Article:

Mızrak, E. and Oberauer, K. (2022) Working memory recruits long-term memory when it is beneficial: Evidence from the Hebb effect. Journal of Experimental Psychology: General, 151 (4). pp. 763-780. ISSN 0096-3445

https://doi.org/10.1037/xge0000934

Working memory recruits long-term memory when it is beneficial: Evidence from the Hebb Effect

Eda Mızrak and Klaus Oberauer

University of Zurich

Author Note: Eda Mızrak and Klaus Oberauer, Department of Psychology, University of Zurich, Switzerland. This research was supported by a grant from the Swiss National Science Foundation to K. Oberauer (project 100014_179002). Correspondence should be addressed to Eda Mızrak, Department of Psychology, Cognitive Psychology Unit, University of Zürich, Binzmühlestrasse 14/22, 8050 Zurich, Switzerland. E-mail: eda.mizrak@psychologie.uzh.ch

Abstract

When encoding task-relevant information in working memory (WM), we can use prior knowledge to facilitate task performance. For instance, when memorizing a phone number, we can benefit from recognizing some parts as known chunks (e.g., 911) and focus on memorizing the novel parts. Prior knowledge from long-term memory (LTM), however, can also proactively interfere with WM contents. Here, we show that WM selectively recruits information from LTM only when it is helpful, not when it would interfere. We used variants of the Hebb paradigm in which WM is tested through immediate serial recall of lists. Some lists were repeated frequently across trials, so they were acquired in LTM, as reflected in increasing serial-recall performance across repetitions. We compared interference conditions in which that LTM knowledge could interfere with holding another list in WM to a neutral condition in which that knowledge could be neither beneficial nor harmful. In Experiments 1-3, lists in the interference conditions shared their items with the learned lists but not their order. We observed no proactive interference. In Experiments 4 and 5, the interference lists' first three items overlapped exactly with the learned lists, and only the remaining items had a new order. This made LTM knowledge partially beneficial and partially harmful. Participants could use LTM flexibly to improve performance for the first part of the list without suffering interference on the second half. LTM-mediated learning of the first part even boosted memory for the unknown second part. We conclude that there is a flexible gate controlling the flow of information from LTM and WM so that LTM knowledge is recruited only when helpful.

Keywords: long-term memory; working memory; proactive interference; proactive facilitation; Hebb effect

Introduction

When learning new information, we often rely on pre-existing knowledge that can facilitate learning. For instance, when memorizing a new phone number, it can help to recognize chunks that are familiar to us (e.g., 911, or one's birth year). By using prior knowledge that resides in our long-term memory (LTM), we can reduce the load on our limited-capacity working memory (WM).

WM maintains goal-relevant information easily accessible for as long as it is needed. It serves as a mental blackboard where we can keep the information most relevant to the goal at hand and build on this information. WM is capacity limited such that only a small amount of information can be held available in WM in an accessible state (see Oberauer et al., 2016 for a review). LTM, by contrast, maintains our life-long memories in a rather stable but less accessible state compared to WM (see Cowan, 2008 for a review of differences between working memory and long-term memory). LTM is often thought to be unlimited in storage capacity. Here we ask: What happens when goal-relevant information consists of both novel elements that need to be encoded and maintained in WM, and known elements that are already stored in LTM? How do WM and LTM interact with each other?

The interplay between WM and LTM has been a matter of debate in the last twenty years. Theories that describe the architecture of WM diverge in their views of how separated (Baddeley, 2000) or interconnected (Cowan, 2012; Oberauer, 2009) WM and LTM are. However, all theorists agree that there is information exchange between the two. Here, we will focus on one side of this exchange; information flow from LTM to WM. Specifically, we will test the *flexible gate hypothesis* (Oberauer, 2009), which states that only LTM information beneficial for the function of WM is allowed into WM, whereas LTM information that is possibly harmful to the current task is blocked.

As in the phone number example, we should be able to use our LTM when it has goal-relevant information, and this should help WM. For instance, several studies found that when asked to recall a list of items immediately after presentation, participants were better at recalling the ones that consisted of well-known chunks such as “PDF”, or a learned pair of words, compared to the ones that did not (Cowan et al., 2004; Portrat et al., 2016; Thalmann et al., 2019). Two recent studies (Norris et al., 2020; Thalmann et al., 2019) showed that not only memory for chunked items but also memory for the items following a chunk were better remembered, compared to lists that did not include any chunks. This suggests that WM can draw on helpful knowledge in LTM, and this can free up capacity in WM for information not directly supported through LTM, improving overall performance.

However, information flow between LTM and WM is not necessarily helpful. For example, if the information in LTM is similar to but nevertheless different from the goal-relevant information, one might erroneously rely on the LTM information (e.g., remembering the chunk 911 when the new number contains the digits 912). If information in LTM related to the current contents of WM is always automatically retrieved into WM, it would often interfere with the WM task at hand. Therefore, to avoid proactive interference from LTM, it would be useful for the cognitive system to flexibly control the flow of information from LTM to WM. Ideally, WM should rely on LTM when LTM knowledge is beneficial, and not when it potentially interferes. According to the flexible-gate hypothesis, this is how the system works. The key prediction flowing from this hypothesis is that LTM information benefits WM performance when it matches the information that needs to be held in WM (i.e., creates proactive facilitation) but does not harm it when it mismatches (i.e., it does not create proactive interference).

Recently, Oberauer et al. (2017) tested this prediction. Participants first learned arbitrary colors of 120 unique objects. For instance, in the learning phase a sofa was learned as being colored in a shade of pink whereas a car was learned as being a shade of green. After the learning phase, participants were given a visual WM test in which three colored objects had to be briefly remembered. One of them was a learned object presented in the same color (i.e., the sofa in the same pink), another a learned object in a different color (i.e., the car in a shade of orange), and the third a new object (i.e., a lamp in green). After a brief delay they were cued with one of the objects, presented in grey, and asked to recall its color in the current trial by choosing it from a color wheel. Memory performance for the objects whose color matched with previous learning (here: the pink sofa) was facilitated in comparison to new objects, suggesting a contribution of LTM. Memory performance for the objects whose color mismatched previous learning (here: the orange car) was not worse than performance for new objects. This finding showed that the learned object did not act as an automatic retrieval cue that brings its learned color back into WM – if that were the case, it would interfere with WM for learned objects' new colors. Nevertheless, participants showed an above-chance tendency to report the LTM color instead of the WM color for these objects (e.g., reporting the car as green), yet this behavior did not decrease overall performance for this condition. These responses might be due to LTM information being used as a last resort in the absence of WM information, rather than an obligatory influx of information from LTM: The LTM information is used only when there is nothing better at hand, so there is nothing to lose from using it.

The study by Oberauer et al. (2017) provided evidence for proactive facilitation of WM by LTM combined with evidence against proactive interference. The results were as predicted from the flexible gate hypothesis. In the present series of experiments, we intend to test the flexible-gate hypothesis further.

The Present Study

The aim of this study is two-fold: (1) We aim to generalize the finding of Oberauer et al. (2017) using verbal material in different memory tasks and test whether WM recruits LTM when it is beneficial whereas there is no proactive interference from LTM to WM. (2) We want to test whether the contribution of LTM to WM is limited to instances in which LTM is fully beneficial, or whether it is possible to benefit from partial LTM information and discard the harmful parts from the same memory representation. To this end we needed to create conditions in which LTM knowledge relevant to the current demand for WM is completely helpful, completely harmful, or contains a mixture of both helpful and harmful information.

We used a variant of the Hebb paradigm that is commonly used to examine long-term learning of sequences in the context of immediate serial recall (Hebb, 1961). In the Hebb paradigm, participants study a list of items on each trial and are asked to recall them in serial order immediately after presentation. On a subset of trials (typically every third or fourth trial), the study list is repeated whereas the rest of the trials use new, randomly generated lists. Serial-recall performance gradually improves for the repeated lists (Hebb lists) compared to the randomized lists (i.e. Filler lists). This effect was replicated in several studies that used this paradigm (Cumming et al., 2003; Cunningham et al., 1984; Fendrich et al., 1991; Oberauer et al., 2015; Oberauer & Meyer, 2009; Page & Norris, 2009). Researchers hypothesized that over the course of repetitions, the memory for the sequence of items in the repeated lists becomes more and more robustly established in LTM. Accordingly, LTM representations of the lists contribute to immediate serial recall for the repeated lists (Burgess & Hitch, 2005, 2006; Page & Norris, 2009).

The Hebb paradigm is suitable for our aims because it already shows that participants benefit from their LTM for the Hebb lists. Modifications were required to test whether the knowledge about the repeated list might create proactive interference with immediate serial recall of other lists. Proactive interference could arise if information is retrieved from LTM that mismatches the information to be maintained in WM. Retrieval of such information is most likely if information involved in the WM task is related to the mismatching information in LTM, so that the former could act as a retrieval cue for the latter. This was the case in the study of Oberauer et al. (2017): An object previously associated with one color was presented with a different color in the WM task. At test, the object was given as retrieval cue for reproduction of its color in the WM task, and therefore could have acted as cue to the previously learned color as well. Here we aim to create similar situations for the Hebb paradigm.

In the Hebb paradigm we know that information from LTM is retrieved when it facilitates performance on the Hebb lists, but we don't know what retrieval cues are responsible for this retrieval. We consider three – not mutually exclusive – possibilities. The first, tested in Experiments 1 and 2, is that the list items of the repeated list individually act as retrieval cues to the LTM trace of that list. If that is the case, then the Filler lists in the standard Hebb paradigm – which consist of the same items as the Hebb lists, but in a shuffled order – would act as retrieval cues to the Hebb list as well, and recalling the Filler lists could be impaired by proactive interference. Therefore, part of the Hebb effect could reflect proactive interference on the Filler trials. In Experiments 1 and 2 we added a third condition – *No-overlap Filler* lists composed of items from a separate set – that served as control condition for proactive facilitation on the Hebb trials, and proactive interference on the traditional Filler trials, which we now refer to as *Overlap Fillers* to mark the fact that they overlap with the repeated Hebb lists (Figure 1). If presentation of a list automatically brings to mind LTM knowledge of learned lists of the same items, then Overlap Filler trials should suffer more proactive interference than No-overlap Filler trials. This is because both kinds of fillers are preceded by an equal number of lists using the same item set, but in the case of Overlap Fillers, there is one such list learned strongly in LTM (i.e., the Hebb list), whereas in the case of No-overlap Fillers, there are only weak LTM traces of never-repeated lists.

The second possibility – tested in Experiment 3 – builds on models of serial recall in which each list item is bound to a representation of its ordinal list position, and at recall, the successive list positions serve as retrieval cues to the items bound to them (Botvinick & Watanabe, 2007; Brown et al., 2000; Lewandowsky & Farrell, 2008). Based on this idea, we considered the possibility that on Hebb trials the position codes acted as retrieval cues not only to the items bound to them in WM but also to the items associated to them in LTM through repeated exposure to the same list. Although long-term learning of individual item-position associations has long been looked at with skepticism based on one study (Cumming et al., 2003), two recent studies have demonstrated long-term learning of item-position associations, raising the possibility that they contribute to the Hebb effect (Majerus & Oberauer, 2019; Nakayama & Saito, 2017)[1]. If that is so, then the position markers used as retrieval cues on Overlap Filler trials could also elicit proactive interference. It is difficult to

[1] It should be noted that in the study of Nakayama and Saito, item-position learning was slower compared to the Hebb learning.

create a control condition that involves serial recall but avoids using position cues, so in Experiment 3 we changed the task to one in which participants had to recall the list position, given an item as cue. This test uses items as retrieval cues and the positions as targets, and therefore, relies on item-position bindings, though using them in the reverse cue-target direction compared to serial recall. The control condition then involved No-overlap Filler lists composed of items from a separate set, so that the items used as retrieval cues in this condition were not associated strongly to a particular list position in LTM (see Figure 1 for an illustration of the conditions and the test procedure).

The third possibility emerges from two computational models of the Hebb effect (Burgess & Hitch, 2006; Page & Norris, 2009). Although differing in many regards, these models share the assumption that Hebb lists are learned in LTM as unified chunks. This implies that their retrieval is not cued by re-experiencing the individual items of the Hebb list in any order, and it is also not cued by re-engaging the same position codes. Rather, retrieval of a chunked list representation relies on a match between the currently presented list and the chunked list in LTM, and the computation of that match is sensitive to the order of items. For instance, a learned Hebb list ABCD has little or no match with a shuffled list CADB.

