# Memory resources recover gradually over time: The effects of word-frequency, presentation rate and list-composition on binding errors and mnemonic precision in source memory

**Authors:** Vencislav Popov[1], Matthew So[1], Lynne M. Reder[1]

[1] Department of Psychology, Carnegie Mellon University

Corresponding author: Vencislav Popov, vencislav.popov@gmail.com

Authors' Note:

Vencislav Popov is now in the Department of Psychology, University of Zurich

*This manuscript has been accepted for publication in Journal of Experimental Psychology: Learning Memory and Cognition*

**Abstract**

Normative word frequency has played a key role in the study of human memory, but there is little agreement as to the mechanism responsible for its effects. To determine whether word frequency affects binding probability or memory precision, we used a continuous reproduction task to examine working memory for spatial positions of words. In three experiments, after studying a list of five words, participants had to report the spatial location of one of them on a circle. Across experiments we varied word frequency, presentation rate and the proportion of low frequency words on each trial. A mixture model dissociated memory precision, binding failure and guessing rate parameters from the continuous distribution of errors. On trials that contained only low- or only high-frequency words, low-frequency words lead to a greater degree of error in recalling the associated location. This was due to a higher word-location binding failure and not due to differences in memory precision or guessing rates. Slowing down the presentation rate eliminated the word frequency effect by reducing binding failures for low-frequency words. Mixing frequencies in a single trial hurt high-frequency and helped low-frequency words. These findings support the idea that word frequency can lead to both positive and negative mnemonic effects depending on a trade-off between a HF encoding advantage and a LF retrieval cue advantage. We suggest that 1) low-frequency words require more resources for binding, 2) that these resources recover gradually over time, and that 3) binding fails when these resources are insufficient.

Keywords: working memory, mixture modeling, continuous reproduction task, word frequency, contiguity effects

## Introduction

Normative word frequency has played a key role in empirical research and theoretical development on human memory (Clark, 1992; Glanzer & Bowles, 1976; Hulme et al., 2003; Mandler et al., 1982; Reder et al., 2000; Popov & Reder, 2020b). Its prominence is due to the fact that it can either facilitate or impair memory performance depending on the nature of the study task, the test, and the details of the study sequence (for a review, see Popov & Reder, 2020a). For example, high-frequency (HF) words are recognized less well in item recognition tests, leading to more false alarms and fewer hits than low-frequency (LF) words (e.g., Glanzer & Adams, 1990; Reder et al., 2000). Despite the LF recognition advantage, HF words lead to better memory in free recall (Deese, 1960), serial recall (Hulme et al., 1997) and associative recognition (Clark, 1992). Crucially, these effects typically appear in pure lists that contain only HF or only LF words, but disappear when both types of words are mixed in the same list (Ozubko & Joordens, 2007; Reder et al., 2007).

While the effects of word frequency on item recognition, associative recognition, free recall, and serial recall are well established, their effects on cued-recall or source recall are less clear and provide a challenge for memory models. In both cued-recall and source memory[1] tasks, HF cues sometimes lead to better performance than LF cues, sometimes to worse performance than LF cues, and sometimes to no difference at all. For example, in cued recall, some researchers have found that only the frequency of the *target* matters and there is either no effect of the cue word's frequency (Madan et al., 2010; Nelson & McEvoy, 2000) or only a small benefit of HF cues that disappears when controlling for other variables (Criss et al., 2011). In contrast, we have

[1] source memory tasks can be considered a version of cued recall tasks where the target is some contextual detail presented simultaneously with the cue during study and test

shown that cued recall performance is better if the cues are familiarized to a greater degree prior to the experiment (Reder et al., 2016). To make matters even more confusing, in source memory the findings are also mixed: recalling the source context is often better with LF cues (e.g., Diana & Reder, 2006; Glanzer et al., 2004; Osth et al., 2018; Reder et al., 2002); however, DeWitt et al., (2012) found an advantage for cues that participants rated as more familiar beforehand.

This inconsistent pattern of results is unsatisfactory because it makes it difficult to evaluate memory models. Numerous models have been proposed to explain word frequency effects in item recognition and free recall (for a review, see Popov & Reder, 2020a) and here we will focus on the Source of Activation Confusion (SAC) model (Popov & Reder, 2020b; Reder et al., 2000, 2007). SAC depends on the following assumptions to explain a wide array of word frequency effects: 1) memory formation depletes a limited resource that *recovers* gradually over time; 2) HF words have stronger preexisting representations in memory compared to LF words; 3) the amount of resources required to process an item is an inverse function of its existing strength; 4) as a result, processing LF words depletes more resources than does HF words, leaving fewer resources to create episodic associations between words and an experiential context; 5) cues associated with more episodic contexts are less effective in retrieving any specific episode; 6) LF words are associated with fewer contexts and, as such, spread more activation to episodes they are connected to. As a result of these assumptions, LF words are stored less well and are less likely to be bound to other study items or to an episodic context; however, once stored, they are more effective as cues during retrieval. These two factors create a trade-off (Popov & Reder, 2020a; Reder et al., 2007), and, depending

on the task conditions (e.g., presentation rate, list-composition, test type), can produce either a positive or a negative effect of frequency[2] (see **Figure 1**).

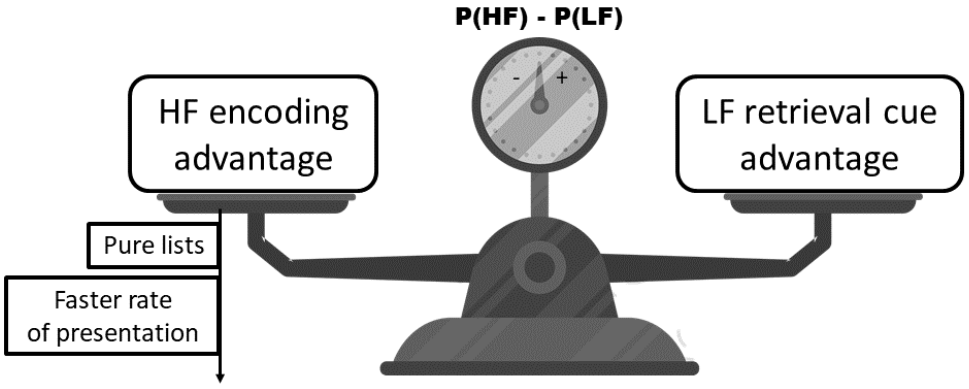


***Figure 1.*** *Illustration of the trade-off between an HF encoding advantage and an LF retrieval cue advantage. A memory task will produce either positive ($P_{HF}>P_{LF}$) or negative ($P_{HF}<P_{LF}$) effects of word frequency depending on whether the HF encoding advantage exceeds the LF retrieval cue advantage. Factors such as pure lists or faster presentation rate increase the HF encoding advantage and as such increase $P_{HF}$ (memory performance for HF words) relative to $P_{LF}$ (memory performance for LF words).*

Cued-recall and source-memory tasks present an ideal opportunity to illustrate and test these principles because they combine properties of both item recognition (using a word as a cue) and recall (generating a feature from memory). As our earlier discussion showed, this sometimes leads to an HF benefit, and sometimes to a LF benefit in these tasks. This puzzling pattern could potentially be explained by SAC. On the one hand, SAC predicts that HF words are more likely to be bound to other items and context; on the other hand, if a HF word is used to cue memory

[2] It is important to note that the model consistently produces either a positive or a negative effect under the same conditions – it is not that it can simply fit any pattern desired. See Popov and Reder (2020b) for simulations and model fits.

afterwards, it will spread less activation to the associated items and context, making retrieval more difficult. Therefore, whether we will observe a positive or a negative word-frequency effect in cued-recall and source-memory tasks depends on whether the storage benefit exceeds the retrieval deficit for HF cues (see **Figure 1**).

One way to test the model is to manipulate factors such as presentation rate and list-composition that are known to affect how strongly items are encoded. Presenting words at a faster rate increases the positive word frequency effect in free recall (Gregg et al., 1980) and decreases or even reverses the negative word frequency effect in item recognition (Criss & McClelland, 2006; Malmberg & Nelson, 2003). In addition, as we noted earlier, the typical effects of word frequency tend to appear only in homogenous frequency lists and disappear or reverse in mixed lists containing both high and low frequency words (Ozubko & Joordens, 2007). Word frequency effects also depend on the order of LF and HF items within the list, such that presenting LF words in the first part of the list and HF words in the second part of the list leads to worse performance than the reverse (Miller & Roodenrys, 2012).

SAC's explanation of these results draws on the first four principles outlined above – speeding up the presentation rate allows less time for resource recovery, which causes words to be stored less well. This hurts LF words more than it does HF words, since LF words require more resources. Thus, faster presentation rates increase the encoding strength difference between HF and LF words, which can in turn override the LF retrieval cue advantage. Conversely, when LF and HF words are mixed in a single list, the presence of LF words hurts the storage of HF words because LF words deplete more resources. On the other hand, in mixed lists LF words benefit from having more resources available compared to pure LF lists. According to SAC, we should observe similar effects in a cued-recall/source-memory task – whether HF or LF word cues would lead to

better performance would depend on the presentation rate and the list-composition. As a result, manipulating these factors would not only allow us to test the SAC model, but also to explain why the current literature shows an inconsistent pattern of results when it comes to the cue word`s frequency in cued-recall and source memory.

