Mem Cogn
DOI 10.3758/s13421-015-0512-8

The Hebb repetition effect in simple and complex memory span

Klaus Oberauer · Timothy Jones · Stephan Lewandowsky

© Psychonomic Society, Inc. 2015

Abstract The Hebb repetition effect refers to the finding that immediate serial recall is improved over trials for memory lists that are surreptitiously repeated across trials, relative to new lists. We show in four experiments that the Hebb repetition effect is also observed with a complex-span task, in which encoding or retrieval of list items alternates with an unrelated processing task. The interruption of encoding or retrieval by the processing task did not reduce the size of the Hebb effect, demonstrating that incidental long-term learning forms integrated representations of lists, excluding the interleaved processing events. Contrary to the assumption that complex-span performance relies more on long-term memory than standard immediate serial recall (simple span), the Hebb effect was not larger in complex-span than in simple-span performance. The Hebb effect in complex span was also not modulated by the opportunity for refreshing list items, questioning a role of refreshing for the acquisition of the long-term memory representations underlying the effect.

Keywords Immediate serial recall · Complex span task · Working memory · Long-term memory

Half a century ago, Donald Hebb (1961) asked participants in an experiment to remember lists of random digits for immediate recall in the order of presentation. Unbeknown to the participants, Hebb presented the same list on every third trial, interspersed with new random lists in the intervening two trials. Across 24 trials, immediate serial recall improved for the repeated but not the random lists. Although participants were not asked to remember the lists beyond the time of immediate test, more long-lasting memory traces accrued with the repetitions, which gradually improved people's ability to remember lists matching those traces.

Tests of immediate serial recall are routinely used to investigate short-term or working memory, also known as primary memory (from here on we will use the term working memory). Most tests of serial recall involve the simple-span procedure, in which people must recall a list of items immediately upon presentation in forward order. Because of its limited capacity, working memory is commonly assumed to hold only the current list, perhaps with a few traces of the immediately preceding one, but it is not thought to be suited to acquire a representation of the commonalities of lists spanning four trials. Therefore, the pervasive Hebb effect documents the contribution of some longer-lasting form of memory, referred to as long-term memory or secondary memory, to tests of immediate serial recall. The Hebb effect implies that lists maintained for immediate recall leave long-term memory traces, and that these traces are used in immediate recall (Burgess & Hitch, 2005, 2006; Page & Norris, 2009).

Here we investigate whether the Hebb repetition effect is also observed with two variants of the complex-span paradigm. The typical complex-span task differs from the simple-span task by the addition of a distractor task that is to be carried out in between pairs of list items during encoding; here we also investigate a less common variant in which distractors are interspersed between items at retrieval. The distractor task usually requires processing without any explicit memory demand, for instance reading sentences (reading span, Daneman & Carpenter, 1980), solving arithmetic

K. Oberauer (✉)
Department of Psychology – Cognitive Psychology, University of Zurich, Binzmühlestrasse 14/22, 8050 Zürich, Switzerland
e-mail: k.oberauer@psychologie.uzh.ch

T. Jones · S. Lewandowsky
University of Bristol, Bristol, UK

S. Lewandowsky
University of Western Australia, Perth, WA, Australia

Published online: 25 February 2015

problems (operation span, Turner & Engle, 1989), or carrying out a series of choice response tasks (Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007). Complex-span tasks have become popular in particular because their psychometric properties render them suitable for measuring working-memory capacity (Oberauer, Süß, Schulze, Wilhelm, & Wittmann, 2000; Wilhelm, Hildebrandt, & Oberauer, 2013), and they are good predictors of fluid intelligence (Conway et al., 2005; Conway, Kane, & Engle, 2003; Engle, Tuholski, Laughlin, & Conway, 1999). Although behavioral phenomena from complex span tests bear many similarities with those from simple-span tests, the two paradigms also differ in some regards. For instance, whereas the majority of errors in simple-span tests are order errors, item errors are more prevalent in complex span tests (Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012; Unsworth & Engle, 2007b). In correlational studies, when multiple simple span and complex-span tasks are used, the two types of tasks load on separate factors (Gathercole, Pickering, Ambridge, & Wearing, 2004; Kane et al., 2004). Those observed differences between the two types of span task render it plausible that simple-span and complex-span performance may also differ with regard to the Hebb effect.

As we discuss next, there are additional reasons to believe that an examination of the Hebb effect in complex span should be theoretically rewarding. Compared to the Hebb effect in simple span, there are equally plausible theoretical reasons to believe that the Hebb effect in complex-span should be greater, or that it might not be present at all. The goal of the present work is to investigate the empirical merits of these competing theoretical expectations.

The Hebb effect could be diminished or even abolished in complex span because the distractor task interrupts encoding of the list, thereby disrupting the formation of an integrated list representation. The Hebb effect in simple span appears to depend on an integrated representation of the list, as demonstrated by an experiment by Cumming, Page, and Norris (2003): After a learning phase with the standard repetition of one list every third trial, they introduced transfer lists matching the previously repeated list in every second list item, whereas the intervening list positions were filled with new items. There was no transfer from the learned list to recall of the repeated items on these transfer lists. Hitch, Fastame, and Flude (2005) investigated Hebb learning with training lists in which only every second item was repeated, rather than the complete list as in the standard Hebb paradigm. There was no evidence of learning in this condition. On the basis of their results, Cumming et al. (2003) as well as Hitch et al. (2005) argued that Hebb learning consists of the formation of a unified (chunked) representation of the list, or at least of segments of the list (for computational models implementing this idea see Burgess & Hitch, 2006; Page & Norris, 2009). The acquisition and use of such chunks is disrupted if repetition is limited to sub-components of learned chunks.

On this hypothesis, one might expect the Hebb effect to be at least diminished – if not absent altogether– in complex span: It is known that people find it difficult to exclude representations used for distractor-task processing from working memory, and those attempts are often not entirely successful (Oberauer, Farrell, Jarrold, Pasiecznik, & Greaves, 2012). If distractor materials are selected at random, as in our experiments reported below, then distractor representations arguably play a similar role to the not-repeated items in the training lists of Hitch et al. (2005), or in the transfer lists of Cumming et al. (2003): When representations of not-repeated distractors are interspersed with representations of repeated list items, formation or application of chunks could be disrupted, thereby diminishing the Hebb effect or abolishing it altogether.

Moreover, Hebb learning is incidental – after all, participants are not asked to remember a list any further after they have finished recalling it immediately after presentation. It is generally assumed that incidental learning does not depend on the person's intention to learn but on the kind and degree of processing of the material (Craik & Lockhart, 1972; Hyde & Jenkins, 1969). Thus, any material that is attended to and processed to some extent becomes a candidate for incidental learning. It follows that incidental learning of events during a complex-span trial is unlikely to be limited to list items at the exclusion of distractors: People must process the distractors just like list items, and indeed most complex-span experiments enforce an accuracy criterion on the distractor task to prevent people from focusing on the list alone (Conway et al., 2005). One would therefore expect Hebb learning to apply non-selectively to all events that a person attends to and processes during a complex-span trial, in which case distractor-task representations should be as much part of the long-term memory trace as the memoranda. When the distractors have no systematic relationship to the memoranda and differ across repetition trials as well – as was the case in all experiments presented below – such a composite trace would be largely useless for improving recall of a repeated list in the current Hebb paradigm. These reasons justify the prediction that the Hebb effect should be diminished or even abolished in the complex-span paradigm.

On the other side of the argument are considerations that cite the presumed greater involvement of secondary or long-term memory in complex span compared to simple span. Unsworth and Engle (2006) have argued that in simple-span tests, up to four list items can be held in working memory, whereas in complex span the distractor task pushes previous list items out of working memory. In consequence, at the point of recall after a complex-span list only the last one or two items can be recalled from working memory, whereas the remainder must be retrieved from long-term memory. If this assumption is correct, the Hebb effect might be expected to be

larger in complex span than in simple span: The Hebb effect reflects a gradually strengthened long-term memory representation of the repeated list, and the impact of that representation on immediate-recall performance should be larger the more that recall depends on long-term memory.

