (1974). Working Memory. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory Vol. 8* (pp. 47–89). New York: Academic Press.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*, 551–581.

Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language*, *61*, 457– 469.

Depoorter, A., & Vandierendonck, A. (2009). Evidence for modality-independent order coding in working memory. *The Quarterly Journal of Experimental Psychology, 62*, 531-549.

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review*, *9*, 59–79.

Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. Psychological *Review*, *119*, 223-271.

Frankish, C. (1985). Modality-specific grouping effects in short-term memory. *Journal of Memory and Language*, *24*, 200–209.

Guérard, K., & Tremblay, S. (2008). Revisiting evidence for modularity and functional equivalence across verbal and spatial domains in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 556 –569.

Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology*, *36*, 73–137.

Hoijtink, H.J.A. (2012). *Informative Hypotheses. Theory and Practice for Behavioral and Social Scientists.* Boca Raton: Chapman & Hall/CRC.

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin*, *140*, 339–373.

Hurlstone, M. J., & Hitch, G. J. (2015). How is the serial order of a spatial sequence represented? Insights from transposition latencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 295-324.

Jones, D. M., Farrand, P., Stuart, G., & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1008-1018.

Klauer, K. C. & Stegmaier, R. (1997). Interference in immediate spatial memory: Shifts of spatial attention or central executive involvement?. *The Quarterly Journal Of Experimental Psychology*, *50A*, 79-99.

Kruschke, J. K. (2011). Bayesian assessment of null values via parameter estimation and model comparison. *Psychological Science*, *6*, 299-312.

Logie, R. H., Saito, S., Morita, A., Varma, S., & Norris, D. (2012). *Recalling visual serial order for verbal sequences*. Paper presented at the meeting of the Experimental Psychology Society, University College London, London, U.K., Abstract p.23-24, January 5-6.

Meiser, T., & Klauer, K. C. (1999). Working memory and changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*, 1272-1299.

Morey, C. C., & Mall, J. T. (2012). Cross-set interference costs during concurrent verbal and spatial serial memory tasks are asymmetric. *The Quarterly Journal of Experimental Psychology*, *65*, 1777-1797.

Morey, C. C., Morey, R. D., van der Reijden, M., & Holweg, M. (2013). Asymmetric cross-domain interference between two working memory tasks: Implications for models of working memory. *Journal of Memory and Language*, *69*, 324-348.

Parmentier, F. B. R., & Jones, D. M. (2000). Functional characteristics of auditory temporal-spatial short-term memory: Evidence from serial order errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 222-238.

Parmentier, F. B. R., & Maybery, M. T. (2008). Equivalent effects of grouping by time, voice, and location on response. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*, 1349 –1355.

Ryan, J. (1969). Grouping and short-term memory: Different means and patterns of grouping. *The Quarterly Journal of Experimental Psychology*, *21*, 137-147.

Saito, S., Logie, R. H., Morita, A., & Law, A. (2008). Visual and phonological similarity effects in verbal immediate serial recall: A test with Kanji materials. *Journal of Memory and Language*, 59, 1–17.

Smyth, M. M. & Scholey, K. A. (1996). Serial order in spatial immediate memory. *The Quarterly Journal of Experimental Psychology*, *49A*, 159-177.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian

measures of model complexity and fit. *Journal of the Royal Statistical Society: Series*

B (Statistical Methodology), *64*, 583–639.

Tamaoka, K., & Makioka, S. (2004). Frequency of occurrence for units of phonemes, morae,

and syllables appearing in a lexical corpus of a Japanese newspaper. *Behavior*

Research Methods, Instruments, & Computers, *36*, 531-547.

Tremblay, S., Parmentier, F. B. R., Guérard, K., Nicholls, A. P., & Jones, D. M. (2006). A

spatial modality effect in serial memory. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, *32*, 1208-1215.

Tremblay, S., Saint-Aubin, J., & Jalbert, A. (2006). Rehearsal in serial memory for visual-

spatial information: Evidence from eye movements. *Psychonomic Bulletin & Review*,

13, 452-457.

Vandierendonck, A. (2015). Modality-independence of order coding in working memory:

Evidence from cross-modal order interference at recall. *The Quarterly Journal of*

Experimental Psychology. Advance Online Publication.

Wickelgren, W. A. (1964). Size of rehearsal group and short-term memory. *Journal of*

Experimental Psychology, *68*, 413-419.