Our first goal, then, is to characterize in detail how the frequency of a cue word affects source memory, specifically, how word frequency interacts with presentation rate and list-composition. Our second goal is to shed more light on the resource depletion mechanism proposed by SAC. Despite SAC's success in modeling word-frequency effects, an unresolved question concerns what happens when there are insufficient resources to store a new item in memory (for example, if encoding the word "chair" in an experiment requires 0.7 resources, and there are only 0.3 resources available). SAC assumes that memory traces differ in strength, and that the probability of recall depends on how high the memory strength is relative to a retrieval threshold (Popov & Reder, 2020b). When an event is not recalled successfully, this could be either because the memory trace does not exist (i.e., the stimulus was not bound to the study context), or because the trace strength (or its precision[3]) did not pass the retrieval threshold. We envision two possible ways in which resource depletion affects memory trace formation. First, if there are insufficient resources, participants could fail to bind the item to the study context or to other items and no memory trace would be created. A second possibility is that a memory trace is still created, but its strength and precision are lower, proportional to the amount of resources remaining. These experiments aim to adjudicate between these two possibilities.

[3] We use the terms strength and precision interchangeably in the manuscript, as SAC assumes that the strength of a memory trace determines the precision of the response in a continuous reproduction paradigm like the one used in these experiments.

## All or None vs. Differential Strength

How would one determine whether a memory trace is formed or not versus whether a trace just varies in strength / precision? This question has been investigated extensively in studies on visual working memory. Working memory can store and maintain a limited number of items, and researchers have tried to identify what happens when this capacity is surpassed (Bays et al., 2009; Ma et al., 2014; Zhang & Luck, 2008). One popular paradigm used to distinguish the probability that an item exists in memory from the precision with which it is stored is the continuous reproduction task (for a review, see Ma et al., 2014). In the original version of this task, participants saw several colored squares presented simultaneously at different locations on the screen; at test, one location at random was highlighted and participants had to indicate on an analog (continuous) circular scale the color that was presented in the cued location (Wilken & Ma, 2004). Rather than testing memory for color when cued by location, we modified the task to present words around a circle and asked participants to indicate location around the circle when a word was cued in the center of the circle (see **Figure 2** for our adaptation of this task). Performance on this task is measured as the angular difference on a circular scale between the target feature and the response; in this case, the difference between the originally presented location and the reported location for the cued word. Typically, responses form a bell-curved distribution centered on the target response, and a mixture model is used to dissociated what sources contribute to the error distribution (see **Figure 3**).

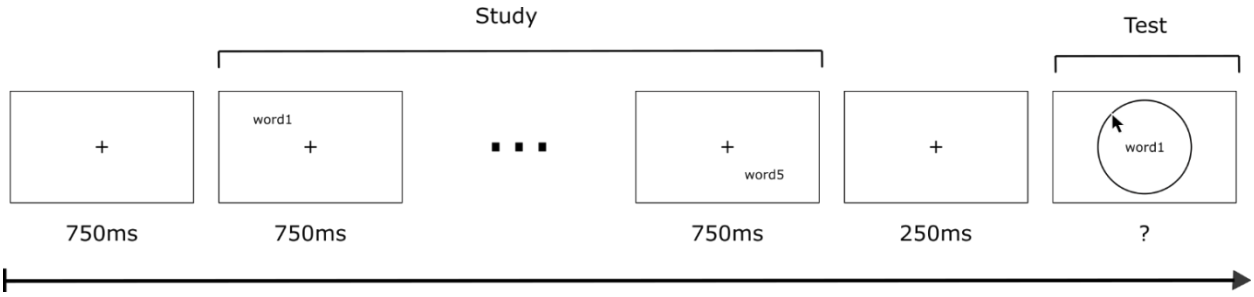


***Figure 2.*** *An illustration of a single trial in the current experiments.*

The key assumption behind this technique is that on any given trial, the size of the angle error could be the result of: 1) correct but noisy recall of the feature from the correct item; 2) a random guess due to the absence of a memory trace; 3) a noisy recall of a feature associated with a different study item, rather than the probed/target item (Bays et al., 2009; Ma et al., 2014). These three sources of error contribute to a mixture distribution with multiple components. Recall of the correct item's feature contributes to a normal distribution of errors centered on the target's feature; incorrect recall of a different item's feature contributes to a normal distribution of errors centered on the non-target item's feature; random guessing results in a uniform distribution of errors (see **Figure 3**, left). A mixture model can be used to estimate four parameters – the proportion of correct recalls, $p_{\text{correct}}$; the proportion of mis-bindings (i.e., recalling the feature of a non-target item), $p_{\text{mis-binding}}$; the proportion of random guesses, $p_{\text{guess}} = 1 - p_{\text{correct}} - p_{\text{mis-binding;}}$ and the precision of the memory trace, which is the standard deviation of the normal distribution components, σ – more precise memory traces have less noise, they lead to responses that are closer to the studied location, and consequently, to a smaller σ. This paradigm has proven useful in dissociating recall success, mis-binding and recall precision in visual working memory.

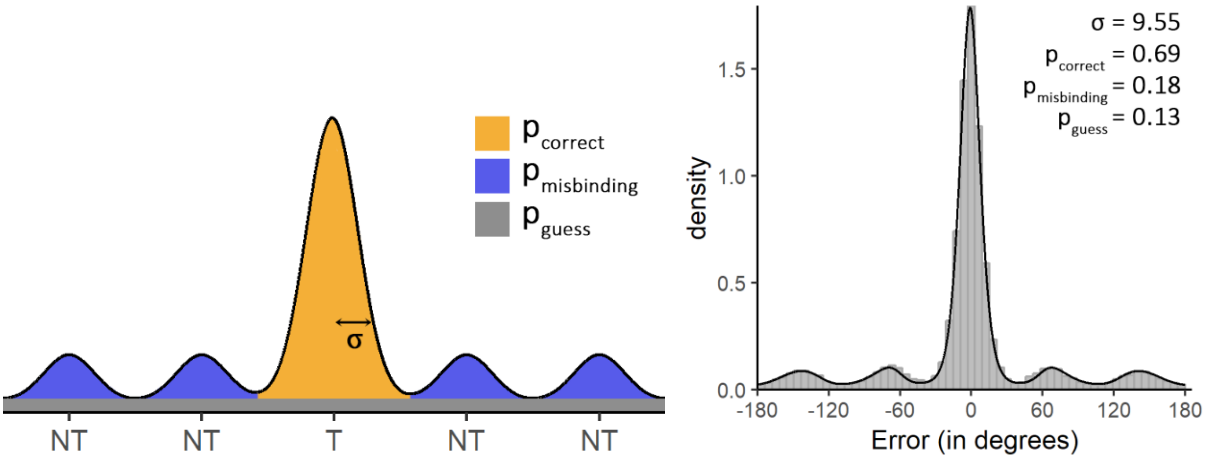


***Figure 3.*** *Left panel - illustration of how the mixture model components contribute to the error response distribution. T = target location, NT = non-target location. Right panel - distribution of location recall error (in degrees) from 31800 trials done by 106 participants. The black line is the distribution fit by the 3-parameter mixture model*

In order to determine whether resource depletion affects the probability of binding an item to its context or just affects the precision of that binding, we adapted this paradigm to the verbal domain. In three experiments, on each of 300 trials per experiment, participants studied five words presented sequentially in different locations, followed by a probe with one of the studied words. A participant had to indicate on a circle where the probed word appeared during study (**Figure 2**). The key question then is whether word frequency and presentation rate affect binding probability, guessing rates or the memory precision parameters. The SAC model claims that both high frequency words and slower presentation rates cause less resource depletion. Thus, both factors should affect the same parameter of the mixture model. SAC assumes that LF words require more resources to be encoded. That could leave fewer resources to bind the word to its location and to store that binding in memory. If there are insufficient resources to store the binding, we should see an increase in the mis-binding or guessing parameters, because participants would not have any

information available about the probe word's location. In contrast, if instead a weaker, less precise trace is established for LF words, word frequency should affect the precision (σ) with which the correct location is recalled. The same logic should apply for the presentation rate manipulation because faster presentation rates allow less time for resource recovery.

In Experiment 1, we varied whether each trial contained HF or LF words. In Experiment 2, we also varied how long each word was presented on the screen (500/750/1000 ms.). The first two experiments had trials of homogenous frequency – all five words on each trial were either HF or LF. In Experiment 3, we again manipulated presentation rate, but this time each trial contained both HF and LF and we varied whether the trial contained 40% or 60% LF words.

## Experiment 1: Main effect of word frequency

### Method

**Participants.** Twenty-four native English-speaking students from Carnegie Mellon University participated for either course credit or for $10 compensation. The sample sizes (the number of participants and the number of trials per participant) were determined through custom simulations of the mixture model. Since we wanted to fit the model to individual participants, our initial parameter recovery simulations showed that at least a 100 observations per condition are necessary to reliably estimate the p_correct, p_misbinding and sigma parameters. Fewer than a 100 observations per condition often lead to failure to recover the correct parameters. Thus, we decided on 300 total trials per participant, such that we could fit the model to the presentation speed manipulation and the frequency manipulation, separately. Given the large number of observations per participant, sample sizes of 10-30 are typical in the visual working memory literature (simulations have shown that number of subjects and number of trials per subject are interchangeable, as long as the between-subject variability does not greatly exceed within-subject

variability, Smith & Little, 2018). As a rule of thumb, we aimed for 20 participants in Experiment 1, but we doubled the target for Experiments 2 and 3 because the number of trials per condition was necessarily smaller due to the increased number of factors.

