How Does Semantic Knowledge Impact Working Memory Maintenance? Computational and Behavioral Investigations

Benjamin Kowialiewski[1,2], Benoît Lemaire[1], Sophie Portrat[1]

[1] Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, LPNC, 38000 Grenoble, France

[2] University of Liège, Liège, Belgium

Correspondence concerning this article should be addressed to Benjamin Kowialiewski, Laboratoire de Psychologie et NeuroCognition (LPNC), Université Grenoble-Alpes, Bâtiment Michel Dubois prev. BSHM, 1251 Avenue Centrale, 38400 Saint-Martin-d'Hères. E-mails: benjamin.kowialiewski@univ-grenoble-alpes.fr; bkowialiewski@uliege.be

Abstract

It is now firmly established that long-term memory knowledge, such as semantic knowledge, supports the temporary maintenance of verbal information in Working Memory (WM). This support from semantic knowledge is well-explained by models assuming that verbal items are directly activated in long-term memory, and that this activation provides the representational basis for WM maintenance. However, the exact mechanisms underlying semantic influence on WM performance remain poorly understood. We manipulated the presence of between-item semantic relatedness in an immediate serial recall task, by mixing triplets composed of semantically related and unrelated items (e.g. *leaf – tree – branch* – wall – beer – dog; hand – father – truck – *cloud – sky – rain*). Compared to unrelated items, related items were better recalled, as had been classically observed. Critically, semantic relatedness also impacted WM maintenance in a complex manner, as observed by the presence of proactive benefit effects on subsequent unrelated items, and the absence of retroactive effects. The complexity of these interactions is well-captured by TBRS*-S, a decay-based computational architecture in which the activation occurring in long-term memory is described. The present study suggests that semantic knowledge can be used to free up WM resources that can be reallocated for maintenance purposes, and supports models postulating that long-term memory knowledge constrains WM maintenance processes.

Keywords: Working Memory, Semantic knowledge, Computational models, TBRS

Introduction

Verbal Working Memory (WM) is the ability to maintain verbal information over a short period of time. It has been shown to be influenced by several semantic factors (Poirier & Saint-Aubin, 1995; Walker & Hulme, 1999), suggesting the existence of close interactions between WM and the long-term memory linguistic system. These interactions are now well-established, but the mechanisms through which they occur are poorly specified. This is an important theoretical question, because many contemporary models assume that WM relies on direct activation within the long-term memory system (Acheson & MacDonald, 2009; Cowan, 1995, 2001; Majerus, 2019; Martin et al., 1996; Oberauer, 2009). Importantly, there had been little attempt previously to model these influences within a formal computational architecture. In the present study, we took advantage of a convergent approach involving behavioral and computational methods to assess the hypothesis that semantic knowledge can be used in an efficient manner to free-up WM resources that can then be reallocated to maintain more information.

Many studies have shown that semantic knowledge supports the short-term maintenance of verbal information. For instance, this is demonstrated by the presence of so-called psycholinguistic effects in immediate serial recall tasks. A recall advantage is observed in the semantic relatedness effect for lists composed of semantically related words (e.g. leaf – tree – branch) over semantically unrelated words (e.g. wall – sky – dog) (Kowialiewski & Majerus, 2018; Monnier & Bonthoux, 2011; Poirier & Saint-Aubin, 1995; Tse, 2009; Tse et al., 2011). Similarly, verbal items associated with concrete or highly imageable semantic features (e.g. table – car – hand) are better recalled than verbal items characterized by abstract or low imageable semantic features (e.g. phase – doubt – link). This is known as the concreteness or imageability effect (Acheson et al., 2010; Campoy et al., 2015; Castellà & Campoy, 2018; Chubala et al., 2018; Kowialiewski & Majerus, 2018; Miller & Roodenrys, 2009; Romani et al., 2008; Walker & Hulme, 1999). Other studies also demonstrated the importance of semantic knowledge during WM maintenance. The classical deleterious impacts of phonological similarity (Baddeley, 1966) and word length (Baddeley et al., 1975) on WM performance can be strongly reduced when participants are explicitly instructed to use a semantic maintenance strategy (Campoy & Baddeley, 2008; Logie et al., 1996). Likewise, WM performance increases when such a maintenance strategy is required (Hanley & Bakopoulou, 2003), or when participants are instructed to perform semantic judgements concerning the memoranda (Savill et al., 2015). Overall, these studies add to the empirical evidence showing an influence of long-term memory knowledge on WM performance, which has been shown to occur both in the verbal (Brener, 1940) and the visual (Oberauer et al., 2017; Xie & Zhang, 2017) domains.

Theoretically, semantic effects in WM can be explained by models presuming a close interaction between WM and long-term memory knowledge (Acheson & MacDonald, 2009; Cowan, 1995, 2001; Majerus, 2013, 2019; Martin et al., 1996; Oberauer, 2002, 2009). For instance, the Embedded-Processes model (Cowan, 1995, 2001) is an influential theoretical framework. This framework assumes that WM processing relies on direct activation in long-term memory, and that this activation provides the representational basis for WM maintenance. In the verbal domain, it has been proposed that the maintenance of verbal information may rely on direct activation within the linguistic system itself (Acheson & MacDonald, 2009; Jefferies et al., 2006; Majerus, 2013, 2019; Martin et al., 1996; Patterson et al., 1994). Semantic effects can be explained by assuming that verbal items receive feedback from higher levels of representations through interactive activation principles (McClelland & Rumelhart, 1981). In interactive activation models, the semantic relatedness effect is explained by considering that words related at the semantic level share a higher number of common semantic features than unrelated words do (Dell et al., 1997). Alternatively, semantically related words may have direct lateral excitatory connections between each other, due to lexical co-occurrence effects (Hofmann & Jacobs, 2014). Thanks to these shared semantic features and/or lateral excitatory connections, semantically related items are thought to reactivate each other, which increases their activation level and makes them easier to maintain.

Although activation-based models can theoretically account for the presence of semantic effects in WM, little effort has been made to build a computational WM model in which long-term memory activation is taken into account. The architecture proposed by Haarmann and Usher (2001) is a two-layer neural network composed of a "posterior system" where the initial activation in long-term memory is sent to a limited-capacity "prefrontal cortex system" in which each item competes via between-item inhibitions. In this architecture, the semantic relatedness effect was modeled by postulating the existence of mutual excitatory connections between semantically related items. However, because no mechanism responsible for the maintenance of serial order information was implemented, this model was strictly limited to simulate performance in free recall paradigms. Recently, Kowialiewski and Majerus (2020) implemented Dell's interactive activation model of language processing (Dell et al., 1997) within a WM architecture that takes into account how serial order information may be represented, i.e. the Start-End Model (Henson, 1998). These authors showed that the semantic relatedness and concreteness/imageability effects could be successfully modeled in immediate serial recall paradigms. At the same time, the WM architecture they used is only descriptive, and does not model WM maintenance processes in a realistic way. This limits the ability of this model to be extended toward a larger set of WM phenomena and experimental designs. Hence, although current computational models make a very good description of how semantic effects may influence overall WM performance, they nevertheless make a poor description of how items may actually be maintained, and especially how long-term memory knowledge may potentially affect and interact with maintenance processes.

Other models account well for the active maintenance processes taking place over time in WM. One such model is the Time-Based Resource Sharing model (TBRS, Barrouillet et al., 2004). This model considers that WM maintenance is constrained by two temporal factors: (1) the constantly decaying WM representations, and (2) the time available to restore such representations. In addition, the decaying WM representations are thought to be restored via the focus of attention, a central bottleneck limited to one operation at a time. This restoration process, called *refreshing*, supposedly occurs very rapidly, outside of explicit awareness (Camos et al., 2018), via constant switching between memoranda (Vergauwe & Cowan, 2015). The TBRS model offers an appealing explanation for classical cognitive load effects; when a distracting task embedded in the inter-item interval of a list to be remembered has to be performed, WM performance decreases proportionally to the attentional capture caused by the distractor (Barrouillet et al., 2011). Theoretically, this result is explained by assuming that when the attention is occupied by a distractor, decaying WM representations cannot be refreshed and are affected by the deleterious effect of decay. A computational implementation of this theoretical model, TBRS* (Oberauer & Lewandowsky, 2011) was shown to be able to account for several important well-established WM phenomena (Oberauer et al., 2018). These include cognitive load effects, serial position curves, omissions and transposition errors. Therefore, the

TBRS* model is an excellent potential architecture that can be used to model semantic effects in WM more realistically.

The original TBRS* model does not consider how long-term memory knowledge potentially interacts with WM maintenance. However, further studies have progressively acknowledged the need to include a rigorous description of long-term memory mechanisms. This accounts for a wider range of cognitive phenomena. For instance, Portrat et al. (2016) implemented a supplementary searching mechanism in long-term memory within TBRS* to model the maintenance and recall of chunks in WM. More recently, Lemaire and Portrat (2018) proposed a hybrid version of TBRS* that included an interference mechanism. This accounted for several interference effects, such as item-distractor similarity effects (Oberauer, Farrell, et al., 2012), which the original TBRS* model is unable to simulate. Despite these new refinements, the TBRS* model does not account for the presence of semantic effects in WM.

In this study, we integrated the core assumptions made by interactive activation models of language processing to model the semantic relatedness effect in TBRS*[1], by considering that items are directly activated in long-term memory. Theoretically, since semantically related items are supposed to reactivate each other in long-term memory, they should be less susceptible to the deleterious effect of decay, leading to a better recall performance than semantically unrelated items. Critically, we took advantage of this new integration to assess a hypothesis directly derived from those combined principles. Indeed, this new version of the model predicts that

[1] It is important to note that the leading cause of forgetting in WM is a matter of intense debate. While we chose the TBRS* architecture to account for the resource freeing hypothesis we are developing, we do not deny interference as a source of forgetting, but this question is out of the scope of the present paper.

semantic knowledge should free up attentional WM resources that can be reallocated to maintain more information. More specifically, we predicted that the presence of a semantic triplet within a list to be remembered should give a beneficial *proactive benefit*, by increasing WM performance for subsequent, semantically unrelated items. This prediction is derived from the assumption that the human cognitive system tries to reallocate attentional resources in an efficient manner (Lemaire et al., 2018). Because semantic triplets benefit from strong activation in the long-term memory knowledge base, fewer refreshing attempts should be required to keep them active. Then more attentional resources will be available for reallocation to maintain subsequent items to be remembered. We tested this *resource freeing hypothesis* directly on human participants, before assessing its plausibility using computational simulations.

Experiment

In this experiment, the semantic content of a list to be remembered was manipulated via the inclusion of a semantic triplet, such that half of the items were semantically related (e.g., leaf – tree – branch), while the other half was composed of items that were unrelated at the semantic level (e.g., wall – sky – dog). The triplet was presented either at the beginning (i.e. <u>leaf</u> – <u>tree</u> – <u>branch</u> – wall – sky – dog) or at the end (i.e. wall – sky – dog – <u>leaf</u> – <u>tree</u> – <u>branch</u>) of the list. These conditions were then compared to a neutral condition in which all items were semantically unrelated (e.g. hammer – jacket – horn – wall – sky – dog). According to the resource freeing hypothesis, we expect that the presence of a semantic triplet at the beginning of the list should have a beneficial proactive effect on WM performance. In other words, the presence of a semantic triplet in the first half of the list should free up attentional resources that can be reallocated to maintain the subsequent items of the list. Recall performance will thus be improved for these items, and this can be compared to the same items in the condition without a semantic triplet.

A similar phenomenon has been observed in previous studies for lists of letters (e.g., "PDFCHDL") (Portrat et al., 2016; Thalmann et al., 2018). However, in these studies, no *retroactive benefit* on WM performance was observed, i.e. an impact on the first half of the list when the triplet was presented in the second half. Based on these previous studies, we do not expect a retroactive effect upon using semantic triplets. This latter prediction derives from the resource freeing hypothesis. Effectively, when the semantic triplet is presented at the end of the list, the participants should become aware of the semantic triplet very late (i.e. from the fifth item). This may not leave enough time to free up WM resources.

Method

Participants. Thirty undergraduate students aged between 18 and 30 years were recruited from the university community of the Université Grenoble Alpes. All participants were French-native speakers, reported no history of neurological disorder or learning difficulty, and gave their written informed consent before starting the experiment. The experiment had been approved by the ethic committee of CER Grenoble Alpes: Avis-2019-04-09-2.

Material. We used a pool composed of 120 French words with a lexical frequency (count per million) of M_{log} = 2.899 and SD_{log} = 1.689. The words were 1–3 syllables long (M = 1.483, SD = 0.594) and were composed of 2–7 phonemes (M = 4.058, SD = 1.11). The stimuli were drawn from 40 different semantic categories, which included taxonomic (e.g., dog – wolf – fox) or thematic (e.g., sky – cloud – rain) relationships. Previous studies had shown that both types of semantic relationships are likely to impact WM performance in a similar way (Kowialiewski & Majerus, 2020; Tse, 2009). All the stimuli were recorded by a French native male speaker using a neutral voice. Each word was exported into an individual .wav file, whose average length was M = 455 ms and SD = 64 ms. Background noise was removed via the noise reduction tool implemented in Audacity.

There were 3 different experimental conditions, labeled as follow:

- In the T1 condition, the semantic triplet was presented in the first half of the list (Triplet in 1st half; e.g., leaf – tree – branch – wall – sky – dog).
- In the T2 condition, the semantic triplet was presented in the second half of the list (Triplet in 2nd half; e.g., wall – sky – dog – leaf – tree – branch).
- In the NT condition, no semantic triplet was presented and all items were semantically unrelated (No Triplet; e.g., hammer – jacket – horn – wall – sky – dog).

The experimental conditions were created by using the 40 semantic categories. Twenty triplets were used to create the T1 condition and 20 triplets were used to create the T2 condition. The 40 semantic categories were used again and randomly combined to create 80 triplets composed of unrelated words. Forty unrelated triplets were used to fill the second and first part of the T1 and T2 conditions, respectively. The remaining 40 triplets were combined to create the NT condition. In the T1 and T2 conditions, special care was taken to avoid the semantically unrelated words having an obvious semantic relationship with the semantic triplet itself. More specifically, we insured that, within each sequence, each word that composes the semantically unrelated triplet had no obvious semantic relationship with any item that belongs to the semantically related triplet. Each word appeared three times throughout the entire experiment: once in a semantic triplet, and twice in a semantically unrelated triplet.

The a priori semantic associations between the words were further assessed in an online survey, in which an independent group of 80 participants was invited to judge on a scale ranging from 0 to 5 to what extent pairs of words were semantically related. The pairs of words were drawn from the experimental lists, by extracting the adjacent words from each trial (e.g. given the list "ABCD", the pairs "AB", "BC" and "CD" were used). The total number of pairs to be judged was 1,018. Due to this large number of pairs of words, the participants were provided with only 250 pairs to judge, and were nonetheless free to stop the survey at any moment. Final data collection indicated that each pair was judged 12,659 times on average. A Bayesian independent samples T-Test (see the statistical analysis section below) confirmed that the a priori defined related and unrelated pairs did differ in term of semantic relatedness judgment, this difference being associated with decisive evidence ($M = 4.463$, $SD = .5$, and $M = .427$, $SD = .601$, for related and unrelated pairs, respectively, $BF_{10} = 9.809e+387$).

The immediate serial recall task was composed of sixty trials in total, twenty for each experimental condition (T1, T2 and NT). To avoid stimulus list effects, we generated 36 different versions of the lists to be remembered. We first generated three versions of the 20 trials that composed each experimental condition. Each version was then combined in a pairwise manner with each version of the other experimental lists to create 9 different versions of the lists. Each of these versions were duplicated, but the positions of the triplets within each list were exchanged (i.e. the T1 condition became the T2 condition; [1:3, 4:6] => [4:6, 1:3]), resulting in 18 different versions. Finally, for each version, a new one was created by re-ordering the items within each triplet (e.g. leaf – tree – branch – sky – wall – dog => branch – leaf – tree – wall – dog – sky). In the NT condition, all items within each list were randomly ordered.

To randomly order the items within each list or triplet, we avoided as far as possible across the entire experiment that the same item would be presented twice in the same position. Although this could not be totally avoided, it was nevertheless minimized by testing all possible permutations within a given trial. Within each version, the lists were presented in a pseudorandom order, such that the same semantic condition could not be repeated across more than three consecutive trials.

Procedure. Each trial began with the presentation of a white fixation cross displayed on a black background for 1000 ms, followed by the six-item memory list. The items were presented aurally at a speed of 1 item every 2 seconds. After the presentation of the list to be remembered, the participants were asked to recall out loud the items in the order in which they appeared. The participants were invited to substitute any item they did not remember by the word “blanc” (the French for “blank”). During each recall attempt, the numbers from 1 to 6 were successively displayed on the screen. When the first screen displayed the number “1”, the participants were invited to synchronize their oral response with a key press. More specifically, the participants were told that each time they began to recall an item or a “blank” out loud, they had to press the spacebar. The number on the screen was increased after each keypress. Once the last item had been recalled, the participants had to press the spacebar to initiate the next trial. Response times were automatically recorded by the computer, which allowed us to approximate recall latencies corresponding to each recall attempt.

The experimenter performed one practice trial to demonstrate the exact procedure to follow. The participants then performed three practice trials before the beginning of the main experiment. None of the stimuli in the practice trials were used in the main experiment. In addition, the stimuli in the practice trials were always semantically unrelated. Finally, the experiment was divided into two blocks, allowing participants to take a short break. Task presentation and timing were controlled using OpenSesame (Mathôt et al., 2012) run on a desktop computer. The auditory stimuli were presented via headphones connected to the computer, in a soundproof booth at comfortable listening level. Participants' responses were transcribed online by a research assistant blind to the main theoretical hypothesis, onto an electronic spreadsheet, and were also recorded using a digital recorder.

Scoring procedure. To determine the impact of the different semantic conditions (T1, T2, NT) on WM processing, recall performance was first assessed using a *strict serial recall criterion*. By this criterion, an item was considered to be correctly recalled only if it was recalled at the correct serial position. For instance, given the target sequence "Item1 – Item2 – Item3 – Item4 – Item5 – Item6" and the recall output "Item1 – Item2 – blank – Item3 – blank – Item5", only "Item1" and "Item2" would be considered as correct, resulting in a score of 0.333. Second, we used an *item recall criterion*, in which an item was considered as correct, regardless of its serial position. For the previous example, "Item1", "Item2", "Item3" and "Item5" would be considered as correct, resulting in a score of 0.667. While the strict serial recall criterion takes into account the ability to recall the position of a given item in a memory list, the item recall criterion is more informative concerning whether the item itself had been maintained in WM or not. This is important, because psycholinguistic effects mostly affect the ability to recall item information, rather than the serial order in which they had been presented (Majerus, 2009). It should be noted that a small but real deleterious effect of semantic relatedness upon memory for serial order information is observed (Tse et al., 2011).

