Journal of Memory and Language 127 (2022) 104377

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

The semantic relatedness effect in serial recall: Deconfounding encoding and recall order[☆]

Benjamin Kowialiewski [*], Julia Krasnoff, Eda Mizrak, Klaus Oberauer

Department of Psychology, University of Zurich, Switzerland

ARTICLE INFO

Keywords:
Working Memory
Serial Recall
Semantics
Semantic Relatedness

ABSTRACT

The ability to store information in verbal working memory (WM) closely interacts with our linguistic knowledge. For instance, we can hold semantically related words (e.g., "cat, dog, bird") better in our WM than unrelated ones (e.g., "desk, pillow, mouse"). This study investigates boundary conditions of the beneficial effect of semantic relatedness of words on immediate memory for lists. Independently varying the encoding and recall order for lists of related words, we unraveled several mechanistic explanations of the semantic relatedness effect. We first tested a semantic cueing mechanism, according to which the recall of an item facilitates the recall of other semantically related items. We next disentangled an interactive activation versus feature overlap account of semantic relatedness. Whereas the former predicts that the semantic relatedness effect emerges from the temporal co-activation of related items, the latter predicts that it emerges from the superposition of semantic features bound to similar contexts. Our results demonstrate that semantic relatedness affects WM performance at the encoding stage of WM processing, which rules out semantic cueing as a plausible mechanism. Further, the temporal order in which words are presented was the most important determinant of the semantic relatedness effect, in agreement with the interactive activation account. This study supports a model in which semantic relatedness supports WM through interactive activation occurring in semantic long-term memory.

Introduction

Working memory (WM) is a core function of the cognitive system responsible for holding information briefly available for further processing. Numerous studies have shown that WM closely interacts with linguistic knowledge. This interaction occurs at all stages of language processing, including the phonological/sub-lexical (Gathercole et al., 1999; Majerus et al., 2004), lexical (Brener, 1940; Guitard et al., 2018; Hulme et al., 1991; Lewandowsky & Farrell, 2000; Roodenrys et al., 2002), and semantic stages (Poirier & Saint-Aubin, 1995). A central finding from this literature is the *semantic relatedness effect*, which describes the recall advantage for lists composed of semantically related words (e.g., "piano, guitar, violin, flute, accordion, saxophone") compared to unrelated words (e.g., "sanctuary, brother, fear, wall, hornet, sea"). This effect is usually large[1] (Kowialiewski & Majerus, 2020), thus providing clear evidence for the interaction between semantic knowledge and WM.

So far, there is no broad consensus regarding why this recall advantage for related vs unrelated lists of words occurs. With the present work we aim to contribute to a better understanding of the mechanisms that drive the semantic relatedness effect in WM. There are at least three different accounts that could explain it which we describe below.

According to the *interactive activation account*, semantic relatedness impacts WM maintenance at encoding when we process the words, through the activation of word representations in the long-term memory linguistic system. This long-term memory activation is then supposed to support the retrieval of information in WM (Cowan, 1999; Nee & Jonides, 2013; Oberauer, 2009). Based on this account, one way we can explain the semantic relatedness effect is by assuming that semantically related words support each other right at encoding. This is assumed to happen via spreading of activation within a semantic long-term memory network, or alternatively via redundant feedback activation between the lexical and semantic levels of language processing (Dell et al., 1997; Hofmann & Jacobs, 2014; Kowialiewski & Majerus, 2020). For instance, when the word "apple" is activated, the related words "plum" and "pear" will also receive some activation from "apple" and vice-versa. Before

☆ The research reported in this article was supported by funding from the Swiss National Science Foundation (SNSF, Grant No 100014_179002).

* Corresponding author at: Department of Psychology, Cognitive Psychology Unit, University of Zurich, Binzmühlestrasse 14/22, 8050 Zurich, Switzerland.
E-mail addresses: bkowialiewski@uliege.be, benjamin.kowialiewski@uzh.ch (B. Kowialiewski).

[1] Cohen's d ranging from ~1 to ~2.

https://doi.org/10.1016/j.jml.2022.104377
Received 30 March 2022; Received in revised form 15 September 2022; Accepted 16 September 2022
Available online 28 September 2022

plum is even presented, it is already activated to a certain extent. This pre-activation will boost plum's activation when plum is presented. Through interactive activation occurring at the encoding stage, related words benefit from higher activation levels compared to unrelated words. This would in turn increase their availability in semantic long-term memory, which in turn can support WM and thereby increase peoples' performance. The way the activation is spread in semantic memory is informed by semantic priming (e.g., the word "apple" can prime the word "pear"). Studies in the psycholinguistic domain show that temporal proximity is critical for a priming effect to emerge, as priming effects are usually reduced if a target and its related prime are interleaved by an unrelated distractor (e.g., "apple" and "pear" separated by "square") (Brunel & Lavigne, 2009; Lavigne et al., 2011, 2012). If semantic relatedness effects emerge from interactive activation, it follows that the temporal order in which semantically related items are encoded should be a driving factor for this effect.

A second explanation is offered by connectionist models of serial recall in which distributed representations of items are bound to distributed representations of contexts (Oberauer, Lewandowsky, et al., 2012). In these models, a word is encoded into WM through an association to its list position as the context, and all word-position associations needed for encoding a list are added up (i.e., superimposed) in a shared matrix of connection weights, as illustrated in Fig. 1. This newly formed item-context association is assumed to be the core content of WM. Both items and their contexts are represented in a distributed fashion. Similar representations, such as two adjacent positions or two similar words, are assumed to be represented by a larger proportion of shared features. In Fig. 1, overlapping features are represented by circles filled in grey. In this *feature overlap account*, between-item similarity increases memory for item information through mutual strengthening of shared connection weights when similar contents are associated to similar contexts. This occurs because the shared features of both representations are associated to parts of the same contexts multiple times, which creates a more robust representation (Oberauer, Farrell, et al., 2012). To illustrate this, consider three different scenarios: A, B and C, as depicted in Fig. 1. In scenario A, two different items, sharing no features, are encoded in two positions sharing a proportion of features. This case would be equivalent to associating "car" and "fit" to positions 1 and 2, respectively. In this case, although the two items are associated to adjacent positions, they share no overlapping representation in the weight matrix. Hence, no mutual strengthening occurs. The same happens in Scenario B, where two similar items are associated to two different contexts. This would be equivalent to encoding "flute" and "piano" at (for instance) positions 1 and 4. In Scenario C, both the items and their context share similar features. This would be the equivalent of encoding "apple" and "pear" at positions 1 and 2. In this case, the shared features at both the item and contextual levels lead to an overlap in the weight matrix, leading to a mutual strengthening between the two items and hence a more robust representation. In Fig. 1, this is illustrated by the thicker weights that associate the items to their context. Hence, memory models with distributed representations for items and contexts predict that similar items should lead to increased recall performance,[2] if and only if those similar items are also associated to sufficiently close contexts and/or serial positions.

The third potential explanation of the semantic relatedness effect is a *semantic cueing mechanism* occurring at recall. Recalling a word n would provide facilitatory semantic cues to recall word $n + 1$ if both words n and $n + 1$ are semantically related. For instance, retrieving "piano" would be easier if the previously recalled word was "guitar" than if it was "universe", if we assume that "guitar" would reactivate "piano" at retrieval. This could occur for instance through interactive activation of word representations as described above. Alternatively, when people retrieve an item, they also retrieve the semantic features associated to it, which might provide a cue to recall other, semantically related items. In the WM literature, the existence of such a cueing mechanism has been suggested following the observation that recall performance for semantically related items decreases less than for semantically unrelated items following a distractor tasks in the retention interval (Kowialiewski & Majerus, 2020; Neale & Tehan, 2007). This could be due to the use of a semantic cueing mechanism helping at retrieval, thereby counteracting the deleterious effect of interference. If most related items have been forgotten due to interference, they can still be retrieved through semantic cueing provided that at least one item of the list can be successfully recalled. Since unrelated items do not benefit from this semantic cueing mechanism, they suffer more strongly to interference. In the free recall literature, this mechanism is supported by the semantic clustering effect (Bousfield, 1953; Bousfield & Sedgewick, 1944; Howard & Kahana, 2002): People tend to recall semantically related words together, even if they have not been presented together. For instance, if multiple words from the "fruit" category are presented in a list, they will be more likely to be recalled at successive output positions, even if they were encoded at distant input positions. Together with the general beneficial effect of semantic relatedness in free recall tests, this has been interpreted as reflecting the existence of semantic cues helping the retrieval of related items. As it has been argued that free recall and serial recall are driven by similar mechanisms (Bhatarah et al., 2009; Ward et al., 2010), it is likely that such a semantic cueing mechanism could also contribute to the semantic relatedness effect in WM tasks. For instance, guitar might activate the "musical instruments" cue which can also be used for "piano" but not for "universe".

Most prior studies have investigated the semantic relatedness effect by comparing performance for purely semantically related lists of words to performance for unrelated lists. Whereas this procedure uncovered the semantic relatedness effect, the data cannot be used to test the three theoretical accounts against each other because all three accounts predict better performance for semantically related than unrelated lists. To test these accounts, we need to manipulate conditions that are predicted to drive the semantic relatedness effect by one of the accounts but not the others. Those critical conditions are described in the following: According to the interactive activation account, encoding the semantically related words temporally close to each other is the critical determinant driving the recall advantage for semantically related vs unrelated words. Hence, the temporal proximity of related words during encoding is crucial for this account. The feature overlap account postulates that the semantic relatedness effect emerges from the fact that related words are associated to a similar context, regardless of its nature (e.g., temporal, positional, spatial). The contextual proximity of related words is therefore the critical determinant for the feature overlap account. Finally, the semantic cueing mechanism predicts that semantically related words must be tested close to each other. Hence, the critical determinant is the proximity of related words in the recall order. Thus, these three accounts differ with regards to (1) whether they assume that the semantic relatedness effect is due to processes taking place at encoding (i.e., interactive activation and feature overlap accounts) or recall (i.e., semantic cueing mechanism) and (2) whether they attribute a special importance to the temporal dimension (i.e., the interactive activation account and the cueing account) as compared to other contexts, such as similar spatial locations (i.e., the feature overlap account). Here, we used these differences between the accounts to test them against each other.

We independently varied the order in which related vs unrelated words are encoded, the order in which they are recalled, and the context to which they are associated (temporal vs spatial). The general procedure is based on Kowialiewski et al. (2021) who presented related word in subgroups of words from two categories. In one condition, related items were presented in a *grouped* fashion, next to each other (e.g., "<u>piano</u>, <u>guitar</u>, <u>flute</u>, **lion**, **leopard**, **puma**"). In another condition, related items were presented in an *interleaved* fashion, at more distant

[2] Note that this holds only when WM performance is assessed at the item level, as opposed to the order level.