**Procedure.** Figure 2 illustrates a single trial. Each trial consisted of a study phase and a source memory test. For the study phase, first a fixation appeared for 750 ms. Then five words appeared one at a time at random locations exactly 400 pixels away from the center of the screen (measured from the center of the word). No two words appeared closer than 60 degrees (measured from the center of the screen to the center of the word). Each word was presented for 750 ms. Participants had to memorize the location for each word but did not know beforehand which word would be tested on a given trial.

Immediately after the last (fifth) word appeared, a fixation cross appeared for 150, 250 or 350 ms (uniformly distributed; this interval varied for reasons beyond the scope of this article, but we report it for completeness). Then a circle with a radius of 400 pixels appeared in the center of the screen with one of the just-studied words in the middle. A mouse cursor also appeared in the center for each test. Participants had to click on the circle where they thought the word had appeared in the study phase. The circle and the cue word remained on screen until a response was made. The angle difference between the word position and the participant's response was recorded for each trial. When a participant responded too quickly (less than 300 ms), a warning was displayed indicating that the response was too fast and to try to answer as accurately as possible.

The experiment took approximately one hour to complete. To reduce fatigue and increase motivation, there was a break after each block of 30 trials. The break lasted as long as the participant wanted. During the break, a score for the previous block appeared, that was based on

the accuracy for each trial, along with the highest block score up to that point. The score did not provide any additional monetary compensation.

Stimuli were presented using E-prime on a 1680x1050 resolution monitor. Viewing distance was approximately 60 cm. Words were displayed in Courier New 18-point font.

**Design and Materials.** We used a within-subject design with one independent variable – word frequency (High vs Low). The experiment had a total of 10 blocks and each block contained 30 trials, half consisting of HF words and half of LF words. In total, there were 150 trials consisting of HF words, and 150 trial consisting of LF words. The order of HF and LF trials was randomized for each participant with the restriction that there were no more than three consecutive trials from the same frequency class. The words selected for each trial as well as which word would be probed was randomly determined for each participant. Probe words were equally likely to come from each of the five serial positions. Words were sampled without replacement from the pool described below.

The full word pool contained 1500 randomly selected nouns from the $SUBTLEX_{US}$ word database (van Heuven et al., 2014). Word length was between five and seven letters. Each word appeared at least 50 times in the 51-million-word corpus. Half of the words appeared less than two times per million words (LF); the other half appeared > 10 times per million words (HF).

**Mixture model selection and validation**

We used maximum likelihood estimation to fit the mixture model described above individually to each participant's data. The density of responses was modeled as a mixture of a von Mises (vM) distribution centered on 0° with a standard deviation σ (this is the circular equivalent of a normal distribution), four vM distributions each centered on the relative locations of the remaining four non-target words, each with the same standard deviation as the main

distribution, σ, and a uniform guessing distribution (**Figure 3**). Each of the four non-target vM distributions contributed a quarter of the $p_{\text{mis-binding}}$ responses.

We also fit a number of alternative models, including: 1) a single vM distribution without a guessing or a mis-binding component; 2) a single vM distribution plus uniform guessing but no mis-binding; 3) a central vM distribution plus four mis-binding vM distributions but no uniform guessing; 4) the full model described above, but with separate precision parameters for the correct recall and the mis-binded recalls. We compared all models using the Akaike Information Criterion (AIC), which takes into account the number of parameters. All models were fit to each individual participant's data.

Table A1 in Appendix A shows the AIC value of each model fit for each participant, relative to the AIC value of the best fitting model for each participant. For example, a value of 0 means that the model fit was the best for that participant; a value of 4 means that the corresponding model's AIC fit was higher by 4 than the best fitting model. Overall, 68 out of 106 (64%) of participants were best fit by three parameter model presented in **Figure 3** (right; Bays et al., 2009), and 22% of participants were best fit by an identical model that used separate precision parameters for correct and for misbinding responses. The two two-parameter models were preferred in only 14% of participants (combined). The two-parameter models were decisively rejected in 50% of participants (Δ AIC > 10), and moderately rejected in 80% of participants (Δ AIC > 4). Even though the most complicated 4 parameter model was preferred in 23 participants, the model's Δ AIC was greater than 3 only for 3 of those participants. This means that for the remaining 20 participants, the 4-parameter model does not fit substantially better than the simpler 3-parameter model of Bays et al (2009). In summary, the data from 83% of participants supported the 3-parameter model. Thus, we used that model for all analyses presented below. We allowed all

parameters to vary across conditions and participants, and we then analyzed the resulting parameters as a function of condition via T tests. **Figure 3** (right) shows the histogram of errors in all experiments, the mean model parameter values, and the simulated density of the model. **Figure 4** shows the distribution of parameter values for all participants.

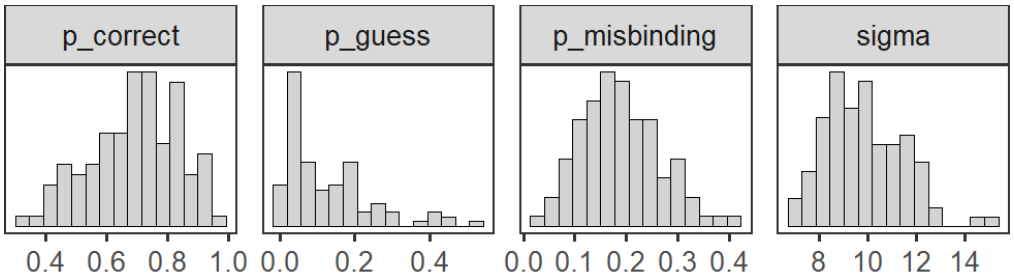


***Figure 4.*** *Distribution of participants' parameter values*

**Results**

The data and analysis code for all three experiments are openly available at https://github.com/venpopov/frequency-rate-interaction.

**Raw error.** Figure 5 shows the main effect of word frequency on the raw error in recalling the associated location. The figure shows the results for all three experiments in different colors to aid comparisons; the report that follows focuses on Experiment 1. Raw errors were analyzed with mixed-effect linear regressions with random intercepts and slopes for each participant. HF word probes led to 4.46 degrees lower overall error (95CI: 1.86-7.06) compared to LF word probes ($\Delta AIC = -5$, $\chi2(1) = 7.456$, $p = .006$). When treating frequency as a continuous variable, increasing frequency over the LF range led to a decrease in error, but the effect plateaued such that changes in frequency over the HF range had no effect (see **Figure 6**, interaction between categorical and

continuous frequency – ΔAIC=-4, $\chi 2(1) = 5.49$, $p = .019$). Thus, in homogenous frequency lists, HF cues were better for retrieving the source location.

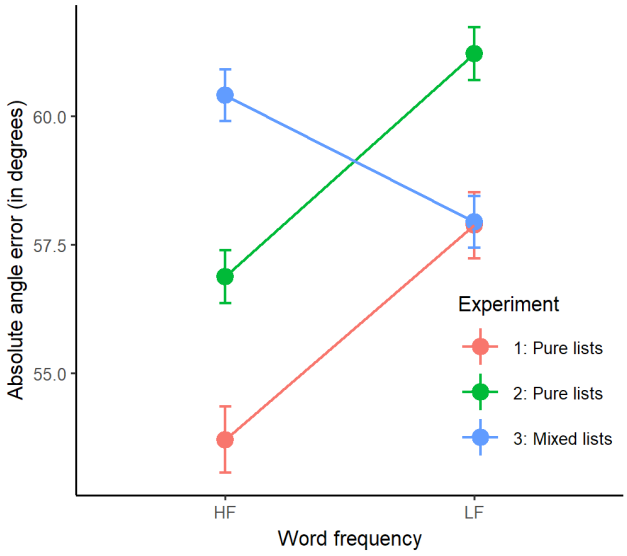


***Figure 5.*** *The effect of word frequency on raw recall error in Experiments 1, 2 and 3. Experiments 1 and 2 used lists of homogenous frequency, while Experiment 3 used mixed lists of both HF and LF words in a single trial. The main difference between Experiments 1 and 2 is that Experiment 2 also varied the presentation rate during study. Error bars represent ±1 Standard Error (SE)*

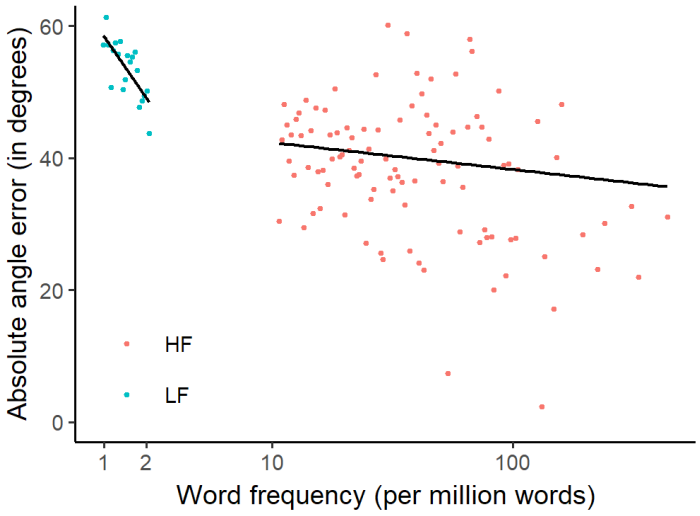


***Figure 6.*** *Continuous effect of word frequency on raw recall error on pure lists (combined data from Experiments 1 & 2)*

**Model parameters.** What gives rise to greater error associated with LF cues? Figure 7 shows the mixture model parameters for HF and LF probes in Experiments 1 and 2. The only two parameters that varied significantly as a function of frequency in Experiment 1 were the probability of remembering the item and the probability of error due to mis-binding. In Exp. 1, HF probes lead to 3.3% greater correct recall compared to LF probes, $t(23) = 2.34$, $p = .029$, and this was because on LF trials there were 4% more errors due to mis-binding, that is, with recalling a location associated with a non-target word, $t(23) = -2.542$, $p = .018$. Neither the guessing rate nor the memory precision differed between HF and LF probes ($ps > .7$).