The two models differ in the details of how a list chunk is cued and retrieved. In the model of Burgess and Hitch (2006), the match computation is incremental: When a new list is presented, it is used continuously as retrieval cue to re-activate in LTM all lists that match the new list building up in WM. With every additional item of the new list that retrieval cue becomes more specific, narrowing down the set of activated LTM representations. As soon as the mismatch between the current list in WM and a chunk in LTM exceeds a threshold (which is a parameter in the model), that chunk is rejected. In this model, the beginning of the new list is the most important retrieval cue to LTM traces of matching lists. Support for this assumption comes from experiments in which only parts of a list were repeated across a subset of trials. This led to improved serial recall only if the beginning of the list was repeated, not when the end or some middle section was repeated (Hitch et al., 2005).

In the model of Page and Norris (2009), chunks compete with representations of individual list items for being selected for maintenance and subsequent recall. Chunk representations have a chance to win that competition only after the list presentation is finished, and the activations of individual-item representations have subsided. For the chunk unit to be sufficiently strongly activated at that point, the items of the learned list must have been presented at least approximately in the correct order – how good that approximation needs to be depends on the model parameters. To conclude, both models imply that retrieval of the LTM representation of a Hebb list requires an order-sensitive match between the new list and at least a substantial part of the Hebb list.

To test this idea, Experiments 4 and 5 consisted of a learning and a transfer phase. In the learning phase the Hebb lists were fully repeated, and there were no overlapping lists. In the transfer phase we introduced the Overlap Filler condition with lists that matched the Hebb lists exactly in the first three positions, whereas the last four items were presented in a new randomized order. We hypothesized that the first three list items of the Overlap Filler trials could act as retrieval cue to the matching Hebb list. We allocated the partial match to the beginning of the list because partial matches at the beginning have been shown to work best for eliciting a Hebb effect (Hitch et al., 2005).

If the partial match of the Overlap Filler list with the Hebb list is sufficient to trigger retrieval of the chunk of the Hebb list, participants could utilize their LTM of the first three items for recalling the Overlap Filler lists, and accuracy for these items should be at similar levels with Hebb lists. If this is the case, we can conclude that the LTM representation of the Hebb list was retrieved. In the models of Burgess and Hitch (2006) and of Page and Norris (2009), a chunk is retrieved and used in an all-or-none fashion. Therefore, if it is used to boost recall of the first three items of an Overlap Filler list, it should also be used for informing recall of the remaining items – and therefore, cause proactive interference. We will test for proactive interference from LTM by comparing the performance of the last four items of Overlap Filler lists to the control lists. If there is proactive interference from the LTM representation of the Hebb list, performance for these items should be lower than control lists. In contrast, if the gate is flexible enough to make partial use of the LTM trace of the Hebb lists to assist recall of the first part of the Overlap Filler list without suffering proactive interference on the second part, then performance for Overlap Filler and No-overlap Filler lists for the final four items should be similar.

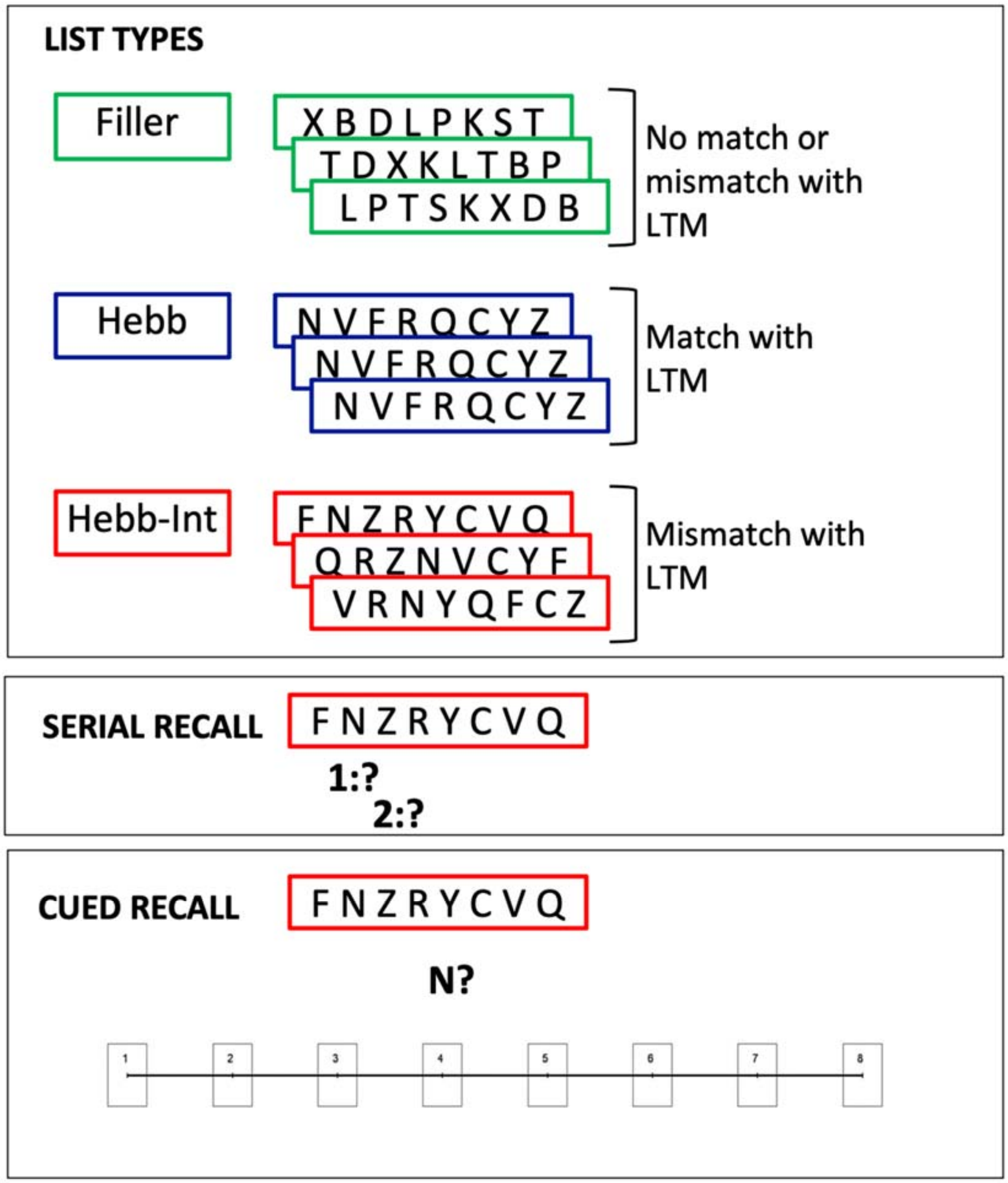

Figure 1. LIST TYPES. Illustration of encoding phase in Experiments 1,2, and 3 for consonants. Trial types and the design was the same for words. Study list items were presented sequentially and centrally one at a time. No-overlap Filler list items are presented in a random order at each trial. Hebb list items are presented always in the same order, and Overlap Filler list items are presented in a random order at each trial. Overlap Filler list contents and Hebb list trials consist of the same items. SERIAL RECALL. Example of encoding and testing phase for an Overlap Filler trial in Experiments 1 and 2. Participants were instructed to type the items they studied in their order of presentation. They were given the serial position number as a reminder, and were asked to type the corresponding item from the current trial. If participants typed in the Hebb list item for position 2 (e.g., “V”) instead of the current trial’s list (e.g., “N”), this would be an instance of Hebb intrusions. CUED RECALL. Example of encoding and testing phase for Overlap Filler trial in Experiment 2.

Participants were cued with the item and were asked to select the position of this item in the current trial from the response scale (for consonants: 1-8, for words: 1-7). If participants selected the Hebb list position for this item (e.g., "1") instead of the current trial's position (e.g., "2"), this would be an instance of interference.

Experiments 1 – 3

General Method

Participants

Twenty-five (6 males), twenty-six (7 males), and twenty-five (9 males) young adults (students from the University of Zurich) participated in Experiments 1,2, and 3, respectively. Data from two participants from Experiment 1, and from one participant of Experiment 3 were excluded due to recording of incomplete data leaving twenty-three participants for Experiment 1 and twenty-four participants for Experiment 3. Experiments lasted 60-75 minutes. Participants were reimbursed with a course credit or 15 Swiss Francs per hour for their time.

Numbers of participants were pre-determined for all experiments based on the previous studies from our lab using similar within-subjects experimental designs. As we analyzed the data with Bayesian methods, we planned to increase the sample size in case of ambiguous results, though we saw no need to do so for the present experiments. .

Materials

Experiments consisted of word blocks and consonant blocks. In word blocks, stimuli were two sets of seven monosyllabic German nouns (drawn from the dlexdb.de lexical database). For each participant, 14 words were randomly chosen from a pool of monosyllabic nouns consisting of 98 words. These 14 words were then randomly assigned to the two sets. We made sure that the lists did not consist of words which were high in semantic relatedness. Stimuli for consonants were two non-overlapping sets of eight consonants randomly drawn from the 21 consonants of the German alphabet. Hebb lists were repeated exactly across trials. Overlap Filler lists consisted of the same set of stimuli as the Hebb lists but shuffled in a new random order for each trial. No-overlap Filler lists consisted of the other stimulus set, shuffled in a new random order for each trial.

Procedure

The experiment consisted of two parts, with 64 trials each. The first part tested lists of consonants, and the second part tested lists of words. Consonant and word trials were analyzed separately. There was one Hebb list for consonants (e.g., N, V, F, R, Q, C, Y, Z) and one Hebb list for words (e.g., door, side, pan, cat, wheel, dust, fan) for each participant. Each part of the experiment consisted of 16 mini-blocks of four trials. A mini-block had two No-overlap Filler lists, one Hebb list, and one Overlap Filler list.

In Experiment 1, the order of trials in a mini-block was always the same: 1) No-overlap Filler list, 2) Hebb list, 3) No-overlap Filler list, 4) Overlap Filler list. One disadvantage of such a fixed order of conditions is that participants could learn when each list type will be presented, and adapt strategies such as using LTM on the second trial of each mini-block and not using it on the fourth trial of each mini-block. To prevent this possibility, in Experiments 2 and 3 the order of the lists within a mini-block was randomized, with one constraint: The last trial of each mini-block must not be the same list type as the first trial of the following mini-block. If one mini-block ended with a Hebb trial, the following mini-block started with either a No-overlap Filler or an Overlap Filler trial.

Each trial began with a central fixation point presented for 1000 ms, followed by the study list presentation. Consonant lists consisted of eight items, whereas word lists had seven items. Each list item was presented for 500 ms for consonants and 1000 ms for words, one after each other, centrally. Following a 1000 ms delay, participants started the recall task.

Immediate Serial Recall of Items. In Experiments 1 and 2 we tested participants with an immediate serial recall test. Participants were asked to recall the list items in the order of their presentation. To remind them of where they are in the list, they were presented with the position of the next item in the list, starting from 1, to the left of an empty box into which they were asked to type the next item. After that, they pressed the return key and continued with the recall of the next item.

Immediate Cued Recall of Positions. In Experiment 3, we used cued recall with the items as the cue and the list positions as the recall targets. Presentation procedure and list manipulations were exactly the same as Experiments 1 and 2. At test, participants were presented with an item from the study list in the middle of the screen. Below the item, there was a scale with 1 to 8 options (for consonant lists) or 1 to 7 options (for word lists). They were asked to select the position in which they had encountered this item in the study list. At each trial, they were tested with all the study list items one after another in random order. Positions already selected could be selected again for another item. Figure 1 depicts examples of a trial from this test in the bottom panel.

Data Analysis

Analysis of Immediate Recall Performance. In order to test our hypotheses, we estimated a Bayesian generalized linear mixed model (BGLMM) that was implemented in the R package rstanarm (Stan Development Team, 2018).

All experiments provided a binary dependent variable for accuracy. For Experiments 1 and 2 (as well as Experiments 4 and 5 below), a study item recalled in the correct serial position was coded as correct, and a study item recalled in an incorrect position, or a response that did not match any of the study list items, was coded as incorrect. Experiment 3 was an item-cued recall task, in which a response was correct if the participant chose the correct position for the give item, and incorrect if they chose another position. Consequently, our model employed a binomial distribution as conditional distribution with predictions based on a linear model with a probit link function (i.e., a repeated measures probit regression).

