Memory & Cognition
https://doi.org/10.3758/s13421-018-0805-9

The effects of refreshing and elaboration on working memory performance, and their contributions to long-term memory formation

Lea M. Bartsch[1,2] · Henrik Singmann[1] · Klaus Oberauer[1]

© Psychonomic Society, Inc. 2018

Abstract
Refreshing and elaboration are cognitive processes assumed to underlie verbal working-memory maintenance and assumed to support long-term memory formation. Whereas refreshing refers to the attentional focussing on representations, elaboration refers to linking representations in working memory into existing semantic networks. We measured the impact of instructed refreshing and elaboration on working and long-term memory separately, and investigated to what extent both processes are distinct in their contributions to working as well as long-term memory. Compared with a no-processing baseline, immediate memory was improved by repeating the items, but not by refreshing them. There was no credible effect of elaboration on working memory, except when items were repeated at the same time. Long-term memory benefited from elaboration, but not from refreshing the words. The results replicate the long-term memory benefit for elaboration, but do not support its beneficial role for working memory. Further, refreshing preserves immediate memory, but does not improve it beyond the level achieved without any processing.

Keywords Working memory · Long-term memory · Refreshing · Elaboration · Bayesian generalized linear mixed models

Introduction

The literature on memory has long made a distinction between working memory and long-term memory (see Cowan, 2008, for a review). Working memory is understood as a system for holding a limited amount of information available for processing (Baddeley, 1986), whereas long-term memory is a system for permanently storing, managing, and retrieving information for later use with a probably unlimited capacity (Tulving, 1972). Theorists have often assumed one or several control processes that people could apply to the current contents of working memory, which are thought to help maintaining information in working memory, establish the information in long-term memory, or both. Three such control processes are being discussed: People could (a) attend to the to-be-remembered information to refresh it; (b) elaborate on it; or (c) engage in articulatory rehearsal. Our study focusses on the experimental manipulation of two of these processes, namely, refreshing and elaboration, and investigates their effects on (a) immediate memory (presumably reflecting maintenance in working memory) and (b) on delayed memory (reflecting episodic long-term memory).

Refreshing as maintenance mechanism in working memory

Refreshing is understood as briefly thinking of a stimulus just after it is no longer physically present but while its representation is still active (Johnson, Reeder, Raye, & Mitchell, 2002). The process was introduced as a general attention-based mechanism (attentional refreshing; Barrouillet & Camos, 2007; Cowan, 1999; Johnson, 1992) and is assumed to be distinct from articulatory rehearsal, which is conceptualized as the specialized mechanism for the verbal domain (Baddeley, 1986; Camos, Lagner, & Barrouillet, 2009; Hudjetz & Oberauer, 2007). Refreshing is a core component of several models of working memory: In the MEM (multiple-entry, modular) memory model, refreshing permits the

We thank Raffael Schmitt and Samuel Pawel for helping with data collection.
The data and the analysis scripts can be accessed in the Open Science Framework (https://osf.io/weuc2).

✉ Lea M. Bartsch
l.bartsch@psychologie.uzh.ch

[1] University of Zurich, Zurich, Switzerland

[2] University Research Priority Program "Dynamics of Healthy Aging", Zurich, Switzerland

Published online: 19 March 2018

reactivation of memory representations (Johnson, 1992). Similarly, Cowan (1995) states that a memory trace could be reactivated by focusing attention on it (again), before its activation is entirely lost. The time-based resource-sharing (TBRS) model proposes refreshing through attentional focusing as a mechanism for reactivating decaying memory traces (Barrouillet, Bernardin, & Camos, 2004; Barrouillet, Bernardin, Portrat, Vergauwe, & Camos, 2007). Refreshing as conceptualized in the above theories has never been observed directly but rather has been inferred from results of experiments varying the opportunity for refreshing (i.e., varying cognitive load; Barrouillet et al., 2007; Camos, Mora, & Barrouillet, 2013; Mora & Camos, 2013). The evidence for the presumed effects of refreshing is therefore less than compelling, as it does not rely on experimentally inducing the process in question. To date, only two studies (Souza and Oberauer, 2017a; Souza, Rerko, & Oberauer, 2015) experimentally induced refreshing in a (visual) working-memory task to test its effect on memory. Our study aimed at closing this gap for verbal material by experimentally manipulating refreshing and investigating its role for working memory.

The role of refreshing for episodic long-term memory

Beside its supposed role in working-memory maintenance, refreshing of information in working memory has also been argued to improve long-term memory (Johnson et al., 2002). Several studies have contrasted refreshing of a single word to repeated reading of a word, and they consistently found a benefit of refreshing on delayed item recognition (Johnson et al., 2002; Johnson, Mitchell, Raye, & Greene, 2004; Raye, Johnson, Mitchell, Greene, & Johnson, 2005; Raye, Johnson et al., 2002).

Additional, more indirect evidence for the effect of refreshing on episodic memory (EM) comes from experiments varying the available time for refreshing. For instance, a study varying cognitive load in a complex-span task found that low cognitive load—providing more free time during the maintenance interval—led to better delayed recall (Camos & Portrat, 2015). The authors attributed this to the necessary involvement of refreshing during working-memory maintenance to build up long-term memory representations. Converging evidence comes from the McCabe effect (McCabe, 2008): Words studied in complex-span tasks are recalled better in a delayed memory test than words studied in simple-span tasks. This effect is often explained by the hypothesis that the secondary task forces people to refresh items after each distraction, thereby generating better episodic retrieval cues (Loaiza, Duperreault, Rhodes, & McCabe, 2015; Loaiza & McCabe, 2012a, b; Loaiza, Rhodes, & Anglin, 2013). Contrary to this explanation, a recent study by Souza and Oberauer (2017b) showed that the long-term memory benefit of items in complex-span tasks can be fully attributed to the longer amount of free time for processing information in working memory (WM) in complex-span compared with standard, simple-span tasks.

What do people do when asked to attend to a word just encoded into WM, or when given free time to process words during maintenance? One possibility is that by focusing attention on an item, it is more deeply encoded into memory and therefore better remembered, as predicted by the levels of processing theory (Craik & Lockhart, 1972). If that is the case, the effect of refreshing would be closely related to that of elaboration.

Elaboration effects on long-term memory

Elaboration refers to processes that more deeply encode and store information for later retrieval (elaborative rehearsal; Craik and Lockhart 1972; Greene, 1987; Klatsky, 1988). Elaboration is thought to enrich the memory representation of an item by activating many aspects of its meaning and by linking it into the preexisting network of semantic associations (Craik & Tulving, 1975). It has repeatedly been shown to improve episodic long-term memory (e.g., Craik & Tulving, 1975; Gallo, Meadow, Johnson, & Foster, 2008). Research has further focussed on the use of various elaborative strategies, such as mental imagery, sentence generation, or chunking, and provided evidence for their long-term memory benefits (e.g., Dunlosky & Hertzog, 2001).