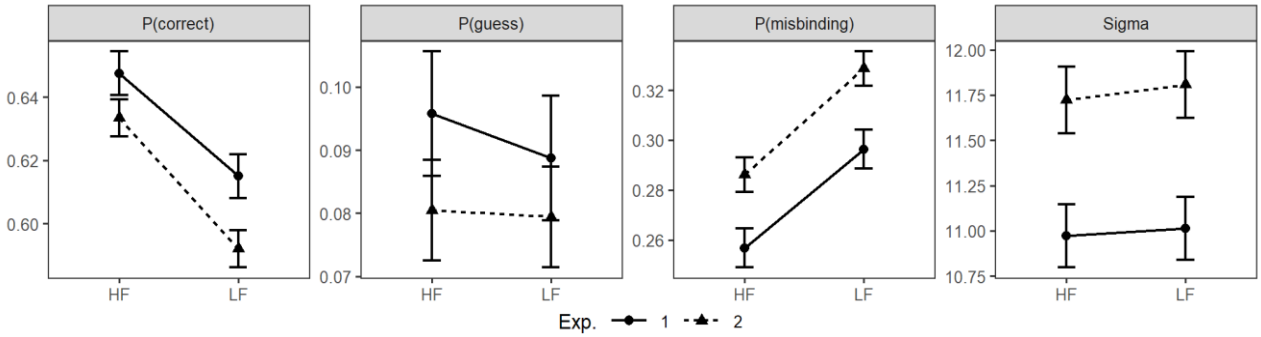


*Figure 7. Mixture model parameter estimates for Experiments 1 and 2 as a function of word frequency. Error bars represent ±1 SE.*

## Discussion

Experiment 1 showed that HF cues are more effective in retrieving the associated location. This finding is contrary to some prior research that showed no effect of cue frequency in cued-recall (Madan et al., 2010), or a negative effect of cue frequency in source-memory (Glanzer et al., 2004; Osth et al., 2018)[4]. Additionally, the mixture modeling revealed that word frequency does not affect memory precision; instead, LF words lead to more mis-binding errors than HF words. Experiment 2 aimed to replicate and extend these results by varying the study duration for each word. SAC predicts that the word frequency effect will increase with faster presentation rates because resources recover at a fixed rate over time. As a result, they would have recovered to a lesser degree with faster presentation rates. This leads to an increased difference between HF and LF words since LF words demand more resources for processing and will be hurt to a greater degree by the faster presentation rate.

## Experiment 2: Interaction between word frequency and presentation rate

[4] These studies used mixed-lists, while Experiment 1 (and Experiment 2) used pure-lists. We will return to this point in Experiment 3, which used mixed-lists.

## Method

**Participants.** Forty-one students from Carnegie Mellon University who were native English speakers participated either for course credit or for a $10 compensation.

**Procedure, materials, and design.** The procedure and the materials were identical to those used in Experiment 1. For Experiment 2, we used a within-subject, 2x3 design. The independent variables were word frequency (High vs Low) and presentation rate (500/750/1000 ms.). Each trial contained only words from the same frequency class and each word within a trial was presented for the same duration. There were a total of 300 trials, with 50 trials in each cell of the experimental design.

## Results

**Raw error.** Figure 5 shows that the overall difference between LF and HF trials on the raw recall error was nearly identical to Experiment 1's. In Experiment 2, HF probes led to 4.78 degrees lower overall error (95CI: 2.56-7.00) compared to LF probes ($\Delta AIC = -11$, $\chi2(1) = 12.86$, $p < .001$). As Figure 8.1 shows, the recall error decreased with slower presentation speeds ($\Delta AIC = -5$, $\chi2(1) = 7$, $p = .008$). Most importantly, the raw error difference between LF and HF trials decreased with slower presentation rates and was almost eliminated when each word was presented for 1s (interaction between word frequency and presentation rate - $\Delta AIC = -2$, $\chi2(1) = 3.84$, $p = .048$). Note that presentation rate does not affect HF trials, but slower rates decrease errors on LF trials.

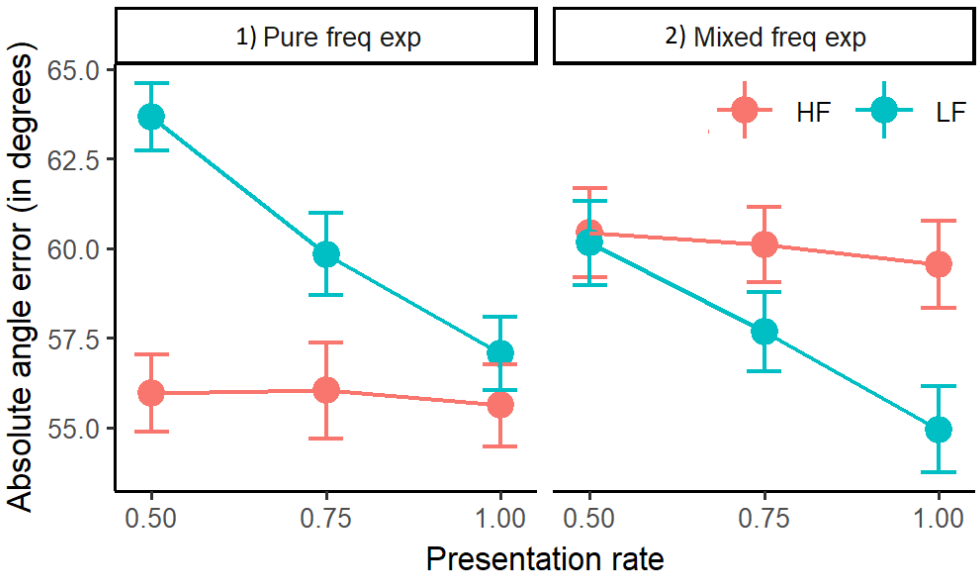


*Figure 8. The effect of word frequency, presentation rate and list-composition on raw recall error. 1) Experiment 2, pure lists of 100% HF or 100% LF; 2) Experiment 3, mixed lists. Error bars represent ±1 SE.*

**Model parameters.** The results replicated Experiment 1 closely (see **Figure 7**). In Exp. 2, HF probes lead to 4.1% greater correct recall compared to LF probes, $t(40) = 3.55$, $p < .001$. This was because there were 4.3% more errors due to mis-binding, $t(40) = -3.08$, $p = .003$. Neither the guessing rate nor the memory precision differed between HF and LF probes (ps > .8). Similarly, the presentation rate affected only the correct recall probability and the mis-binding probability, but not the guessing rate nor the memory precision (**Figure 9**, top row). A linear regression revealed that the only parameter that varied significantly as a function of duration was the probability of misbinding errors, $F(1) = 5.51$, $p = .025$ (all other *ps* > *.19*). We did not explore the interaction between frequency and presentation rate with the model because there were only 50 observations per participant per condition and parameter recovery simulations showed us that the estimated model parameters would not be reliable.