Statistical analysis. We performed a Bayesian analysis, as this reduces Type-1 false error probabilities relative to frequentist statistics (Schönbrodt et al., 2017). The Bayesian approach has the further advantage of computing continuous values against or in favor of a given model, rather than deciding for the presence of an effect based on an arbitrary statistical threshold. Evidence in favor of a model is given by the Bayesian Factor (BF). This reflects the likelihood ratio of a given model relative to other models, including the null model. The null model and the effect of interest can be tested simultaneously, by directly comparing the alternative hypothesis against the null hypothesis, and vice versa. The BF_{10} is used to determine the likelihood ratio for the alternative model (H_1) relative to the null model (H_0), and the BF_{01} to determine the likelihood ratio for H_0 relative to H_1. We use the classification of strength of evidence proposed in previous studies (Jeffreys, 1998; Wagenmakers et al., 2011): a BF of 1 provides no evidence, $1 < BF < 3$ provides anecdotal evidence, $3 < BF < 10$ provides moderate evidence, $10 < BF < 30$ provides strong evidence, $30 < BF < 100$ provides very strong evidence and $100 < BF$ provides extreme/decisive evidence. In Bayesian ANOVAs, we performed Bayesian model comparisons using a top-down testing procedure, which first computes the BF value for the most complex model possible (i.e. the model including all main effects and all possible interactions). The BF value for each term is then assessed by directly comparing the full model against the same model, but by dropping the term under investigation. To minimize error of model estimation, the number of Monte Carlo simulations generated was set to $N_{iterations} = 100{,}000$. For some critical contrasts of interest, we also report the 95% Bayesian Credible Intervals using the highest density intervals of the sampled posterior distribution of the model under investigation ($N_{iterations} = 100{,}000$). All analyses were performed using the BayesFactor package (Morey & Rouder, 2014) implemented in R (R Development Core Team, 2008) using the default wide Cauchy prior distribution of $r = \frac{\sqrt{2}}{2}$.

On each graph we report the 95% Confidence Intervals for each mean. We follow the recommendations made by (Baguley, 2012). After correcting the data for between-subject variability (Cousineau, 2012; Morey, 2008), the confidence intervals of each mean j were computed using the following formula:

(1) $\hat{\mu}_j \pm t_{n-1,1-a/2}\sqrt{\frac{2J}{4(J-1)}}\hat{\sigma}'_{\hat{\mu}_j}$

where $\hat{\mu}_j$ is the j^{th} mean, $t_{n-1,1-a/2}$ is the two-tailed critical t value with $n-1$ degrees of freedom, J is the number of means included in the graph, and $\hat{\sigma}'_{\hat{\mu}_j}$ is the standard error of the j^{th} mean.

Results

We assessed recall performance as a function of semantic condition (T1, T2, and NT) and serial position (1 through 6). Using the strict serial recall criterion, we found decisive evidence supporting main effects of both the semantic condition (BF_{10} = 1.905e+18) and the serial position (BF_{10} = 2.497e+84). Likewise, the interaction term was associated with decisive evidence (BF_{10} = 5.826e+4). Similar results were observed using the item recall criterion, with decisive evidence being associated to both main effects of semantic condition (BF_{10} = 1.572e+14), serial position (BF_{10} = 8.743e+50) and the interaction term (BF_{10} = 5.175e+13).

The presence of an interaction suggests that the semantic condition impacted recall performance differently across serial positions, as shown in **Figure 1**. This interaction was further explored using Bayesian paired-samples T-Tests. To reduce the number of statistical contrasts, we averaged recall performance across the first (i.e. positions 1 through 3) and second (i.e. positions 4 through 6) halves of the lists, within each semantic condition.

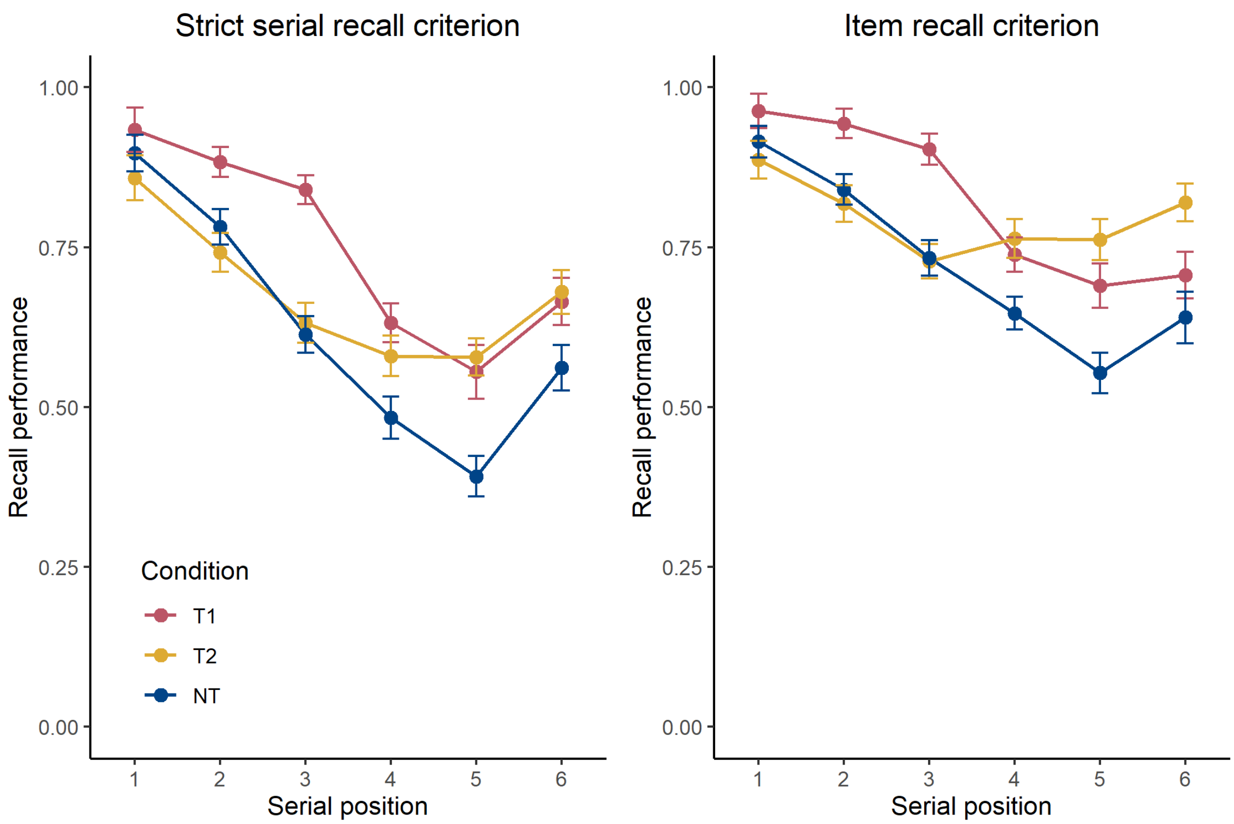

Figure 1. Recall performance as a function of serial position for each semantic condition. T1 = Semantic Triplet in the first half of the list. T2 = Semantic Triplet in the second half of the list. NT = No Triplet. Left panel: strict serial recall criterion. Right panel: item recall criterion.

Semantic relatedness effect. First, the overall impact of the semantic relatedness effect was assessed. Recall performance over the first half of the list was higher in T1 than in NT, and this difference was associated with decisive evidence (Strict serial recall criterion: BF_{10} = 2.604e+5, $CI_{95\%}$ = [0.756; 1.75], d = 1.323, M_{diff} = 0.122; Item recall criterion: BF_{10} = 7.179e+4, $CI_{95\%}$ = [0.693; 1.652], d = 1.228, M_{diff} = 0.107). Likewise, recall performance over the second part of the list was higher in T2 than in NT, and this difference was associated with decisive evidence (Strict serial recall criterion: BF_{10} = 1.551e+4, $CI_{95\%}$ = [0.586; 1.507], d = 1.117, M_{diff} = 0.134; Item recall criterion: BF_{10} = 1.48e+7, $CI_{95\%}$ = [1.017; 2.129], d = 1.636, M_{diff} = 0.169).

Hence, as has been classically observed, semantically related words were associated overall with better recall performance compared to semantically unrelated words.

Proactive benefit of the semantic triplet. Next, we assessed whether the semantic triplet had a proactive benefit on recall performance, as predicted by the resource freeing hypothesis. Critically, recall performance over the second half of the list was higher in T1 than in NT, and this difference was associated with decisive evidence (Strict serial recall criterion: BF_{10} = 4.178e+4, $CI_{95\%}$ = [0.661; 1.608], d = 0.189, M_{diff} = 0.139; Item recall criterion: BF_{10} = 2.282e+3, $CI_{95\%}$ = [0.48; 1.353], d = 0.98, M_{diff} = 0.099). Therefore, a proactive benefit of the semantic triplet on recall performance has been observed.

Retroactive effect of the semantic triplet. Finally, the retroactive effect of the semantic triplet was assessed. Recall performance over the first half of the list did not improve in T2 compared to NT, this analysis being associated with anecdotal evidence slightly favoring the null hypothesis (Strict serial recall criterion: BF_{10} = 0.446, BF_{01} = 2.244, $CI_{95\%}$ = [-0.574; 0.122], d = -0.248, M_{diff} = -0.021; Item recall criterion: BF_{10} = 0.557, BF_{01} = 1.794, $CI_{95\%}$ = [-0.611; 0.091], d = -0.28, M_{diff} = -0.019). This analysis suggests that the semantic triplet did not retroactively impact recall performance.

Discussion

To sum up the results of this experiment, we observed that the presence of a semantic triplet enhanced recall performance specifically for the semantically related items that compose the triplet, as had been observed classically in immediate serial recall WM tasks (Poirier & Saint-Aubin, 1996). The novelty of our experiment is that it showed that the presence of this semantic triplet has a proactive benefit on recall performance. In other words, the semantic triplet improved recall performance for subsequent, semantically unrelated items of the same list to be remembered. Interestingly, no retroactive effect on recall performance was observed. This pattern of results is consistent with what has previously been observed in studies using triplets composed of letters (e.g. "PDFCHDL") (Portrat et al., 2016; Thalmann et al., 2018).

Overall, these results support a resource freeing hypothesis, according to which semantic relatedness should free up attentional WM resources. Indeed, since semantically related items benefit from the fact that they reactivate each other (Dell et al., 1997; Hofmann & Jacobs, 2014; Kowialiewski & Majerus, 2020), their activation level is thought to be very high. Due to this high activation level, these items would require fewer refreshing attempts to counteract the deleterious effect of decay, thereby freeing up attentional resources that could then be reallocated to refresh other, less activated items of the list (i.e. the unrelated items). Notwithstanding the support provided by this experiment to the resource freeing hypothesis, the exact mechanisms that might be responsible for the proactive benefit of the semantic triplet have yet to be specified. In the next section, we described the implementation within TBRS* of some principles derived from activation-based models of WM, to investigate to what extent the mechanisms we think might be responsible for the resource freeing hypothesis are plausible.

Computational modeling

TBRS*: General architecture

The TBRS* model (Oberauer & Lewandowsky, 2011) is a fully interconnected neural network composed of two layers. One layer codes for positional information and the other one codes for item information (see **Figure 2a**). Positional information in the positional layer is represented in a distributed fashion across 54 units. Each position is coded by a subset of 9 units, also called a *positional marker*. Each position inherits a proportion of units from the previous adjacent position, defined by a probability P. The remaining units are randomly assigned to the positional vector. In other words, adjacent positions overlap to some extent, and this overlap decreases exponentially as their separation increases. Item information is represented in a unitary fashion within an 81-unit vector. The WM representations are stored within the weight matrix w that connects the items to positional units (see **Figure 2a**).

Due to this initial choice of implementation, which is potentially problematic as we will see later, the nature of the information itself (i.e. the item) and the position cannot be dissociated when stored in WM. The two elements can be considered as two faces of the same coin. At the beginning of a trial, the weight matrix w that contains the WM representations is empty (i.e. all values are set to zero), and the associations are formed through different phases.

Encoding. Encoding is performed by activating (i.e. setting the values to 1) the nodes in the item and position layers. This means that during encoding an item is co-activated with the positional marker corresponding to the current position. The values in the weight matrix w are then updated following a Hebbian learning rule:

(2) $\Delta w_{ij} = (L - w_{ij})\eta a_{pos} a_{itemj}$

where L is an asymptotic value, fixed at 1/9 to obtain values ranging from 0 to 1 during retrieval. The term $a_{posi} a_{itemj}$ is the product between activation of unit i in the position layer and activation of unit j in the item layer, whose values are either 0 or 1. Hence, a link within the weight matrix is created if two units in both layers are co-activated. Each w_{ij} connection is then updated by adding Δw_{ij} to its current value. The learning strength η is a scaling factor:

(3) $\eta = 1 - exp(-rt)$

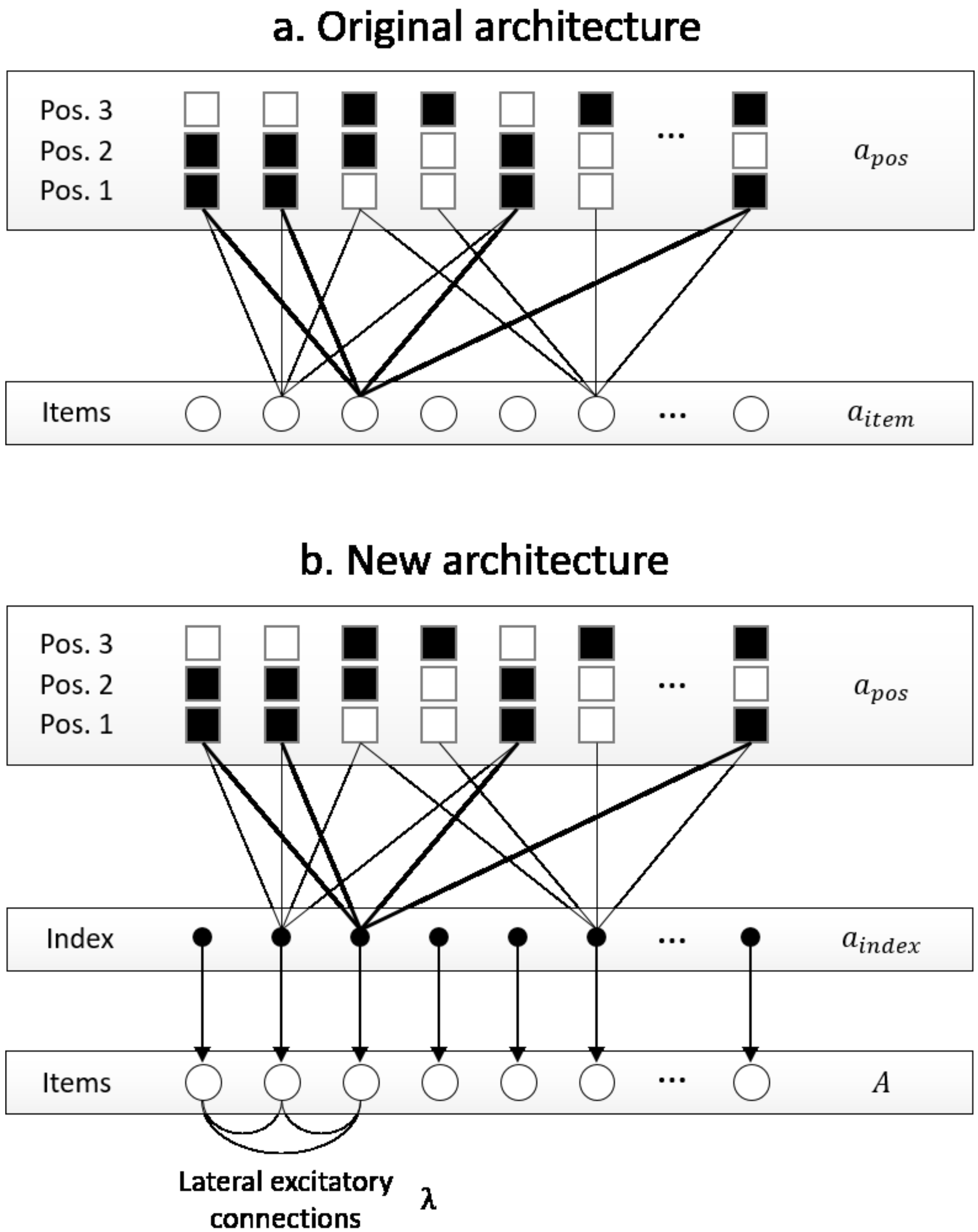

Figure 2. Illustration of the original TBRS* (a) and new TBRS*-S (b) architectures. Both architectures maintain serial order information within a positional layer (a_{pos}). In the original architecture, encoding is performed by creating associations between the positional and the item (a_{item}) layers. In the item layer, which makes a minimal description of item information, each node represents a given item and can take two possible values: 0 or 1. In the revised architecture,

encoding is performed by first activating information in the long-term memory, item layer (A). Each node in this new layer still represents a given item but takes continuous values between 0 and 1. During encoding, an association is created between the positional and the index layer (a_{index}) in order to indicate which information in the A layer is being encoded. The semantic relatedness is modeled within the A layer using lateral excitatory connections, through which activation spreads from one node to another.

This η parameter follows an exponential function; it increases as the time spent on encoding increases, and progressively reaches an asymptote. Encoding lasts until activation reaches 95% of the highest value. Since it is assumed that encoding takes about 500 ms (Jolicœur & Dell'Acqua, 1998), it gives a mean encoding rate R $=$ 6 (because $1 - exp(-6 * 0.5) = 0.95$). However, to model some variability, it is not the value R which is used but rather the outcome of a random draw from a normal distribution centered at R, with a standard deviation of 1. This term is called r. Implementation details of this procedure can be found in the Matlab code associated to Oberauer and Lewandowsky (2011) or in our own Julia code available on the Open Science Framework (https://osf.io/y386u/).