Elaboration and working memory

Only little research has focussed on the effects of elaboration on immediate memory, and the results are inconclusive. Several studies (Bailey, Dunlosky, & Kane, 2008, 2011; Dunlosky & Kane, 2007; Kaakinen & Hyönä, 2007) investigated the effect of strategy use on performance on complex-span tasks, a popular paradigm for testing working memory, and revealed a positive correlation between elaborative strategies and performance. In contrast, Morrison, Rosenbaum, Fair, and Chein (2016) found no correlation between semantic strategies and performance in a working-memory task. However, this evidence is merely correlational. The only experimental evidence for a beneficial effect of elaboration for working memory has been shown in a very specific case by Jonker and Macleod (2015). In their study, they showed that an orienting task inducing relational processing of words during encoding resulted in equivalent memory for the serial order of a study list compared with silent reading. In contrast, any other orienting task to be performed on the items, including semantic judgement of the individual items, resulted

in disruption of memory for order. Therefore, relative to other cognitive operations performed on the current contents of working memory, relational processing appears to help maintenance.

Are refreshing and elaboration distinguishable?

Conceptually, refreshing and elaboration are different processes, but it is not clear that their effects on memory are actually separable. It is conceivable that focusing attention on a representation in working memory leads to deeper and richer encoding so that the effect of refreshing on memory is mediated by elaboration. If that is the case, refreshing and elaboration should have equivalent effects on both working memory and long-term memory. To date, no research has been reported on the relationship of refreshing and elaboration; with the present study, we aimed at providing some initial insight into their relation.

Taken together, the evidence for the beneficial effects of refreshing on working memory—so far, is only indirect—and the evidence for elaboration benefits on WM is only correlational. In the following experiments, we aimed at closing that gap and experimentally controlled refreshing and elaboration through instruction. We investigated the effects of each process, and their combination, on an immediate and a delayed test of memory and investigated whether both processes show similar result patterns.

Design of experiments

As in the studies of Johnson and colleagues, we compared instructed refreshing to a repeating (re-reading) baseline during the maintenance phase of a working-memory task. In two additional conditions, we instructed participants to elaborate a subset of the items they held in memory. Elaboration logically entails attending to the words, either in memory or in the environment. When elaboration is applied to words just encoded into working memory, but no longer presented, it entails refreshing, whereas when elaboration is applied to words while they are presented, it entails (re-)reading, as in the repeat condition. Therefore, we realized two elaboration conditions: One in which words are repeated and elaborated, and one in which they are refreshed and elaborated. In this way, we can gauge the effects of elaboration on its own by comparing elaboration of repeated words to the repeating baseline. In addition, we can ask whether combining elaboration with refreshing is more effective than each of them alone.

How can we measure the effect of refreshing in our paradigm? The Johnson et al. studies—testing the effect of refreshing on EM—used repeat as the baseline, and therefore we follow their precedent for assessing the effect of refreshing on EM. For assessing the effect of refreshing on WM, the repeat condition is probably not a suitable baseline because it provides a second chance for encoding the word into WM. Therefore, we assess the effect of refreshing against two baselines: The one used in Souza et al. (2015; i.e., comparison within the memory set between items refreshed more vs. less) and a comparison of refreshed items to a no-processing baseline. For the first comparison, we compare the items that were processed in refreshing trials to the items within the same trial that were not further processed after initial encoding. The second comparison, against a no-processing baseline, is important to assess whether refreshing actually leads to an improvement of memory after encoding. This cannot be assessed with the first comparison.

Methods

In the two experiments presented here, we asked participants to remember six nouns in serial order. After list presentation, either the first three words or the last three words were to be processed again in one of four ways, depending on the experimental condition. During encoding it was not predictable which half of the items would have to be processed. In the repeat-without-elaboration condition, the three words appeared again sequentially on the screen, and the subjects had to simply re-read them silently. In the refresh-without-elaboration condition, the to-be-processed words were replaced by refreshing prompts appearing at the same location. The subjects were instructed to "think of" the corresponding words as soon as the prompts were shown. In the repeat-with-elaboration condition, the three to-be-processed words were shown again sequentially on the screen, and subjects were instructed to generate a vivid mental image of the three objects interacting. The stimuli appearing on the screen in that condition did not differ from the repeat-without-elaboration condition, leaving the instruction to form a vivid mental image as the only difference between these conditions. Finally, in the combined refresh-with-elaboration condition, the participants had to "think of" the words replaced by the prompts and additionally form a vivid mental image of those items. Again, the event sequence of this condition does not differ from the refresh-without-elaboration condition apart from the instruction to form a mental image. The experiments used a 2 × 2 × 2 (repeat/refresh [repeat, refresh] × elaboration [with elaboration, without elaboration] × processing [processed triplet, not-processed triplet]) within-subject design. Orienting the processing task to only a subset of the words in memory allows us to draw inferences about the effect of each of the processing conditions on memory by comparing the memory performance of the triplet of words that was not further

processed after encoding to the triplet that was processed according to one of the four experimental conditions.

Experiment 1

Participants

We recruited 20 students from the University of Zurich as participants (10 female) at the age of 19–28 years (M = 22.11 years, SD = 5.32). They were compensated with either 15 Swiss Francs (about 15 USD) or partial course credit for the 1-hour experiment.

Materials and procedure

The stimuli were nouns randomly drawn from a pool of 863 German abstract and concrete nouns for each subject. The nouns were between two and 15 letters long and had a mean normalised lemma frequency of 30.81/million (drawn from the dlexdb.de lexical database).

The sequence of an experimental trial is illustrated in Fig. 1. The six to-be-remembered words in each trial were sequentially presented in boxes from top to bottom on the screen, each for 500 ms. Depending on the experimental condition, a cue was presented 1,000 ms after the last memory item, indicating whether the first half or the second half of the list had to be processed again. In the repeat-with-elaboration and repeat-without-elaboration conditions, the to-be-processed first word of the cued triplet was shown again in the same box as during encoding for 1,400 ms, followed by a 600 ms interstimulus interval (ISI). The respective second and third word of that triplet followed at the same pace successively in their corresponding boxes. In the refresh-without-elaboration and refresh-with-elaboration conditions, each to-be-processed word of a triplet was replaced by a refreshing prompt (#?#?#) in its corresponding box, and participants were instructed to "think of" the word in that box. In the repeat-with-elaboration and refresh-with-elaboration conditions, participants were additionally instructed to form a vivid mental image of the three words interacting with each other.[1]

After processing the words in the cued triplet, participants' memory for each list item was tested in their order of presentation using a four-alternative forced-choice procedure. For that purpose, four words were presented from which the subject could choose the correct word in the currently tested list position with a button press. All tests trials had the following four response options: the target (i.e., correct) word, one lure from the same triplet of words within the present list, one lure from the other triplet of the present list, and one new word. This choice had to be made for each of the serial positions successively. We applied this four-alternatives forced-choice recognition task in order to test both memory for items (i.e., discriminating between items that have been presented in the current memory list and new items) and for serial order (i.e., discriminating between the item in the tested position and other list items).

Within each block of four trials, the same type of processing was instructed throughout, and a screen repeating the instructions of the particular condition was shown prior to the beginning of each block. The order of the condition blocks was randomized between subjects. After a total of 16 blocks, with four blocks of each condition, the participants performed a mental arithmetic task for 2 minutes, where they had to judge the correctness of multiplication statements (e.g., 3 × 6 = 18). After this distractor task, we assessed participants' long-term memory for the words they had encoded into working memory throughout the experiment. To this end, we presented in each trial the first word of one word triplet in one memory list, and asked participants to choose the word that had followed the given word within the same triplet from four different options. These included the correct word (i.e., which could be either the word in the second or third position of the target triplet for the first prompt and the fifth or sixth word for the second prompt), a word from the other triplet of the same list, a word from another list, and a new word. This allowed us to keep the format of delayed recognition very similar to the immediate test, and furthermore to compare in each trial the memory performance for processed to not-processed items. As for the immediate test, the delayed test provided information about both item memory (i.e., which words have been presented in the experiment) and relational memory (i.e., which words have been together in a triplet). The participants were made aware of the delayed memory test before the start of the experiment.