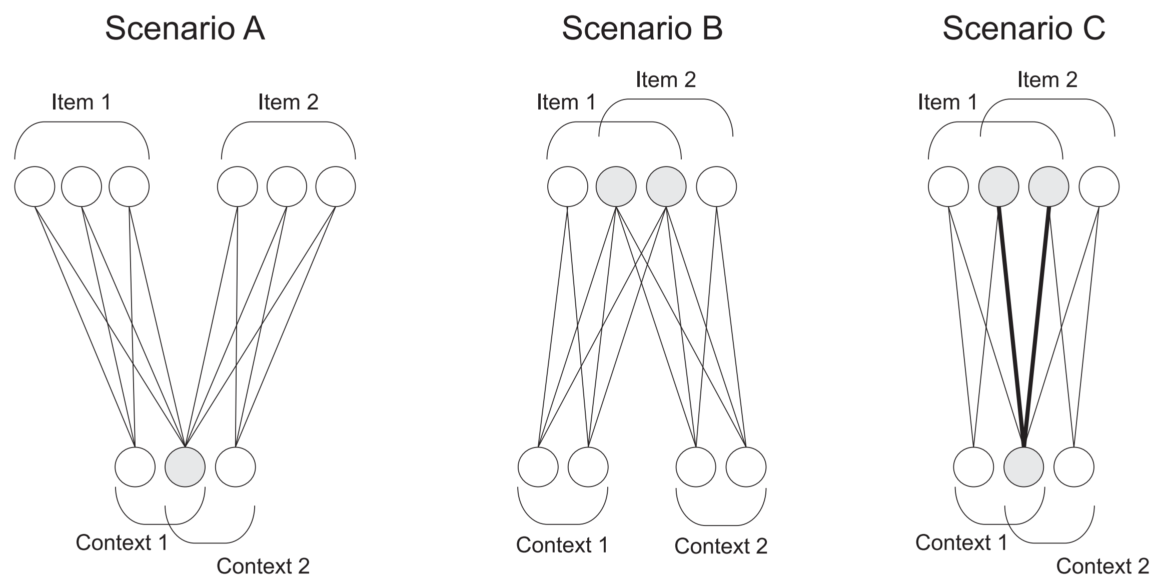


Fig. 1. Illustration of the Feature Overlap Phenomenon. *Note*. For illustration purpose, the model is illustrated using an unrealistically small number of features. Overlapping features are filled in grey.

serial positions (e.g., "piano, **lion**, guitar, **leopard**, flute, **puma**"). The semantic relatedness effect was reduced to half its size in the interleaved compared to the grouped condition. This result shows that proximity of the related word is important for the relatedness benefit, but because encoding and recall followed the same order, this study cannot be used to differentiate between the three theoretical accounts.

In the current work, we presented participants with lists of words from three semantic categories. In one condition, the related words were presented in groups (e.g., "*piano*, *flute*, leopard, tiger, **arm**, **leg**"). In another condition, the related words were presented following an interleaved pattern (e.g., "*piano*, leopard, **arm**, *flute*, tiger, **leg**"). Recall performance for these lists was compared to an unrelated condition (e.g., "rum, church, tongue, fox, wind, sister"). Experiment 1 solely aimed at validating the robustness of our experimental material, by replicating the classical semantic relatedness effect using this new procedure. In Experiment 2, we deconfounded the order of presentation from the order of recall, which allowed us to dissociate between a purely encoding vs recall origin of the semantic relatedness effect. This, in turn, enabled us to test the interactive activation and feature overlap accounts (which both assume encoding to be crucial for the semantic relatedness effect) against the semantic cueing account (which assumes retrieval to be crucial for the effect). Finally, in Experiment 3 we factorially varied the spatial and temporal contexts in which the words were encoded, which allowed to adjudicate between the interactive activation and feature overlap accounts. Note that these grouping manipulations somewhat differ from the standard semantic relatedness manipulation, in which all words of a list are semantically related, or unrelated. However, there is no a priori reason why the semantic-relatedness effects in our paradigm should not generalize to all types of semantic relatedness manipulations.

Experiment 1

We designed a variant of the task Kowialiewski, Gorin, et al. (2021) used. This was done to deconfound the proximity of semantically related word in the encoding order, the recall order, and the context they are associated to. The purpose of the first experiment was to test whether we still find a semantic relatedness effect with this task. In particular, we needed to ensure that the effect is still found when the words in the semantically related lists have only pair-wise relations (e.g., "*piano*, *flute*, leopard, tiger, **arm**, **leg**").

Data availability

All the materials, codes, data, and data analyses across all experiments have been made available on the Open Science Framework: https://osf.io/bqfy5/.

Method

Participants. Thirty-five young adults aged between 18 and 35 were recruited on the online platform Prolific (https://prolific.co/). Sample sizes for this and the subsequent experiments were first estimated based on previous studies investigating the impact of semantic and phonological similarity, leading to a base sample size of 35. In case the Bayes Factor (see statistical procedure) did not reach a sufficient level of evidence (BF $>$ 10 for either the null or the alternative hypothesis) concerning the critical effects of interest, we planned to recruit thirty-five more participants; this turned out to be unnecessary. All participants were English native speakers, reported no history of neurological disorder or learning difficulty, and gave their written informed consent before starting the experiment. The experiment was carried out in accordance with the ethical guidelines of the Faculty of Arts and Social Sciences at the University of Zurich.

Material. The stimuli involved a set of 144 English nouns. There were four nouns from each of 36 semantic categories. The stimuli have been made available on OSF (https://osf.io/bqfy5/). To construct the related and unrelated lists, we first identified categories that were potentially related to each other. The semantically related lists were created by selecting two words from each category, then combining three different categories to form 6-word lists. The related lists always followed the following pattern: AABBCC, with each letter referring to a semantic category (e.g., "*piano*, *flute*, leopard, tiger, **arm**, **leg**"). The unrelated lists were built by combining words from six different categories. Hence, the unrelated lists always followed the following pattern: ABCDEF (e.g., "rum, church, tongue, fox, wind, sister"). In both the related and unrelated condition, we avoided combining words from two categories that were semantically related to each other when selecting words from different categories. For instance, words from the "alcohol" category (i.e., rum – cognac – tequila – whiskey) were identified as being related with those from the "container" category (i.e., glass – cup – bowl – goblet). Each word appeared twice across the whole experiment, once in a related list, and once in an unrelated list. Using this procedure, it was possible to produce a total of 48 trials, with 24 trials in each condition (related, unrelated).

After creating the 48 lists, we ensured that a given word could not be presented more than once in each serial position (e.g., the word "green" twice at position 3). This was minimized by testing all possible within-list permutations without modifying the grouping structure of the related condition. The presentation order of the lists across the whole experiment was then randomized, with the further constraint that a given semantic condition (related, unrelated) could not be presented on more than three consecutive trials. The sequences were then exported to an external file in the ".JSON" format that could be read by the JavaScript programming language. This whole process repeated 60 times to create 60 different versions of the set of lists. Each participant was assigned randomly to one of these versions, by drawing a number from a random uniform distribution at the beginning of the experiment.

Our lists were created using semantic categories. The rationale behind this idea is that members of the same category are more closely related in a semantic feature space than items from different categories because humans naturally categorize information based on their semantic distance. We furthermore report the semantic relatedness of our lists using LSA-cosine values. These values were computed for three different types of pairs. First, we computed LSA values for all possible pairwise comparisons within each list of unrelated items. Second, we computed the LSA values for all pairwise comparisons of unrelated items in the related conditions. As our related lists were made up of separate pairs of related items, they contained both items that were related and items that were unrelated to any given word on the list. Third, we extracted LSA values for the related pairs themselves. Independent-samples T-Test show that the unrelated pairs in the related and unrelated conditions did not differ in terms of LSA values, and this absence of difference was supported by strong evidence ($BF_{01} = 12.3271$). The unrelated pairs in the unrelated condition differed from the related pairs, and this difference was supported by decisive evidence ($BF_{10} = 9.861e + 4751$). Finally, the unrelated pairs in the related condition differed from the related pairs, and this difference was supported by decisive evidence ($BF_{10} = 2.591e + 4056$).

Procedure. Throughout the whole experiment, participants were presented with six boxes arranged in a circular fashion, as displayed in Fig. 2. The words of each list were visually presented at the center of each box in lowercase letters. The words were presented at a pace of 600 ms/word. This presentation rate was identified as producing recall accuracy at a reasonable level (i.e., ~70 %), as indicated by a pilot study. The order in which the words appeared was always in a clockwise direction, as illustrated in Fig. 2, left panel. The first word appeared in the upper left box, followed by the second word in the upper right box, and so forth until all six words were presented. At retrieval, the words were cued following the same order as encoding by filling the corresponding square in grey. A prompt box appeared in the middle of the screen, prompting the participants to type their response, which appeared in that box. If participants did not know an item, they could leave the prompt box empty, resulting in an omission error. The retrieval phase is illustrated in Fig. 2, right panel. To submit a response, participants had to press the "Enter" key of their keyboard. This automatically led to the cueing of the next word. Participants performed three training trials before the beginning of the main experiment. The words used in the training trials were not used in the main experiment. Stimulus presentation was controlled via the JavaScript programming language, using the jQuery library, which allows efficient communications with HTML and CSS.

The cued recall procedure we used differs from the standard immediate serial recall procedure commonly used in the WM literature. This way of measuring WM however made the next experimental manipulations possible (see Experiments 2 & 3) to disentangle between the three theoretical accounts presented in the introduction. In Experiment 1, we first wanted to test whether the semantic relatedness effect can be reliably observed using this experimental procedure.

Scoring procedure. To determine the impact of the different semantic conditions (related, unrelated) on WM performance, we used a *strict serial recall criterion*. By this criterion, an item was considered 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. Additionally, we analyzed the data using an *item recall criterion*, in which an item was considered as correct, even if recalled at a wrong serial position. For the previous example, "Item1", "Item2", "Item3" and "Item5" would be considered as correct. Finally, we also report results using an *order recall criterion*. For this criterion, we computed the proportion of items recalled at their correct position out of the number of items recalled regardless of their position. As all scores converged toward the same conclusions across all experiments, we only report the strict serial recall analyses here for simplicity. Analyses using the item and order recall criteria are reported in **Appendix A** and **Appendix B**, respectively. We furthermore report in **Appendix C** the descriptive results after discarding participants with recall performance (i.e., items recalled in their correct serial position) above 90 %. These post-hoc analyses were performed to confirm that our results hold when we excluded any participants that might have cheated (which would have resulted in very high recall performance).

Statistical analysis. We conducted Bayesian analyses using the BayesFactor package implemented in R (Morey & Rouder, 2014). We use

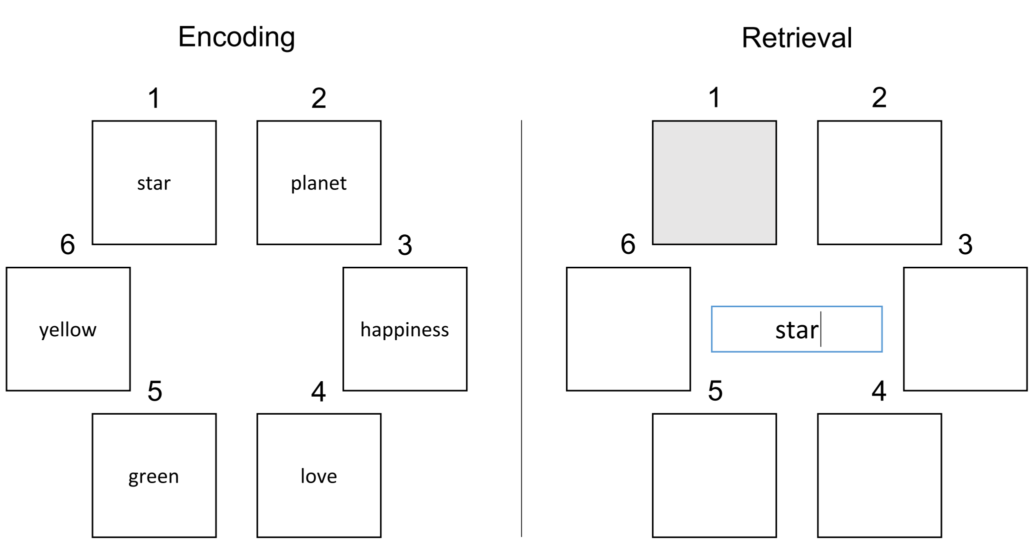


Fig. 2. Illustration of the Cued Recall Procedure Used in Experiment 1. *Note.* Left panel: During encoding, each word appeared sequentially following a clockwise direction. The numbers above each box (not displayed in the experiment) refer to the order of encoding. This illustration shows an example of a related condition. Right panel: At retrieval, the to-be-recalled words were cued, by filling the corresponding boxes in grey. Participants had to type the to-be-recalled word in a prompt box. The numbers above the boxes (not displayed in the experiment) refer to the order of retrieval.

the classification of strength of evidence proposed in previous studies (Jeffreys, 1998): 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. We ran Bayesian ANOVAs for within-subjects designs (i.e., mixed-effects models) as developed by Rouder et al. (2012). Specifically, we performed Bayesian model comparisons using a top-down testing procedure. We first started with the most complex possible structure including all effects, their interactions, a random intercept, and the random slopes for each main effect. We then progressively reduced the model's complexity by comparing the current best model to the same model without the specific effect of interest. We first started by testing the random slopes associated with each main effect, followed by the interactions and the main fixed effects until reaching the best possible model. We assessed each effect of interest by comparing the best model to the same model with or without the effect in question. To minimize error of model estimation, the number of Monte Carlo simulations generated was set to $N_{iterations} = 10^4$. 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 ($N_{iterations} = 10^4$). We used the default Cauchy prior distribution with a medium scale, $r = \frac{\sqrt{2}}{2}$. Note that to keep our models and the manuscript as straightforward as possible, we didn't include the serial-position effect in these analyses as we were merely interested in the overall semantic relatedness effect. Recall performance across serial positions is nonetheless reported in each graph. In addition, each graph reports the 95 % within-subject Confidence Intervals for each mean, following the recommendations made by Baguley (2012).

