Available online at www.sciencedirect.com
ScienceDirect
Journal of Memory and Language 55 (2006) 601–626
Journal of Memory and Language
www.elsevier.com/locate/jml

A formal model of capacity limits in working memory ☆

Klaus Oberauer [a,*], Reinhold Kliegl [b]

[a] Department of Experimental Psychology, University of Bristol, 12A Priory Road, Bristol BS8 1TU, UK
[b] University of Potsdam, Germany

Received 2 March 2006; revision received 2 August 2006

Abstract

A mathematical model of working-memory capacity limits is proposed on the key assumption of mutual interference between items in working memory. Interference is assumed to arise from overwriting of features shared by these items. The model was fit to time-accuracy data of memory-updating tasks from four experiments using nonlinear mixed effect (NLME) models as a framework. The model gave a good account of the data from a numerical and a spatial task version. The performance pattern in a combination of numerical and spatial updating could be explained by variations in the interference parameter: assuming less feature overlap between contents from different domains than between contents from the same domain, the model can account for double dissociations of content domains in dual-task experiments. Experiment 3 extended this idea to similarity within the verbal domain. The decline of memory accuracy with increasing memory load was steeper with phonologically similar than with dissimilar material, although processing speed was faster for the similar material. The model captured the similarity effects with a higher estimated interference parameter for the similar than for the dissimilar condition. The results are difficult to explain with alternative models, in particular models incorporating time-based decay and models assuming limited resource pools.

Keywords: Working memory; Interference; Capacity; Mathematical model; Non-linear mixed effects

Introduction

It is a common and sometimes annoying experience—human cognitive capacities are limited. Problem solving is hard work and fails every so often; complex sentences must be read twice and still escape comprehension; logical and mathematical derivations need to be double-checked because humans, unlike computers, are highly error-prone calculators. We observe that some people cope better with hard reasoning problems and others worse, but individual variability in performance only underscores the general fact that we all experience severe limitations in our thinking. What is it that constrains our cognitive abilities?

A preliminary answer is that most complex tasks require working memory, which has a severely limited capacity (e.g., Carpenter, Just, & Shell, 1990; Cowan, 2005). Working memory can be characterized as a system to remember several briefly presented items (e.g.,

☆ This research was funded by Deutsche Forschungsgemeinschaft (DFG), Grant INK 12, Project C, and Grant KL 955/4-1. We thank Petra Grüttner and Hannelore Gensel for their assistance in conducting the experiments, and Mike Page and two anonymous reviewers for their helpful comments on a previous version of the paper.
* Corresponding author.
E-mail address: k.oberauer@bristol.ac.uk (K. Oberauer).

0749-596X/$ - see front matter
doi:10.1016/j.jml.2006.08.009

words, digits, pictures of objects, spatial locations) and to manipulate these or other items at the same time (Baddeley, 1986). The term capacity limit, as we understand it here, refers to the observation that people's performance declines rapidly with an increase in memory demand in a wide variety of experimental tasks. By memory demand we mean the number of independent items that must be held simultaneously available for processing. The capacity limit of working memory appears to be highly general, as evidenced by the large amount of common variance shared between many different tasks (Kane et al., 2004; Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000). Furthermore, working memory capacity is of central importance for complex cognition, as shown by its strong correlation with reasoning ability (Engle, Tuholski, Laughlin, & Conway, 1999; Kyllonen & Christal, 1990; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002). Nonetheless, explaining limited cognitive abilities with recourse to working memory is only a preliminary answer as long as we do not know why working memory capacity is limited. The purpose of this article is to suggest an answer to this question.

The capacity limit of working memory is traditionally explained by limited resources. For example, Just and Carpenter (1992) assume that working memory has a limited pool of activation that must be shared for all memory and processing tasks within broad content domains (e.g., there is one resource pool for verbal tasks and another one for spatial tasks). Similarly, Anderson, Reder, and Lebiere (1996) attribute capacity limits to a limited amount of source activation that must be shared among those chunks that are held in working memory at any time.

Resource accounts of capacity limits in human cognition have been criticized for being too unconstrained and empirically empty (Navon, 1984). Several alternatives to resource limits have been proposed to explain the limited capacity of working memory. Several authors have attributed forgetting from working memory to rapid time-based decay (Baddeley, Thomson, & Buchanan, 1975; Barrouillet, Bernardin, & Camos, 2004; Page & Norris, 1998; Kieras, Meyer, Mueller, & Seymour, 1999). Others have argued that the capacity limit arises from mutual interference between representations held in working memory (Nairne, 1990; Saito & Miyake, 2004), or that it reflects a limitation of the focus of attention to be directed to a maximum of about four chunks (Cowan, 2001).

In a previous paper (Oberauer & Kliegl, 2001), we formalized several alternative accounts of working memory capacity within a common framework, varying only the critical theoretical assumptions that are specific to each account. These models were evaluated with a particularly rich data set: all models were fit to individual time-accuracy functions for 16 young and 17 old adults for six levels of memory demand of a numerical memory-updating task. It turned out that only two models, one based on decay of memory traces over time and the other one based on interference between partially overlapping representations, could fit the data. The interference model (described below) was the more parsimonious one in terms of the number of free parameters, and it had a slightly better fit. The purpose of the present paper is to present an improved version of the interference model, to extend its domain of application by fitting it to new data, and to test the model with experimental manipulations assumed to affect its most important parameter, the degree of mutual interference between two representations.

We will apply the models to variants of the memory-updating task, originally designed by Salthouse, Babcock, and Shaw (1991). Each trial starts with the presentation of a variable number of initial memory items. Each item is set in its own frame on the computer screen (see Fig. 1 for examples). Items are digits for the numerical version and dots at particular locations within the frames for the spatial version. Participants are required to remember which item is displayed in which frame. They are then required to update individual items according to operations displayed one after another, each in one of the frames. The operations are single-digit additions or subtractions in the numerical version, and arrows indicating mental shifts of the dots within their frames in the spatial version. In the numerical example of Fig. 1, for instance, the first updating operation appears in the left frame. The person must retrieve the current digit remembered for the left frame ("3"), apply the arithmetic operation to it, and then replace the previous digit with the result ("5") in working memory. At the end of each trial, the final results for each frame must be recalled. Thus, throughout the updating sequence participants must remember one item in each frame. The number of items to be remembered at any time defines the memory demand (MD) of a task. In the factor-analytic study of Oberauer et al. (2000), the memory-updating task with digits had a high loading on the verbal–numerical working memory factor, whereas the spatial version loaded highly on the spatial working memory factor. Hence, the task is a valid measure of working memory capacity.

All experiments reported here determined time-accuracy functions (TAFs) for individual participants and experimental conditions. To this end, we tested each participant on a large number of trials in each condition. The presentation time for updating operations was varied across these trials, ranging from very short times yielding accuracies close to chance to six seconds per operation, where accuracy reaches an asymptotic maximum level. The performance of a person in a specific condition can then be described as a function relating accuracy of reporting the final values to presentation

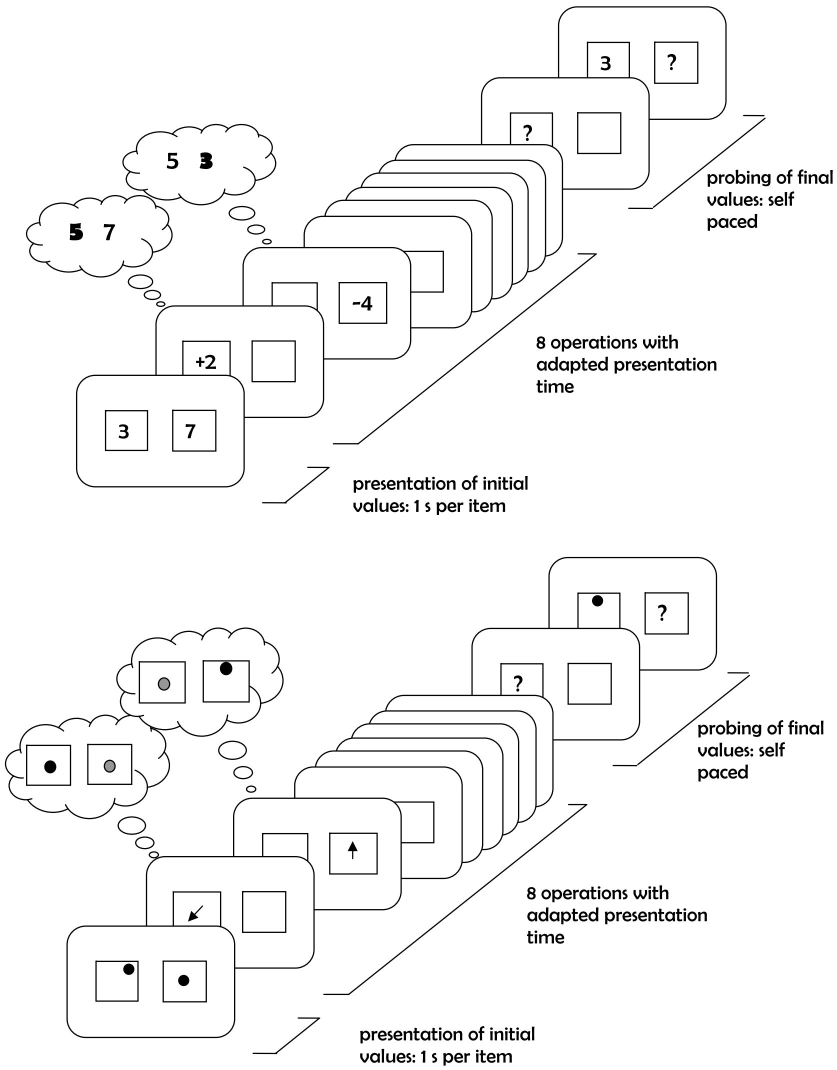

Fig. 1. Task examples for memory-updating with digits (top) and spatial positions (bottom), with a memory demand of two elements. Thought bubbles illustrate the content of working memory after completing the first and the second updating step; in the numerical task the just updated digit is printed in bold, and in the spatial task, the just updated dot is filled in black.

time for individual operations (cf. Kliegl, Mayr, & Krampe, 1994).

The remainder of this article is organized as follows: first, we introduce our strategy of data analysis and model fitting, nonlinear mixed effects (NLME) models. Next, we describe an experiment measuring TAFs with the spatial version of the memory-updating task. Then we introduce the improved interference model and apply it to the data of Experiment 1, as well as to the data from the parallel numerical task version used by Oberauer and Kliegl (2001). After that, we present two experiments with manipulations of one critical parameter in the model, that is, the degree of overlap between representations held in working memory, and fit the model to these data.

Nonlinear mixed effects (NLME) models

The data from our experiments can be described with respect to the effects of experimental conditions on the sample mean (e.g., fixed effects of presentation time, storage demand, and item similarity on accuracy) and with respect to interindividual differences in these effects (i.e., so-called random effects). The adequate method of analysis and modeling for this kind of data is multi-level regression. Moreover, time-accuracy curves are known to be nonlinear—previous studies have shown that the increase of accuracy over processing time can be described well by a negatively accelerated exponential function (Kliegl et al., 1994). Therefore, we used nonlinear mixed effects (NLME) models (Pinheiro & Bates, 2000) as a framework for data analysis and model fitting. NLME is a nonlinear extension of multilevel regression models to nonlinear regression. Parameters are estimated simultaneously as fixed or random effects. Each fixed effect translates into a parameter mean (across participants) of a linear or nonlinear regression equation. Its associated random effect represents its variability across individuals; it can be omitted if there is no significant variability. Random effects are assumed to be normally distributed with a mean of zero around their fixed effects. Importantly, instead of estimating separate parameters for each individual, the model estimates the variances and covariances of those parameters for which random effects are specified. Thus, the number of parameters estimated is independent of the number of participants in an experiment, and obviously much smaller compared to fitting individual participants' data.

Nevertheless, we can obtain "estimates" of random effects for each individual (i.e., actually a prediction of each individual's deviation from the corresponding fixed effect based on the estimated variances and covariances). These random "effects" are called best linear unbiased predictions (BLUPs). BLUPs differ from direct parameter estimates for each individual because they take into account the individual variance and the sample mean: if the individual estimate is very precise, the BLUP will not differ much from the direct estimate; if the estimate is noisy the BLUP will shrink towards the overall estimate of the corresponding fixed effect. Technically, BLUPs are the conditional modes of the random effects, evaluated at the conditional estimate of fixed effects (Pinheiro & Bates, 2000). Thus, an important advantage of NLME (e.g., compared to linear or nonlinear repeated-measures regression analyses) is that it reduces the risk of overfitting the model to unreliable differences between individuals.

We used NLME as a data-analysis tool by fitting descriptive time-accuracy functions (described in the context of Experiment 1) and as a model-fitting tool for applying the interference model to the data. For both purposes we used the *nlme* package (Pinheiro, Bates, DebRoy, & Sarkar, 2005) as provided in the R language and environment for statistical computing (R Development Core Team, 2005).

Experiment 1

The goal of the first experiment was to obtain time-accuracy functions for a spatial version of the memory-updating task, which could be used to test the generality of the interference model. If the model describes the dynamics and capacity limitations of working memory in general, it should be applicable to tasks with spatial content as well as to tasks with verbal or numerical content.

Method

Participants

Twenty-one students from high schools in Potsdam, Germany, participated in the experiment. Their mean age was 18.8 years ($SD = 1.28$); there were 10 men and 11 women. Participants were reimbursed with 12,- DM (about US$ 6) for each 1-h session.