Experiment 2

Experiment 2 was identical to Experiment 1, but added two further conditions without any instructed processing. These conditions served as baselines against which we could gauge whether processing a subset of a memory list (according to the instructed process) improves or impairs memory for the processed triplet, and memory for the not-processed triplet, relative to a standard test of immediate memory. In the short baseline condition, recognition followed directly 1,000 ms after the initial presentation of the list of six words and allowed us to measure the level of memory directly after encoding. In the long baseline, a blank screen interval was inserted after list

[1] The timing parameters where chosen based on a pilot experiment, which allowed participants to process the items in each of the four experimental conditions in a self-paced mode. The mean processing times (PT) where PT = 1,419 ms in the repeat-without-elaboration condition, PT = 1,491 ms in the repeat-with-elaboration condition, PT = 1,197 ms in the refreshing-without-elaboration, and PT = 1,198 ms in the refreshing-with-elaboration condition.

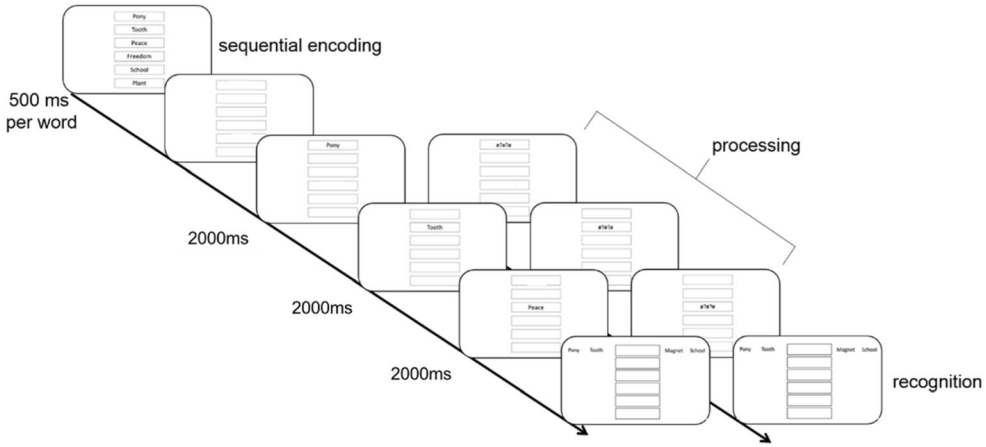

Fig. 1 Working memory paradigm of Experiment 1. Subjects were shown a list of six words sequentially, followed by either the first or second triplet being processed according to the four experimental conditions

presentation for the same amount of time (6 seconds) that the processing steps in the four experimental conditions took, before participants' memory was tested. The long baseline condition allowed us to investigate the impact of time on immediate memory without any processing manipulation, so participants were free to use it for any process on the memory items they might find helpful, or do nothing.

Participants

For Experiment 2, we recruited 30 students from Zurich University (21 female), ages 19–28 years (M = 23.82 years, SD = 3.82). They were compensated with either 15 Swiss Francs (about 15 USD) or partial course credit for the 1-hour experiment.

Materials and procedures

The six conditions (the four experimental conditions of Experiment 1 and the two baseline conditions) were implemented within condition-pure mini blocks of four trials, resulting in 12 trials per condition throughout the whole experiment. Everything apart from adding the two baseline conditions and reducing the number of trials per condition was held constant between Experiment 1 and Experiment 2.

Results

To draw inferences about the effect of refreshing and elaboration on working memory as well as these processes' impact on long-term memory formation, we first focus on the results of the four processing conditions of Experiments 1 and 2. We subsequently evaluate these results in comparison with the baseline conditions of Experiment 2. All data and analysis scripts can be assessed on the Open Science Framework (https://osf.io/weuc2).

Data analysis

We analysed Experiments 1 and 2 jointly using a Bayesian generalized linear mixed model (BGLMM) implemented in the R package rstanarm (Stan Development Team, 2018). The dependent variable was the number of correct and incorrect responses in each cell of the design per participant. Correct responses were defined as choosing the target item from the four alternatives. Therefore, we assumed a binomial data distribution predicted by a linear model through a probit link function (i.e., a repeated-measures probit regression). The fixed effects were processing (processed vs. not-processed triplet), repeat/refresh (repeated vs. refreshed items), elaboration (with vs. without elaboration instruction), and all their interactions. Following the recommendation of Barr and colleagues (Barr, Levy, Scheepers, & Tily, 2013; see also Schielzeth & Forstmeier, 2009), we implemented the maximal random-effects structure justified by the design; by-participant random intercepts and by-participant random slopes for all fixed effects (as all factors were within-subject factors). In addition, we estimated the correlation among the random-effects parameters. As factor coding, we used the orthonormal contrasts described in Rouder, Morey, Speckman, and Province (2012; section 7.2) that guarantee that priors affect all factor levels equally. For factors with two levels as employed here this corresponds to contrasts with values of $\sqrt{2}/2$ and $-\sqrt{2}/2$.

Following Gelman et al. (2013), the regression coefficients were given weakly informative Cauchy priors, with location 0 and scale 5. We used completely noninformative priors for the correlation matrices, so-called LKJ priors, with shape parameter 1 (Stan Development Team, 2017). Bayesian procedures provide posterior probability distributions of the model parameters (i.e., the regression weights) that express uncertainty about the estimated parameters. The highest density regions (HDRs) of these posteriors can be used for statistical

inference. A 95% HDR represents the range in which the true value of a parameter lies with probability 0.95, given model and data (Morey, Hoekstra, Rouder, Lee, & Wagenmakers, 2016). If zero lies outside the Bayesian HDR, then there is strong evidence for the existence of the corresponding effect; although the strength of evidence varies continuously, for simplicity we will describe effects as "credible" if their HDRs exclude zero. We used an MCMC algorithm (implemented in Stan; Carpenter et al., 2017) that estimated the posteriors by sampling parameter values proportional to the product of prior and likelihood. These samples are generated through four independent Markov chains, with 1,000 warm-up samples each, followed by 1,000 samples drawn from the posterior distribution which were retained for analysis. Following Gelman et al. (2013), we confirmed that the four chains converged to the same posterior distribution by verifying that the $\hat{R}$ statistic—reflecting the ratio of between-chain variance to within-chain variance—was <1.02 for all parameters, and we visually inspected the chains for convergence.

