Set Size and Repetition Effects in Cued-Recall from Long-Term Memory

Elif Gezen (elif.gezen@psy.lmu.de)[1], Susanne Haridi[1], Mirko Thalmann[2] & Eric Schulz[2]
[1]Department of Psychology, Ludwig Maximilian University of Munich, Germany
[2]Institute for Human-Centered AI, Helmholtz Munich, Germany

Abstract

The set size effect refers to the decline in memory performance as the number of to-be-remembered items increases. However, memory retrieval in everyday life is often fast and accurate, suggesting the existence of factors that support efficient retrieval. One such factor may be repetition of items in memory. Theoretical views differ in how repetitions affect memory representations. Single-stronger image views suggest that each repetition strengthens an existing memory trace without increasing effective set size. Multiple-image views suggest that each repetition creates separate memory traces, inflating effective set size. In this study, we used a similarity-based sequential sampling model to generate hypotheses distinguishing these two views and tested them in a cued-recall paradigm. Results showed a robust set size effect on recall accuracy and a main effect of repetition on response times and accuracy, but no interaction. These results suggest that repetitions strengthen existing memory traces, without alleviating set size effects.

Keywords: set size effects; cued-recall; long-term memory; repetition; search

Introduction

Human memory has been likened to many things, such as a wax tablet, a gramophone, or a box (Roediger, 1980). These metaphors commonly refer to a storage of information that can be retrieved upon demand. If memory indeed functions as such, then retrieving any given item should get harder as the number of items increases. This phenomenon is known as the set size effect, and is typically reflected in behavioral performance by decreased accuracy and increased latency with larger memory sets (e.g., Gillund & Shiffrin, 1981; Haridi et al., 2025; Ratcliff & Murdock, 1976; Ratcliff et al., 2005; Tulving & Pearlstone, 1966). This effect has been demonstrated across major long-term memory tasks, namely, recognition (e.g., Ratcliff & Murdock, 1976; Strong, 1912), free-recall (e.g., Gillund & Shiffrin, 1981; Grenfell-Essam et al., 2013), and cued-recall (e.g., Haridi et al., 2025; Tulving & Pearlstone, 1966).

However, people efficiently retrieve information in daily life, suggesting human memory can overcome the set size effect. Many researchers have proposed that human memory owes its ability for rapid and accurate retrieval to its semantic organization (e.g., Kumar, 2021; Shiffrin & Atkinson, 1969). Recently, Haridi et al. (2025) manipulated item similarity and context in a series of cued-recall experiments and showed that an increased number of semantically or temporally related items slows down retrieval, even when the overall number of items remains the same, revealing that the set size effect is not only dependent on the overall number of items, but also on the number of semantically or temporally related items. Moreover, they proposed the Similarity-Based Sequential Sampling Model (SimSS), which successfully accounted for their results and provided a framework for investigating set size effects on cued-recall from long-term memory.

Notably, human memory includes many items that neither share a context nor are similar, which can also be efficiently remembered. Here, we propose that a memory could be easier to retrieve depending on how strongly a cue is linked to that specific memory, termed association strength. One way to manipulate association strength is to repeat items (Jang & Lee, 2019; Verde, 2009; Wilson & Criss, 2017).

Repetition Effects in Long-Term Memory

Theoretical accounts differ in explaining the effects of repetition on memory. One view suggests that repetitions help retrieval by strengthening existing individual item representations (Hintzman, 1976; Ratcliff et al., 1990; Shiffrin et al., 1990; Verde, 2009; Wilson & Criss, 2017). Under this view, each time an item is encountered, its underlying memory trace becomes more robust. These stronger item representations would have a higher probability of being retrieved compared to other items in the memory set. By making a specific repeated item stand out, repetition might therefore serve as a mechanism by which human memory systematically overcomes the set size effect.

Another view suggests that every repetition of a memory item is stored as a separate memory trace (Hintzman, 1976; Jang & Lee, 2019). Under this view, encountering an item multiple times does not strengthen a single representation, but rather accumulates distinct representations of that same item in the overall memory set.

There is support in the literature for both views: repetitions either strengthen existing items or add new items to the list (Gillund & Shiffrin, 2006; Jang & Lee, 2019; Raaijmakers & Shiffrin, 1981; Ratcliff et al., 1990; Shiffrin & Steyvers, 1997; Shiffrin et al., 1990; Verde, 2009; Wilson & Criss, 2017). If repetitions add new items to the list, this too might benefit the recall of the currently repeated item, because having multiple traces increases the chances of retrieving it. However, a systematic benefit for overcoming the set size effect would not be predicted, since here each repetition actively inflates the

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

effective set size. Therefore, if repetitions strengthen items, they may alleviate the set size effect; but if they add new items, they would increase the effective set size and fail to overcome the set size effect.