Results

Recall performance as a function of semantic condition (related, unrelated) was assessed using a Bayesian Repeated Measures ANOVA. Descriptive results are displayed in Fig. 3. Post-hoc analyses revealed that three participants had an average recall performance above 90 %. Descriptive results after discarding these participants are reported in

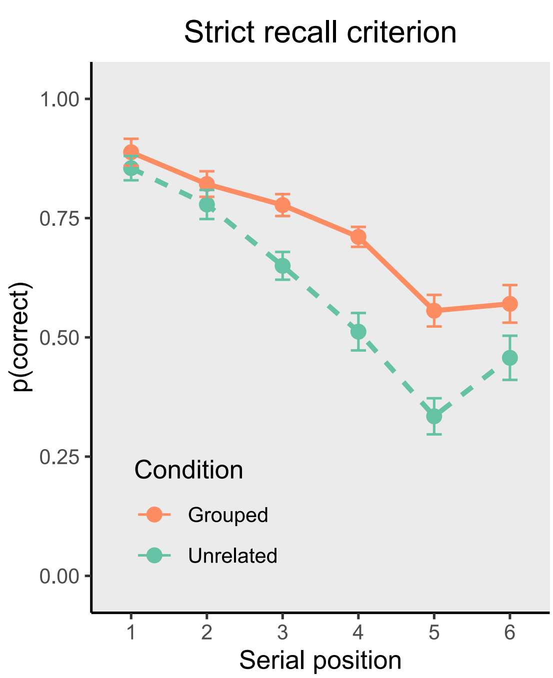


Fig. 3. Recall Performance as a Function of Serial Positions and Semantic Condition – Experiment 1. *Note.* Error bars represent 95% within-subject confidence intervals.

Appendix C.

The best model was the model including the main effect of condition and the random intercept. We found decisive evidence supporting this best model as compared to the model not including the main effect of condition ($BF_{10} = 1.832e + 6$). Specifically, there was a recall advantage in the related compared to the unrelated condition ($d = 1.433$, $CI_{95\%} = [0.894; 1.849]$, $M_{diff} = 0.12$).

Discussion

The results of this first experiment show that the semantic relatedness effect also occurs when semantic relatedness is manipulated using semantic sub-groups composed of two words instead of whole lists of semantically related words. In the following experiments, we made use of this semantic relatedness manipulation to test the different theoretical accounts presented in the introduction.

Experiment 2

In Experiment 2, we tested whether the semantic relatedness effect is modulated by the encoding and/or retrieval order of semantically related words. According to the interactive activation account, for semantically related words to activate each other at encoding, they should be presented temporally close to each other. The semantic relatedness effect should therefore be modulated by the temporal order in which the related items are presented at encoding, but not the retrieval order. The feature overlap account also predicts a larger semantic relatedness effect when related words are presented closely together, because in that case they are associated to similar temporal contexts. The order in which the related items are retrieved should not impact the magnitude of the effect. By contrast, the semantic cueing account predicts a stronger semantic relatedness effect when related words are recalled closely together, because recalling one word makes it easier to recall another, semantically related word. The temporal order in which the related items are presented should not influence the magnitude of the semantic relatedness effect, because the semantic cueing account assigns a specific role of retrieval in driving this semantic effect. These predictions are summarized in Table 1. As can be seen, manipulating the proximity between related items at encoding vs retrieval allowed us to disentangle only the semantic cueing account from the two other accounts (i.e., interactive activation and feature overlap). Note that we consider only pure accounts. Hybrid accounts combining multiple mechanisms are also possible. They lead to the prediction of additive effects of some of our proximity manipulations, or interactions. As none of these more complex patterns was observed, we will not consider such hybrid accounts.

To test the roles of encoding and of retrieval we made use of the same semantic sub-groups that were already used in Experiment 1 (e.g., "*piano*, *flute*, leopard, tiger, **arm**, **leg**"). We added an interleaved condition, in which the same words were presented in an interleaved fashion (e.g., "*piano*, leopard, **arm**, *flute*, tiger, **leg**"), so that semantically related words were separated much more during encoding. These conditions were compared to a neutral condition in which all the words were unrelated (e.g., "rum, church, tongue, fox, wind, sister"). Previous

Table 1
Summary of the Predictions Derived from Each Account – Experiment 2.

	Proximity manipulation	
Accounts	Encoding	Recall
Interactive activation	Yes	No
Feature overlap	Yes	No
Semantic cueing	No	Yes

Note. The values (Yes or No) in each cells indicate whenever the magnitude of the relatedness effect should be modulated as a function of the manipulated proximity (encoding, recall).

research has reported a reduced relatedness effect when increasing the positional distance between related words at encoding (Kowialiewski, Gorin, et al., 2021; Saint-Aubin et al., 2014). Based on these prior results, we expected the semantic relatedness effect to be substantially reduced in the interleaved vs grouped encoding condition.

The novel aspect of this experiment lies in the de-confounding of presentation order and recall order. The cued recall procedure we implemented (see for instance Fig. 2) allowed us to control the order in which participants recalled the words independently of the order in which they encoded them. On half of the trials, participants were asked to recall the words following a grouped pattern in which semantically related words were recalled in immediate succession (e.g., "*piano*, *flute*, leopard, tiger, **arm**, **leg**" OR "leopard, tiger, **arm**, **leg**, *piano*, *flute*"). On the other half of the trials, participants recalled the words following an interleaved pattern (e.g., "*piano*, leopard, **arm**, *flute*, tiger, **leg**" OR "**arm**, leopard, *piano*, **leg**, tiger, *flute*"), so that semantically related items were recalled far apart. The manipulation of presentation order (grouped vs interleaved) was orthogonal to the manipulation of recall order (grouped vs interleaved), creating four combinations of order for the semantically related lists.

The semantic cueing account predicts a stronger semantic relatedness effect in the grouped vs interleaved recall condition. In statistical terms, this should be reflected in an interaction between semantic relatedness and recall condition. If the semantic relatedness occurs exclusively at encoding, as predicted by the interactive activation and feature overlap accounts, this effect should only be modulated as a function of the encoding condition. In this case, we expect stronger semantic relatedness effects in the grouped vs interleaved encoding condition, resulting in an interaction between semantic relatedness and encoding condition.

Method

Participants. Thirty-five young adults aged between 18 and 35 were recruited on the online platform Prolific. All participants were English native speakers, reported no history of neurological disorder or learning difficulty, and gave their written informed consent before starting the experiment. The experiment was carried out in accordance with the ethical guidelines of the Faculty of Arts and Social Sciences at the University of Zurich.

Material. We used the same set of stimuli as in Experiment 1. The procedure used to create the grouped and unrelated sequences was identical to Experiment 1, with the exception that both conditions were now generated twice, resulting in a total of 96 to-be-remembered lists. There were therefore 48 grouped and 48 unrelated lists. Half of the grouped lists were used to create the interleaved condition. To do this, we changed the within-list order of the words so that the related words were now presented in an interleaved fashion (i.e., AABBCC → ABCABC). The unrelated lists were generated twice so that each related condition (grouped, interleaved) had an unrelated condition to which the recall order was matched (see procedure below). Hence, each word was presented four times across the whole experiment: once in a grouped condition, once in an interleaved condition, and twice in an unrelated condition. In each condition, half of the lists had to be recalled by the participants following a grouped structure, and half of the lists had to be recalled in an interleaved structure (see details in the procedure below).

Procedure. The procedure used in Experiment 2 is illustrated in Fig. 4. The encoding procedure was identical to the one used in Experiment 1. The recall procedure involved participants recalling the words in a specific order, depending on the different recall conditions. In the grouped encoding condition, on half the trials participants had to recall the related words following a grouped pattern (i.e., grouped recall condition). On the other half of the trials of the grouped encoding condition, they had to recall the items following an interleaved pattern (i.e., interleaved recall condition). A similar procedure was applied to the interleaved encoding condition, in which participants had to recall the words following either a grouped or an interleaved pattern. Note that participants were not informed beforehand of the direction of recall, and all conditions were presented in random order. This ensured that any potential expectation effects regarding the way items were recalled would be neutralized and would therefore not influence the way participants encoded the items. This is important, because previous experiments have shown that foreknowledge of recall condition influence recall performance (Guitard & Saint-Aubin, 2021).

To clarify, suppose participants had to recall words 1 through 6 and these words are related, following a semantically grouped structure such as "AABBCC". In cases where participants had to recall the words in a grouped pattern, recall order could be "341256", where the digits refer to the ordinal position of the items in the input sequence (i.e.,

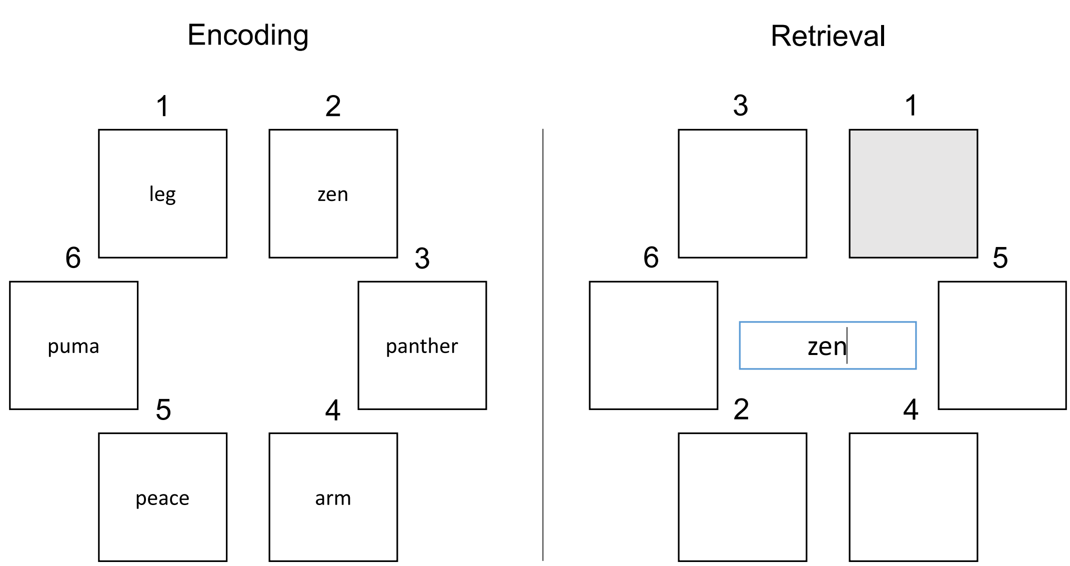


Fig. 4. Illustration of the Cued Recall Procedure Used in Experiment 2. *Note.* Left panel: During encoding, each word appeared sequentially in each box following a clockwise direction. The numbers above each box refer to the order of encoding. This illustration shows an example of a related – interleaved condition. Right panel: At retrieval, the to-be-recalled words were cued by filling the corresponding boxes in grey. Participants had to type the to-be-recalled word in a prompt box. The numbers above the boxes refer to the order of retrieval. In this example, the items had to be retrieved following a grouped pattern (starting to recall "zen" in input position 2 followed by "peace" in input position 5, and so on).

participants first had to recall the third-presented item, B, followed by the fourth-presented item, the other B, and so on). Another grouped recall order could be “561234”. If they had to recall the words in an interleaved pattern, recall order could be “135246” or “531642”. Next, suppose participants encoded a semantically interleaved sequence such as “ABCABC”. In cases where they had to recall the words in a grouped pattern, recall order could be “142536” or “361425”. If they had to recall the words following an interleaved pattern, a possible recall order is “312645”. Another possible recall order might be “231564”. The full list of recall pattern is reported in Table 2.