For analysing the baseline conditions in Experiment 2, we estimated a second binomial BGLMM with two fixed-effects factors and their interaction on the same dependent variable. We again estimated the maximal random-effects structure and the correlation among the random effects parameters. The first factor was processing with two levels: processed or long baseline versus not-processed or short baseline. This factor groups the conditions in which processing of words during maintenance is instructed or at least enabled, and the conditions providing no opportunity for such processing. The second factor combined repeat/refresh, elaboration, and baseline into a single grouping factor with five levels: repeat with elaboration, repeat without elaboration, refresh with elaboration, refresh without elaboration, and baseline. Note that we will focus here on the pairwise comparisons of the baseline conditions against the other factor levels. To test these pairwise comparisons, we calculated difference distributions (i.e., posterior distributions of the differences of parameter values between two conditions) and report their 95% HDRs. Again, if this HDR does not contain zero, then this constitutes evidence for a highly credible difference. Again, we verified that the $\hat{R}$ statistic was <1.01, and visually inspected the chains for convergence.

Results

Figure 2 shows the estimated proportion of correct responses and their corresponding 95% highest posterior density regions for the immediate and delayed memory data from the core design shared by both experiments. The posterior effect estimates are presented in Tables 1 and 2. As the BGLMM for the experimental conditions of Experiment 1 and 2 revealed evidence for the same pattern of results, we combined the data in a single BGLMM, which is presented in the following.[2] We also ran a BGLMM including word concreteness as a factor, but that factor had only a main effect, without entering into any interactions, and therefore we present the simpler model without concreteness.

A first question was whether our manipulation of processing half of a memory list had an effect on memory. The analysis for immediate memory supported an effect of our manipulation, as it showed a credible main effect of processing, implying that participants had better memory for items that were processed again after initial encoding than for items from the not-processed triplets (see Table 1 and Fig. 2, upper panel). This was also true for delayed memory: Items that were processed again after encoding were better remembered than not-processed items (see Table 2 and Fig. 2, lower panel).

In what follows, we first report effects of refreshing and of elaboration on immediate memory, followed by the effects of these two processing manipulations on delayed memory. In each section, we first focus on the effects of each processing manipulation within the core design, drawing on the joint analysis of both experiments. Next, we ask how each experimental condition of the core design compares to the baseline conditions of Experiment 2. We compared the immediate memory performance in each processing condition of Experiment 2 to the baselines using pairwise comparisons of the difference distributions taken from the second BLGMM described above. Figure 3 shows the estimated response probabilities in the immediate and delayed memory test in the six conditions of Experiment 2. The analysis revealed that there was no difference between the two baselines, indicating that free time after encoding had no effect on memory (neither immediate, $\Delta = 0.02$, 95% HDR = [−0.03, 0.07], nor delayed, $\Delta = 0.04$, 95% HDR = [−0.01, 0.10]).[3] We therefore pooled the baselines for all following comparisons to the processing conditions reported below.[4]

Working memory

Refreshing effects on working memory We first tested how the effect of refreshing a subset of words in working memory compares to the effect of repeated reading of these words. This is the comparison through which Johnson and colleagues evaluated the effect of refreshing on delayed memory

[2] We have also analysed the accuracy data with a standard mixed ANOVA. The pattern of significant and nonsignificant effects matched that of credible and noncredible effects reported for the BGLMM. Note, however, that analysing accuracy data with a linear model (such as ANOVA) is highly discouraged and can lead to spurious results (e.g., Jaeger, 2008).

[3] For all pairwise comparisons reported in the text, results are on the probability scale (i.e., $\Delta = 0.02$ corresponds to an effect of 2%).

[4] In other words, all comparisons against the baseline reported below are based on the second BGLMM. To obtain the results, we averaged the posterior samples of the short and the long baseline and used this distribution for the calculation of the difference distributions for each pairwise comparison.

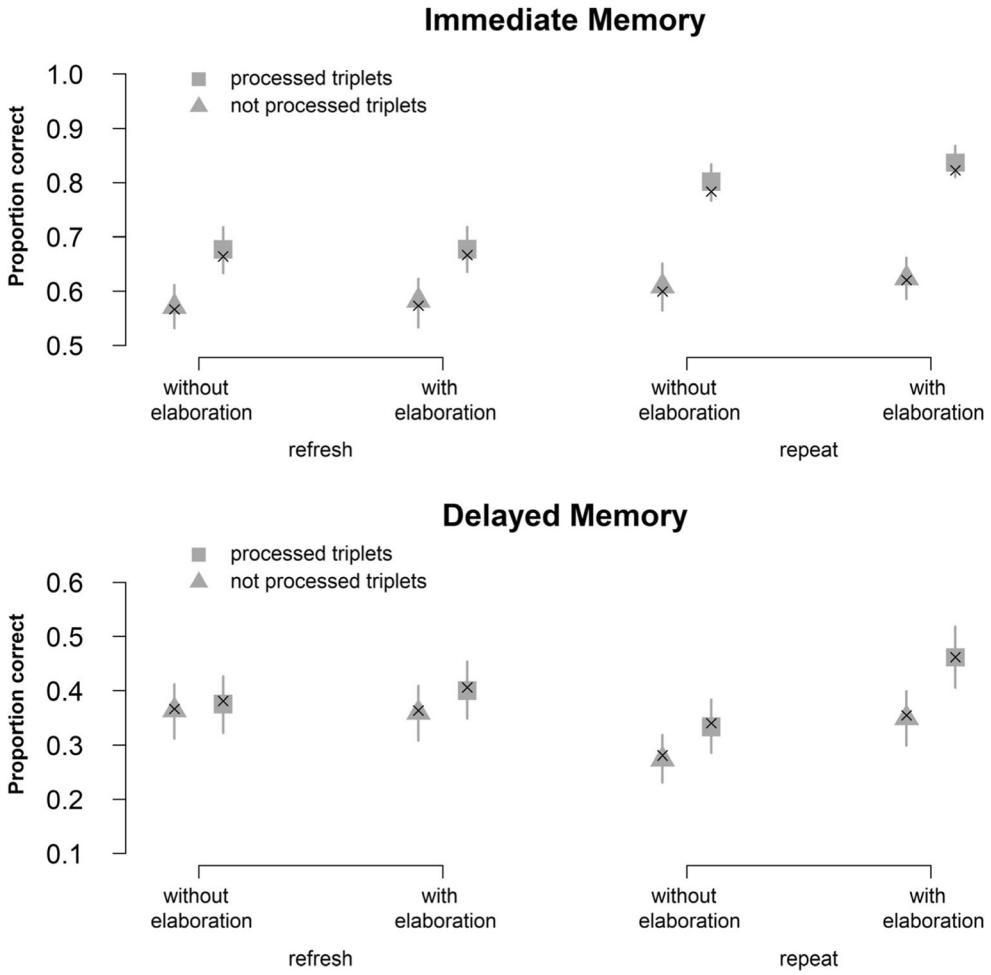

Fig. 2 Proportion correct of the combined data from Experiments 1 and 2 in the immediate (WM, upper graph) and delayed (LTM, lower graph) memory task. Grey symbols and error bars represent estimated proportions and their 95% HDRs from the BGLMM. Crosses represent observed proportions. Overlap indicates the model adequately describes the data

(Johnson et al., 2002; Raye et al., 2007). The BGLMM shows a credible main effect of repeat/refresh (see Table 1) but in the direction of an overall advantage for trials with a repeated triplet over trials with a refreshed triplet. Thus, the direction of this effect is in opposite direction to what was observed by Johnson et al. for delayed memory. This main effect was further qualified by the two-way interaction of processing and repeat/refresh, indicating that repeated words benefited more

