An examination of refreshing in between-category sequences

Naomi Langerock[1], David Wisniewski[2], Marcel Brass[2], and Evie Vergauwe[1]

1. University of Geneva, Switzerland

2. Ghent University, Belgium

Word Count : 5394 words

Corresponding Author : Naomi Langerock, Université de Genève, Faculté de Psychologie et des Sciences de l'Education, 40 bd du Pont d'Arve, 1211 Genève 4, Switzerland, Naomi.Langerock@unige.ch.

Short Title: Refreshing in between-category sequences

Abstract

The present study focused on refreshing within a working memory context. Refreshing refers to the mechanism that brings back information into the focus of attention in order to counteract forgetting of memory traces. Despite quite some research on this topic, the exact nature of refreshing remains unclear. The present study investigated refreshing by means of the cognitive load effect. This effect is typically observed in complex span tasks which combine processing and storage demands. It refers to the observation that <u>working memory</u> performance depends on the cognitive load of concurrent processing, defined as the proportion of time between list items that is occupied by concurrent processing and therefore not available to refresh memory items. Traditionally, the cognitive load effect has been demonstrated <u>using within-category memory sequences, in which all memory items are drawn from one category (e.g. all words). Here, we show that the cognitive load effect also applies to between-category memory sequences, in which memory items are drawn from different categories (e.g. words, orientations, faces, etc.). The ensemble of the results adds to the domain-generality of the cognitive load effect. Implications concerning the specific nature of refreshing and future research directions are discussed.</u>

Introduction

Working memory (WM) refers to the ensemble of processes with as a common goal to keep information in a state of activation, ready for immediate use, in the presence of ongoing cognitive activities. It thus often involves a maintenance and a processing component, and the interplay between these two processes constitutes a main topic of research. One of the main questions is concerned with the nature of the resources supporting these processing and maintenance activities. In the early versions of one of the most influential models of WM, the multiple-component model[1,2], processing and maintenance activities are assumed to rely on different resources, with maintenance supported by domain-specific resources (e.g., articulatory rehearsal or domain-specific visual or spatial resources) and processing supported by domain-general attentional resources as well. Because maintenance is assumed to rely exclusively on domain-specific resources, it is expected that short-term maintenance will be disrupted by concurrent processing that involves material pertaining to the same domain, for example when both components involve verbal information or when both components involve visual or spatial information, but not, or very little, by concurrent processing that involves material pertaining to another domain, for example when maintenance involves verbal information and processing involves visuo-spatial information[3-6].

Several other models of WM, however, assume the existence of a common resource shared between maintenance and processing in WM, and this resource is assumed to be central attention[7-9]. This attentional resource is suggested to be domain-general and thus, to be involved in maintenance and processing regardless of the nature of the information to be maintained or processed[Footnote1]. As a result, it is expected that maintenance will be disrupted by any attention-demanding concurrent processing task, regardless of the nature of the information to be maintained or processed. Thus, based on whether or not domain-general attention is assumed to support maintenance, concurrent processing of information pertaining

to another domain is expected to interfere with short-term maintenance or not. The role of attention in short-term retention is thus of crucial importance to understand the interplay between maintenance and processing activities in WM.

The role of attention in maintenance: The effect of Cognitive Load

It is within the debate of domain-specific versus domain-general WM resources that the cognitive load (CL) effect was first described[7,12]. The CL effect refers to the observation that the length of memory sequences that can be remembered decreases as the CL of concurrent processing activities increases, with CL defined as the proportion of time during which processing captures attention in such a way that attentional maintenance cannot take place[7,13-14]. The observation that short-term memory performance depends heavily on the attentional demands of concurrent processing lends strong support to the idea that short-term maintenance is critically dependent on attention.

The CL effect has mainly been examined using the complex span task. In this task, participants have to maintain series of memory items which are presented in alternation with processing phases (see Figure 2). During each processing phase, several processing items are presented, and CL can be manipulated by, for example, manipulating the number of items to be processed within a fixed time interval; as more items need to be processed, less time remains available for attentional maintenance. Making use of complex span tasks, it has been demonstrated that the CL effect applies not only to the maintenance of verbal material[7,13,15,16] but also to the maintenance of visual and spatial material[17-19], and to the maintenance feature associations like letters and locations[20]. The CL effect appears thus to be a robust phenomenon that generalizes across different memory materials. This supports the idea that, in general, short-term maintenance depends on attention. Moreover, consistent with the idea that it is domain-general attention that is shared between maintenance and processing, the CL effect has been demonstrated, regardless of whether processing and maintenance materials

pertain to the same domain or to different domains[17-19]. For example, the CL effect still appeared when letters to be maintained were combined with a spatial processing task or when locations to be maintained where combined with a semantic judgment task. Together, these findings strongly suggest that short-term maintenance is critically dependent on attention, a domain-general resource. The goal of the present study was to provide further evidence for the domain-generality of the CL effect by investigating this effect in memory sequences that combine different categories of information in a same memory sequence (e.g., words, orientations, places, faces).

Cognitive load and attentional refreshing

One prominent way in which the CL effect has been interpreted is by assuming that, as soon as attention is available in between processing items, it is used to reactivate decaying memory traces through attentional focusing[7,13-15,23-25]. This process is often referred to as *refreshing* and the assumption is that attentional refreshing reactivates information represented in WM by briefly bringing representations back into the focus of attention[7,24-26]. Refreshing is assumed to be different from articulatory rehearsal; whereas refreshing is by definition assumed to rely on attention, rehearsal uses the speech system and is assumed to operate independently from attention[15,22,27,28,30].