Whereas for the grouped and interleaved encoding conditions the recall order depends on the semantic structure, the unrelated lists, by definition, do not have such a semantic structure. A grouped or interleaved recall pattern can therefore not be directly applied to these lists. Instead, recall order in the unrelated lists was yoked to that of corresponding related lists. For instance, if participants encoded one sequence in the order “AABBCC” and had to recall it in the order “135246”, one corresponding unrelated list had to be recalled in the identical order. This way of recalling the unrelated lists ensured that any difference between any two semantic conditions (related or unrelated) could not be attributed to the recall order imposed. In other words, each related condition (grouped, interleaved) had a corresponding unrelated condition to which the recall order was matched. Hence, there were eight experimental conditions throughout the experiment: two relatedness conditions (related vs unrelated), two encoding conditions (grouped vs interleaved), and two recall conditions (grouped vs interleaved).

Results

Recall performance as a function of semantic relatedness (related, unrelated), encoding condition (grouped, interleaved) and recall condition (grouped vs interleaved) was assessed using a Bayesian Repeated Measures ANOVA. Results are displayed in Fig. 5. Post-hoc analyses revealed that one participant had an average recall performance above 90 %. Descriptive results after discarding this participant are reported in **Appendix C.**

Strict recall criterion. The best model was the model including all main effects, the interaction between encoding order and relatedness, and the random intercept. Model comparison with this best model provided decisive evidence supporting the main effect of encoding order (BF_{10} = 2.166e + 9), the main effect of relatedness (BF_{10} = 1.108e + 17) as well as the interaction between relatedness and encoding (BF_{10} = 6e + 6). The main effect of recall order was supported by moderate evidence (BF_{10} = 4). Critically, we found strong evidence *against* the interaction between relatedness and recall order (BF_{01} = 14). Moderate evidence was found against the triple interaction (BF_{01} = 7.333).

Specific Bayesian paired samples T-Tests indicated that the relatedness effect was robustly observed when items were semantically grouped during encoding. This occurred both when the recall pattern was grouped (BF_{10} = 2.613e + 4, d = 1.031, $CI_{95\%}$ = [0.559; 1.388], M_{diff} = 0.139) and when recall was interleaved (BF_{10} = 4.817e + 8, d = 1.657, $CI_{95\%}$ = [1.081; 2.121], M_{diff} = 0.146). In contrast, when items were presented in an interleaved fashion during encoding, the relatedness effect barely emerged. This was observed regardless of whether the items were recalled in a grouped (BF_{10} = 2.762, d = 0.425, $CI_{95\%}$ = [0.053; 0.726], M_{diff} = 0.038) or interleaved (BF_{10} = 0.411, d = 0.226, $CI_{95\%}$ = [-0.105; 0.536], M_{diff} = 0.021) fashion.

Discussion

The semantic relatedness effect was strongly observed when words were encoded following a grouped structure. When the words were encoded following an interleaved structure, the relatedness effect almost disappeared. Critically, the order in which participants recalled the items did not credibly influence the magnitude of the relatedness effect.

These results have two major implications. First, they suggest that the influence of semantic relatedness occurs at the encoding stage of WM. Second, they fail to support the existence of a semantic cueing mechanism acting at retrieval. If recalling the word “piano” provided additional cues to retrieve the word “guitar”, we would expect stronger semantic relatedness when participants recalled the related words close to each other than when they recalled these items in an interleaved fashion.

For simplicity, we considered a pure cueing account assuming that semantic relatedness affects WM performance only at retrieval. We did not envision an account in which people already detect semantic relationships at encoding and make use of those relationships at retrieval. This hybrid account, even if more plausible, would have predicted a larger semantic relatedness when participants recalled the items in a grouped vs interleaved fashion, and this specifically when the items were semantically grouped at encoding. This was not observed. Hence, regardless of the cueing account we considered, we failed to provide evidence for a hybrid account.

The present results are equally well predicted by the interactive activation and feature overlap accounts of semantic relatedness. For the interactive activation account, it does not matter how items are recalled. For a relatedness effect to emerge, the temporal order in which the related items are presented is critical to lead to their co-activation. For the feature overlap account, when semantically related words are presented in a grouped fashion, they are bound to similar contexts, which is a necessary condition to observe a beneficial effect at the item level. In the present experiment, the relevant context that people use as retrieval cues for the items is likely to be a mixture of temporal contexts (i.e., the temporal position of an item in the presented list) and spatial contexts (i.e., the location of the item’s frame), as temporal and spatial contexts were perfectly correlated. For semantically related lists presented in a grouped fashion, related words were presented both temporally and spatially close together, so for the feature overlap account it does not

Table 2
Encoding and recall patterns used in Experiment 2.

Encoding pattern											
Grouped						Interleaved					
A1	A2	B1	B2	C1	C2	A1	B1	C1	A2	B2	C2
Recall pattern											
Grouped						Interleaved					
A1	A2	B1	B2	C1	C2	A1	B1	C1	A2	B2	C2
A1	A2	C1	C2	B1	B2	A1	C1	B1	A2	C2	B2
B1	B2	A1	A2	C1	C2	B1	A1	C1	B2	A2	C2
B1	B2	C1	C2	A1	A2	B1	C1	A1	B2	C2	A2
C1	C2	A1	A2	B1	B2	C1	A1	B1	C2	A2	B2
C1	C2	B1	B2	A1	A2	C1	B1	A1	C2	B2	A2

Note. Each letter refers to a semantic category. The numbers “1″ and “2” refers to the first and second item within a given semantic category. Recall order in the unrelated lists (not reported here) was modelled based on the recall pattern reported in this table.

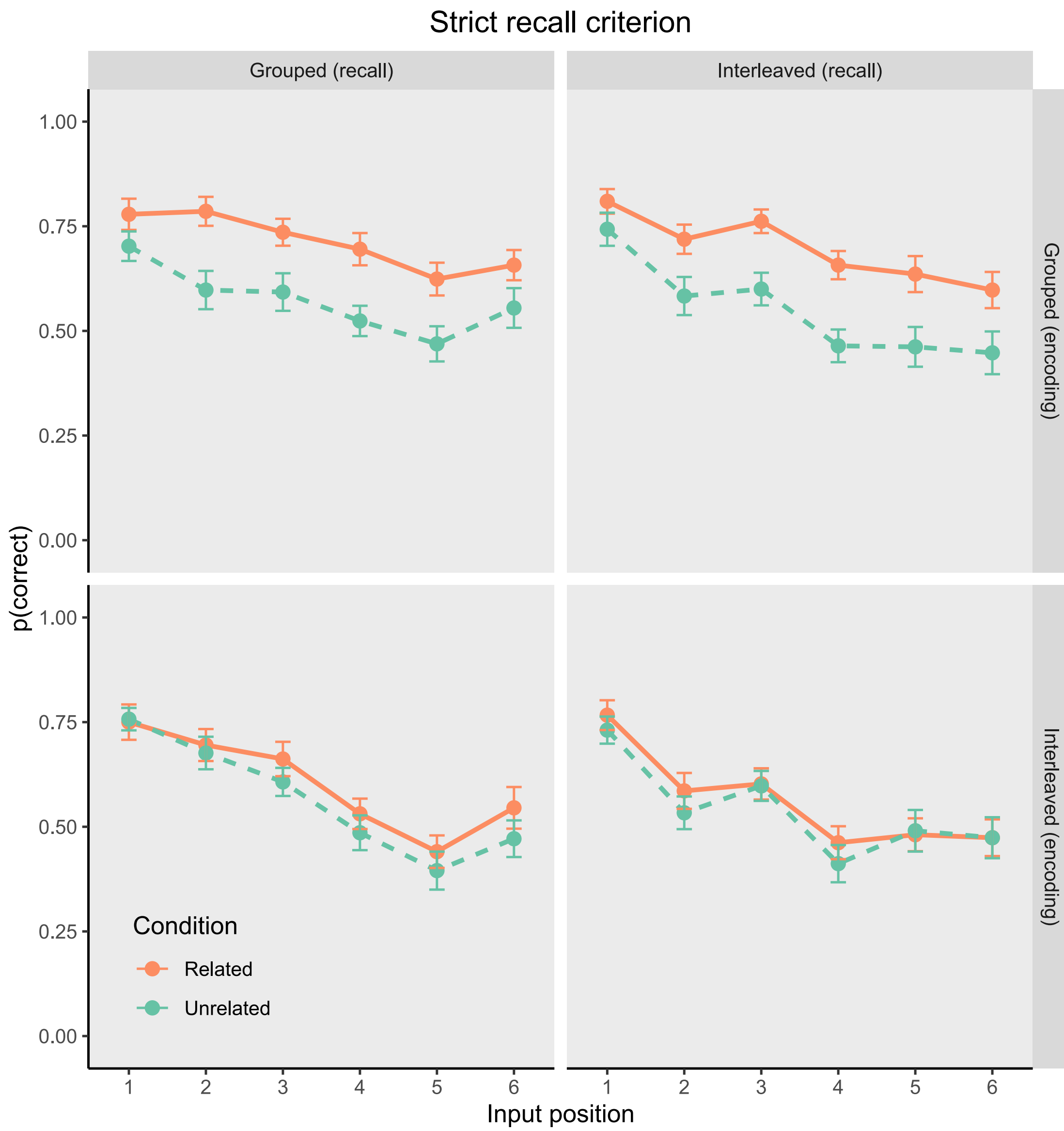


Fig. 5. Recall Performance as a Function of Input Position, Encoding Condition, Relatedness Condition and Recall Condition – Experiment 2. *Note*. Left panels: Recall following a grouped pattern. Right panels: Recall following an interleaved pattern. Upper panels: Encoding following a grouped pattern. Lower panels: Encoding following an interleaved pattern. Error bars represent 95% within-subject confidence intervals.

matter whether people predominantly relied on a temporal or a spatial context as retrieval cue – for both contexts, the grouped encoding condition leads to stronger mutual strengthening of semantically related items.

All in all, the present results are consistent with both the interactive activation and feature overlap accounts and challenge the semantic cueing account. In the next experiment, we test the interactive activation vs feature overlap accounts of semantic relatedness.