Materials and procedure

A spatial version of memory updating was used. Participants saw n frames arranged on an imaginary circle on the screen, where n corresponded to the memory demand (MD) and varied between one and four (in condition MD 1, the frame was displayed centrally). At the beginning of each trial a dot appeared in each frame in one of nine possible locations within the frame. The locations were defined by an invisible 3×3 grid subdividing each frame. The dots appeared simultaneously for n seconds, and participants were instructed to remember their initial locations. After that, arrows were presented in the middle of the frames, one at a time, to indicate the direction for a mental update of the dot position in the respective frame (see Fig. 1). The arrows could be vertical, horizontal, or diagonal, and participants were required to mentally shift the dot one step in the grid in the direction of the arrow. They were informed that a dot would never leave its frame. After eight mental shift operations, appearing in a clockwise sequence through the n frames, the final positions of all n dots were probed for recall in random order. Participants responded using the number pad on the right of the standard computer keyboard. The nine number keys were mapped consistently to the nine possible locations in a frame, and in addition little icons representing each dot position were attached on top of the keys, covering the numbers.

The procedure was the same as in the experiment with numerical memory updating (Oberauer & Kliegl, 2001). There were 20 blocks for each MD condition. Each block consisted of one trial with a fixed presentation time of 6 s per arrow, and twelve trials with presentation times determined by an adaptive algorithm described below. In total there were 260 trials for each MD condition. Memory demand was fixed within each block and varied over blocks in the sequence 1-4-2-3-3-2-4-1. The first 40 trials in each condition were regarded as practice and excluded from analysis. The complete procedure required eight sessions of about one hour.

During the first four sessions we adapted presentation times for the operations in each trial using Kaernbach's (1991) algorithm, that is we increased times by a certain amount after an error and decreased times after a correct response (taking into account only the first probed item). The proportion of the increment to the decrement determines the accuracy level to which this algorithm converges. We used the algorithm with three different target accuracies (33%, 66%, and more than 66%), alternating between trials so that one of the three targets was applied to every third trial. This way, each participant received items in each condition with a set of presentation times tailored to his or her ability that covered his or her entire range between chance and asymptotic maximum performance. After half the sessions, we determined the 12 presentation times with the highest frequency of trials and selected presentation times for the remaining sessions such that in the end they filled up each of these 12 time categories to a minimum of 10 trials. TAFs were fitted to the mean accuracies of each participant in each experimental condition across these 12 presentation times.

Results and discussion

Data of individual participants are shown in Fig. 2, together with fits of the interference model to be described below.[1] To present data on an aggregated level we had to find a way of averaging the data of individual participants. Because of the adaptive manipulation of presentation times we could not average accuracies for a fixed set of presentation times common to all participants. Therefore, we ordered the 12 presentation times on which each participant was tested in each condition. We then averaged across participants the accuracies as well as the presentation times in each of the 12 presentation time categories. The mean accuracies are plotted as a function of mean presentation times in Fig. 3, together with the mean predictions of the interference model obtained in the same way from predictions for individual participants.

To summarize the data further we fitted descriptive time-accuracy functions (TAFs) to the data of each MD condition. We used negatively accelerated exponential functions of the form

$$p = d + (c - d)(1 - \exp(-b(t - a))), \tag{1}$$

where d is a parameter for chance performance, c represents asymptotic performance, b is the rate of approaching asymptote, and a the point in time t where accuracy p raises above chance.[2] For the present purpose, we fixed d to 1/9, because there were nine possible spatial positions; this leaves three free parameters for each participant and condition.

The TAFs serve the same function here as the General Linear Model in an analysis of variance or linear regression, in that they provide a nonlinear descriptive model through which we can estimate the effects of experimental manipulations. One TAF was fitted to each MD condition and differences between conditions were modeled through differences in the parameters a, b, and c. In the present experiment we tested condition effects through linear and quadratic contrasts of the MD variable. We evaluated the significance of these contrasts by introducing them first as fixed effects, then removing them one by one and testing whether the model fit declined significantly. A contrast was removed permanently if its removal did not lead to a significant loss of fit, as assessed by the Likelihood ratio test in *nlme*. After testing the fixed effects, random effects were added to the model one by one and maintained if they improved the fit significantly. After two or three random effects were added, we tested whether the estimated covariances between the random effects were significant, and if they were not, all covariances were fixed to zero to simplify model estimation.

The best fitting descriptive TAF model resulting from this procedure had 8 fixed effects (the three intercepts of a, b, and c, the three linear contrasts of MD, and the quadratic contrasts on b and c) and seven random effects (the three intercepts, the three linear contrasts, and the quadratic contrast on b). The overall fit of the model was assessed by the adjusted R^2 statistic:

$$R^2_{\text{adj}} = 1 - \frac{\sum_{i=1}^{n}(d_i - \hat{d}_i)^2/(n-k)}{\sum(d_i - \bar{d})^2/(n-1)} \tag{2}$$

where d_i represents the observed values, $\hat{d}_i$ are the predicted values, $\bar{d}$ is the mean, n is the number of data points, and k indicates the number of free parameters (McElree & Dosher, 1989).[3] The final model had an R^2_{adj} of .852.

We extracted two indicators of performance for each participant from the descriptive TAFs. First, parameter c from Eq. (1) reflects the asymptotic accuracy reached when processing time is not externally limited. Second, we computed criterion-referenced presentation times (CPTs) relative to the asymptotes. CPTs can be derived from Eq. (1) by setting p to c times the desired relative criterion k and solving for t:

[1] The raw data of this and the other experiments described in this paper can be obtained from the first author on request.

[2] To constrain intercept and rate to positive values, the equation actually fitted contained exp(a) and exp(b) in the places of a and b, respectively.

[3] The development of R^2 statistics or equivalent indicators of goodness of fit for multilevel regression models is an active field of research (e.g., Roberts & Monaco, 2006). For example, one problem with the formula employed in this article is that the inclusion of level-2 predictors (i.e., subject-level predictors) could theoretically render the adjusted R^2 statistic negative. We included this statistic only for reasons of rough comparability of these results with related earlier research and as an additional index for the comparison of nested models involving different level-1 predictors (i.e., item-level predictors).

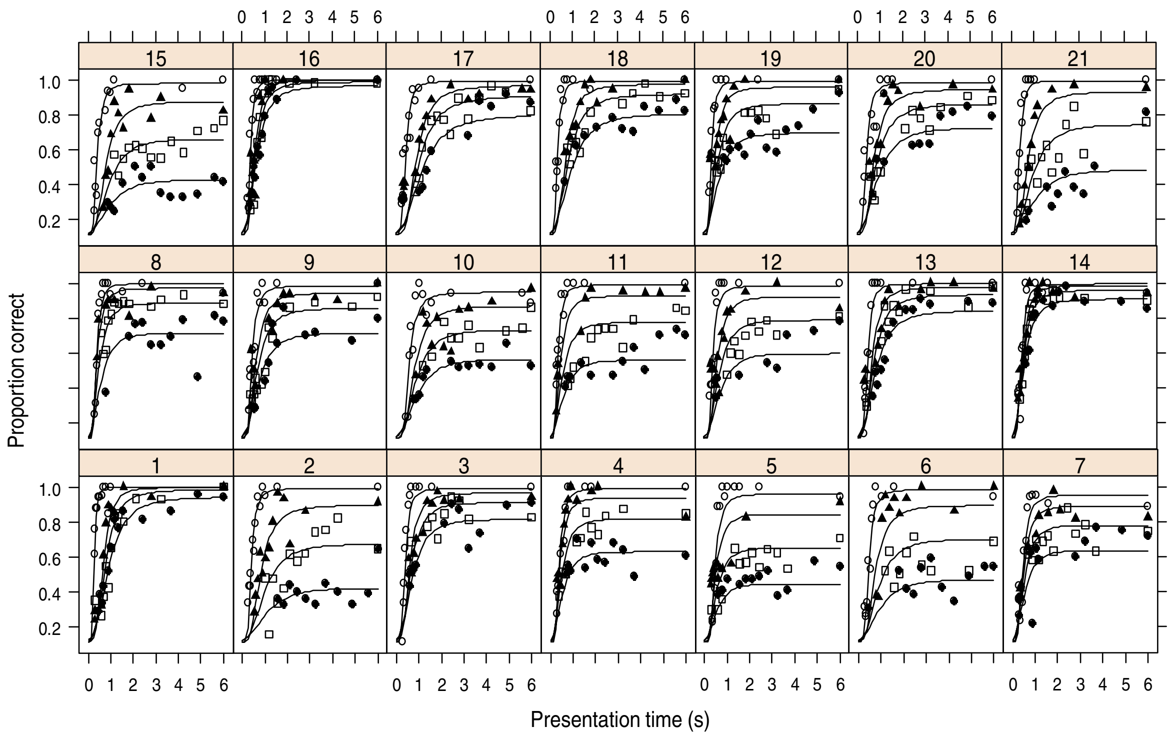

Fig. 2. Data and predictions for individual participants in Experiment 1 (spatial memory updating) generated by NMLE. Data are represented by points (unfilled circles, MD 1; filled triangles, MD 2; unfilled squares, MD 3; filled diamonds, MD 4). Predictions are represented by lines, with MD 1 to MD 4 ordered from left to right in each panel. The figure was produced with the lattice package for *R* (Sarkar, 2005).

$$\mathrm{CPT} = a + b\left[\ln\left(\frac{c-d}{c}\right) - \ln(1-k)\right] \tag{3}$$

Relative CPTs represent the time a participant needs to reach a given proportion k of his or her asymptote in a condition. Thus, CPTs reflect processing speed conditional on asymptotic accuracy. The asymptotes and CPTs were computed for each individual and each MD condition and then averaged within conditions. The asymptotes derived from the descriptive models together with those predicted by the interference model are summarized in the upper panel of Fig. 4; the CPTs for 80% of the asymptote are shown in the lower panel. The data showed the same pattern as that described by Oberauer and Kliegl (2001) for the numerical task version: asymptotic accuracy declined with increasing MD in an accelerated fashion, and relative CPT increased in a roughly linear manner, although in the present data the increase ended at MD 3. This pattern has proven to be highly diagnostic for distinguishing between models of capacity limits in working memory. In particular, the accelerated decline of asymptotic accuracies over MD, together with an increase of relative CPTs, could not be reproduced by models assuming a constant pool of resources and by a model explaining capacity limits through confusion between items at recall (the "cross-talk model"). The best fit was provided by an interference model. We now introduce an improved version of that model.

The interference model

Assumptions

The central assumption of the interference model is that items in working memory interfere with each other through interactions of their features. Items are represented by sets of features that are activated together. Fig. 5 illustrates the representational assumptions of the interference model. The activated features of each item are a relatively small subset of all features in the system (technically speaking, they form sparse distributed representations). Because different items are represented as different patterns of activation across the same set of features, their representations can interact and thereby degrade each other. One such form of

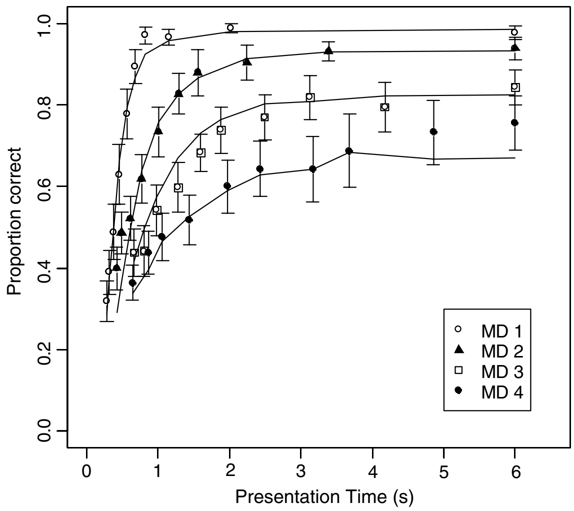

Fig. 3. Data from Experiment 1 (points) and predictions from the interference model (lines), averaged over participants for 12 categories of presentation times. Placement of points on the x-axis reflects mean presentation time in each category. Error bars represent 95% confidence intervals.