Table captions

1) Table 1: Results of the model selection procedure for the accuracy data in Experiment 1. Shown are the models with the three lowest DICs. In the strict analyses (lower part), only trials with 100% accuracy in the concurrent task were included. In the raw analyses (upper part) all trials were included. Reliable effects in these models are indicated by "+" for a positive coefficient and "-" for a negative coefficient. Non-reliable effects included in the model are indicated by "?".

Embedded verbal task				
Analysis	DIC	Verbal	Spatial	Verbal x Spatial
Raw	6248	-		
	6249	-	?	
	6250	-	?	?
Strict	2575	-		
	2576	-	?	
	2577	-	?	?
Outer verbal task				
Analysis	DIC	Verbal	Spatial	Verbal x Spatial
Raw	5855	-	-	+
	5863	-		
	5863	-	?	
Strict	3531	-	?	
	3532	-		
	3535	-	?	?
Embedded spatial task				
Model	DIC	Spatial	Verbal	Spatial x Verbal
Raw	5368	-	-	+
	5397	-		
	5397	-	?	
Strict	3011	-	?	
	3012	-		
	3012	-	?	?
Outer spatial task				
Model	DIC	Spatial	Verbal	Spatial x Verbal
Raw	8683	-	-	+
	8697	-	-	
	8698	-		
Strict	5497	-	-	+
	5500	-		
	5500	-	?	

2) Table 2: Results of the model selection procedure for the response time data in Experiment 1. Shown are the models with the three lowest DICs. In the strict analyses (lower part), only trials with 100% accuracy in both tasks were included. In the raw analyses (upper part) all trials were included. Reliable effects in these models are indicated by "+" for a positive coefficient and "-" for a negative coefficient. Non-reliable effects included in the model are indicated by "?".

		Embedded verbal task		
Analysis	DIC	Verbal	Spatial	Verbal x Spatial
Raw	6532	+		
	6533	+	?	
	6534	+	?	?
Strict	1946	+		
	1946	+	?	
	1948	+	?	?
		Outer verbal task		
Analysis	DIC	Verbal	Spatial	Verbal x Spatial
Raw	7014	+		
	7014	+	?	
	7014	+	?	?
Strict	3328	+	?	?
	3330	+	?	
	3330	+		
		Embedded spatial task		
Model	DIC	Spatial	Verbal	Spatial x Verbal
Raw	5582	+	?	?
	5584	+	?	
	5585	+		
Strict	2596	+	?	?
	2597	+	?	
	2597	+		
		Outer spatial task		
Model	DIC	Spatial	Verbal	Spatial x Verbal
Raw	7697	+	+	-
	7707	+		
	7707	+	?	
Strict	2535	+		
	2535	+	?	?
	2536	+	?	

3) Table 3: Results of the model selection procedure for the accuracy data in Experiment 2. Shown are the models with the three lowest DICs. In the strict analyses (lower part), only trials with 100% accuracy in the concurrent task were included. In the raw analyses (upper part) all trials were included. Reliable effects in these models are indicated by "+" for a positive coefficient and "-" for a negative coefficient. Non-reliable effects included in the model are indicated by "?".
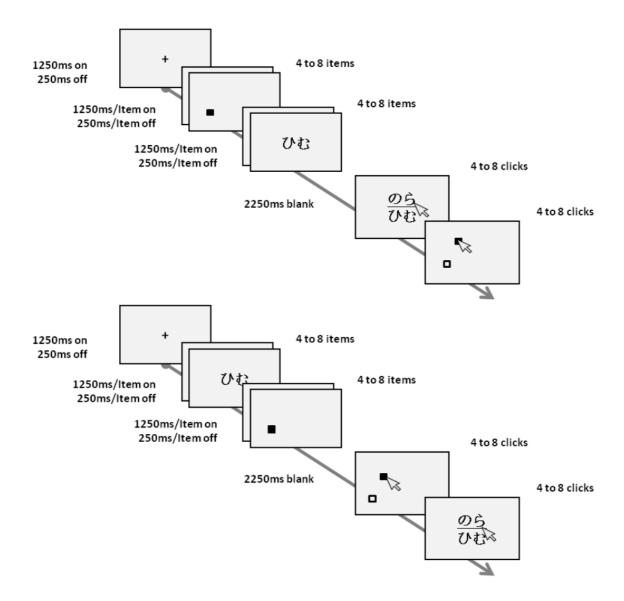
Embedded verbal task

Analysis	DIC	Verbal	Outer	Domain	Verbal x Outer	Verbal x Domain	Outer x Domain	Verbal x Outer x Domain
	10662	-	-	-				?
Raw	10663	-	?	-	?			?
	10664	-	-	-		?		?
	2791	-	-	-			+	
Strict	2794	-	-	-			+	?
	2795	-	-	?		?	+	