Because the CL effect is assumed to reflect the interplay between memory decay and refreshing within a WM context (but see Ref. 29 for an alternative interpretation), it has frequently been used to study the process of refreshing. For example, *(1)* The CL effect has been demonstrated for different types of memory material (verbal, visual, spatial; Refs. 7,17-20). It is hence typically proposed that refreshing is a domain-general mechanism of maintenance that can be used for almost all types of materials. *(2)* Camos and colleagues have used the CL effect to examine how refreshing relates to articulatory rehearsal. By demonstrating that the CL effect did not interact with effects typically linked to articulatory

rehearsal (e.g., articulatory suppression, phonological similarity, or word length; Refs. 15, 27, 28, 30, for a review, see Ref. 21), it was concluded that refreshing and articulatory rehearsal are independent. Similarly, *(3)* Bayliss and colleagues have used the CL effect to examine how refreshing relates to short-term consolidation. By demonstrating that the CL effect did not interact with an effect typically linked to consolidation (i.e., providing more time between item presentation and the first processing item in a complex span task), it was concluded that refreshing is different from short-term consolidation[31]. And finally, *(4)* several studies have shown a CL effect in children between 7 and 14 years old, suggesting hence that the use of refreshing as a maintenance strategy is already used by children at the age of seven years old[32.33,28]. In the present study, we propose to use a similar approach, using the CL effect to examine refreshing. In particular, we propose to examine the generalizability of refreshing by examining whether the CL effect, and thus refreshing, occurs for between-category memory sequences.

Refreshing in between-category sequences

Up until now, the CL effect has only been demonstrated for memory sequences that consist of items that pertain to the same category, i.e., within-category sequences such as series of letters, series of words, series of spatial locations, and so on. It is currently unclear whether the CL effect can also be observed for memory sequences that consist of items that pertain to different categories of information, i.e. between-category sequences <u>such as a sequence consisting of a face, a word, an orientation and a place</u>. Based on the assumption that the CL effect can be used to infer the occurrence of refreshing, and on the assumption that refreshing is a general mechanism that can be used to maintain almost all types of material, there is no apparent reason to expect that the CL effect would only be observed for within-category sequences and thus, there is no apparent reason to expect that the CL effect would not or to a lesser account occur for between-category sequences. Moreover, refreshing

is typically proposed as a mechanism that counteracts decay[7,13,25] and, to our knowledge, there is currently no theory assuming that decay only occurs in within-category memory sequences. Therefore, based on common assumptions of refreshing, decay and the CL effect, one would expect a CL effect for between-category sequences.

Nevertheless, time-based decay as the main determinant of short-term forgetting (see Ref. 34 for an overview) is not accepted by all and several researchers argue rather in favor of interference-based forgetting[35-37]. One could then assume that refreshing is not necessarily tied to decay, but might rather serve to counter interference caused by new-incoming information that would act on the memory representations. In a recent study of particular relevance for this point, Farrell and Oberauer[38] had compared recall for same-category sequences (six items drawn from one category) and between-category sequences (six items drawn from three categories, thus two items per category). They had observed that participants made generally less errors in the between-category sequence, and the errors made in these between-category sequences concerned relatively more within-category errors than expected based on simulations from the same-category sequences. In their experiment, in contrast to the complex span paradigm, no specific refreshing opportunities were inserted, nor was there any concurrent attention-demanding task. Thus, in a simple short-term memory task, recall of items of different categories was observed to be better than recall of items drawn from a single category. A similar result had been observed by Young and Supa[39]. They presented participants with sequences of either only digits, only words, or a combination of both. The task also concerned an immediate recall task and they observed, like Farrell and Oberauer[38], better memory performance for between-category sequences than for same-category sequences. It is likely that remembering items drawn from a single category is more difficult than remembering items drawn from different categories because of within-category confusion. If one assumes that refreshing counteracts confusion-based forgetting, rather than

decay-based forgetting, then it would be possible that the CL effect observed so far in within-category sequences reflects the need to refresh confusable items. If refreshing serves to reduce or eliminate this confusion, then the need to refresh would be drastically reduced in between-category series. As a consequence, using list items from fairly distinct categories may abolish, or at least strongly minimize, the CL effect.

The current study tested whether a CL effect can be observed for between-category sequences. Observing a CL effect for between-category sequences would contribute to the growing body of evidence for CL effects in WM and testify to its generality. On the other hand, not observing a CL effect for between-category sequences would cast doubt on common assumptions of refreshing, decay and the CL effect.

Method

Participants

Thirty-five students from the University of Geneva (mean age = 21.31 years, SD = 1.96, 24 females) have participated in this experiment for partial course credits.

Tasks and Materials

The task was created using E-Prime 2.0 software[40] and administered to the participants on a computer. A complex span task was created by combining series of four memory items belonging to four different categories (maintenance component) with an auditory discrimination task (processing component). The task procedure is shown in Figure 2. The categories used in this study were pseudo-words, line orientations, faces and outdoor places. These categories pertain to different domains (verbal, spatial, visual) and are well differentiated from each other[Footnote 2].