The assumption that long-term memory is more involved in complex span than in simple span received support from an observation first reported by McCabe (2008): When tested with a final free recall test for the words on all memory lists encountered in the experiment, participants were found to recall more words from complex-span than from simple-span lists. This effect has been replicated several times (Loaiza & McCabe, 2012; Loaiza, McCabe, Youngblood, Rose, & Myerson, 2011). According to McCabe, the effect arises because in complex span people cannot hold the entire list in working memory. They are thus forced to temporarily outsource parts of the list to long-term memory, and to bring them back into working memory through "covert retrieval" during the distractor-task phases. Because covert retrieval serves as retrieval practice, stronger long-term memory traces are established in complex-span than in simple-span, which does not require covert retrieval during list presentation. If covert retrieval during complex span contributes to the Hebb effect, then the strength of the Hebb effect in complex span should depend on the opportunity for covert retrieval. This opportunity can be varied through the "cognitive load" imposed by the distractor task. Cognitive load, as defined by Barrouillet et al. (2007), refers to the proportion of time available for the distractor task during which attention is actually occupied by the distractor task. When distractors are presented at a leisurely pace (e.g., one arithmetic step every 2 s), then cognitive load is said to be low because processing only takes up a fraction of the available time. When distractors are presented at a fast pace (e.g., one arithmetic step every 500 ms), cognitive load is high because the entire available time is required for processing. According to Barrouillet and colleagues, any remaining time in between distractors that is not taken up by processing can be used to attend to the representations of the memoranda, thereby refreshing them. The concept of refreshing as used by Barrouillet and colleagues (c.f. Raye, Johnson, Mitchell, Greene, & Johnson, 2007) is very similar to the concept of covert retrieval, as McCabe (2008) recognized. The two processes might not be the same, but for both it is assumed that they can be carried out only when attention is not occupied by a distractor task. It follows that varying cognitive load arguably also varies the opportunity for covert retrieval during a complex-span task. If covert retrieval plays a role in building the long-term memory representations underlying the Hebb effect, then the Hebb effect in complex span should be larger at low than at high cognitive load.

To summarize, theoretical considerations and existing evidence provide equally strong reasons for predicting that the Hebb effect in complex span should be larger than in simple span, or that it should be smaller or even non-existent. Through the following experiments we tested these contrasting predictions. Experiments 1 and 2 established that there is a Hebb effect with complex span, suggesting that distractors did not disrupt the formation of integrated list representations. Experiment 3 generalizes this finding to a version of complex span in which distractor processing is interspersed between recall rather than encoding of memoranda, thereby showing that the effect is resilient to disruptions at test. Finally, Experiment 4 directly compared the Hebb effect for simple and complex span. In addition, Experiment 4 varied the opportunity for covert retrieval (or refreshing) in a complex span paradigm, thereby testing the assumption of McCabe (2008) about the role of covert retrieval in that paradigm. We found that the size of the Hebb effect was unaffected by cognitive load. We conclude that the Hebb effect is a highly robust attribute of list learning that is unaffected by disruptive distractors at encoding or test.

Experiments 1 and 2

Participants performed a complex-span task, which required them to remember lists of consonants and to make size judgments on words displayed after each consonant. The same list was used on every third trial; we refer to those trials as the repetition trials. Each set of three consecutive trials, including one repetition trial and two non-repetition trials, will be called a cycle. We expected a Hebb repetition effect, that is, better immediate recall for repetition trials than new trials, especially at later cycles. Experiments 1 and 2 differed in only two regards: Experiment 1 involved memory lists of seven consonants; for Experiment 2 we increased list length to eight to create more room for improvement through learning. To compensate for the longer duration of trials, in Experiment 2 we reduced the number of trials from 27 to 24.

Method

Participants Participants were 32 (Experiment 1) and 27 (Experiment 2) members of the University of Western Australia community. They took part in a single 1-hour session in exchange for AUD$10 or course credit.

Materials For each new trial a memory list was constructed by sampling the required number of consonants without replacement from the set of all consonants except Q and Y. The list for the first repeated trial was constructed in the same way and then held constant for all repetitions. The repeated list was used in every third trial, beginning with trial three.

Materials for the distractor task consisted of 264 English nouns referring to concrete objects. They were selected from a larger set of nouns referring to objects varying across a broad

range of size, from "ladybird" to "sun." The participants' task was to judge for each word whether the object was smaller or larger than a soccer ball. To make the task unambiguous, we selected only the words referring to the 25 % largest and the 25 % smallest objects in the original set. Each word from the experimental set was used three times throughout the experiment. Words for each size judgment were drawn at random on every trial, including for the repetition trials. Thus, repetition trials had a constant memory list but variable distractor-task stimuli. The random selection of distractors maximizes the chance of distractor representations disrupting the formation of an integrated list representation, thereby creating a condition for which there are good theoretical reasons to expect the Hebb effect to disappear.

Procedure Each trial started with a fixation cross, followed after 3 s by the first letter displayed centrally in red for 1.5 s. The letter was immediately replaced by the first distractor word, displayed centrally in black. Participants judged whether the word referred to an object larger or smaller than a soccer ball by pressing the "/" (slash) key or the Z key, respectively, on the computer keyboard. Once a response was made, or after the maximum time of 2 s elapsed, the distractor disappeared and was replaced by the next word. Each letter was followed by four size judgments. The fourth size judgment was immediately followed by the next to-be-remembered letter, and so on until presentation of the list was completed. The very last size judgment was followed by a red question mark, prompting participants to commence recall by entering the first letter on the keyboard. The entered letter was displayed for 0.3 s, and was then replaced by the question mark again to prompt recall of the second letter, and so on until participants had given as many responses as there were letters in the list. Omissions were not allowed. The next trial commenced 2.5 s after the last recall response.

Results

We first report memory accuracy to test for the classic Hebb repetition effect. Next we ask whether repetition of the memory list had an impact on speed and accuracy of the distractor task. We analyzed all data with a Bayesian linear regression model, using the *BayesFactor* package (Morey & Rouder, 2012; Rouder, Morey, Speckman, & Province, 2012) for *R* (R Development Core Team, 2012). The *lmBF* function in the *BayesFactor* package estimates linear models and returns the Bayes factor (BF) of the model relative to a null model that predicts the data by the intercept alone. Two alternative models M_1 and M_2 can be compared to each other by dividing their BFs (relative to the null model). The ratio of the BFs of M_1 vs. null and M_2 vs. null is the BF of M_1 vs. M_2.

For each analysis we investigated two predictors, cycle and repetition. Cycle refers to the ordinal number of the eight sets of three consecutive trials, each including one repeated and two non-repeated lists. Cycle was entered as a continuous variable, centered on zero. For each analysis we estimated four models: M_c, with only a main effect of cycle; M_r, with only a main effect of repetition versus new trials; M_{add}, with additive effects of cycle and repetition; and M_{full}, with both additive effects and their interaction. Each of these models included subjects as a random effect, and therefore we also estimated M_b as a baseline model with only the intercept and the random effect of subjects. We assessed the strength of evidence for the main effect of cycle by $BF(M_c)/BF(M_b)$, and the main effect of repetition by $BF(M_r)/BF(M_b)$. Evidence for the interaction was assessed by $BF(M_{full})/BF(M_{add})$. BFs larger than 1 reflect evidence in favor of the model in the numerator; Bayes factors smaller than 1 reflect evidence in favor of the model in the denominator. The strength of evidence for the model in the denominator can be gauged by the reciprocal of the BF. For instance, if $BF(M_{full})/BF(M_{add}) = 0.5$, then the BF in favor of the additive model is 2. BFs <3 are usually regarded as evidence "barely worth mentioning;" BF between 3 and 10 as "substantial evidence," BF between 10 and 100 as "strong evidence," and BF >100 as "decisive" (Kass & Raftery, 1995).