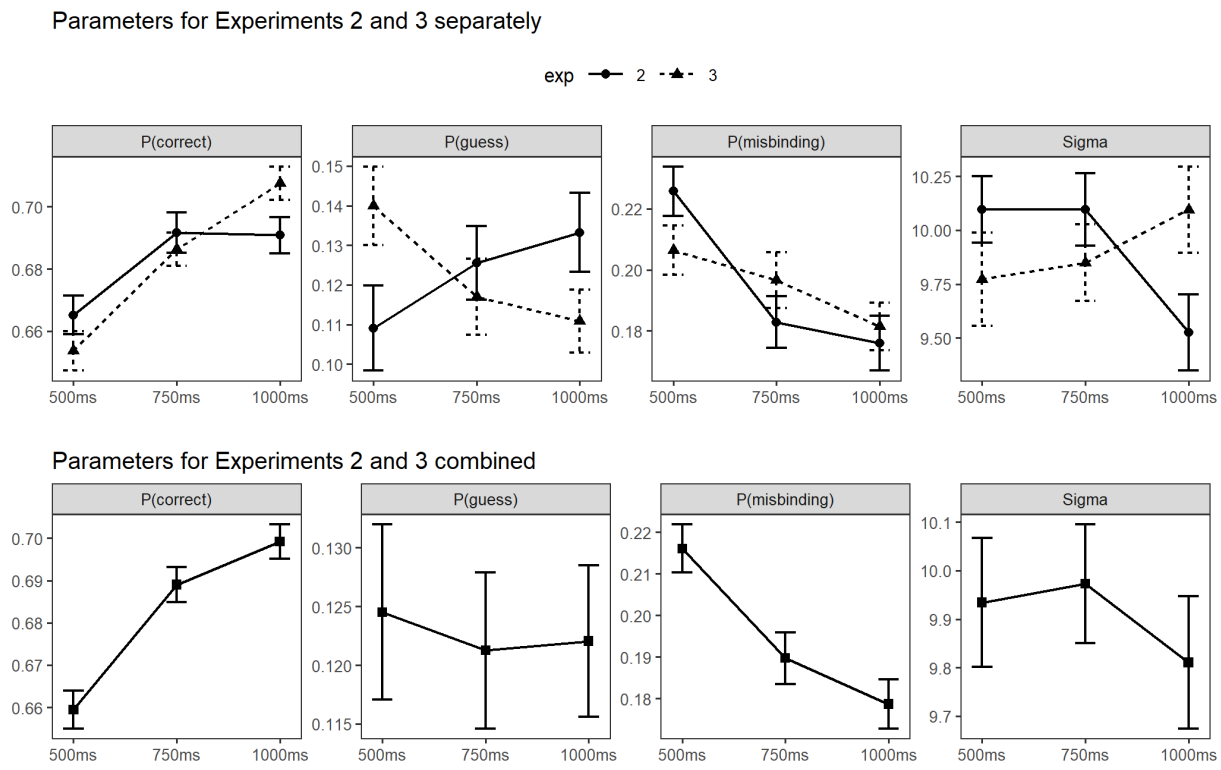


***Figure 9.*** *Model parameter estimates as a function of presentation rate. Top row shows the parameters separately for Experiments 2 and 3; the bottom row shows the parameters for the two experiments combined. Error bars represent ±1 SE.*

## Discussion

Experiment 2 replicated both the raw error findings and the model parameter estimates for the positive effect of word frequency found in Experiment 1 – in trials of pure LF or pure HF composition (i.e., all words are of the same frequency class), HF words lead to better source recall due to fewer mis-binding errors. Slowing down the presentation rate had a similar effect – it increased performance by reducing misbinding errors, but it did not affect precision. Since SAC posits that both the effect of word frequency and the effect of presentation rate are due to the same mechanism of differential resource depletion and recovery, it would have been surprising if the effects of these variables on the model parameters differed. Finally, consistent with SAC's

predictions, the effects of word frequency on source memory were reduced by slowing down the presentation rate, which allows more time for resource recovery.

As we noted before, our results are opposite to those found by prior work (Glanzer et al., 2004; Osth et al., 2018). These studies used mixed lists, and our first two experiments used pure lists; as list-composition usually changes the word frequency effect in free recall, list-composition is a likely cause of this discrepancy. SAC suggests that LF cues have a retrieval benefit that could be masked by their storage deficit. Mixing HF and LF in a single list reduces the resource demands relative to pure lists, as does slowing down the presentation rate. Thus, LF words might lead to fewer errors than HF words in mixed lists with slow presentation rates. Experiment 3 explored how the frequency composition of the trial affects the interaction between word frequency and presentation rate. In addition to varying frequency of the probe word and the presentation rate, we varied whether each trial consisted of 40% LF words (i.e., 2LF and 3HF words per trial) or of 60% LF words (i.e., 3LF and 2HF words per trial).

## Experiment 3: Word frequency, presentation rate and list-composition

### Method

**Participants.** Forty-one students from Carnegie Mellon University who were native English speakers participated either for course credit or for a $10 compensation. Their ages ranged from 18 to 36 years.

**Procedure, materials and design.** The procedure and the materials were identical to those used in Experiment 1. For Experiment 3, we used a within-subject, 2x3x2 design. The independent variables were probe frequency (High vs Low), presentation rate (500/750/1000ms) and proportion of LF per trial (40% vs 60%). As in Experiments 1 and 2, each word within a trial was presented

for the same duration. In contrast to Experiments 1 and 2, words within a trial varied in frequency. There were a total of 300 trials, with 25 trials in each cell of the experimental design.

**Results**

**Raw error.** Consistent with prior research, mixing HF and LF in a single trial reversed the HF advantage (**Figure 5**). Overall, LF probes resulted in 2.4 degrees less errors than HF probes (95CI: 0.68-4.05), $\Delta$AIC = -6, $\chi 2(1) = 7.59$, $p = .006$. This effect was in the opposite direction compared to the pure list experiments. When we combined the data from Experiments 2 (pure lists) and 3 (mixed lists), the regression model revealed that there was a significant interaction between word frequency and list composition, $\Delta$AIC = -19, $\chi 2(1) = 20.43$, $p < .001$. As can be seen from **Figure 5**, performance for HF cues was hurt by the presence of LF cues in the mixed-lists; in addition, performance for LF cues was boosted by the presence of HF cues in the mixed-lists. When we looked at the interaction between word frequency and whether the mixed-lists were composed of 40% or 60% LF words, we found no significant effect, $\Delta$AIC = 1, $\chi 2(1) = 0.86$, $p = .35$. Thus, while it mattered whether the list was pure or mixed, it made no statistical difference whether there were 2 or 3 LF words on the list (out of five).

Furthermore, as in Experiment 2, slowing down the presentation rate significantly improved performance, $\Delta$AIC = -10, $\chi 2(1) = 11.65$, $p < .001$; however, the interaction between frequency and presentation rate was not significant in Experiment 3 ($p = .35$), although the pattern shown in Figure 8 showed a similar effect as in Experiment 2. Combining data from experiments revealed a significant frequency by serial position interaction overall (see the Combined Results section). Figure 8 shows how probe frequency, list-composition and presentation rate interact. There are several notable patterns. First, presentation rate does not affect source recall for HF probes ($\Delta$AIC = -1, $\chi 2(1) = 2.71$, $p = .10$, for Experiment 3) and this holds regardless of the

frequency composition of the trial. Second, slowing down the presentation rate improves recall for the location of LF probes in both pure and mixed lists ($\Delta$AIC = -8, $\chi 2(1) = 9.42$, $p < .002$ for Experiment 3).

In summary, Experiments 2 and 3 provide support for the claim that disparate findings in the literature could be due to differences in presentation rate, list-composition, or both. The way the frequency effect changes with presentation rate and with list-composition is consistent with the encoding-retrieval frequency trade-off suggested by SAC (Figure 1; Reder et al., 2007; Popov & Reder, 2020a).

**Model parameters.** Similar to the first two experiments, the only parameter that was affected by word frequency was the probability of retrieving the correct location, $t(40) = 2.12$, $p = .04$. Consistent with the raw error results, LF words lead to better recall of the correct location (2.4%). The effect of duration on model parameters was qualitatively similar to what we found in Experiment 2 (see Figure 9); however, none of these differences were statistically significant (all *ps* > 0.09).

## Combined results

### Raw Error – Frequency by serial position interaction

Given the frequency by presentation rate interaction is one of the key predictions of SAC, it is concerning that Experiment 2 showed only a marginally significant interaction and that Experiment 3 failed to find statistical evidence for the interaction. It is possible that on their own, each of the two experiments was underpowered to detect the interaction. To increase the power of our statistical test, we combined the data from Experiments 2 and 3 and repeated the mixed-effects regression model. The analysis of the combined dataset showed evidence for a significant frequency by presentation rate interaction, $\Delta$ AIC = -5, $\chi 2(1) = 5.59$, $p = .018$. We acknowledge

that this was an unplanned post-hoc analysis. However, analyzing the combined data makes sense from a theoretical perspective – SAC predicts that the interaction should appear in both pure lists (Experiment 2), and in mixed lists (Experiment 3). Thus, combining the data from the two experiments is a reasonable procedure, even if it was not planned beforehand.

**Model parameters**

Even though *presentation duration* significantly affected the raw error similarly in Experiments 2 and 3, the analyses of the model parameters did not reach statistical significance when analyzing those experiments separately. To increase the analysis power, we analyzed the parameter estimates for the effect of duration of both experiments combined (**Figure 9**, bottom row). The probability of making misbinding errors decreased significantly as the presentation rate increased, $F(1) = 6.44$, $p = 0.012$; conversely, the probability of selecting the correct location marginally increased as the presentation rate increased, $F(1) = 3.70$, $p = 0.057$; there were no effects of duration on either the guessing rate, $F(1) = 0.015$, $p = 0.903$, or memory precision, $F(1) = 0.151$, $p = 0.694$.

Experiments 1 and 2 both showed no statistically significant effect of *word frequency* on memory *precision*. However, a non-significant p-value is not evidence *against* an effect of frequency. We followed up on the previous analysis by performing a Bayesian T-test on the sigma parameter as a function of word frequency after combining the data from both experiments via the *BayesFactor* package (Morey & Rouder, 2015). We used the default prior settings in the package. The Bayes Factor in favor of the null hypothesis was $BF_{null} = 6.8$, which is considered moderate evidence for the null (Lee & Wagenmakers, 2013) and evidence *against* a word frequency effect on precision. Similarly, a Bayesian linear regression revealed that there was moderate evidence

*against* a presentation duration effect on precision ($BF_{null}$ = 6.6; combining data from Experiments 2 and 3).