Experiment 3

In this experiment, our goal was to disentangle the interactive activation account from the feature overlap account. We systematically varied the dimension – temporal or spatial – along which the items were semantically grouped or semantically interleaved at encoding. Participants were presented with lists of pair-wise related vs unrelated words. As in Experiment 2, the related words were either presented in a grouped or in an interleaved fashion. The novel aspect of this experiment lies in the dimensions along which the grouped vs interleaved manipulation was performed. We took advantage of the fact that the words were presented in a spatial configuration, so that we could de-correlate their temporal from their spatial proximity at encoding by presenting the words in a random order across the boxes, rather than in a clockwise order. This enabled us to factorially vary the way the words were presented along the temporal vs spatial dimensions. One manipulation involved presenting the words in a grouped vs interleaved fashion along the temporal dimension, by presenting the related words temporally close or distant to each other. In another independent manipulation, the words were presented in a grouped vs interleaved fashion along the spatial dimension, by presenting the related words at adjacent vs distant spatial locations. The retrieval phase involved cueing each position in random order by their spatial location.

Straightforward predictions can be derived following this manipulation. These predictions are summarized in Table 3. According to the interactive activation account, the spatially grouped vs interleaved manipulation shouldn't affect the magnitude of the semantic relatedness effect. Instead, this account predicts that the relatedness effect should be maximally observed in the temporally grouped condition. Studies on

Table 3
Summary of the Predictions Derived from Each Account – Experiment 3.

	Proximity manipulation	
Accounts	Temporal	Spatial
Interactive activation	Yes	No
Feature overlap	Yes	Yes
Semantic cueing	No	No

Note. The values (Yes or No) in each cells indicate whenever the magnitude of the relatedness effect should be modulated as a function of the manipulated proximity (temporal, spatial).

semantic priming have shown that priming between two related words is disrupted by an intervening unrelated word (Brunel & Lavigne, 2009; Lavigne et al., 2011, 2012). This suggests that the mutual activation of semantically related words in long-term memory is disrupted by intervening unrelated words. In statistical terms, this prediction translates into an interaction between semantic relatedness and the temporally grouped vs interleaved conditions, and an absence of interaction between semantic relatedness and the spatially grouped vs interleaved conditions.

According to the feature overlap account, the relatedness effect should be stronger in the spatially grouped vs interleaved conditions. This is because the spatial dimension is now a relevant context which the to-be-remembered items are associated to, as the words are cued for retrieval by their spatial location. This should lead to stronger encoding into WM when items share common semantic and contextual (i.e., spatial) features (Oberauer et al., 2016). In statistical terms, we should therefore observe an interaction between semantic relatedness and the spatially grouped vs interleaved conditions. The feature overlap account could also predict an interaction between semantic relatedness and the temporal dimension. Even though our procedure made the temporal dimension less relevant, previous studies have nonetheless shown that the retrieval of positional information in WM is strongly anchored to the temporal dimension (Gilbert et al., 2017; Hartley et al., 2016). It is hence likely that participants will also use the temporal dimension as a relevant context. Critically, however, they could not rely exclusively on temporal context as retrieval cue because the spatial location is the only cue given to them at test. Therefore, the feature overlap account does not rule out an interaction between temporal proximity and semantic relatedness, but it must predict a (perhaps additional) interaction between spatial proximity and semantic relatedness. Although we already ruled out the semantic cueing account in the previous experiment, Experiment 3 gives the further opportunity to test this account. As the order in which the related items are recalled is completely random in Experiment 3, we expect that both the temporal and spatial manipulations should have no effect on the size of the relatedness effects, that is, an absence of interactions between any proximity manipulations and semantic relatedness.

Method

Participants. Thirty-five young adults aged between 18 and 35 were recruited on the online platform Prolific. All participants were English native speakers, reported no history of neurological disorder or learning difficulty, and gave their written informed consent before starting the experiment. The experiment was carried out in accordance with the ethical guidelines of the Faculty of Arts and Social Sciences at the University of Zurich.

Material. The list of stimuli was identical to Experiment 2.

Procedure. Related and unrelated word lists were presented. In the related lists we manipulated the temporal and spatial proximity of the related words. To manipulate the *temporal proximity*, the semantically related words were presented either temporally grouped or interleaved. To manipulate the *spatial proximity,* the semantically related items were presented spatially grouped or interleaved. These two dimensions (temporal and spatial proximity) were factorially manipulated. To make this more concrete, examples illustrating all four possible conditions are reported in Fig. 6. The presentation pattern followed the same constraints as in Experiment 2 (see Table 4). As in Experiment 2, the unrelated lists cannot be labelled as being "grouped" or "interleaved", as no related words were presented in these lists. We applied a similar constraint as in Experiment 2, by modelling the presentation schedules of the unrelated lists on those from the grouped and interleaved conditions.

As Experiment 2 failed to find any influence of recall order on the magnitude of the relatedness effect, we decided to set recall order as random. This procedure was intended to discourage participants from relying exclusively on the temporal dimension to perform the task, as the orders of encoding and recall were orthogonal to each other. To create the recall order for each trial, we first generated all possible permutations (without replacement) of the sequence [1,2,3,4,5,6]. We then randomly selected a subset of those permutations – one for each trial – which served to define the recall order. The first permutation was excluded, as it merely corresponds to the original presentation order. As in Experiment 2, participants were not informed beforehand of the direction of recall, and all conditions were presented in random order.

There were eight experimental conditions throughout the experiment, formed by crossing three variables: relatedness (related vs unrelated), temporal proximity (grouped vs interleaved), and spatial proximity (grouped vs interleaved).

Results

We assessed recall performance as a function of semantic relatedness (related, unrelated), temporal proximity (grouped, interleaved) and spatial proximity (grouped vs interleaved) using a Bayesian Repeated Measures ANOVA. Results from this analysis are displayed in Fig. 7. Post-hoc analyses revealed that no participant had an average recall performance above 90 %.

The best model was the model including all main effects, the interaction between relatedness and temporal proximity, the interaction between spatial proximity and temporal proximity, the random intercept, the random effect of temporal proximity, and the random effect of relatedness. When tested against this best model, we found decisive evidence supporting the main effect of temporal proximity (BF_{10} = 3.446e + 6), the main effect of spatial proximity (BF_{10} = 1,514) and the main effect of relatedness (BF_{10} = 5,192). We found decisive evidence supporting the interaction between temporal proximity and spatial proximity (BF_{10} = 1.891e + 10). There was decisive evidence supporting the interaction between temporal proximity and relatedness (BF_{10} = 2.394e + 7). We found strong evidence *against* the interaction between relatedness and spatial proximity (BF_{01} = 14.593). Finally, we found moderate evidence against the triple interaction (BF_{01} = 8.966).

Specific effects were explored using Bayesian paired-samples t-tests. When the related items were temporally grouped, there was a robust recall advantage for related vs unrelated lists, and this was observed to a similar extent regardless of whether the items were spatially grouped (BF_{10} = 5.078e + 5, d = 1.186, $CI_{95\%}$ = [0.715; 1.577], M_{diff} = 0.136) or spatially interleaved (BF_{10} = 3.571e + 4, d = 1.029, $CI_{95\%}$ = [0.58; 1.39], M_{diff} = 0.129). When items were temporally interleaved, the impact of semantic relatedness was much reduced, and evidence about its existence ambiguous, regardless of whether the items were spatially grouped (BF_{10} = 0.462, d = 0.24, $CI_{95\%}$ = [-0.1; 0.538], M_{diff} = 0.03) or spatially interleaved (BF_{10} = 1.041, d = 0.332, $CI_{95\%}$ = [-0.017; 0.629], M_{diff} = 0.039).

We found robust evidence supporting an interaction between spatial proximity and temporal proximity. Although this was not part of our main hypotheses, this interaction was also explored. Recall performance was higher when the sequences were overall encoded in a temporally grouped vs interleaved fashion, and this was observed both when the sequences were spatially grouped (BF_{10} = 1.894e + 7, d = 1.407, $CI_{95\%}$

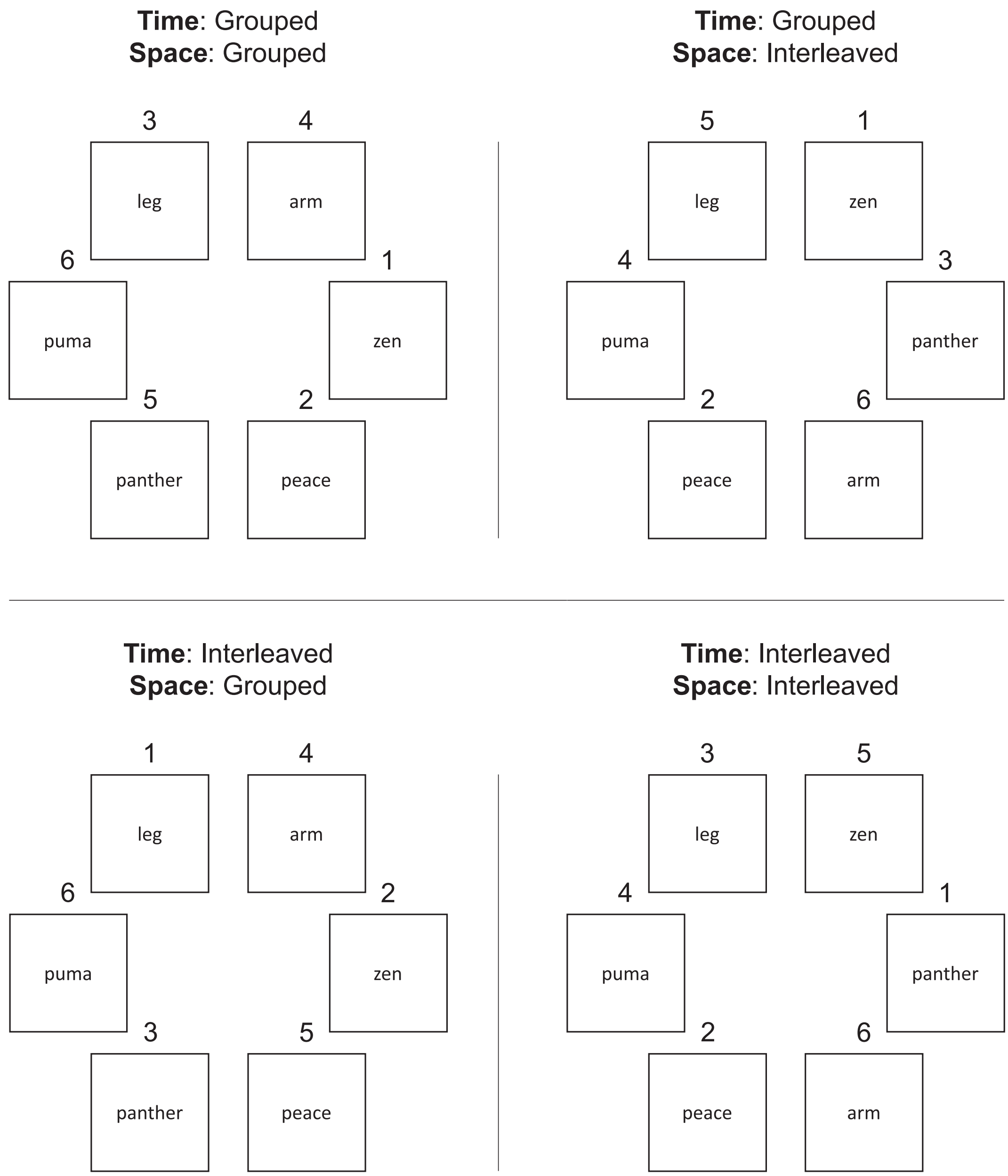


Fig. 6. Illustration of the Encoding Procedure Used in Experiment 3. *Note*. The numbers above each box correspond to the temporal order in which the items were presented. Left panels: Related items are spatially grouped. Right panels: Related items are spatially interleaved. Upper panels: Related items are temporally grouped. Lower panels: Related items are temporally interleaved.