The fixed effects were list type (Hebb vs. No-overlap Filler vs. Overlap Filler) and stimulus type (words vs. consonants), and their interaction. We used the maximal random-effects structure justified by the design suggested by Barr et al. (2013) that included by-participant random intercepts and by-participant random slopes for all fixed effects (as all factors were within-subject factors). We also estimated the correlations between the by-participant random-effects parameters.

We evaluated the findings from the BGLMM by examining the posterior probability distributions of the estimated marginal means of the fixed effects on the probability scale. In line with the Bayesian framework, for statistical inference, we will inspect the 95% credibility intervals based on the .025 and .975 quantiles from the posterior distribution of the parameters. The true effect is inferred to lie within that interval with a posterior probability of .95.

For factor coding of both fixed and random effects, we used the orthonormal contrasts suggested by Rouder et al. (2012) that ensure an equal effect of the priors on all factor levels. Following Gelman et al. (2014) we used weakly informative priors for all parameters to facilitate convergence of the model. The priors for the fixed effects were weakly informative t-distributions with 4 degrees of freedom and scale 4. The variance-covariance matrix was split into priors for the variances with Cauchy (0, 4) and for the correlation matrix with LKJ (1) using the method of Lewandowski, Kurowicka, and Joe (2009).

The model was estimated using a version of Hamiltonian Monte-Carlo in Stan (Carpenter et al., 2017). We estimated a total of 2000 samples in 4 independent chains with different random start values. The first 1000 samples were discarded as warmup samples. Model convergence was assessed based on $\hat{R}$ (all $\hat{R} < 1.01$) and visual assessments of chain convergence.

Our aim was to test whether there was proactive facilitation and/or proactive interference from LTM to WM. Proactive facilitation would be signaled by higher accuracy in the Hebb condition compared to the No-overlap Filler condition. Proactive interference would be shown by poorer performance in the Overlap Filler condition compared to the No-overlap Filler condition. We expect to also replicate the standard Hebb effect – better memory in the Hebb than the Overlap Filler condition – but depending on how our new No-overlap Filler condition compares to those two conditions, the Hebb effect could arise from proactive facilitation, proactive interference, or a combination of both.

Analysis of responses given to Overlap Filler list. We wanted to test whether participants solely relied on their WM, which consists of the current trial list, or also draw on information from LTM (i.e., the Hebb list), even if it is not relevant to the current trial in the Overlap Filler condition. To understand whether there was any contribution from the LTM representation of the Hebb list during recall of Overlap Filler lists, we compared the output position of recalled items (i.e., the list position in the recalled sequence) to two independent input positions, the input position of the recalled item in the trial's study list, and its input position in the Hebb list. For example, consider the first Overlap Filler list trial shown in

Figure 1 (F – N – Z – …) and imagine a participant correctly recalls N as the second letter. In this case, the output position of N is 2, the input position of N from the trial's study is 2 as well, and the input position from the Hebb list (see Figure 1) is 1. If the LTM trace of the Hebb list has an influence on recall of Overlap Filler lists, we should see that items are recalled closer to their input positions in the Hebb list than expected by chance.

For analysis, we calculated two distance measures between the output and the two input positions: *WM distance* and *LTM distance*. In each case, we subtracted the input position from the output position (i.e., distance = output – input). WM distance is the difference between output and input position from the trial's study list. Continuing the example above, WM distance for N is 0 (2 – 2). LTM distance is the difference between output position and input position from the Hebb list. Continuing from the example above, LTM distance for N is 1 (2 – 1).

The distance scale for words (shown in the lower row of Figure 3) ranges from -6 to 6, the largest distances in a 7-item list (7-1 or 1-7). However, the probability of obtaining each of the possible values of the distance scale is not uniform. There is only one possible combination of input and output positions resulting in a distance of -6 or 6, respectively. In contrast, there are 7 possible combinations that would result in a distance of 0 (i.e., all items recalled in the correct positions). Calculating the number of possible combinations for each value on the distance scale shows that this number has its maximum at 0, linearly decreases to each side, and has a minimum at -6 and 6. Thus, if responses were random (i.e., not based on study list input position) we would expect a triangular probability distribution of the distances[2]. In contrast, if participants use the study list position to guide their recall, they have a higher likelihood of recalling an item in the correct output position (i.e., distance = 0) which would substantially increase the probability of distance 0.

Additionally, the most common errors during serial recall are confusions with close neighbors (i.e., predominantly distances of -1 or +1, and less often -2 or + 2, etc.). Thus, if participants use the study list position to guide their recall, we would expect to see more shorter and fewer long distances than expected by chance. Consequently, if participants use the study list position to guide their recall, we expected that the distance distribution is centered at 0 and steeply decreases towards -6 and 6, so that it is more peaked and has longer tails than the triangular distribution. We model this using a double exponential distribution, which is also known as the Laplace distribution (see Figure 3 for an illustration of the shapes of the two distributions).

For consonants, the same logic and predicted probability distributions follows. The only difference is that there are 8 input and output positions, so that the distance scale ranges from -7 to 7.

We fitted both a triangular distribution and a Laplace distribution to the distributions of WM and LTM distances shown in Figure 3, separately for words and consonants. Each of the distributions only had one free parameter and was centered at 0. The triangular distribution was constrained to be symmetric with the free parameter determining both

[2] To visualize random responses, we generated data by randomly sampling integers from 1 to the number of list items (e.g., for words, a random sample from 1 to 7: 2-5-4-7-1-3-6), and calculated distances of these responses from input positions 1 to 7. We plotted the distance distribution based on this data and the shape of the distribution was triangular (See Supplemental Material Figure S1). The script for generating the data and plotting the distance distribution is provided at the OSF files.

minimum and maximum (i.e., minimum: the *x*-axis value where $y = 0$). For the Laplace distribution, the free parameter determined the steepness of the distribution. We used maximum-likelihood estimation to fit the distributions to the vector of distances across participants. Thus, a better fit of a distribution is indicated by a larger maximum log-likelihood value.

Results from Experiments 1-3

Recall Performance

Model convergences for all experiments were inspected, and all $\hat{R}$ were below 1.01. Figure 2 presents performance across list types. Immediate recall was the best for Hebb lists, and was at comparable levels for Overlap Filler and No-overlap Filler lists for both words and consonants. Table 1 presents the evidence for the comparisons between lists that were our main interests. The pairwise comparison of Hebb and No-overlap Filler lists provided evidence for proactive facilitation (i.e., Hebb > No-overlap Filler), whereas performance of Overlap Fillers and No-overlap Fillers was similar, providing evidence for the absence of proactive interference.

These findings indicate that participants learned the Hebb lists and were able to use what they learned (LTM) to guide their performance. As explained above, if LTM interfered with WM, we would expect lower performance for Overlap Filler lists compared to the No-overlap Filler lists. However, this was not the case in any of the Experiments 1 to 3.

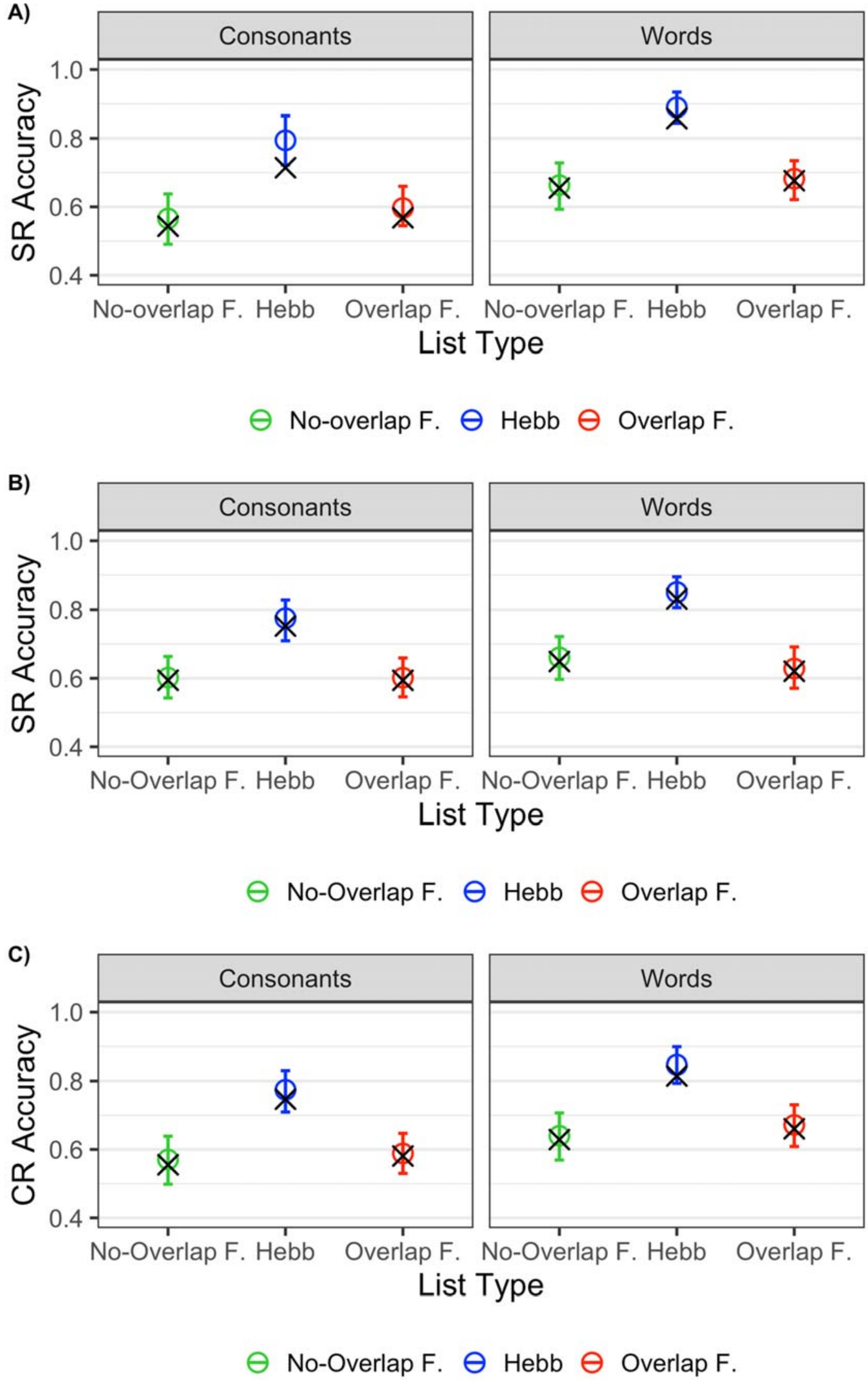
A)
Consonants
Words
SR Accuracy
1.0
0.8
0.6
0.4
No-overlap F.
Hebb
Overlap F.
List Type
No-overlap F.
Hebb
Overlap F.
B)
Consonants
Words
SR Accuracy
1.0
0.8
0.6
0.4
No-Overlap F.
Hebb
Overlap F.
List Type
No-Overlap F.
Hebb
Overlap F.
C)
Consonants
Words
CR Accuracy
1.0
0.8
0.6
0.4
No-Overlap F.
Hebb
Overlap F.
List Type
No-Overlap F.
Hebb
Overlap F.

Figure 2. Proportion correct of the immediate serial recall of items (row A and B) and immediate cued recall of positions (row C) of all lists of Experiment 1 (row A), 2 (row B), and 3 (row C). Colored circles and error bars represent estimated proportions and 95% CIs, respectively. X marks represent observed proportions. Overlap indicates the model adequately describes the data.