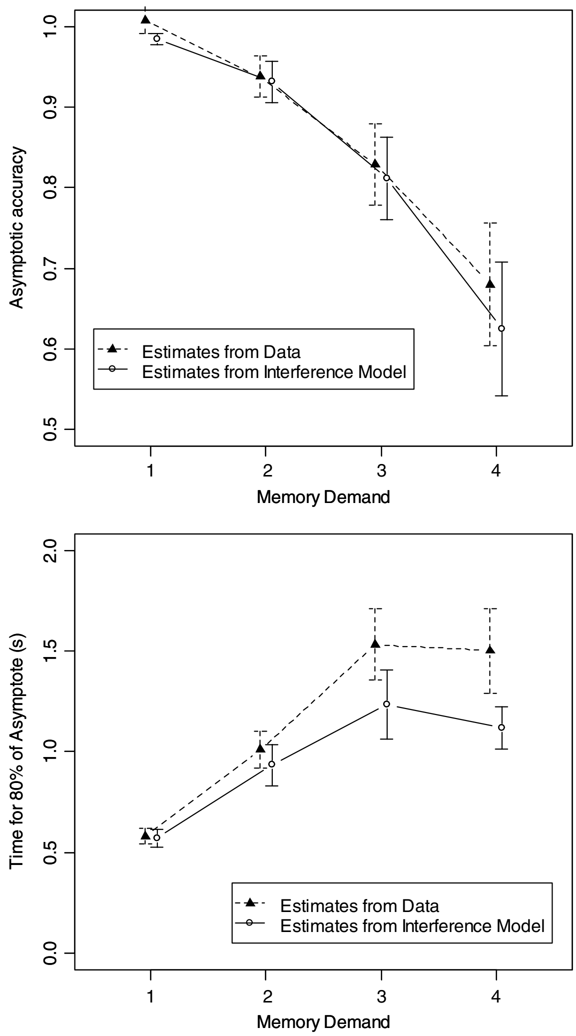

Fig. 4. Asymptotic accuracies (top panel) and time demands for 80% of the asymptote (CPTs) for spatial memory updating (Experiment 1), estimated from fits of descriptive time-accuracy functions to the data and to the predictions of the interference model (BLUPs). Error bars represent 95% confidence intervals.

interaction is feature overwriting (Lange & Oberauer, 2005; Nairne, 1990; Neath, 2000): if two items share a feature, they compete for this feature, which can lead to the loss of that feature in one of the representations. A rationale for feature overwriting can be found in models that use synchronized firing of model neurons as a mechanism of binding together the features that belong to the representation of an item (Raffone & Wolters, 2001). In this kind of models, feature units representing features that belong to the same representation fire at the same time, whereas units belonging to different representations fire out of synchrony. A feature unit can belong to only one representation because it can fire in synchrony with only one set of other feature units.[4]

Between any two items there is a certain degree of feature overlap, with more similar items sharing a larger proportion of feature units. Thus, each representation loses a certain proportion of its feature units through overwriting. The mean proportion of feature units shared between any two items in a set is expressed as parameter C. Thus, when there are two items of that set in working memory at the same time, each one can be expected to lose a mean proportion of $C/2$ of its feature units.[5] In general, when there are n items in working memory at the same time, each item suffers interference from $n - 1$ other items. Assuming that the proportions of overlapping feature units of all pairs in a set of n items are stochastically independent, the proportion of feature units remaining for each item after the loss due to overwriting computes as

[4] It is useful to distinguish between features (e.g., phonemes in a word, or the color of an object) and feature units (i.e., the units in a distributed representation as illustrated in Fig. 5). If a feature is represented by a multitude of units, overlap between two items in that feature does not necessarily result in complete loss of that feature in one of them, but to loss of a subset of the feature units representing that feature in both items.

[5] Readers of Oberauer and Kliegl (2001) should be aware that in that paper we used parameter C to represent the average proportion of overwritten features, which we express as $C/2$ here, and this explains why our previous estimates for C in young adults were about half of the present estimates.

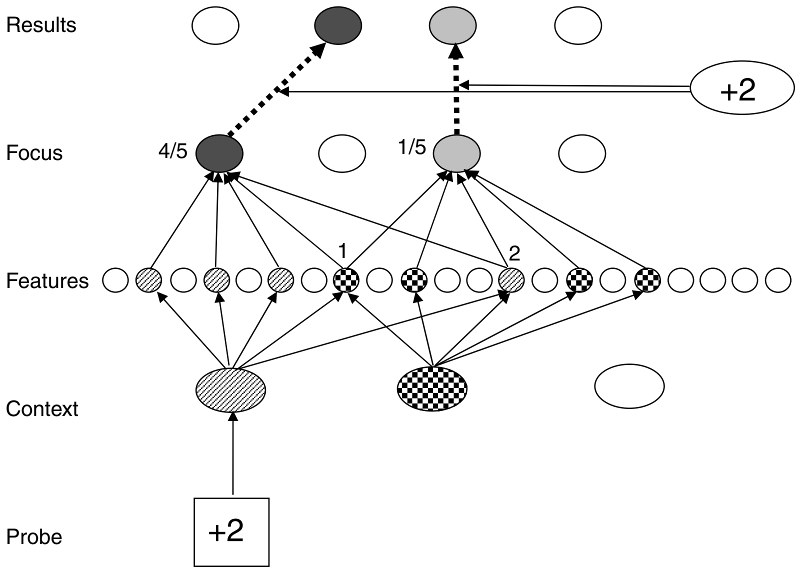

Fig. 5. Illustration of the assumptions of the interference model. The model consists of four layers of units, one of which (the feature layer) is assumed to have sparse distributed representations. The other layers are sketched as using localist representations, but this is not meant to rule out that they use distributed representations as well. The items currently held in working memory are represented as sets of active feature units in the feature layer. Feature units belonging to the same item are bound together by firing in synchrony with each other (i.e., sharing a common phase of firing). They also fire in synchrony with a context unit in the context layer; this binds each item to a representation of its location on the screen (i.e., its frame). This binding mechanism is illustrated by the patterning of activated feature units (hatched vs. checkered)—units filled with the same pattern are firing in synchrony with each other, as well as with the context unit they are bound to. The figure illustrates a state in which two elements have been encoded; these elements overlap in two feature units (labeled 1 and 2) out of their five feature units, so that $C = 2/5$. Feature unit 1 is grabbed by the first item, whereas feature unit 2 is grabbed by the second item. Therefore, each item loses one of its five units due to overwriting. Each feature unit belonging to an item is associated with a representation of that item in the focus layer; the focus layer represents the current content of the focus of attention in working memory. At retrieval, the retrieval cue selects the corresponding context unit, which in turn selects all feature units in the feature layer that fire in its phase (i.e., have the same fill pattern in the figure). A gating mechanism allows these feature units to pass on activation to the focus units to which they are connected. Because of feature loss from overwriting, only four of the five feature units belonging to the target item can activate the correct focus unit, so that its asymptotic activation is 4/5. Due to feature overlap, one of these feature units (unit 2) also activates one of the competing focus units, which therefore has an asymptotic activation of 1/5. The activated focus units are used as input for computing the result. The connections from the focus units to the result units (dotted lines) are modulated by a representation of the operation (here, "+2"). In the example illustrated here, the target item is activated higher than any competitor, and therefore the correct result receives the highest activation level and is selected. Due to noise, the co-activated representation of a competitor item in the focus layer could reach the highest activation level, and as a consequence, a wrong result unit would win the competition.

$$\mathrm{Pr}_i = (1 - C/2)^{n-1} \tag{4}$$

Further, we assume that the activation each item receives is distributed equally among its feature units. The proportional feature loss due to interference therefore results in an equivalent loss of an item's activation in working memory, A_i:

$$A_i = (1 - C/2)^{n-1} \tag{4a}$$

When items are encoded into working memory, they become bound to specific context representations by which they can be retrieved. In serial recall, for example, each item is often assumed to be linked to a segment on a temporal or positional context representation (for models along this line see, e.g., Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1999). In a task such as memory updating, it is plausible to assume that items are bound to representations of their frames on the screen. In a model using synchrony of firing as a binding mechanism, feature units belonging to an item fire in synchrony not only with each other but also with the context they are bound to.

When an item is to be retrieved, it is cued by its frame. The representation of that context serves to select the features currently bound to it as input for the updating operation (or for the final recall operation). One possibility for implementing the cued retrieval of an item from working memory is to let the feature units of the cued item activate a representation of that item in the focus of attention (Oberauer, 2002). Hence, retrieval of an item from working memory is the gradual activation of a representation of that item in the focus layer (see Fig. 5). The asymptotic activation that can be reached for a representation in the focus is determined by the sum of the activation of the feature units that provide input to it. The activation is then passed on from the focus representation to a representation of the result of the updating operation. The gradual activation of the target item i in the focus of attention can be described by a negatively accelerated function (McClelland, 1979):

$$a_i = A_i(1 - \exp(-tr)) \quad (5)$$

where a_i is the activation of the target item in the focus layer, A_i is the activation of that item in the feature layer (acting as an asymptote of a_i), t is the time since beginning of the retrieval process, and r is the rate of activation. Inserting Eq. (4a) for A_i we obtain:

$$a_i = (1 - C/2)^{n-1}(1 - \exp(-tr)) \quad (5a)$$

Due to the partial overlap of items the cue not only activates the associated target item, but also partially activates competing items in the focus layer. Each competitor is activated by those features of the target that are shared with the competitor and are bound to the target and its cue, so that they are selected for being forwarded to the focus layer. In Fig. 5, feature units 1 and 2 are shared between the target item and the competitor, but only feature unit 2 sends activation to the focus layer, because unit 1 has been grabbed by the competitor. Therefore, feature unit 2 contributes to activation of the competitor in the focus layer. In general, each competitor will have grabbed away $C/2$ of the C feature units it shares with the target. Of the remaining $C/2$, other competitors (if there are any in working memory) are expected to grab away a proportion of $C/2$. The proportion of feature units of the target that are shared with any competitor *and* remain bound to the target can, therefore, be expressed as $C/2$ times $(1 - C/2)^{(n-2)}$ for $n > 1$. The activation equation for competitor items therefore becomes:

$$a_j = (C/2)(1 - C/2)^{n-2}(1 - \exp(-tr)) \quad (6)$$

The focus of attention selects only one item at a time for processing (Oberauer, 2002), the one receiving the highest activation input. The probability that among n items in working memory the target item i forwards the highest activation to the focus layer can be expressed by the Boltzmann equation (Anderson & Lebiere, 1998, p. 90) :

$$p_i = \frac{\exp(a_i/T)}{\sum_{j=1}^{n} \exp(a_j/T)} \quad (7)$$

In this equation, p_i is the probability that the activation of a target item i, expressed as a_i, is higher than the activation of all other items transmitted to the focus layer. Parameter T reflects the noise in the system, expressed as the "temperature", which is related to the standard deviation of activation by $T = \mathrm{sqrt}(6)\,\sigma/\pi$; in the model we treat the activation noise σ as a free parameter.

Besides the other items held in working memory, there is also competition from the remaining items in the set of possible items, in particular during the computation of a new item from the retrieved item. There were nine possible locations in each frame in the spatial updating task and nine possible results in the arithmetic updating task. We assume that all items in the set compete for being generated as results of the updating operation according to their current level of activation. Therefore, the denominator of Eq. (7) must sum over the activation levels of nine items, which can be broken down into three categories, the target item with activation a_i, the other items currently in working memory, which have activation a_j, and the remaining items in the set which have activation 0. Therefore, Eq. (7) can be expanded:

$$p_i = \frac{\exp(a_i/T)}{\exp(a_i/T) + (n-1)\exp(a_j/T) + (9-n)\exp(0/T)} \quad (8)$$

The processing model for memory updating assumes that the cognitive system encodes all initial values without errors. The first updating cycle starts with the presentation of the first updating operation in one frame. The selected frame position is used as a retrieval cue, and the gradual accumulation of activation for the target item i and the competitors in the focus layer begins. This accumulation of activation proceeds until the presentation time for the operation ends. During the whole time, the activation of i is immediately used as a continuous input for a procedure that performs the updating operation. The updating operation consists of gradually activating a representation of the operation's result. The shifts in mental space (for dots) or arithmetic computations (for digits) are assumed to yield the correct result if and only if the activation of the correct result is higher than the activation of all competing items (i.e., the other items in working memory and the remaining items in the set).

Retrieval of an item and transformation of that item according to the presented operation hence form a cascade of two operations. We initially attempted to model this cascade explicitly by distinguishing two accumulating activations (one for the retrieved item in the focus layer, one for the result of the operation in the result layer) driven by two rates (i.e., one for retrieval into the focus layer, and one for producing the new item in the result layer). That model, however, led to estimation problems (i.e., failure to converge), which we took as indicating that the model was overparameterized for the present data. Therefore, we simplified the model by collapsing the two processes into one. The gradual activation of the target item i by rate r therefore represents the joint process of retrieving the target item into the focus of attention, and of activating the result of the updating operation. Likewise, the gradual activation of competing items j reflects the joint activation of other items currently in working memory and the activation of competing operation results, possibly arising from these items.

We assume for simplicity that the probabilities of succeeding with each updating step are independent. The probability to recall an item correctly at the end of a trial, therefore, is the product of the probabilities of successful individual updating operations on that item, times the probability that retrieval succeeds in response to the item's final probing. In the absence of any useful memory, participants are still forced to select one of nine response alternatives, therefore we set chance performance to 1/9. We can estimate accuracy in recalling each item i as:

$$P_i = 1/9 + (1 - 1/9)p_i^m p_i', \tag{9}$$

with m expressing the number of updating operations applied to item i, p_i the probability of a single successful updating step, and P_i the probability of recalling the correct item in frame i at the end of a trial. The probability to succeed in the final retrieval, denoted p_i', is computed just like p_i, but with processing time t set to infinity, because there was no time limit for retrieval. In the experiments reported here, m depends on the memory demand condition, because the total number of updating operations was held constant for each trial, so that with increasing MD, each individual item became updated less often (e.g., with a total of eight operations, $m = 4$ for MD 2 and $m = 2$ for MD 4).