The memory materials consisted in pools of 16 instances per category (Figure 1 shows two examples for each category). The 16 pronounceable French pseudo-words were

drawn from a list of 120 pseudo-words used in a recent study by Camos, Mora, Oftinger & Vergauwe[48]. We selected pseudo-words that had been created by changing a vowel from an existing word, with priority given to the words with lowest frequency in order to avoid semantic elaboration. The images of line orientations were based on the stimuli used by Lewis-Peacock et al.[41] and Lewis-Peacock and Postle[42]. These always consisted of a combination of two line segments. These line segments could be oriented 30° to the left, 30° to the right, 60° to the left, or 60° to the right. Their combination led to 16 line orientation images. The faces category consisted in 16 pictures of male faces, drawn from the pool of stimuli of Lewis-Peacock et al.[41]. The places category consisted in a pool of 16 pictures of outdoor places, also drawn from the pool of stimuli of Lewis-Peacock et al.[41].

We created 24 trial types in total (see Figure 2 for a general example of a trial). For each trial type, participants had to remember four memory items in the correct order. The trial types were created in such a way that each category appeared six times in each of the four serial positions, and each category in a certain serial position was always followed by any of the other categories on exactly two trials (e.g. category 1 in serial position 2 is twice followed by category 3, twice by category 4, and twice by category 2). Each category is hence completely independent from the next upcoming category. These 24 trial types make up one block and were repeated in each of the three blocks. That is, the same 24 trial types were used in each block, but the specific exemplars used for each category were randomly determined on each trial out of the 16 possible category items. Furthermore, the order of the different trial types was randomized in each block and thus trial order was different in each block. The association between trial type and CL was counterbalanced across participants and each block had 12 trials with a low CL and 12 trials with a high CL, randomly intermixed.

Participants were instructed to remember the information in order of presentation for final recall at the end of the trial. This recall was administered as a probed recall (see Figure

2), i.e., one of the four squares representing the four serial positions was highlighted in red and participants had to select the correct item for the highlighted serial position from the eight options per category displayed below. We opted for probed recall of only one item instead of serial recall of all the items in order to avoid an unnecessary lengthening of the experiment. Regarding this probed recall, on the recall screen eight possible options per category were selected in such a way that for each of the four categories the correct category item was part of it, together with seven other, randomly determined items from the same category. The eight options were randomly distributed over the eight squares per category. The order of the categories was fixed: from upper to lower screen participants saw places, orientations, faces and pseudo-words. The probed memory item was selected in such a way that within each block, each serial position was probed 6 times, 3 times in a low CL and 3 times in a high CL trial. Regarding the different categories, over the three blocks each category was probed 4 times in each serial position (twice in a low and twice in a high CL condition; 4 categories *4 serial positions *4 times = 64 trials). In the remaining 8 trials (64 + 8 = 72) each category was probed two more times (once in the low and once in the high CL condition) in two randomly defined serial positions.

For the auditory discrimination task to be performed during the processing phase, participants had to discriminate between a low (262 Hz) and a high tone (524 Hz) by pressing keys (see Refs. 20 and 49-51 for other studies using this task as attention-demanding processing task in a WM context during maintenance). Participants were to press 1 on the keyboard if they thought the tone was low in frequency, or 2 if they thought the tone was high in frequency, immediately upon hearing each tone.

Procedure

The procedure for a single trial was as follows, see Figure 2. After a 500-ms fixation cross, a first memory item was presented for 750 ms, followed by a blank screen for 250 ms.

Then a processing phase of 4000 ms started during which participants heard two (low CL) or four tones (high CL) to discriminate through a headphone (lasting for 200 ms). These tones were presented at a regular rhythm (1 tone every 2000 ms or 1 tone every 1000 ms in the low and high CL conditions, respectively). After the first processing phase, the second memory item was shown for 750 ms, followed by a blank screen for 250 ms and a second processing phase. This sequence was repeated for each memory item, until the four memory items had been presented. All memory items were presented in the center of the screen, against a white background, within a virtual rectangle of 20*25 cm.

The fourth processing phase was followed by the probed recall phase. Participants were presented on screen with four empty boxes (i.e., as much as items to remember), and one of these boxes was highlighted in red (i.e., red border instead of black). The participant was instructed to recall the item that was presented in that specific serial position. Thus, if the second box was highlighted, the participant should recall the item that was presented in the second serial position. Participants had to click on the item presented below that had occupied this serial position in the memory sequence and this item appeared then in the highlighted box. After a short jittered inter trial interval (logarithmic distribution with mean 6000 ms, min 700 ms, max 14500 ms), the next trial started.

Participants started the experiment with a training phase. First of all, the general outline of the experiment was explained and participants were shown the 64 possible memory items, presented by category to familiarize them with these images. Then the maintenance task was explained, followed by two examples. Next participants were trained on the tone discrimination (processing) task until they reached a minimum score of 80% correct. Finally, participants performed three training trials combining the maintenance and processing task. All participants indicated that they had well understood the task and then started the experimental trials.

Results

The current study and main analysis pertaining to our research question, i.e., the presence of a CL effect in between-category memory sequences, was preregistered on aspredicted.com (https://aspredicted.org/2wh92.pdf). As planned, participants who did not perform well on the tone-discrimination task were excluded from further analysis, as it could not be ascertained that they followed the instructions inducing the manipulation of the CL. The cut-off criterion was set at 70 percent of correct responses on the tone discrimination task. Initially and as preregistered, 24 participants were tested. Five of these were excluded based on the 70% cut-off criterion. The results of the remaining 19 participants resulted in clear evidence for the main goal of the present study, i.e., the presence of a clear CL effect in between-category memory sequences. Because several additional, exploratory analyses remained inconclusive with the data of only 19 participants, we decided to add more participants. Given Bayesian analyses, there is little danger from collecting additional data[52,53]. In particular, we added 11 participants (total of 35) in order to arrive at a final sample of 30 participants that passed the 70% cut-off criterion for tone discrimination task. The mean accuracy score on the tone discrimination task for these remaining 30 participants was 91 % (SD = 5).