Embedded spatial task

Model	DIC	Spatial	Outer	Domain	Spatial x Outer	Spatial x Domain	Outer x Domain	Spatial x Outer x Domain
	10974	-	-	-	?			+
Raw	10974	-	-	-	+	+	+	
	10975	-	-	-	+	?		+
	4045	-		-		+		
Strict	4045	-	?	-		+		
	4045	-	?	-	?	+		

4) Table 4: Results of the model selection procedure for the response time data in Experiment 2. Shown are the models with the three lowest DICs. In the strict analyses (lower part), only trials with 100% accuracy in both tasks were included. In the raw analyses (upper part) all trials were included. Reliable effects in these models are indicated by "+" for a positive coefficient and "-" for a negative coefficient. Non-reliable effects included in the model are indicated by "?".
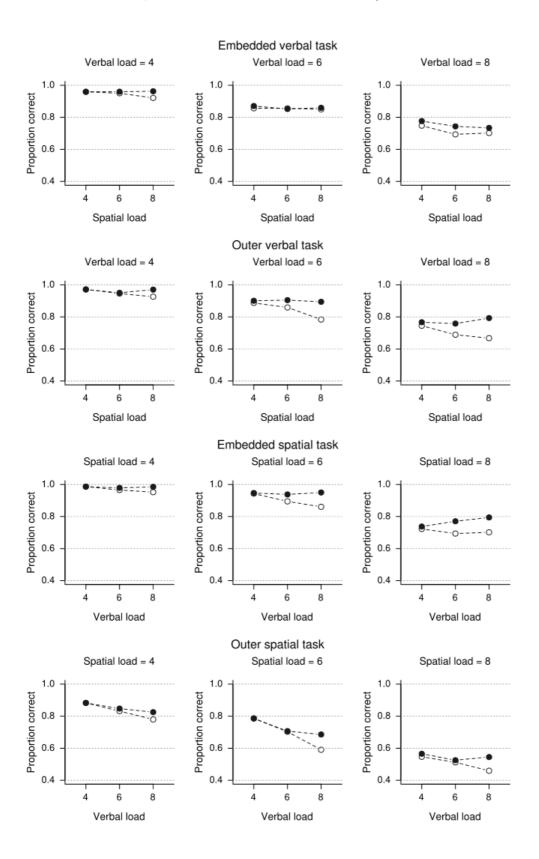
Embedded verbal task								
Analysis	DIC	Verbal	Outer	Domain	Verbal x Outer	Verbal x Domain	Outer x Domain	Verbal x Outer x Domain
Raw	11139	+	?	?		+		-
	11140	+	?	+			?	?
	11140	+	?	?	?	+		-
Strict	2598	+	?	+	+			-
	2599	+	?	+	+		?	?
	2599	+	?	+	+	?	-	

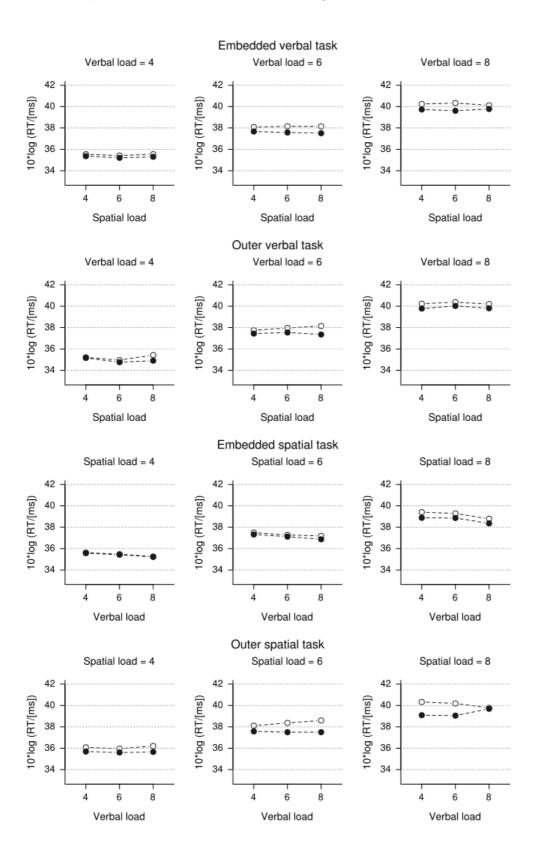| **Embedded spatial task**								
Model		Spatial	Outer	Domain	Spatial x Outer	Spatial x Domain	Outer x Domain	Spatial x Outer x Domain
Raw	8951	+	?	?	-		+	
	8953	+						
	8953	+						
Strict	1949	+						
	1950	+	?					
	1950	+		?				