The Present Study

Previous research on the set size effect on long-term memory suggests that the semantic organization of memory may enable rapid and accurate cued-recall. The SimSS model by Haridi et al. (2025) accounts for these findings, supporting a serial, sequential search process in which memories are retrieved sequentially based on their semantic similarity and shared context with the presented cue. However, to our knowledge, the direct effect of item repetitions has not been investigated within this framework. In the present study, we aimed to explore whether repetitions can alleviate the set size effect in cued-recall, depending on whether a repetition strengthens a given item (by increasing the association strength between the cue and the target) or adds it to the list as a separate memory trace. This was directly possible through varying the association-strength parameter in the SimSS model. Accordingly, we ran simulations of the model to generate hypotheses. To test these hypotheses, we ran an online behavioral experiment employing a cued-recall paradigm.

Model Predictions & Hypotheses

The SimSS model, an implementation of the (ARC)-REM model for cued-recall (Diller et al., 2001), was chosen for the present study because it was designed specifically as a cued-recall model that accounts for the set size effect. Importantly, it directly implements item-item association strength, thereby allowing us to model repetition effects. It is based on the assumptions that a) memory retrieval arises from a sequential search through memory representations, b) this search is guided by semantic similarity, and c) the probability of correct retrieval scales with set size. Based on the values of semantic similarity (sem) and association strength (a), the total similarity (s) is calculated as:

$$s(k,c) = \begin{cases} \text{sem}(k,c)^{\beta} + a & \text{if } k = t_c, \\ \text{sem}(k,c)^{\beta} & \text{otherwise,} \end{cases} \tag{1}$$

where k stands for the current item, c stands for cue, t_c stands for the target, and β controls the weight of semantic similarity (set to 1 for this study). The probability of sampling any given item (k), then, is calculated as:

$$p(k) = \frac{s(k,c)}{\sum_{j=1}^{N} s(j,c)} \tag{2}$$

where N stands for the total number of items in memory. Sampling continues until an item is accepted as correct, or the search is abandoned. RT is determined by the number of samples that have been taken until sampling stops. For details of the SimSS model, see (Haridi et al., 2025).

Based on the two theoretical views regarding repetition effects, we implemented the SimSS in two distinct ways. To model the strengthening of a single trace, each repetition was treated as causing an increase in the association strength (a) between the cue and the target item, as a function of the number of repetitions. We hereby refer to this implementation as the single-stronger image (Shiffrin et al., 1990) implementation (Figure 1, panel A). Simulations using this implementation of the model predicted a main effect of set size and also of repetitions. They also predicted an interaction between repetitions and set size, such that as repetitions increased, the effect of set size on RTs and accuracy was alleviated.

To model multiple traces, each repetition was treated as a new item added to the list with association strength (a) (without changing the association strength of that item to the cue), increasing the set size (N) by the number of repetitions. We hereby refer to this implementation as the multiple-image (Shiffrin et al., 1990) implementation (Figure 1, panel B). Simulations run using this implementation of the model predicted, as expected, a main effect of set size. However, they did not predict a main effect of repetition or an interaction between repetitions and set size.

We deliberately implemented the models such that all items in the list are repeated, rather than only target items. We observed that this approach eliminates the prediction of the multiple-image view that repetitions might benefit the recall of the repeated item. This created a divergence between the two views where the single-stronger image view predicts a benefit of repetitions on recall, whereas the multiple-image view does not, allowing us to empirically separate the two accounts.

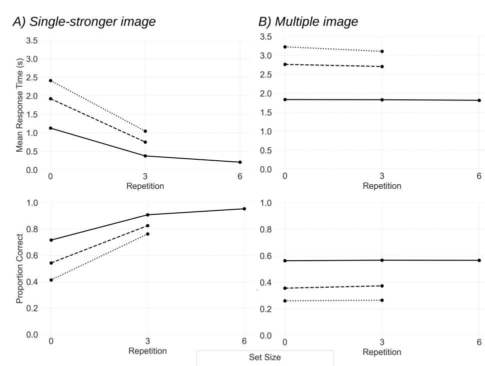

Figure 1: SimSS model predictions. **A)** shows model times and proportion of correct responses across set sizes for the single-stronger image implementation of the model. **B)** shows model times and proportion of correct responses across set sizes for the multiple-image implementation of the model.