Memory accuracy Memory performance was scored as the proportion of letters reported in their correct list positions. Figure 1 shows proportion correct by cycle and repetition (new vs. repeated). Table 1 summarizes the BFs reflecting the strength of evidence for the main effects and the interaction. The evidence for a main effect of cycle was substantial in Experiment 1 but weak in Experiment 2. There was compelling evidence for the main effect of repetition in both experiments. The interaction was supported only weakly in both cases.

The Hebb effect was primarily reflected in the main effect of repetition. Its size can be estimated by sampling from the posterior distribution, using the *posterior* function in the *BayesFactor* package (Morey & Rouder, 2012). The sample provides information about the mean and the 95 % credible interval of the effect, which are given in Table 2. The 95 % credible interval is the range in which the true effect size lies with a posterior probability of .95. Based on the findings from the first two experiments, we can say that the Hebb effect increases memory performance in complex span by 6–16 percentage points over 8–9 list repetitions.

Size-judgment performance Failures to respond to a size-judgment trial were scored as errors. Response times (RTs) of correct trials only were analyzed. We estimated Bayesian linear models with the same predictors as for memory accuracy. The resulting BFs are reported in Table 1; the data are plotted in Fig. 2. In both experiments accuracy improved and RTs declined over cycles. The BFs for the main effects of repetition show that list repetition had a beneficial effect on

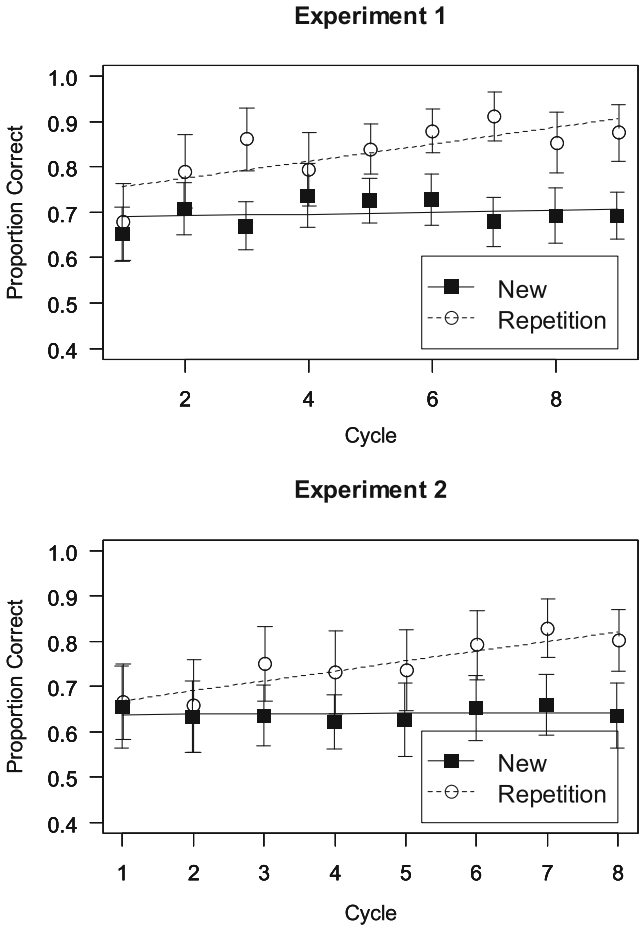

Fig. 1 Memory accuracy in Experiment 1 (**top**) and Experiment 2 (**bottom**). Error bars are 95 % confidence intervals (CIs) for within-subject comparisons (Bakeman & McArthur, 1996). The CIs can be interpreted in terms of classical null-hypothesis tests for pair-wise comparisons between data points: Two means differ significantly ($p < .05$) when their CIs overlap by less than 50 % of the interval between each mean and the corresponding CI boundary (G. Cumming & Finch, 2005). The straight lines are regression lines from the mean posterior parameters of linear models with cycle as predictor, applied separately to each repetition condition

RTs in Experiment 1, and on both accuracies and RTs in Experiment 2. Evidence for the interaction was non-existent in Experiment 1 and modest at best in Experiment 2.

Discussion

Experiments 1 and 2 established that there is a Hebb effect with complex span. Memory was better for repeated than for new lists. The beneficial effect of repetition emerged fairly rapidly – by the third cycle it was already strong – and this explains why there was only weak evidence for the interaction of repetition and the linear effect of cycle that would be expected from more gradual learning.

The repetition benefit extended to the distractor task: Size judgments were made faster, and in Experiment 2 also more accurately, in the context of repeated lists. This is a novel finding that we did not predict. Several post-hoc explanations could be offered. From the perspective of a resource theory, it could be argued that encoding and maintaining repeated lists consumes a smaller share of a limited resource, leaving more of that resource for concurrent processing. Other explanations could start from the assumption that people notice the list repetition and find the repeated lists easier to encode and maintain. In previous experiments with the Hebb paradigm, the majority of participants became aware of the list repetitions at some point during the experiment (McKelvie, 1987; Sechler & Watkins, 1991). McCabe (2010) has shown that merely anticipating an easier memory task leads people to respond faster to a concurrent processing task in a complex-span paradigm. When people perceive the memory task to be harder, they apparently devote more of the time in between items to further processing of the memoranda – this could involve consolidation, refreshing, covert retrieval, or elaboration – and therefore delay responding to the distractors.

Experiment 3

Before moving on to a direct comparison of the Hebb effect in simple and complex span we need to examine one possible explanation for the results of the first two experiments. It has been claimed that the Hebb effect arises primarily from learning of the output sequence, as opposed to the presented list (Cunningham, Healy, & Williams, 1984). Cunningham et al. (1984) found a Hebb effect only for lists that were initially recalled, not for lists repeatedly encoded but not recalled, suggesting that Hebb learning occurs only during recall. A later study with better control of learning times observed a robust Hebb effect also for not-recalled lists, but the effect was slightly larger for recalled lists (Oberauer & Meyer, 2009), implying that learning occurred both during encoding and recall. To the extent that the Hebb effect arises from learning during recall, our finding of a Hebb effect in Experiments 1 and 2 would be unsurprising, because in the standard complex-span paradigm that we used in those experiments, the output sequence consisted of uninterrupted recall of all list items, just like in simple span.

To test the possibility that the Hebb effect in Experiments 1 and 2 relied on learning during uninterrupted list recall, in Experiment 3 we used a variant of complex span in which the distractor episodes interrupt the output sequence instead (Lewandowsky, Duncan, & Brown, 2004): Recall of each item was preceded by a brief series of distractor operations. If Hebb learning occurred primarily during output, and if distractors disrupt the formation of associations between list items that support the Hebb effect, then we might expect the Hebb effect to disappear in Experiment 3.

Table 1 Bayes Factors for the linear models for Experiments 1–3, and the three span conditions of Experiment 4

Effect	E1	E2	E3	E4 Simple	E4 Complex low CL	E4 Complex high CL
	Memory accuracy					
Cycle	11.7	1.7	4.3×10^{7}	1.5×10^{5}	1.1	49.2
Repetition	1.0×10^{12}	4446.7	6.6×10^{7}	1.3×10^{17}	1723.2	863.7
Cycle × Repetition	4.6	2.1	47.8	2.4	0.2	3.2
	Size-judgment accuracy					
Cycle	6922.2	38.3	0.2		25.4	34.9
Repetition	0.16	121.3	0.3		2.6×10^{5}	0.3
Cycle × Repetition	1.6	0.15	0.2		0.4	0.1
	Size-judgment response time					
Cycle	4×10^{26}	5×10^{17}	3.0×10^{17}		6×10^{5}	2×10^{32}
Repetition	243.3	139.9	112.5		0.7	0.2
Cycle × Repetition	2.8	5.3	0.11		0.2	0.9

Note. E1–E4 denote Experiments 1–4; CL = cognitive-load condition

Method

Participants Twenty-three members of the University of Western Australia campus community took part in a single 1-hour session in exchange for AUD$10 or course credit.