One further feature of the model needs to be introduced: the fusion of two cascaded processes (retrieval and the updating operation) into one is most likely unproblematic for MDs larger than one. Two exponential accumulation processes in cascade would give rise to a cumulative Gamma function (McClelland, 1979), whereas we merge them into a single process modeled as an exponential function. The two functions differ only in their early sections, that is, at the shortest presentation times, for which the present data are relatively scarce and unreliable, so that there is no realistic chance for distinguishing between the two functions. Our simplification could be problematic, however, for MD 1, because it glosses over a difference in task demands between MD 1 and higher demands. With a MD of 1, each operation can be immediately applied to the result of the previous one, so that the input to the operation does not have to be retrieved before being manipulated. With higher MDs, each operation involves a switch to another frame, so that the new frame's content must be retrieved before it can be operated on. Several experiments with self-paced versions of the memory-updating task (Garavan, 1998; Oberauer, 2002, 2003) have shown that switching from one object in working memory to another is associated with substantial time costs, compared to operations that can be applied directly to the result of the previous operation. These switch costs can be interpreted as the time for retrieving a new item from working memory into the focus of attention. In the present context, this means that conditions with MDs larger than 1 would involve a retrieval step before the operation proper, whereas the MD 1 condition would not involve such a step. To capture this difference, we allowed different rate parameters for MD 1 (denoted r_1) and for higher MDs (simply denoted r) in all applications of the interference model.

To summarize, the interference model has four free parameters: the interference parameter C, the two rate parameters r and r_1 and the standard deviation of activation, σ. These parameters can vary between people or between tasks and experimental conditions, or both. Variations of parameters over tasks reflect task characteristics. For example, C reflects the mean degree of feature overlap of the task's items; it can then be interpreted as a proxy for inter-item similarity in the task. Likewise, the rate parameter r reflects the typical speed of an updating operation as required in the task. Variation of parameter means over conditions in an experiment reflects the effect of the manipulation on the parameter. Variation over persons, on the other hand, reflects individual differences. Model parameters with individual differences can be thought of as theoretically meaningful latent variables that explain individual differences in observed variables (i.e., measured performance). The rate parameter r, for instance, can be interpreted as an indicator of a person's processing speed, σ reflects the level of noise in a person's working memory system, and C can be interpreted as the person's susceptibility to interference. Within the NLME framework, both dimensions of variation are taken into account: variation over experimental condition is captured by fixed effects, and variation over persons is captured by random effects.

Application to the spatial memory updating data (Experiment 1)

We fitted the model to the data with NLME, progressing through a series of models with increasingly relaxed constraints on individual differences in model parameters (i.e., allowing more parameters to have random effects) up to a point where no further increase of fit could be achieved. The model fit was assessed by the log-Likelihood statistic returned by the NLME algorithm (higher log-Likelihoods represent better fits), the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the adjusted R^2 statistic. The AIC and BIC are derived from log-Likelihood, penalizing

Table 1
Model fits for spatial memory updating (Experiment 1)

Nr	Fixed effects	Random effects	Par	AIC	BIC	log-Lik	R^2_{adj}	Sign
0	C, r, σ	C	5	−1269	−1244	639.3	0.713	
1	C, r, r_1, σ	C	6	−1387	−1358	700.0	0.747	0
1a	C, r, r_1, r_2, σ	C	7	−1427	−1393	720.6	0.757	1
2	C, r, r_1, σ	C, σ	8	−1483	−1444	749.7	0.785	1
3	C, r, r_1, σ	C, r	8	−1469	−1430	742.6	0.782	1
4	C, r, r_1, σ	C, r, r_1	11	−1589	−1535	805.3	0.823	3
5	C, r, r_1, σ	C, r, r_1, σ	15	−1607	−1533	818.6	0.831	4
5a	C, r, r_1, r_2, σ	C, r, r_1, σ	16	−1639	−1560	835.6	0.836	5

Legend: Par, number of free parameters; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; log-Lik, log-Likelihood; Sign, model fit is significantly better than the nested model with the number in this column.
Note. Estimates are based on 999 data points.

for the number of free parameters, and therefore are recommended for comparing model fits while taking parsimony into account (lower values of AIC and BIC, i.e., higher absolute values, indicate a better fit).[6] When two models with the same fixed effects are nested (i.e., one can be derived from the other by dropping a subset of the random effects), the model fits can be compared by a test of whether the less parsimonious model had a significantly better fit than the more constrained model. We used this last criterion, where possible, and AIC and BIC otherwise, to determine which model version presents the best compromise between fit and parsimony. The fit statistics of the model versions tested are summarized in Table 1.

Besides the theoretically motivated versions we also tested two exploratory model versions that introduce a difference in the rate parameter between MD 2 and the two higher levels of MD. These model versions (1a and 5a) have three different rate parameters, r_1, r_2, and r (for MD > 2). The theoretical model does not provide any rationale for assuming different rates for MD levels larger than one, but it does not exclude them either. The purpose of investigating these models is to test whether processing rates in the MD range from two to four are actually constant. If this were the case, it would suggest a qualitative difference between processing at MD 1 and processing at higher MD levels. A qualitative difference in processing rates would be strong support for the assumption that with MD 1 each item to be updated is already in the focus of attention and, therefore, can be processed immediately, whereas with higher MD levels, a retrieval step must precede processing. A similar rationale underlies the argument for a focus of attention put forward by McElree and Dosher (1989), who investigated speed-accuracy function—partially analogous to our time-accuracy functions—for the Sternberg recognition task. It turned out that with the present data a model with different processing rates for MD 2 and MD > 2 yielded a small but significant improvement of fit. Therefore, there was no evidence for a qualitative difference in processing dynamics between MD 1 and higher MD levels. This finding does not contradict the idea of a focus of attention in working memory, but does not support it either. To anticipate a result from the next data set, no evidence for an increase in rate from MD 2 to 4 was found with the numerical memory updating data, thereby lending some support to a qualitative difference between MD 1 and higher levels of MD.

Among the theoretically motivated models, version 5 provided the best fit. The parameter estimates of that model are summarized in Table 2; the first row presents the sample means (i.e., fixed effects), the second row their standard deviations (i.e., random effects), and the third to last row their correlations. Standard deviations and correlations are estimated parameters of the NLME model in which the interference model is embedded.

The fitted model yielded predictions for the performance of individual participants (BLUPs), which are plotted together with the data in Fig. 2. The averaged predictions are plotted alongside the data in Fig. 3. We modeled the predicted data by descriptive TAFs (Eqs. (1) and (3)) in the same way as the empirical data to obtain model predictions for asymptotic accuracies and CPTs for each participant; these predictions are plotted with the corresponding empirical results in Fig. 4. The model reproduced the qualitative trends in the data well, with an accelerated decline of asymptotic accuracy and an approximately linear increase of CPT with increasing levels of MD up to 3.

Remodeling the numerical memory-updating data

Next, we tested whether the modified interference model still fits the original data from numerical memory

[6] Readers might wonder why our log-Likelihoods are positive. This is because with continuous distributions log-Likelihood is the logarithm of a probability density, which can be larger than 1. As a consequence, AIC and BIC, which are defined as −2 log(Likelihood) plus a penalty term, are negative.

Table 2
Parameter estimates of interference model (Version 5) for spatial memory updating (Experiment 1)

	C	r	$r_1 - r$	σ
Mean	0.39 [0.36, 0.43]	1.40 [1.19, 1.60]	0.44 [0.25, 0.63]	0.15 [0.14, 0.16]
Standard deviation	0.071 [0.05, 0.11]	0.447 [0.31, 0.64]	0.386 [0.26, 0.57]	0.022 [0.01, 0.03]
C		0.12 [−0.34, 0.53]	−0.13 [−0.60, 0.40]	−0.04 [−0.51, 0.45]
r			−0.78 [−0.95, −0.27]	0.57 [0.11, 0.83]
$r_1 - r$				−0.16 [−0.65, 0.42]

Note: The top two rows of the table contain estimates for the parameter means and their standard deviations in the participant sample; the bottom part of the table contains estimates for the parameter correlations across participants. Approximate 95% confidence intervals computed by the *nlme* package are given in brackets. Instead of estimating r_1 directly, NLME estimates the slope of r from MD > 1 to MD1, which is given here as $r_1 - r$; r_1 can be obtained by adding this slope to r.

updating (Oberauer & Kliegl, 2001). The data were time accuracy data from 18 young adults who worked on a numerical updating task with MDs from one to four.[7] Descriptive TAFs were fit to the data in the same way as in Experiment 1 ($R^2_{adj} = .888$). There were significant fixed effects for the three intercepts, the three linear contrasts, and the quadratic contrasts on a and b; random effects were significant for the three intercepts, and the linear and quadratic contrasts on a and b. The significant effects are plotted in Fig. 6. The downward accelerated decline of asymptotes with MD is not apparent in these data (i.e., the quadratic contrast on c was not significant), probably because young adults' asymptotes were close to ceiling up to MD 4. Old adults showed an accelerated downward trend in asymptotes for MD 1 to 4, and young adults showed the same trend for higher levels of MD in the numerical updating task (Oberauer & Kliegl, 2001).

We applied the set of progressively relaxed versions of the interference models used for Experiment 1 to the data from numerical updating; the results are summarized in Tables 3 and 4. The models with more than one random effect had all correlations constrained to zero because freely estimating these correlations led to no significant improvement of fit. Model version 5 provided the best fit; no further improvement could be reached by allowing rate to increase from MD 2 to MD > 2 (version 5a). The model fit the data well, as is illustrated in Fig. 6. One systematic deviation from the data was that the model overpredicted asymptotic accuracies at low MDs. This overprediction probably reflects the assumption that people do not make mistakes in the arithmetic computations; this assumption might be overly optimistic.

A comparison of the parameter estimates for the numerical task with those for the spatial task shows a larger mean value of C, a faster processing rate r and a larger increase of processing rate from r to r_1, as well as a higher standard deviation σ for the spatial task. The differences in parameter estimates were significant, as

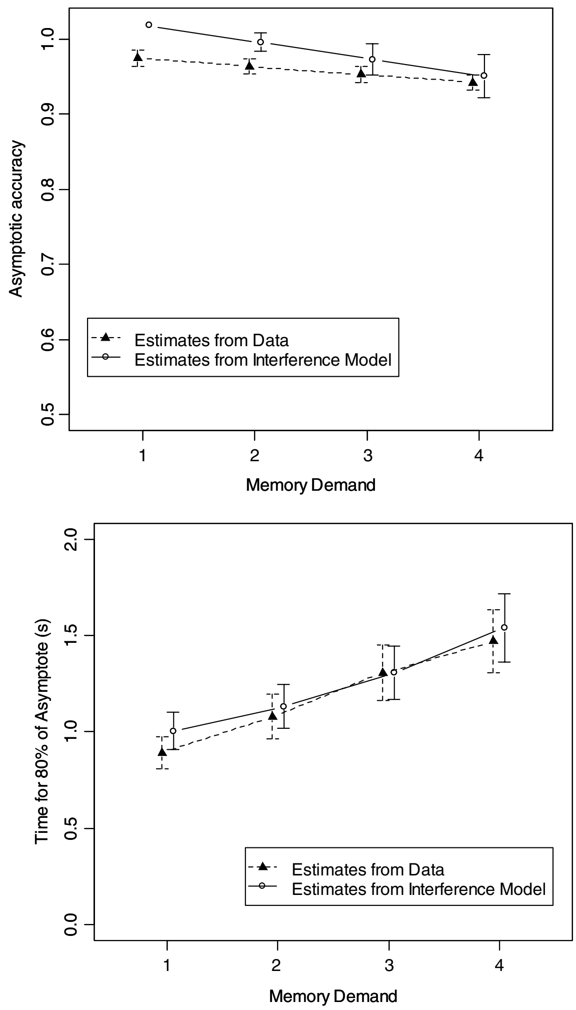

Fig. 6. Asymptotic accuracies (top panel) and time demands for 80% of the asymptote (CPT 80) for numerical memory updating, estimated from fits of descriptive time-accuracy functions to the data and to the predictions of the interference model. Error bars represent 95% confidence intervals.

[7] Only data from young adults and MDs up to four are modeled here for comparability with the other data presented in this article. This selection enabled us to include two participants who did not complete the higher levels of MD and, therefore, had been excluded in Oberauer and Kliegl (2001).

Table 3
Model fits for numerical memory updating

Nr	Fixed effects	Random effects	Par	AIC	BIC	log-Lik	R^2_{adj}	Sign
0	C, r, σ	C	5	−1015	−991	512.4	.775	
1	C, r, r_1, σ	C	6	−1042	−1013	527.2	.783	0
1a	C, r, r_1, r_2, σ	C	7	−1045	−1012	529.5	.783	1
2	C, r, r_1, σ	C, σ	7	−1364	−1331	689.2	.860	1
3	C, r, r_1, σ	C, r	7	−1365	−1331	689.4	.860	1
4	C, r, r_1, σ	C, r, r_1	8	−1391	−1353	703.5	.871	3
5	C, r, r_1, σ	C, r, r_1, σ	9	−1403	−1360	710.4	.877	4
5a	C, r, r_1, r_2, σ	C, r, r_1, σ	10	−1398	−1350	709.0	.875	4, not 5

Legend: Par, number of free parameters; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; log-Lik, log-Likelihood; Sign, model fit is significantly better than the nested model with the number in this column.
Note. Estimates are based on 868 data points. Correlations between random effects were fixed to zero.

Table 4
Parameter estimates of interference model (Version 5) for numerical memory updating

	C	r	$r_1 - r$	σ
Mean	0.29 [0.27, 0.32]	0.69 [0.58, 0.81]	0.08 [0.02, 0.14]	0.12 [0.10, 0.13]
Standard deviation	0.035 [.016, .074]	0.20 [0.12, 0.36]	0.10 [0.06, 0.16]	0.022 [0.012, 0.042]

Note: The table contains estimates for the parameter means and their standard deviations in the participant sample (with approximate 95% a confidence intervals in brackets); the parameter correlations were fixed to 0. Instead of estimating r_1 directly, NLME estimates the slope of r from MD > 1 to MD 1, which is given here as $r_1 - r$; r_1 can be obtained by adding this slope to r.

shown by the non-overlapping confidence intervals (compare Tables 2 and 4). This means that there was substantially more mutual interference between spatial than between numerical items in working memory. As a result, the decrease of asymptotic performance with memory load was steeper in the spatial version, as was the increase in time demand to reach 80% of the asymptote (compare Figs. 4 and 6).