In summary, we used the SimSS model to generate predictions for the single-stronger-image and multiple-image views, which are clearly distinct. In particular, the single-stronger image view predicts:

1. A main effect of set size will be observed, where as set sizes get larger, response accuracies decrease and RTs increase

(also predicted by the multiple-image implementation of the SimSS model).

2. A main effect of repetition will be observed, where as repetitions increase, response accuracies will increase and RTs will decrease (not predicted by the multiple-image implementation of the SimSS model).
3. An interaction effect between set size and repetition will be observed, where as repetitions increase, the effect of set size on RTs and accuracies will be alleviated (not predicted by the multiple-image implementation of the SimSS model).

Methods

Participants

We recruited 303 consenting adults aged 18 to 45 years (*M* = 34.49, *SD* = 6.77; 148 female, 3 non-binary/other) via the online research platform Prolific (https://prolific.com). All participants were native English speakers located in the UK or the USA with no reported memory disorders.

Following preregistered criteria (https://osf.io/9p83k/), we excluded participants who answered fewer than half the learning trials in a given block (*N* = 2; see Experimental Design), removed single trials with RTs exceeding 20 seconds (*N* = 36), and excluded participants who had fewer than 2 trials remaining in a given recall block (*N* = 1), leaving data from 300 participants for analyses.

Participants were compensated with a base fee of £6.50, plus a performance-based bonus (*M* = 1.17, *SD* = 0.25, max = £1.50). The study was approved by the Ethical Board of the Department of Psychology and Educational Sciences, Ludwig Maximilian University of Munich.

Experimental Design

The experimental task was programmed using jsPsych Version 8.2.1 (Leeuw, 2015). Participants could run the experiment only on a computer, using an up-to-date version of Chrome (Google LLC, 2024) or Firefox (Mozilla Foundation, 2024), and the experiment appeared in full-screen mode.

We used a 3 (SET SIZE: between-subjects) × 3 (REPETITION: within-subjects) mixed-design. Set sizes could be 5, 10, and 15; and repetitions could be 0, 3, and 6. Each participant was randomly assigned to one of the three set sizes. To maintain feasible task durations, participants assigned to set size 5 (*N* = 100) completed two randomly chosen repetition conditions, while those in set sizes 10 (*N* = 101) and 15 (*N* = 99) completed only the 0 and 3 repetition conditions. Thus, each participant completed two blocks of the same set size but different repetitions (see Figure 2). To confirm that this imbalance between set size - repetition combinations did not bias our results, we ran the main analyses (See Results) once again without the 6-repetition condition, post-hoc.

Procedure & Stimuli

At the start of the experiment, participants received instructions and a practice block to familiarize themselves with the task. They were instructed not to use any memory aids throughout the experiment.

At the start of each block, participants performed a filler task, included to control for lag, with its duration adjusted so that all participants reached the recall task after the same amount of time regardless of the length of the learning phase. In each trial of the filler task, participants were presented with two illustrations of cartoon aliens and asked to choose "Which alien is more friendly?" (5 s) by clicking the corresponding button. The stimuli were taken from Adobe Stock (Adobe Stock, 2023a, 2023b, 2024).

Next, the participants started a learning task where in each trial, they were shown one word pair (4 s), followed by a fixation cross (1 s). Then, they were shown one word from the presented pair and asked to press the spacebar to type the other (5 s). Upon answering, they received feedback (2 s). Then, the same process was repeated with the words in the pair reversed. Presentation order within each pair was randomized. To avoid rejecting correct answers due to typos or time constraints, correctly typing the first three letters of the target word was considered a correct response. For the filler and learning tasks, each trial had a fixed duration, to match block durations between and within subjects, allowing us to control for lag by varying the number of trials included in the filler task.