**Serial position effect**

Thus far we demonstrated that word frequency and presentation rate affect memory by affecting the mis-binding probability; in contrast, they do not affect the precision of the memory trace. One could argue that our mixture model is simply not sensitive to differences in memory precision, which would invalidate the conclusions presented above. Thus, we need to demonstrate that differences in precision can be detected using our mixture model. For this reason, it is worth noting how the model parameters differ as a function of the temporal serial position of the probe word in the study sequence. Figure **10** shows that the combined results of the three experiments produced a typical serial position curve – lower performance for probes presented farther back in the sequence (a recency effect), except for the first studied word (a primacy effect). Similar serial position curves occur for probability of correct recall, probability of mis-binding and most important – for memory precision. That is, recently studied probes are associated with much higher precision than probes presented earlier in the trial. Thus, the absence of an effect of word frequency on precision is not due to an inability of the model to detect differences in precision.

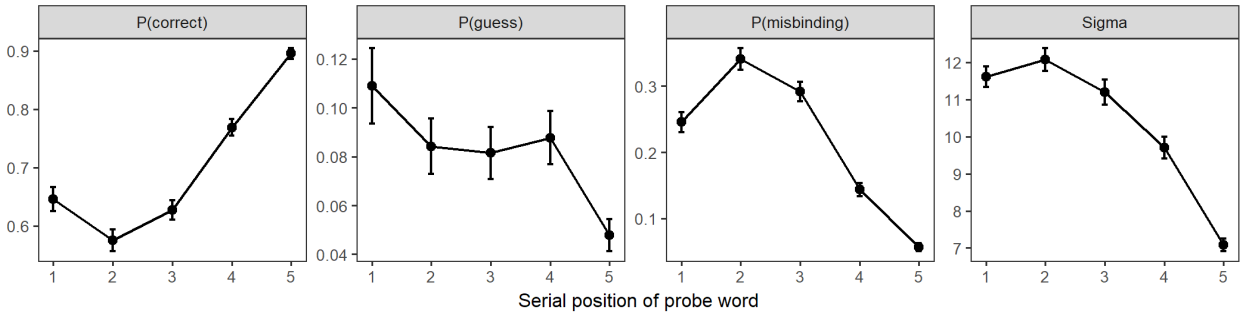


*Figure **10.** Mixture model parameter estimates as a function of serial position. Error bars represent ±1 SE.*

**Source of mis-binding errors (contiguity effects)**

The key finding that LF words increase mis-binding errors rather than decreasing the precision of the recalled locations suggests that when resources are depleted, the binding between the word and the location fails altogether. However, there is an alternative interpretation of this result. The mixture model operates on retrieval data and, as such, it is not possible to determine whether a mis-binding error occurs because a binding was not created in the first place or whether a weak binding failed to pass a retrieval threshold. The absence of an effect of word frequency or duration on recall precision indicates that when a binding can be retrieved, the precision of the associated location does not differ. However, for cases in which the binding cannot be retrieved, a memory trace could still exist even if no information was successfully accessed from it[5].

In order to shed more light on this question, we decided to investigate the source of the mis-binding errors using the law of contiguity (Davis et al., 2008; Kahana, 1996). Prior work has established that events experienced closely in time become associated with one another. For example, contiguity effects in free recall show that when participants recall an item that was

[5] We thank Adam Osth for positing this counter-argument

studied in temporal serial position X, they are much more likely to then recall the item that was studied in the subsequent temporal position X+1 than any other item studied on the list (Kahana, 1996). Similar effects occur in cued recall tasks in which participants first learn a list of paired associates ($A_1$-$B_1$, $A_2$-$B_2$, $A_3$-$B_3$,…), and they are then asked to recall a B item associated with a cued A item. Specifically, if cued with $A_i$ and unable to retrieve the correct $B_i$ associate, participants are more likely to make an intrusion error with a temporally neighboring $B_{i+1}$ item (Davis et al., 2008; although see Osth & Fox, 2019 for failure to find similar effects in associative recognition). These temporally proximate intrusion errors suggest that even though the correct $A_i$-$B_i$ binding could not be retrieved, a memory trace with associated temporal information existed.

Based on this prior work, we expected that if no information is stored in memory for trials in which mis-binding errors occur, participants should respond with one of the other four stored locations at random. Alternatively, if mis-binding errors in source memory exhibit contiguity effects, this would indicate that despite failure to retrieve the correct word-location binding, a memory trace with associated temporal information was in fact stored.

In order to test these alternatives, we performed a contiguity analysis in the following way. For each trial, we calculated the probability that the response comes from each of the five encoded locations, based on the von Mises likelihood:

$$P_{i,j} = \frac{VM\left(\hat{x}; \theta_{i,j}, k\right)}{\sum_{j=1}^{5} VM\left(\hat{x}; \theta_{i,j}, k\right)}$$

where $P_{i,j}$ is the probability that the response $\hat{x}$ on trial *i* comes from the location presented at serial position *j*, $\theta_{i,j}$. VM is the von Mises distribution and *k* is the memory precision estimated by the mixture model fitted to each participant's overall data. **Figure 11** shows the result of this analysis, separately for probes in each serial position. As can be seen from the figure, participants

were most likely to recall the correct location (there is a match between the serial position of the word probe and the most probable serial position of the recalled location as estimated by the model). However, when a mis-binding error occurred, that error was most likely to come from locations presented in neighboring serial positions.

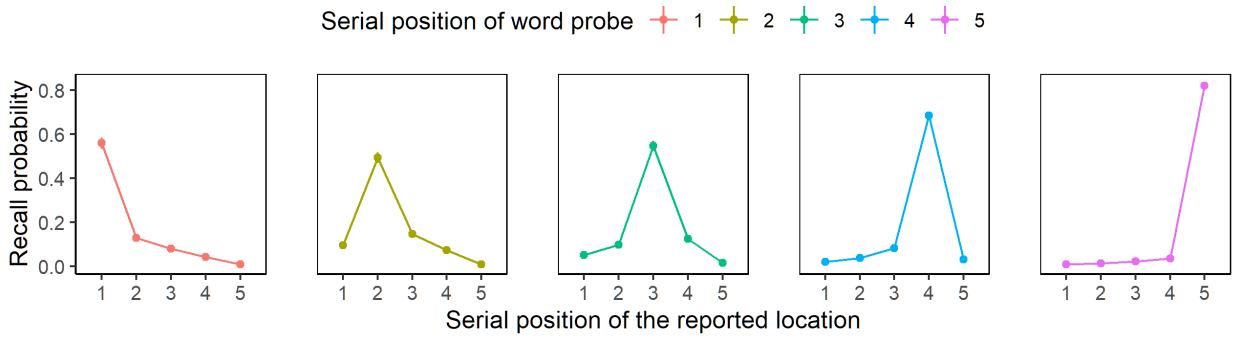


***Figure 11.*** Probability of recalling each of the five studied locations depending on the serial position of the word probe. Recall probability of each studied location was estimated by a mixture model (see main text for details).

Typically, contiguity effects are plotted in terms of the temporal lag between a target item and the recalled item. To do this, for each trial we calculated the same probability described above, but we coded the temporal distance between the probe's serial position during study, and the serial positions associated with each possible misbinding error. The black dots and lines in **Figure 12** show this effect collapsed over serial positions and focuses only on the mis-binding errors – lag of 1 means that the mis-binding error was a location that was studied in a serial position immediately following the probed word, and lag of -1 means that it was studied in the immediately preceding serial position. Thus, even when participants failed to retrieve the correct word-location binding, they did not respond with one of the other 4 studied locations at random. Instead, they were more likely to respond with locations that appeared in temporal proximity to the probe, indicating that a

binding between the word probe and the temporal context was stored and accessed despite failure to retrieve the correct location.

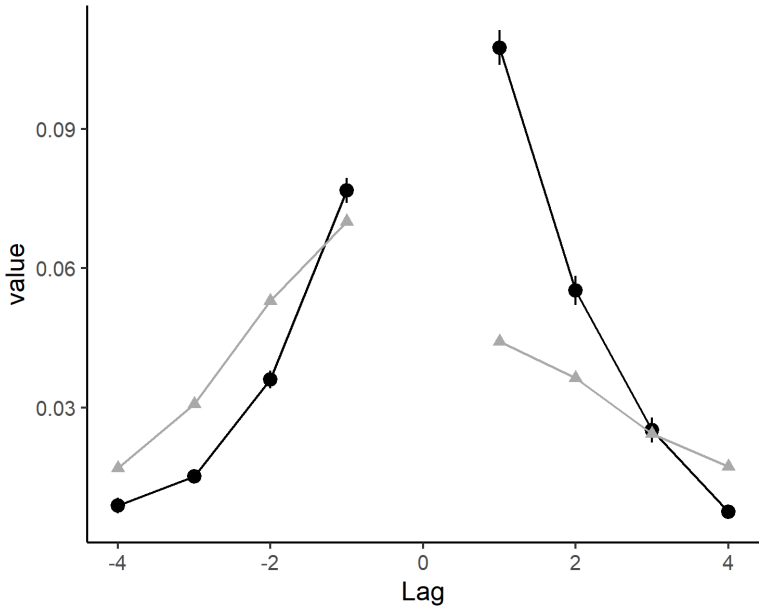


***Figure 12.*** Contiguity effect in source memory. The probability of a mis-binding error/intrusion from locations studied at serial positions i+lag relative to the serial position of the probe. Error bars represent ±1 SE. The black dots represent the observed data. The grey triangles represent the expected data if participants made a mis-binding error/intrusion from a random serial position.