Discussion

The modified interference model was successfully applied to time-accuracy data from a spatial and a numerical version of the memory-updating task. This result supports the notion that the same basic principles can be used to model capacity limitations in numerical and spatial working memory. The form of the functions that relate memory load to time demand and asymptotic accuracy are the same in both domains, only the parameter values are different. In particular, the spatial locations yielded higher interference and more noise than the digits, and the processing rates were faster for spatial than for numerical updating. These differences might be due to the specific materials used here, or they might reflect general differences between spatial and numerical working memory. The difference in the noise parameter σ was somewhat surprising and indicates that σ is not a general parameter of the memory system, but content dependent. One interpretation of this result is that the noise parameter does not reflect a constant property of the cognitive architecture, but a combination of noise intrinsic to the cognitive system and noise arising from the representations involved in a cognitive process. If, for instance, representations of dot positions consist of fewer feature units than representations of digits, then the activation levels of items in working memory, A_i, would be expected to vary more from item to item and from one updating step to the next in the spatial than in the numerical task, because with fewer feature units the actual loss of units in each instance would have a larger variability around its mean.

The direct comparison of memory updating with numbers and with spatial positions yielded another interesting observation. The time demand for a spatial shift operation was shorter than the time demand for an arithmetic operation as long as memory demand was minimal. This is evident in the CPTs for MD 1, which was 581 ms ($SD = 90$) for the spatial task, and 891 ms ($SD = 178$) for the numerical task. However, time demands showed a steeper increase with complexity in the spatial than in the numerical task. At MD 4, a spatial shift operation required as much time for 80% of the asymptote (1501 ms) as an arithmetic operation (1470 ms). The slope of the CPTs with memory load obviously does not reflect an increase due to a constant or proportional factor. Instead, it is linked to the difficulty of holding several items in working memory simultaneously, which is a function of the interference parameter. The interference model captures this to some degree by the assumption that overlapping representations are partially overwritten, thus reducing their overall activation, so that more time is required to raise their activation reliably above that of the competitors.

Experiment 2

The second experiment combined the numerical and the spatial memory-updating task to investigate interference between items from the same domain and interference between domains. One of the best established findings in working-memory research is that dual-task costs are much reduced when the two tasks come from different content domains—in particular, when one of them uses verbal (including numerical) material and the other visuo-spatial material (Baddeley, 1986; Farmer, Berman, & Fletcher, 1986). These findings, together with dissociations from neuropsychology and neuroimaging (eg., Smith & Jonides, 1997), have been used to argue for separate subsystems of verbal and visuo-spatial working memory. Further research has revealed that double dissociations can also be found within both broad domains. Impairment of memory by a concurrent processing task is larger when both tasks involve words, or both involve digits, than when one involves words and the other digits (Conlin, Gathercole, & Adams, 2005). Dual-task costs are also smaller when one task involves visual and the other involves spatial information than when both involve visual, or both involve spatial information (Klauer & Zhao, 2004). Rather than postulating ever more separate subsystems, an alternative explanation of these findings is by assuming a single working memory system operating on various kinds of representations in different parts of the brain (Cowan, 2005). The interference model fits well with this alternative view. We assume that representations in different content domains share, on average, fewer features than representations within a content domain. In dual-task situations that require representations of both tasks to be held in working memory simultaneously, there will be more interference through feature overwriting when the tasks rely on representations from the same domain.

To test this idea we modeled the design of Experiment 2 on the logic of dual-task experiments used to demonstrate the double dissociation of content domains. There were three conditions with a memory demand of four items, one with four digits, one with four spatial positions, and one combining two digits with two spatial positions. If we think of each pair of items as constituting one task, these three conditions correspond to three dual-task combinations, two within-domain combinations and one cross-domain combination. In addition, there were two conditions with MD 2, one with two digits and one with two spatial positions, which can be interpreted as the "single-task" control conditions. Whereas in dual-task experiments the two tasks usually differ in several regards, only one of which is the content domain of the representations involved, our design manipulates content domain (i.e., the kind of representation and the corresponding kind of operation) without changing other task features, thereby enabling a more controlled assessment of how the different combinations of content domains affect performance. To our knowledge, the present experiment is the first to investigate different combinations of contents in a working memory task while holding constant the task paradigm.

The purpose of Experiment 2 is to test whether the interference model can reproduce the effects of memory demand within a domain and the effects of dual-task combinations across domains within a common framework and a small number of free parameters. We assume three overlap parameters, C_{num}, C_{spat}, and C_{mix}. The first two parameters reflect the degree of overlap within each domain, and the third parameter reflects the degree of overlap across domains. In addition, we used two rate parameters, r_n and r_s, for numerical and spatial updating operations, respectively, and one noise parameter σ.

Method

Participants

Participants were 10 high-school students from Potsdam. Their mean age was 19.1 years ($SD = .88$), and four were female. They participated in 10 1-h sessions and received 12,- DM (about US$ 6) for each session.

Design and procedure

There were five conditions. Two "single-task" conditions involved memory-updating tasks with a MD of two items, either digits (numerical 2) or spatial positions (spatial 2). The two frames in which the stimuli appeared were presented adjacent to each other in the center of the screen. As in the previous experiments, the digits had to be updated according to eight successive arithmetic operations, and the spatial positions were to be updated according to eight successively presented arrows. The mixed condition combined two frames with digits and two frames with dots (mixed 4). The four frames were arranged in a 2×2 matrix in the middle of the screen, with the dots displayed in the top row and the digits in the bottom row. The digits were updated through arithmetic operations displayed in the respective frames, and the dot positions through arrows. The sequence of updating operations again followed a clockwise order, so that two spatial operations were followed by two arithmetic operations, then again two spatial operations and two arithmetic operations. The mixed condition was contrasted with two content-homogeneous conditions of equal memory demand, one with two rows of digits (numerical 4), and one with two rows of dots (spatial 4).

There were 18 blocks for each condition, each block consisting of 13 trials. One trial had a fixed presentation time of 6 s, the other presentation times were determined by the same adaptive algorithm as in the previous experiments. In the mixed condition, the same presentation time was used for the spatial and the numerical operations. The blocks were ordered according to the sequence spatial 2 – numerical 4 – mixed 4 – numerical 2 – spatial 4, followed by its reverse, and then repeated until the end.

Results

The averaged accuracies as a function of presentation time, produced in the same way as for Experiment 1, are displayed in Fig. 7. We first analyzed the data by fitting descriptive TAFs jointly to the three numerical conditions (including the numerical accuracies from mixed 4) and to the three spatial conditions (including the spatial accuracies from mixed 4; combining numerical and spatial conditions in a single model was not feasible because the large number of parameters made estimation computationally extremely costly). Within each content domain, the intercept of each parameter represented the MD 2 condition, and two contrasts were defined to capture the differences between conditions: the *dual-task contrast* represented the difference between MD 2 and the mixed condition, that is, dual-task costs from combining contents across domains. The *domain contrast* represented the difference between the cross-domain mixed condition and the within-domain MD 4 condition, thereby reflecting the increase in dual-task costs when combining contents from the same domain.

The best fitting model for the numerical conditions ($R^2_{\text{adj}} = .854$) had fixed effects for the intercepts on a, b, and c, and for both contrasts on a and on c. Random effects were significant for the intercept of b and c, and for both contrasts on c. For the spatial conditions we had to fix the a parameter to 0 because otherwise a would have been estimated to negative values, with the absurd implication of above-chance performance at time 0. The best fitting model ($R^2_{\text{adj}} = .817$) had fixed effects on the intercepts of b and c, as well as the dual-task contrast on b and both contrasts on c. All these fixed effects were associated with significant random effects.

The asymptotes and CPTs for all conditions are shown in Fig. 8. The significant domain contrast for the asymptote parameter c in both content domains reflect the typical pattern of dual-task studies: digits could be recalled better when combined with additional spatial material (mixed numerical) than when combined with more digits (numerical 4). Conversely, spatial positions could be recalled better when combined with digits (mixed spatial) than when combined with additional spatial information (spatial 4). The significant dual-task contrast shows that, nonetheless, combining representations from the numerical and the spatial domain in working memory also engenders dual-task costs (for similar findings see Morey & Cowan, 2004). The pattern of CPTs across conditions was less clear and somewhat clouded by the large variability in two of the spatial conditions.

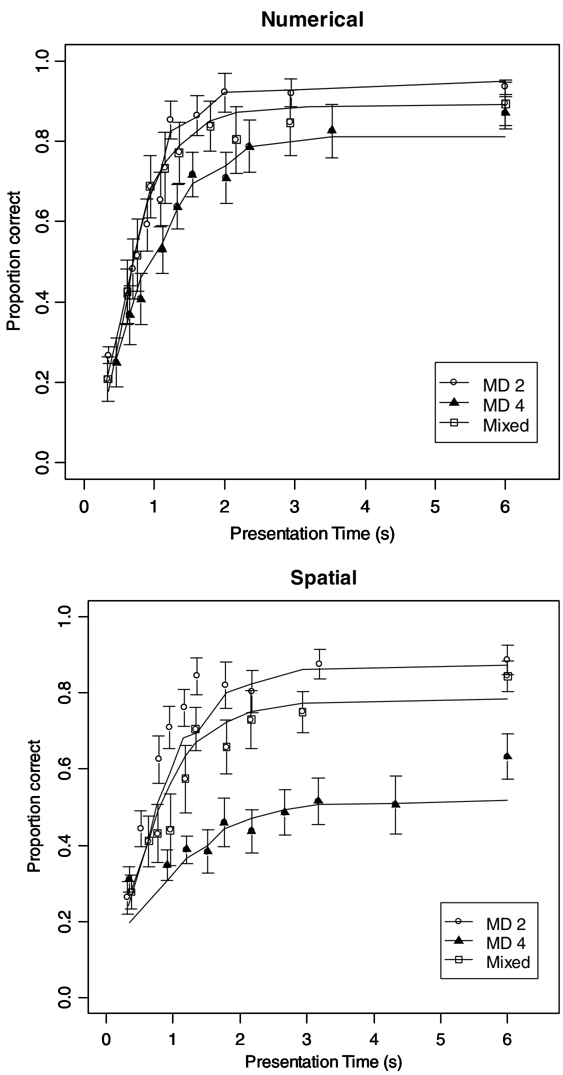

Fig. 7. Data from Experiment 2 (points) and predictions of the interference model (lines), averaged over participants for 12 categories of presentation times. Placement of points on the x-axis reflects mean presentation time in each category. Error bars represent 95% confidence intervals.

The fits of a series of progressively relaxed versions of the interference model are summarized in Table 5. With the small N of this study correlations cannot be estimated reliably (see the large confidence intervals of

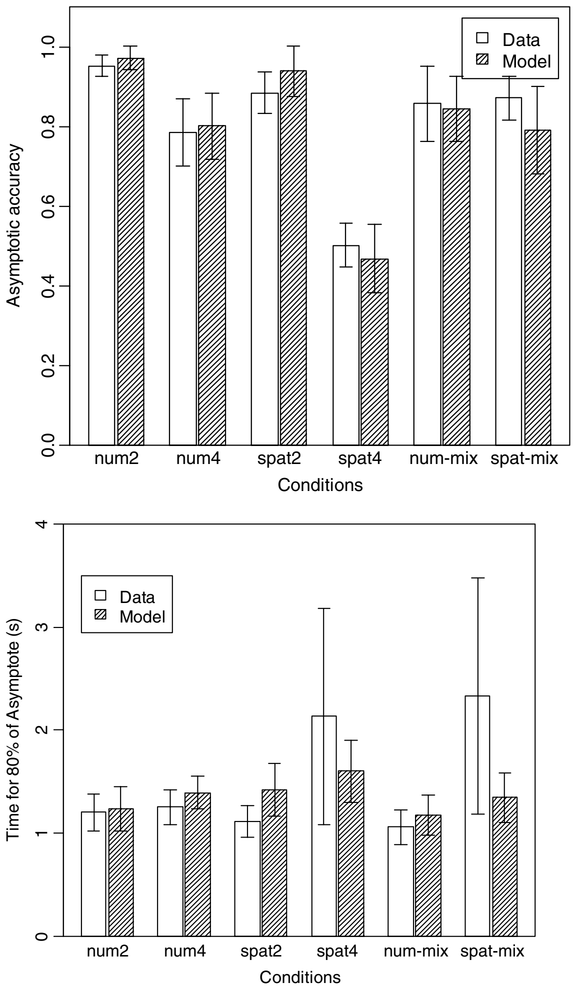

Fig. 8. Asymptotic accuracies (top panel) and time demands for 80% of the asymptote for the six conditions of Experiment 2, estimated from fits of descriptive time-accuracy functions to the data and to the predictions of the interference model. Num, numerical; spat, spatial; mix, mixed. Error bars represent 95% confidence intervals.

the correlations in Experiment 1); therefore we fixed all correlations to zero. Version 5 provided a significantly better fit than the models nested below it. The parameter estimates of this model are shown in Table 6. The estimated means and standard deviations match closely those of the previous experiments with spatial updating (Table 2) and numerical updating (Table 4), with the exception of the numerical rate, which was larger than in the previous experiment and had a smaller standard deviation (although the latter came with a very broad confidence interval). The model predictions are added as lines to Fig. 7, and the TAF parameters estimated from them are displayed alongside those from the data in Fig. 8. The model reproduced the critical pattern of dual-task effects on asymptotic accuracies well. The pattern of CPTs across conditions was reproduced less accurately; the model underpredicted the means and standard deviations of CPTs in the spatial part of the mixed condition and the spatial memory-demand-4 condition. This deviation from the data points to possible individual differences in the spatial conditions that are not yet captured by the model. With sample sizes as small as the present ones it is not possible to further analyze these individual differences.