The words presented in the learning (and later, recall) task were nouns taken from the THINGS dataset (Hebart et al., 2019). We imported all words, and filtered them to include words made up of four to six letters, to control for the confounding effect of word length on recall performance (Katkov et al., 2014). Later, we used pre-trained Word2Vec embeddings (Mikolov et al., 2013) from the `en_core_web_md` model provided by the spaCy Python package to calculate word-to-word cosine similarities, following Haridi et al. (2025). Since prior work has shown that similarity facilitates recall (Epstein et al., 1975; Haridi et al., 2025), we excluded similar words (threshold = 0.5), and also removed outlier words, yielding a final pool of 220 words. For each participant, words from this pool were randomly paired. We randomly sampled these pairs to construct three word pair lists for set size 5, and two lists for set sizes 10 and 15. Each list was repeated the appropriate number of times: in every repetition, all word pairs were presented once, in a newly randomized order, before the next repetition cycle began. This yielded three distinct final learning sequences for set size 5 (corresponding to 0, 3, and 6 repetitions) and two for set sizes 10 and 15 (corresponding to 0 and 3 repetitions). Participants assigned to set size 5 were randomly administered two of their three generated sequences, while those in set sizes 10 or 15 received both.

After the learning task was a distractor task, adapted from Witte et al. (2025). This letter span task was included to ensure the upcoming recall task would not engage working memory processes. Participants were presented with alternating single letters (1 s) and simple math equations (maximum

6 s, response ended trial). They were required to remember the letters in the order of presentation, and to indicate whether the equations were true or false by pressing the corresponding button. Afterwards, they saw a grid of all possible letters and were asked to click on the letters in presentation order. The task consisted of four different block sizes (4, 5, 6, or 7 trials), presented in random order. After each block size, participants received feedback. The single letters were 18 consonants of the English alphabet. The simple mathematical equations were either subtraction or multiplication operations involving two randomly generated numbers between 0 and 10.

Lastly, participants had to complete the recall task. To overrule possible recency effects across different set sizes, we employed a proactive design where participants were only tested on the last five pairs of the preceding learning block. In each trial, they were first presented with a fixation cross (1 s), followed by a cue word from one of these last five pairs, and were asked to type the other word in the pair. These recall trials did not have a predetermined duration, as our main dependent variable was RT. To ensure that we would get a clean RT, participants had to press space before they could type, and correctly typing the first three letters (in maximum 3 seconds) was considered a correct answer.

The main part of the experiment consisted of two consecutive blocks, each comprising all four tasks (see Figure 2 for a schematic representation). Participants could take a break between blocks (max. 1 min). After the main experiment, participants were asked to indicate their age and gender and answer honesty-check questions to reveal whether they used any memory aids. Finally, we asked for any strategies they might have used and general comments. The whole experiment took approximately 45 minutes.

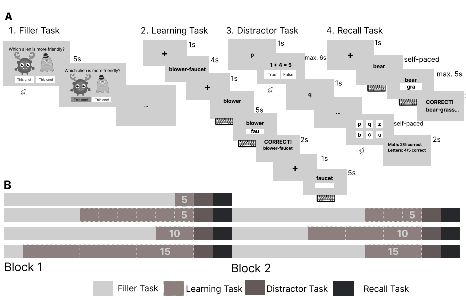

Figure 2: Schematic representation of trial and block structures. **A)** shows the trial structures for each task. **B)** shows the overall experimental structure for each different set size. The numbers indicate the set size for the shown block. The dashed lines differentiate repetitions. The shades differentiate each task within the block. Block order was randomly determined for each participant.

Results

Following SimSS model simulations, first, we expected a set size effect, as predicted by both the single-stronger image and multiple-image interpretations of the SimSS model. Second, to distinguish between the two theoretical views regarding repetitions, we hypothesized that if repetitions strengthen a single image, there would be a main effect of repetition. Third, under the same assumption, we expected an interaction between repetition and set size on RTs and response accuracies. We estimated these effects from the empirical RTs using a mixed-effects model with the following model structure (preregistered):

$$\text{RT} \sim \text{set size} \times \text{repetition} + (1 \mid \text{participant})$$

The preregistered random slope for repetition failed to converge, and therefore was removed. For RT analyses, we only included correct responses. For accuracy analyses, we used the same model structure with the corresponding dependent variable.

See Figure 3 for a visualization of the results. We fitted Bayesian hierarchical linear regressions using the brms package (Burkner, 2017, 2018) in R (R Core Team, 2024). Given the right-skewed distribution of RTs that can be efficiently accounted for by lognormal likelihood functions (Ulrich & Miller, 1993; van der Linden, 2006), we used this function for the RT models, and a Bernoulli likelihood function for the accuracy models. Both models were fitted with four MCMC chains, 5000 warm-up and 5000 stored iterations.