Maintenance. Immediately after an item is encoded, the model enters a dynamic balance state constrained by two phenomena: decay and refreshing. When the focus of attention is driven away from the WM content, WM representations decay and all the w_{ij} are updated:

(4) $\Delta w_{ij} = -w_{ij}\left(1 - exp(-Dt)\right)$

Decay is controlled by the decay rate D, and it depends on the time t during which the central bottleneck is occupied, either by encoding or refreshing another item, or by a distracting task. Importantly, during the time a WM representation is being encoded or refreshed, all other WM representations are affected by decay. During refreshing, the items are first retrieved (see below), then re-encoded using the same principles as in Eq. 2. Refreshing occurs through rapid switching between memoranda. In the original TBRS* model, each refreshing attempt lasts 80 ms, which is close to the empirical estimation of 50 ms (Vergauwe & Cowan, 2015).

Retrieval & recall. Before being recalled and/or refreshed, an item first needs to be retrieved. Retrieval is performed by cueing the weight matrix using positional markers:

(5) $a_{item\,j} = \sum_{i=1}^{n} a_{pos\,i} w_{ij} + n \text{ where } n \sim N(0, \sigma)$

In this formula, the activation $a_{item\,j}$ of item j is the sum of all the weights that connect item j to the positional unit i that correspond to the current position. The item most strongly associated to the cued position is then retrieved. In most cases, this is the item that was initially encoded. However, for the model to produce transposition errors (i.e. retrieving a wrong item), a zero-centered random Gaussian noise with a standard deviation σ is added to each activation value $a_{item\,j}$. Modeled this way, items associated to more similar positions (e.g., positions 2 and 3) are more likely to be transposed than most distant positions are (e.g., positions 1 and 5). In a second step, if the activation value of the retrieved item is below the retrieval threshold θ, an omission is produced.

During recall, retrieval is performed by cueing the required position. While an item is being recalled, all other WM representations decay following Eq. 4, but by assuming a recall time t_r = 0.5 s. After selection and recall of an item, response suppression is applied to the weight matrix using Hebbian anti-learning:

(6) $\Delta w_{ij} = -L.\, a_{pos\,i} a_{item\,j}$

When applying Hebbian anti-learning, the $a_{item\,j}$ vector retrieved from Eq. 5 is used. In the case of an omission, no response suppression is applied. Then the model moves on to the next position. It should be noted that the WM representations also decay during the production of omissions. An example of the time course produced by the model is displayed in **Figure 3**.

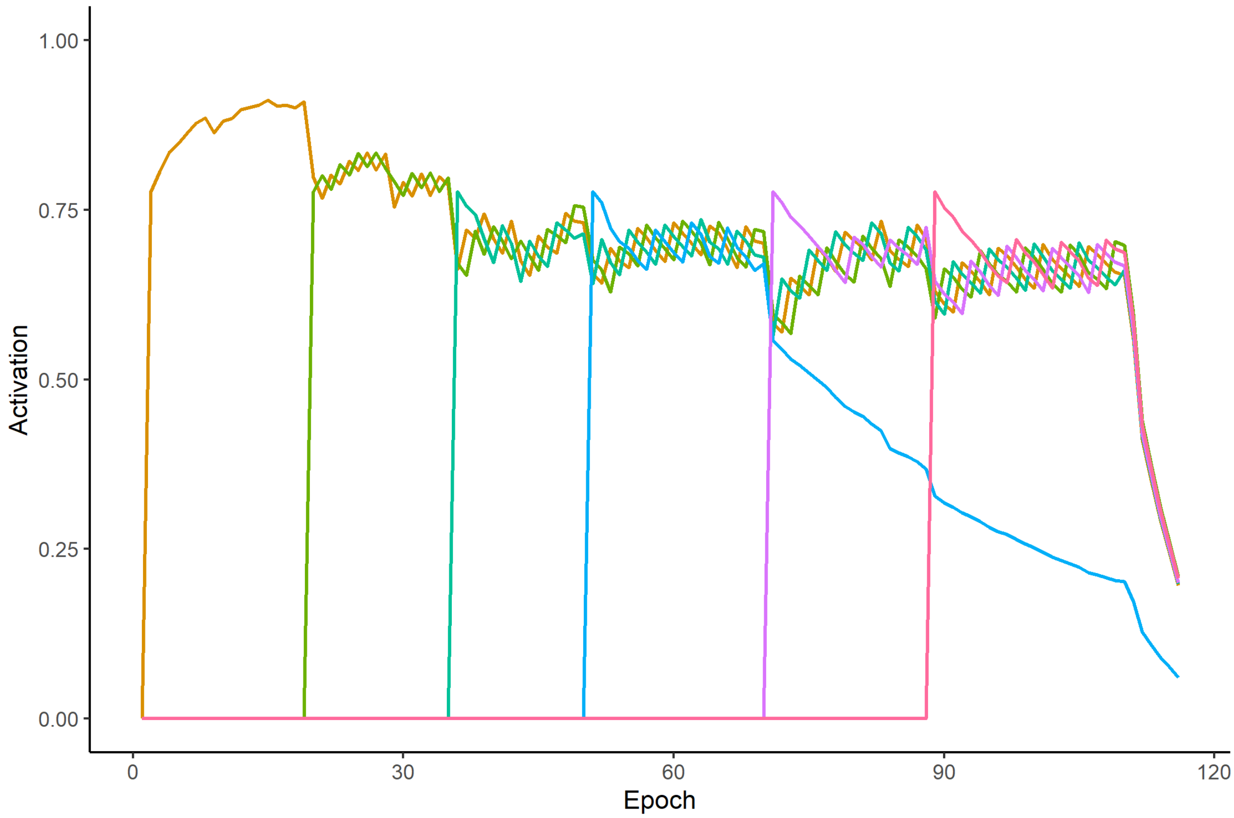

Figure 3. Activation values of the model across the different epochs of one trial.

Refreshing schedule. There are controversies as regards the refreshing schedule that participants may use during maintenance (Vergauwe et al., 2016). For instance, participants could refresh items cumulatively (1, 2 – 1, 2, 3 – 1, 2, 3, 4…) just like one rehearses verbal information (Tan & Ward, 2008), but they may use other schedules as well. In an extensive investigation of the TBRS* model, Lemaire et al. (2018) tested the ability of several refreshing schedules to fit 3 different datasets. Among the different schedules investigated, they found that a *Least Activated First* schedule provided the best fit and outperformed the cumulative refreshing one. In this refreshing schedule, the model refreshes in priority the least activated items. The rationale behind this mechanism is to consider that the human cognitive system is very efficient, and try to use the available resources for optimizing purposes. Accordingly, we used this refreshing schedule throughout our simulations.

Basically, the system performs a series of short refreshing episodes, provided there are no external events such as encoding a new item or recalling all items. Each of these episodes is devoted to refreshing a single item as mentioned previously. To select this item, the system scans each position and retrieves the most associated item for each one. The item to be refreshed is the least activated one. Its weights are thus strengthened and the system engages in the next refreshing episode. Scanning positions sequentially is probably not cognitively plausible but it is the way it is implemented on a von Neumann sequential computer. It does not preclude any brain-level parallel mechanism which is outside the level of description of our model.

TBRS*-S: A new architecture to model semantic relatedness

As mentioned in the introduction, although recent efforts have been made to model long-term memory phenomena in TBRS* (Lemaire & Portrat, 2018; Portrat et al., 2016), the model is unable to simulate between-item semantic relatedness on recall performance. One reason is that the item layer does not incorporate or describe the phenomena that may occur in long-term memory, and especially the complexity of interactions occurring in the linguistic knowledge base. At a theoretical level, the semantic relatedness effect can be explained by assuming that semantically related items reactivate each other, for instance by spreading the activation in a semantic network from one node to another (Hofmann & Jacobs, 2014; Neely, 1977). In the following paragraphs, we describe the stages of a first tentative adaption of the TBRS* architecture to account for the complex interactions between semantic knowledge and WM maintenance. We first adapt the TBRS* architecture by assuming the existence of this basic spreading activation principle. We then consider the dissociation between the ability to maintain item and serial order information, without which semantic relatedness effects cannot be modeled. Finally, a new, more psychologically plausible recall mechanism was implemented.

Modeling spreading activation. In a first attempt, we tried to keep the architecture as close to the original one as possible. We implemented a model in which the encoding and/or refreshing of a memorandum leads to the automatic co-activation of all semantically related items, with this co-activation being constrained by a new parameter, λ. More specifically, we assumed that items are represented within a semantic network, with semantically related items linked by a connection of strength λ[2]. For instance, consider the two semantically related memoranda “item A” and “item B” that need to be encoded in position 1 and 2, respectively. The activation of “item A” in position 1 also activates “item B”, scaled by λ, resulting in the creation of an association for both items at this current position 1.

We reasoned that this implementation should enhance recall performance by reducing the rate of omissions, because semantically related items would be more strongly encoded overall. Indeed, because adjacent positions share in average a proportion P of positional markers, if item B is encoded in position 1, it would also be encoded to some extent in position 2, resulting in higher activation levels at the time of retrieval. This intuitive interpretation was met in the model in a condition where all the items were semantically related: When λ varied between 0 and 0.5, the proportion of omissions decreased linearly from 0.163 to 0.046. However, this reduction of

---

[2] This way of representing semantic relationships in an all-or-nothing fashion appears to allow the description of the semantic effects we are interested in. This could be extended later on to take into account various degrees of semantic relationships instead of only one.

omission errors was accompanied by a strongly deleterious impact on the maintenance of order information, which produced a dramatic drop of performance when a strict serial recall criterion was considered (from 0.652 to 0.262). Increasing λ also slightly increased repetition errors. Hence, implemented this way, the presence of semantic relationships at the whole-list level *decreased* recall performance[3].

This drop of performance is inherently linked to the model's fundamental properties. Since there is no way to store item information without also storing positional information, the obligatory co-activation of the semantically related items also results in their association to a wrong position, leading to a strong increase of transposition errors, far beyond the small advantage observed at the item level. The output of this first attempt of modeling semantic relatedness in TBRS* is reported in **Appendix A**.

Following this first modeling attempt, it appeared that the original TBRS* architecture is severely limited due to the lack of dissociation between the item and positional information stored in WM. Indeed, although the two pieces of information are coded within distinct layers, once encoded they form a unique WM representation (i.e. the weight configuration within the *w* matrix). At a theoretical and empirical level, this is problematic because previous studies have shown a dissociation between the ability to recall item and serial order information in WM (Gorin et al., 2016; Henson et al., 2003; Majerus, 2009, 2013, 2019), suggesting the existence of different mechanisms to maintain both type of information. This means that using the original

---

[3] We also implemented a distributed version of this model, with semantically related items sharing similar nodes at the item level. This model behaved in a similar way.

TBRS* architecture with a basic spreading activation principle is not suited to capture the semantic relatedness effect, as revealed by our first modeling attempt.

Dissociating item and positional information. To solve this problem, the model must be adapted to meet the assumptions made by the Embedded-Processes model of WM (Cowan, 1995, 2001), which is particularly well-suited to explain the impact of long-term memory knowledge in WM, especially within a decay-based architecture. According to this theoretical framework, items activated in long-term memory are constantly decaying, unless they can be actively maintained using the focus of attention. Hence, the Embedded-Processes model assumes that what is maintained via the focus of attention is the sustained temporary activation of the items themselves, not a temporary WM representation. This principle is now adopted in the new TBRS*-S (S = Semantic) architecture we propose, whose general structure is displayed in **Figure 2**b. More specifically, we assumed that items are associated with their own activation values A. Encoding is performed by directly activating the node of the current item:

(7) $\Delta A_i = (1 - A_i)\eta$

This formula is identical and follows the same rules as for the item-position connections in Eq. 2, with the exception that activation of each item i is now represented within the vector A, and the activation of each item reaches an asymptote of 1.

To model the semantic relatedness, we chose to approximate the assumptions made by interactive activation models of language processing (Dell et al., 1997; McClelland & Rumelhart, 1981), by assuming that semantically related items constantly reactivate each other. At a theoretical level, this reactivation process could occur via redundant activation feedbacks between the lexical and semantic levels of language processing (Dell et al., 1997), or by postulating the existence of between-item lateral excitatory connections (Hofmann & Jacobs, 2014). A recent implementation of such interactive activation models showed that both phenomena produce similar outcomes on WM recall performance (Kowialiewski & Majerus, 2020). Accordingly, in the present study we only assumed the existence of lateral excitatory connections. This solution is more efficient computationally speaking, and is sufficient for the purpose of the present study. Therefore, once an item gains a given amount of activation at the moment of encoding and/or refreshing, all other semantically related items A_j also receive a proportion of this activation via spreading activation:

$$(8)\ A_{j,t} = min\left(\left[1, \left(A_{j,t-1} + A_{i,t-1}\lambda\right)\right]\right)$$

Where λ is the value of the weight that connects the semantically related items A_j and A_i, and t refers to the timestamp of the ongoing iteration. We also included a min function to ensure that activation values will not exceed 1. This modeling choice is more generally consistent with semantic priming effects, whereby the presentation of a prime (e.g., "boat") facilitates the processing of a target (e.g., "captain"), as classically observed in the psycholinguistic domain (Zwitserlood, 1989). In addition, in agreement with interactive activation models we approximated the persistence of this spreading activation during decay by updating A_i:

$$(9)\ \Delta A_i = (1 - A_i).tanh\left(\lambda \sum A_j\right)$$

In this equation, the activation value A_i of item i at time t is increased with the activation spreading from all its semantically related neighbors A_j. The second factor is scaled by a hyperbolic tangent, as classically made to ensure that the final activation of A_i will not exceed 1. Note that the total activation received by A_i at time t is computed *before* decay is applied, and is actually applied to A_i *after* decay.

Modeled this way, the new TBRS*-S architecture should be able to handle the influence of semantic relatedness on recall performance. However, recent theoretical debates have highlighted the fact that WM cannot rely exclusively on activated long-term memory (Norris, 2017, 2019). It also needs temporary WM representations. One particularly critical aspect is the maintenance of serial order information. Without specific mechanisms to maintain the serial order of a sequence to be remembered, the model would have no information whatsoever regarding the relevant representation to recall at a given position. Therefore, the maintenance of serial order information in this new implementation follows the same principles as the original TBRS* architecture, i.e. via the creation of item-position connections. These item-position connections are automatically created when encoding an item and they are updated during maintenance. They are not influenced by activation values in long-term memory. During retrieval, these item-position connections are used to select the index of the corresponding item in long-term memory. Once the index is selected, retrieval is constrained by the activation level of item A_i in long-term memory. The item is correctly retrieved if its activation value in the long-term memory layer is above the retrieval threshold θ. Otherwise an omission is produced.

This implementation assumes that an item can be recalled in WM with little or no knowledge about its position, provided that its temporary activation value in long-term memory is sufficiently high. Inversely, knowledge about the position of a given item can be retrieved, even with poor information regarding the nature of the information itself. In addition, this implementation solves the problem of storing multiple tokens of the same representation, as also mentioned by Norris (2017). If item A is presented in the first and fourth positions, it is still possible to encode it twice, because the item-position connections can be created multiple times over the same item in long-term memory. Indeed, it is not the long-term memory representation itself that is associated to a position, but rather a unique index on that representation.

Weighted retrieval mechanism during recall. In addition to these substantial changes at the level of WM representations, we considered a new recall mechanism that would significantly enhance the model's behavior. In the original TBRS* architecture, items are recalled one by one using the positional markers that represent each position in isolation. Due to this implementation, an item can be recalled at a wrong position at the beginning of the list, even though it is more strongly associated to its original position. Cognitively speaking, it makes sense that participants would notice that items never appear twice in a list. Therefore, they may decide not recall an item at one position, because this item is much more associated to another position. For instance, consider "item A" and "item B" encoded in positions 1 and 2, respectively. When trying to retrieve an item at position 1, the system may sometimes retrieve a wrong item because of noise, in this case "item B". However, because the association between "position 2" and "item B" is so strong and unambiguous, it makes sense to consider that "item B" will not compete for selection when trying to retrieve "item A" at position 1.

In TBRS*-S, given a position at which retrieval has to be performed, all items with a lower activation value at that position than at another position are excluded from the competition. If all items are excluded, the model produces an omission. This implementation requires a retrieval to be performed for each of the positions. There are probably more cognitively plausible retrieval implementations, in which the existing strong associations directly pops up and inhibits it as a candidate, but we stuck to a simple implementation for this retrieval mechanism. In addition to being more cognitively plausible, this *weighted retrieval* mechanism enhanced the general model's behavior via the production of realistic serial position curves across the strict serial recall criterion and the item recall criterion, but also critically the omission rate. With the standard recall mechanism, the omission rate was very low over the first three positions, and increased substantially across positions 4 through 6, as can be seen in **Appendix B**. This phenomenon is not observed in the empirical data. The most activated item is always considered, so in the original model little room is left for omissions at the beginning of the list. Indeed, most of the time there will be a very strongly activated item. This weighted retrieval mechanism fixed this issue, as we will see in the next section.

Table 1. Range of values explored within the grid search. Note that lambda (λ) has been estimated separately.

Fixed parameters					
Parameter	Meaning	Value			
s	Standard deviation of processing rates	1			
Te	Mean duration of an encoding episode	0.5			
Tr	Mean duration of a refreshing episode	0.08			
$Trec$	Mean duration of a recall episode	0.5			
n	Number of items in long-term memory	81			
ISI	Inter-stimulus interval	1.5			
Free parameters					
Parameter	Meaning	Minval	Maxval	Steps	Best
R	Processing rate	1	9	1	**3**
σ	Noise added at retrieval	0.0	.1	.01	**.01**
θ	Retrieval threshold	0.0	.3	.025	**.25**
D	Decay rate	.1	.9	.1	**.3**
P	Overlap between positions	.2	.8	.1	**.6**
λ	Lateral connections	0.0	.05	.001	**.013**

It is important to note at this point that despite these substantial theoretically-driven changes in the original TBRS* architecture, we did not add any extra free parameters, apart from λ. Without λ, it would be impossible to manipulate the between-item semantic relatedness.

Parameter estimation. To identify the set of parameters that would reproduce basic WM behaviors, namely primacy and recency effects as well as the amount of recalled items observed in the empirical data, parameters were estimated using a grid search method exploring 81,081 points of the parameter space. The grid search method considered 5 different free parameters across a wide range of plausible values (see **Table 1**). Each combination of parameters was estimated by running the model 1,500 times in the neutral condition only, with no semantic relatedness (i.e. the NT condition, with $\lambda = 0.0$). To reproduce effects and not only magnitudes, the objective function used consisted in giving a score to each model. More specifically, each configuration was rewarded with a notation system as follows.