Discussion

Experiment 2 provided a successful generalization of the interference model to a dual-task combination. The model provided a satisfactory quantitative fit to the data, and the parameter estimates closely matched those from the previous two studies. The data of Experiment 2 match the typical finding from dual-task combinations of verbal (or numerical) with spatial tasks, in that they exhibit a double dissociation between the two content domains. This pattern is usually interpreted as evidence for two separate working memory systems. The interference model reproduced the double dissociation with a single parameter specific

Table 5
Model fits for numerical, spatial, and dual-task memory updating (Experiment 2)

Nr	Random effects	Par	AIC	BIC	log-Lik	R^2_{adj}	Sign
0	C_{num}, C_{spat}	9	−904	−863	460.9	0.771	
1	C_{num}, C_{spat}, C_{mix}	10	−999	−954	509.3	0.810	0
2	C_{num}, C_{spat}, σ	10	−1037	−992	528.4	0.819	0
3	C_{num}, C_{spat}, r_n, r_s	11	−991	−942	506.7	0.815	0
4	C_{num}, C_{spat}, C_{mix}, σ	11	−1055	−1005	538.4	0.827	1,2
5	C_{num}, C_{spat}, C_{mix}, r_n, r_s, σ	13	−1089	−1031	557.7	0.842	4

Legend: All models use C_{num}, C_{spat}, C_{mix}, r_n, r_s, and σ as fixed effects. Par, number of free parameters; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; log-Lik, log-Likelihood; Sign, model fit is significantly better than the nested model with the number in this column.

Note. Estimates are based on 682 data points. Correlations were fixed to zero.

Table 6
Parameter estimates of interference model (Version 5) for numerical, spatial, and dual-task memory updating (Experiment 2)

	C_{num}	C_{spat}	C_{mix}	r_n	r_s	σ
Mean	0.30 [0.26, 0.34]	0.47 [0.43, 0.51]	0.25 [0.21, 0.30]	1.02 [0.95, 1.12]	1.38 [1.12, 1.64]	0.15 [0.14, 0.17]
Standard deviation	0.041 [0.023, 0.090]	0.056 [0.033, 0.101]	0.056 [0.036, 0.107]	0.068 [0.001, 1.140]	0.410 [0.223, 0.659]	0.023 [0.013, 0.034]

Note: Data are estimates for the parameter means and their standard deviations in the participant sample (95% confidence intervals in brackets); as there were only ten subjects, correlations can not be estimated reliably and therefore were fixed at zero.

to the dual-task combination (C_{mix}), showing that the present data can be explained parsimoniously in the context of a single working memory system. The interference model, therefore, offers a way of reconciling evidence for a dissociation of content domains in working memory with evidence for a strong association of working memory performance across domains (the latter coming mostly from correlational studies, e.g., Kane et al., 2004; Oberauer, Süß, Wilhelm, & Wittmann, 2003). We can think of working memory as a unitary system, described by domain-general principles and some domain-general parameters, operating on different kinds of representations with different operations, which require parameters specific to them. One consequence of this view is that there should be positive correlations between individual's parameter estimates (in particular C and r) across domains, which could explain the correlations of performance in verbal and in spatial working memory tasks observed in factor-analytic studies. Unfortunately, our present sample size is much too small to test this prediction.

The interference model treats the double dissociation of the verbal–numerical and the spatial domain as a special case of the general principle that the degree of interference is a function of representational overlap. Thus, it should be possible to construct similar double dissociations between any two content categories when the inter-item overlap within a category is substantially larger than the overlap of two items from different categories. Support for this assumption comes from dual-task studies combining different kinds of verbal materials. Combining a sentence reading task with memory for numbers, or an arithmetic processing task with memory for words, leads to less interference than combining sentence reading with memorizing words or arithmetic with recall of digits (Conlin et al., 2005; Li, 1999; Turner & Engle, 1989). Serial recall of consonants was better when a set of digits had to be remembered simultaneously, than with a second set of consonants. Conversely, short-term memory for digits suffered more from a secondary load with digits than with consonants (Sanders & Schroots, 1969).

The results of Experiment 2 challenge a model of working memory that explains capacity limits purely in terms of time-based decay and rehearsal. In our competitive test of various formal models, a decay model was the only model besides the interference model to provide a good fit of time-accuracy functions from the numerical memory updating task (Oberauer & Kliegl, 2001). A model along similar lines has been advanced and successfully tested by Barrouillet et al. (2004). So far, that model has been applied only to data from tasks with verbal or numerical contents. Our decay model as well as the time-based resource sharing model of Barrouillet et al. assume that only one process can occur at any time, either an updating operation on an item in working memory or a rehearsal process. This account explains MD effects by the increasing demand for rehearsal when more items are held in working memory. With a single processing mechanism—a "retrieval bottleneck" in Barrouillet et al.'s theory—this model predicts equal performance for pure and for mixed sets of four items. The asymptote and the CPT of the mixed condition, therefore, should only depend on the difficulty of the tasks combined in that condition, that is, they should be intermediate between the conditions of the numerical MD 4 and the spatial MD 4 condition, contrary to our findings (and those of many other dual-task studies). One way around this problem would be to allow separate rehearsal mechanisms for verbal and for spatial contents of working memory that can run in parallel. Splitting the rehearsal mechanism into two would raise the question whether operations that manipulate working memory contents in the two domains (e.g., an arithmetic operation and a spatial shift) could also run in parallel. Evidence with self-paced versions of the mixed memory-updating task suggest that this is possible only after several hours of practice (Oberauer & Göthe, 2006; Oberauer & Kliegl, 2004). If two rehearsal mechanisms are assumed to operate in parallel even in unpracticed participants, we must assume that rehearsal is not constrained by the same domain general bottleneck that constrains other working memory operations in unpracticed people. This conclusion, however, clashes with a basic assumption of our decay model (Oberauer & Kliegl, 2001) as well as the decay based model of Barrouillet et al. (2004), namely, the assumption that the updating operations in the present task, as well as other concurrent processing tasks, compete with rehearsal for time. It is not clear at the moment whether another model based on decay and rehearsal can be formulated that can account for the double dissociation of verbal and spatial working memory, but it seems clear that such a model would have to differ substantially from the one that we have previously fitted with good success to the numerical updating data.

Experiment 3

If interference is a function of representational overlap, it should be possible to vary the degree of interference by working with similar and dissimilar memory sets from the same content domain. This is a specific prediction of the interference model. Working-memory models assuming that capacity limits arise from a limited quantity of resources—either from a general pool or from multiple domain specific pools—do not predict differential interference effects as long as all memory sets come from the same content domain and tap into the same resource. Likewise, models based on decay as the cause of limited capacity in working memory do not predict similarity effects within domains. Experiment 3 is an attempt to manipulate the interference parameter by the degree of similarity between memory items within a content domain.

We used the well-known phonological similarity effect (Conrad & Hull, 1964) to manipulate the degree of overlap between representations that must be held in working memory simultaneously. To this end, we designed a memory-updating task with letters as material. The interference model predicts that the interference parameter C will be larger with a set of phonologically similar letters than with a set of phonologically dissimilar letters.

Method

Participants

Seven high-school students from Potsdam and one student from the University of Potsdam participated in the experiment, which consisted of 16 1-h sessions. Their mean age was 20.0 years ($SD = 2.07$), and four of them were female. Participants received 12,- DM for each session.

Design and procedure

A memory-updating task with letters was used. In the similar condition, the letters were from the set "B, C, D, E, F, G". In the dissimilar condition, the set was "H, I, J, K, L, M".[8] Letters were updated by letter arithmetic operations (e.g., $C + 2 = E$). The operations ranged from −2 to +2, and their results always remained within the respective letter set. Therefore, not only the starting values, but also the intermediate and final results of updating operations were phonologically similar in one condition, and dissimilar in the other.

Before starting with the memory-updating task, participants practiced letter arithmetic for two 1-h sessions. The goal of the training was to bring participants as closely as possible to automatic performance of the letter arithmetic operations, and to equate their performance on the different equations as much as possible. Participants responded to single letter equations (e.g., "B + 2 = ?") by typing the correct letter key on the keyboard. Each session comprised 24 blocks with 30 trials each. Within one block, the same set of letters was used for all equations. All equations started with a letter from the set, combined with one of the four operations (+1, −1, +2, −2), and resulted in another letter of the set. The equations within a block were generated at random, with the constraint that all possible equations were presented once before one was repeated. The letter set used in each block was indicated by the color of the question mark in the equations; the corresponding letters were marked with a colored transparent patch on the computer keyboard (red for the similar and blue for the dissimilar letters). The sequence of blocks was determined adaptively. After the first two blocks, one with each letter set, the next block always used the letter set on which the performance score on the last block with the same set had been worse. The performance score was the number of errors if more than three errors were made in one of the two blocks compared, and the mean reaction time otherwise. This algorithm provided more

[8] In German, the similar letter set is pronounced: "/be:/ /tse:/ /de:/ /e:/ /Ef/ /ge:/". The dissimilar set is pronounced "/ha:/ /i:/ /jOt/ /ka:/ /El/ /Em/". The manipulation of feature overlap achieved by contrasting these sets is not ideal because most letters in the similar set share the same phoneme, /e:/, contrary to the assumption of stochastic independence of feature overwriting. If anything, this violation of a model assumption should work against the hypothesis of a good model fit. The violation is not as severe as it might look because phonemes are certainly not the only features by which letters are represented, and not the only features shared between letters in a set (other features would include the alphabet position of a letter, its visual shape, and features coding for the categories consonant and vowel). We have run simulations in which we manipulated similarity by having all stimuli in the similar set share the same 20% of all feature units—generating an increase in C larger than estimated in Experiment 3. The decrease of A_i with MD from 1 to 4 could still be extremely well approximated by Eq. (4), $R^2 > .98$. Apparently, the model is robust against a modest degree of violation of the stochastic independence assumption.

training for the letter set with which a participant performed worse.

After the training sessions, participants worked on memory-updating tasks with letters. There were four conditions of memory load (1-4 letters) crossed with two conditions of phonological similarity. Each of the 14 sessions comprised eight blocks in a fixed order of memory loads (1-4-3-2-2-3-4-1), with the similarity conditions alternating between blocks. Different from the previous experiments, presentation times were fixed. There were ten presentation times for each memory-demand condition; they ranged from 250 ms to 4 s for MD 1, and from 400 ms to 5 s for MD 4, with the other conditions in between. Presentation times were assigned to the trials in a semi-random order with the constraint that each presentation time was used in each condition 14 times. The first two sessions on the memory-updating task were regarded as practice and not included in the analysis. This leaves 12 trials for each combination of MD, similarity condition, and presentation time.

One additional change from previous experiments was that we lined up the frames in a row in the middle of the screen instead of the circular arrangement used so far. This was an attempt to manipulate the spatial discriminability of the frame positions. In a study of short-term memory for letters in spatial arrays, Healy (1977) found better memory for letters at the ends of the array than in the middle. No such effect was observed in the present experiment, so we will not discuss this manipulation further. A further change was that the locations of the updating operations were determined at random with the constraint that they be equally distributed across all locations within each sequence of eight operations. Therefore, not every transition from one operation to the next involved an object switch (i.e., the probability of repeating the object was $1/n$), and the next object to be accessed was not predictable.

Results

Letter arithmetic training

Due to computer failure, training data from two participants were lost. For the remaining six participants, reaction times for correct responses were analyzed after trimming values that exceeded the individual's mean in a session by three standard deviations. Four participants received more training sessions on the dissimilar than on the similar letter set, and two received more training on the similar set. We aggregated reaction times of the first three blocks and the last three blocks within each letter set. There was no difference in mean reaction times between sets at the beginning of training. The means were 2.5 s for similar letters ($SD = 0.3$) and 2.3 s for dissimilar letters ($SD = 0.5$). Over the last three blocks, however, there was a tendency for similar letters to be responded to faster, which just failed the conventional significance level, $t(5) = 2.46$, $p = .057$. The means were 1.3 s ($SD = 0.1$) for similar and 1.5 s ($SD = 0.3$) for dissimilar letters. Thus, we have reasons to believe that the letter-updating task on its own is easier with the letters from the similar set. This may be the case because the similar set was taken from the beginning of the alphabet, where the sequence might be represented more accurately than in the middle part from which the dissimilar set was taken. If this is the case, it should be manifest as different rate parameters in the model for the two letter sets. The difference should point in the opposite direction from that expected for the interference parameter: similar letters should yield larger interference, but a higher processing rate than dissimilar letters.