Figures

1) Figure 1: Trial flow in Experiment 1 (upper panel = embedded-verbal condition, lower panel = embedded-spatial condition).
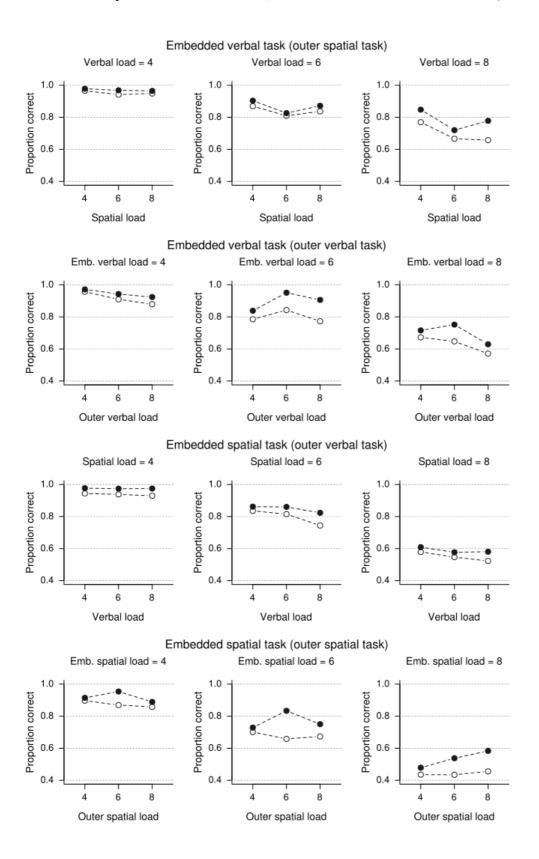
2) Figure 2: Serial reconstruction accuracy in Experiment 1 separately plotted for each combination of embedded and outer task loads (filled circles = trials with 100% accuracy in the concurrent task, unfilled circles = all trials included).
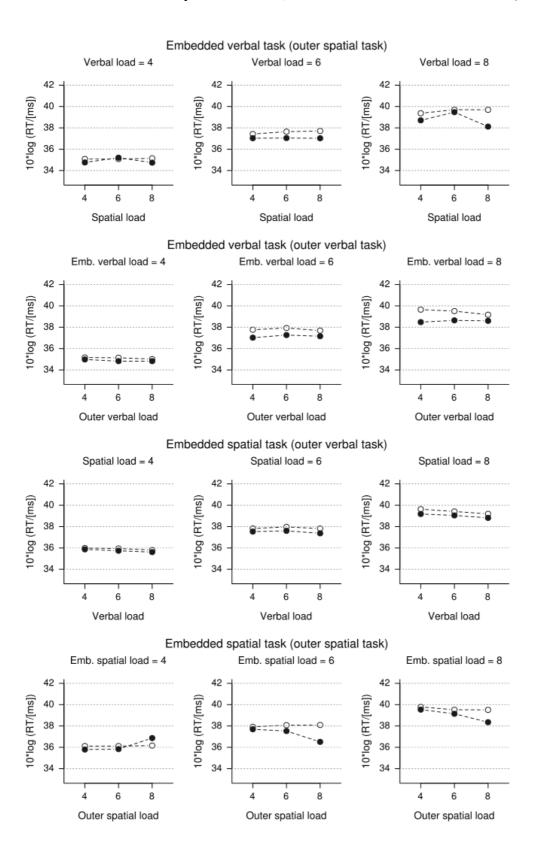
3) Figure 3: Serial reconstruction response times in Experiment 1 separately plotted for each combination of embedded and outer task loads (filled circles = trials with 100% accuracy for both tasks, unfilled circles = all trials included).

4) Figure 4: Serial reconstruction accuracy in the embedded tasks in Experiment 2 separately plotted for each combination of embedded and outer task loads (filled circles = trials with 100% accuracy in the concurrent task, unfilled circles = all trials included).

5) Figure 5: Serial reconstruction response times for the embedded tasks in Experiment 2 separately plotted for each combination of embedded and outer task loads (filled circles = trials with 100% accuracy for both tasks, unfilled circles = all trials included).

Supplement captions

A1) Descriptive statistics for the accuracy data in Experiment 1. Shown are the number of trials, proportion correct, and standard error for each combination of embedded and outer task loads. In the strict analyses (lower part), only trials with 100% accuracy in the concurrent task were included. In the raw analyses (upper part) all trials were included.