Ten points were attributed if the configuration of parameters correctly produced the primacy effect:

- Four points for Position 1 > Positions 2 through 5
- Three points for Position 2 > Positions 3 through 5
- Two points for Position 3 > Position 4 through 5
- One point for Position 4 > Position 5

One point was attributed if the configuration of parameters correctly produced the recency effect:

- Position 6 > Position 5

In addition, if the average performance level produced by a specific combination of parameters was below 0.4 or above 0.8 (i.e. floor and ceiling effects, respectively) using the strict serial recall criterion, this combination was automatically discarded from the selection.

This notation system resulted in a maximum possible 33 points: 11 points for each recall criterion, including strict serial recall, item recall, and omissions. It should be noted that the production of omission errors is a particularly critical aspect of the resource freeing hypothesis, because this hypothesis predicts that the presence of a semantic triplet should prevent other items from being lost (i.e. an activation level below the retrieval threshold). Among all combinations of parameters that produced the highest score, we selected the combination that minimized the Root Mean Squared Error (RMSE). The RMSE was computed across 18 data points: the 6 serial positions across the three recall criteria considered.

After the set of parameters minimizing the RMSE was identified, we fitted the semantic conditions on this set of parameters by systematically varying λ across a range of plausible values (see **Table 1**), and this was done by running the model 10,000 times[4] for each value of λ. To select the appropriate value of λ, we averaged the recall differences between the NT and T1 conditions across positions 1 through 3, and between the NT and T2 conditions across positions 4 through 6. This resulted in a mean difference that represents the overall impact of the semantic triplet on recall performance[5]. The value of λ that minimized this difference with the empirical data was then selected. Hence, the proactive benefit of the semantic triplet was *never* used during the selection of parameters. We reasoned that the presence of a proactive benefit (and the equivalent absence of a retroactive effect) should be a direct consequence of the model's general behavior, not the result of a specific combination of parameters across the parameter space.

---

[4] Since only 51 values of λ were estimated, we increased the number of simulations per estimation.

[5] With the strict serial recall criterion, the model did not reproduce well the serial recall performance. Therefore, the item criterion was used for this fitting (see simulation results).

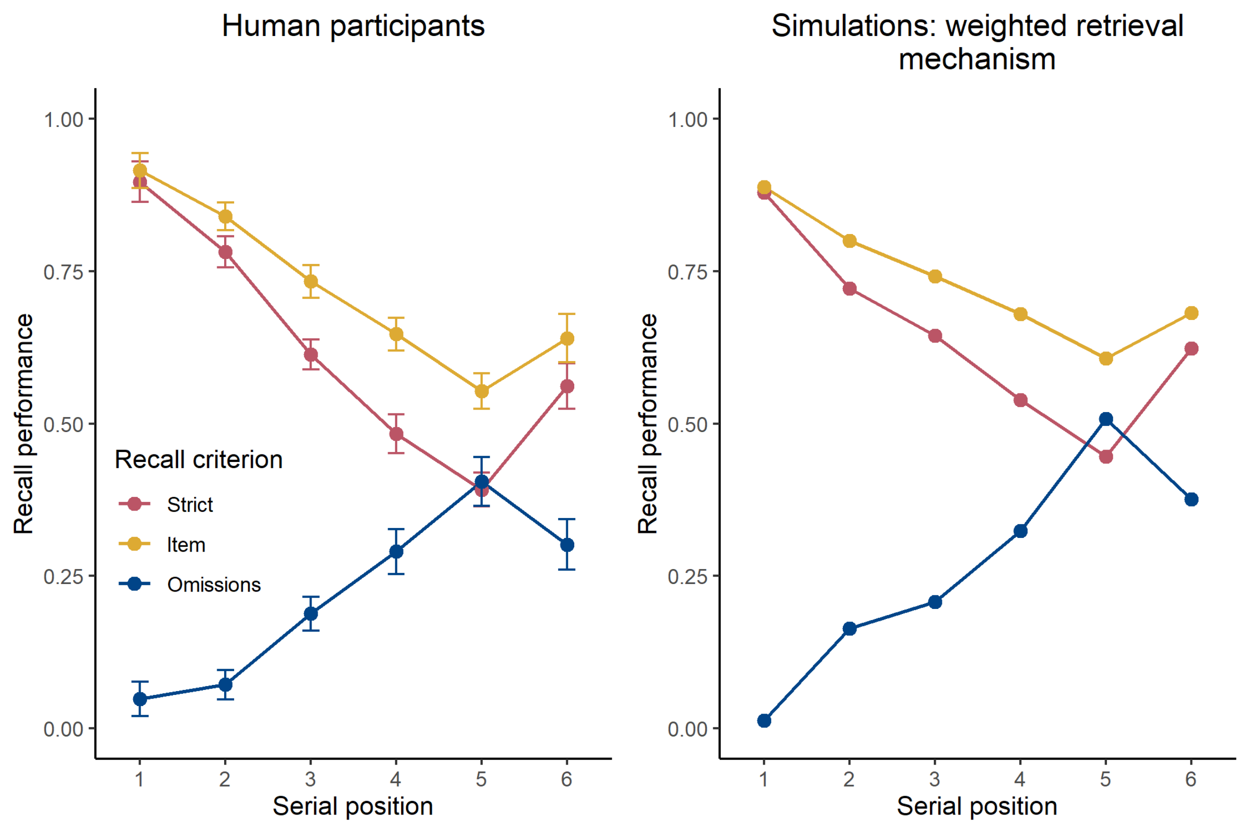

Figure 4. Recall performance for the three recall criteria (strict, item and omission errors) as a function of the serial position observed in humans (left panel) and produced by the model using the Least Activated First mechanism (right panel), NT condition only.

Simulation results

Among the 81,081 sets of parameters estimated in the grid search method, a total of 584 combinations resulted in a score of 33 points. These models, when fitted against the NT experimental data had an average RMSE of 0.130, with SD = 0.037. The set of parameters that minimized the error was associated with a RMSE of 0.052 (see **Table 1**)[6]. As can be seen in **Figure 4**, the model was nearly indistinguishable from the empirical data, and successfully reproduced the pattern of performance observed across the different recall criterion considered, including primacy and recency effects. Hence, the model was able to correctly capture recall performance in a general manner. To model semantic relatedness, we varied the value of λ while keeping all other parameters unchanged. We found that a value of 0.013 best fitted the empirical data. The evolution of the RMSE as a function of λ is provided in **Appendix C**.

Semantic relatedness effect. As can be seen in **Figure 5** (lower panels), the model successfully captured the general impact of semantic relatedness on recall performance. Compared to the NT condition, semantically related items were associated with better recall performances at the item level. However, when a strict serial recall criterion was considered, the pattern of results differed slightly. We will return to this issue below.

Proactive benefit of the semantic triplet. When omission errors and the item recall criterion were considered, the model correctly produced the proactive benefit of the semantic triplet. Using the strict serial recall criterion, this was also observed, albeit to a lesser extent. In fact, the proactive benefit of the semantic triplet increased in a linear fashion with the value of λ (see **Appendix C**).

---

[6] In this estimation, a very small value of σ (.01) minimized the error, suggesting that little noise was required to produce transposition errors. In fact, the σ parameter is not the sole source of transposition errors. Other factors contribute to the production of transpositions, including the stochastic sampling of the R parameter (see equation 3) and the overlap P between contextual markers.

Retroactive benefit of the semantic triplet. Similarly, the model successfully predicted an absence of retroactive benefit of the semantic triplet on recall performance, and this was consistent using the strict serial recall criterion and the item recall criterion.

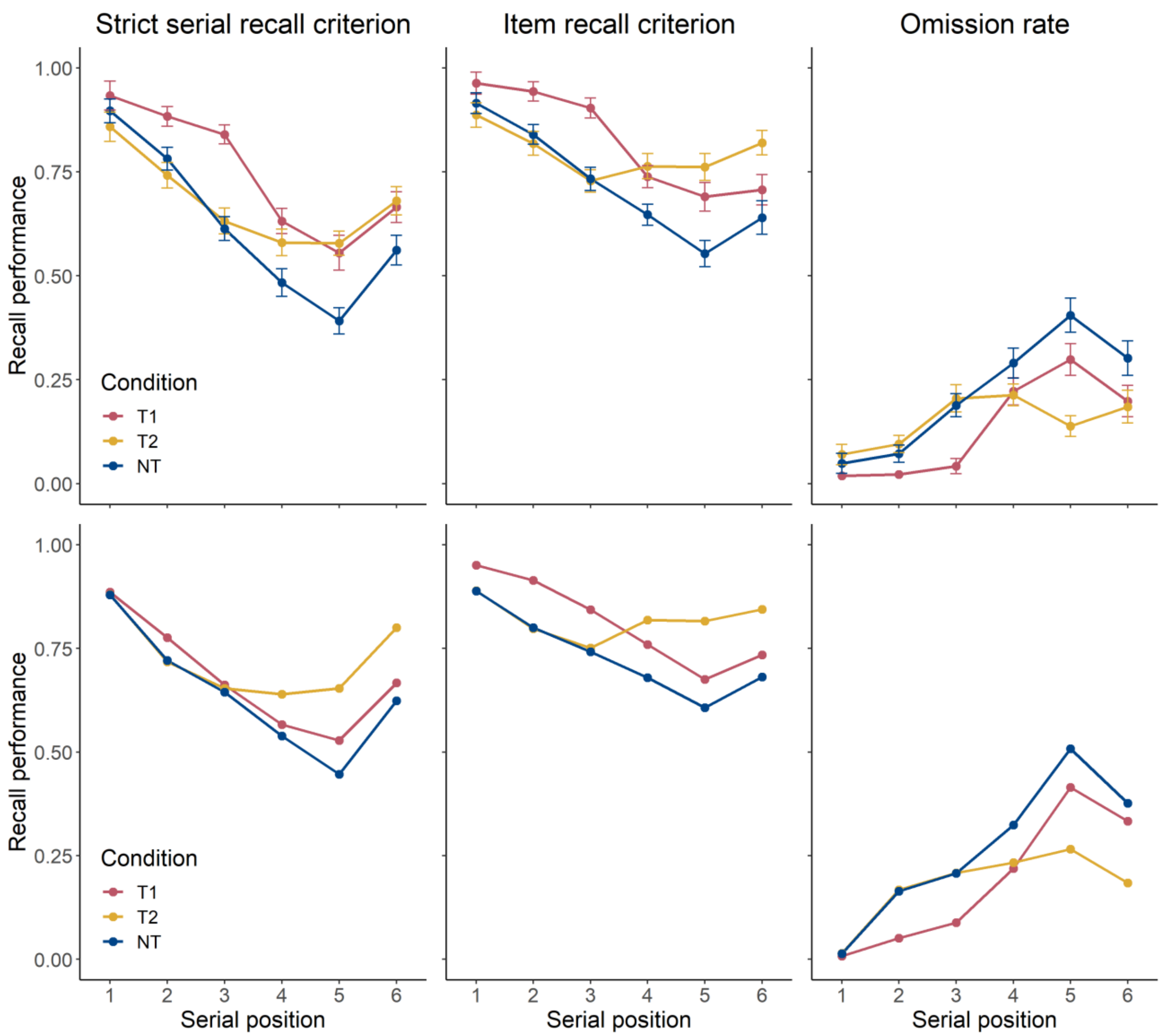

Figure 5. Recall performance as a function of serial position for each semantic condition (T1, T2 and NT). From left to right: Strict serial recall, item recall and omissions criteria. Upper panels: human subjects. Lower panels: TBRS*-S.

Model's diagnosis

Overall, the model behaved as expected following the resource freeing hypothesis. In the following paragraphs, we propose to describe a deeper investigation of the model's different components to analyse why it produced this behaviour.

Least Activated First mechanism. One of our main critical assumptions as to why the model should behave in agreement with the resource freeing hypothesis involves the Least Activated First mechanism. We assumed that since semantically related items have an overall higher activation level than unrelated items, the subsequent semantically unrelated items should be refreshed more often, with the Least Activated First mechanism being directly responsible for this redistribution of attentional resources. If this supposition is true, the proactive benefit of the semantic triplet should no longer be observed when using a cumulative refreshing schedule. This prediction was indeed met in the model, as can be seen in **Figure 6**[7], suggesting that the Least Activated First mechanism played an essential role to free up attentional resources.

[7] We re-estimated the value of λ (= 0.01) when simulating the data using the cumulative refreshing schedule. A higher value of λ did not change the absence of proactive benefit. The remaining parameters were kept identical.

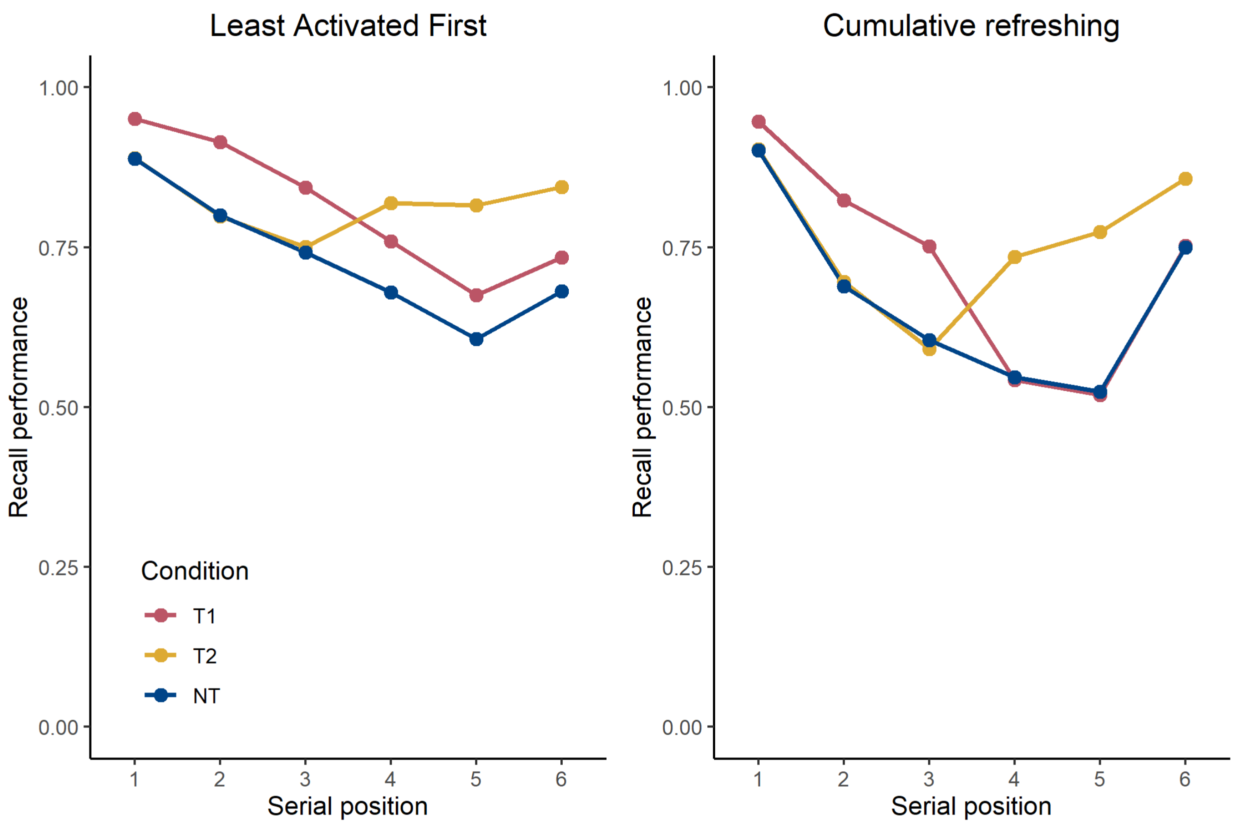

Figure 6. Pattern of recall performance produced by the model across the different semantic conditions (T1, T2 and NT) when using the Least Activated First mechanism (left panel) or when considering a cumulative refreshing schedule (right panel). Item recall criterion only.

Pattern of refreshing episodes. To understand the impact of the Least Activated First mechanism on WM performance as a function of the semantic condition, **Figure 7** should be examined. This displays the pattern of refreshing episodes that directly follows the encoding phase of the last three items, averaged across 10,000 simulations. Each panel represents the average number of refreshing episodes (y axis) over each item (x axis) during the free period of time directly following the encoding of each newly presented item, starting from item 4[8]. It can clearly be seen that the number of refreshing episodes over each item is unevenly distributed across the semantic conditions. In the T1 condition, items 4, 5 and 6 progressively benefit from more refreshing episodes to the expense of items 1, 2 and 3. In the T2 condition, the pattern is reversed: Items 4, 5 and 6 are subject to fewer refreshing episodes, and these refreshing episodes are reallocated toward items 1, 2 and 3.

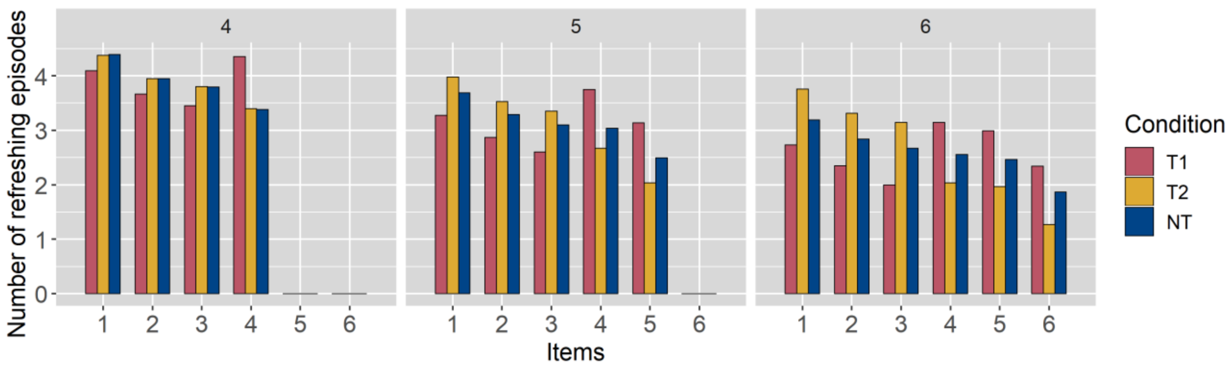

Figure 7. Mean number of refreshing episodes (across 10,000 simulations) over the different items for each semantic condition (T1, T2, NT). Each panel (numbered 4, 5 and 6) represents the maintenance phase that directly follows the encoding of a given item. For instance, panel 5 represents the maintenance phase between the encoding of item 5 and item 6.