Materials and procedure The experiment was identical to Experiment 2 with one exception: The distractor episodes were moved from the encoding to the recall phase. Specifically, recall of each letter – prompted by a red question mark – was preceded by four size judgments on words.

Results

We analyzed the data in the same way as for the first two experiments.

Memory accuracy The proportion of letters recalled in correct order is plotted in Fig. 3 as a function of cycle and repetition (new vs. repeated). The comparison of Bayesian regression models returned strong evidence for the main effects of both cycle and repetition (repeated vs. new), as well as their interactions. The BFs are given in Table 1. The size of the Hebb effect, reflected in the posterior density of the main effect of repetition, was comparable to the effect sizes of the first two experiments (see Table 2) – numerically it was somewhat larger than in the preceding two experiments, perhaps because there was more room for improvement, given that accuracy started at a lower level in the first cycle. That said, the credible intervals of the posterior distributions of the effect sizes overlap considerably, so that this numerical difference is unlikely to be systematic. The conservative conclusion is that the Hebb effect in Experiment 3 was at least as large as in Experiments 1 and 2. Clearly, interrupting recall by a distractor task does not reduce or abolish the Hebb effect.

Table 2 Means and 95 % credible intervals for the Hebb Effect

Experiment	Mean	95 % credible interval
1	0.13	[0.10, 0.16]
2	0.10	[0.06, 0.14]
3	0.15	[0.11, 0.19]
4: Simple span	0.12	[0.09, 0.15]
4: Complex span, low CL	0.08	[0.05, 0.12]
4: Complex span, high CL	0.08	[0.04, 0.11]

CL = cognitive-load

Size-judgment performance Proportion correct and mean RTs of the size judgments are presented in Fig. 4. The BFs for the linear models on these data are included in Table 1. There was no evidence for any main effect or their interaction on the judgment accuracies. Participants responded increasingly faster over the course of the experiment, reflecting practice with the size judgments. The main effect of repetition shows shorter RTs for distractors in trials with repeated lists. This result replicates the Hebb effect on distractor RTs already observed in the first two experiments.

Discussion

Experiment 3 demonstrates the Hebb effect for a variant of complex span in which distractor processing is interleaved with recall rather than study. To the extent that the Hebb effect relies on learning during retrieval, interruption of the recalled list sequence does not disrupt learning the list, and does not disrupt application of what has been learned.

Experiments 1–3 provide an existence proof for the Hebb effect in complex span. This is a novel result that rules out the possibility that distractors – whether at encoding or at test –

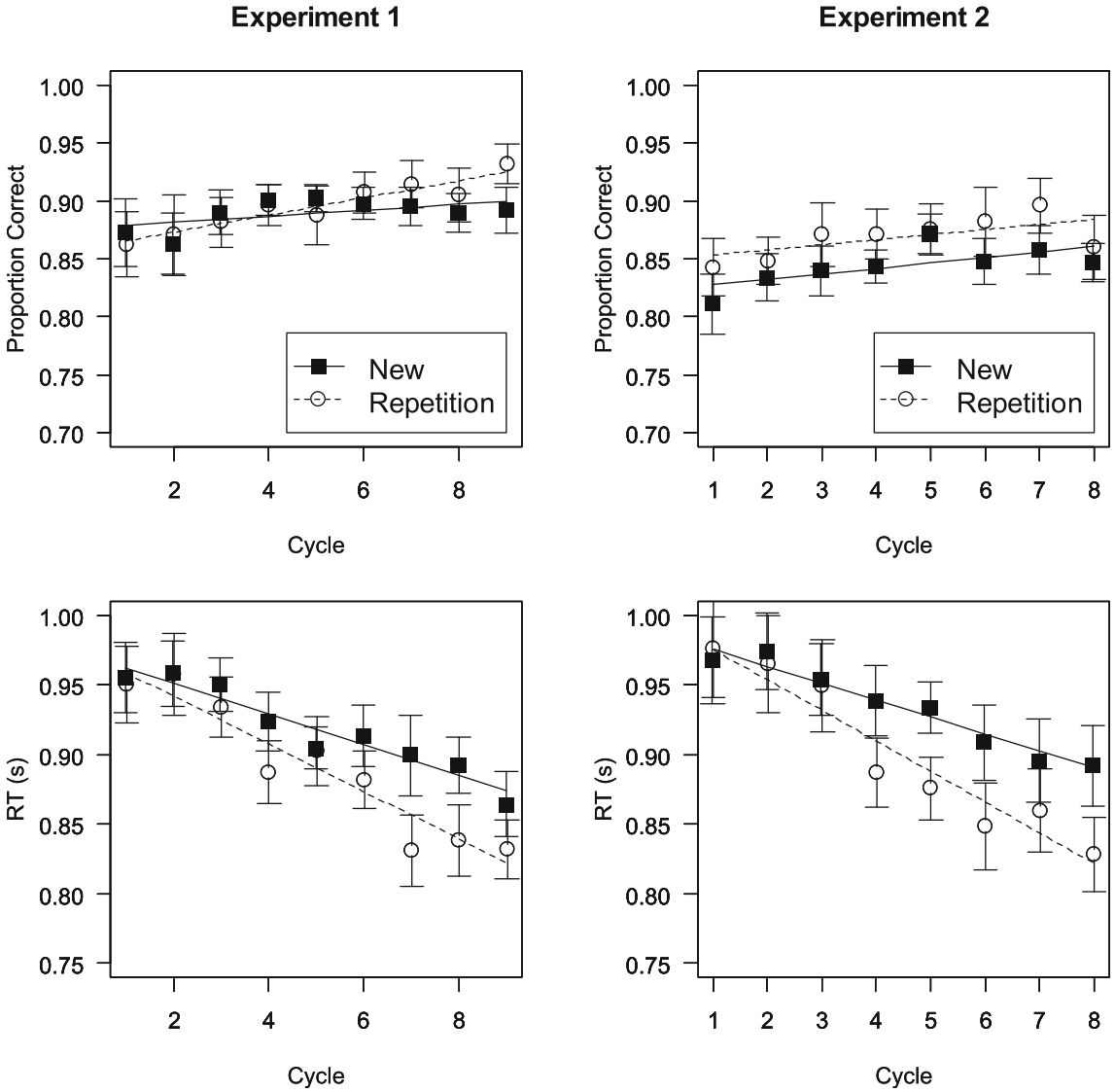

Fig. 2 Performance in the size-judgment task in Experiment 1 (**left**) and Experiment 2 (**right**). Error bars are 95 % confidence intervals for within-subject comparisons. The straight lines are regression lines from the mean posterior parameters of linear models with cycle as predictor, applied separately to each repetition condition

disrupt the formation of integrated list representations that are thought to support Hebb learning. The next experiment examined whether the Hebb effect might benefit from covert "retrieval practice" during encoding, by manipulating the opportunity for such covert retrieval in a conventional complex-span paradigm.

Experiment 4

In Experiment 4 we compared the Hebb effect in simple- and complex-span tasks, using the standard complex-span paradigm with distractors during encoding. In addition, we varied the cognitive load in the complex-span task to manipulate the opportunity for covert retrieval during a complex-span task. If covert retrieval causes long-term memory traces to be laid down and strengthened, and if these memory traces underlie the Hebb effect, then the Hebb effect should be reduced when we increase cognitive load.

Method

Participants Thirty students from the University of Western Australia took part in three 1-hour sessions for financial reimbursement (at the rate of AUD$10/hr) or course credit. Two participants took part in only one session, and therefore we excluded their data from analysis.