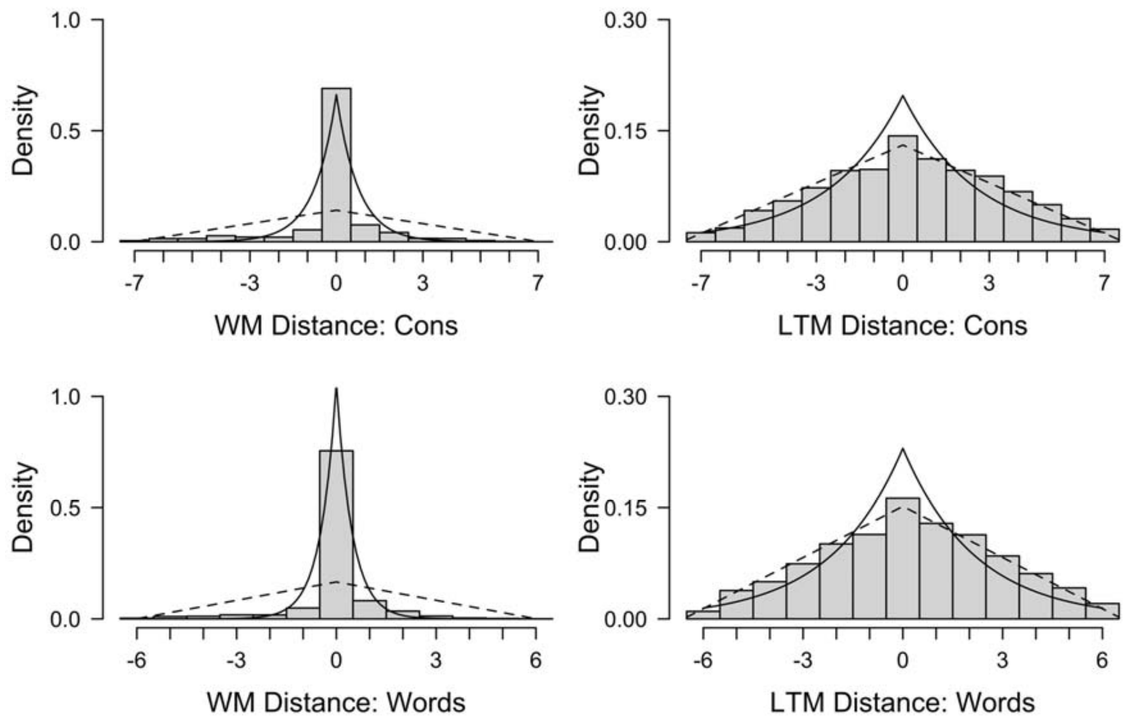

Figure 3. Histograms of the distributions of WM (left column) and LTM (right column) distances for consonants (upper row) and words (lower row) in Experiment 1. Lines show predictions from best fitting Laplace distributions (solid lines) and triangular distributions (dashed lines). WM distance is better fit with a Laplace distribution whereas LTM distance is better fit with a triangular distribution. Experiment 2 distributions were very similar to those from Experiment 1.

Source of responses to Overlap Filler Lists

Figure 3 shows the predictions from the best fitting Laplace (solid line) and triangular distribution (dashed line) on top of the distributions of distances from Experiment 1. Results strongly suggest that participants used only WM information in the Overlap Filler trials, and that there was no intrusions from LTM representation of the same items in different positions (i.e., position-item interference)[3]. For the WM distances (left column) the Laplace

[3] We also examined the possibility of intrusions driven by item-item associations in LTM in Experiments 1 and 2. Item-item interference is when one recalls, for instance item V from the first Overlap Filler list of Figure 1, and the next recalled item is item F, which is the item that followed item V in the Hebb list, instead of item Q

distribution clearly fits best; the difference in log-likelihood was over 1800 in its favor for both consonants and words. For the LTM distances (right column) we see the opposite pattern, suggesting that participants did not use the LTM information in Overlap Filler trials. The difference in log-likelihood was 211 for consonants and 182 for words favoring the triangular distribution. Experiment 2 and Experiment 3 provided equivalent results (see Table 1 in Appendix for the log-likelihood values for each experiment). Figure 4 shows the predictions from the best fitting Laplace (solid line) and triangular distribution (dashed line) on top of the distributions of distances from Experiment 3.

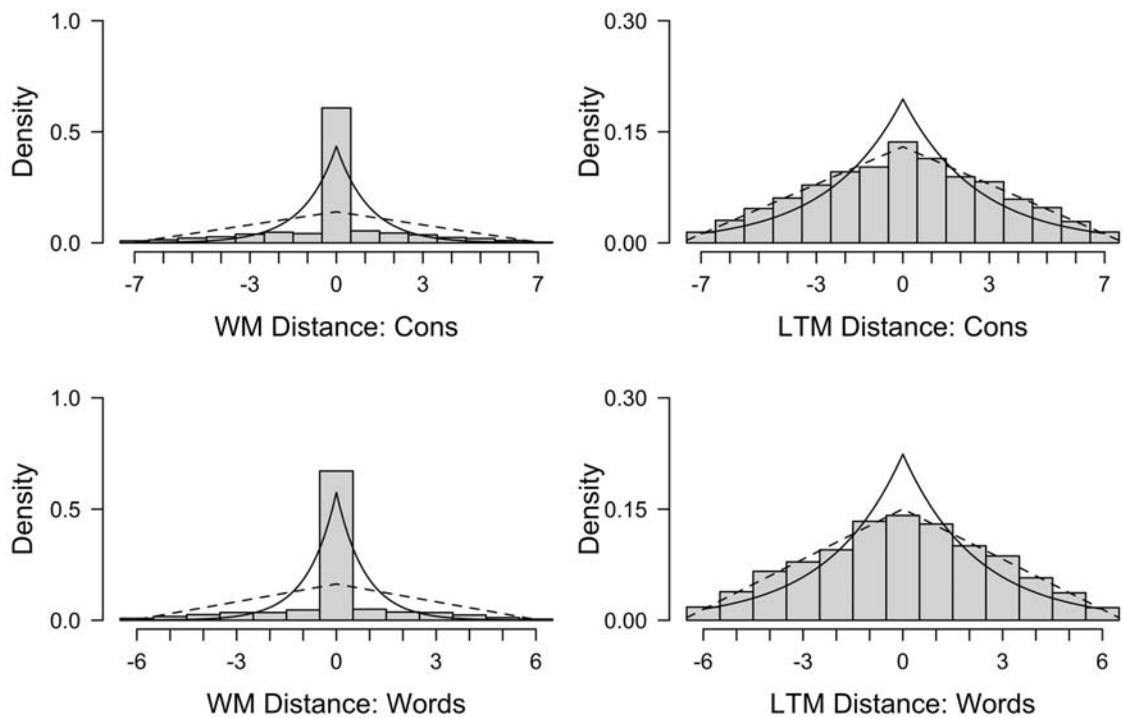

Figure 4. Histograms of the distributions of WM (left column) and LTM (right column) distances for consonants (upper row) and words (lower row) in Experiment 3. Lines show predictions from best fitting Laplace distribution (solid lines) and triangular distributions (dashed lines). WM distance is better fit with a Laplace distribution whereas LTM distance is better fit with a triangular distribution.

Discussion

which is the correct item in the Overlap Filler list. We found that these types of intrusions did not exceed chance level. This analysis is provided in the Supplemental Material.

Results from Experiments 1-3 provided evidence for proactive facilitation, reflected in better performance for Hebb lists compared to No-overlap Filler lists, and no evidence for proactive interference, as shown by comparable performance between Overlap Filler and No-overlap Filler lists. Additionally, when we examined the error distributions, we found evidence against the assumption that responses in Overlap Filler trials were influenced by the LTM trace of the Hebb list.

Apparently, participants used their LTM knowledge of the Hebb list only when it was useful – on Hebb trials – but not when it could have harmed performance – on Overlap Filler trials. Collectively, our findings so far suggest that participants can flexibly use LTM to improve their performance in a test of WM.

Experiment 4

All experiments above indicated evidence for the use of LTM when needed, and no use of LTM when it was irrelevant. This suggests LTM supports WM when it is beneficial and does not interfere with it when it is harmful. However, the three lists could be differentiated from each other as soon as the list presentation started. Already after seeing the first one or two list items, participants could have started using LTM for the Hebb lists, and decided to rely solely on WM after eliminating the usefulness of LTM for the Overlap Filler list, as envisioned by Burgess and Hitch (2006). This strategy could make the control of information flow between LTM and WM relatively easy. In a similar vein, in the model of Page and Norris (2009) the mismatch of order between the Overlap Filler list and Hebb list as a sequence would lead to at best weak activation of the LTM representation of the Hebb list during Overlap Filler lists. This weak activation might not have reached the threshold for opening the LTM gate for WM. Collectively, the findings from Experiments 1-3 support the adaptive use of LTM in a WM test, but do not provide insight about how flexible the gate between

LTM and WM is, because the gate could be opened or closed for an entire episode as soon as the utility of LTM was assessed.

Therefore, we wanted to go one step further and increase the overlap between the two lists, making it harder to determine the usefulness of LTM information during list presentation. In Experiment 4, we introduced an overlap between Hebb and Overlap Filler lists such that they shared the first three items in the same positions. If the list-initial items of the current list are particularly important as retrieval cues to LTM traces of the Hebb list, as suggested by the findings of Hitch et al. (2005), then this list-initial overlap could lead to retrieval of the entire Hebb list in Overlap Filler trials. Using that LTM information would be beneficial for recalling the first three items of the Overlap Filler lists but potentially interfere with recalling the remaining four. Alternatively, the fact that the initial match is followed by a mismatch between the current list and the LTM representation of the Hebb list could block the latter from being retrieved at all because the overlap between the two lists is below a threshold (set to 60% by Burgess and Hitch, 2006) In that case, neither proactive facilitation nor interference should be observed on Overlap Filler lists.

In the models of Burgess and Hitch (2006) and Page and Norris (2009), a chunk encapsulating the entire Hebb list is retrieved and used in an all-or-none fashion. Here we also consider the possibility that WM can draw adaptively and flexibly on LTM even within an episode, on an item-by-item basis, using it only when it is beneficial. If that is the case, we should see improved performance for the first three items for the Overlap Filler list compared to the No-overlap Filler lists, without impairment of recall of the last four items.

Table 1. Summary of results of BGLMM for all experiments. The experiment, the name of the pairwise comparison between list types, the stimuli type, Δ, 95% CI of difference distribution, the hypothesis the results provide evidence for, and the blocks the data for the analysis came from. Δ = Mean of difference distribution for two conditions given in column "Comparison", positive values indicate the first condition is larger than the second.

Exp.	Comparison	Stimuli	Serial Positions	Δ	95% CI	Evidence for	Blocks
1	Hebb > No-overlap Filler	Cons	1:7	0.154	0.109, 0.198	Proactive Facilitation	1-4
	Hebb > No-overlap Filler	Words	1:7	0.232	0.173, 0.293	Proactive Facilitation	5-8
	Overlap ~ No-overlap Filler	Cons	1:7	0.0201	-0.011, 0.053	No Proactive Interference	1-4
	Overlap ~ No-overlap Filler	Words	1:7	0.0121	-0.023, 0.047	No Proactive Interference	5-8
2	Hebb > No-overlap Filler	Cons	1:7	0.116	0.070, 0.162	Proactive Facilitation	1-4
	Hebb > No-overlap Filler	Words	1:7	0.137	0.096, 0.179	Proactive Facilitation	5-8
	Overlap ~ No-overlap Filler	Cons	1:7	-0.001	-0.025, 0.023	No Proactive Interference	1-4
	Overlap ~ No-overlap Filler	Words	1:7	-0.0201	-0.051, 0.011	No Proactive Interference	5-8
3	Hebb > No-overlap Filler	Cons	1:7	0.136	0.093, 0.177	Proactive Facilitation	1-4
	Hebb > No-overlap Filler	Words	1:7	0.146	0.107, 0.184	Proactive Facilitation	5-8
	Overlap ~ No-overlap Filler	Cons	1:7	0.011	-0.0129, 0.034	No Proactive Interference	1-4
	Overlap ~ No-overlap Filler	Words	1:7	0.019	-0.0122, 0.045	No Proactive Interference	5-8
4	Hebb > No-overlap Filler	Words	1:7	0.235	0.187, 0.277	Proactive Facilitation	1-3
	Hebb > No-overlap Filler	Words	1:7	0.255	0.171, 0.342	Proactive Facilitation	4-6
	Overlap (SP4-7) > No-overlap Filler (SP 1-7)	Words	1:3	0.142	0.068, 0.224	Proactive Facilitation	4-6
	Overlap (SP4-7) > No-overlap Filler (SP 1-7)	Words	4:7	0.214	0.118, 0.311	Proactive Facilitation	4-6
5	Hebb > No-overlap Filler (SP1-7)	Words	1:7	0.191	0.157, 0.226	Proactive Facilitation	1-4
	Hebb > No-overlap Filler (SP1-7)	Words	1:7	0.207	0.153, 0.265	Proactive Facilitation	5-8
	Overlap (SP4-7) > No-overlap Filler (SP1-7)	Words	1:3	0.120	0.067, 0.175	Proactive Facilitation	5-8
	Overlap (SP4-7) > No-overlap Filler (SP 1-7)	Words	4:7	0.123	0.059, 0.187	Proactive Facilitation	5-8
	Overlap (SP4-7) ~ No-Overlap Filler (SP4-7)	Words	1:3	0.002	-0.041, 0.045	No proactive interference	5-8
	Overlap (SP4-7) ~ No-Overlap Filler (SP4-7)	Words	4:7	0.0107	-0.042, 0.0618	No proactive interference	5-8

Methods

Participants

Twenty young adults (4 males, students from the University of Zurich) participated in this study, which lasted 60-65 minutes. They were reimbursed with course credit or 15 Swiss Francs per hour for their time.