Before we move on, it is important to show that these contiguity effects are not an artifact of the analysis procedure. As Healey et al. (2019) demonstrated, contiguity effects in free recall could be produced artificially due to the serial position curve, even if temporal order does not influence recall. Artificial contiguity effects can occur when there is autocorrelation in the availability of items at different serial positions – if a participant recalls an item from serial position 1, then they will be more likely to recall an item from serial position 2 simply due to the primacy effect. To avoid this problem, the typical analysis of contiguity effects in free recall – the conditional-lag probability – computes the probability of transition from position i to position i+1

conditional on the successful recall of both items. Unfortunately, this correction is not possible to do for our task, because only a single item is probed on each trial, and we do not know which of the remaining four items/locations have been successfully encoded in memory.

To determine whether our contiguity effects could be an artifact of the serial position curve, we instead followed Healey et al (2019) and performed a Monte Carlo simulation. We started with the estimates of correct recall probabilities as a function of serial position presented in Figure **10**. We then used this curve to define a binomial distribution of recall probability for each serial position. From this curve we simulated 10,000 trials in the following way. For each serial position, we determined whether the item-location binding is available in memory by drawing a sample from the corresponding binomial distribution. If the correct item-location binding for a particular serial position was not available in memory, we randomly selected one of the remaining unavailable item-location bindings as the response. Finally, we calculated over all simulated trials, the probability to incorrectly recall a location associated with items at different lags. The grey triangles and lines on **Figure 12** show the results. Indeed, this simulation produced an artificial contiguity effect even when mis-binding errors were selected at random from the unavailable item-location bindings. However, this artificial contiguity effect was much less pronounced than the effect in the observed data, with nearly three times as many misbindings for the actual data at lag +1 relative to the simulated data. This suggests that the observed contiguity effect was not simply an artifact of the serial position curve.

**Performance decline over successive trials.**

Finally, we also investigated how performance changes over time in the task. The resource depletion and recovery assumptions predict that performance should gradually decline over

successive trials of a memory test, and it should bounce back after a break[6]. While this effect has not been previously observed (Hitch et al., 2009; Page et al., 2013), it should be noted that whether such an effect would occur depends on the presentation rate and the inter-trial-interval. The median time for full resource recovery in all our model simulations was 4.83 seconds (see Popov & Reder, 2020), meaning that this prediction would be observed only with very short inter-trial intervals. That was not the case in the above-mentioned studies (for example, Hitch et al., 2009, had a 12 second break between trials).

Even though the current study was not specifically designed to test this prediction, several aspects of its design make it more suitable than the studies discussed above to address this question. The inter-trial-interval (the time between the location response on one trial and the beginning of the next trial) was only 750 ms long, and the median response time to the source memory test at the end of each trial was 1837 ms. Thus, the interval between trials was very short compared to the above-mentioned studies, making this dataset more appropriate for this analysis.

We looked at the angle error between the target location and the response location as a function of 1) the trial position within a block (1-30), and 2) as a function of block number. There were self-paced breaks after every block of 30 trials, so if the prediction is correct, we should see the performance deteriorate as a function of trial number within a block. **Figure 13**, left panel, shows exactly that result (a mixed-effects linear regression with random intercepts for each participant confirmed that this effect is significant, $\Delta$ AIC = -4, $\chi 2(1) = 6.23$, $p = .013$). Since there was a break after each block, this result means that while performance declined as each block progressed, it bounced back after the break for the beginning of a new block. This result was not due to general fatigue – if we look at the average performance across the 10 blocks (**Figure 13**,

[6] We are grateful to the anonymous reviewer who pointed out this prediction

right panel), we see that performance improved over blocks due to practice, while it declined within blocks.

Thus, our data show that when the inter-trial interval is kept short and participants have to respond to a single probe item between trials, we observe exactly what the resource-recovery-and-depletion assumptions predict – a decline over successive trials followed by a bounce back after a break, even when there is no general fatigue over time.

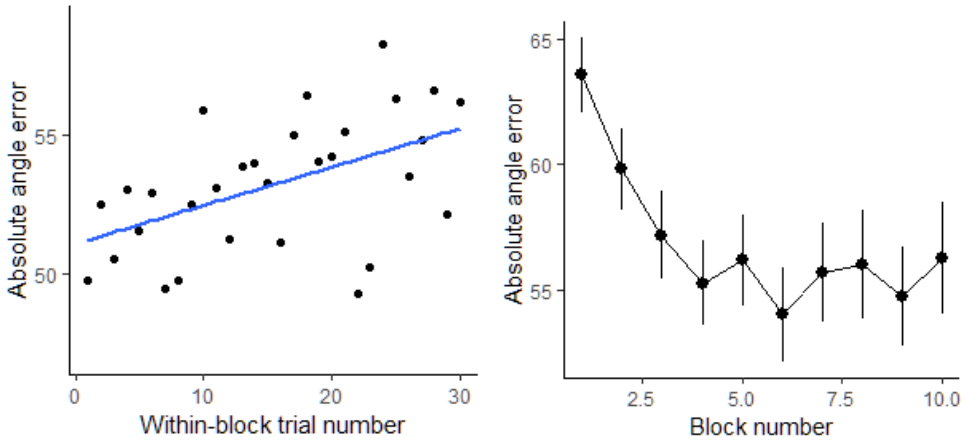


***Figure 13.*** *Left panel – Absolute angle error increases with each trial within a block. Right panel – absolute angle error decreases across blocks. Error bars represent ±1 SE.*

## General Discussion

The current experiments provided a rich dataset concerning the effects of word frequency, presentation rate and list composition on source memory for word locations in working memory. Several notable patterns emerged:

- Better source memory for HF words, but only in pure rather than mixed frequency trials. This effect was a continuous function of word frequency.
- The frequency effect decreased linearly with slower presentation rates because source memory for LF words, but not for HF words, improved with slower presentation rates.

- The greater the proportion of LF words on the study trial, the worse the source memory for all items. This list-composition effect did not interact with presentation rate.
- Source recall was worse for LF words because they were more likely to be mis-bound to an incorrect location; when the correct location was recalled, there was no difference between the precision of the memory for HF vs. LF words.
- When a mis-binding error occurred, the incorrectly recalled location was more likely to come from neighboring temporal positions during study
- Performance declined over successive trials within each block, but it bounced back after each break. At the same time, performance improved over blocks

Our interpretation of these results is based on the SAC model of memory (Popov & Reder, 2020; Reder et al, 2007) – LF words require (and thus consume) more resources for their processing, leaving fewer resources for binding them to the location in which they are presented. Given that (a) these resources are depleted with each word, (b) that the rate of resource recovery is constant, and (c) that LF words require more resources, it follows that the effect of word frequency should increase with faster presentation rates. For the same reason, increasing the proportion of LF words in the trial decreases memory performance for all items on the trial – with more LF items in the study set more resources are consumed.

These results, that cue frequency can have either positive, null or negative effects in cued-recall/source memory depending on the presentation rate and list composition, could potentially explain why prior studies have shown conflicting results (Criss et al., 2011; DeWitt et al., 2012; Diana & Reder, 2006; Glanzer et al., 2004; Madan et al., 2010; Osth et al., 2018; Reder et al., 2016). As this study has demonstrated, whether a HF or a LF cue will lead to better performance depends on the resource demands during encoding – reducing the resource demands, either via

slower presentation rate or via the presence of HF words on a mixed list, allows LF words to be stored well, and reveals their retrieval advantage. To be clear, our argument is not that prior studies have confounded presentation rate, list-composition and cue frequency – our argument is that the specific levels of presentation rate and list-composition chosen in prior studies might have contributed to whether they will show a positive or a negative cue frequency effect.

The fact that the LF cue retrieval advantage can be masked depending on list-composition and presentation rate is theoretically important. The LF retrieval advantage in item recognition has been attributed to less contextual competition for LF words because they have been experienced in fewer contexts (for a recent review, see Popov & Reder, 2020a; Osth & Dennis, 2015; Reder et al., 2000, 2007; Dennis & Humphreys, 2001; Criss et al., 2011). As Criss et al. (2011) argued, a null or a positive effect of high frequency cues is inconsistent with most current memory models. They argued against the idea that LF cues have less contextual competition because they found little-to-no difference in performance depending on the frequency of the cue (Criss et al., 2011; Nelson & McEvoy, 2000). Our results put doubt on this conclusion by showing that under the right circumstances, LF cues lead to better memory, as predicted by contextual competition models. One of the main takeaways is that the LF cue benefit could be masked by the encoding-retrieval trade-off we described in the introduction. Thus, contrary to Criss et al's (2011) claim, SAC can explain these results by positing that under high resource demands conditions, LF words and their context bindings are formed weaker relative to HF words, which masks the LF cue advantage.

One limitation of the method used in these experiments is that there was always at least 60 degrees distance between every two words on the circle. Since five words were presented on each trial, this means that words were relatively equidistant from each other. This was done to avoid visual overlap between the strings for different words. However, it could be argued that this spatial

arrangement could cause people to encode locations verbally (for example, using labels for clock positions – 7 O'Clock, 11 O'Clock, etc), rather than spatially. We do not believe this is a major problem for the interpretation of our results for two reasons. First, even though words were presented with a minimum of 60 degrees between them, the overall positions were randomized on each trial. That is, any of the 360 positions on the circle were equally likely to be presented over the course of the experiment. The five locations on a trial were also presented randomly, and not in a specific order (clockwise or counterclockwise). Since the five positions were different on each trial, and a participant could not predict where the next word would appear, it is unlikely that such a strategy would be beneficial. Recoding a spatial position, which is perceived directly, into a verbal label would also require extra time and effort. The relatively fast presentation rates (500/750/1000 ms per word) would make that difficult.