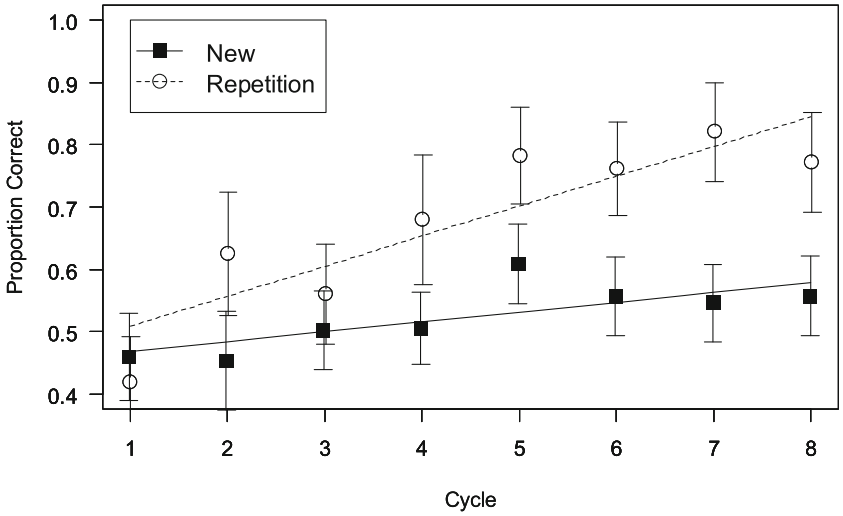

Fig. 3 Memory accuracy in Experiment 3. Error bars are 95 % confidence intervals for within-subject comparisons. The straight lines are regression lines from the mean posterior parameters of linear models with cycle as predictor, applied separately to each repetition condition

Materials and procedure The materials and procedure were as in Experiment 2, with the following modifications: In one session, participants were tested with the complex-span task exactly as in Experiment 1; here we will refer to this as the low cognitive-load condition. In a second session participants were tested on the complex-span task with high cognitive load. We increased cognitive load by shortening the time window for each size-judgment trial from 2 s to 1.2 s. Thus, the next stimulus already appeared 1.2 s after each word, irrespective of whether people had made a response during that time. In the third session, participants were tested on a simple-span task, created by cutting out the distractor-task periods. That is, presentation of each to-be-remembered letter was immediately followed by presentation of the next letter. Because simple-span trials took less time than complex-span trials, we ran two blocks of 24 trials of simple span; each block used a different repeated list. The order of sessions was counterbalanced across participants.

Results

We ran two Bayesian linear-model comparisons for each dependent variable. One included span type (simple vs. complex) as a third predictor (in addition to cycle and repetition),

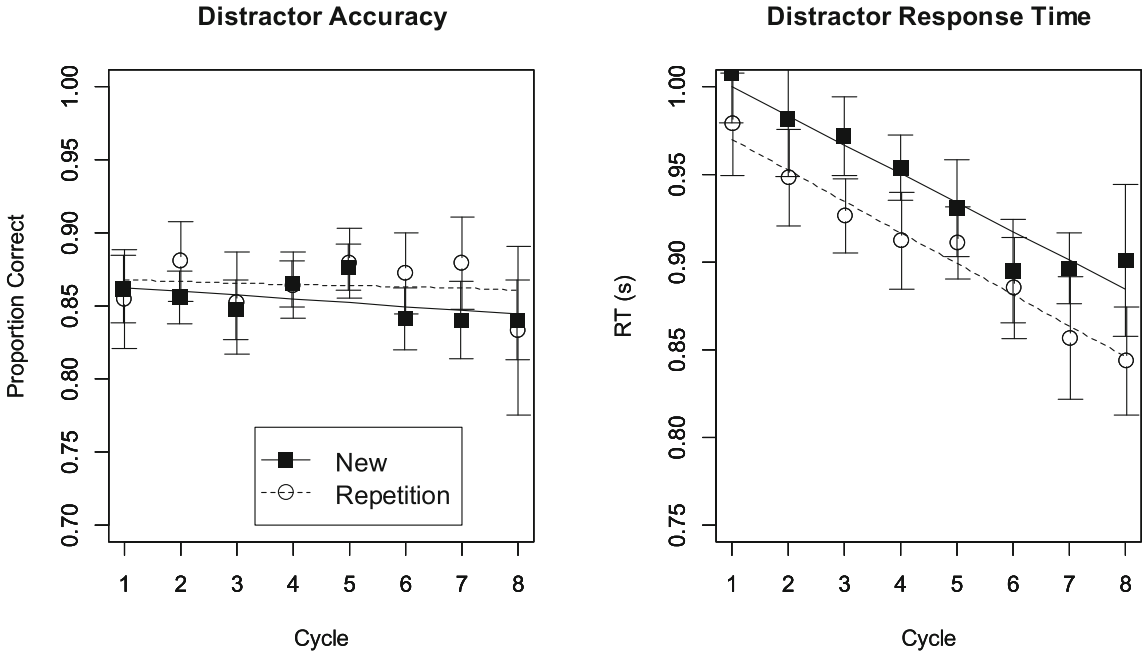

Fig. 4 Performance in the size-judgment task in Experiment 3. Error bars are 95 % confidence intervals for within-subject comparisons. The straight lines are regression lines from the mean posterior parameters of linear models with cycle as predictor, applied separately to each repetition condition

collapsing over cognitive load. The other focused on the complex-span conditions, contrasting high and low cognitive load, omitting the data from the simple-span session. Both analyses now involved three predictors, so that a more complex set of model comparisons was needed to assess all interactions. We estimated the full model, including all main effects and all interactions, and compared that against a progression of reduced models created by eliminating first the three-way interaction, then additionally each of the two-way interactions. We determined the BF for each interaction by dividing the BF of the model including that interaction by the BF for the model that eliminated that interaction but was otherwise identical. For instance, evidence for the interaction of repetition and cognitive load was assessed by the ratio of the BF for a model including all three two-way interactions (but not the three-way interaction) to the BF for a model including only the other two two-way interactions (cycle × repetition, and cycle × cognitive load), but eliminating the repetition × cognitive-load interaction. Table 3 summarizes the resulting BFs for the main effects and interactions in both analyses.

In addition, we ran the regression analysis with cycle and repetition as predictors within each of the three span-type conditions separately. The resulting BFs are added to Table 1 for comparison with the preceding experiments.

Table 3 Bayes Factors for linear models, Experiment 4

	Memory accuracy	Size-judgment accuracy	Size-judgment response time
Simple span vs. complex span			
Cycles	76,353.0		
Repetition	1.8×10^{14}		
Simple vs. complex (SC)	9.3×10^{6}		
Cycles × Repetition	5.6		
Cycles × SC	0.18		
Repetition × SC	0.56		
Three-way	0.14		
Complex span: low vs. high CL			
Cycles	57.0	101.8	1.5×10^{6}
Repetition	470,926.2	21.1	0.17
CL	1.4×10^{6}	9.6×10^{57}	6.6×10^{147}
Cycles × Repetition	1.5	0.20	0.39
Cycles × CL	0.20	0.16	0.32
Repetition × CL	0.11	0.56	0.17
Three-way	0.26	0.21	0.13

CL = cognitive-load

Memory performance Memory accuracy (plotted in Fig. 5) increased over cycles, and was better for repeated than for new lists; the BFs for those two main effects imply decisive evidence. The interaction of cycle with repetition was moderately supported (BF >5), implying continued Hebb learning of the repetition list over cycles. Unsurprisingly, participants did better in simple- than in complex-span tasks.

None of the interactions involving span type was supported. In fact, the BFs provided some evidence against these interactions, and in favor of an additive model excluding the interactions. The evidence in favor of the model excluding an interaction compared to the same model including it can be gauged by the reciprocal of the BF for that interaction in Table 3. The BF in favor of omitting the interaction of repetition with span type was 1/0.56 = 1.7. Thus, the data speak somewhat more in favor of an additive model of repetition with span type than against it, although the evidence is weak. The analyses of each span-type condition separately yielded strong evidence for the main effect of repetition in each condition, demonstrating that the repeated lists benefited from learning regardless of span type.