For all analyses, we have used JASP[54]. JASP uses the BayesFactor package[55] as a back-end computational engine. In JASP, the null hypothesis is a point-null hypothesis with all mass on 0. For the alternative models, we have used the default settings proposed by JASP (i.e., t-test: r scale = .707; ANOVA: r scale = .5 for fixed effects, r scale = 1 for random effects, and r scale = .354 for covariates).

Planned analysis

In line with the preregistered plan of analysis, we calculated mean recall scores for the probed memory item in the low and high CL conditions. In the low CL condition, participants

reported on average 74% (SD = 16) of the probed memory items correctly, while in the high

CL condition, this was only 61% (SD = 17). As preregistered, we used Bayesian analysis to

examine the evidence in the data for the predicted difference in mean recall scores. To do so,

we used the Bayes Factor approach (in contrast to the parameter estimation approach). We

performed the planned and preregistered one-sided Bayesian t-test in JASP revealing extreme

evidence[Footnote 3] in favor of the expected CL effect in between-category memory sequences,

BF = 604731. In the supplementary material 1, the sequential BF analysis is shown (as we

added more participants later on) as well as a robustness check.

Further analyses

The following analyses were not preregistered and were performed to explore and

qualify the observed CL effect. In a first additional analysis, we examined the CL effect for

each category and over the different trial blocks. In a second additional analysis, we explored

the recall error pattern in more detail.

In the first additional analysis, we performed a 2 (CL: low and high) * 4 (Category:

places, faces, orientations and words) * 3 (Block: first – second – third) Bayesian repeated

measure ANOVA on the percentage correct recall with CL, Category and Block as within-

subject factors. The best model included the main effects of CL and of Category, but no main

effect of Block, nor any interaction terms. This best model (CL + Category) was preferred by

a BF of 1.04^{19} over the null model. Additionally, the best model was preferred by a BF of

about 17 over a model that also included the interaction between CL and Category. There is

thus strong evidence against the hypothesis that the CL effect differed among the different

memory categories [Footnote 4]. Next, we compared the best model (CL + Category) against a

model that also includes the main effect of Block. The best model was preferred by a BF of

about 30 over the model including the factor Block as well. This strong evidence for the

In the second additional set of analyses, we analyzed the error patterns in the low and the high CL conditions. Each answer given by the participant can be categorized into four response types, of which one corresponds to the correct response and the three correspond to errors (Figure 4). If the answer given corresponds to the probed item, this results in a "Correct response". Regarding our example in Figure 2, this would correspond to answering "PAUF". If the answer given by the participant is not the correct response, then it was first verified whether the participant reported the category correctly. If this was the case, this means the participant reported an incorrect instance of the correct category, corresponding hence to an "Instance error". In our example this would correspond to answering for example "GING", which corresponds to the correct category of pseudo-words, but it is not the correct instance. If on the other hand, the category was not correct, then it was checked whether the participant reported a correct instance of another category. This would come down to an "Order error". In our example, reporting the two lines tilted 30° towards each other on the upper side would correspond to an order error. The participant gave the correct instance but of a non-probed category. If, on the other hand, the participant reported an incorrect instance of a non-probed category, this was classified as an "Other error". Answering the two lines both oriented 30° to the right would correspond to this kind of error. The participants thus reported an instance of an incorrect category, and additionally, for the reported category, the participant reported an instance that had not been shown in that trial.

The main goal of analyzing the error patterns was to explore whether reducing the opportunity for refreshing (going from low to high CL) would result in a disproportional increase in one of these error types. If that were the case, this could inform us more about the nature of refreshing. For example, observing a higher proportion of order errors in the high

CL than in the low CL condition would suggest that refreshing specifically focusses on order information. Or, observing a higher relative proportion of instance errors in the high than in the low CL condition would indicate that refreshing acts rather on the item level, without specifically taking into account category information. Figure 5 shows the distribution of the error types in the both CL conditions. It can immediately be seen that the relative proportions of the different error types were not drastically different between the low and high CL conditions.

We performed Bayesian t-tests in JASP to compare the proportion of each of these error types in the low and the high CL conditions. Regarding the instance errors, we obtained inconclusive ("anecdotal") evidence against a difference (BF = 2.62) between the CL conditions. If any difference was observed, then it was more instance errors in the low CL condition. Regarding the order errors, we obtained moderate evidence against a difference between the low and the high CL condition (BF = 4.98). Moderate evidence was also obtained against a difference for the other errors between the low and the high CL conditions (BF = 3.17). Thus, taken together, there is no evidence in the data for a particular relation between refreshing and error types.

Discussion

The main outcome of the present study concerns the observation of a CL effect for between-category memory sequences. The observation of a CL effect for between-category sequences adds to the robustness of the CL effect, which previously had only been shown using memory material belonging to one same category within a sequence (e.g., all verbal, spatial, or visual memory items). Moreover, as the CL effect is typically interpreted as the result of the act of refreshing, observing the effect for between-category memory sequences adds to the generalizability of refreshing as a domain-general mechanism to maintain information in WM.