Table 1 Posterior effect estimates and their 95 % HDRs of the generalized linear mixed model for binomial response variables for the immediate serial memory data of Experiments 1 and 2

	Mean parameter on probit scale	95% HDR
(Intercept)	**0.46**	**[0.37, 0.56]**
Processing	**0.32**	**[0.27, 0.37]**
Repeat/refresh	**−0.20**	**[−0.26, −0.15]**
Elaboration	−0.04	[−0.09, 0.01]
Processing × Repeat/refresh	**−0.18**	**[−0.23, −0.13]**
Processing × Elaboration	−0.02	[−0.07, 0.03]
Repeat/refresh × Elaboration	0.04	[−0.02, 0.10]
Processing × Repeat/refresh × Elaboration	0.04	[−0.02, 0.11]

Note. Credible effects, defined as HDRs excluding zero, are printed in boldface. HDR = highest density region

Table 2 Posterior effect estimates and their 95 % HDRs of the generalized linear mixed model for binomial response variables for the delayed memory data of Experiments 1 and 2

	Mean parameter on probit scale	95% HDR
(Intercept)	**−0.35**	**[−0.42, −0.27]**
Processing	**0.11**	**[0.06, 0.16]**
Repeat/refresh	0.04	[−0.01, 0.10]
Elaboration	**−0.10**	**[−0.16, −0.05]**
Processing × Repeat/refresh	−0.07	[−0.15, 0.00]
Processing × Elaboration	−0.05	[−0.12, 0.03]
Repeat/refresh × Elaboration	**0.12**	**[0.05, 0.19]**
Processing × Repeat/refresh × Elaboration	0.02	[−0.08, 0.11]

Note. Credible effects, defined as HDRs excluding zero, are printed in boldface. HDR = highest density regions

from being processed again than refreshed words did. Nevertheless, the effect of processing appeared for both, repeated words (Δ = 0.21, 95% HDR = [0.18, 0.24]) and refreshed words (Δ = 0.10, 95% HDR = [0.07, 0.13]).

Next, we compared the effect of guided refreshing to the baseline level of memory (i.e., pooled short and long baseline), separately for processed and not processed words. Whereas the processed words in the repeat condition were remembered better than the words in the baseline conditions (Δ = 0.18, 95% HDR = [0.14, 0.21]), the processed words in the refresh condition were not—memory for refreshed words was about equal to memory in the baseline conditions (Δ = 0.03, 95% HDR = [−0.005, 0.06]). Further, the comparison of the not-processed triplets in the refreshing condition to the baseline suggests that refreshing a subset of items in working memory harms the memory for the remaining items (i.e., the

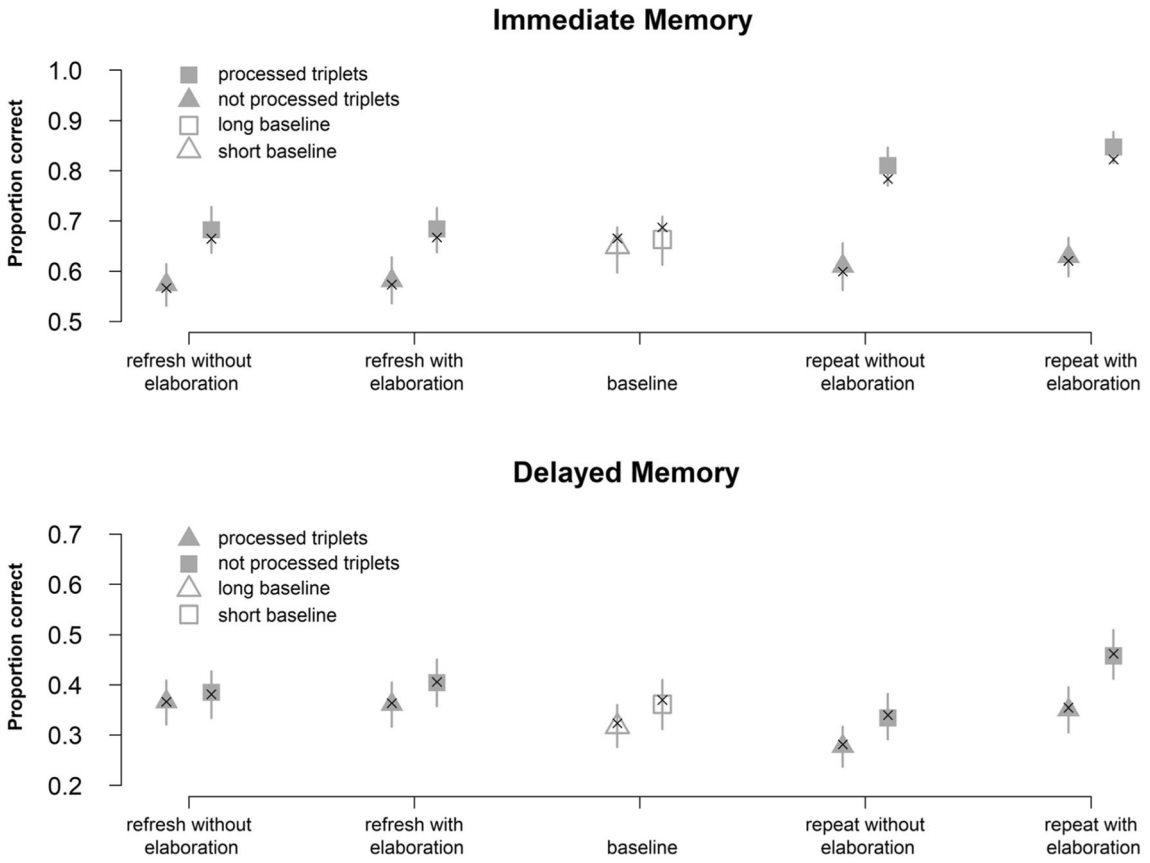

Fig. 3 Proportion correct of the immediate serial (WM, upper graph) and delayed cued (LTM, lower graph) memory performance of all conditions of Experiment 2. Grey symbols and error bars in the background represent estimated proportions and 95% HDRs. Crosses represent observed proportions. Overlap indicates the model adequately describes the data

not-processed triplet; Δ = −0.08, 95% HDR = [−0.11, −0.04]). This did not appear to be the case for the not-processed triplet in the repeat condition (Δ = −0.03, 95% HDR = [−0.07, 0.00]). To summarize, whereas repeating a subset of words in working memory boosts their availability in memory above a no-processing baseline, refreshing merely maintains memory at the baseline level, while at the same time not-refreshed words drop below the baseline level.

Effects of elaboration on working memory In the BGLMM for the core design of both experiments, there was no credible evidence for a main effect of elaboration on working memory performance, or for any of the interactions involving elaboration (see Table 1 and Fig. 2). However, the result is not completely clear-cut. For the main effect of elaboration, 94% of the posterior mass of the effect was in the direction of better memory in the elaboration than the no-elaboration conditions. For the interaction of repeat/refresh with elaboration, 97% of the posterior indicated better memory in the repeat with elaboration than in the repeat without elaboration condition. For the three-way interaction of processing, repeat/refresh, and elaboration, 78% of the posterior indicated that the processed words in the repeat-with-elaboration condition tended to be recalled better than words in all other conditions (see Fig. 2). Note that the direction of these effects was the same as the ones observed for delayed memory (compare Table 1 to Table 2), but appeared to be somewhat smaller.