Table 2 shows that the posterior mean of the Hebb effect was slightly smaller for the complex span (for both levels of cognitive load) than for the simple span. We can assess the strength of evidence against each directed hypothesis separately by computing one-sided BFs (Wagenmakers & Morey, 2013). First, we calculated the one-sided BF for the hypothesis that the Hebb effect for complex span is *smaller* than that for simple span, compared to the alternative that it is equal to or larger than that for simple span. This BF was 1.09, implying ambiguous evidence. Second, we calculated the one-sided BF for the hypothesis that the Hebb effect for complex span is *larger* than that for simple span, compared to the alternative that it is equal to or smaller than that for the simple span. This BF was 0.03, implying strong evidence against the directed hypothesis (BF = 1/0.03 = 33.1). Thus, the present data provide evidence that is equally compatible with the hypotheses that the Hebb effect is equal for both span types or that it is smaller for complex than for simple span; at the same time, the data provide strong evidence against the assumption that the Hebb effect is larger for complex span.

The analysis focusing only on complex span confirmed the main effects of cycles and of repetition, with scant evidence for their interaction. As expected, memory was better with lower cognitive load. There was no evidence for any interaction involving cognitive load. Rather, the results provided fairly substantial evidence against the possibility that the Hebb effect was reduced with higher cognitive load: The BF against the interaction of repetition with cognitive load was 1/0.11 = 9.1.

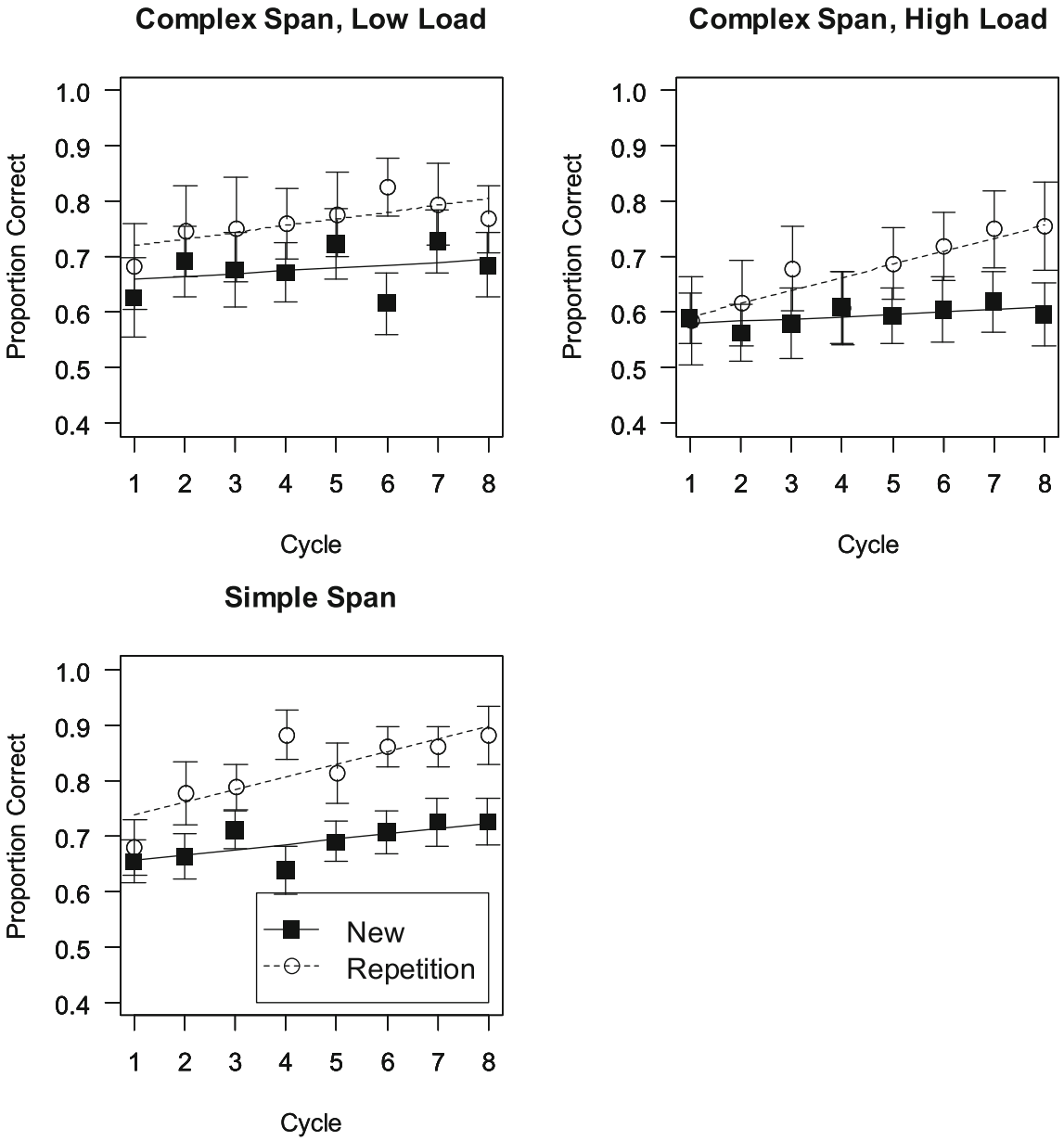

Fig. 5 Memory accuracy in Experiment 4. Error bars are 95 % confidence intervals for within-subject comparisons. The straight lines are regression lines from the mean posterior parameters of linear models with cycle as predictor, applied separately to each repetition condition

Size-judgment performance Accuracy and mean RTs of size judgments in the complex-span conditions are plotted in Fig. 6. The BFs for the linear models on these data (see Table 3) reflect a simple pattern: Performance improved over cycles. With higher cognitive load, responses were faster but less accurate, reflecting the increased time pressure. On repetition trials accuracy was slightly improved but RT was unaffected by list repetition. None of the interactions was supported by the data. In particular, there was no evidence that cognitive load modulated the repetition effect, and for RTs there was even modest evidence against that proposition, BF = 1/0.17 = 5.9.

Discussion

Experiment 4 replicated the Hebb effect in complex span. The effect was observed for memory performance, and for the distractor task it was reflected in accuracy but not RT. We found no evidence that the Hebb effect differed in magnitude between complex span and simple span. The effect was numerically smaller in complex span, but this trend must be seen in light of the fact that the effect of list repetition in complex span was smaller in Experiment 4 than in Experiments 1 and 2, which used the same complex-span task (see Table 2). The true effect is therefore probably somewhat larger than estimated in Experiment 4, and hence even closer to the effect