In the present experiment, we obtained very straightforward evidence for a CL effect, i.e., better memory performance in the low than in the high CL condition. This effect was observed for each of the different categories separately: pseudo-words, faces, places, and line orientations. To our knowledge, we are not only the first to demonstrate the CL effect for between-category sequences, but we are also the first to demonstrate the CL effect for these specific memory materials (except for non-words for which a CL effect was recently demonstrated[46]) <u>Footnote 5</u>. Based on current assumptions on refreshing, i.e. a domain-general attentional maintenance mechanism counteracting decay-based forgetting of memory representations, we had hypothesized to observe a CL effect for between-category sequences. As stated in the introduction, there was nevertheless at least one reason why one could have expected not to observe a CL effect for between-category memory sequences. Instead of refreshing being tied to decay, one could assume refreshing to serve as counteracting interference and as such serve to keep representations distinguished from one another. The studies of Farell and Oberauer[38] and Young and Supa[39] for example had shown that sequences composed of items belonging to different categories are easier to remember than sequences of items belonging to a same category. This outcome may be the result of within-category confusion, as items showing more overlap are more confusable. If refreshing serves to eliminate this confusion, by making them more distinct, then we should not or to a much lesser extent have observed a CL effect for between-category sequences. The current study contradicts this reasoning by showing strong evidence for a CL effect when the memory items belong to different categories. Thus, it appears that refreshing is not limited to the maintenance of sequences of memory items that can easily be confused.

<u>Now that we have demonstrated, for the first time, a CL effect in between-category memory sequences, it might be interesting, in future studies, to directly compare the CL effect in between-category and within-category memory sequences. This will allow testing</u>

whether the magnitude of the CL effect is similar in between-category and within-category memory sequences. As a first comparison, in the supplementary material 2 we present a between-study comparison using the data of an earlier study of ours in which within-category sequences were used[20]. This comparison appears to reveal similar CL effects in the two types of memory sequences.

In exploratory, follow-up analyses of the main results of the present study, we have searched for some further clues on the nature of refreshing by exploring the error patterns in the low and high CL conditions. Three types of errors were explored: instance errors (i.e., correct category but erroneous category instance), order errors (i.e., correct instance of a category not probed in that serial position) and other errors (i.e., incorrect instance of a category not probed in that serial position). We observed that the proportion of these errors in the low and high CL condition was fairly similar; almost half of the observed errors in both conditions were order errors, about one third of the observed errors in both conditions were instance errors and the remaining 20% of the observed errors in both conditions were other errors. Thus, none of the error types was disproportionally represented when the CL of concurrent processing was increased. This indicates that refreshing affects item and order information similarly, and is in line with recent findings of Camos, Lagner, and Loaiza[57] who directly compared item-based and order-based recall in verbal WM. Thus, refreshing seems to act on a representation as a whole, taking into account the totality of information linked to the memory item instead of having its focus on either item or order information only.

It was interesting to observe that the modifications that we made in the typical complex span paradigm in the present study did not alter the typically observed findings, confirming hence the robustness of the CL effect. As a main manipulation, we had changed the typical same-category memory items to between-category memory items and we still observed a CL effect. Next to this main manipulation, (1) we made use of categories that had

not been used before in a complex span paradigm, i.e., line orientations, faces and outdoor places. We obtained nevertheless a typical CL effect for these, as well as for the non-novel categories (i.e., pseudo-words[46]). (2) We changed the typical serial recall procedure to a probed recall procedure and still obtained a CL effect. (3) Last but not least, we increased the number of trials typically used in the complex span paradigm (usually not exceeding 36 trials) to 72 trials in the present study and the CL effect remained present from the beginning until the end of the experiment. This indicates that the spontaneous use of refreshing as a maintenance strategy was not reduced by fatigue that might arise by the end of the experimental trials. We can hence conclude that the CL effect and the assumed act of refreshing underlying this effect are robust phenomena, resisting to a number of variations in the paradigm used and strongly suggesting a domain-general character.

"The results of the present study fit well with those WM frameworks that assume the existence of domain-general resources to support maintenance[7-9]. On the contrary, strictly speaking the results do not fit with the early versions of the multi-component account of WM[1,2], in which maintenance is supposed to be driven by domain-specific resources and not necessarily by attention. A more recent version of the multi-component model[58] added the episodic buffer as a fourth component. This buffer is conceived as a storage device that is capable of maintaining different types of information and might well be driven by attentional resources. Such a kind of intermediate account was also suggested by a number of recent studies[10,15], with maintenance typically supported by attentional processes, but verbal maintenance possibly additionally supported by domain-specific verbal resources. No domain-specific visuo-spatial resources were proposed in these studies, leading to an asymmetry between verbal and visuo-spatial resources. The present results could fit with these WM accounts, as long as there is an explicit attentional component involved in the maintenance of information within WM."

To conclude, the present study has shown that a typical CL effect is observed when using the complex span paradigm in which memory items belonging to different categories have to be maintained. This observation adds to the robustness of the CL effect and the domain-generality of refreshing as a maintenance strategy.

Footnotes

**1.** It should be noted, however, that there are also data and theories indicating that attentional resources could be domain-specific, or at the very least, asymmetric[10,11].

2. Furthermore, these categories respond to an additional criterion which is that they are suitable for use in Multi Voxel Pattern Analysis (MVPA) of functional Magnetic Resonance Imaging (fMRI) data[41,42]. The technique of MVPA[43-47] has only recently been developed and several studies have provided evidence that MVPA of fMRI data can be used to decode the content of the focus of attention[41,42]. This technique consists in two steps: (1) to train the pattern classifier, category-labeled brain activity patterns are fed to the classifier and the classifier fits a decision boundary that separates brain activity patterns of different memory categories, and (2) to test the classifier, a novel, independent brain activity pattern is given to the classifier and it is tested whether the trained classifier can correctly determine the category associated with that pattern. Given the current state of the method, between-category decoding works best. Using MVPA, it can be tracked which category is actively thought about at different points in time. Making use of MVPA in future studies as a complement to behavioral studies could allow us to obtain more information about what is typically going on during a WM task. We opted hence for the use of MVPA suitable categories to see whether this technique could be used in the future within the CL paradigm.