Embedded verbal task

Embedded Load	4	4	4	6	6	6	8	8	8
Outer Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	230	230	230	230	230	230	230	230	229
Prop. Correct (Raw)	0.96	0.95	0.92	0.86	0.86	0.85	0.75	0.69	0.70
Standard Error (Raw)	0.01	0.01	0.02	0.02	0.02	0.02	0.03	0.03	0.03
Trials (Strict)	195	142	53	173	120	46	166	82	44
Prop. Correct (Strict)	0.96	0.96	0.96	0.87	0.85	0.86	0.78	0.74	0.73
Standard Error (Strict)	0.01	0.02	0.03	0.03	0.03	0.05	0.03	0.05	0.07

Outer verbal task

Outer Load	4	4	4	6	6	6	8	8	8
Embedded Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	230	230	230	230	230	230	230	230	230
Prop. Correct (Raw)	0.97	0.95	0.93	0.89	0.86	0.78	0.75	0.69	0.67
Standard Error (Raw)	0.01	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.03
Trials (Strict)	224	201	101	216	181	91	213	167	87
Prop. Correct (Strict)	0.97	0.95	0.97	0.90	0.91	0.89	0.77	0.76	0.79
Standard Error (Strict)	0.01	0.02	0.02	0.02	0.02	0.03	0.03	0.03	0.04

Embedded spatial task

Embedded Load	4	4	4	6	6	6	8	8	8
Outer Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	230	230	230	230	230	230	230	230	230
Prop. Correct (Raw)	0.99	0.97	0.95	0.94	0.90	0.86	0.72	0.69	0.70
Standard Error (Raw)	0.01	0.01	0.01	0.02	0.02	0.02	0.03	0.03	0.03
Trials (Strict)	219	179	119	209	168	103	203	139	100
Prop. Correct (Strict)	0.99	0.98	0.99	0.95	0.94	0.95	0.74	0.77	0.79
Standard Error (Strict)	0.01	0.01	0.01	0.02	0.02	0.02	0.03	0.04	0.04

Outer spatial task

Outer Load	4	4	4	6	6	6	8	8	8
Embedded Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	230	230	230	230	230	230	230	230	229
Prop. Correct (Raw)	0.88	0.83	0.78	0.79	0.70	0.59	0.55	0.51	0.46
Standard Error (Raw)	0.02	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
Trials (Strict)	213	162	110	210	159	90	199	154	98
Prop. Correct (Strict)	0.88	0.85	0.83	0.79	0.71	0.69	0.57	0.53	0.55
Standard Error (Strict)	0.02	0.03	0.04	0.03	0.04	0.05	0.04	0.04	0.05

A2) Descriptive statistics for the response time data in Experiment 1. Shown are the number of trials, mean response time, and standard error for each combination of embedded and outer task loads. In the strict analyses (lower part), only trials with 100% accuracy in the concurrent task were included. In the raw analyses (upper part) all trials were included.

Embedded verbal task

Embedded Load	4	4	4	6	6	6	8	8	8
Outer Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	230	230	230	230	230	230	230	230	229
$10*\log_{10}RT$ (Raw)	35.54	35.42	35.54	38.08	38.15	38.15	40.24	40.33	40.12
Standard Error (Raw)	0.08	0.08	0.09	0.09	0.09	0.09	0.09	0.09	0.09
Trials (Strict)	181	132	49	125	82	31	86	42	25
$10*\log_{10}RT$ (Strict)	35.36	35.21	35.30	37.67	37.58	37.51	39.72	39.61	39.77
Standard Error (Strict)	0.07	0.09	0.19	0.10	0.12	0.21	0.12	0.13	0.20

Outer verbal task

Outer Load	4	4	4	6	6	6	8	8	8
Embedded Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	230	230	230	230	230	230	230	230	230
$10*\log_{10}RT$ (Raw)	35.21	34.96	35.43	37.75	37.96	38.15	40.23	40.38	40.20
Standard Error (Raw)	0.09	0.10	0.13	0.09	0.09	0.11	0.10	0.10	0.10
Trials (Strict)	213	184	95	173	145	70	116	90	56
$10*\log_{10}RT$ (Strict)	35.16	34.76	34.90	37.44	37.54	37.35	39.78	40.03	39.81
Standard Error (Strict)	0.08	0.09	0.16	0.09	0.10	0.14	0.11	0.16	0.14