The initial RT model did not show an effect of set size on RT (β = 0.01, 95% CI [-0.00, 0.03]). There was a marginal effect of repetition with the credible interval barely excluding zero (β = -0.03, 95% CI [-0.06, 0.00]), suggesting faster responses with increased repetitions. We found no interaction between set size and repetition (β = -0.00, 95% CI [-0.01, 0.00]).

There was a decrease in response accuracies with increasing set size (β = -0.13, 95% CI [-0.19, -0.06]), indicating the presence of a set size effect, as hypothesized. Moreover, there was an effect of repetition (β = 0.31, 95% CI [0.17, 0.47]), supporting our hypothesis that repetitions benefit recall performance. Consistent with RT results, we found no interaction between set size and repetitions (β = 0.01, 95% CI [-0.00, 0.03]).

The RT results were unexpected given prior literature (e.g., Haridi et al., 2025). Specifically, we found that while performance declined from set size 5 to 10, participants in set size 15 unexpectedly outperformed those in set size 10. Such unexpected patterns can arise in between-subjects designs given substantial individual differences (*SD* = 0.39, 95% CI [0.35, 0.43]) (Efird, 2011; Gorvine, 2018), possibly masking a set size effect on RTs (see Discussion for details). Therefore, we re-estimated both models with set size entered as a categorical predictor to examine the set size effect more directly through pairwise comparisons. For the RT model, we increased the number of iterations to 7500 warm-up and 7500 stored iterations to aid convergence.

We found an RT increase only for set size 10 (β = 0.22, 95% CI [0.08, 0.36]), but not for set size 15 (β = 0.09, 95% CI [−0.06, 0.23]), both compared to set size 5, supporting our observation that participants in set size 15 performed unexpect-

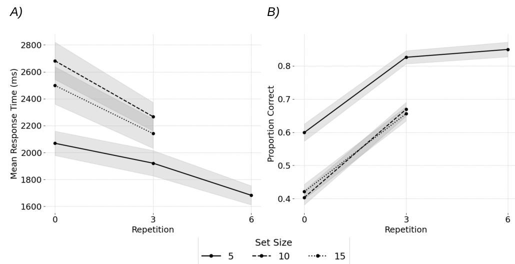

Figure 3: Results. **A)** shows results on RTs, and **B)** shows results on response accuracies. In both panels, lines represent empirical means, and shades represent the standard error of the mean.

edly well. As expected, there was a main effect of repetition, with RTs significantly decreasing from no repetitions to 3 (β = −0.14, 95% CI [−0.23, −0.05]) and 6 (β = −0.20, 95% CI [−0.30, −0.09]). Importantly, no interaction between set size and repetitions was observed (all 95% confidence intervals included zero).

There was a significant decrease in response accuracies for both set size 10 (β = −1.15, 95% CI [−1.77, −0.53]) and set size 15 (β = −1.05, 95% CI [−1.67, −0.42]), both compared to set size 5. These results align with the initial model results, revealing a set size effect as hypothesized. Similarly, there was a significant main effect of repetition on response accuracies, where accuracies increased with 3 (β = 1.52, 95% CI [1.14, 1.91]) and 6 (β = 1.82, 95% CI [1.28, 2.36]) repetitions, compared to the no-repetition condition. Similar to the RT results, we found no interaction between set size and repetitions (all 95% confidence intervals included zero, see Table 1 for full statistics).

Overall, repetition reliably improved recall performance, but this effect did not overcome any decreases caused by larger set sizes. Although the main effects mostly align with the predictions of the single-stronger image view, the lack of interaction aligns with the predictions of the multiple-image view.

Discussion

In the present study, we tested whether repetitions strengthen a memory representation or add separate traces to the memory set within a sequential sampling framework. Specifically, we investigated whether repeating all items in a list alleviates the set size effect: a benefit predicted only by the single-stronger image view, but not the multiple-image view.

Our results demonstrated a clear set size effect on response accuracies across all three set sizes, as expected. However, the set size effect on RTs was present only when comparing set size 5 versus 10, but not set size 5 versus 15. Although these findings indicate that memory performance generally declines as the number of distinct items increases, the unexpected absence of a set size effect on RTs at the largest set size warrants consideration. If this absence reflects a true underlying effect, these results would not align with the SimSS model framework, upon which this study is based. Instead, a plausible explanation for not observing a set size effect on RTs at the largest set size is the substantial baseline RT variability (mean difference 0.13, *SD* = 0.38, 95% CI [0.34, 0.42]), especially considering that we manipulated set size between-subjects. In between-subjects designs, random assignment can produce such baseline variability across conditions. Participants in the set size 15 condition may have happened to have faster baseline RTs, masking the expected set size effect. It was argued that in such between-subjects designs, detecting the effect of interest is harder, particularly when individual differences are large (Efird, 2011; Gorvine, 2018). We adopted a between-subjects design despite this limitation to limit task duration and therefore the risk of fatigue and loss of attention.