Materials

For this experiment, stimuli were only words. We made two sets of seven words, one for the Hebb and Overlap Filler lists, and one for the No-overlap Filler lists. The words were similar to the ones from the Experiments 1-3. Hebb and Overlap Filler lists were constructed from one set of words, No-overlap Filler lists from the other. Overlap Filler lists were identical to Hebb lists in the first three positions; the items in the remaining four positions were ordered randomly for every trial. Therefore, we now call this condition Overlap Filler (SP4-7) because only items from serial positions 4 to 7 were randomized. The fourth item in the Overlap Filler list could not be the same as the fourth item in the Hebb list. This made it unambiguous on the presentation of the fourth item that this list was not the Hebb list.

Procedure

Experiment 4 differed from Experiments 1-3 in the following ways. The experiment consisted of six blocks of 16 trials, resulting in 96 trials.

The first half of the experiment (Blocks 1-3), referred to as the *learning phase*, consisted of only Hebb and No-overlap Filler lists. We wanted participants to first learn the Hebb lists well before we introduced Overlap Filler lists. Each block consisted of 4 mini-blocks, and each mini-block was comprised of 2 Hebb and 2 No-overlap Filler trials. The order of Hebb and No-overlap Filler trials in a mini-block was randomized with the constraint that the first trial of a mini-block was not the same type of list as the last trial of the previous mini-block.

The second half of the experiment (Blocks 4-6), referred to as the *transfer phase*, introduced the Overlap Filler (SP4-7) condition. Each mini-block had one Hebb, one Overlap Filler (SP4-7), and two No-overlap Filler trials, presented in random order with the same constraint as explained above. In total there were 48 No-overlap Filler, 36 Hebb, and 12 Overlap Filler (SP4-7) trials.

Data Analysis

Immediate Recall Accuracy. Our analytic approach was similar to the first three experiments, with some changes owing to the design changes: The analysis focused on comparing the three conditions in the transfer phase. We dropped material as fixed effect because there were only word lists.

Additionally, all lists were divided into two parts for the analysis. Performance was averaged across serial positions 1-3 for the first part, and serial positions 4-7 for the second part. This distinction was included in the BGLMM as a fixed effect. Evidence for proactive facilitation comes from performance for the first part of Overlap Filler (SP4-7) lists being better than the first part of the No-overlap Filler lists, and comparable to the Hebb lists. Evidence for proactive interference comes from performance for the second part of Overlap Filler (SP4-7) lists being worse than for the No-overlap Filler lists.

Responses from WM and LTM for Overlap Filler lists. We again calculated two distance measures, WM distance and LTM distance, for the Overlap Filler (SP4-7) lists for Experiment 4. The procedure for calculating the distances was the same as Experiments 1-3 with one difference. For this analysis, we only examined the distances for the last four serial positions, as those were the mismatching information between WM and LTM. For this reason, we changed the distance range from -3 to 3. Figure 6 (upper row) shows the distributions for the two distances.

Results

Immediate Recall Accuracy

Model convergence was inspected and all $\hat{R}$ were below 1.01. Figure 5A depicts the results from Experiment 4. Table 1 shows evidence for all pairwise comparisons of interest for Experiment 4. Pairwise comparison of Hebb and No-overlap Filler lists from the learning phase indicated evidence for proactive facilitation (i.e., Hebb > No-overlap Filler). For the transfer phase, we found that the first three list items were better recalled in Overlap Filler (SP4-7) trials compared to No-overlap Filler trials.

To test for proactive interference, we compared the second part of the lists. Performance for the Overlap Filler (SP4-7) list was not worse but better than the No-overlap Filler lists.

Together, these findings indicate that participants were able to draw on LTM for the first serial positions, and exclusively use their WM for the rest of the list. Use of LTM for the first part not only improved recall of that part of the list but also boosted performance for the subsequent items, providing further evidence for proactive facilitation.

Source of responses to Overlap Filler Lists

We again found that the Laplace distribution fits WM distance better than the triangular distribution (difference in log-likelihood: 730), and the triangular distribution fits LTM distance better than the Laplace distribution (difference in log-likelihood: 59). This indicates that Overlap Filler responses for the last four items likely came exclusively from

WM representations, and were not contaminated with LTM representation of how these items were ordered on the Hebb list.

Discussion

Results from Experiment 4 indicated that participants were able to flexibly use or refrain from using LTM representations within the same episode to guide their serial recall performance. However, we were not able to fully rule out proactive interference for the second half of the Overlap Filler (SP4-7) lists, for the following reason. Overlap Filler (SP4-7) list performance for the last four items was better than No-overlap Filler list performance. This suggests that use of LTM for the first three items boosted performance for the subsequent items in Overlap Filler (SP4-7) lists. We hypothesize that the use of LTM reduced the load on WM, which made performance better for the subsequent items (Norris et al., 2020; Thalmann et al., 2019). If this is the case, then even if there was proactive interference from the use of LTM to WM, its effect might not be visible because it is counteracted by the boosted performance. We will address this issue with Experiment 5.

Experiment 5

In Experiment 5, we aimed to introduce a control list that had the same load on WM as the Overlap Filler (SP4-7) list to have a better control condition to test for proactive interference. This No-Overlap Filler (SP4-7) list was constructed from a third set of words not used for any other kind of list. In the No-Overlap Filler (SP4-7) list the first three items were always the same, in the same order, whereas the last four items were presented in random positions throughout the experiment. Hence, over the course of the experiment participants could learn the first triplet of the No-Overlap Filler (SP4-7) lists but could not learn anything about how they continued. In this way, we equate the probability of retrieving the first three items from LTM and WM for Overlap Filler (SP4-7) and No-Overlap

Filler (SP4-7) lists. Different from the Overlap Filler (SP4-7) lists, for the No-Overlap Filler (SP4-7) lists there was no pertinent LTM information about the final four items, so they could not suffer proactive interference. If there is interference from LTM to WM, we expect the Overlap Filler (SP4-7) lists to have worse performance than the No-Overlap Filler (SP4-7) lists. If indeed there is no interference from LTM to WM, and LTM can be flexibly used together with WM, then performance to No-Overlap Filler (SP4-7) lists and Overlap Filler (SP4-7) lists should be similar. This would mean we should also see a boost for No-Overlap Filler (SP4-7) Filler lists' performance for the last four items, which we will measure by comparing it to the No-overlap Filler (SP 1-7) lists whose items were always presented in a random order across all seven serial positions.

Methods

Participants

Twenty young adults (5 males, students from the University of Zurich) participated in this study, which lasted 70-75 minutes. They were reimbursed with course credit or 15 Swiss Francs per hour for their time.

Materials

For this experiment, stimuli were only words. We made three sets of seven words, one for the Hebb and Overlap Filler (SP4-7) lists, one for the No-overlap Filler (SP1-7) lists, and one for the No-Overlap Filler (SP4-7) lists. The words were similar to the words used in the experiments presented above. Hebb and Overlap Filler (SP4-7) lists were identical for the first three items, and the last four items diverged in their assignment of items to positions in the same way as in Experiment 4. No-Overlap Filler (SP4-7) Filler lists did not overlap with any of the other lists. In this list the first three items were always repeated in the same order, whereas the last four items were presented in a random order in each trial.

Procedure

The experiment consisted of eight blocks of 16 trials each, resulting in 128 trials. The learning phase (Blocks 1-4) consisted of only Hebb and No-overlap Filler lists. Each block consisted of 4 mini-blocks, and each mini-block was comprised of 2 Hebb and 2 No-overlap Filler List trials. The order of Hebb and No-overlap Filler trials in a mini-block was randomized with the constraint that the first trial of a mini-block was not the same type of list as the last trial of the previous mini-block.

In the transfer phase (Blocks 5-8), we introduced Overlap Filler (SP4-7) and No-Overlap Filler (SP4-7) lists in the mini-blocks. A mini-block had one Hebb, one Overlap Filler (SP4-7), one No-Overlap Filler (SP4-7) Filler, and one No-overlap Filler (SP1-7) list. The order of the lists in a mini-block were again randomized with the same constraint as explained above. In total there were 48 No-overlap Filler (SP1-7), 48 Hebb, 16 Overlap Filler (SP4-7), and 16 No-Overlap Filler (SP4-7) Filler trials.

The test procedure was the same as the one in Experiments 1, 2, and 4.

Data Analysis

We used the same analytic approach as for Experiment 4, with the exception that there were four levels of condition for the transfer phase, and we tested proactive interference by comparing accuracy on the last four positions between Overlap Filler (SP4-7) and No-Overlap Filler (SP4-7) lists.

Results

Results from Experiment 5 were similar to Experiment 4: Better performance for Hebb lists than No-overlap Filler (SP1-7) lists, and better performance for Overlap Filler (SP4-7) lists than No-overlap Filler (SP1-7) lists for both the first and the second parts of the lists. The new finding was that performance for the No-Overlap Filler (SP4-7) lists was similar to that for Overlap Filler (SP4-7) lists, and both showed better performance than No-overlap Filler lists (SP1-7). The fact that the last four items of Overlap Filler (SP4-7) lists were recalled as well as those of No-overlap Filler (SP4-7) lists shows that there was no proactive interference from LTM to WM for Overlap Filler (SP4-7) lists. Moreover, the finding that recall of the last four lists in both these conditions was superior to that in the No-overlap Filler (SP1-7) showed that the recruitment of LTM for the first three list items led to a comparable reduction of WM load for the remaining four items on those lists.

One could expect higher performance for the Overlap Filler (SP4-7) lists compared to No-Overlap Filler (SP4-7) Filler lists for the following reason. Participants had to learn the first three items of the No-Overlap Filler (SP4-7) Filler list during the transfer phase to benefit from LTM, whereas they could immediately benefit from LTM of Hebb list for the Overlap Filler (SP4-7) list. As we did not see such a difference in the transfer phase, we need to consider an alternative possibility: Upon entering the transfer phase, participants might have noticed the risk of proactive interference from using their knowledge of the Hebb list on the new Overlap Filler (SP4-7) lists. To prevent proactive interference, they could have stopped using their existing knowledge, and relearned the first three items of the Overlap Filler (SP4-7) lists as a separate chunk.

To understand why the Overlap Filler (SP4-7) and the No-Overlap Filler (SP4-7) Filler conditions had equivalent performance, we visually examined performance of these conditions as a function of mini-blocks. Figure 6 (left panel) shows that even in the first mini-block of the transfer phase, when the Overlap Filler (SP4-7) condition is first introduced, performance for SP 1-3 in this condition is as good as Hebb list performance. This suggests that participants benefited from their knowledge of the Hebb list from the start, rather than starting to learn a new chunk. In comparison, No-Overlap Filler (SP4-7) lists started with somewhat weaker performance in the first two mini-blocks. However, this difference quickly dissipates. Participants seem to have learned the first three items of the new No-Overlap Filler (SP4-7) Filler lists very quickly and benefit from this knowledge already at the third repetition (i.e., mini-block 18). This is why, on average across the mini-blocks of the transfer phase, performance of these two conditions was indistinguishable. Even in the first mini-blocks their performance was so close that the evidence for their difference is ambiguous.