Embedded spatial task

Outer Load	4	4	4	6	6	6	8	8	8
Embedded Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	230	230	230	230	230	230	230	230	230
$10*\log_{10}RT$ (Raw)	35.62	35.48	35.26	37.49	37.27	37.19	39.41	39.29	38.79
Standard Error (Raw)	0.07	0.07	0.06	0.08	0.07	0.06	0.09	0.09	0.07
Trials (Strict)	213	173	116	184	145	90	95	70	56
$10*\log_{10}RT$ (Strict)	35.57	35.42	35.22	37.32	37.12	36.87	38.88	38.86	38.36
Standard Error (Strict)	0.07	0.07	0.08	0.07	0.07	0.09	0.13	0.14	0.11

Outer spatial task

Outer Load	4	4	4	6	6	6	8	8	8
Embedded Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	230	230	230	230	230	230	230	230	229
$10*\log_{10}RT$ (Raw)	36.08	35.96	36.20	38.11	38.37	38.59	40.32	40.19	39.78
Standard Error (Raw)	0.12	0.12	0.13	0.10	0.13	0.12	0.12	0.12	0.11
Trials (Strict)	181	125	86	132	82	42	49	31	25
$10*\log_{10}RT$ (Strict)	35.69	35.60	35.65	37.57	37.50	37.50	39.09	39.05	39.69
Standard Error (Strict)	0.11	0.13	0.16	0.11	0.15	0.18	0.21	0.25	0.31

A3) Descriptive statistics for the accuracy data in the embedded tasks in Experiment 2. Shown are the number of trials, proportion correct, and standard error for each combination of embedded and outer task loads. In the strict analyses (lower part), only trials with 100% accuracy in the concurrent task were included. In the raw analyses (upper part) all trials were included.

Embedded verbal task (outer spatial task)

Embedded Load	4	4	4	6	6	6	8	8	8
Outer Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	180	180	180	180	180	180	180	180	177
Prop. Correct (Raw)	0.97	0.94	0.95	0.87	0.81	0.84	0.77	0.67	0.66
Standard Error (Raw)	0.01	0.02	0.02	0.03	0.03	0.03	0.03	0.04	0.04
Trials (Strict)	137	71	14	128	45	13	91	33	9
Prop. Correct (Strict)	0.98	0.97	0.96	0.90	0.83	0.87	0.85	0.72	0.78
Standard Error (Strict)	0.01	0.02	0.05	0.03	0.06	0.09	0.04	0.08	0.14

Embedded verbal task (outer verbal task)

Embedded Load	4	4	4	6	6	6	8	8	8
Outer Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	180	180	180	180	180	180	180	180	180
Prop. Correct (Raw)	0.96	0.91	0.88	0.79	0.84	0.77	0.67	0.65	0.57
Standard Error (Raw)	0.02	0.02	0.02	0.03	0.03	0.03	0.04	0.04	0.04
Trials (Strict)	141	70	46	96	68	39	91	65	30
Prop. Correct (Strict)	0.97	0.94	0.92	0.84	0.95	0.91	0.72	0.75	0.63
Standard Error (Strict)	0.01	0.03	0.04	0.04	0.03	0.05	0.05	0.05	0.09

Embedded spatial task (outer verbal task)

Embedded Load	4	4	4	6	6	6	8	8	8
Outer Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	174	174	174	174	174	174	174	174	174
Prop. Correct (Raw)	0.94	0.94	0.93	0.84	0.82	0.74	0.58	0.55	0.52
Standard Error (Raw)	0.02	0.02	0.02	0.03	0.03	0.03	0.04	0.04	0.04
Trials (Strict)	150	124	89	153	137	82	147	120	87
Prop. Correct (Strict)	0.98	0.97	0.97	0.86	0.86	0.82	0.61	0.58	0.58
Standard Error (Strict)	0.01	0.01	0.02	0.03	0.03	0.04	0.04	0.05	0.05

Embedded spatial task (outer spatial task)

Embedded Load	4	4	4	6	6	6	8	8	8
Outer Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	174	174	174	174	174	174	174	173	168
Prop. Correct (Raw)	0.90	0.87	0.86	0.70	0.66	0.67	0.44	0.43	0.46
Standard Error (Raw)	0.02	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.04
Trials (Strict)	70	27	9	59	21	8	52	20	6
Prop. Correct (Strict)	0.91	0.95	0.89	0.73	0.83	0.75	0.48	0.54	0.58
Standard Error (Strict)	0.03	0.04	0.10	0.06	0.08	0.15	0.07	0.11	0.20

A4) Descriptive statistics for the response time data in the embedded tasks in Experiment 2. Shown are the number of trials, mean response time, and standard error for each combination of embedded and outer task loads. In the strict analyses (lower part), only trials with 100% accuracy in the concurrent task were included. In the raw analyses (upper part) all trials were included.