**3.** According to Jeffreys'[56] classification scheme of BF's.

4. For each of the four categories, we performed a one-sided Bayesian t-tests. All BF's were in favor of a higher recall performance in the low CL conditions and these BF's were all above 10, providing strong evidence that recall was indeed better in the low CL condition, for each of the four memory categories.

5. As stated in Footnote 1, the categories chosen in this study were categories that had previously been shown to be suitable for use within an MVPA-paradigm (up until now typically making use of between-category sequences). The fact that the CL effect applies to between-category memory sequences with these specific categories suggests that future MVPA studies could use the paradigm of the current study to examine refreshing in working memory.
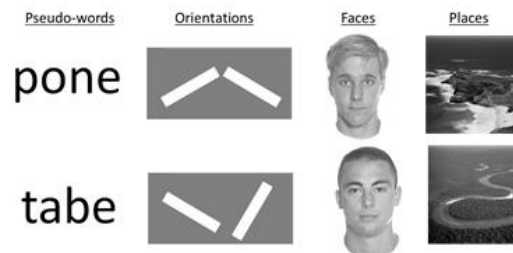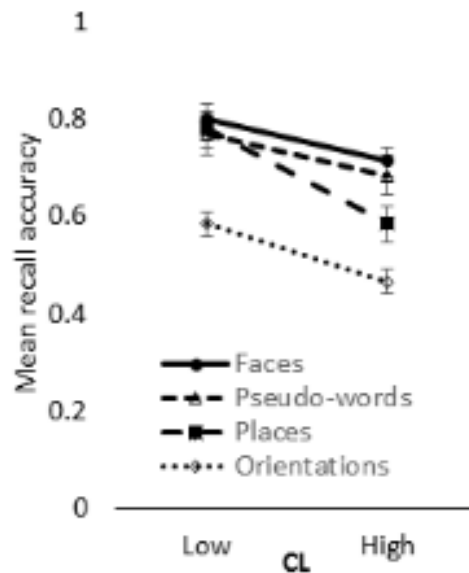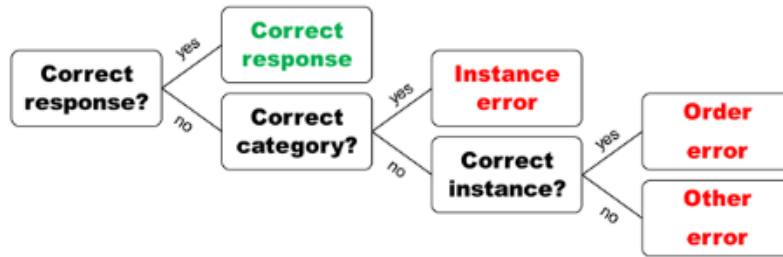
Figure 1. Two examples of instances of each of the four memory categories
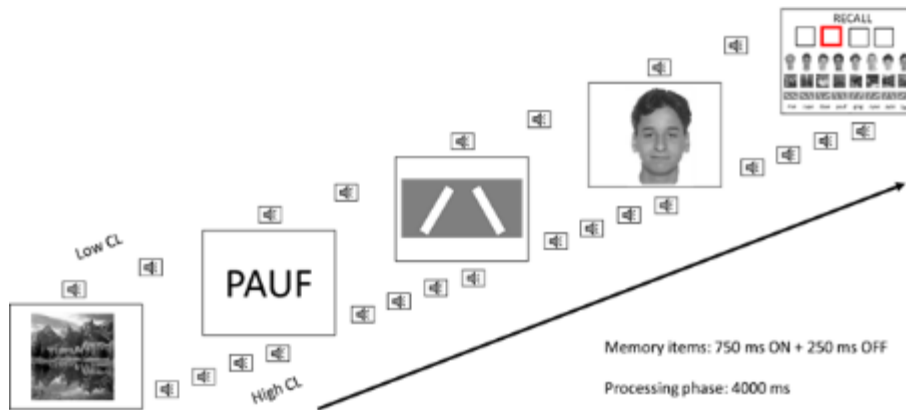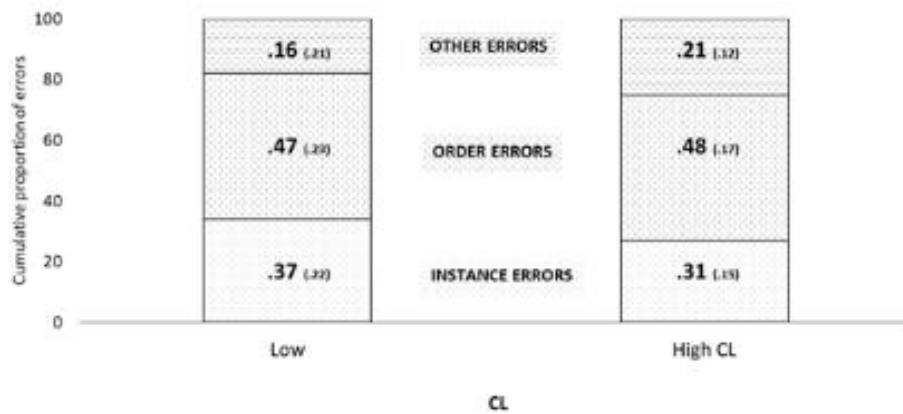

Figure 2. Example of a trial in the Low and the High CL condition.