Notwithstanding the absence of credible effects involving elaboration, we next compared memory for words in the elaboration conditions, as well as memory for words in the conditions without elaboration, separately against the (pooled) baseline in Experiment 2. Words in the elaboration conditions were better remembered than the baseline (Δ = 0.04, 95% HDR = [0.01, 0.07]). Again, this effect was mainly driven by the words in the repeat-with-elaboration condition (see Fig. 3). In contrast, the words in the conditions without elaboration showed no credible difference to the baseline (Δ = 0.03, 95% HDR = [−0.01, 0.05]). In summary, the results do not provide convincing evidence for an effect of elaboration on working memory. If anything, the effect was small and confined to the condition in which the words were repeated.

Long-term memory

Effects of refreshing on long-term memory The joint BGLMM of both experiments (see Table 2) revealed no evidence for a main effect of repeat/refresh on delayed memory performance, and also no evidence for an interaction of processing with repeat/refresh. Hence, contrary to the findings of Johnson and colleagues, refreshing did not lead to better long-term memory than repeated reading. We nevertheless compared words in the refresh and repeat conditions separately against the baseline. The comparisons with the (pooled) baseline conditions indicates that the processed triplets of the refreshing without elaboration condition were remembered at about the same level as the baselines (Δ = 0.04, 95% HDR = [−0.01, 0.09]), as was memory for the processed words of the repeat without elaboration (Δ = 0.00, 95% HDR = [−0.05, 0.05]). Note that the above pattern of results also holds for a lenient score of performance in the delayed memory task, counting all responses showing correct item memory (i.e., the target, same-list items, and other-list items) as correct responses. To summarize, our results provide no evidence for an effect of refreshing on long-term memory.

Long-term memory effects from elaboration The analysis of the delayed memory data revealed evidence for a main effect of elaboration, and an interaction of repeat/refresh with elaboration. Follow-up analyses of the interaction revealed that the elaboration effect appeared for words in the repeat conditions (repeated with elaboration vs. repeat without elaboration: Δ = 0.10, 95% HDR = [0.06, 0.14]), but not in the refresh conditions (refresh with elaboration vs. refresh without elaboration: Δ = 0.01, 95% HDR = [−0.03, 0.05]). In sum, memory was better for trials with instructed elaboration, but only when elaboration was accompanied by repetition. Furthermore, in Experiment 2, the triplets in the repeat-with-elaboration condition (Δ = 0.07, 95% HDR = [0.02, 0.11]) showed better memory than the (pooled) baseline. Likewise, the triplets in the refreshing with elaboration condition showed better memory than the baseline (Δ = 0.05, 95% HDR = [0.004, 0.08]). Together, this suggests a beneficial effect of elaboration for long-term memory compared with conditions without any processing instruction, or without time for processing. None of the other differences to the (pooled) baseline were credible. The above evidence speaks for a beneficial effect of elaboration on long-term memory.

Discussion

Refreshing and elaboration have been proposed as control processes on the contents of working memory, potentially serving to maintain them in working memory and to lay the ground for successful long-term memory formation. Our main goal was to investigate the impact of experimentally induced refreshing and elaboration on working memory and episodic long-term memory. Comparing the effect patterns of refreshing and elaboration should help to answer the question whether the two processes are distinguishable. In the following, we will first discuss the effects of both processes on long-term memory and second their effects on working memory.

How do elaboration and refreshing affect long-term memory?

Long-term memory retrieval was highest for repeated items that were elaborated, confirming that participants followed the elaboration instruction, and replicating the long-term memory benefit of elaboration. Furthermore, elaboration of repeated words resulted in long-term memory benefits compared with the level of memory directly after encoding (short baseline) and compared with the level after additional free time after encoding (long baseline), demonstrating that instructed elaboration added something over and above people's spontaneous encoding behaviour.

Refreshing showed no effects on long-term memory. We need to ask why, in contrast to the present study, previous studies reported long-term memory benefits for refreshing (Johnson et al., 2002; Johnson et al., 2004; Raye, Johnson, Mitchell, Greene, & Johnson, 2007; Raye, Johnson, Mitchell, Reeder, & Greene, 2002). Some methodological differences might have contributed to the discrepancy between our finding and those of Johnson and colleagues (Johnson et al., 2002; Raye et al., 2002). First, Johnson and colleagues tested episodic long-term memory through a standard yes–no item-recognition test, whereas we used a four-alternative forced-choice test tapping both item and relational memory. We do not think that this explains the different outcomes because we also found no effect of refreshing on a lenient score for delayed memory that reflected only item memory.

Second, the present study asked participants to refresh three items in working memory, whereas previous studies included refreshing memory sets of only one to two items per trial. Therefore, it is possible that refreshing benefits on long-term memory occur only when working memory load is low, so that people have a good chance to remember the words when they are asked to refresh them. Against that possibility speaks the study by Souza et al. (2015), which showed (short-term) memory refreshing benefits with six-colour arrays. Future research has to clarify whether refreshing verbal material only benefits long-term memory when the load on working memory is low.

A third discrepancy between the present study and previous research is that the present memory test was not incidental but well expected by the participants. The mere knowledge about a delayed memory test might have changed the participants' behaviour in our experiment. Furthermore, previous research by Johnson and colleagues never included an immediate memory test, as was the case here. Thus, we might have created a testing effect for all items, whereas Johnson et al.'s method might have created a testing effect only for the refreshed item. The instruction of previous studies to "think of" included also the instruction to recall the refreshed item aloud, which confounded refreshing with recall. The increased long-term memory for refreshed items in previous studies therefore could have been due to recall—in line with the testing effect (Roediger & Karpicke, 2006)—rather than due to refreshing per se. One possible conclusion from our study is that the refreshing benefit observed by Johnson et al. can only be observed for long-term memory with their experimental design and is not universal. At least, our study demonstrates a clear boundary condition of said refreshing benefit.

Does elaboration benefit working memory?

As elaboration benefited long-term memory, in this as well as in previous studies, we were interested also in its immediate effects on memory. Despite its clear advantage at the delayed test, there was no compelling evidence for an effect of elaboration on immediate memory. If anything, elaboration tended to be helpful only when the to-be-remembered words were presented again in the retention interval—which is typically not the case in tests of working memory. Our findings therefore fail to provide experimental support for the conclusion from previous correlational studies, which found that higher performance on complex-span tasks was related to individuals' use of so-called normatively effective elaboration strategies such as imagery and sentence generation (Bailey et al., 2008, 2011; Bailey et al., 2009; Dunlosky & Kane, 2007). Reasons for this could be due to the present study using a simple-span paradigm and previous research relying on complex-span tasks. Bailey et al. (2011), following Unsworth and Engle (2007), have argued that in complex span, retrieval comes mostly from "secondary memory" (i.e., episodic LTM), and they have provided initial evidence that self-reported elaboration predicted performance only on those trials of a simple-span test that also involved retrieval from episodic LTM. Therefore, our findings could be reconciled with those of previous strategy self-report studies by assuming that elaboration improves only episodic LTM. An alternative explanation, of course, would be the reverse causality: If participants have good memory, they have more information in memory to elaborate on. They would subsequently also show better LTM.