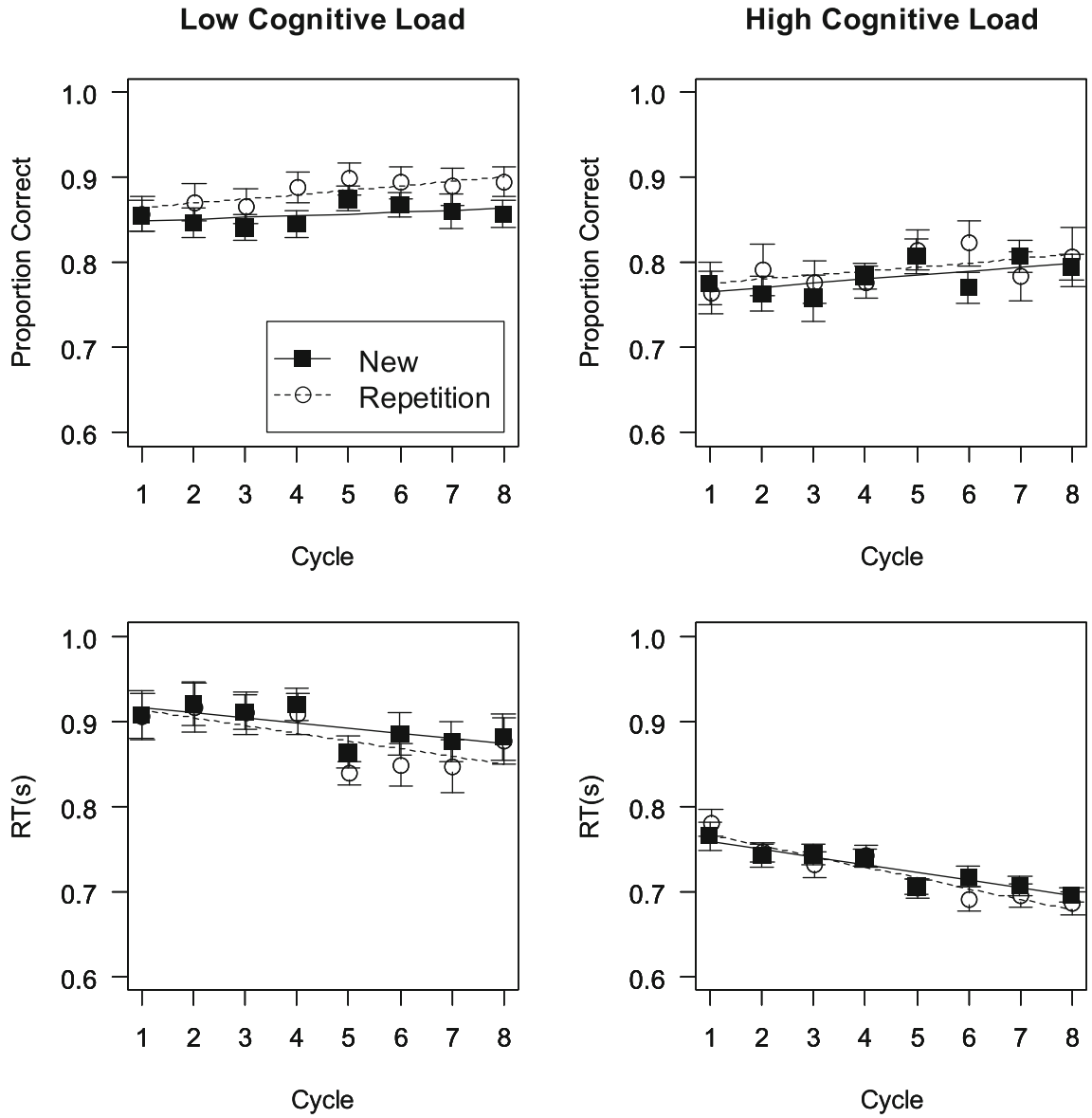

Fig. 6 Performance in the size-judgment task in Experiment 4. Error bars are 95 % confidence intervals for within-subject comparisons. The straight lines are regression lines from the mean posterior parameters of linear models with cycle as predictor, applied separately to each repetition condition

estimated for simple span. In addition, we obtained substantial evidence against the hypothesis that the Hebb effect is modulated by cognitive load.

When focusing on the low cognitive-load condition in isolation, the absence of a Cycle × Repetition interaction (BF = 0.2) might lead some readers to conclude that in this condition there was no Hebb effect. This conclusion would be premature for two reasons. First, this interaction was statistically supported in the joint analysis of all three conditions (BF = 5.6), and the evidence against the three-way interaction (BFs = 0.14 and 0.26, see Table 3) speaks against the possibility that the Cycle × Repetition interaction differed between conditions. Note that unlike conventional frequentist statistics, which can at best *fail* to find evidence *for* a three-way interaction, our Bayesian analysis provided evidence *against* its existence. Second, throughout all experiments and conditions of this article, the Hebb effect manifested itself most clearly through the main effect of repetition, whereas the evidence for the Cycle × Repetition interaction was much weaker, as reflected in the substantially smaller BFs. One reason for the relative weakness of the interaction, supported by the learning curves, is that the growth of accuracy over cycles in the repeated condition is not linear but rather decelerating, so that the Hebb effect is not well captured by the interaction of repetition with a linear trend over cycles. We argue that the main effect of repetition is sufficient evidence for the existence of a Hebb effect, because Hebb learning is the only conceivable explanation for why serial recall is better for repeated than for non-repeated lists: The repeated lists were chosen at random for each participant, so the only systematic difference between

repeated and non-repeated lists was the fact that the former occurred at trial numbers divisible by three, and that they were repeated.

One potential concern with the present comparison of the Hebb effect between simple and complex spans, and between two levels of cognitive load in complex span, could be that the three span types differed in trial duration. By implication, the time between successive repetitions of the repeated list was shortest in simple span, and longest in the low cognitive-load condition of complex span. These time differences were problematic if we had reasons to suspect that the memory traces that gradually build up for the repeated list are sensitive to the passage of time. This is not the case: When interference across lists is minimized, the Hebb effect is not affected by the number of non-repeated lists between two successive repetitions (for at least up to 12 intervening trials), and by implication, by the time between repetitions (Page, Cumming, Norris, McNeil, & Hitch, 2013).

General discussion

The present experiments provide an existence proof for the Hebb repetition effect in complex-span tasks. The effect is of approximately the same size as for a comparable simple-span task, and is not modulated by cognitive load. These findings weaken both the theoretical arguments for expecting a reduced or abolished Hebb effect in complex span, and the arguments for predicting an even larger effect than in simple span.

The interruption of list encoding (Experiments 1, 2, and 4) or of list recall (Experiment 3) by the distractor task did not hinder the formation of a long-term representation of the memory list. Despite being incidental, long-term learning of repeated lists appears to largely exclude the distractor-task materials. If the long-term representation of lists included representations of the distractor task, transfer across repeated lists would have been impaired because the distractors were unrelated across list repetitions. One possible explanation for the selective long-term learning of lists is that long-term learning applies only to information held in working memory for some time, not to material that is just briefly processed. Against that proposition stands the well-established finding that mere processing without intention to learn, especially semantic processing such as the size judgments used in our experiments, generates long-term memory traces (Craik & Lockhart, 1972). Moreover, even the content of working memory does not consist exclusively of memory items – distractors are involuntarily encoded into working memory (Oberauer, Farrell, et al., 2012), and removing them – by unbinding their representations from their context – takes time, so that working-memory representations are contaminated with distractor information, particularly at high cognitive load, when there is little opportunity for removing distractor representations (Oberauer, Lewandowsky, et al., 2012). Therefore, it is unlikely that long-term learning during complex-span tasks entirely excludes distractors.

We conclude that the Hebb effect in complex span relies on a memory trace that, despite its contamination with elements of the processing task, is strong enough to support the gradual improvement of list recall. Perhaps the memory traces of words from the size-judgment task were sufficiently distinct from the letter lists to keep interference at a minimum. Future studies could investigate the contribution of the processing task to the long-term memory trace by manipulating the similarity between memory and processing-task materials (e.g., using word lists combined with size-judgment tasks), and manipulating whether in the repeated condition the distractor-task materials also repeat across trials.

Our findings provide no support for the proposition that long-term memory contributes more strongly to complex span than simple span performance (Unsworth & Engle, 2007a). Whereas better long-term retention for complex span than simple span materials was observed in final free recall tests (McCabe, 2008), there was no indication in the present experiments of better long-term memory for repeated lists in complex than in simple-span tasks. A possible resolution of this discrepancy is that stronger long-term memory traces are created during complex span trials, but that these traces are not used in subsequent complex span trials using the same list. One conceivable reason why that might be the case is that the McCabe effect arises from elaborative encoding and rehearsal, which is known to improve delayed recall (Craik & Tulving, 1975). Due to their longer duration, complex-span trials provide more opportunity for elaboration than simple-span trials, and therefore could result in better memory for list words in final free recall, whereas immediate recall might not benefit much from elaboration (Rose & Craik, 2012). In contrast, the Hebb effect is unlikely to rely on elaboration because it has been observed with lists of digits and, in the present case, letters, which are less amenable to elaboration than words, the material used to demonstrate the McCabe effect.