Memory updating

Time-accuracy functions averaged over participants are presented in Fig. 9. The data were analyzed by fitting descriptive TAFs as in the previous experiments. We tested differences between conditions with a contrast for similarity, the linear and quadratic contrasts of MD, and their interactions with similarity. The best fitting model included significant linear contrasts on all three parameters as well as a quadratic contrast on b. The main effect of similarity was not significant on any of the TAF parameters, but there was a significant interaction of similarity with the linear contrast of the c parameter. As shown in Fig. 10 (upper left panel), asymptotic accuracy declined faster over MD in the similar than in the dissimilar condition. Random effects were significant for the linear contrasts on b and c, and for the main effect of similarity on c. The descriptive model had an adjusted R^2 of .811.

The interference model was fitted to each participant's data for similar and dissimilar letter sets simultaneously. There were five free parameters: two overlap parameters for similar letters (C_{sim}) and dissimilar letters (C_{dis}), the rate parameters r and r_1, and the noise parameter σ (C_{sim} was not estimated directly but through a multiplicative factor F_{sim}, with $C_{sim} = F_{sim} \times C_{dis}$). The model fits are presented in Table 7. Given the small N, we again constrained all correlations to zero. A model with four random effects (Model 5) provided the best fit. Table 8 summarizes the parameter estimates. As expected, C_{sim} was estimated to be larger than C_{dis} (i.e., F_{sim} was significantly larger than 1). The model predictions for asymptotes and CPTs are shown in the bottom panel of Fig. 10. The model reproduced the interaction between memory load and similarity on the asymptotes, as well as the roughly linear increase of CPTs over MD.

Because in the training data a difference in processing speed between similar and dissimilar sets emerged, we also tested a model version that allowed different rates for the two similarity conditions. This led to no significant improvement in fit over model version 5b. Allowing

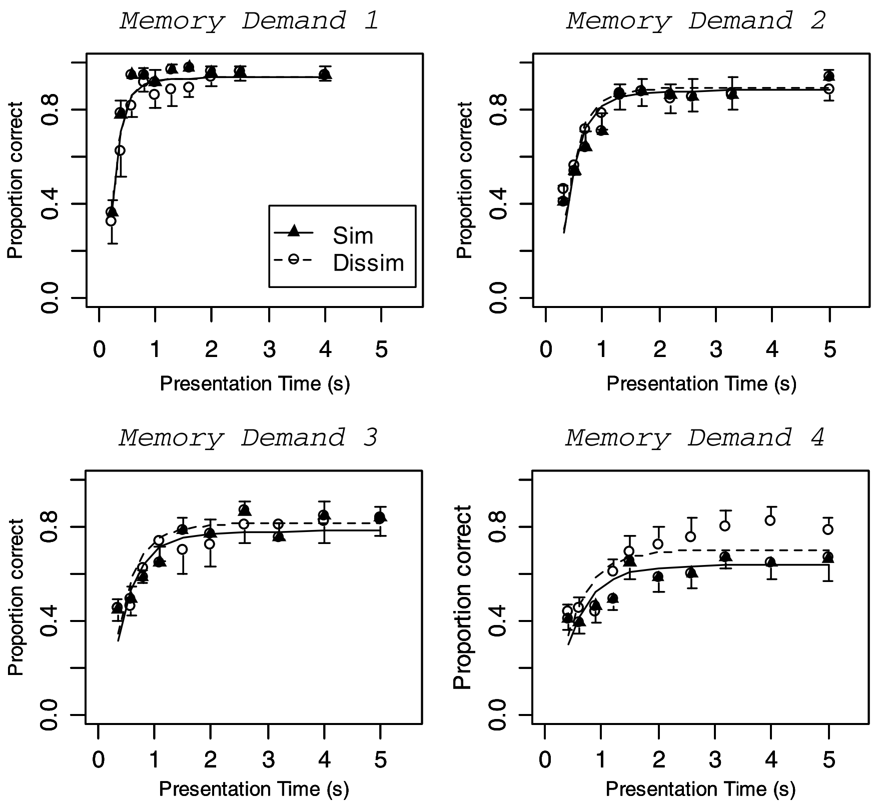

Fig. 9. Data from Experiment 3 (points) and predictions of the interference model (lines), averaged over participants. Error bars represent 95% confidence intervals.

different noise parameters, however, improved the fit significantly, although not dramatically (AIC = −963, BIC = −913, log-Likelihood = 492.4, $R^2_{adj} = .763$). The noise parameter for similar items was estimated at .171, whereas σ for dissimilar items was .184. In this model version, F_{sim} increased to 1.20 [CI = 1.13, 1.27], reflecting a more pronounced difference between the C parameters in the two similarity conditions. One explanation for this set of results is as follows. The larger degree of feature overlap in the similar set was in part counteracted by the smaller amount of noise in that set, possibly because the early part of the alphabet was more clearly represented in participants' long-term memory. When the model is not permitted to distinguish the two noise parameters, it uses an average noise parameter that leads to underprediction of performance in the similar condition, and this is partly compensated by understating the degree of overlap in that condition.

The model estimates contained one surprising outcome, the high value estimated for the rate parameter. Letter arithmetic should be a relatively slow process compared to digit arithmetic or shifts of spatial positions. The high r does not reflect a misspecification of the model, however, but truly represents a feature of the data. The CPTs for MD 1, which can be regarded as the purest assessment of processing speed in the present paradigm, were substantially lower in letter arithmetic than in digit arithmetic (compare Fig. 10 with Fig. 6). For some reason, participants in Experiment 3 accomplished letter arithmetic operations faster than participants in the numerical updating experiment accomplished digit arithmetic. One reason for this could be that the letter arithmetic task involved a smaller set of operations (4) and a smaller set of possible results (6), so that retrieval competition for arithmetic facts was lower in the present experiment. Another reason could be the task-specific practice prior to the assessment of memory updating.

Discussion

Experiment 3 provided a successful parameter manipulation for the interference model. As predicted, phonologically similar memory contents yielded a larger interference parameter than phonologically dissimilar contents. This happened despite the fact that dissimilar items were associated with slower processing, as became

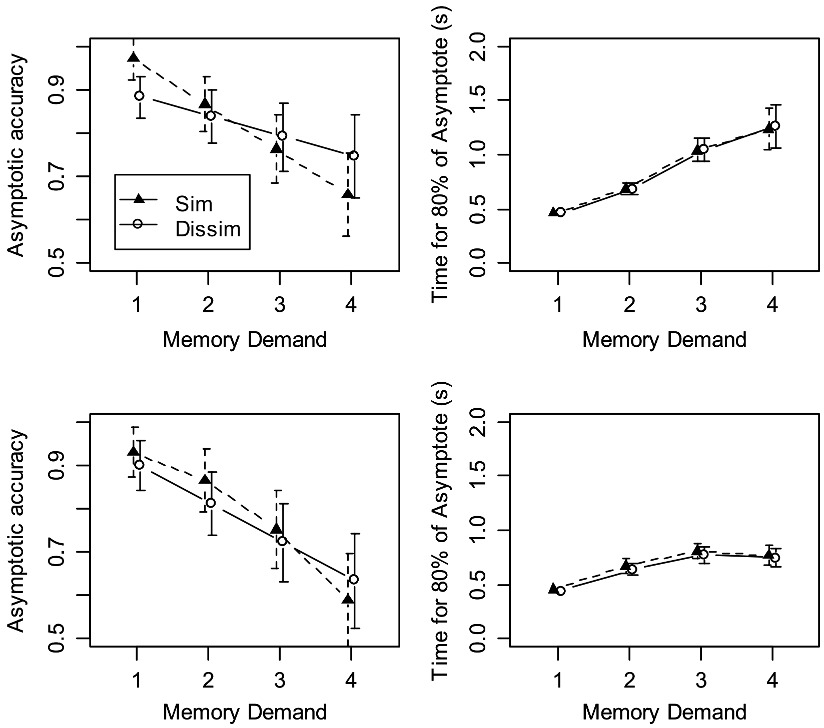

Fig. 10. Asymptotic accuracies and time demands for 80% of the asymptote (CPT 80) for similar and dissimilar letter sets in Experiment 3, estimated from fits of descriptive time-accuracy functions to the data (top panels) and to the predictions of the interference model (bottom panels). Error bars represent 95% confidence intervals.

Table 7
Model fits for memory updating with similar and dissimilar letters (Experiment 3)

Nr	Fixed effects	Random effects	Par	AIC	BIC	log-Lik	R^2_{adj}	Sign
0	C_{dis}, F_{sim}, r, σ	C_{dis}	6	−787	−761	400.0	0.671	
1	C_{dis}, F_{sim}, r, r_1, σ	C_{dis}	7	−845	−813	429.3	0.700	0
2	C_{dis}, F_{sim}, r, r_1, σ	C_{dis}, σ	8	−944	−909	480.1	0.748	1
3	C_{dis}, F_{sim}, r, r_1, σ	C_{dis}, r	8	−852	−816	433.9	0.711	1
4	C_{dis}, F_{sim}, r, r_1, σ	C_{dis}, r, r_1	9	−860	−821	439.5	0.720	3
5	C_{dis}, F_{sim}, r, r_1, σ	C_{dis}, r, r_1, σ	10	−952	−908	486.2	0.759	4
6	C_{dis}, F_{sim}, r, r_1, σ	C_{dis}, F_{sim}, r, r_1, σ	11	−950	−901	486.2	0.759	not 5

Legend: Par, number of free parameters; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; log-Lik, log-Likelihood; Sign, model fit is significantly better than the nested model with the number in this column (and all worse fitting nested models).

Note. Estimates are based on 868 data points. Correlations were fixed to zero.

Table 8
Parameter estimates of interference model (Version 5) for memory updating with similar and dissimilar letters (Experiment 3)

	C_{dis}	F_{sim}	r	$R_1 - r$	σ
Mean	0.34 [0.32, 0.36]	1.09 [1.05, 1.13]	2.10 [1.82, 2.39]	0.89 [0.63, 1.15]	0.18 [0.16, 0.20]
Standard deviation	0.013 [0.003, 0.062]	(Fixed)	0.36 [0.19, 0.68]	0.16 [0.03, 0.91]	0.026 [0.015, 0.044]

Note: The top two rows of the table contain estimates for the parameter means and their standard deviations in the participant sample; as there were only eight subjects, correlations can not be estimated reliably and therefore were fixed at zero.

apparent in the training data. The faster processing of items from the similar set provided an opportunity to dissociate storage and processing in working memory. The same experimental manipulation that increased the processing speed for single items reduced the capacity to hold several items in working memory simultaneously. Such a dissociation constitutes another challenge for a decay based model of capacity limits in working memory. A crucial assumption of decay models is that processing of items competes with rehearsal (Barrouillet et al., 2004; Salthouse, 1996; Towse, Hitch, & Hutton, 2000). If an experimental manipulation speeds up processing, then there should be more time for rehearsal. In the present experiment, there is no reason to believe that similar items take longer to rehearse, or decay faster than dissimilar items. Therefore, decay theories have no obvious means to explain why similar items are remembered worse than dissimilar ones under conditions of high memory load.

Of course, decay based models—as well as models assuming limited resource pools—can account for effects of within-domain similarity on recall accuracy, but only by introducing an additional mechanism generating similarity-based interference (e.g., Page & Norris, 1998). The success of our model shows that the particular interference mechanism it incorporates is sufficient to explain the observed capacity limits in at least one working memory task in great detail. This finding raises the possibility that other mechanisms such as decay or resource limitations are not necessary to explain the limited capacity of working memory.

General discussion

The goal of this work was to explore a simple formal model of capacity limits and processing dynamics in working memory. We focused on the interference model because that model had emerged as the most successful one in a previous study that comparatively fitted six models to time-accuracy functions of young and old adults on the numerical memory-updating task (Oberauer & Kliegl, 2001). Here we developed an improved version of the interference model based on a more detailed explication of its representation and processing assumptions. The model explains the effects of increasing memory demand—defined as the number of independent items to be held available at the same time—on both accuracy and processing speed. With increasing memory demand, each individual item suffers from interference from more other items. Interference means that part of an item's feature-based representation is overwritten, resulting in less overall activation of the representation, and thus to a reduced probability of recalling the item, as well as slower processing.

Instead of fitting the model separately to each individual, we used NLME as an integrated framework for fitting simultaneously fixed effects (i.e., parameter means) and random effects (i.e., standard deviations and correlations). This approach uses substantially fewer free parameters (e.g., only 15 parameters for model version 5 to fit 999 data points in Experiment 1, compared to four parameters for each of 21 participants = 84 parameters for individual fits and increasing with each additional participant), and thereby reduces the chance of overfitting (i.e., adapting parameter estimates to random fluctuations across participants). We successfully extended the model to a spatial version of memory updating and to a combination of numerical and spatial updating, and we provided an experimental validation of the overlap parameter C by manipulating phonological similarity in a letter updating task.

Evaluation of the interference model

How good is the quantitative fit of the interference model to the present data sets? It is important to note that using the NLME framework enabled us to account for the variance associated with a large number of data points with an extremely small number of free parameters. The ratio of free parameters to data points (1:50 or more) is an order of magnitude smaller than in most conventional model fits. Given that there is considerable noise in the individual data (see Fig. 2), we feel that accounting for 80–90% of the variance is about as much as one can expect with a framework such as NLME that protects against overfitting (i.e., fitting unsystematic variance). In the following we use the adjusted R^2 values of the descriptive TAF fits to the data as a benchmark for how much variance we can expect a model such as our interference model to explain. The R^2_{adj} of descriptive functions for numerical updating was .888, compared to .877 for the interference model (version 5). For spatial updating (Experiment 1), the R^2_{adj} were .852 (descriptive) and .830 (model); for the dual task, the descriptive values were .854 and .817 (numerical and spatial, respectively) and .842 (model), and for the letter updating task, .811 (descriptive) and .759 (model). Thus, except for letter updating (Experiment 3), the model came reasonably close to the maximum of systematic variance that it can be expected to account for.