Understanding the absence of retroactive effect. These deep investigations of the model's behavior demonstrate the role of the Least Activated First mechanism during the reallocation of attentional WM resources, and how this reallocation process interacts in a complex manner with

[8] During the free time that follows the encoding of the first three items, the refreshing episodes are evenly distributed, and hence are not informative.

semantic relatedness. However, the absence of a retroactive effect of the semantic triplet on recall performance seems at odds with the pattern observed in **Figure 7**. Indeed, in the T2 condition, once the items that compose the semantic triplet are presented, the attentional resources begin to be reallocated toward items 1, 2 and 3. All things being equal, this attentional reallocation should have led to a retroactive impact, an absent pattern in the model's output. To understand why, **Figure 8** displays the proportion of trials (over 10,000 simulations) for which the items were forgotten. These are the activation values for the T2 and NT conditions in long-term memory below the retrieval threshold, after the end of each maintenance phase (i.e. just before encoding the next item).

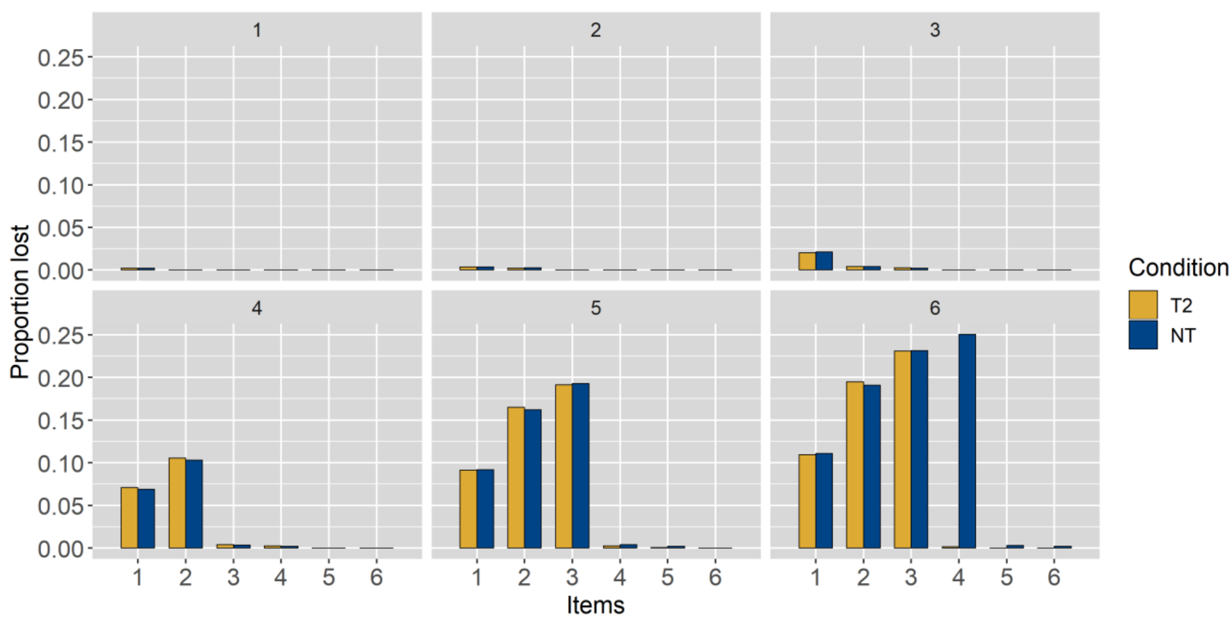

Figure 8. Proportion of trials (across 10,000 simulations) for which a given item has been lost, shown for the T2 and NT semantic conditions. Each panel represents the maintenance phase that directly follows the encoding of an item.

During the first maintenance phase, the first item was hardly ever forgotten, a pattern that remained relatively constant across the first three maintenance phases. After the fourth maintenance phase, WM overload began, and it became difficult to maintain all items. During the fifth maintenance phase, most of the items that could potentially be saved through attentional reallocation (i.e. those forgotten in the NT condition) were already lost. This is shown in **Figure 8,** fifth panel, by the equivalent number of forgotten items in NT compared to T2. At this stage, the semantic relatedness only started to affect recall performance. After the last maintenance phase (i.e. just before recalling the items, sixth panel), semantic relatedness did not prevent the first items of the list from being lost. If the semantic triplet did prevent any item from being lost, we would have expected a decrease in T2 compared to NT over items 1-2-3 in the sixth panel, which is clearly not the case. This is because there is nothing left to be saved; the only items that may benefit from the reallocation of attentional resources are those that are strongly enough activated to survive this far. Put another way, there is no retroactive benefit of the semantic triplet, because by the time that the attentional reallocation potentially allows these items to be maintained, a large part of them are already lost. The only items that could benefit from this reallocation process are those that are already well-above the retrieval threshold.

Illustrated in a more concrete example, let's suppose the last author of this study is juggling with a bunch of balls, but her lack of expertise only allows her to juggle with 3 balls. In this imaginary example, the balls represent the items, throwing a ball in the air is the equivalent of refreshing, and gravity represents decay. Each time an item is encoded, the second author is throwing a ball at her and she has to deal with a new ball. At the beginning, it is really easy to juggle with 1, 2 or even 3 balls. When a new ball is added, it becomes too difficult to follow the rhythm, and one ball has to be dropped from the game. Overall, recall performance (i.e. the total number of balls one is able to juggle with) will be equivalent to around 3 items, regardless of set size. Now let's suppose she is given a new type of "magic balls". These balls are lighter and are also less subject to gravity. This makes them easier to juggle with. These are the semantically related items. If three of these magic balls are introduced at the beginning (i.e. the T1 condition), it gets really easy to juggle with them. So easy that this leaves a lot of time to properly juggle with a new bunch of three balls. This produces the proactive effect observed in the data. Now let's suppose these magic balls are introduced at the end (i.e. the T2 condition). It is very likely that some of the regular balls would have already been dropped at that time. The introduction of the new magic balls would not save those that are already dropped. This is very similar to what happens in the model.

Discussion

In these series of simulations, we proposed a new TBRS*-S architecture that integrates the core principles assumed by activation-based models of WM. More specifically, we provided an item layer in which the decaying long-term memory representations were directly activated. This activation was independent from the processes involved in the maintenance of serial order information. To model the semantic relatedness effect, we implemented lateral excitatory connections between semantically related items, through which items constantly reactivated each other in long-term memory. This new integration successfully captured the overall recall advantage usually observed for semantically related over semantically unrelated items.

Furthermore, when combined with the Least Activated First mechanism which assumes that the least activated items are refreshed as a priority, the model handled the proactive benefit of semantic triplets very well. A closer inspection within the model's behavior suggested that this proactive benefit was directly caused by the interaction between the Least Activated First mechanism and the high activation level of semantically related items. However, this pattern was only observed when using an item recall criterion that does not take the model's ability to recall serial order information into account. The proactive benefit of the semantic triplet was also accompanied by a deleterious impact to recall serial order of the triplet itself, suggesting that there might be an issue in the way serial order information is represented and maintained in the current implementation.

Finally, the model also successfully captures the absence of retroactive impact of semantic triplets. A fine-grained diagnosis of the model explained the reason for this pattern. When the semantic triplet began to impact WM maintenance, the items most likely to be forgotten were already lost during the inter-item interval because of WM overload. Hence, it was impossible to maintain these items through attentional reallocation. Instead, attentional resources benefited items that were active enough to survive.

Overall, simulations show the plausibility of the resource freeing hypothesis to account for the results collected in the experiment. However, so far this study hinted exclusively at a role of the maintenance phase to account for the semantic triplet proactive effect. Alternatively, the recall phase may play a role to account for proactive effects (Cowan et al., 1992, 2002). In the next section, we described our extensive analysis of participants' recall latencies, and how these may impact WM performance across the different semantic conditions we manipulated.

A closer inspection of recall latencies

So far, the recall process was relatively underspecified within current implementations of TBRS*. This can be attributed to the fact that researchers neither measured nor took into account recall latencies in verbal WM tasks in a systematic manner. However, recall latencies are critical for decay-based architectures, including TBRS*. Many errors produced by the model can only be attributed to the time it takes to recall the items, during which not only do items decay, but also maintenance via refreshing is prevented. Empirically, recall latencies differ across serial positions, and researchers in the WM domain even consider recall latencies as a tool that should be used to compare different competing computational models (Farrell & Lewandowsky, 2004; Hurlstone & Hitch, 2015). It appears that some stimuli take longer to be recalled, such as long compared to short words (Cowan et al., 1992), or nonwords compared to word stimuli (Walker & Hulme, 1999). In the psycholinguistic domain, studies have shown reliable semantic priming effects on response latencies in tasks involving word production or shadowing (Python et al., 2018; Slowiaczek, 1994). It is therefore very likely that the semantically related items in the present experiment were also associated with shorter recall latencies. Critically, any model that assumes the existence of time-based forgetting should also assume that items that take less time to be recalled should produce a proactive benefit on recall performance. Indeed, given a sequence “ABCDEF” the to be remembered, if “ABC” takes less time to be recalled than a control condition does, the items “DEF” are thought to decay to a lesser extent, and should therefore be associated with higher recall performance. In this section, we explore whether recall latencies can explain the semantic effects observed in the experimental data.

Empirical data

Data preprocessing. In the following analysis, recall latencies were first log-transformed (natural logarithm) to reduce the skew of the latency distribution, as reported in **Appendix D**. We were first interested in overall recall latencies across serial position and semantic conditions, regardless of response type (correct responses, omissions, transpositions). Therefore, for each participant we took the median of these log-transformed recall latencies for each position and each semantic condition as dependent variables[9]. Also, cumulative recall latencies across serial

[9] Using the mean of log-transformed recall latencies did not significantly change the overall results.

positions were considered, by applying the log-transformation after summing the data across the different serial positions. For illustrative purposes, we used median raw recall latencies when plotting the results (see **Figure 9**), as log-transformed values are likely to be poorly informative from a psychological perspective.

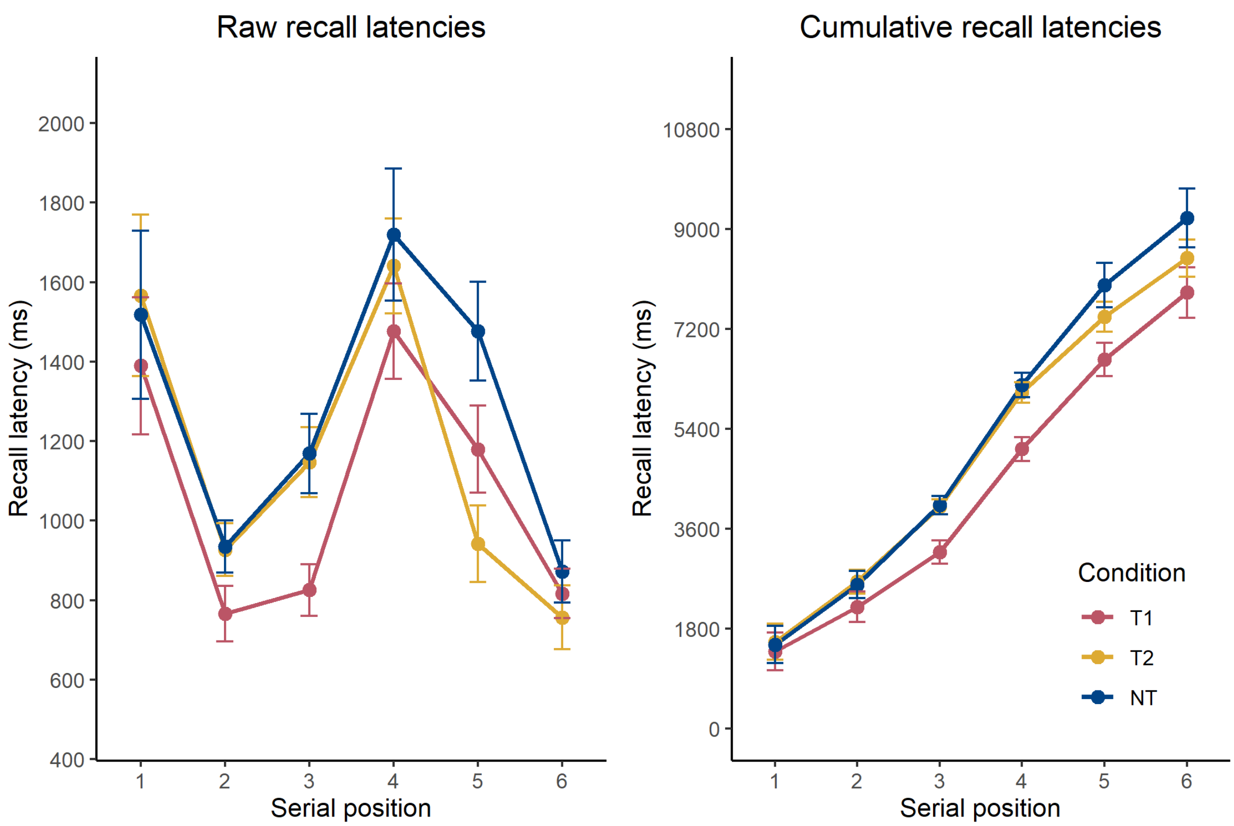

Figure 9. Median recall latencies as a function of the serial position for each semantic condition (T1, T2 and NT). Left panel: raw recall latencies. Right panel: cumulative recall latencies. It should be noted that since untransformed response latencies are used for illustrative purposes, error bars are potentially misleading.

Recall latencies. In a first analysis, we assessed recall latencies as a function of the semantic condition (T1, T2, NT) and serial position (1 through 6) using a Bayesian Repeated Measures ANOVA. We found decisive evidence supporting both main effects of semantic condition (BF_{10} = 19224.64) and serial position (BF_{10}= 1.532e+55). Likewise, the interaction term was also associated with decisive evidence (BF_{10} = 598.541). Bayesian Paired-Samples T-Tests indicate faster recall latencies over the positions that correspond to semantically related items (T1: $BF_{10} > 100$ over positions 2 and 3; T2: $BF_{10} > 100$ and BF_{10} = 68.944 over positions 5 and 6, respectively), as is also evidenced by **Figure 9** (left panel).

Cumulative recall latencies. Positions in which semantically related items were presented were associated with faster recall than positions of unrelated item. The critical question we ask here is whether the proactive benefit of the semantic triplets on recall performance could be explained by the fact that when recalling items 4, 5 and 6, less time elapsed *before* recalling these items in T1 compared to NT. To answer this question, we performed Bayesian Paired-Samples T-Tests over cumulative recall latencies throughout the different positions. Results show that the time elapsed in the T1 condition was always shorter than that in NT, and this was true throughout the different serial positions ($BF_{10} > 100$ for positions 2 through 5, and BF_{10} = 79.995 for position 6), as shown in **Figure 9** (right panel). This analysis confirms that less time elapsed before recalling items 4, 5 and 6 in the T1 condition compared to the NT condition.

Influence of recall latencies on recall performance. The analyses detailed above show that positions 1, 2 and 3 in the T1 condition were associated overall with shorter recall latencies, which also shorten the time elapsed when recalling the items that directly followed the semantic triplets. However, these analyses do not inform us whether faster recall predicted recall performance over the items that followed the semantic triplets. If so, are recall latencies *alone* responsible for the proactive benefit of the semantic triplet, or rather do they have an additive effect with the semantic influence during maintenance?

To explore these questions, we performed a Mixed Effect Bayesian ANOVA using the *brms* R package (Bürkner, 2017). The use of a mixed model approach was motivated by the possibility of performing the analysis on individual trials instead of data averaged across participants, while also including participants as a random factor. We reasoned that the time elapsed on items 1, 2 and 3 might be informative about recall performance over the second half of the list, and this on an individual trial basis, which is a critical variance that aggregate data cannot take into account. Parameters of the models were estimated using 4 independent Markov Chains, each with 50,000 samples, including 5,000 warmup samples. In the analyses reported above, the Markov Chains always converged, as indicated by R-hat < 1.05. Bayes Factors for the effects of interest were obtained using the *bayes_factor()* function implemented in the *brms* package, by directly contrasting the full model against the same model without the effect of interest.

We used mean recall performance (collapsed across positions 4, 5 and 6) as a dependent variable. Fixed-effects included the semantic condition (T1, NT), and recall latencies cumulated across positions 1, 2 and 3 (then log-transformed) as a predictor. The random effect included the by-participant random intercept. We were not able to include the by-participant random slope for the effect of cumulative recall latencies and semantic conditions due to convergence problems. However, the same analysis performed under a frequentist approach revealed an identical outcome when using the maximal random-effect structure (see **Appendix E**). If the total time elapsed before recalling items 4, 5 and 6 was important, then it should be a good predictor of recall performance for these items. In addition, if the total time elapsed is responsible alone for the proactive benefit of the semantic triplet, then adding complexity in the model by considering the effect of semantic conditions should not provide any further evidence. Fixed-level effects indicate that both semantic condition (M = -0.09, SD = 0.02, $CI_{95\%}$ = [-0.12; -0.05]) and recall latencies (M = -0.2, SD = 0.02, $CI_{95\%}$ = [-0.24; -0.16]) credibly impacted recall performance over the second half of the list. In addition, using the bayes_factor() function, we found decisive evidence supporting both the impact of cumulative recall latencies (BF_{10} = 9.828e+19) and of semantic condition (BF_{10} = 1.07e+4). Therefore, the results of this analysis showed that the total time that elapsed when recalling items 1, 2 and 3 predicted recall performance averaged over positions 4, 5 and 6. In addition, the impact of the semantic condition was robust, even after considering cumulative recall latencies as a predictor, suggesting that recall latencies alone cannot fully explain the proactive benefit of the semantic triplet.