Finally, we observed no modulation of the Hebb effect in complex span by the opportunity for rehearsing or refreshing the memory items, as afforded by cognitive load. With low cognitive load, there is more opportunity for refreshing than with high cognitive load, and yet the magnitude of the Hebb effect was indistinguishable for the two loads in Experiment 4. This finding strengthens our conclusion that the Hebb effect does not arise from rehearsal (elaborative or otherwise), and that it does not arise from refreshing or covert retrieval. It is possible – and we think plausible – that the McCabe effect arises from one of these processes. If this is the case, it implies that there are two forms of long-term learning during complex-span tasks, one driven by elaborative rehearsal or covert retrieval and underlying the McCabe effect, and the

other not relying on any of those processes, underlying the Hebb effect.

We propose that the McCabe effect reflects episodic long-term memory for items, whereas the Hebb effect reflects the acquisition of semantic long-term memory for sequences. More specifically, the Hebb effect reflects a form of long-term learning that results in unified representations of lists, or sub-sequences in lists, akin to the learning of words by forming unified representations of sequences of phonemes or letters (Page & Norris, 2009). Converging evidence supports the close link between the Hebb effect and word learning: The size of the Hebb effect correlates with individual differences in the ability to learn new word forms (Mosse & Jarrold, 2008; Szmalec, Loncke, Page, & Duyck, 2011). Transfer experiments showed that syllable sequences learned in a Hebb repetition procedure interfered with access to similar words in a lexical-decision task, showing that the sequences acquired during Hebb learning were integrated into the mental lexicon (Szmalec, Duyck, Vandierendonck, Mata, & Page, 2009; Szmalec, Page, & Duyck, 2012).

To conclude, we demonstrated that the Hebb effect is observed with equal strength in complex- and simple-span tasks, regardless of cognitive load. This is remarkable because there are good reasons to expect that interleaving list items with distractor episodes disrupts the formation of an integrated chunk of the list in long-term memory. The Hebb effect reported here is the first direct evidence that long-term learning of memory lists improves immediate recall in a complex-span test, and as such opens a window into investigating the role of long-term memory in tests of working memory.

Acknowledgments The research reported in this article was supported by a grant from the URPP "Dynamics of Healthy Aging" at the University of Zurich to the first author, by a World University Network grant to the first and the second authors, and a Discovery Grant and a Discovery Outstanding Researcher Award from the Australian Research Council to the third author.

References

Bakeman, R., & McArthur, D. (1996). Picturing repeated measures: Comments on Loftus, Morrison, and others. *Behavior Research Methods, Instruments, & Computers, 28,* 584–589.

Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 570–585.

Burgess, N., & Hitch, G. J. (2005). Computational models of working memory: Putting long-term memory into context. *Trends in Cognitive Sciences, 9,* 535–541.

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language, 55,* 627–652.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12,* 769–786.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences, 7,* 547–552.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684.

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104,* 268–294.

Cumming, G., & Finch, S. (2005). Inference by eye: Confidence intervals and how to read pictures of data. *American Psychologist, 60,* 170–180.

Cumming, N., Page, M., & Norris, D. (2003). Testing a positional model of the Hebb effect. *Memory, 11,* 43–63.

Cunningham, T. F., Healy, A. F., & Williams, D. M. (1984). Effects of repetition on short-term retention of order information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 575–597.

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

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.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental Psychology, 40,* 177–190.

Hebb, D. O. (1961). Distinctive features of learning in the higher animal. In J. F. Delafresnaye (Ed.), *Brain mechanisms and learning* (pp. 37–46). Oxford: Blackwell.

Hitch, G. J., Fastame, M. C., & Flude, B. (2005). How is the serial order of a verbal sequence coded? Some comparisons between models. *Memory, 13,* 247–258.

Hyde, T. S., & Jenkins, J. J. (1969). Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology, 82,* 472–481.

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.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90,* 773–795.

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.

Loaiza, V. M., & McCabe, D. P. (2012). Temporal-contextual processing in working memory: Evidence from delayed cued recall and delayed free recall tests. *Memory & Cognition, 40,* 191–203.

Loaiza, V. M., McCabe, D. P., Youngblood, J. L., Rose, N. S., & Myerson, J. (2011). The influence of levels of processing on recall from working memory and delayed recall tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1258–1263.

McCabe, D. P. (2008). The role of covert retrieval in working memory span tasks: Evidence from delayed recall tests. *Journal of Memory and Language, 58,* 480–494.

McCabe, D. P. (2010). The influence of complex working memory span task administration methods on prediction of higher level cognition and metacognitive control of response times. *Memory & Cognition, 38,* 868–882.

McKelvie, S. J. (1987). Learning and awareness in the Hebb digits task. *Journal of General Psychology, 114,* 75–88.

Morey, R. D., & Rouder, J. N. (2012). BayesFactor (Version 0.9.0). Retrieved from http://cran.at.r-project.org/web/packages/BayesFactor/index.html

Mosse, E. K., & Jarrold, C. (2008). Hebb learning, verbal short-term memory, and the acquisition of phonological forms in children. *Quarterly Journal of Experimental Psychology, 61,* 505–514.

Oberauer, K., Farrell, S., Jarrold, C., Pasiecznik, K., & Greaves, M. (2012a). Interference between maintenance and processing in working memory: The effect of item-distractor similarity in complex span. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 665–685. doi:10.1037/a0026337

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012b). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review, 19,* 779–819.

Oberauer, K., & Meyer, N. (2009). The contributions of encoding, retention, and recall to the Hebb effect. *Memory, 17*, 774–781.

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.

Page, M. P. A., Cumming, N., Norris, D., McNeil, A. M., & Hitch, G. J. (2013). Repetition-spacing and item-overlap effects in the Hebb repetition task. *Journal of Memory and Language, 69,* 506–526. doi:10.1016/j.jml.2013.07.001

Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society of London, 364,* 3737–3753.

R Development Core Team (2012). R: A language and environment for statistical computing (Version 2.15.2). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from URL: http://www.R-project.org

Raye, C. L., Johnson, M. K., Mitchell, K. J., Greene, E. J., & Johnson, M. R. (2007). Refreshing: A minimal executive function. *Cortex, 43,* 135–145.

Rose, N. S., & Craik, F. I. M. (2012). A processing approach to the working memory/long-term memory distinction: Evidence from the levels-of-processing span task. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38,* 1019–1029.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology, 56,* 356–374.

Sechler, E. S., & Watkins, M. J. (1991). Learning to reproduce a list and memory for the learned. *American Journal of Psychology, 104,* 357–394.

Szmalec, A., Duyck, W., Vandierendonck, A., Mata, A. B., & Page, M. P. A. (2009). The Hebb repetition effect as a laboratory analogue of novel word learning. *Quarterly Journal of Experimental Psychology, 62,* 435–443.

Szmalec, A., Loncke, A., Page, M. P. A., & Duyck, W. (2011). Order or disorder? Impaired Hebb learning in dyslexia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1270–1279. doi:10.1037/a0023820

Szmalec, A., Page, M. P. A., & Duyck, W. (2012). The development of long-term lexical representations through Hebb repetition learning. *Journal of Memory and Language, 67,* 342–354. doi:10.1016/j.jml.2012.07.001

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

Unsworth, N., & Engle, R. E. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language, 54,* 68–80. doi:10.1016/j.jml.2005.06.003

Unsworth, N., & Engle, R. E. (2007a). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114,* 104–132.

Unsworth, N., & Engle, R. E. (2007b). On the division of short-term and working memory: An examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin, 133*, 1038–1066.

Wagenmakers, E.-J., & Morey, R. D. (2013). *Simple relation between one-sided and two-sided Bayesian point-null hypothesis tests*. Unpublished Manuscript, University of Amsterdam.

Wilhelm, O., Hildebrandt, A., & Oberauer, K. (2013). What is working memory capacity, and how can we measure it? *Frontiers in Psychology, 4*. Retrieved from doi:10.3389/fpsyg.2013.00433