We did not formally compare the interference model to competing models in this article, so readers might wonder whether other models of about equal complexity could give an equal or even better account of the data. We doubt this. In our previous study (Oberauer & Kliegl, 2001), we found that four other models—two based on a constant limited amount of resources, one assuming a "magical number" of chunks that can be held in working memory, and one based on crosstalk at retrieval of target items from cues—had a common problem: they could not reproduce the accelerated decline of asymptotic accuracy with increasing memory demand, together with a slowing of processing speed relative to that asymptote with increasing MD (as reflected by the CPTs). Only the interference model and the decay model reproduced this pattern. The present data from Experiment 1 showed the same shape of the MD effects on asymptotic accuracy and CPTs. The models that did not do well on the numerical updating data would therefore show qualitative misfits to the data of the present Experiment 1 as well. The decay model can handle the shape of the MD function on asymptotic accuracy, but it has problems with accounting for the data of Experiments 2 and 3, because it has no mechanism for explaining why asymptotic accuracy varies systematically with the similarity of the material held in working memory. Among the models that have so far been formalized for application to time-accuracy functions, the interference model is the only one that reproduces the qualitative patterns in the data presented here and in Oberauer and Kliegl (2001).

The interference model in its present form does, however, show one systematic quantitative misfit, it underpredicts the increase of CPTs over memory demand in the spatial task (Experiments 1 and 2) and in the letter arithmetic task (Experiment 3). There are several possibilities for improving the model in this regard. One straightforward amendment would be to let rate increase with memory demand. There is no theoretical rationale for doing this, however. A second possibility is to let the noise parameter increase with memory demand. With an increasing amount of feature overlap we can expect that not only the mean proportion of overwritten feature units increases but also the trial-to-trial variability in that proportion, because not every pairwise combination of items has the same degree of overlap, and also because there might be some randomness in which of two items sharing a feature unit will lose that unit. Variability in the proportion of overwritten features naturally translates into variability of available activation A_i, so that increasing the noise parameter σ as a function of overall overlap is a plausible way of improving the model. A third possible source of the increasing time demand with higher levels of memory load could be that rehearsal improves memory at long presentation times. Rehearsal could be beneficial even in the absence of decay if it involves processes going beyond pure maintenance rehearsal. Maintenance rehearsal in itself would not be expected to improve memory when feature overwriting is the main cause of forgetting—repeating an item to oneself could help to re-establish some of that item's lost features, but only at the expense of other items that had previously grabbed these features. Elaborative rehearsal, however, could improve memory by enriching the representations with more features, or by recoding items into more distinct representations with less feature overlap. In the spatial updating task, for instance, people might relate dot positions to each other across several frames, forming an integrated pattern. In the letter updating task, they might try to find a word that integrates the current set of letters.

Implications for capacity limits in working memory

What does the present research imply for the general question of what causes capacity limits it working memory? It should be clear, not only from our Experiments 2 and 3 but also from the large literature on similarity effects in immediate recall paradigms (from Conrad & Hull, 1964; to Conlin et al., 2005) that any viable model must include some mechanism of similarity-based interference. Our present work shows that a model based on two forms of similarity-based interference—overwriting of shared features and confusion of similar items at recall—is sufficient to account for a particularly rich set of data from one working memory paradigm. Our research does not rule out that, in addition, other factors such as time-based decay or resource limitations contribute to forgetting, and thereby to capacity limits in working memory. Other data might force us to consider a more complex model including multiple determinants of capacity limitations. Our reading of the literature on whether there is time-based decay in working memory is that there is as yet no compelling evidence for a substantial contribution of decay to forgetting in immediate memory tasks (Nairne, 2002), and fairly strong evidence against it (Lewandowsky, Duncan, & Brown, 2004). It still remains to be shown, however, that the interference mechanisms built into our model are sufficient to explain forgetting in other paradigms used to study short-term and working memory.

One popular paradigm used to study working memory is known as the complex span paradigm, in which participants must alternate between encoding memory items and processing other information. The presently most elaborated model for that paradigm is the time-based resource sharing model of Barrouillet et al. (2004). This model assumes time-based decay as the cause of forgetting. Support for the model comes from four related findings: (1) memory performance in the complex span paradigm is unaffected by the number of processing steps to be performed between encoding of successive items; (2) memory is unaffected also by the total time spent on the processing task, but (3) it declines with increasing rate of processing (i.e., number of steps per time), and (4) memory also declines when each processing step requires more time. These findings are explained by the time-based resource sharing model as follows: items in working memory decay over time. An attentional bottleneck can be devoted either to executing a processing step or to retrieving a memory item, thereby refreshing that item's memory trace. With a slower processing rate or shorter duration of individual processing steps, there are more opportunities for the bottleneck to switch rapidly between executing a processing step and refreshing some of the memory items, thereby counteracting decay. The main determinant of forgetting, therefore, is the proportion of time in which the attentional bottleneck is occupied by the processing task, so that it cannot combat decay.

Although the evidence for this model is impressive, none of it forces us to assume decay as the source of forgetting. The findings obtained with the complex span paradigm strongly suggest that the time available for refreshing memory traces in between processing steps is the main determinant of memory performance. This idea is equally compatible with interference as the source of forgetting. We can think of a version of the time-based resource sharing model in which each processing step generates an amount of interference that depends on the degree of overlap between the current memory items and the representations generated for the processing task. Whenever the attentional bottleneck is available, it can be used to repair partially degraded memory traces through retrieval and redintegration (Lewandowsky, 1999; Schweickert, 1993). One advantage of this interference version of Barrouillet et al's idea is that it can explain why the similarity between the memory items and the material involved in the processing tasks affects performance (Conlin et al., 2005; Li, 1999). An obvious next step on the route to a more comprehensive formal model of capacity limits in working memory is to develop a formal model of the decay-based and of the interference-based version of the time-based resource sharing framework, and to focus empirical investigations on their differential predictions.

The details of our assumptions about the representations and processes in working memory, for instance the binding mechanism using synchronous firing, and the feature overwriting mechanism it entails, are probably not essential to the success of the interference model. Other forms of feature interaction, for instance the migration of features from one item to another (cf. Jefferies, Frankish, & Lambon Ralph, 2006), might underlie the partial degradation of representations. (The parameter C would then represent the probability of a feature to migrate from one representation to one particular other representation currently held in working memory). The formal model is cast on a very abstract level, leaving open many details and making many simplifying assumptions. The advantage of having a simple model with few free parameters comes at the cost of underspecifying many aspects of processes in working memory, and hopefully misspecifying only a few of them. Models on such an abstract level should therefore be complemented by detailed simulations (i.e., computational models) that implement the assumptions that we outlined here only verbally. Fine-grained implementations of the interference model would enable us to make specific predictions arising from different ways in which features interact, and how this interaction gives rise to the capacity limit of working memory.

Conclusions

To conclude, the interference model provides satisfactory quantitative fits to four data sets, covering several basic phenomena associated with working memory capacity. No other formal model we are aware of gives an equally accurate and parsimonious account for these or another comparatively rich set of data. We take the success of the interference model as support for its underlying assumptions: the capacity of working memory is limited by mutual interference between the items held available simultaneously. Interference arises from interactions between features of item representations, which lead to partially degraded memory traces. The degradation of representations in turn leads to slower processing and to retrieval errors. In addition, other items in working memory compete with the target item for recall, and that competition becomes larger as more items are held in working memory and as they are more similar to each other.

References

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Mahwah, NJ: Erlbaum.

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., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short term memory. *Journal of Verbal Learning and Verbal Behavior, 14*, 575–589.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133*, 83–100.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107*, 127–181.

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.

Carpenter, P., Just, M. A., & Shell, P. (1990). What one intelligence test measures: a theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97*, 404–431.

Conlin, J. A., Gathercole, S. E., & Adams, J. W. (2005). Stimulus similarity decrements in children's working memory span. *Quarterly Journal of Experimental Psychology, 58A*, 1434–1446.

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

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

Cowan, N. (2005). *Working memory capacity*. New York: Psychology Press.

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.

Farmer, E. W., Berman, J. V. F., & Fletcher, Y. L. (1986). Evidence for a visuo-spatial scratch-pad in working memory. *Quarterly Journal of Experimental Psychology, 38A*, 675–688.

Garavan, H. (1998). Serial attention within working memory. *Memory & Cognition, 26*, 263–276.

Healy, A. F. (1977). Pattern coding of spatial order information in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 16*, 419–437.

Jefferies, E., Frankish, C., & Lambon Ralph, M. A. (2006). Lexical and semantic binding in verbal short-term memory. *Journal of Memory and Language, 54*, 81–98.

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

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics, 49*, 227–229.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working-memory capacity: a latent-variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189–217.

Kieras, D. E., Meyer, D. E., Mueller, S., & Seymour, T. (1999). Insights into working memory from the perspective of the EPIC architecture for modeling skilled perceptual-motor and cognitive human performance. In A. Miyake & P. Shah (Eds.), *Models of working memory. Mechanisms of active maintenance and executive control* (pp. 183–223). Cambridge: Cambridge University Press.

Klauer, K. C., & Zhao, Z. (2004). Double dissociations in visual and spatial short-term memory. *Journal of Experimental Psychology: General, 133*, 355–381.

Kliegl, R., Mayr, U., & Krampe, R. T. (1994). Time-accuracy functions for determining process and person differences: an application to cognitive aging. *Cognitive Psychology, 26*, 134–164.

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

Lange, E. B., & Oberauer, K. (2005). Overwriting of phonemic features in serial recall. *Memory, 13*, 333–339.

Lewandowsky, S. (1999). Redintegration and response suppression in serial recall: a dynamic network model. *International Journal of Psychology (Special Issue: Short-Term/Working Memory), 34*, 434–446.

Lewandowsky, S., Duncan, M., & Brown, G. D. A. (2004). Time does not cause forgetting in short-term serial recall. *Psychonomic Bulletin & Review, 11*, 771–790.

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

McClelland, J. L. (1979). On the relations of mental processes: an examination of systems of processes in cascade. *Psychological Review, 86*, 287–330.

McElree, B., & Dosher, B. A. (1989). Serial position and set size in short-term memory: the time course of recognition. *Journal of Experimental Psychology: General, 118*, 346–373.

Morey, C. C., & Cowan, N. (2004). When visual and verbal memories compete: evidence of cross-domain limits in working memory. *Psychonomic Bulletin & Review, 11*, 296–301.

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

Nairne, J. S. (2002). Remembering over the short-term: the case against the standard model. *Annual Review of Psychology, 53*, 53–81.

Navon, D. (1984). Resources—a theoretical soupstone? *Psychological Review, 91*, 216–234.

Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review, 7*, 403–423.

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. (2003). Selective attention to elements in working memory. *Experimental Psychology, 50*, 257–269.

Oberauer, K., & Göthe, K. (2006). Dual-task effects in working memory: interference between two processing tasks, between two memory demands, and between storage and processing. *European Journal of Cognitive Psychology, 18*, 493–519.

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., & Kliegl, R. (2004). Simultaneous execution of two cognitive operations—evidence from a continuous updating paradigm. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 689–707.

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.

Page, M. P. A., & Norris, D. (1998). The primacy model: a new model of immediate serial recall. *Psychological Review, 105*, 761–781.

Pinheiro, J. C., & Bates, D. M. (2000). *Mixed-Effect Models in S and S-Plus*. Berlin: Springer.

Pinheiro, J. C., Bates, D. M., DebRoy, S., & Sarkar, D. (2005). nlme: linear and nonlinear mixed effects models. R package (Version 3.1-73).

R Development Core Team (2005). R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. ISBN: 3-900051-07-0, URL: http://www.R-project.org.

Raffone, A., & Wolters, G. (2001). A cortical mechanism for binding in visual working memory. *Journal of Cognitive Neuroscience, 13*, 766–785.

Roberts, J. K. & Monaco, J. P. (2006). Effect-size measures for the two-level linear multilevel model. Paper presented at the Annual Meeting of the American Educational Research Association, April 9, 2006.

Saito, S., & Miyake, A. (2004). An evaluation of the task switching account of working memory span scores: evidence against a temporal decay assumption. *Journal of Memory and Language, 50*, 425–443.

Salthouse, T. A. (1996). The processing speed theory of adult age differences in cognition. *Psychological Review, 103*, 403–428.

Salthouse, T. A., Babcock, R. L., & Shaw, R. J. (1991). Effects of adult age on structural and operational capacities in working memory. *Psychology and Aging, 6*, 118–127.

Sanders, A. F., & Schroots, J. J. F. (1969). Cognitive categories and memory span III. Effects of similarity on recall. *Quarterly Journal of Experimental Psychology, 21*, 21–28.

Sarkar, D. (2005). lattice: Lattice Graphics. R package (Version 0.13-8).

Schweickert, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition, 21*, 168–173.

Smith, E. E., & Jonides, J. (1997). Working memory: a view from neuroimaging. *Cognitive Psychology, 33*, 5–42.

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.

Towse, J. N., Hitch, G. J., & Hutton, U. (2000). On the interpretation of working memory span in adults. *Memory & Cognition, 28*, 341–348.

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