Response type. The analyses described so far show that the time spent recalling the first part of the list was a good predictor of recall performance in the second part of the list. However, this observation was based on total recall latencies, regardless of the type of response that may have caused the proactive benefit of the semantic triplet. To derive **Figure 10**, we divided recall latencies across three different response types: correct responses associated with unrelated (NT) and semantically related (T1) stimuli, and omission errors[10]. It should be noted that recall latencies were considered across position 1, 2 and 3 only, because we were interested in the time elapsed *before* recalling positions 4, 5 and 6. As predicted, correct responses for semantically related items were associated with faster recall latencies (Median = 826 ms) compared to correct responses for semantically unrelated stimuli (Median = 938 ms), and this difference was

---

[10] There were few omission errors in the T1 condition over positions 1, 2 and 3. We therefore considered omission errors regardless of the semantic condition (NT, T1). We also looked at transposition errors, but these errors differed only slightly from correct responses, and were therefore poorly informative. We decided not to include these errors for the sake of simplicity.

supported by decisive evidence, as shown by a Bayesian Paired-Samples T-Test on log-transformed recall latencies ($BF_{10} = 250$). However, the most striking difference was of omission errors, which took more than twice the time to be produced compared to correct responses for semantically unrelated stimuli (Median = 2,200 ms, $BF_{10} = 1.437e+7$).

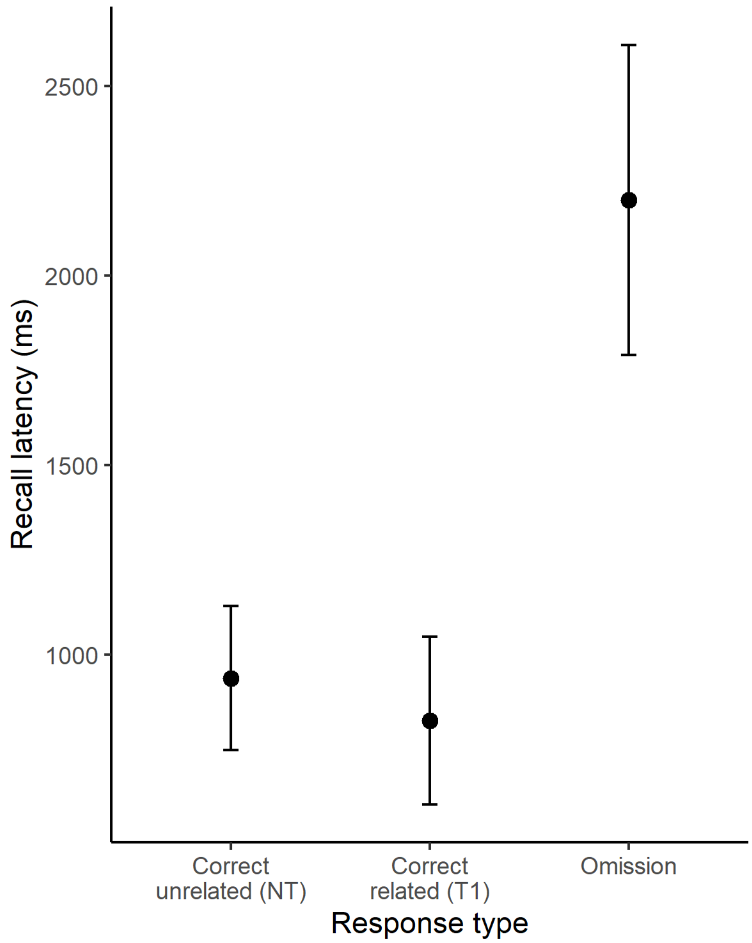

Figure 10. Median recall latencies as a function of response type: correct responses in the unrelated and related conditions, and omission errors. Only responses across the first three positions are considered. Since untransformed response latencies are used for illustrative purposes, error bars are potentially misleading.

Focus on omissions. The previous analysis suggests that if the proactive benefit of the semantic triplet is to be explained by recall latencies, then omission errors should be an important cause of this proactive benefit. A general prediction derived from this idea is that the number of omissions produced at the beginning of a list to be remembered should predict recall performance for the subsequent items within the same list[11]. We explored this possibility using a Bayesian Mixed effect regression, with recall performance (using a strict serial recall criterion) in the second part of the list as the dependent variable, and omission rate in the first part of the list as a predictor. To avoid confounding factors with the semantic triplets themselves, we excluded T1 and T2 conditions from the analysis. Random effects included the by-participant random intercept and the by-participant random slope for the omission rate in the first part of the list. Random-level effects indicate that the number of omissions in the first part of the list credibly predicted recall performance over the second half of the list ($M = 0.28$, $SD = 0.12$, $CI_{95\%} = [0.06; 0.54]$). When comparing this model against the intercept-only model, the Bayes Factor was associated with decisive evidence ($BF_{10} = 1.8e+5$). Hence, the omission rate produced at the beginning of a list to be remembered was a robust predictor of recall performance for the remaining items in the list.

It appears that the time it takes to recall the items is a good predictor of subsequent recall performance. More generally, the proactive benefit of the semantic triplet may stem from a decrease of omission errors, as was suggested from previous studies (Lovatt et al., 2002). This

[11] It is important to note at this point that the Mixed effect analysis is required, because it allows the regression analysis to be performed trial by trial, and the consideration that the slope of this influence may vary for each participant. A regression analysis on aggregated data would merely show that WM capacity is correlated among subjects, which is neither surprising nor the focus of the question under investigation.

observation is in agreement with studies showing a strong effect of output interference on WM recall performance (Cowan et al., 2002; Oberauer, 2003). In the next section, we discuss the theoretical implications of our results.

General discussion

In this study, we investigated interactions between attentional maintenance processes in WM and semantic long-term memory knowledge. We determined experimentally that the presence of a semantic triplet in a list to be remembered freed up WM resources, which in turn enhanced recall performance for the other, semantically unrelated items, and this compared to a condition without a semantic triplet. Critically, the within-list position of the semantic triplet produced strikingly different patterns of recall performance: When the semantic triplet was presented at the beginning of the list to be remembered, a proactive benefit of the semantic triplet was observed. However, when the semantic triplet was presented at the end of the list to be remembered, no retroactive impact was observed. These phenomena were successfully captured by a WM architecture integrating a separate long-term memory layer in which we assumed a direct activation of items within the long-term memory system. Further exploratory analysis of recall latencies suggested that the time spent recalling the items could also be a critical factor to proactively impact WM performance.

Proactive benefit through refreshing

Using the principles assumed from interactive activation models of language processing, and those from the TBRS* architecture, we were able to produce a novel prediction derived from the integration of both accounts. More specifically, we predicted that the presence of a semantic triplet should free up attentional WM resources that could be reallocated to prevent the other items of the list from being lost due to time-based forgetting. In this study, not only did we show that this prediction was empirically confirmed, but we also demonstrated the plausibility of those theoretical principles through their formal implementation in the TBRS*-S computational architecture. A fine-grained diagnosis of the model's behavior showed that this freeing of attentional WM resources is a direct consequence of the interaction between the high activation level associated with semantically related items, and the Least Activated First mechanism. Because semantically related items are the less likely to be forgotten due to their high activation level, they are maintained with fewer refreshing episodes. Since these refreshing episodes are not dedicated to refresh the semantic triplet, they can be reallocated to refresh other, semantically unrelated items. An important aspect of the simulations is that the model using the proactive benefit as a criterion was never fitted. Instead, the proactive benefit of the semantic triplet naturally emerged after implementing inter-item excitatory connections within the long-term memory layer.

An interesting aspect of the model's behavior is the fact that it did not produce a retroactive impact of the semantic triplet on recall performance. As mentioned by Thalmann et al. (2018), an intuitive prediction from decay and rehearsal/refreshing models is that the presence of a chunk - semantic or not - should free up attentional WM resources. This would be regardless of the chunk's position, because there would still be room for attentional reallocation, especially if the refreshing process operates in a very fast manner, as in TBRS. However, this interpretation is based merely on intuition about the theory, not on a formal implementation. The absence of a retroactive impact of the semantic triplet was due to the fact that when the semantic triplet started to impact recall performance, all the items that could have been saved through attentional reallocation were already lost during the previous between-item maintenance phases. Hence, the items that benefited from the attentional reallocation process were those that were robust enough to survive, leading to an absence of retroactive impact of the semantic triplet. This phenomenon illustrates the fact that computational models are fruitful tools to guide reasoning about a theory, which can be biased by our intuitions (Farrell & Lewandowsky, 2010).

In the model, the proactive benefit of the semantic triplet was accompanied by a deleterious impact on the ability to recall order information over the semantic triplet itself, leading to an absence of serial recall advantage over positions 1, 2 and 3 in T1 compared to NT. This issue is due to a property of our model. Semantically related items reactivate each other and need fewer refreshing attempts due to their high activation level in long-term memory, as already explained. At the same time, a side-effect of this reallocation process is that during the time spent not refreshing these semantically related items, their positional information (i.e. the item-position associations) was also lost proportionally, leading to an absence of recall advantage when a strict serial recall criterion was considered. However, this only happened in the T1 condition since in the T2 condition, the model behaved as expected. Indeed, because the semantic triplet appeared later in the list, the reallocation process also appeared much later, leaving less opportunity for the serial order information to be lost. Although this issue is symptomatic of a weakness in the model, we nevertheless see it as an opportunity for future research to further investigate and gain a better comprehension of the interactions between semantic knowledge and the maintenance of serial order information, which are still poorly understood. In fact, the question of serial order information is critical for WM models, and there is no clear consensus regarding the nature of the codes that are actually used to maintain serial order information (Majerus, 2019), proposals ranging from positional/contextual (Burgess & Hitch, 2006; Henson, 1998) to spatial (Abrahamse et al., 2014), temporal (Hartley et al., 2016) or even associative (Lewandowsky & Murdock, 1989) ones.

Proactive benefit through recall latencies

We also explored the possibility that the locus of the proactive benefit of the semantic triplet could stem from recall rather than maintenance. A first overall assessment of recall latencies demonstrated that the recall latencies were indeed shorter over the positions where semantically related items were presented, and this time spent recalling the items was a good predictor of recall performance over the remaining items of the list. It was demonstrated in a previous systematic investigation of recall latencies that omission errors take twice the time to be recalled than correct responses (Haberlandt et al., 2005). Thus, the number of omission errors produced should be a critical factor for subsequent recall performance. This latter prediction was met when conducting a mixed-effect regression analysis.

The idea that the time taken to output items should affect recall performance is not new. One of the first investigations of this idea is similar to our experimental manipulation, and dates back to the study conducted by Cowan et al. (1992). They observed that long words, when presented in the first half of a to-be-remembered list, proactively (but not retroactively) impeded WM performance as compared to short words, the latter being faster to recall. A further investigation of this effect (Lovatt et al., 2002) suggested that this proactive interference effect provoked by the word length effect could have been explained by the number of errors produced at the beginning of the list to be remembered, results which converge with the observations made in the present study.

It is important to emphasize that these analyses on recall latencies are merely correlational. There are many ways through which recall latencies could have predicted recall performance (see Lewandowsky & Oberauer, 2008 for a discussion). More generally, recall performance and latencies are two sides of the same coin, and hence they measure the same construct (Vandierendonck, 2017). Critically, these analyses on recall latencies should not be taken as evidence to support decay-based models. In fact, direct manipulations of response speed have sometimes yielded to a complete absence of impact on recall performance (Cowan et al., 2006). Our results suggest that if time-based forgetting exists, then recall latency is a critical factor that proponents of the decay theory should be particularly aware of when conducting experiments, taking a special care concerning the production of errors.

Relationship with other phenomena

Similar proactive benefits have been observed through the manipulation of word frequency. More specifically, Miller and Roodenrys (2012) embedded high frequency words in lists composed of low frequency ones (i.e. HHH-LLL or LLL-HHH). Compared to lists composed only of low frequency words (i.e. LLL-LLL), these authors observed not only that the high frequency words were better recalled, but also that they had a proactive benefit on the subsequent low frequency words of the list. In contrast, no retroactive benefit was observed. More generally, it appears that proactive effects in WM follow a general rule, as depicted in **Figure 11**.

To account for the proactive benefit of psycholinguistic variables (i.e. semantic relatedness, word length and word frequency), and their lack of retroactive impact, we suggest that they may follow this general principle: Highly activated items are less likely to be refreshed, because the focus of attention refreshes the least activated items in priority. This automatically leads to a reallocation of attentional resources over the items associated with the lowest activation values. Although we already demonstrated the plausibility of this mechanism as regards the semantic relatedness in our first simulation, one may wonder how this principle would apply to the word frequency and word length effects.

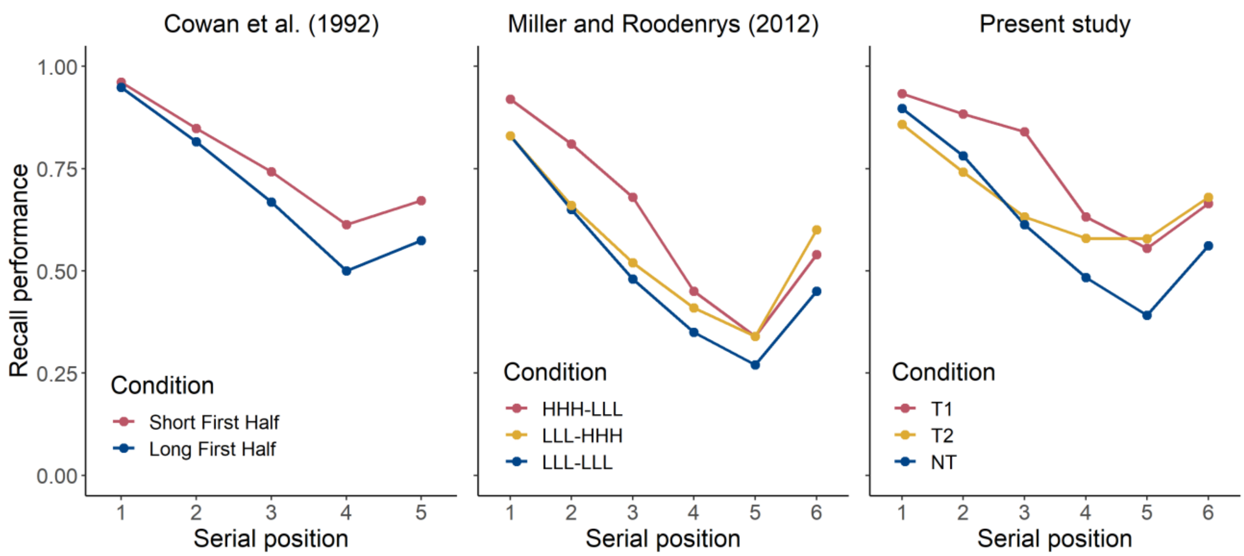

Figure 11. Illustration of proactive effects observed in previous studies (left an central panels, adapted from Cowan et al., 1992 and Miller & Roodenrys, 2012, respectively) and in our own study (right panel).

From the perspective of interactive activation models of language processing, the word frequency effect can be explained by presuming that high frequency words have stronger connection weights than low frequency words. This would be between their lexical and phonological representations (Besner & Risko, 2016), for instance due to a mere exposure effect (Zevin & Seidenberg, 2002). Thanks to these stronger connection weights, high frequency words receive stronger redundant feedback activation from their phonological representation compared to low frequency words, leading to overall stronger activation values. When embedded in a list of low frequency words, high frequency words naturally have a stronger activation level than the low frequency words. This results in a reallocation of attentional resources toward low frequency words and hence a proactive benefit.

The proactive effect caused by the word length effect can be explained in the same way. There has been much debate surrounding the origin of the word length effect, and this is partly due to its hypothetical role in distinguishing decay-based forgetting from interference-based forgetting (Lewandowsky & Oberauer, 2008), a critical theoretical question in the WM domain. One supposed origin of the word length effect is that it arises from a confounding factor: the neighborhood density effect. This effect is characterized by better recall performance for words that share many phonological neighbors compared to words sharing fewer phonological neighbors. For example, the word “cat” which has “fat”, “bat”, “mat”, “rat”, etc. as neighbors has better recall than words with fewer phonological neighbors (Roodenrys et al., 2002). It appears that long words are characterized by fewer neighbors than short words (Jalbert et al., 2011). Recent evidence suggests that the word length effect might be mostly attributed to this confounding factor (Guitard et al., 2018). The neighborhood density effect is explained by interactive activation models, by assuming that the phonological neighbors of a target word are co-activated. Thus, the activation of “cat” results in the obligatory activation of “fat”, “bat”, “mat”, etc. Then, more activation is sent back to the original target via redundant feedbacks between the lexical and phonological levels of language processing (Chen & Mirman, 2012; Dell et al., 1997; Vitevitch & Luce, 2016). Logically, this results in higher activation levels for short words drawn from dense neighborhoods compared to long words found in sparse neighborhoods. As for the semantic similarity and lexical frequency effects, this should lead to a reallocation of attentional resources toward sparse neighborhood words when dense neighborhood words are embedded in the same list.

Alternative accounts

This study is framed by decay-based models of WM. As such, we adapted the original TBRS* architecture and supplemented it with a linguistic system according to decay-based principles. However, other phenomena have been proposed to account for WM capacity limitations, such as interference. For instance, this is hypothesized by the SOB-CS model (Oberauer, Lewandowsky, et al., 2012), a direct competing architecture to TBRS*. The SOB-CS model already demonstrated an excellent explanatory power to simulate many benchmark phenomena observed in WM tasks. Technically, it should be possible to simulate semantic effects in SOB-CS by coding the items along a semantic dimension, and by considering that semantically related items are represented across a similar set of overlapping semantic features. However, this implementation is likely to lead to a reversed semantic relatedness effect, because between-item similarities should always result in poorer recall performance, due to novelty-gated encoding (see Chekaf et al., 2016 for a related interpretation), at least in the actual version of SOB-CS. It should be noted that this prediction still needs to be formally assessed through a computational model, which is beyond the scope of the present study. More generally, over sixty years of debate precedes the question of decay versus interference-based forgetting in WM (Ricker et al., 2016), and this question remains heavily debated (Dagry & Barrouillet, 2017; Farrell et al., 2016; Lemaire & Portrat, 2018; Oberauer, 2019; Oberauer et al., 2016; Ricker et al., 2020). Given the strong and robust impact that psycholinguistic variables have on WM performance (see Kowialiewski & Majerus, 2018, 2020 for meta-analyses), psycholinguistic effects can be systematically modeled in different computational architectures to assess their plausibility. This in turn may inform us about the cause of forgetting in WM in a novel, refreshing manner.