Table 4
Encoding patterns used in Experiment 3.

Spatial dimension											
Grouped						Interleaved					
A1	A2	B1	B2	C1	C2	A1	B1	C1	A2	B2	C2
Temporal dimension											
Grouped						Interleaved					
A1	A2	B1	B2	C1	C2	A1	B1	C1	A2	B2	C2
A1	A2	C1	C2	B1	B2	A1	C1	B1	A2	C2	B2
B1	B2	A1	A2	C1	C2	B1	A1	C1	B2	A2	C2
B1	B2	C1	C2	A1	A2	B1	C1	A1	B2	C2	A2
C1	C2	A1	A2	B1	B2	C1	A1	B1	C2	A2	B2
C1	C2	B1	B2	A1	A2	C1	B1	A1	C2	B2	A2

Note. Each letter refers to a semantic category. The numbers “1″ and “2″ refers to the first and second item within a given semantic category. Recall order in the unrelated lists (not reported here) was modelled based on the recall pattern reported in this table.

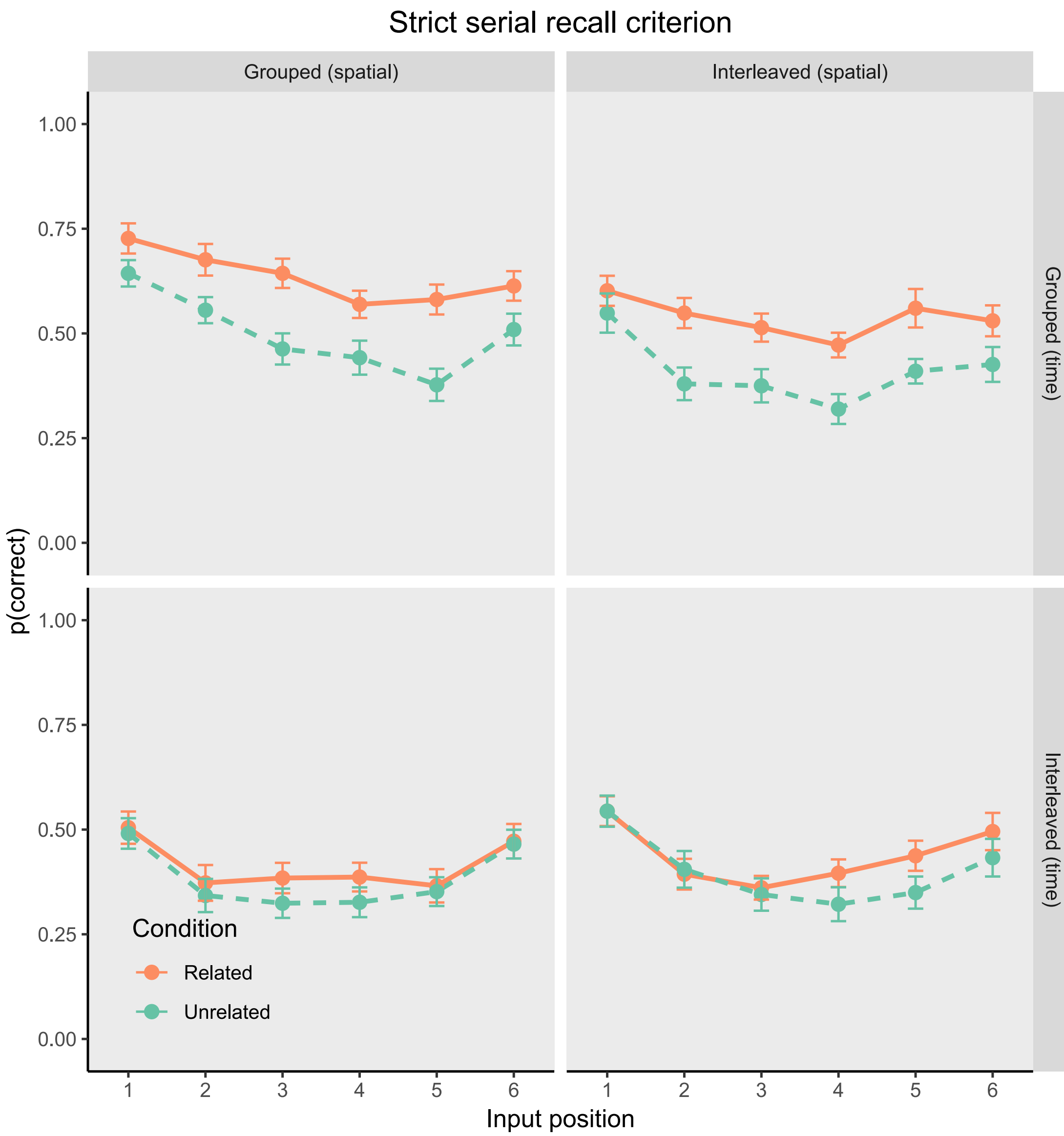


Fig. 7. Recall Performance as a Function of Input Position, Relatedness Condition, Timing Dimension and Spatial Dimension – Experiment 3. *Note.* Left panels: Spatial dimension following a grouped pattern. Right panels: Spatial dimension following an interleaved pattern. Upper panels: Timing dimension following a grouped pattern. Lower panels: Timing dimension following an interleaved pattern. Error bars represent 95% within-subject confidence intervals.

= [0.886; 1.82]) and spatially interleaved (BF_{10} = 495, d = 0.774, $CI_{95\%}$ = [0.355; 1.097]). We also found better recall performance when the sequences were spatially grouped vs interleaved. While this effect was credibly observed when the sequences were temporally grouped (BF_{10} = 5,217, d = 0.915, $CI_{95\%}$ = [0.482; 1.261]), no such credible evidence was found when the sequences were temporally interleaved (BF_{10} = 0.788, d = 0.303, $CI_{95\%}$ = [-0.604; 0.043]). Hence, the condition in which the sequences were grouped both along the temporal and spatial dimensions led to highest recall performance.

Discussion

The results of Experiment 3 showed that the magnitude of the semantic relatedness effect was modulated as a function of the temporal, but not spatial, dimension. Whereas the relatedness effect was credibly observed when the items were temporally grouped, it was not observed anymore when related words were temporally interleaved. The grouped vs interleaved presentation of the items along the spatial dimension did not modulate the magnitude of the semantic relatedness effect.

These results support the interactive activation account for the semantic relatedness effect: The effect is assumed to emerge through interactive activation of items within the linguistic system during encoding (Dell et al., 1997; Hofmann & Jacobs, 2014; Kowialiewski & Majerus, 2020). According to this account, temporal proximity of related words is the critical determinant leading to facilitative effects of semantic relatedness. The importance of temporal proximity for interactive activation in the semantic network was shown in previous studies reporting reduced semantic priming effects when a prime and its related target were interleaved by an unrelated distractor (Brunel & Lavigne, 2009; Lavigne et al., 2011, 2012).

Our results speak against the feature overlap account. The key result for this conclusion is the absence of interaction between semantic relatedness and spatial proximity. If semantic relatedness emerged from the superposition of semantic features, the overlapping contextual features provided by the spatial dimension should have resulted in stronger encoding into WM when the related words were spatially grouped (Oberauer et al., 2016; Oberauer & Lin, 2017). This was not observed. It could be argued that in our experiment, participants did not use the spatial dimension to perform the task and hence that semantic features were exclusively encoded along the temporal dimension. The fact that we observed a robust overall recall advantage in the spatially grouped vs interleaved encoding condition counters that claim, because it shows that participants did use the spatial context as retrieval cue. In the next section, we discuss the theoretical implications of these results.

The results of Experiment 3 further argue against the semantic cueing mechanism. In this experiment, encoding and recall were fully orthogonal to each other. Therefore, we should have observed a consistent relatedness effect, regardless of the way participants encoded the items. The fact that the semantic relatedness effect specifically interacted with the time of encoding further rejects an exclusive retrieval origin of the semantic relatedness effect.

General discussion

We tested three mechanistic explanations for the semantic relatedness effect in WM: the interactive activation account, the feature overlap account, and the semantic cueing account. To this end, we compared lists with pairs of semantically related words to lists of unrelated words and found (a) we replicate the relatedness benefit, (b) the relatedness benefit depends on presenting the related words in temporal proximity. It does not depend on recalling the words in temporal proximity, and not on the words being associated to spatially close contexts. Taken together, these results support the assumption that related words activate each other at encoding through spreading activation in a semantic network. They rule out the hypothesis that semantically related words act as retrieval cues for each other during test, and they also rule out an explanation of the semantic relatedness benefit through mutual strengthening of similar contents associated to similar contexts.

Semantic relatedness benefits WM through activated long-term memory

Our results suggest that the semantic relatedness emerges from a reactivation process occurring exclusively at encoding, in which related items reactivate each other in a semantic network, or via their shared semantic features (Dell et al., 1997; Hofmann & Jacobs, 2014), as assumed by the interactive activation account. During encoding of words in WM, the semantic network (i.e., semantic long-term memory) gets activated and semantically related words benefit from the reactivation of the same network as compared to unrelated items. This idea is generally congruent with models assuming that long-term memory supports the short-term retention of information (Cowan, 1999; Majerus, 2019; Nee & Jonides, 2013; Oberauer, 2002, 2009). According to these models, when an item is presented, this item becomes activated in the long-term memory system. Items that benefit from a stronger long-term memory activation will be easier to access at the retrieval stage of WM due to their higher availability. The plausibility of such a mechanism has already been shown through computational modeling (Haarmann & Usher, 2001; Kowialiewski, Lemaire, et al., 2021; Kowialiewski & Majerus, 2020). The novelty of the present study was to provide evidence supporting this mechanism by showing that the semantic relatedness effect depends on the temporal proximity in which related items are encoded and does not occur at retrieval.

Previously, it was shown that semantically related items are subject to less forgetting than unrelated items in WM tasks involving secondary-task processing during the retention interval (Kowialiewski & Majerus, 2020; Neale & Tehan, 2007). This finding has been explained through a semantic cueing mechanism, according to which the retrieval of an item could provide cues to recall the other semantically related items of the list. However, our results do not support the semantic cueing mechanism. An alternative explanation for the reduced interference effects for semantically related items could be the following: Semantically related items strengthen each other through activated semantic long-term memory. If the content of WM is strongly degraded by interference, recall can still be performed by retrieving the items from long-term memory, which is known to be relatively insensitive to interfering tasks (Glanzer & Cunitz, 1966). Hence, related items could be more resistant to interference due to their greater availability in long-term memory as compared to unrelated items.

Why does temporal proximity matter?

Similar to semantic priming studies, we showed that the semantic relatedness benefit depends on temporal proximity during encoding. When related items were not presented as temporal neighbors, we did not observe the semantic relatedness benefit. To explain why, studies from multiple priming experiments in the psycholinguistic domains can be informative. In these multiple priming experiments, a target (e.g., "tiger") is preceded by two primes, both of which can be related ("lion, puma"), unrelated ("tree, computer"), or only one item being related ("cheetah, shelve" or "taxi, panther"). These studies have shown substantially reduced priming effects when a related prime and its target are interleaved with an unrelated item (see Lavigne et al., 2013 for a review). In contrast, when the prime is temporally close to its neighbor, priming effects are maximally observed. These patterns of results were successfully modeled by Brunel and Lavigne (2009) using a cortical network model. They have shown that the reduced priming effect in an interleaved condition can occur if inhibition processes (implemented through global inhibition in their model) between semantically unrelated words are assumed. Basically, when presenting the word "lion", the word "tiger" will be pre-activated. If an unrelated word such as "computer" is interleaved between "lion" and "tiger", this unrelated word exerts an inhibitory effect, thereby reducing potential priming effects afforded through spreading activation. A similar mechanism could explain the temporal proximity effect of semantic relatedness in the context of WM tasks: Having unrelated items interleaved in between the related items might exert some inhibitory processes, thereby reducing the related items' activation level, or might reduce the probability to detect the presence of semantic relationships between items. If this explanation is correct, the critical variable would not be temporal proximity, but the number of intervening events.