Embedded verbal task (outer spatial task)

Embedded Load	4	4	4	6	6	6	8	8	8
Outer Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	180	180	180	180	180	180	180	180	177
$10*\log_{10}RT$ (Raw)	35.08	35.08	35.14	37.42	37.65	37.71	39.36	39.70	39.69
Standard Error (Raw)	0.11	0.10	0.11	0.11	0.11	0.11	0.11	0.12	0.12
Trials (Strict)	131	67	13	100	29	10	58	17	4
$10*\log_{10}RT$ (Strict)	34.75	35.21	34.73	37.03	37.04	37.02	38.70	39.47	38.12
Standard Error (Strict)	0.10	0.16	0.28	0.12	0.20	0.29	0.11	0.40	0.61

Embedded verbal task (outer verbal task)

Embedded Load	4	4	4	6	6	6	8	8	8
Outer Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	180	180	180	180	180	180	180	180	180
$10*\log_{10}RT$ (Raw)	35.16	35.14	35.01	37.76	37.93	37.69	39.65	39.51	39.17
Standard Error (Raw)	0.10	0.09	0.12	0.14	0.14	0.14	0.14	0.15	0.17
Trials (Strict)	134	62	39	64	61	29	38	31	11
$10*\log_{10}RT$ (Strict)	34.98	34.82	34.82	37.02	37.27	37.16	38.48	38.64	38.60
Standard Error (Strict)	0.11	0.15	0.20	0.19	0.16	0.28	0.20	0.21	0.37

Embedded spatial task (outer verbal task)

Embedded Load	4	4	4	6	6	6	8	8	8
Outer Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	174	174	174	174	174	174	174	174	174
$10*\log_{10}RT$ (Raw)	35.99	35.95	35.81	37.82	37.97	37.82	39.63	39.42	39.19
Standard Error (Raw)	0.09	0.10	0.09	0.10	0.10	0.09	0.10	0.09	0.09
Trials (Strict)	144	118	85	108	97	55	37	31	21
$10*\log_{10}RT$ (Strict)	35.87	35.74	35.61	37.53	37.60	37.37	39.18	39.05	38.82
Standard Error (Strict)	0.09	0.09	0.10	0.07	0.10	0.14	0.18	0.16	0.25

Embedded spatial task (outer spatial task)

Embedded Load	4	4	4	6	6	6	8	8	8
Outer Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	174	174	174	174	174	174	174	173	168
$10*\log_{10}RT$ (Raw)	36.12	36.12	36.16	37.92	38.07	38.09	39.78	39.52	39.51
Standard Error (Raw)	0.10	0.08	0.09	0.09	0.10	0.10	0.10	0.08	0.09
Trials (Strict)	60	25	7	31	13	3	8	3	1
$10*\log_{10}RT$ (Strict)	35.79	35.83	36.87	37.69	37.52	36.51	39.53	39.14	38.35
Standard Error (Strict)	0.12	0.16	0.40	0.17	0.24	0.87	0.37	0.57	-

A5) Descriptive statistics for the accuracy data in the outer tasks in Experiment 2. Shown are the number of trials, proportion correct, and standard error for each combination of embedded and outer task loads. In the strict analyses (lower part), only trials with 100% accuracy in the concurrent task were included. In the raw analyses (upper part) all trials were included.

Outer spatial task (embedded verbal task)

Outer Load	4	4	4	6	6	6	8	8	8
Embedded Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	180	180	180	180	180	180	180	180	177
Prop. Correct (Raw)	0.84	0.78	0.64	0.62	0.50	0.45	0.36	0.35	0.30
Standard Error (Raw)	0.03	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.03
Trials (Strict)	169	128	91	162	111	63	162	123	69
Prop. Correct (Strict)	0.85	0.83	0.74	0.64	0.52	0.53	0.37	0.37	0.36
Standard Error (Strict)	0.03	0.03	0.05	0.04	0.05	0.06	0.04	0.04	0.06

Outer verbal task (embedded verbal task)

Outer Load	4	4	4	6	6	6	8	8	8
Embedded Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	180	180	180	180	180	180	180	180	180
Prop. Correct (Raw)	0.86	0.68	0.65	0.62	0.61	0.55	0.56	0.51	0.39
Standard Error (Raw)	0.03	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
Trials (Strict)	169	109	65	153	129	60	146	102	47
Prop. Correct (Strict)	0.86	0.72	0.71	0.63	0.69	0.66	0.57	0.56	0.48
Standard Error (Strict)	0.03	0.04	0.06	0.04	0.04	0.06	0.04	0.05	0.07

Outer verbal task (embedded spatial task)