How does refreshing affect working memory?

When a subset of the words in working memory is refreshed, the refreshed words are remembered better in an immediate test than the words from the not-refreshed subset. Therefore, refreshing is effective in prioritizing the refreshed representations within working memory. This effect of refreshing, however, is weaker than that of repeating the words. Our results from Experiment 2 further show that, in contrast to repeating, refreshing a subset of items in working memory did nothing to improve memory for the refreshed items relative to the baselines without any instructed processing. At the same time, the remaining items (i.e., the not-processed triplet) were

remembered worse compared with the baselines. In sum, refreshing a triplet of words in working memory has no beneficial effect on immediate memory beyond maintaining performance at the level that is achieved when tested immediately after encoding. At the same time, refreshing some items negatively affects the remaining items in memory.

Our study was strongly inspired by a recent direct experimental manipulation of refreshing (Souza et al., 2015), which showed refreshing benefits for visual material in working memory. So, why did we not find a refreshing benefit above baseline level? One explanation could be that in our study items were only refreshed once. Souza et al. found that the beneficial effect of refreshing increases with its frequency: The more refreshing opportunities an item receives, the higher the probability of recalling that item. Nevertheless, the discrepancy between our results and those of Souza et al. (2015) is not that big: The Souza et al. refreshing-frequency effect is a comparison between memory for items receiving zero, one, and two refreshings within a memory set. This comparison is analogous to that between processed vs. not-processed triplets in our study, for which we obtained a credible difference: Within the refreshing condition, processed triplets (refreshed once) were recalled better than not-processed triplets. One difference still remains: In Souza et al.'s study, the refreshed items were recalled *better* than the baseline conditions without any processing instruction, which was not the case in our Experiment 2. This difference between baseline and refreshing conditions, however, was not found in a more recent study of guided refreshing of visual memory items (Souza & Oberauer, 2017a). Taken together, the previous studies and the present experiments converge on the following conclusion: Refreshed items in working memory are remembered better than not-refreshed items within the same trial, but when these refreshed items are compared with baseline performance in separate trials without any refreshing instruction, refreshing yields no beneficial effect.

Apart from the Souza et al. studies discussed above, most studies on the effect of refreshing on WM performance have used complex-span tasks. One could argue that the lack of a beneficial effect of refreshing in our experiments was because we did not use a complex-span task. However, there is no theoretical reason why refreshing should be limited to complex-span paradigms. One effect often attributed to refreshing, the McCabe effect, has recently been shown to arise from opportunities to refresh in both simple-span and complex-span tasks (Souza & Oberauer, 2017b). We therefore see no reason why our findings should not generalize to other paradigms, including complex span.

Our results are consistent with the idea that refreshing maintains representations in working memory, whereas not-refreshed representations are forgotten over time. On this assumption, refreshing can at best be expected to keep performance at the level obtained at an immediate test (i.e., the short baseline of Experiment 2). This was the case for the refreshed triplets, whereas memory for the not-refreshed triplets dropped below that level. During a retention interval without instructed processing (i.e., the long baseline of Experiment 2), participants can be assumed to refresh all six words. One might wonder why concentrating the entire time on refreshing a subset of three words did not lead to better memory for these three words than distributing the same refreshing time over all six words. Whereas our data showed a trend in that direction (compare the refreshed triplets to the long baseline in Fig. 3), there was no statistical evidence supporting it. This could be explained by assuming that refreshing is so efficient that it can maintain the memory strength of all six words, and concentrating refreshing on a subset of three items means to refresh them more than necessary, without appreciable further benefit.

Our results are also consistent with the alternative view that refreshing a subset of items prioritizes these items in working memory over the not-refreshed ones. Prioritization could mean that these items are strengthened, but nothing happens to the not-refreshed items. At test, when items compete for retrieval, the not-refreshed items tend to be blocked by their stronger competitors. Again, one might ask why selectively strengthening a subset of items does not improve memory for them above the baseline conditions. Again, the answer could be that strengthening has diminishing returns, so that the benefit of strengthening some items beyond their initial level is smaller than the cost for the remaining items.

Are refreshing and elaboration distinct processes?

If refreshing and elaboration are the same, then the pattern of effects of both processes on working memory and on long-term memory should be the same. The effects on working memory are not informative in this regard: If elaboration were different from refreshing, we could expect it to add something to memory performance over and above mere refreshing—it did not. This could mean that when asked to refresh a set of items, people already elaborate them, so that additionally instructing them to elaborate the items makes no difference. However, elaboration also did not add much to mere re-reading in the repeat condition, either. Therefore, it could also be that refreshing and elaboration are different processes, but elaboration simply has no effect on immediate memory.

Evidence for a distinction of refreshing and elaboration lies in the long-term memory results: Compared with mere re-reading the words in the repeat condition, elaboration benefited memory after a delay, whereas refreshing did not. Likewise, in comparison to the baseline conditions of Experiment 2, elaboration improved delayed memory whereas refreshing did not. It is particularly telling that elaboration had this beneficial effect only in conjunction with repeating, but not in conjunction with refreshing. Apparently, the need to attend to a set of words in working memory—as opposed to in

the environment—undermines the effectiveness of elaboration. This is not what we would expect if refreshing and elaboration were two terms for the same process.

Conclusion

We replicated the beneficial effect of elaboration on long-term memory. Contrary to previous findings, we found no such effect for refreshing, suggesting that refreshing and elaboration are not the same process. Further, neither refreshing nor elaboration did much to improve working-memory performance, compared with no processing of the memory representations after encoding. Whatever the working-memory system does spontaneously—whether engaging in some maintenance activity, or doing nothing—appears to be hard to beat by any experimentally induced process.

References

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

Bailey, H., Dunlosky, J., & Hertzog, C. (2009). Does differential strategy use account for age-related deficits in working-memory performance? *Psychology and Aging*, *24*(1), 82–92. https://doi.org/10.1037/a0014078

Bailey, H., Dunlosky, J., & Kane, M. J. (2008). Why does working memory span predict complex cognition? Testing the strategy affordance hypothesis. *Memory & Cognition*, *36*(8), 1383–1390. https://doi.org/10.3758/MC.36.8.1383

Bailey, H., Dunlosky, J., & Kane, M. J. (2011). Contribution of strategy use to performance on complex and simple span tasks. *Memory & Cognition*, *39*(3), 447–461. https://doi.org/10.3758/s13421-010-0034-3

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), 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time Constraints and Resource Sharing in Adults' Working Memory Spans. *Journal of Experimental Psychology: General*, *133*(1), 83–100. https://doi.org/10.1037/0096-3445.133.1.83

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*(3), 570–585. https://doi.org/10.1037/0278-7393.33.3.570

Barrouillet, P., & Camos, V. (2007). The time-based resource-sharing model of working memory. In N. Osaka, R. H. Logie, & M. D'Esposito (Eds.), *The cognitive neuroscience of working memory* (pp. 59–80). https://doi.org/10.1093/acprof:oso/9780198570394.003.0004

Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language*, *61*(3), 457–469. https://doi.org/10.1016/j.jml.2009.06.002

Camos, V., Mora, G., & Barrouillet, P. (2013). Phonological similarity effect in complex span task. *The Quarterly Journal of Experimental Psychology*, *66*(10), 1927–1950. https://doi.org/10.1080/17470218.2013.768275