Figure 3. Mean percentage correct recall for the probed item as a function of CL and

Category. Error bars represent the standard error of the mean.

Figure 4. Schematical representation of the correct responses and different error types.

Figure 5. Distribution of the error types according to the CL condition. Each cell reports the mean proportion of errors and its standard deviation.

Acknowledgments

Author Contributions: All authors contributed to the study design. The experiment was programmed and the data collected by Naomi Langerock. Analysis of the data was done by Naomi Langerock and Evie Vergauwe. The first draft of the manuscript was written by Naomi Langerock, the remaining authors provided critical feedback and edited the manuscript accordingly.

References

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

2. Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). New York: Cambridge University Press.

3. Cocchini, G., Logie, R. H., Della Sala, S., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition, 30*, 1086-1095. doi:10.3758/Bf03194326

4. Duff, S. C., & Logie, R. H. (2001). Processing and storage in working memory span. *Quarterly Journal of Experimental Psychology, 54*, 31-48. doi:10.1080/02724980042000011

5. Logie, R. H. (1986). Visuospatial Processing in Working Memory. *Quarterly Journal of Experimental Psychology, 38,* 229-247. doi:10.1080/14640748608401596

6. Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Interference with Visual Short-Term-Memory. *Acta Psychologica, 75*, 55-74. doi:10.1016/0001-6918(90)90066-O

7. Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology-General, 133*, 83-100. doi:10.1037/0096-3445.133.1.83

8. Cowan, N. (1988). Evolving Conceptions of Memory Storage, Selective Attention, and Their Mutual Constraints within the Human Information-Processing System. *Psychological Bulletin, 104*, 163-191. doi:10.1037/0033-2909.104.2.163

9. Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999). Individual difference in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 102-134). Cambridge: Cambridge University Press.

10. Morey, C. C., & Miron, M. D. (2016). Spatial sequences, but not verbal sequences, are vulnerable to general interference during retention in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*, 1907-1918. doi: 10.1037/xlm0000280

11. van Dijck, J. P., Abrahamse, E. L., Majerus, S., & Fias, W. (2013). Spatial attention interacts with serial-order retrieval from verbal working memory. Psychological Science, 24, 1854-1859. doi:10.1177/0956797613479610

12. Barrouillet, P., & Camos, V. (2001). Developmental increase in working memory span: Resource sharing or temporal decay? *Journal of Memory and Language, 45*, 1-20. doi:10.1006/jmla.2001.2767

13. Barrouillet, P., Bernardin, S., Portrat, S., Vergauwe, E., & Camos, V. (2007). Time and cognitive load in working memory. *Journal of Experimental Psychology-Learning Memory and Cognition, 33*, 570-585. doi:10.1037/0278-7393.33.3.570

14. Barrouillet, P., Portrat, S., & Camos, V. (2011). On the Law Relating Processing to Storage in Working Memory. *Psychological review, 118,* 175-192. doi:10.1037/A0022324

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

16. 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,* 1675-1684. doi:10.3758/Bf03193501

17. Vergauwe, E., Barrouillet, P., & Camos, V. (2009). Visual and Spatial Working Memory Are Not That Dissociated After All: A Time-Based Resource-Sharing Account. *Journal of Experimental Psychology-Learning Memory and Cognition, 35*, 1012-1028. doi:10.1037/a0015859

18 Vergauwe, E., Barrouillet, P., & Camos, V. (2010). Do Mental Processes Share a Domain-General Resource? *Psychological science, 21*, 384-390. doi:10.1177/0956797610361340

19. Vergauwe, E., Dewaele, N., Langerock, N., & Barrouillet, P. (2012). Evidence for a central pool of general resources in working memory. *Journal of Cognitive Psychology, 24*, 359-366. doi:10.1080/20445911.2011.640625

20. Langerock, N., Vergauwe, E., & Barrouillet, P. (2014). The Maintenance of Cross-Domain Associations in the Episodic Buffer. *Journal of Experimental Psychology-Learning Memory and Cognition, 40,* 1096-1109. doi:10.1037/a0035783

21. Camos, V. (2017). Domain-Specific Versus Domain-General Maintenance in Working Memory: Reconciliation Within the Time-Based Resource Sharing Model. *The Psychology of Learning and Motivation, 67,* 135-171. doi:10.1016/bs.plm.2017.03.005

22. Vergauwe, E., & Cowan, N. (2014). A common short-term memory retrieval rate may describe many cognitive procedures. *Frontiers in Human Neuroscience, 8*. doi: 10.3389/fnhum.2014.00126

23. Vergauwe, E., & Cowan, N. (2015). Working Memory Units Are All in Your Head: Factors That Influence Whether Features or Objects Are the Favored Units. *Journal of Experimental Psychology-Learning Memory and Cognition, 41,* 1404-1416. doi:10.1037/xlm0000108

24. Cowan, N. (1992). Verbal Memory Span and the Timing of Spoken Recall. *Journal of Memory and Language, 31*, 668-684. doi:10.1016/0749-596x(92)90034-U

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

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

27. Mora, G., & Camos, V. (2013). Two Systems of Maintenance in Verbal Working Memory: Evidence from the Word Length Effect. *Plos One, 8*. doi:10.1371/journal.pone.0070026