The theoretical account developed by Popov and Reder (2020) could, at least theoretically, also explain the results we observed in the present study. This account suggests that encoding in WM depletes a limited pool of resources. This depletion is furthermore inversely related to item strength: items associated with stronger representations in long-term memory are assumed to deplete fewer resources. The amount of available resources in the pool also constrains how strongly an item is encoded, with many resources allowing stronger encoding. If interactive activation principles are represented in this model, it could be assumed that semantically related items would be more easy to encode and would deplete fewer resources. In the case where the semantic triplet is presented at the beginning of the to-be-remembered list, these resources could be reallocated to encode more strongly the subsequent to-be-remembered items of the list. This should lead to a proactive benefit as compared to a condition in which all the items are semantically unrelated. Since encoding strength depends on the amount of resources depleted from previous items, it naturally predicts a proactive benefit, but no retroactive benefit. The plausibility of this account has been demonstrated through computational simulations, and successfully reproduced the word frequency manipulations made by Miller and Roodenrys (2012) that we presented above.

Finally, an obvious alternative mechanism to explain the proactive benefit observed in our experiment boils down to a chunking account, as previously observed in studies using chunks composed of letters (Norris et al., 2020; Portrat et al., 2016; Thalmann et al., 2018). After encoding the words “leaf - tree - branch”, the participants might rapidly become aware of the super-ordinate semantic category that characterizes the items. Only “nature” or “forest” may be maintained. This semantic category could then be kept on as a single item, and used as retrieval cue at the moment of recall (Kowialiewski & Majerus, 2020; Saint-Aubin & Ouellette, 2005).

This is a highly intuitive explanation. An informal assessment of participants' strategies at the end of the experiment confirms that the use of this strategy was indeed common. Participants reported "not carrying" on the semantically related words, because they "knew" that these words were - for instance - "about the nature theme". This chunking account does not change the overall conclusions of this study: the maintenance of this semantic triplet should free up attentional WM resources, and these resources could then be reallocated to maintain more information, as already demonstrated through the use of computational simulations (Portrat et al., 2016).

Conclusion

In this study, we used a convergent approach involving behavioral experiments and computational modeling to get a better understanding of the mechanisms underlying the influence of semantic knowledge on WM performance. We showed that semantic knowledge frees up attentional WM resources that can be reallocated for maintenance purpose. This suggests that semantic knowledge interacts in a complex manner with WM maintenance processes. Furthermore, recall latencies appear to be a potentially critical factor, but it is frequently neglected in WM paradigms. This study brings novel evidence supporting strong interactions between WM and the long-term memory system.

Acknowledgements

This study was funded by the French National Research Agency (ANR) under the CHUNKED project #ANR-17-CE28-0013-03. We thank all the participants for their time devoted to this study.

Appendix

Appendix A

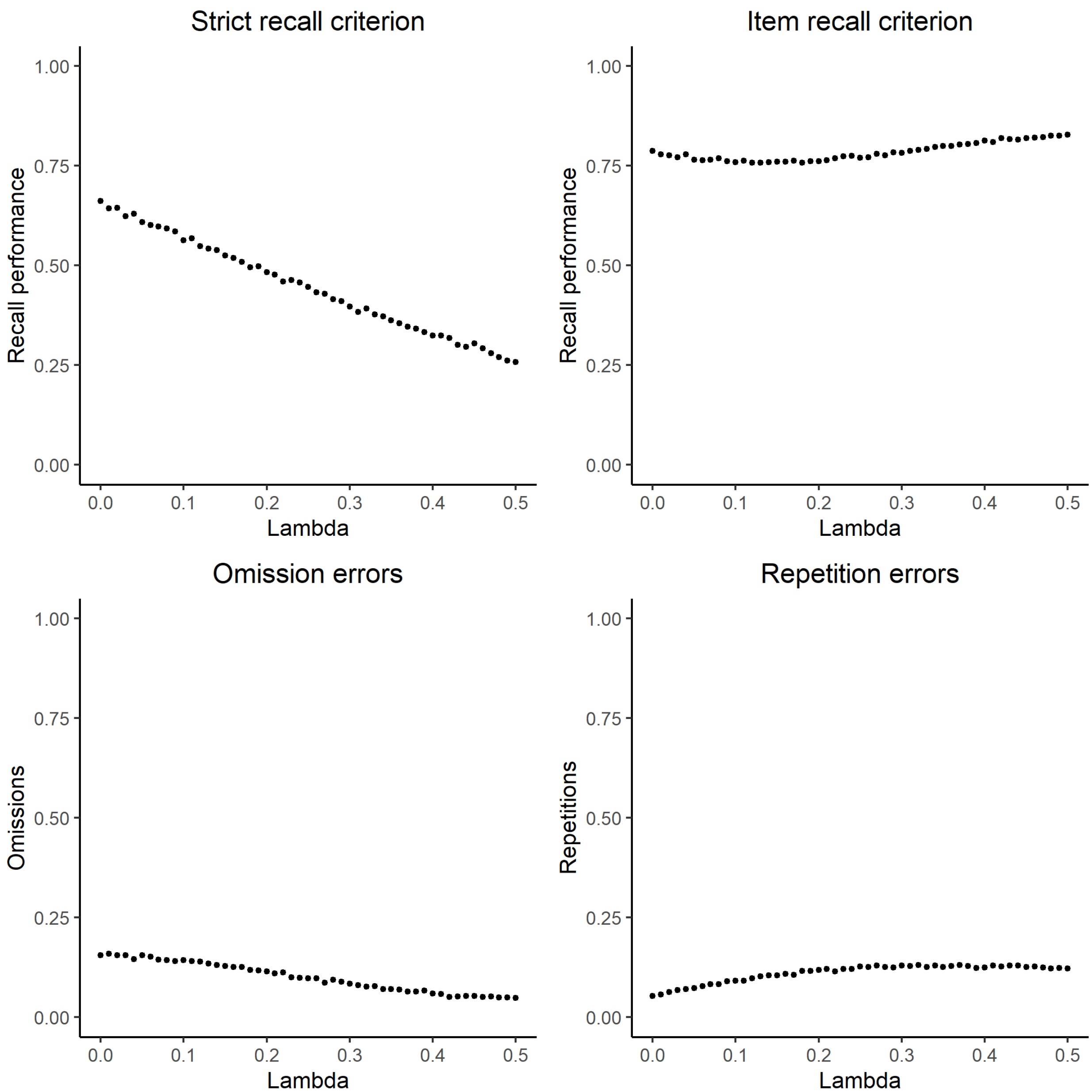

Figure A1. Recall performance (top panels), omission errors (bottom left panel) and repetition errors (bottom right panel) as a function of the value of the Lambda parameter.

Appendix B

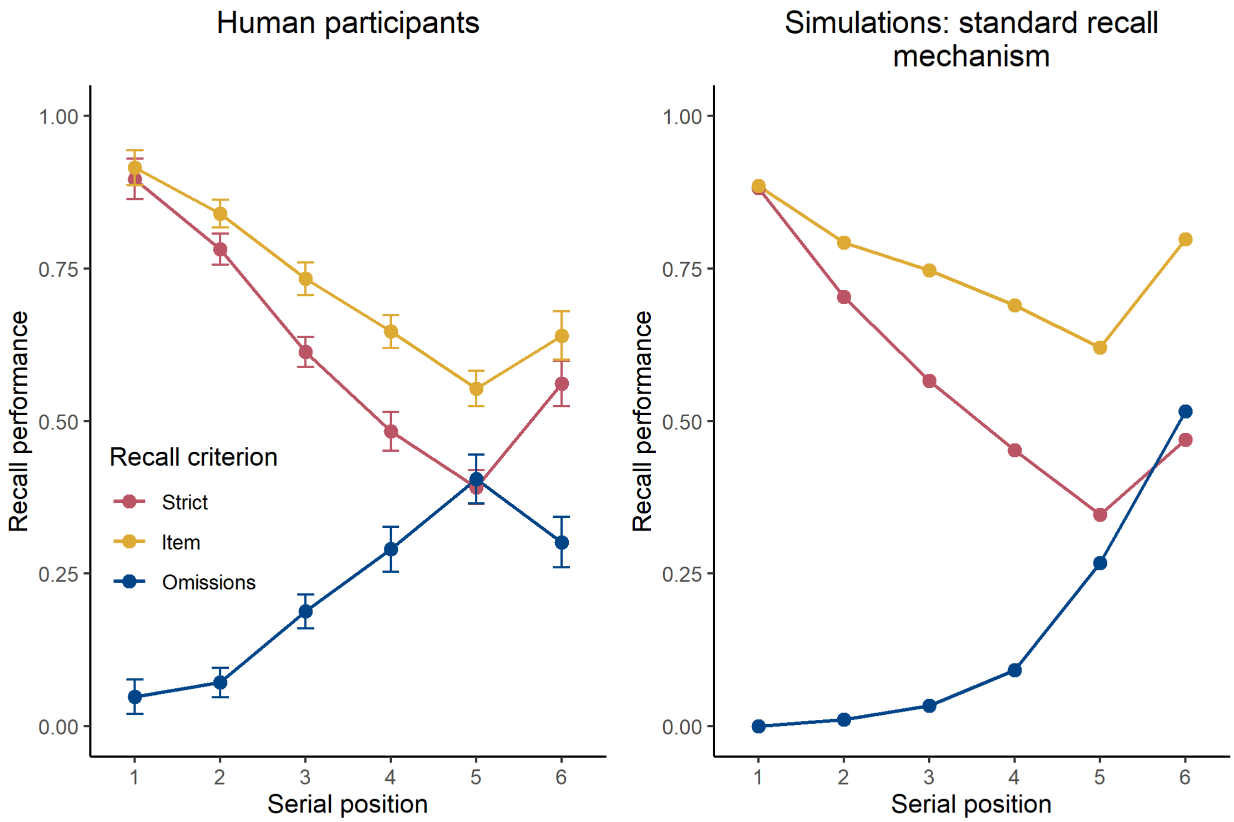

Figure B1. Recall performance for the three recall criteria (strict, item and omission errors) as a function of the serial position observed in humans (left panel) and produced by the model with a standard recall mechanism (right panel).

Appendix C

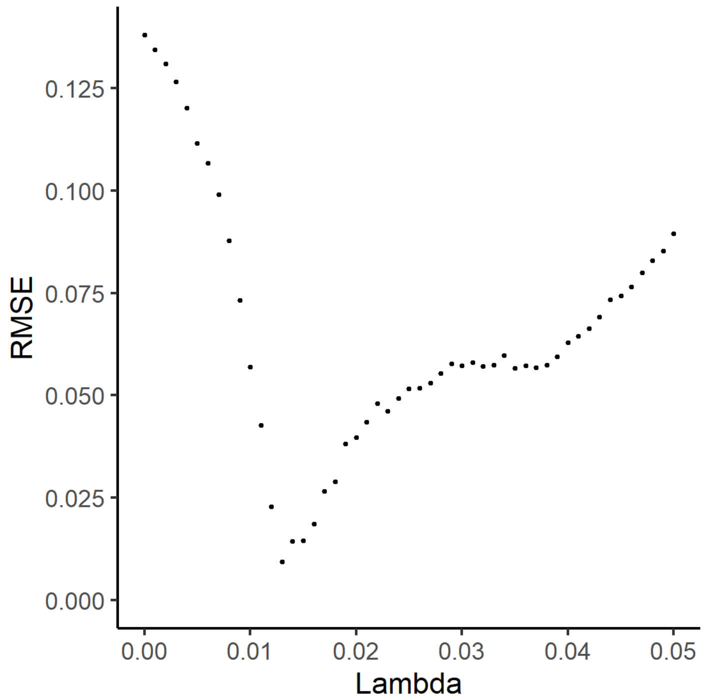

Figure C1. Root mean square errors (RMSE) as a function of the value of the lambda parameter.

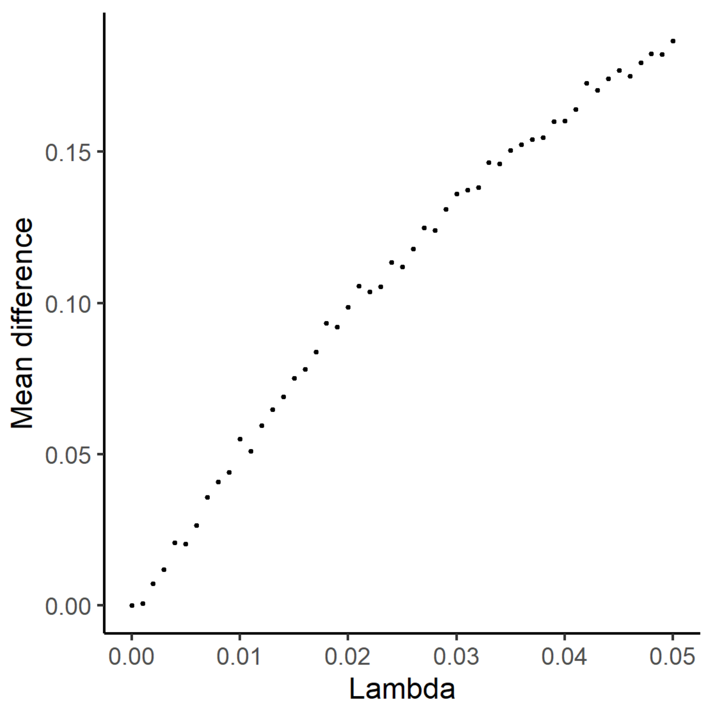

Figure C2. Mean recall performance differences between T1 and NT in positions 4-5-6 as a function of the values of the lambda parameter.

Appendix D

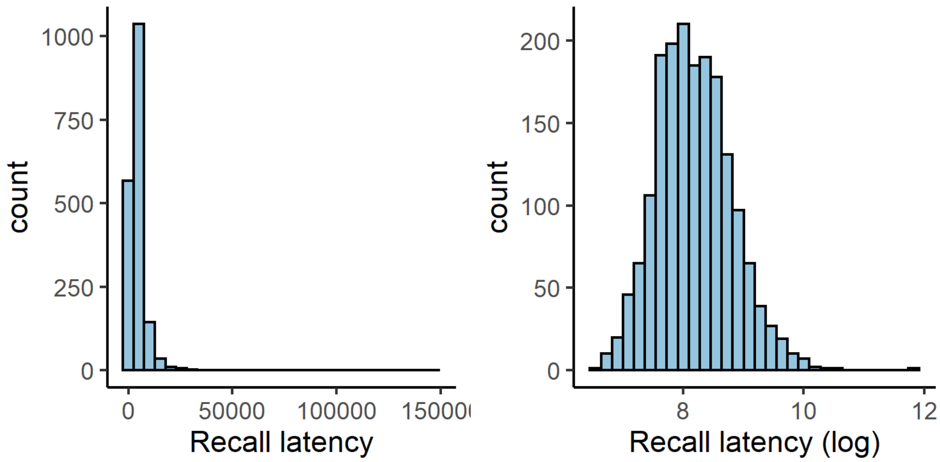

Figure D1. Shift of the response latencies distribution before (left panel) and after (right panel) log-transformation.

Appendix E: Linear mixed-effect model using a frequentist approach

Mean recall performance (collapsed across positions 4, 5 and 6) as a function of recall latencies (cumulated across positions 1, 2 and 3) and semantic condition (T1, NT) were assessed using the *lmer()* function of the *lme4* package (Bates et al., 2015). The statistical significance of each fixed effect of interest was obtained via Satterthwaite's degrees of freedom method using the *lmerTest* package (Kuznetsova et al., 2017). The raw cumulative recall latencies were log-transformed twice to solve convergence problems. This double log-transformation did not dramatically impact the shape of data distribution, which followed the classical Gaussian distribution, as can be seen in **Figure E1**. The maximal random-effect model including the by-participant random slope for the effect of recall latencies and semantic condition indicated a significant effect of recall latencies (t = -9.228; df = 37.319; p = 3.6e-11) and semantic condition (t = -3.986; df = 28.514; p = 4.25e-4).

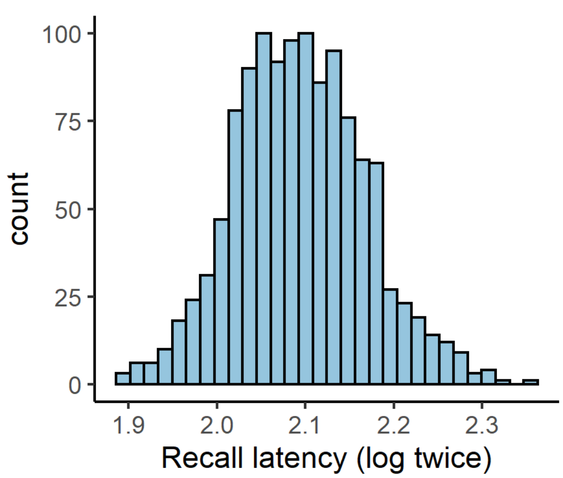

Figure E1. Distribution of recall latencies after the log-transformation has been applied twice.

References

Abrahamse, E., van Dijck, J.-P., Majerus, S., & Fias, W. (2014). Finding the answer in space : The mental whiteboard hypothesis on serial order in working memory. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00932

Acheson, D. J., & MacDonald, M. C. (2009). Verbal Working Memory and Language Production : Common Approaches to the Serial Ordering of Verbal Information. *Psychological Bulletin*, *135*(1), 50-68. https://doi.org/10.1037/a0014411

Acheson, D. J., Postle, B. R., & MacDonald, M. C. (2010). The Interaction of Concreteness and Phonological Similarity in Verbal Working Memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *36*(1), 17-36. https://doi.org/10.1037/a0017679

Baddeley, A. D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *The Quarterly Journal of Experimental Psychology*, *18*(4), 362-365. https://doi.org/10.1080/14640746608400055

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*(6), 575-589.

Baguley, T. (2012). Calculating and graphing within-subject confidence intervals for ANOVA. *Behavior Research Methods*, *44*(1), 158-175. https://doi.org/10.3758/s13428-011-0123-7

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., Portrat, S., & Camos, V. (2011). On the Law Relating Processing to Storage in Working Memory. *Psychological Review*, *118*(2), 175-192. https://doi.org/10.1037/a0022324

Besner, D., & Risko, E. F. (2016). Thinking outside the box when reading aloud : Between (localist) module connection strength as a source of word frequency effects. *Psychological Review*, *123*(5), 592-599. https://doi.org/10.1037/rev0000041

Brener, R. (1940). An experimental investigation of memory span. *Journal of Experimental Psychology*, 467-482.