Semantic cueing mechanism: Serial recall vs Free recall

The lack of evidence for the semantic cueing mechanism contrasts with what has been observed in the long-term memory domain with free recall tasks. Free recall over the long-term appears to be strongly constrained by between-item semantic relatedness. This has been shown through the semantic clustering effect (Bousfield, 1953; Bousfield & Sedgewick, 1944; Howard & Kahana, 2002), the effect that related items tend to be recalled at successive output positions. These results have led to the development of long-term memory models in which semantic cues are encoded and help retrieving other items sharing the same semantic cues (Polyn et al., 2009). The reason why we do not see a similar pattern might be due to the differences between serial recall and free recall, and specifically the way items need to be retrieved. In free recall tasks, semantic cues can be used to recover list items, and the order in which these items are recovered does not matter. However, in serial recall, using an item's meaning as a cue to other list items is risky because it potentially messes up the order; this is especially the case when the items' order is arbitrary, as in most serial recall tasks. Therefore, participants probably choose not to rely on semantic cueing in tasks that demand memory for order. Relying on context cues to maintain items' positions is a much safer strategy, which is assumed by most models of serial recall (Burgess & Hitch, 1999; Henson, 1998; Lewandowsky & Farrell, 2008; Oberauer, Lewandowsky, et al., 2012; Oberauer & Lewandowsky, 2011). In contrast, in free recall it doesn't matter in which order the items are recalled. Using items' meaning to cue the next to-be-remembered word is therefore an efficient strategy.

Feature overlap: The special case of semantic similarity

Our study challenges the feature overlap account as an explanation for the semantic relatedness effect. In connectionist models of serial recall assuming encoding into WM as occurring between sets of distributed representations, similarity benefits item memory via mutual strengthening of similar contents associated to similar contexts. If the similar items are encoded at distant contexts, the beneficial effect of similarity should be strongly reduced. The fact that the semantic relatedness effect was not more strongly observed when items were spatially grouped in Experiment 3, despite participants using the spatial dimension to perform the task, does not support feature overlap as a plausible explanation.

However, our study does not rule out feature overlap as a general WM mechanism to explain similarity effects. Feature overlap provides a powerful and comprehensive explanation of manipulations of similarity along other dimensions than the semantic one. Feature overlap predicts two effects of between-item similarity: Similar items, when bound to overlapping contexts, support each other, thereby improving item memory. At the same time, similar items are more likely to be confused with each other, thereby reducing order memory (Oberauer, Lewandowsky, et al., 2012). Studies in the visual (Guitard & Cowan, 2020; Jalbert et al., 2008; Logie et al., 2016; Saito et al., 2008), phonological (Fallon et al., 2005; Gupta et al., 2005; Lian et al., 2004; Neale & Tehan, 2007; Nimmo & Roodenrys, 2004), and auditory (Visscher et al., 2007; Williamson et al., 2010) domains observed that between-item similarity, while enhancing the ability to recall item information, also increases the prevalence of order errors, in line with the predictions of the feature overlap account. What our results show is that the core principles of feature overlap do not apply to semantic similarity, and therefore do not explain the influence of semantic relatedness.

This conclusion is reaffirmed by another observation: Semantic similarity does not affect order memory. In connectionist models of serial recall, encoding similar items results in increased order errors due to a problem of discriminability between WM traces, a phenomenon called interference by confusion (Oberauer et al., 2016). Hence, according to the feature overlap account, similarity between list items always comes with a cost for serial order memory. This effect is regularly observed for between-item similarity on other dimensions. By contrast, semantic similarity does not increase serial order errors in a reliable manner (Baddeley, 1966; Neale & Tehan, 2007; Saint-Aubin & Poirier, 1999; Tse et al., 2011), confirming that the principles of the feature overlap account do not apply to semantic similarity.

One possible way to explain why semantic similarity a special case is that meaning is encoded through the activation of lexical and semantic representations in long-term memory, and WM does not encode meaning. Specifically, representations of meaning are not bound to the contexts – for instance, temporal position or spatial position – through which information in WM is accessed. If we make this simplifying assumption, then it logically follows that increasing items' semantic relatedness should not decrease their discriminability in WM tasks. This idea fits well with the explanation that semantic similarity exclusively impacts WM through higher availability in long-term memory. Such an account predicts that semantic similarity should specifically impact item memory, by making related items more available. In contrast, this higher activation does not lead to impaired order memory, as it is supposed to remain outside of the core WM representation.

Conclusion

In summary, semantic relatedness benefit in WM emerged when related words were in temporal proximity during encoding. Neither the spatial proximity nor recall proximity modulated the magnitude of the semantic relatedness benefit. These results can be explained by an interactive activation account that suggests semantically related words boost each other's activation during encoding.

Open Science statement:

All the data and codes have been made available on the Open Science Framework: https://osf.io/bqfy5/.

CRediT authorship contribution statement

Benjamin Kowialiewski: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Julia Krasnoff:** Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Eda Mizrak:** Conceptualization, Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **Klaus Oberauer:** Conceptualization, Methodology, Software, Validation, Formal analysis, Resources, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

Results using the item recall criterion across all experiments.

Experiment 1.

The best model was the model including the main effect of condition and the random intercept. Decisive evidence supported this best model over the model not including the main effect of condition ($BF_{10} = 5.03e + 5$). Items in the grouped condition were better recalled compared to the items in the unrelated condition ($CI_{95\%} = [1.16; 2.226]$, d = 1.756).

Experiment 2.

The best model was the model including the main effects of relatedness and encoding, the interaction between relatedness and encoding and the random intercept. Model comparison with this best model provided decisive evidence supporting the effect of encoding order ($BF_{10} = 1.698e + 6$), relatedness ($BF_{10} = 9.328e + 16$), and the interaction between relatedness and encoding ($BF_{10} = 7.06e + 5$). Importantly, we found moderate evidence against the interaction between relatedness and recall ($BF_{01} = 9.736$).

Experiment 3.

The best model was the model including the main effects of relatedness and temporal proximity, the interaction between semantic relatedness and temporal proximity, and the random intercept. When compared to this best model, we found decisive evidence supporting the effect of relatedness ($BF_{10} = 3.265e + 24$) and decisive evidence supporting the effect of temporal proximity ($BF_{10} = 6.244e + 6$). Moderate evidence was found against the effect of spatial proximity ($BF_{01} = 6.582$). We found strong evidence against the interaction between relatedness and spatial proximity. There was strong evidence supporting an interaction between relatedness and temporal proximity ($BF_{10} = 26.978$). The interaction between temporal and spatial proximity was ambiguous ($BF_{10} = 1.136$). Finally, there was moderate evidence against the triple interaction ($BF_{01} = 8.568$).

Appendix B

Results using the order recall criterion across all experiments.

Experiment 1.

The best model was the full model. Strong evidence supported this best model over the model not including the main effect of condition ($BF_{10} = 10.73$). Items in the grouped condition were better recalled compared to the items in the unrelated condition ($CI_{95\%} = [0.175; 0.878]$, d = 0.568).

Experiment 2.

The best model was the model including the main effects of relatedness and encoding, the interaction between relatedness and encoding, the random intercept and the random slope of encoding. Model comparison with this best model provided moderate evidence supporting the effect of encoding order ($BF_{10} = 3.018$), and decisive evidence supporting the relatedness effect ($BF_{10} = 844.537$). The interaction between relatedness and encoding was ambiguous ($BF_{10} = 1.425$). We found strong evidence against the interaction between relatedness and recall ($BF_{01} = 12.999$).

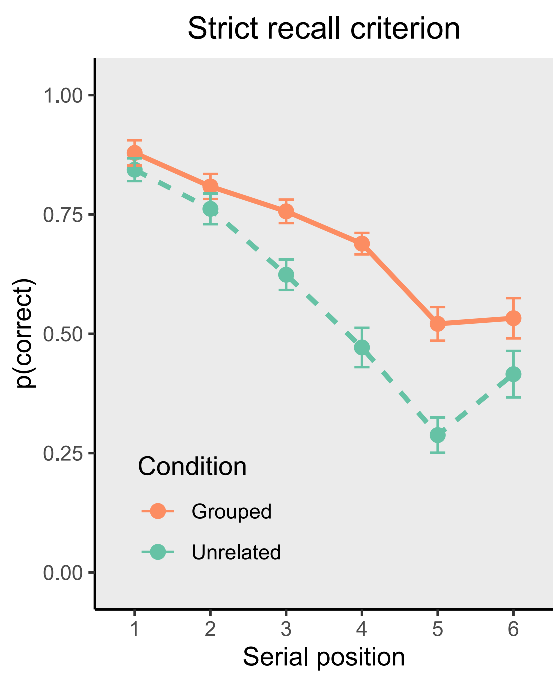
Experiment 3.

The best model was the model including all main effects, the interaction between semantic relatedness and temporal proximity, the interaction between temporal proximity and spatial proximity, the random intercept and the random slope of semantic relatedness. When compared to this best model, we found decisive evidence supporting the effect of relatedness ($BF_{10} = 336.27$), decisive evidence supporting the effect of temporal proximity ($BF_{10} = 7.156e + 25$) and strong evidence supporting the effect of spatial proximity ($BF_{10} = 46.079$). We found strong evidence supporting the interaction between spatial and temporal proximity ($BF_{10} = 3.483e + 8$). The interaction between semantic relatedness and temporal proximity was supported by decisive evidence ($BF_{10} = 3.941e + 4$). Critically, we found strong evidence against the interaction between semantic relatedness and spatial proximity ($BF_{01} = 14.436$). There was moderate evidence against the triple interaction ($BF_{01} = 6.341$).