28. Oftinger, A.-L., & Camos, V. (2016). Developmental improvement in strategies to maintain verbal information in working memory. *International Journal of Behavioral Development,* Advance online publication. doi:10.1177/0165025416679741

29. Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic bulletin & review, 19*, 779-819. doi:10.3758/s13423-012-0272-4

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

31. Bayliss, D. M., Bogdanovs, J., & Jarrold, C. (2015). Consolidating working memory: Distinguishing the effects of consolidation, rehearsal and attentional refreshing in a working memory span task. *Journal of Memory and Language, 81*, 34-50. doi:10.1016/j.jml.2014.12.004

32. Barrouillet, P., Gavens, N., Vergauwe, E., Gaillard, V., & Camos, V. (2009). Working Memory Span Development: A Time-Based Resource-Sharing Model Account. *Developmental* Psychology, 45, 477-490. doi:10.1037/a0014615

33. Gaillard, V., Barrouillet, P., Jarrold, C., & Camos, V. (2011). Developmental differences in working memory: Where do they come from? *Journal of Experimental Child Psychology, 110*, 469-479. doi:10.1016/j.jecp.2011.05.004

34. Ricker, T. J., Vergauwe, E., & Cowan, N. (2016). Decay theory of immediate memory: From Brown (1958) to today (2014). *Quarterly Journal of Experimental Psychology, 69*, 1969-1995. doi:10.1080/17470218.2014.914546

35. Lewandowsky, S., Oberauer, K., & Brown, G. D. A. (2009). No temporal decay in verbal short-term memory. *Trends in Cognitive Sciences, 13*, 120-126. doi:10.1016/j.tics.2008.12.003

36. Oberauer, K., & Lewandowsky, S. (2008). Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological review, 115*, 544-576. doi:10.1037/0033-295x.115.3.544

37. Oberauer, K., & Lange, E. B. (2008). Interference in verbal working memory: Distinguishing similarity-based confusion, feature overwriting, and feature migration. *Journal of Memory and Language, 58*, 730-745. doi:10.1016/j.jml.2007.09.006

38. Farrell, S., & Oberauer, K. (2014). Working memory for cross-domain sequences. *Quarterly Journal of Experimental Psychology, 67*, 33-44. doi:10.1080/17470218.2013.795177

39. Young, C. W., & Supa, M. (1941). Mnemic inhibition as a factor in the limitation of the memory span. *American Journal of Psychology, 54*, 546-552. doi:10.2307/1417204

40. Psychology Software Tools, Inc. [E-Prime 2.0]. (2012). Retrieved from https://www.pstnet.com.

41. Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural Evidence for a Distinction between Short-term Memory and the Focus of Attention. *Journal of Cognitive Neuroscience, 24,* 61-79. doi: 10.1162/jocn_a_00140

42. Lewis-Peacock, J. A., & Postle, B. R. (2012). Decoding the internal focus of attention. *Neuropsychologia, 50*, 470-478. doi:10.1016/j.neuropsychologia.2011.11.006

43. Lewis-Peacock, J.A., Norman, K.A., 2014. Multivoxel pattern analysis of functional MRI data. In: Gazzaniga, M.S., Mangun, G.R. (Eds.), *The Cognitive Neurosciences*. MIT Press, Cambridge, Massachusetts, London, England, 911–920.

44. Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience, 7*, 523-534. doi:10.1038/nrn1931

45. Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 3863-3868. doi:10.1073/pnas.0600244103

46. Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences, 10*, 424-430. doi:10.1016/j.tics.2006.07.005

47. Pereira, F., Mitchell, T., & Botvinick, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *Neuroimage, 45*, S199-S209. doi:10.1016/j.neuroimage.2008.11.007

48. Camos, V., Mora, G., Oftinger, A., & Vergauwe, E. (under revision). Does semantic long-term memory impact refreshing in verbal working memory?

49. Elsley, J. V., & Parmentier, F. B. R. (2009). Is verbal-spatial binding in working memory impaired by a concurrent memory load? *Quarterly Journal of Experimental Psychology, 62*, 1696-1705. doi:10.1080/17470210902811231

50. Vergauwe, E., Langerock, N., & Barrouillet, P. (2014). Maintaining Information in Visual Working Memory: Memory for Bindings and Memory for Features are Equally Disrupted by Increased Attentional Demands. *Canadian Journal of Experimental Psychology-Revue Canadienne De Psychologie Experimentale, 68*, 158-162. doi:10.1037/cep0000025

51. Morey, C. C., & Bieler, M. (2013). Visual short-term memory always requires general attention. *Psychonomic bulletin & review, 20*, 163-170. doi:10.3758/s13423-012-0313-z

52. Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic bulletin & review, 21,* 301-308. doi: 10.3758/s13423-014-0595-4

53. Schönbrodt, F. D., & Wagenmakers, E.-J. (2016). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic bulletin & review,* Advance online publicationf. doi: 10.3758/s13423-017-1230-y

54. JASP Team (2017). JASP (Version 0.8.2)[Computer software].

55. Morey, R. D., & Rouder, J. N. (2012). BayesFactor: Computation of Bayes factors for simple designs [Computer software].

56. Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford: Oxford University Press, Clarendon Press.

57. Camos, V., Lagner, P., & Loaiza, V. M. (2017). Maintenance of item and order information in verbal working memory. *Memory, 25*, 953-968. doi:10.1080/09658211.2016.1237654

58. Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in cognitive sciences, 4,* 417-423. doi:10.1016/s1364-6613(00)01538-2