Supporting our hypothesis regarding the main effect of repetition, participants produced faster and more accurate responses with increasing repetitions, suggesting that repeated exposure during encoding enhances subsequent memory performance. This pattern was predicted by the single-stronger image implementation of the SimSS model, which also offers a mechanism behind it. Recall that the SimSS model includes the parameter association strength (a), which, in the single-stronger image implementation, increased as a function of the number of repetitions. Accordingly, it can be argued that the observed repetition effects may reflect strengthened associations between the words within a pair, resulting from repeated exposure.

Contrary to our hypothesis, repetitions did not alleviate the set size effect: performance declined similarly across repetitions. This finding diverges from predictions of the single-stronger image implementation of the SimSS model. One possible explanation for the set size effect not being alleviated despite a main effect of repetition lies in our experimental design, in which all word pairs were repeated during encoding rather than only target items. As mentioned before, this was a deliberate choice to empirically differentiate the two theoretical views. However, according to the organizational structure of memory as explained and modeled via the SimSS model, recall depends not only on associations within a word pair but also on associations between each pair and the whole search set (i.e., the temporal context; Haridi et al., 2025). Following this, in our experiment, every pair being strengthened may have rendered each item equally salient rather than rendering a single repeated pair more salient within the study set. This would prevent relative advantages during recall. This interpretation resonates with findings from Jang and Lee (2019), who showed that when all items were repeated, repetition improved memory for individual items, while RTs and accuracies nonetheless dropped with increasing set sizes. They argued that repeating all items does not increase a given item's strength relative to the other items in the same set (i.e., its competitors). This interpretation is consistent with the well-known relative-strength competition principle, which holds that the presence of strong competitors decreases the proba-

Table 1: Model Results. β = posterior mean estimate; CI = credible interval. Reference categories: Set size = 5; Repetition = none (0).

	RT		Accuracy	
	β	95% CI	β	95% CI
Fixed Effects				
Intercept	7.51	[7.42, 7.61]	0.64	[0.17, 1.09]
Set Size (10 × 5)	0.22	[0.08, 0.36]	-1.15	[-1.77, -0.53]
Set Size (15 × 5)	0.09	[-0.06, 0.23]	-1.05	[-1.67, -0.42]
Repetition (3 × 0)	-0.14	[-0.23, -0.05]	1.52	[1.14, 1.91]
Repetition (6 × 0)	-0.20	[-0.30, -0.09]	1.82	[1.28, 2.36]
Set Size 10 × Rep. 3	-0.10	[-0.22, 0.02]	0.08	[-0.36, 0.56]
Set Size 15 × Rep. 3	-0.01	[-0.13, 0.12]	-0.07	[-0.55, 0.38]
Set Size 10 × Rep. 6	0.00	[-0.59, 0.59]	-0.01	[-3.15, 3.07]
Set Size 15 × Rep. 6	0.00	[-0.58, 0.58]	0.02	[-3.07, 3.18]
Random Effects				
SD(Subject Intercept)	0.38	[0.34, 0.42]	1.76	[1.55, 2.00]

bility that the target item will be correctly recalled (Mensink & Raaijmakers, 1988; Raaijmakers & Shiffrin, 1981; Verde, 2009).

A limitation of the present study was that we did not explicitly model encoding processes, since the SimSS model was designed as a model of retrieval. Although the learning phase of the experiment was structurally identical across conditions, its length varied with the number of repetitions, even within the same set size. To overrule possible recency effects (which could confound experiments with different set sizes), we only tested the last five word pairs of each learning block, and the duration between the learning and test phases was held constant across all between- and within-subjects conditions. Nevertheless, longer learning phases might have caused encoding interference, as increases in set size could limit the attentional resources allocated to each item, resulting in inefficient encoding (Haridi et al., 2025; Pajkossy & Racsmány, 2019). Although we tried to mitigate these effects through practice trials and an active learning phase, we cannot fully rule out their possible influence on memory performance.