To further investigate whether the Overlap Filler (SP4-7) benefit from already existing LTM of Hebb list, we also examined response times difference between the Overlap Filler (SP4-7) and No-overlap Filler (SP4-7) lists with Bayesian t-tests. Overlap Filler (SP4-7) had shorter response times (mean = 1.48) compared to No-overlap Filler (SP4-7) lists (mean = 1.69, BF = 19.5). This difference lasted throughout the transfer phase. It suggests that response times were sensitive to strength of learning, and participants retrieved faster for the Overlap Filler (SP4-7) lists, with its longer history of learning the first three items, than the No-overlap Filler (SP4-7) lists. These results indicate that participants did not throw away their knowledge of the Hebb lists when confronted with the Overlap Filler (SP4-7) lists, but rather continued to use it to support memory of the first three items.

Similar to Experiment 4, the Laplace distribution fits WM distances better than the triangular distribution (difference in max log-likelihood: 962) and the triangular distribution fits the LTM distances better than the Laplace distribution (difference in max log-likelihood: 123).

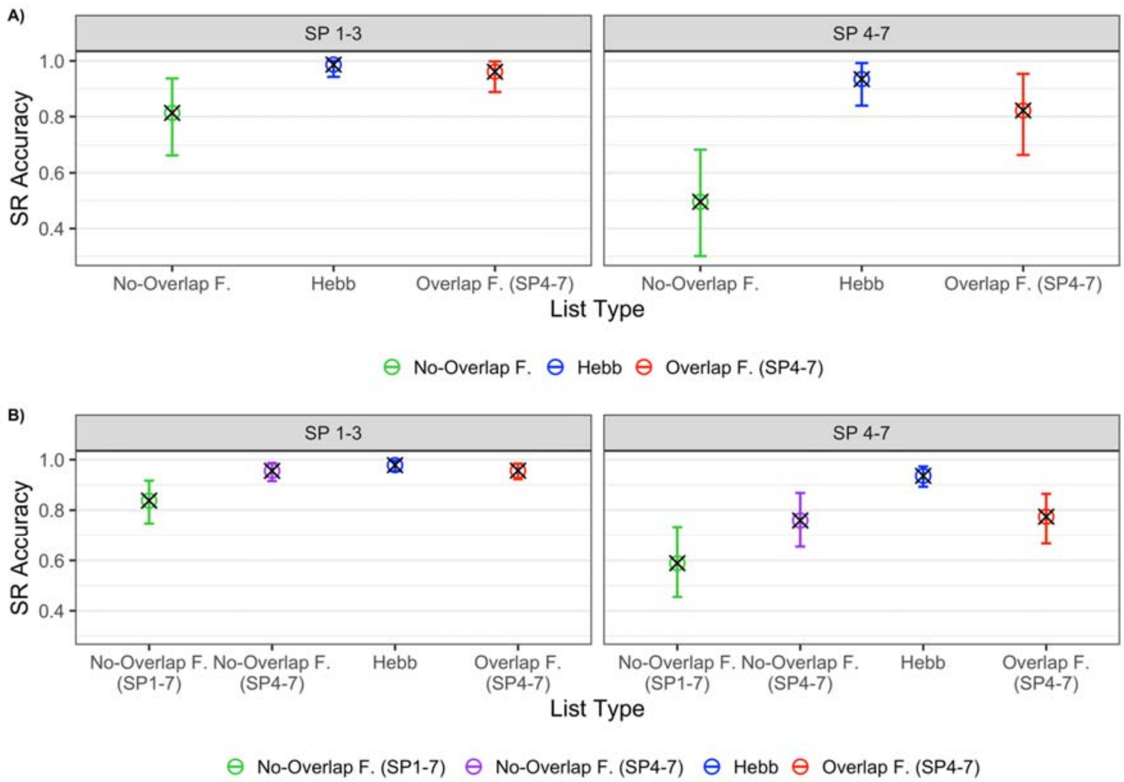

Figure 5. Proportion correct of the immediate serial recall (row A and B) memory performance of all lists of Experiment 4 (row A), and 5 (row B). Colored squares and error bars represent estimated proportions and 95% CIs. X marks represent observed proportions. Overlap indicates the model adequately describes the data.

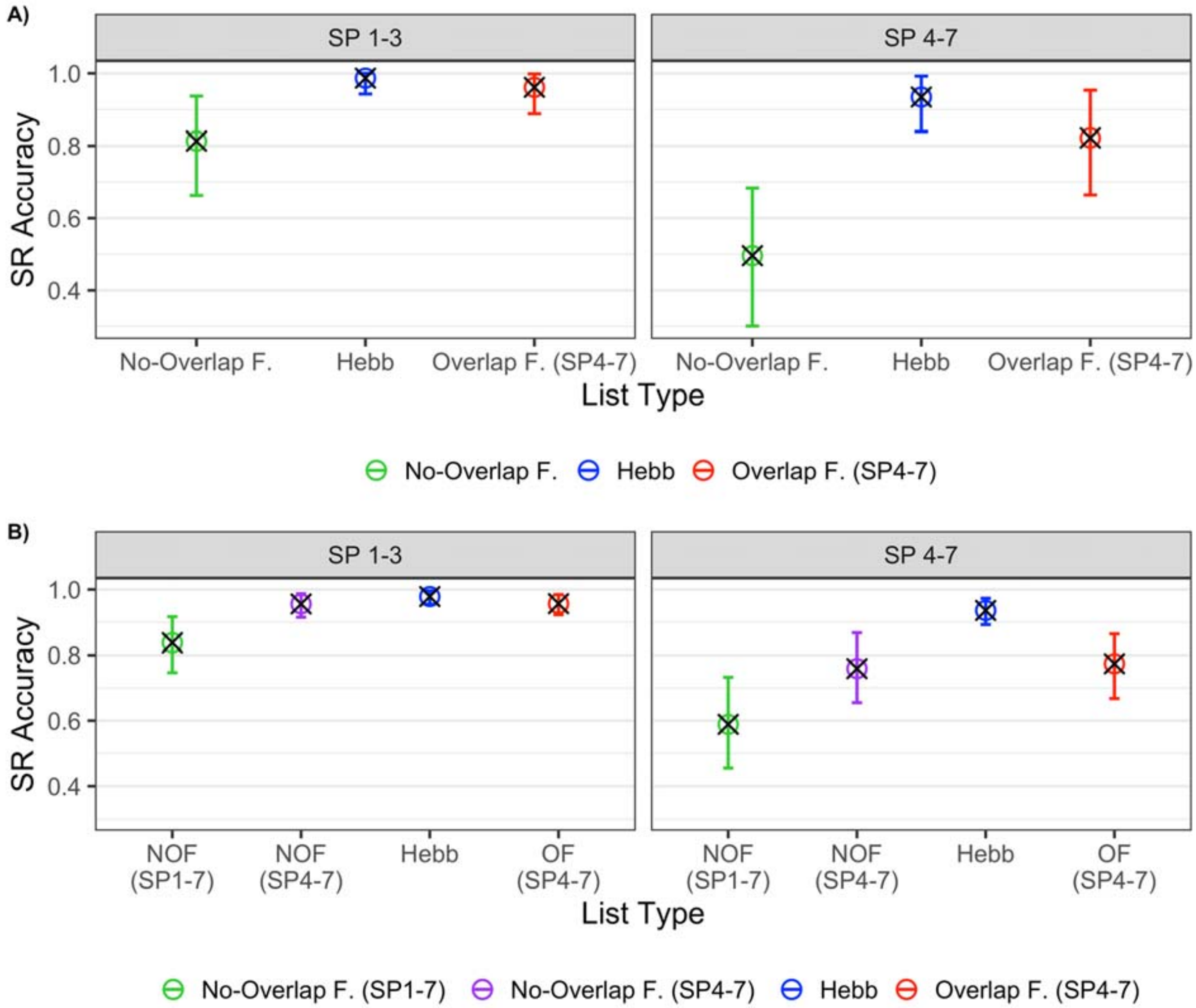

Figure 5. Proportion correct of the immediate serial recall (row A and B) memory performance of all lists of Experiment 4 (row A), and 5 (row B). Colored squares and error bars represent estimated proportions and 95% CIs. X marks represent observed proportions. Overlap indicates the model adequately describes the data. OF = Overlap Filler, NOF = No-Overlap Filler, SP = serial position.

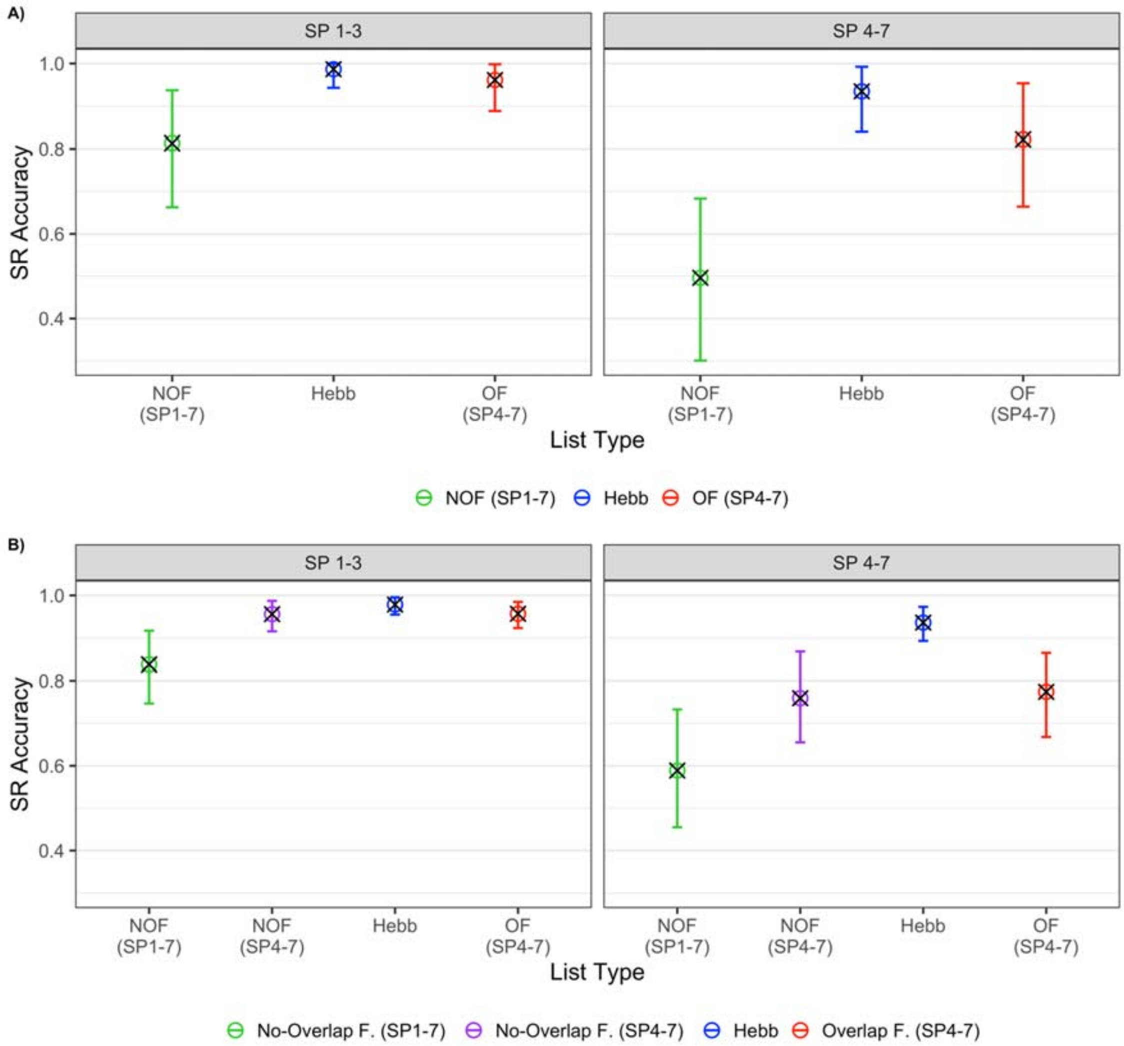

Figure 5. Proportion correct of the immediate serial recall (row A and B) memory performance of all lists of Experiment 4 (row A), and 5 (row B). Colored squares and error bars represent estimated proportions and 95% CIs. X marks represent observed proportions. Overlap indicates the model adequately describes the data. OF = Overlap Filler, NOF = No-Overlap Filler, SP = serial position.