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, *55*(4), 627-652. https://doi.org/10.1016/j.jml.2006.08.005

Bürkner, P.-C. (2017). brms : An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Camos, V., Johnson, M., Loaiza, V., Portrat, S., Souza, A., & Vergauwe, E. (2018). What is attentional refreshing in working memory? *Annals of the New York Academy of Sciences*, *1424*(1), 19-32. https://doi.org/10.1111/nyas.13616

Campoy, G., & Baddeley, A. D. (2008). Phonological and semantic strategies in immediate serial recall. *Memory*, *16*(4), 329-340. https://doi.org/10.1080/09658210701867302

Campoy, G., Castellà, J., Provencio, V., Hitch, G. J., & Baddeley, A. D. (2015). Automatic semantic encoding in verbal short-term memory : Evidence from the concreteness effect. *The Quarterly Journal of Experimental Psychology*, *68*(4), 759-778. https://doi.org/10.1080/17470218.2014.966248

Castellà, J., & Campoy, G. (2018). The (lack of) effect of dynamic visual noise on the concreteness effect in short-term memory. *Memory*, *0*(0), 1-9. https://doi.org/10.1080/09658211.2018.1476550

Chekaf, M., Cowan, N., & Mathy, F. (2016). Chunk formation in immediate memory and how it relates to data compression. *Cognition*, *155*, 96-107. https://doi.org/10.1016/j.cognition.2016.05.024

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations : Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*, *119*(2), 417-430. https://doi.org/10.1037/a0027175

Chubala, C., Surprenant, A. M., Neath, I., & Quinlan, P. T. (2018). Does dynamic visual noise eliminate the concreteness effect in working memory ? *Journal of Memory and Language*, *102*(May), 97-114. https://doi.org/10.1016/j.jml.2018.05.009

Cousineau, D. (2012). Erratum to Cousineau (2005) : Confidence intervals in within-subject designs : A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *8*(3), 182-182. https://doi.org/10.20982/tqmp.08.3.p182

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

Cowan, N. (2001). The magical number 4 in short-term memory : A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87-185. https://doi.org/10.1017/S0140525X01003922

Cowan, N., Day, L., Saults, J. S., Keller, J., Johnson, T., & Flores, L. (1992). The role of verbal output time in the effects of word length on immediate memory. *Journal of Memory and Language*, *31*(1), 1-17. https://doi.org/10.1016/0749-596X(92)90002-F

Cowan, N., Elliott, E. M., Saults, J. S., Nugent, L. D., Bomb, P., & Hismjatullina, A. (2006). Rethinking Speed Theories of Cognitive Development. Increasing the Rate of Recall Without Affecting Accuracy. *Psychological Science*, *17*(1), 67-73. https://doi.org/10.1111/j.1467-9280.2005.01666.x

Cowan, N., Saults, J. S., Elliott, E. M., & Moreno, M. V. (2002). Deconfounding Serial Recall. *Journal of Memory and Language*, *46*(1), 153-177. https://doi.org/10.1006/jmla.2001.2805

Dagry, I., & Barrouillet, P. (2017). The fate of distractors in working memory : No evidence for their active removal. *Cognition*, *169*, 129-138. https://doi.org/10.1016/j.cognition.2017.08.011

Dell, G. S., Schwartz, M. F., Martin, N., Saffran, E. M., & Gagnon, D. A. (1997). Lexical Access in Aphasic and Nonaphasic Speakers. *Psychological Review*, *104*(4), 801-838.

Farrell, S., & Lewandowsky, S. (2004). Modelling transposition latencies : Constraints for theories of serial order memory. *Journal of Memory and Language*, *51*(1), 115-135. https://doi.org/10.1016/j.jml.2004.03.007

Farrell, S., & Lewandowsky, S. (2010). Computational Models as Aids to Better Reasoning in Psychology. *Current Directions in Psychological Science*, *19*(5), 329-335. https://doi.org/10.1177/0963721410386677

Farrell, S., Oberauer, K., Greaves, M., Pasiecznik, K., Lewandowsky, S., & Jarrold, C. (2016). A test of interference versus decay in working memory : Varying distraction within lists in a complex span task. *Journal of Memory and Language*, *90*, 66-87. https://doi.org/10.1016/j.jml.2016.03.010

Gorin, S., Kowialiewski, B., & Majerus, S. (2016). Domain-generality of timing-based serial order processes in short-term memory : New insights from musical and verbal domains. *PLoS ONE*, *11*(12), 1-25. https://doi.org/10.1371/journal.pone.0168699

Guitard, D., Gabel, A. J., Saint-Aubin, J., Surprenant, A. M., & Neath, I. (2018). Word length, set size, and lexical factors : Re-examining what causes the word length effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(11), 1824-1844. https://doi.org/10.1037/xlm0000551

Haarmann, H., & Usher, M. (2001). Maintenance of semantic information in capacity-limited item short-term memory. *Psychonomic Bulletin & Review*, *8*(3), 568-578.

Haberlandt, K., Thomas, J. G., Lawrence, H., & Krohn, T. (2005). Transposition asymmetry in immediate serial recall. *Memory*, *13*(3-4), 274-282. https://doi.org/10.1080/09658210344000297

Hanley, R. J., & Bakopoulou, E. (2003). Irrelevant speech, articulatory suppression, and phonological similarity : A test of the phonological loop model and the feature model. *Psychonomic Bulletin & Review*, *10*(2), 435-444.

Hartley, T., Hurlstone, M. J., & Hitch, G. J. (2016). Effects of rhythm on memory for spoken sequences : A model and tests of its stimulus-driven mechanism. *Cognitive Psychology*, *87*, 135-178. https://doi.org/10.1016/j.cogpsych.2016.05.001

Henson, R. N. A. (1998). Short-term memory for serial order : The start-end model. *Cognitive Psychology*, *36*(2), 73-137.

Henson, R. N. A., Hartley, T., Burgess, N., Hitch, G., & Flude, B. (2003). Selective interference with verbal short-term memory for serial order information : A new paradigm and tests of a timing-signal hypothesis. *The Quarterly Journal of Experimental Psychology*, *56A*(8), 1307-1334. https://doi.org/10.1080/02724980244000747

Hofmann, M. J., & Jacobs, A. M. (2014). Interactive activation and competition models and semantic context : From behavioral to brain data. *Neuroscience and Biobehavioral Reviews*, *46*(P1), 85-104. https://doi.org/10.1016/j.neubiorev.2014.06.011

Hurlstone, M. J., & Hitch, G. J. (2015). How is the serial order of a spatial sequence represented? Insights from transposition latencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 295-324. https://doi.org/10.1037/a0038223

Jalbert, A., Neath, I., Bireta, T. J., & Surprenant, A. M. (2011). When does length cause the word length effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 338-353. https://doi.org/10.1037/a0021804

Jefferies, E., Frankish, C. R., & Lambon Ralph, M. A. (2006). Lexical and semantic binding in verbal short-term memory. *Journal of Memory and Language*, *54*(1), 81-98. https://doi.org/10.1016/j.jml.2005.08.001

Jeffreys, H. (1998). The theory of probability. *Oxford, UK: Oxford University Press*.

Jolicœur, P., & Dell'Acqua, R. (1998). The Demonstration of Short-Term Consolidation. *Cognitive Psychology*, *36*(2), 138-202. https://doi.org/10.1006/cogp.1998.0684

Kowialiewski, B., & Majerus, S. (2018). The non-strategic nature of linguistic long-term memory effects in verbal short-term memory. *Journal of Memory and Language*, *101*, 64-83. https://doi.org/10.1016/j.jml.2018.03.005

Kowialiewski, B., & Majerus, S. (2020). The varying nature of semantic effects in working memory. *Cognition*.

Lemaire, B., Pageot, A., Plancher, G., & Portrat, S. (2018). *What is the time course of working memory attentional refreshing ?* 370-385. https://doi.org/10.3758/s13423-017-1282-z

Lemaire, B., & Portrat, S. (2018). A Computational Model of Working Memory Integrating Time-Based Decay and Interference. *Frontiers in Cognitive Psychology*, *9*, 1-16. https://doi.org/10.3389/fpsyg.2018.00416

Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review*, *96*(1), 25-57.

Lewandowsky, S., & Oberauer, K. (2008). The word-length effect provides no evidence for decay in short-term memory. *Psychonomic Bulletin & Review*, *15*(5), 875-888. https://doi.org/10.3758/PBR.15.5.875

Logie, R. H., Della Sala, S., Laiacona, M., Chalmers, P., & Wyn, V. (1996). Group aggregates and individual reliability : The case of verbal short-term memory. *Memory & Cognition*, *24*(3), 305-321. https://doi.org/10.3758/BF03213295

Lovatt, P., Avons, S. E., & Masterson, J. (2002). Output Decay in Immediate Serial Recall : Speech Time Revisited. *Journal of Memory and Language*, *46*(1), 227-243. https://doi.org/10.1006/jmla.2001.2806

Majerus, S. (2009). Verbal short-term memory and temporary activation of language representations : The importance of distinguishing item and order information. In *Interactions between short-term and long-term memory in the verbal domain* (p. 244).

Majerus, S. (2013). Language repetition and short-term memory : An integrative framework. *Frontiers in Human Neuroscience*, *7*, 357. https://doi.org/10.3389/fnhum.2013.00357

Majerus, S. (2019). Verbal working memory and the phonological buffer : The question of serial order. *Cortex*, 1-12. https://doi.org/10.1016/j.cortex.2018.04.016

Martin, N., Saffran, E. M., & Dell, G. S. (1996). Recovery in deep dysphasia : Evidence for a relation between auditory–verbal STM capacity and lexical errors in repetition. *Brain and Language*, *52*(1), 83-113.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame : An open-source , graphical experiment builder for the social sciences. *Behavior research methods*, *44*(2), 314-324. https://doi.org/10.3758/s13428-011-0168-7

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception : Part 1. An account of basic findings. *Psychological Review*, *88*(5), 375.

Miller, L. M., & Roodenrys, S. (2009). The interaction of word frequency and concreteness in immediate serial recall. *Memory & cognition*, *37*(6), 850-865. https://doi.org/10.3758/MC.37.6.850

Miller, L. M., & Roodenrys, S. (2012). Serial recall, word frequency, and mixed lists : The influence of item arrangement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1731-1740. https://doi.org/10.1037/a0028470

Monnier, C., & Bonthoux, F. (2011). The semantic-similarity effect in children : Influence of long-term knowledge on verbal short-term memory. *British Journal of Developmental Psychology*, *29*(4), 929-941. https://doi.org/10.1111/j.2044-835X.2010.02024.x

Morey, R. D. (2008). Confidence Intervals from Normalized Data : A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*(2), 61-64.

Morey, R. D., & Rouder, J. N. (2014). *BayesFactor : Computation of Bayes factors for common designs (Version 0.9.2+)* (p. https://richarddmorey.github.io/BayesFactor/).

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory : Roles of inhibitionless spreading activation and limited-capacity attention. *Journal of Experimental Psychology: General*, *106*(3), 226-254. https://doi.org/10.1037/0096-3445.106.3.226

Norris, D. (2017). Short-Term Memory and Long-Term Memory are Still Different. *Psychological Bulletin*, *143*(9), 992.

Norris, D. (2019). Even an activated long-term memory system still needs a separate short-term store : A reply to cowan (2019). *Psychological Bulletin*, *145*(8), 848-853. https://doi.org/10.1037/bul0000204

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

Oberauer, K. (2002). Access to information in working memory : Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 411-421. https://doi.org/10.1037//0278-7393.28.3.411

Oberauer, K. (2003). Understanding serial position curves in short-term recognition and recall. *Journal of Memory and Language*, *49*(4), 469-483. https://doi.org/10.1016/S0749-596X(03)00080-9

Oberauer, K. (2009). Design for a Working Memory. In *The Psychology of Learning and Motivation* (1re éd., Vol. 51, Numéro 09). Elsevier Inc. https://doi.org/10.1016/S0079-7421(09)51002-X

Oberauer, K. (2019). Is Rehearsal an Effective Maintenance Strategy for Working Memory? *Trends in Cognitive Sciences*, *23*(9), 798-809. https://doi.org/10.1016/j.tics.2019.06.002

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

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*(7), 758-799. https://doi.org/10.1037/bul0000046

Oberauer, K., Farrell, S., Jarrold, C., Pasiecznik, K., & Greaves, M. (2012). Interference between maintenance and processing in working memory : The effect of item-distractor similarity in complex span. *Journal of Experimental Psychology: Learning Memory and Cognition*, *38*(3), 665-685. https://doi.org/10.1037/a0026337

Oberauer, K., & Lewandowsky, S. (2011). Modeling working memory : A computational implementation of the Time-Based Resource-Sharing theory. *Psychonomic Bulletin & Review*, *18*(1), 10-45. https://doi.org/10.3758/s13423-010-0020-6

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A. R. A., Cowan, N., Donkin, C., Farrell, S., Hitch, G. J., Hurlstone, M. J., Ma, W. J., Morey, C. C., Nee, D., Schweppe, J., Vergauwe, E., & Ward, G. (2018). Benchmarks for models of short term and working memory. *Psychological Bulletin*, 1-154. https://doi.org/10.1037/h0021468

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory : An interference model of complex span. *Psychonomic Bulletin & Review*, *19*(5), 779-819. https://doi.org/10.3758/s13423-012-0272-4

Patterson, K., Graham, N., & Hodges, J. R. (1994). The impact of semantic memory loss on phonological representations. *Journal of cognitive neuroscience*, *6*(1), 57-69. https://doi.org/10.1162/jocn.1994.6.1.57

Poirier, M., & Saint-Aubin, J. (1995). Memory for related and unrelated words : Further evidence on the influence of semantic factors in immediate serial recall. *The Quarterly Journal of Experimental Psychology*, *48*(2), 384-404. https://doi.org/10.1080/14640749508401396

Poirier, M., & Saint-Aubin, J. (1996). Immediate Serial Recall, Word Frequency, Item Identity and Item Position. *Canadian Journal of Experimental Psychology*, *50*(4), 408-412.

Popov, V., & Reder, L. M. (2020). Frequency effects on memory : A resource-limited theory. *Psychological Review*, *127*(1), 1-46. https://doi.org/10.1037/rev0000161

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

Python, G., Fargier, R., & Laganaro, M. (2018). ERP evidence of distinct processes underlying semantic facilitation and interference in word production. *Cortex*, *99*, 1-12. https://doi.org/10.1016/j.cortex.2017.09.008

R Development Core Team. (2008). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Ricker, T. J., Sandry, J., Vergauwe, E., & Cowan, N. (2020). Do familiar memory items decay? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(1), 60-76. https://doi.org/10.1037/xlm0000719

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*(10), 1969-1995. https://doi.org/10.1080/17470218.2014.914546

Romani, C., McAlpine, S., & Martin, R. C. (2008). Concreteness effects in different tasks : Implications for models of short-term memory. *The Quarterly Journal of Experimental Psychology*, *61*(2), 292-323. https://doi.org/10.1080/17470210601147747

Roodenrys, S., Hulme, C., Lethbridge, A., Hinton, M., & Nimmo, L. M. (2002). Word-frequency and phonological-neighborhood effects on verbal short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1019-1034. https://doi.org/10.1037/0278-7393.28.6.1019

Saint-Aubin, J., & Ouellette, D. (2005). Semantic similarity and immediate serial recall : Is there an effect on all trials. *Psychonomic Bulletin & Review*, *12*, 171-177.

Savill, N., Metcalfe, T., Ellis, A. W., & Jefferies, E. (2015). Semantic categorisation of a word supports its phonological integrity in verbal short-term memory. *Journal of Memory and Language*, *84*, 128-138. https://doi.org/10.1016/j.jml.2015.06.003

Schönbrodt, F. D., Wagenmakers, E.-J., Zehetleitner, M., & Perugini, M. (2017). Sequential Hypothesis Testing With Bayes Factors : Efficiently Testing Mean Differences. *Psychological Methods*, *22*(2), 322. https://doi.org/10.1037/met0000061

Slowiaczek, L. M. (1994). Semantic Priming in a Single-Word Shadowing Task. *The American Journal of Psychology*, *107*(2), 245. https://doi.org/10.2307/1423039

Tan, L., & Ward, G. (2008). Rehearsal in immediate serial recall. *Psychonomic Bulletin & Review*, *15*(3), 535-542. https://doi.org/10.3758/PBR.15.3.535

Thalmann, M., Souza, A. S., & Oberauer, K. (2018). How Does Chunking Help Working Memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Tse, C.-S. (2009). The role of associative strength in the semantic relatedness effect on immediate serial recall. *Memory*, *17*(8), 874-891. https://doi.org/10.1080/09658210903376250

Tse, C.-S., Li, Y., & Altarriba, J. (2011). The effect of semantic relatedness on immediate serial recall and serial recognition. *Quarterly journal of experimental psychology*, *64*(12), 2425-2437. https://doi.org/10.1080/17470218.2011.604787

Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance : A rejoinder on the binning procedure. *Behavior Research Methods*, *49*(2), 653-673. https://doi.org/10.3758/s13428-016-0721-5

Vergauwe, E., & Cowan, N. (2015). Attending to items in working memory : Evidence that refreshing and memory search are closely related. *Psychonomic Bulletin & Review*, *22*(4), 1001-1006. https://doi.org/10.3758/s13423-014-0755-6

Vergauwe, E., Hardman, K. O., Rouder, J. N., Roemer, E., Mcallaster, S., & Cowan, N. (2016). Searching for serial refreshing in working memory : Using response times to track the content of the focus of attention over time. *Psychonomic Bulletin & Review*, *23*(6), 1818-1824. https://doi.org/10.3758/s13423-016-1038-1

Vitevitch, M. S., & Luce, P. A. (2016). Phonological Neighborhood Effects in Spoken Word Perception and Production. *Annual Review of Linguistics*, *2*(1), 75-94. https://doi.org/10.1146/annurev-linguistics-030514-124832

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). *Why psychologists must change the way they analyze their data : The case of psi : Comment on Bem (2011).* https://doi.org/10.1037/a0022790

Walker, I., & Hulme, C. (1999). Concrete Words Are Easier to Recall Than Abstract Words : Evidence for a Semantic Contribution to Short-Term Serial Recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(5), 1256-1271.

Xie, W., & Zhang, W. (2017). Familiarity increases the number of remembered Pokémon in visual short-term memory. *Memory & Cognition*, *45*(4), 677-689. https://doi.org/10.3758/s13421-016-0679-7

Zevin, J. D., & Seidenberg, M. S. (2002). Age of Acquisition Effects in Word Reading and Other Tasks. *Journal of Memory and Language*, *47*(1), 1-29. https://doi.org/10.1006/jmla.2001.2834

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, *32*(1), 25-64.