Camos, V., & Portrat, S. (2015). The impact of cognitive load on delayed recall. *Psychonomic Bulletin & Review*, *22*(4), 1029–1034. https://doi.org/10.3758/s13423-014-0772-5

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M.,. .. Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York, NY, Oxford University Press.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). https://doi.org/10.1017/CBO9781139174909

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

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268–294. https://doi.org/10.1037//0096-3445.104.3.268

Dunlosky, J., & Hertzog, C. (2001). Measuring strategy production during associative learning: The relative utility of concurrent versus retrospective reports. *Memory & Cognition*, *29*(2), 247–253. https://doi.org/10.3758/BF03194918

Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *The Quarterly Journal of Experimental Psychology*, *60*(9), 1227–1245. https://doi.org/10.1080/17470210600926075

Gallo, D. A., Meadow, N. G., Johnson, E. L., & Foster, K. T. (2008). Deep levels of processing elicit a distinctiveness heuristic: Evidence from the criterial recollection task. *Journal of Memory and Language*, *58*(4), 1095–1111. https://doi.org/10.1016/j.jml.2007.12.001

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian data analysis (3rd ed.). New York, NY: Chapman & Hall/CRC.

Greene, R. L. (1987). Effects of maintenance rehearsal on human memory. *Psychological Bulletin*, *102*(3), 403–413. https://doi.org/10.1037/0033-2909.102.3.403

Hudjetz, A., & Oberauer, K. (2007). The effects of processing time and processing rate on forgetting in working memory: Testing four models of the complex span paradigm. *Memory & Cognition*, *35*(7), 1675–1684. https://doi.org/10.3758/BF03193501

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434–446. https://doi.org/10.1016/j.jml.2007.11.007

Johnson, M. K. (1992). MEM: Mechanisms of recollection. *Journal of Cognitive Neuroscience*, *4*(3), 268–280.

Johnson, M. K., Mitchell, K. J., Raye, C. L., & Greene, E. J. (2004). An age-related deficit in prefrontal cortical function associated with refreshing information. *Psychological Science*, *15*(2), 127–32. https://doi.org/10.1016/j.neuroscience.2008.09.014

Johnson, M. K., Reeder, J. A., Raye, C. L., & Mitchell, K. J. (2002). Second thoughts versus second looks: An age-related deficit in reflectively refreshing just-activated information. *Psychological Science: A Journal of the American Psychological Society/APS*, *13*(1), 64–67. https://doi.org/10.1111/1467-9280.00411

Jonker, T. R., & Macleod, C. M. (2015). Disruption of relational processing underlies poor memory for order. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, *41*(3), 831–840. https://doi.org/10.1037/xlm0000069

Kaakinen, J. K., & Hyönä, J. (2007). Strategy use in the reading span test: An analysis of eye movements and reported encoding strategies. *Memory*, *15*(6), 634–646. https://doi.org/10.1080/09658210701457096

Klatsky, R. (1988). Theories of information processing and theories of aging. In L. Light & D. M. Burke (Eds.), *Language, memory, and aging* (pp. 1–16). New York, NY: Cambridge University Press.

Loaiza, V. M., Duperreault, K. A., Rhodes, M. G., & McCabe, D. P. (2015). Long-term semantic representations moderate the effect of attentional refreshing on episodic memory. *Psychonomic Bulletin & Review*, *22*(1), 274–280. https://doi.org/10.3758/s13423-014-0673-7

Loaiza, V. M., & McCabe, D. P. (2012a). Temporal–contextual processing in working memory: Evidence from delayed cued recall and delayed free recall tests. *Memory & Cognition*, *40*(2), 191–203. https://doi.org/10.3758/s13421-011-0148-2

Loaiza, V. M., & McCabe, D. P. (2012b). The influence of aging on attentional refreshing and articulatory rehearsal during working memory on later episodic memory performance. *Neuropsychology, Development, and Cognition: Section B, Aging, Neuropsychology and Cognition*, *20*(4)471–493. https://doi.org/10.1080/13825585.2012.738289

Loaiza, V. M., Rhodes, M. G., & Anglin, J. (2013). The influence of age-related differences in prior knowledge and attentional refreshing opportunities on episodic memory. *The Journals of Gerontology: Series B, Psychological Sciences and Social Sciences*, *70*(5), 729–736. https://doi.org/10.1093/geronb/gbt119

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*(2), 480–494. https://doi.org/10.1016/j.jml.2007.04.004

Mora, G., & Camos, V. (2013). Two systems of maintenance in verbal working memory: Evidence from the word length effect. *PLOS ONE*, *8*(7). https://doi.org/10.1371/journal.pone.0070026

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*, *23*(1), 103–123. https://doi.org/10.3758/s13423-015-0947-8

Morrison, A. B., Rosenbaum, G. M., Fair, D., & Chein, J. M. (2016). Variation in strategy use across measures of verbal working memory. *Memory & Cognition*, *44*(6), 922–936. https://doi.org/10.3758/s13421-016-0608-9

Raye, C. L., Johnson, M. K., Mitchell, K. J., Reeder, J. A., & Greene, E. J. (2002). Neuroimaging a single thought: Dorsolateral PFC activity associated with refreshing just-activated information. *NeuroImage*, *15*(2), 447–453. https://doi.org/10.1006/nimg.2001.0983

Raye, C. L., Johnson, M. K., Mitchell, K. J., Greene, E. J., & Johnson, M. R. (2005). Special Issue: Original Article Refreshing : a Minimal Executive Function. Control, 43(1), 135–145. https://doi.org/10.1016/S0010-9452(08)70451-9

Raye, C. L., Johnson, M. R. M. K., Mitchell, K. J., Greene, E. J., & Johnson, M. R. M. K. (2007). Refreshing: A minimal executive function. *Cortex*, *43*(1), 135–145. https://doi.org/10.1016/S0010-9452(08)70451-9

Roediger, H. L., III, & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational. *Perspectives on Psychological Science*, *1*(3), 181–210.

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

Schielzeth, H., & Forstmeier, W. (2009). Conclusions beyond support: Overconfident estimates in mixed models. *Behavioral Ecology*, *20*(2), 416–420. https://doi.org/10.1093/beheco/arn145

Souza, A. S., & Oberauer, K. (2017a). The contributions of visual and central attention to visual working memory. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-017-1357-y

Souza, A. S., & Oberauer, K. (2017b). 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

Souza, A. S., Rerko, L., & Oberauer, K. (2015). Refreshing memory traces: Thinking of an item improves retrieval from visual working memory. *Annals of the New York Academy of Sciences*, *1339*(1), 20–31. https://doi.org/10.1111/nyas.12603

Stan Development Team. (2017). Stan Modeling Language: User's guide and feference manual (Version 2.17.0) [Computer software]. Retrieved from http://mc-stan.org/users/documentation/

Stan Development Team. (2018). rstanarm: Bayesian applied regression modeling via Stan (R Package Version 2.17.3) [Computer software]. Retrieved from http://mc-stan.org/

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 381–403). New York, NY: Academic Press.

Unsworth, N., & Engle, R. W. (2007). 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*(6), 1038–1066. https://doi.org/10.1037/0033-2909.133.6.1038