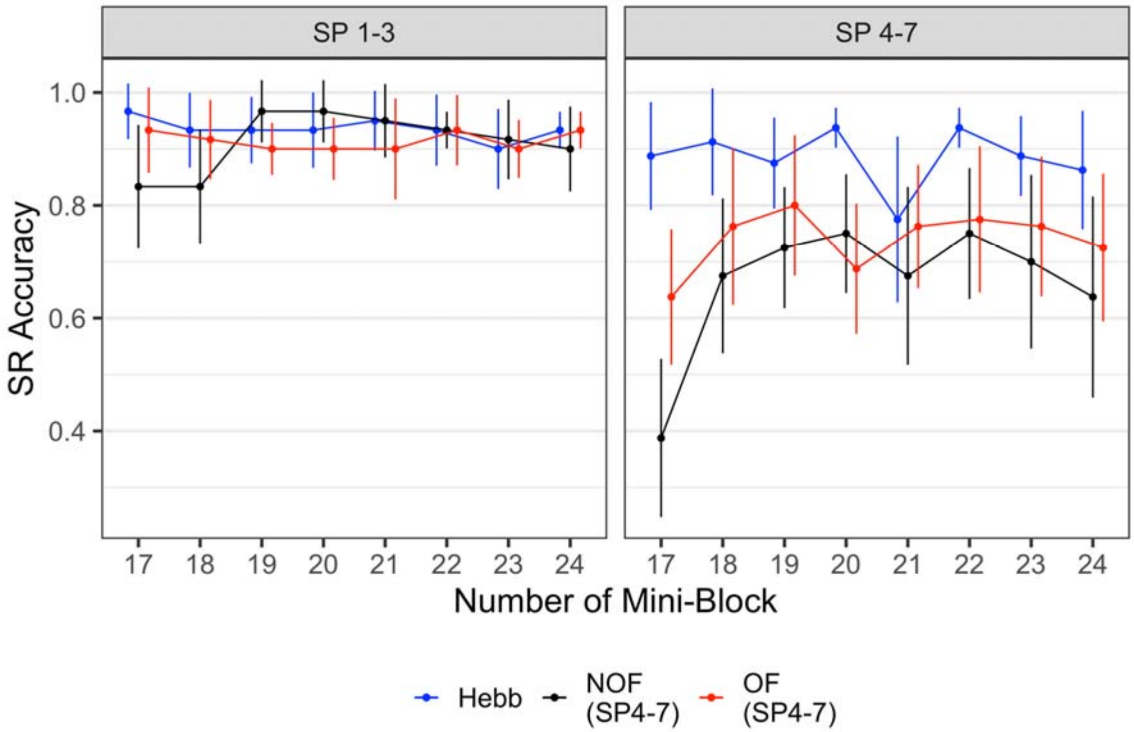

Figure 6. Proportion correct of the immediate serial recall memory performance of Hebb list, Overlap Filler (SP4-7), and No-Overlap Filler (SP4-7) lists as a function of mini-block. The data is from the second half of the study, therefore, mini-blocks start from 17. We stopped plotting at mini-block 24 to show the initial learning phase of the newly introduced lists. Different panels correspond to performance of different parts of the lists, SP 1-3 = serial positions 1 to 3, SP4-7 = serial positions 4 to 7. Errors bars show within-subjects %95 confidence interval. OF = Overlap Filler, NOF = No-Overlap Filler, SP = Serial position.

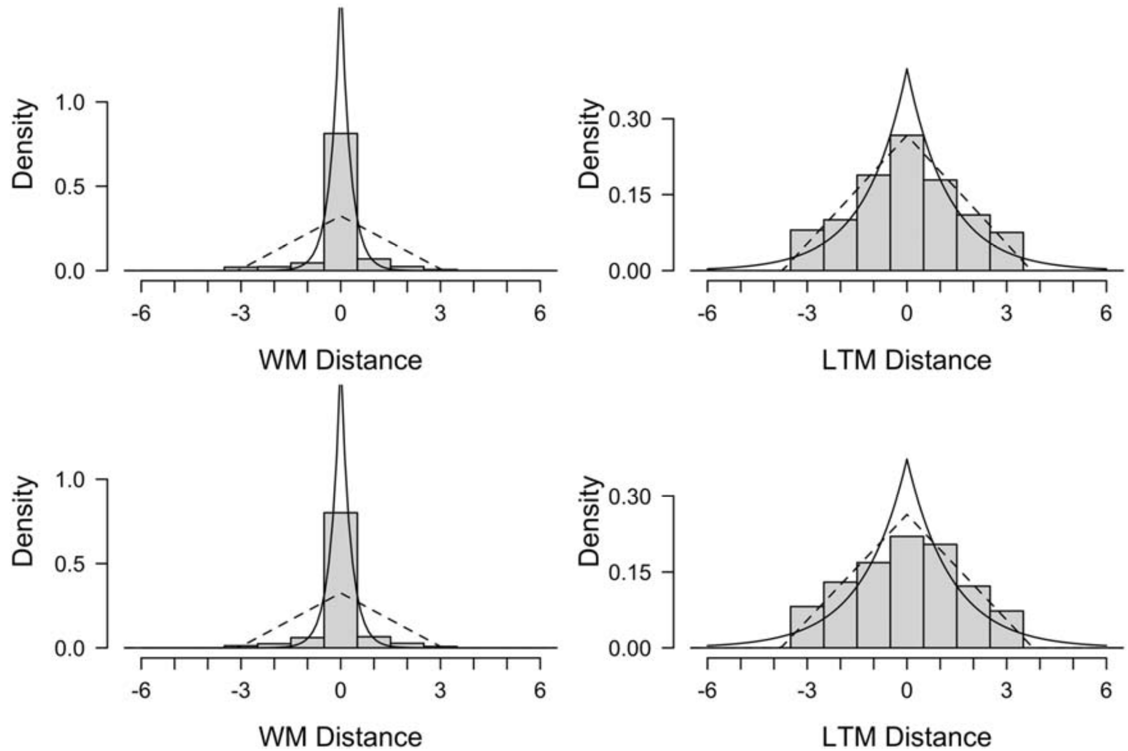

Figure 7. Histograms of the distributions of WM (left column) and LTM (right column) distances for Experiment 4 (upper row) and Experiment 5 (lower row). Lines show predictions from best fitting Laplace distribution (solid lines) and triangular distributions (dashed lines). WM distance is better fit with a Laplace distribution whereas LTM distance is better fit with a triangular distribution.

Discussion

Experiment 5 confirmed the ability to flexibly use beneficial information from LTM while discarding harmful information during Overlap Filler trials. We again found no evidence for proactive interference from LTM to WM. These results collectively support the conjecture that WM recruits LTM only when it is beneficial. According to these results, participants were able to recognize which parts of the LTM traces of the Hebb list were helpful and which were not, and partition the task of remembering the Overlap Filler (SP4-7) and No-Overlap Filler (SP4-7) Filler lists into LTM (old part of the list) and WM (new part). This in turn helped them to remember both parts of these lists better than No-overlap Filler (SP 1-7) lists. Because performance on Overlap Filler (SP4-7) lists was equivalent to that on No-Overlap Filler (SP4-7) lists, we confidently eliminate the possibility of proactive interference from Hebb lists to Overlap Filler (SP4-7) lists.

General Discussion

Across five studies, we examined the contribution of long-term memory to working memory. Our goal was to test whether information flow from long-term memory to working memory was controlled with a flexible gate that admits information when it is useful and blocks it when it is potentially harmful. If information flow between WM and LTM is controlled and selective, only useful information should be admitted that WM benefits from (i.e., proactive facilitation but no proactive interference). If information flow is not controlled, and LTM contribution is more obligatory than voluntary, LTM information both useful and harmful can be admitted into WM. An obligatory contribution would result in both a beneficial and harmful effect on WM performance (both proactive facilitation and interference), or else, LTM information is consistently blocked out and none of these effects should be found.

Findings from each experiment showed evidence for proactive facilitation and no evidence for proactive interference. We suggest that working memory recruits long-term memory depending on the utility of information such that only beneficial long-term memory contents were used to guide memory performance.

Evidence against proactive interference from LTM to WM

We showed no evidence of obligatory contributions of LTM to WM. An obligatory contribution would mean that LTM contents are automatically retrieved whenever there is a sufficient resemblance of those contents to the current information in WM. This would result in both facilitation when the WM and LTM contents match, and interference when they mismatch. However, we only observed facilitation. This finding supported the hypothesis that information flow from LTM to WM is gated in a flexible and adaptive manner.

These findings extend the findings of Oberauer et al. (2017) that provided evidence for proactive facilitation without proactive interference. Our results differ from theirs in one respect. Our error pattern analyses showed no evidence for an LTM contribution to memory performance in the Overlap Filler condition. Oberauer et al. (2017) showed that on the mismatch trials, participants sometimes responded with information from LTM instead of WM. When we examined responses given to LTM mismatch lists in the present experiments, we did not observe any evidence for responses coming from LTM. We will offer an explanation for this discrepancy after we discussed the mechanisms explaining the present findings.

Divide and conquer: Prior knowledge in long-term memory helps maintenance of novel information in working memory

In Experiments 4 and 5, LTM facilitated immediate recall not only of the learned list segment but also for subsequently encoded items. Apparently, WM load was reduced by the use of LTM for half of the list, so that WM was able to support the remaining items in the list better[4].

This finding is very similar to the finding of Thalmann et al. (2019), who showed a similar boost for the items that came after a chunk within a list. It is possible that in our study, Hebb learning resulted in the formation of chunks encapsulating an entire list, or of groups of 3 to 4 items within a list. Similar to the chunks that reduced the load on WM in the Thalmann et

[4] We also evaluated the possibility that the boost for the subsequent items following the LTM items in Overlap Filler lists might be due to an experimental artifact. When the first three items are correctly retrieved from LTM, the probability of correctly guessing the remaining items is greater compared to the No-overlap Filler lists (1/4 for Overlap Filler lists and 1/7 for No-overlap Filler lists). The heightened probability of guessing could inflate the performance for these items in the Overlap Filler lists. Therefore, we corrected performance for the last four items in these lists taking these guessing probabilities into account and performed a Bayesian t-test to test the difference between the two lists. Test results provided evidence for better performance for the last four items in Overlap Filler lists compared to No-overlap Filler lists (Exp.4: $BF_{10} = 5.2$; Exp.5: $BF_{10} = 7.3$). This assured us that our findings were not due to an experimental artifact.

al. (2019) study, when part of a new list matched a chunk, it was held in WM as a chunk, reducing the load on WM, which in turn improved memory for other information maintained concurrently.

This reduction of load implies that the use of LTM to assist WM commences already at encoding: Incoming stimuli go through a recognition process that examines whether this information exists in LTM or not. If an LTM representation is detected, utility of encoding this information in WM is lower than for information that does not match because the LTM representation can be used instead. This utility assessment could work as a filtering system in which stimuli are encoded differently depending on their similarity to any existing representation in LTM. The WM system can decide, on an item-by-item or group-by-group basis (if the list is grouped), whether to encode the item/group fully into WM or instead encode a pointer to LTM which carries a very small WM load. This mechanism is similar to the mechanism proposed by Burgess and Hitch (2006) in which use of LTM representation to guide serial recall on Hebb lists is decided based on a cumulative matching process during encoding.

This utility assessment at encoding can either be done on an item-by-item basis (checking for each item individually how much it needs to be encoded) or on a more aggregate level, looking at a group or sub-sequence of items. Hebb learning suggests that the mechanism operates on an aggregate level: The first item of a (partial or full) Hebb list on its own does not match any LTM knowledge much more than the first item of a No-overlap Filler list. Only when looking at a sequence of at least the first 2 items can a person detect a substantially better-than-chance match to an existing chunk in LTM. Moreover, Cumming et al. (2003) found no transfer of learning a Hebb list to lists in which every second item of the Hebb list was repeated. If the utility assessment was made on an item-by-item basis, it would recognize every second item-position relation as already learned, and recall of lists repeating every second item from the Hebb list should benefit in the same way as our Overlap-Filler (SP4-7) lists, which repeat the first three items.