Appendix C

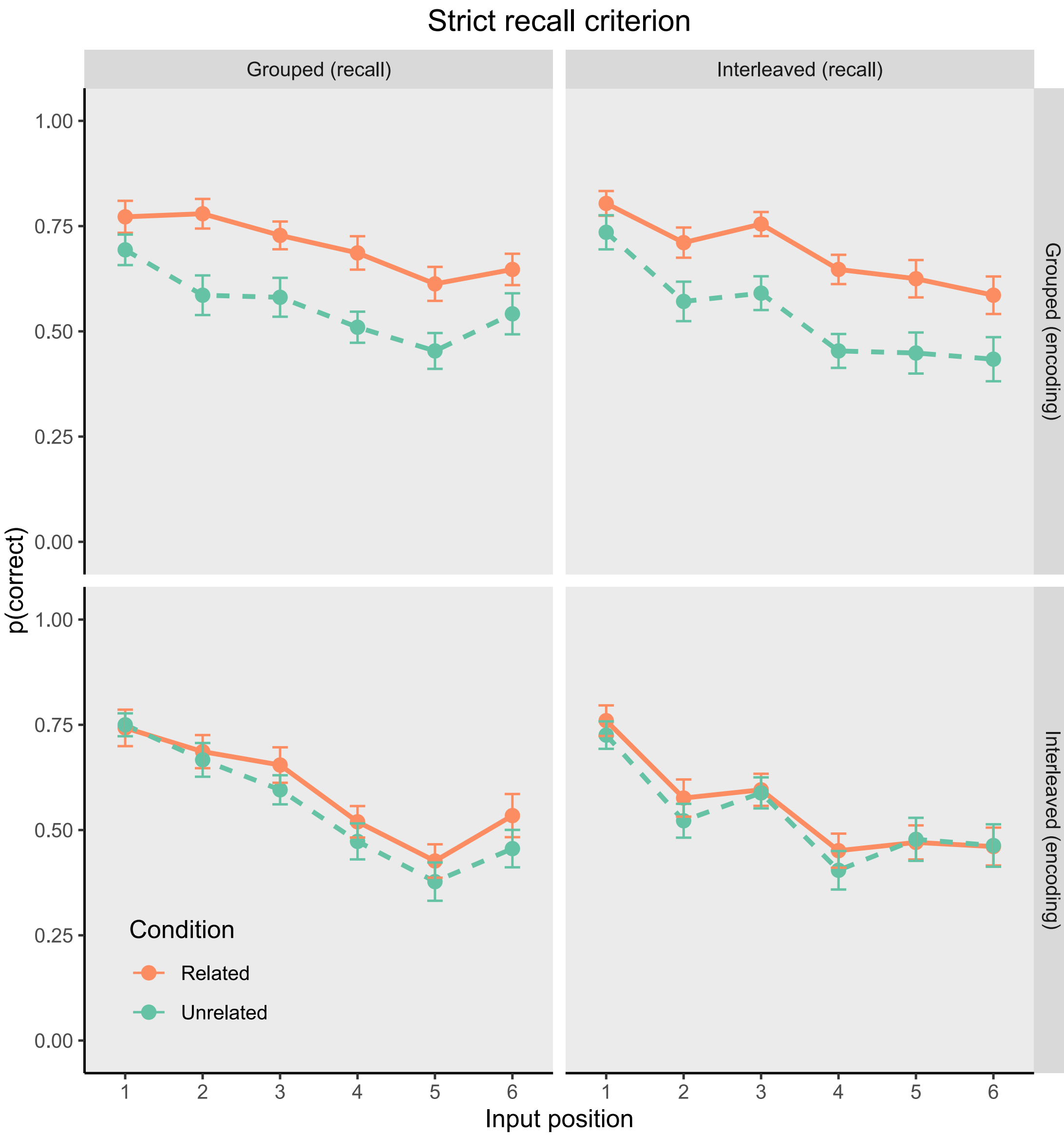
Results after discarding participants with recall performance above 90 %.

Experiment 1.



.

Experiment 2.



.

References

Baddeley, A. D. (1966). Short-term Memory for Word Sequences as a Function of Acoustic, Semantic and Formal Similarity. *Quarterly Journal of Experimental Psychology, 18*(4), 362–365. https://doi.org/10.1080/14640746608400055

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

Bhatarah, P., Ward, G., Smith, J., & Hayes, L. (2009). Examining the relationship between free recall and immediate serial recall : Similar patterns of rehearsal and similar effects of word length, presentation rate, and articulatory suppression. *Memory & Cognition, 37*(5), 689–713. https://doi.org/10.3758/MC.37.5.689

Bousfield, W. A. (1953). The Occurrence of Clustering in the Recall of Randomly Arranged Associates. *The Journal of General Psychology, 49*(2), 229–240. https://doi.org/10.1080/00221309.1953.9710088

Bousfield, W. A., & Sedgewick, C. H. W. (1944). An Analysis of Sequences of Restricted Associative Responses. *The Journal of General Psychology, 30*(2), 149–165. https://doi.org/10.1080/00221309.1944.10544467

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

Brunel, N., & Lavigne, F. (2009). Semantic Priming in a Cortical Network Model. *Journal of Cognitive Neuroscience, 21*(12), 2300–2319. https://doi.org/10.1162/jocn.2008.21156

Burgess, N., & Hitch, G. J. (1999). Memory for Serial Order : A Network Model of the Phonological Loop and Its Timing. *Psychological Review, 106*(3), 551–581.

Cowan, N. (1999). *An Embedded-Processes Model of working memory. In Models of working memory : Mechanisms of active maintenance and executive control* (pp. 62–101). (Cambridge University Press.

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. https://doi.org/10.1037/0033-295X.104.4.801

Fallon, A. B., Mak, E., Tehan, G., & Daly, C. (2005). Lexicality and phonological similarity : A challenge for the retrieval-based account of serial recall? *Memory, 13* (3–4), 349–356. https://doi.org/10.1080/09658210344000215

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic Influences on Short-Term Memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*(1), 84–95.

Gilbert, R. A., Hitch, G. J., & Hartley, T. (2017). Temporal precision and the capacity of auditory–verbal short-term memory. *Quarterly Journal of Experimental Psychology, 70* (12), 2403–2418. https://doi.org/10.1080/17470218.2016.1239749

Glanzer, M., & Cunitz, A. (1966). Two Storage Mechanisms in Free Recall. *Journal of Verbal Learning and Verbal Behavior, 5*, 351–360.

Guitard, D., & Cowan, N. (2020). Do we use visual codes when information is not presented visually? *Memory & Cognition, 48*(8), 1522–1536. https://doi.org/10.3758/s13421-020-01054-0

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

Guitard, D., & Saint-Aubin, J. (2021). The irrelevant speech effect in backward recall is modulated by foreknowledge of recall direction and response modality. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale, 75*(3), 245–260. https://doi.org/10.1037/cep0000248

Gupta, P., Lipinski, J., & Aktunc, E. (2005). Reexamining the phonological similarity effect in immediate serial recall : The roles of type of similarity, category cuing, and item recall. *Memory & Cognition, 33*(6), 1001–1016. https://doi.org/10.3758/BF03193208

Haarmann, H., & Usher, M. (2001). Maintenance of semantic information in capacity-limited item short-term memory. *Psychonomic Bulletin & Review, 8*(3), 568–578. https://doi.org/10.3758/BF03196193

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.

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

Howard, M. W., & Kahana, M. J. (2002). When Does Semantic Similarity Help Episodic Retrieval? *Journal of Memory and Language, 46*(1), 85–98. https://doi.org/10.1006/jmla.2001.2798

Hulme, C., Maughan, S., & Brown, G. D. A. (1991). Memory for familiar and unfamiliar words : Evidence for a long-term memory contribution to short-term memory span. *Journal of Memory and Language, 30*(6), 685–701. https://doi.org/10.1016/0749-596X(91)90032-F

Jalbert, A., Saint-Aubin, J., & Tremblay, S. (2008). Short Article : Visual Similarity in Short-Term Recall for Where and When. *Quarterly Journal of Experimental Psychology, 61*(3), 353–360. https://doi.org/10.1080/17470210701634537

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

Kowialiewski, B., Gorin, S., & Majerus, S. (2021). Semantic knowledge constrains the processing of serial order information in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*. https://doi.org/10.1037/xlm0001031

Kowialiewski, B., Lemaire, B., & Portrat, S. (2021). How does semantic knowledge impact working memory maintenance? Computational and behavioral investigations. *Journal of Memory and Language, 117*(104208). https://doi.org/10.1016/j.jml.2020.104208

Kowialiewski, B., & Majerus, S. (2020). The varying nature of semantic effects in working memory. *Cognition, 202*, Article 104278. https://doi.org/10.1016/j.cognition.2020.104278

Lavigne, F., Chanquoy, L., Dumercy, L., & Vitu, F. (2013). Early dynamics of the semantic priming shift. *Advances in Cognitive Psychology, 9*(1), 1–14. https://doi.org/10.5709/acp-0126-9

Lavigne, F., Dumercy, L., Chanquoy, L., Mercier, B., & Vitu-Thibault, F. (2012). Dynamics of the semantic priming shift : Behavioral experiments and cortical network model. *Cognitive Neurodynamics, 6*(6), 467–483. https://doi.org/10.1007/s11571-012-9206-0

Lavigne, F., Dumercy, L., & Darmon, N. (2011). Determinants of Multiple Semantic Priming : A Meta-analysis and Spike Frequency Adaptive Model of a Cortical Network. *Journal of Cognitive Neuroscience, 23*(6), 1447–1474. https://doi.org/10.1162/jocn.2010.21504

Lewandowsky, S., & Farrell, S. (2000). A redintegration account of the effects of speech rate, lexicality, and word frequency in immediate serial recall. *Psychological Research Psychologische Forschung, 63*(2), 163–173. https://doi.org/10.1007/PL00008175

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

Lian, A., Karlse, P. J., & Eriksen, T. B. (2004). Opposing effects of phonological similarity on item and order memory of words and nonwords in the serial recall task. *Memory, 12*(3), 314–337. https://doi.org/10.1080/09658210344000042 6

Logie, R. H., Saito, S., Morita, A., Varma, S., & Norris, D. (2016). Recalling visual serial order for verbal sequences. *Memory & Cognition, 44*(4), 590–607. https://doi.org/10.3758/s13421-015-0580-9

Majerus, S. (2019). Verbal working memory and the phonological buffer : The question of serial order. *Cortex, 112*, 122–133. https://doi.org/10.1016/j.cortex.2018.04.016

Majerus, S., der Linden, M. V., Mulder, L., Meulemans, T., & Peters, F. (2004). Verbal short-term memory reflects the sublexical organization of the phonological language network : Evidence from an incidental phonotactic learning paradigm. *Journal of Memory and Language, 51*(2), 297–306. https://doi.org/10.1016/j.jml.2004.05.002

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

Neale, K., & Tehan, G. (2007). Age and redintegration in immediate memory and their relationship to task difficulty. *Memory & Cognition, 35*(8), 1940–1953. https://doi.org/10.3758/BF03192927

Nee, D. E., & Jonides, J. (2013). Neural evidence for a 3-state model of visual short-term memory. *NeuroImage, 74*, 1–11. https://doi.org/10.1016/j.neuroimage.2013.02.019

Nimmo, L. M., & Roodenrys, S. (2004). Investigating the phonological similarity effect : Syllable structure and the position of common phonemes. *Journal of Memory and Language, 50*(3), 245–258. https://doi.org/10.1016/j.jml.2003.11.001

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

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., 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

Oberauer, K., & Lin, H.-Y. (2017). An interference model of visual working memory. *Psychological Review, 124*(1), 21–59. https://doi.org/10.1037/rev0000044

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 Section A, 48*(2), 384–404. https://doi.org/10.1080/14640749508401396

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116*(1), 129–156. https://doi.org/10.1037/a0014420

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

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

Saint-Aubin, J., Guérard, K., Chamberland, C., & Malenfant, A. (2014). Delineating the contribution of long-term associations to immediate recall. *Memory, 22*(4), 360–373. https://doi.org/10.1080/09658211.2013.794242

Saint-Aubin, J., & Poirier, M. (1999). Semantic Similarity and Immediate Serial Recall : Is There a Detrimental Effect on Order Information? *The Quarterly Journal of Experimental Psychology Section A, 52*(2), 367–394. https://doi.org/10.1080/713755814

Saito, S., Logie, R. H., Morita, A., & Law, A. (2008). Visual and phonological similarity effects in verbal immediate serial recall : A test with kanji materials. *Journal of Memory and Language, 59*(1), 1–17. https://doi.org/10.1016/j.jml.2008.01.004

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

Visscher, K. M., Kaplan, E., Kahana, M. J., & Sekuler, R. (2007). Auditory Short-Term Memory Behaves Like Visual Short-Term Memory. *PLoS Biology, 5*(3), Article e56. https://doi.org/10.1371/journal.pbio.0050056

Ward, G., Tan, L., & Grenfell-Essam, R. (2010). Examining the relationship between free recall and immediate serial recall : The effects of list length and output order. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(5), 1207–1241. https://doi.org/10.1037/a0020122

Williamson, V. J., Baddeley, A. D., & Hitch, G. J. (2010). Musicians' and nonmusicians' short-term memory for verbal and musical sequences : Comparing phonological similarity and pitch proximity. *Memory & Cognition, 38*(2), 163–175. https://doi.org/10.3758/MC.38.2.163