Although repetitions did not alleviate the set size effect as predicted by SimSS model simulations, our results favor aspects of the single-stronger image view over a pure multiple-image account: repetitions improved, rather than impaired, recall performance. At the same time, the absence of an interaction between set size and repetition indicates that the current single-stronger image implementation of the SimSS model is not sufficient to fully account for the data. Explaining the full pattern of results may therefore require extensions to the SimSS model, such as incorporating relative-strength competition dynamics, alternative theoretical approaches, or quantitative model fitting (which was beyond the scope of the present study).

Our results highlight substantial individual differences in memory performance, consistent with prior findings from similar experiments (Haridi et al., 2025). Future work could therefore consider adapting the SimSS model or employing other models that explicitly account for such variability. Further research may also consider theoretical issues related to repetitions. For instance, it is an established finding that spaced, rather than massed, repetitions benefit memory (Hintzman, 1976; Jang & Lee, 2019; Shiffrin et al., 1990). Although we employed spaced repetitions in the present design, where there was at least one different word pair between the repetitions, it remains an interesting research question whether the temporal spacing of repetitions differentially affects memory performance.

Finally, we would like to emphasize that these conclusions are conditional on the assumptions embedded in the SimSS model, or related frameworks such as the SAM (Raaijmakers & Shiffrin, 1981) and REM (Shiffrin & Steyvers, 1997) models. Future research should assess their robustness across other modelling assumptions.

Conclusion

Using cognitive modeling and an online experiment, we show that repeating word pairs improves recall of those pairs. However, this increase cannot alleviate the speed and accuracy limitations imposed by larger set sizes. Our results help differentiate between two theoretical views of the effect of repetitions on recall performance under a sequential sampling account of memory retrieval, showing that repetitions most likely strengthen memory representations rather than accumulate a separate memory trace. Moreover, we show that the SimSS model accounts for most of our results, suggesting a meaningful role for association strength in retrieval processes. However, the current implementation cannot account for the lack of interaction, and further work should explore potential alternatives to address this pattern of results.

Acknowledgements

During the preparation of this work, the authors used ChatGPT (OpenAI), Claude (Anthropic), and Gemini (Google) for improving code, figures, and readability of certain paragraphs; and Elicit AI for efficient literature search. After using these tools, the authors reviewed and edited the content. They take full responsibility for the content and guarantee that this paper does not contain part of purely AI-generated text.

References

Adobe Stock. (2023a). Monster icons set cartoon vector animal troll cute mascot [Image ID: 553288321. Retrieved from https://stock.adobe.com/de/images/monster-icons-set-cartoon-vector-animal-troll-cute-mascot/553288321].

Adobe Stock. (2023b). Set of cute colorful monsters cartoon baby monster heads nursery design vector tidy flat illustration [Image ID: 581455165. Retrieved from https://stock.adobe.com/de/images/set-of-cute-colorful-monsters-cartoon-baby-monster-heads-nursery-design-vector-tidy-flat-illustration/581455165].

Adobe Stock. (2024). Cartoon cute funny monster characters cute bacteria monster stump and devil isolated vector personages [Image ID: 641013522. Retrieved from https://stock.adobe.com/de/images/cartoon-cute-funny-monster-characters-cute-bacteria-monster-stump-and-devil-isolated-vector-personages/641013522].

Burkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *J. Stat. Softw.*, *80*(1).

Burkner, P.-C. (2018). Advanced Bayesian multilevel modeling with the R package brms. *The R Journal*, *10*(1), 395–411. https://doi.org/10.32614/RJ-2018-017

Diller, D. E., Nobel, P. A., & Shiffrin, R. M. (2001). An ARC-REM model for accuracy and response time in recognition and recall. *J. Exp. Psychol. Learn. Mem. Cogn.*, *27*(2), 414–435.

Efird, J. (2011). Blocked randomization with randomly selected block sizes. *Int. J. Environ. Res. Public Health*, *8*(1), 15–20.

Epstein, M. L., Phillips, W. D., & Johnson, S. J. (1975). Recall of related and unrelated word pairs as a function of processing level. *J. Exp. Psychol. Hum. Learn.*, *1*(2), 149–152.

Gillund, G., & Shiffrin, R. M. (1981). Free recall of complex pictures and abstracts words. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 575–592. https://doi.org/10.1016/S0022-5371(81)90192-4

Gillund, G., & Shiffrin, R. M. (2006). A retrieval model for both recognition and recall. *Psychological Review*, *91*(1), 1–67. https://doi.org/10.1037/0033-295x.91.1.1

Google LLC. (2024). Google Chrome [computer software] [Version 125.0 or later].