Therefore, we envision a mechanism that continuously matches the last *n* items of a list to potentially matching chunks in LTM, similar to the chunk recognition mechanism described in the TRACX model (French et al., 2011), which was proposed for sequence segmentation and chunk extraction in implicit sequence-learning tasks. When this mechanism detects a chunk matching the last *n* items, the decision on how strongly the last *n* items are encoded into WM has already been made – one could only retro-actively weaken these items by removing them from WM and replacing them by a pointer to the LTM representation.

One question that arises in this context is how much a chunk adds to the load on WM. To sketch an answer to this question, we compared memory performance for the last four serial positions in Overlap Filler (SP4-7) lists to serial-recall performance of a four-item list (Loaiza & Halse, 2019) and a five-item list (Souza & Oberauer, 2017) from recent serial recall studies that were similar to our experimental setting. Performance averaged across serial positions for the last four items in the Overlap Filler (SP4-7) list was close to serial recall of a five-item list, and worse than serial recall of a four-item list. This observation hints that there is some form of information maintained in WM for the learned list segment, and that information loads WM about as much as a single item. One plausible way of maintaining a set of chunked items is maintaining a pointer to the LTM representation, similar to what has been suggested for maintaining chunks in WM (Chen & Cowan, 2009). This pointer could be anything that serves as a reliable retrieval cue to the learned chunk in LTM – for instance, an ad-hoc label participants gave to the Hebb list – such that at test that chunk is retrieved and unpacked, and its first three elements are used to output positions 1 to 3.

Implications for chunking models of Hebb learning

We envision a continuous assessment of the match between WM contents and LTM knowledge to gauge the utility of encoding and maintaining information in WM. As we noted above, this is similar to the incremental match computation in the model of Burgess and Hitch (2006). However, there is also an important difference: Our results suggest that the decision on whether to fully encode information into WM or to rely on LTM can be made for only a part of a list, whereas the model of Burgess and Hitch (2006) decides whether or not to use LTM only on the level of an entire list. Similarly, in the model of Page and Norris (2009) a Hebb list, learned as a chunk, can only be retrieved and used in an all-or-none fashion.

The models of Burgess and Hitch (2006), and of Page and Norris (2009), could accommodate the findings of our Experiments 4 and 5 by assuming that people acquire not only a chunk encapsulating the entire Hebb list, but also smaller chunks incorporating only subsets of that list. If participants spontaneously grouped the lists in the training phase according to a 3-4 pattern, they could have learned chunks of each group of the Hebb list, and re-use the chunk of the first group when encountering the Overlap Filler (SP4-7) lists. A more radical assumption would be that every possible subset of a list is learned as its own chunk in parallel. The models in their current form do not assume the acquisition of sub-list chunks, but they could perhaps be modified to enable that possibility. At least the model of Page and Norris (2009) would require revision of some of its assumptions, because in its present form, learning occurs only for representations that win the competition for representing the input. A fledgling chunk can win the competition against individual-item representations only after the entire list has been presented, and the activations of item-representations have subsided.

At that point, learning associates the chunk to the entire list, not to any subset of it.

A flexible gate between working memory and long-term memory

Our findings indicate that in the present serial-recall experiments the flexible gate between working memory and long-term memory is implemented differently than in the visual WM tests used by Oberauer et al. (2017). In our experiments, an assessment of the utility of matching LTM representations is already made during encoding of a new list, so that when useful LTM information is found, less information needs to be maintained in WM. The decision whether or not to use LTM information therefore is already made at encoding. In contrast, Oberauer et al. (2017) envisioned a mechanism that opens or closes the gate at retrieval: If (and only if) retrieval from WM returns little useful information, an attempt is made to retrieve relevant information from LTM.

These differences in when the gating mechanism operates could explain the discrepancy between our results and those of Oberauer et al. (2017): A gating mechanism at retrieval draws on LTM when information in WM is poor, thereby generates an above-chance proportion of responses reflecting LTM knowledge even when that knowledge mis-matches the information to be maintained in WM, as observed by Oberauer et al. (2017). By contrast, a gating mechanism deciding at encoding whether to use LTM rejects all LTM knowledge that mismatches the to-be-encoded stimuli. There is no process by which LTM knowledge is drawn upon as a fallback option if, and only if, retrieval from WM fails.

In both cases, the function of the gate remains the same – effective use of LTM without proactive interference – but implemented in a different way. Our findings support the flexible-gate hypothesis but also show a different mechanism that the gating is implemented with. During encoding, a filtering mechanism leads to differential encoding of LTM items in WM. This leaves no need for the gate to operate during retrieval.

Conclusion

Using long-term memory knowledge that was beneficial for a WM task improved overall task performance. Even in the instances in which LTM knowledge was partially beneficial, only the helpful part was used to guide performance with no interference from the harmful part. Our results generalize the finding by Oberauer et al. (2017) and support the flexible gate hypothesis: The flow of information between WM and LTM is controlled by a gate that is adaptive and selective, admitting information from LTM into WM depending on its utility.

References

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3). https://doi.org/10.1016/j.jml.2012.11.001

Botvinick, M., & Watanabe, T. (2007). From Numerosity to Ordinal Rank: A Gain-Field Model of Serial Order Representation in Cortical Working Memory. *Journal of Neuroscience*, *27*(32), 8636–8642. https://doi.org/10.1523/JNEUROSCI.2110-07.2007

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, *107*(1), 127–181. https://doi.org/10.1037/0033-295X.107.1.127

Burgess, N., & Hitch, G. (2005). Computational models of working memory: Putting long-term memory into context. *Trends in Cognitive Sciences*, *9*(11), 535–541. https://doi.org/10.1016/j.tics.2005.09.011

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*(4), 627–652. https://doi.org/10.1016/j.jml.2006.08.005

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). *Stan*: A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Chen, Z., & Cowan, N. (2009). Core verbal working-memory capacity: The limit in words retained without covert articulation. *Quarterly Journal of Experimental Psychology*, *62*(7), 1420–1429. https://doi.org/10.1080/17470210802453977

Cowan, N. (2008). Chapter 20 What are the differences between long-term, short-term, and working memory? In *Progress in Brain Research* (Vol. 169, pp. 323–338). Elsevier. https://doi.org/10.1016/S0079-6123(07)00020-9

Cowan, N. (2012). *Working Memory Capacity* (1st ed.). Psychology Press. https://doi.org/10.4324/9780203342398

Cowan, N., Chen, Z., & Rouder, J. N. (2004). Constant Capacity in an Immediate Serial-Recall Task: A Logical Sequel to Miller (1956). *Psychological Science*, *15*(9), 634–640. https://doi.org/10.1111/j.0956-7976.2004.00732.x

Cumming, N., Page, M., & Norris, D. (2003). Testing a positional model of the Hebb effect. *Memory*, *11*(1), 43–63. https://doi.org/10.1080/741938175

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*(4), 575–597. https://doi.org/10.1037/0278-7393.10.4.575

Fendrich, D. W., Healy, A. F., & Bourne, L. E. (1991). Long-term repetition effects for motoric and perceptual procedures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(1), 137–151. https://doi.org/10.1037/0278-7393.17.1.137

French, R. M., Addyman, C., & Mareschal, D. (2011). TRACX: A recognition-based connectionist framework for sequence segmentation and chunk extraction. *Psychological Review*, *118*(4), 614–636. https://doi.org/10.1037/a0025255

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian data analysis*. http://public.eblib.com/choice/publicfullrecord.aspx?p=1438153

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

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*(3–4), 247–258. https://doi.org/10.1080/09658210344000314

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, *100*(9), 1989–2001. https://doi.org/10.1016/j.jmva.2009.04.008

Lewandowsky, S., & Farrell, S. (2008). Short-Term Memory: New Data and a Model. In *Psychology of Learning and Motivation* (Vol. 49, pp. 1–48). Elsevier. https://doi.org/10.1016/S0079-7421(08)00001-7

Loaiza, V. M., & Halse, S. C. (2019). Where working memory meets long-term memory: The interplay of list length and distractors on memory performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(8), 1455–1472. https://doi.org/10.1037/xlm0000652

Majerus, S., & Oberauer, K. (2019). Working memory and serial order: Evidence against numerical order codes but for item–position associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000792

Nakayama, M., & Saito, S. (2017). Position-element frequency learning is dissociable from Hebb repetition learning. *Journal of Memory and Language*, *94*, 235–253. https://doi.org/10.1016/j.jml.2016.11.007

Norris, D., Kalm, K., & Hall, J. (2020). Chunking and redintegration in verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(5), 872–893. https://doi.org/10.1037/xlm0000762

Oberauer, K. (2009). Chapter 2 Design for a Working Memory. In *Psychology of Learning and Motivation* (Vol. 51, pp. 45–100). Elsevier. https://doi.org/10.1016/S0079-7421(09)51002-X

Oberauer, K., Awh, E., & Sutterer, D. W. (2017). The role of long-term memory in a test of visual working memory: Proactive facilitation but no proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(1), 1–22. https://doi.org/10.1037/xlm0000302

Oberauer, K., Jones, T., & Lewandowsky, S. (2015). The Hebb repetition effect in simple and complex memory span. *Memory & Cognition*, *43*(6), 852–865. https://doi.org/10.3758/s13421-015-0512-8

Oberauer, K., & Meyer, N. (2009). The contributions of encoding, retention, and recall to the Hebb effect. *Memory*, *17*(7), 774–781. https://doi.org/10.1080/09658210903107861

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 B: Biological Sciences*, *364*(1536), 3737–3753. https://doi.org/10.1098/rstb.2009.0173

Portrat, S., Guida, A., Phénix, T., & Lemaire, B. (2016). Promoting the experimental dialogue between working memory and chunking: Behavioral data and simulation. *Memory & Cognition*, *44*(3), 420–434. https://doi.org/10.3758/s13421-015-0572-9

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356–374. https://doi.org/10.1016/j.jmp.2012.08.001

Souza, A. S., & Oberauer, K. (2017). Time to process information in working memory improves episodic memory. *Journal of Memory and Language*, *96*, 155–167. https://doi.org/10.1016/j.jml.2017.07.002

Thalmann, M., Souza, A. S., & Oberauer, K. (2019). How does chunking help working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(1), 37–55. https://doi.org/10.1037/xlm0000578

Appendix

Table 1. Summary of results of response analyses for Overlap Filler lists. The experiment, Distance (Working Memory vs. Long-term Memory, Distribution, the stimuli type, Loglikelihood values, and the difference in log likelihood values between the two distributions. The difference is always a positive number and the comparison before the number indicates which distribution had higher loglikelihood values (L > T means Laplace is a better fit, T > L means Triangular is a better fit).

Exp.	Distance	Distribution	Stimulus	Loglikelihood	Difference
1	WM	Laplace	Cons	-3588.48	L > T: 1812
	WM	Triangular	Cons	-5400.792	
	LTM	Laplace	Cons	-6660.562	T > L: 211
	LTM	Triangular	Cons	-6449.645	
	WM	Laplace	Words	-2150.553	L > T: 2221
	WM	Triangular	Words	-4372.165	
	LTM	Laplace	Words	-5626.211	T > L: 182
	LTM	Triangular	Words	-5444.59	
2	WM	Laplace	Cons	-4206.855	L > T: 2007
	WM	Triangular	Cons	-6213.81	
	LTM	Laplace	Cons	-7523.019	T > L: 191
	LTM	Triangular	Cons	-7332.261	
	WM	Laplace	Words	-2832.253	L > T: 2046
	WM	Triangular	Words	-4878.06	
	LTM	Laplace	Words	-6207.731	T > L: 189
	LTM	Triangular	Words	-6018.918	
3	WM	Laplace	Cons	-5729.761	L > T: 1260
	WM	Triangular	Cons	-6989.9	
	LTM	Laplace	Cons	-8249.329	T > L: 245
	LTM	Triangular	Cons	-7975.983	
	WM	Laplace	Words	-4279.976	L > T: 1390
	WM	Triangular	Words	-5669.613	
	LTM	Laplace	Words	-6876.085	T > L: 246
	LTM	Triangular	Words	-6630.33	
4	WM	Laplace	Words	-371.0396	L > T: 730
	WM	Triangular	Words	-1101.373	
	LTM	Laplace	Words	-1607.849	T > L: 59
	LTM	Triangular	Words	-1546.519	
5	WM	Laplace	Words	-520	L > T: 962
	WM	Triangular	Words	-1483	
	LTM	Laplace	Words	-2262	T > L: 123
	LTM	Triangular	Words	-2139	