Outer Load	4	4	4	6	6	6	8	8	8
Embedded Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	174	174	174	174	174	174	174	174	174
Prop. Correct (Raw)	0.92	0.92	0.89	0.87	0.88	0.83	0.76	0.75	0.74
Standard Error (Raw)	0.02	0.02	0.02	0.03	0.02	0.03	0.03	0.03	0.03
Trials (Strict)	159	116	41	157	110	39	154	95	37
Prop. Correct (Strict)	0.95	0.96	0.92	0.89	0.94	0.89	0.79	0.81	0.77
Standard Error (Strict)	0.02	0.02	0.04	0.03	0.02	0.05	0.03	0.04	0.07

Outer spatial task (embedded spatial task)

Outer Load	4	4	4	6	6	6	8	8	8
Embedded Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	174	174	174	174	174	173	174	174	168
Prop. Correct (Raw)	0.54	0.51	0.48	0.44	0.38	0.31	0.31	0.29	0.25
Standard Error (Raw)	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.03
Trials (Strict)	144	89	20	135	75	23	138	75	24
Prop. Correct (Strict)	0.55	0.49	0.54	0.48	0.41	0.34	0.31	0.30	0.32
Standard Error (Strict)	0.04	0.05	0.11	0.04	0.06	0.10	0.04	0.05	0.10

A6) Descriptive statistics for the response time data in the outer tasks in Experiment 2. Shown are the number of trials, mean response time, and standard error for each combination of embedded and outer task loads. In the strict analyses (lower part), only trials with 100% accuracy in the concurrent task were included. In the raw analyses (upper part) all trials were included.

Outer spatial task (embedded verbal task)

Outer Load	4	4	4	6	6	6	8	8	8
Embedded Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	180	180	180	180	180	180	180	180	177
$10*\log_{10}RT$ (Raw)	36.51	36.68	36.92	39.00	39.11	39.10	40.50	40.32	40.53
Standard Error (Raw)	0.16	0.14	0.14	0.14	0.14	0.13	0.14	0.13	0.15
Trials (Strict)	131	100	58	67	29	17	13	10	4
$10*\log_{10}RT$ (Strict)	35.86	36.30	36.58	38.68	38.11	38.20	39.87	39.51	39.08
Standard Error (Strict)	0.14	0.17	0.27	0.22	0.28	0.32	0.44	0.42	0.41

Outer verbal task (embedded verbal task)

Outer Load	4	4	4	6	6	6	8	8	8
Embedded Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	180	180	180	180	180	180	180	180	180
$10*\log_{10}RT$ (Raw)	35.71	36.61	36.74	38.80	38.76	38.59	40.21	40.36	40.02
Standard Error (Raw)	0.17	0.18	0.18	0.18	0.18	0.21	0.15	0.15	0.22
Trials (Strict)	134	64	38	62	61	31	39	29	11
$10*\log_{10}RT$ (Strict)	35.22	34.98	35.71	37.56	37.65	38.77	39.36	39.16	41.03
Standard Error (Strict)	0.16	0.22	0.38	0.20	0.28	0.45	0.25	0.24	0.61

Outer verbal task (embedded spatial task)

Outer Load	4	4	4	6	6	6	8	8	8
Embedded Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	174	174	174	174	174	174	174	174	174
$10*\log_{10}RT$ (Raw)	35.29	35.14	35.38	37.77	37.62	37.79	39.80	39.87	39.60
Standard Error (Raw)	0.13	0.14	0.16	0.14	0.15	0.15	0.15	0.14	0.15
Trials (Strict)	144	108	37	118	97	31	85	55	21
$10*\log_{10}RT$ (Strict)	34.99	34.73	34.89	37.10	36.94	37.33	38.78	38.94	38.20
Standard Error (Strict)	0.12	0.13	0.24	0.13	0.15	0.22	0.13	0.20	0.30

Outer spatial task (embedded spatial task)

Outer Load	4	4	4	6	6	6	8	8	8
Embedded Load	4	6	8	4	6	8	4	6	8
Trials (Raw)	174	174	174	174	174	173	174	174	168
$10*\log_{10}RT$ (Raw)	36.87	37.23	37.23	39.28	39.27	38.99	40.68	40.57	40.24
Standard Error (Raw)	0.15	0.15	0.14	0.16	0.13	0.13	0.13	0.13	0.13
Trials (Strict)	60	31	8	25	13	3	7	3	1
$10*\log_{10}RT$ (Strict)	36.14	37.05	35.91	38.37	39.18	37.54	40.76	40.23	40.61
Standard Error (Strict)	0.18	0.33	0.42	0.36	0.50	1.59	0.53	0.64	-