Gorvine, B. (2018). *Research methods: From theory to practice*. Oxford University Press. https://books.google.de/books?id=Ara6nQAACAAJ

Grenfell-Essam, R., Ward, G., & Tan, L. (2013). The role of rehearsal on the output order of immediate free recall of short and long lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(2), 317.

Haridi, S., Schulz, E., & Thalmann, M. (2025). Context size and set size effects: The relevance of specific cues when searching long-term memory. *Computational Brain & Behavior*, 1–33.

Hebart, M. N., Dickter, A. H., Kidder, A., Kwok, W. Y., Corriveau, A., Wicklin, C. V., & Baker, C. I. (2019). THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PloS One*, *14*(10), e0223792.

Hintzman, D. L. (1976). Repetition and memory. In G. H. Bower (Ed.). Academic Press. https://doi.org/10.1016/S0079-7421(08)60464-8

Jang, Y., & Lee, H. (2019). Item repetition and retrieval processes in cued recall: Analysis of recall-latency distributions. *Memory & Cognition*, *47*(4), 792–815. https://doi.org/10.3758/s13421-019-00902-y

Katkov, M., Romani, S., & Tsodyks, M. (2014). Word length effect in free recall of randomly assembled word lists. *Frontiers in computational neuroscience*, *8*, 129.

Kumar, A. A. (2021). Semantic memory: A review of methods, models, and current challenges. *Psychonomic Bulletin & Review*, *28*(1), 40–80.

Leeuw, J. R. D. (2015). jsPsych: A JavaScript ibrary for creating behavioral experiments in a Web browser. *Behavior Research Methods*, *47*(1), 1–12.

Mensink, G.-J., & Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychol. Rev.*, *95*(4), 434–455.

Mikolov, T., Chen, K., Corrado, G. S., & Dean, J. (2013). Efficient estimation of word representations in vector space. *International Conference on Learning Representations*. https://api.semanticscholar.org/CorpusID:5959482

Mozilla Foundation. (2024). Mozilla Firefox [computer software] [Version 127.0 or later].

Pajkossy, P., & Racsmány, M. (2019). How the size of the to-be-learned material influences the encoding and later retrieval of associative memories: A pupillometric assessment. *PLoS One*, *14*(12), e0226684.

R Core Team. (2024). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. https://www.R-project.org/

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. data and discussion. *J. Exp. Psychol. Learn. Mem. Cogn.*, *16*(2), 163–178.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (2005). List-strength effect: I. data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 163–178. https://doi.org/10.1037/0278-7393.16.2.163

Ratcliff, R., & Murdock, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review*, *83*(3), 190.

Roediger, H. L. (1980). Memory metaphors in cognitive psychology. *Memory & Cognition*, *8*(3), 231–246.

Shiffrin, R. M., & Atkinson, R. C. (1969). Storage and retrieval processes in long-term memory. *Psychological Review*, *76*(2), 179.

Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). List-strength effect: II. theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 179.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM–Retrieving Effectively from Memory. *Psychonomic Bulletin & Review*, *4*(2), 145–166.

Strong, E. K. (1912). The effect of length of series upon recognition memory. *Psychological Review*, *19*(6), 447–462. https://doi.org/10.1037/h0069812

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, *5*(4), 381–391. https://doi.org/10.1016/S0022-5371(66)80048-8

Ulrich, R., & Miller, J. (1993). Information processing models generating lognormally distributed reaction times. *Journal of Mathematical Psychology*, *37*(4), 513–525. https://doi.org/10.1006/jmps.1993.1032

van der Linden, W. J. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics*, *31*(2), 181–204. https://doi.org/10.3102/10769986031002181

Verde, M. F. (2009). The list-strength effect in recall: Relative-strength competition and retrieval inhibition may both contribute to forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(1), 205–220. https://doi.org/10.1037/a0014275

Wilson, J. H., & Criss, A. H. (2017). The list strength effect in cued recall. *Journal of Memory and Language*, *95*, 78–88. https://doi.org/10.1016/j.jml.2017.01.006

Witte, K., Thalmann, M., & Schulz, E. (2025). Model-based exploration is measurable across tasks but not linked to personality and psychiatric assessments. *Sci. Rep.*, *15*(1), 27479.