Second, even if participants did use verbal labels to encode the spatial position of words, this would not be an issue for the interpretation of the results - as we note in the introduction, source memory is a variant of a paired-associate learning task – the key aspect of the task is that a binding must be formed between two features of the stimulus. In a source memory task, one of those features is a contextual detail, in this case, the spatial location. In SAC, a binding is an episodic node that connects other nodes representing features of the stimuli and context (Popov & Reder, 2020; Reder et al., 2000; Reder et al., 2007). This binding does not differ depending on the nature of the features it connects – that is, the same episodic node is formed regardless of whether two words are being bound together, or whether a word and a contextual feature are being bound together. The main claim we investigated in this paper is that encoding LF words leaves fewer resources for forming the episodic bindings. Testing this claim would not be influenced by whether

participants encode the locations spatially, as we assume, or whether they first recode them into verbal labels.

**Is any information stored when resources are insufficient?**

The current study presents clear and novel evidence that LF words and short study times lead to more mis-binding errors rather than to diminished precision of the memory trace. While this result is important in and of itself, a careful consideration revealed that it cannot conclusively answer one of our original motivating questions, namely, whether people fail to store information in memory when they have insufficient resources available. The fact that a binding fails to be retrieved does not necessarily mean that it does not exist – it is possible that the memory trace is so weak that it fails to pass a retrieval threshold. Even though there were no differences in memory precision as a function of word frequency or study duration, it should be noted that the current methodology can only reveal the precision of a memory trace if the trace could be retrieved. This is because precision is measured as the angle error specifically for responses that are considered correct. Thus, the current results show that if a memory trace passes a retrieval threshold, it is equally precise regardless of condition. However, the motivating question concerns the precision of memory traces that cannot be recalled.

The analysis of contiguity effects provides some clarity concerning this issue. We showed that just like in cued recall, mis-binding errors are more likely to come from neighboring serial positions (in time, not space). This suggests that even though in these cases the word-location binding could not be retrieved, some temporal information, and possibly a temporal-spatial mapping, must have been stored together with each probed word[7]. If no memory trace were created on these trials, mis-binding errors would not have exhibited contiguity effects. One possibility is

[7] Alternatively, direct associations among neighboring probes could have also been stored.

that the memory trace for each probe by default contains a binding (or multiple bindings) between the probe, the location, and a temporal context vector. It is possible that when resources are running low, only some portion of those bindings are created (e.g., the probe-time binding and not the probe-location or the location-time binding). Further research is required to answer this question more conclusively.

**Contiguity effects**

The contiguity effects we uncovered are important in their own right. While contiguity effects are well established for free recall (for a review, see Healey et al., 2018; Kahana, 1996), only one study so far has shown contiguity effects in cued recall (Davis et al., 2008, 2008) and recent research has found no such effects in multiple associative recognition experiments (Osth & Fox, 2019). Furthermore, we are not aware of any contiguity effect demonstrations in source memory. Why is this important? As Osth et al. (2019) notes, temporal context models such as TCM (Howard & Kahana, 2002) predict that contiguity effects should occur in all of these tasks, because associations are automatically formed between temporally contiguous events, regardless of what type of memory test is expected. In contrast, in models such as SAM (Gillund & Shiffrin, 1984), due to a capacity-limited buffer, associations in paired associate tasks are formed only among elements that are concurrently presented.

Finding contiguity effects in the current task is theoretically important because temporal information is not helpful or required to perform the current task successfully (in contrast, temporal information benefits free recall performance because it guides the memory search). Specifically, since only one item is probed at test in our task, it does not matter where that item was presented relative to other items in the trial. In fact, since the formation of associations is resource intensive (Popov et al., 2019; Popov & Reder, 2020b; Reder et al., 2007), one could argue that forming

temporal associations in this task might be counterproductive because it would leave fewer resources for encoding the item-location bindings. The fact that temporal associations are nevertheless formed provides support for the claim that such associations are automatic and central to the representation of information in memory.

**Word frequency vs context diversity**

Word frequency is a quasi-experimental variable because words cannot be randomly assigned to be of either low or high frequency. For this reason, word frequency is often confounded with other lexical variables such as semantic distinctiveness, concreteness, age of acquisition, contextual diversity and so forth (Maddox & Estes, 1997). It is possible, in principle, that some of these other factors could explain the effect of word frequency in our experiments, as we only controlled for word length differences between HF and LF words. There are several reasons why we did not apply controls to the selection of HF and LF words and why we believe the results presented here reflect the effect of frequency rather than other confounding factors.

Our prior work has repeatedly demonstrated that experimentally manipulating frequency of exposure by differential pretraining of pseudowords (Reder et al., 2002) or of unfamiliar Chinese characters (Reder et al.,2015; Shen et al., 2018) leads to the same effects as word frequency in recognition tasks, cued-recall tasks, and tests of working memory. Based on this prior work, we know that frequency of exposure affects memory in the absence of other lexical confounds, and thus we felt it was appropriate to vary word frequency without controlling for other factors. The second reason why we did not apply such controls is that fitting the mixture models to individual participants requires a large number of trials, and the current experiments presented a total of 750 HF and 750 LF words to each participant. As we wanted these words to be unique nouns of average length, this left us with a small pool of potential words, most of which were used

in the experiment. Applying controls to the word pool would have required us to select a smaller set of words which would have compromised our ability to apply the mixture models to the distribution of errors. Since we know based on our prior work that frequency directly affects memory in the absence of confounds, we chose not to control for potential confounds and use a broad set of words in order to apply our mixture models.

One particular lexical confound deserves a more extended discussion, namely, contextual diversity. Contextual diversity is a measure of how often a word appears in different contexts (Brysbaert & New, 2009), and it is highly correlated with word frequency (Steyvers & Malmberg, 2003). Despite this high correlation, studies that have manipulated these factors independently have found that both have separate effects on recognition and recall tasks (Parmentier et al., 2017; Steyvers & Malmberg, 2003). These two factors correlated highly in our stimuli ($r = 0.94$), which prevents us from disentangling their effects. Nevertheless, several things are worth noting. First, contextual diversity cannot explain the positive effect of word frequency in the pure lists used in Experiments 1 and 2 – HF words have higher contextual diversity than LF words and high contextual diversity is typically associated with worse recognition and recall (Hicks et al., 2005; Parmentier et al., 2017; Steyvers & Malmberg, 2003)[8], and source memory (Marsh et al., 2006), which is the opposite of what we found.

What about the negative frequency effect in the mixed-lists of Experiment 3? One of the main claims of the SAC model, that we discussed in the introduction, is that word frequency effects are the results of a trade-off between a HF encoding advantage and a LF retrieval advantage. We claim that whether one observes a positive or a negative effect of word frequency depends on the

[8] Note that the effect of contextual diversity on serial recall has been questioned in a recent failed replication of Parmentier et al. (Guitard et al., 2019). However, the effects do appear in recognition (Steyvers & Malmberg, 2003) free recall (Hicks et al., 2005) and particularly important, in source memory (Marsh et al., 2006).

balance of these two factors, which itself depends on the encoding demands of the task. The LF retrieval advantage in the model occurs because LF words are associated with a smaller variety of previous contexts, which makes it more likely that cueing memory with a LF word would retrieve the correct episodic context. This is the definition of contextual diversity. Thus, the fact that contextual diversity and word frequency are confounded in the stimuli is not a bug but a feature of the current experiments, as it allows us to observe this trade-off shifting as the encoding demands of the task change.

**Verbal memory and precision**

In the current experiments we did not find any effects of either word frequency or presentation rate on the precision parameter of the mixture model. We did find an effect of serial position on memory precision, which suggests that the model could detect differences in precision when they exist. We imported the concept of memory precision from the visual working memory literature, and it is reasonable to ask whether it has any relevance to verbal memory. The key assumption that we made in this study was that this precision parameter is a proxy for the strength of memory traces. SAC and many other models of memory assume that memory traces vary continuously in strength, and that recall depends on whether this continuous strength passes a retrieval threshold. In typical verbal recall or recognition paradigms, one cannot observe memory strength directly, only being able to measure what proportion of traces are above a participant's retrieval threshold. By analogy to the visual working memory literature, we assumed that the p_correct parameter of the mixture model corresponds to the same categorical measure typically used for recall accuracy. In contrast, we assumed that the precision parameter reflects the continuous strength of those traces that were successfully retrieved, with stronger traces leading to more precise location memories. Thus, our conclusions depend on this interpretation of the

mixture model parameters, which we believe are plausible given the assumptions of the SAC model.

**Comparison to the Item-Order Hypothesis**

SAC's explanation of the mixed-list paradox is similar to the item-order hypothesis (DeLosh & Mcdaniel, 1996; Serra & Nairne, 1993) according to which low-frequency items require more resources for processing, leaving fewer resources for encoding the order of items. Since participants often use information about serial order to guide their free recall (Healey et al., 2018), insufficient resources to encode order information would diminish recall performance for HF words in mixed-lists. One difference between their account and the SAC account is that SAC proposes a more general resource depletion mechanism – depleting more resources for LF items leaves fewer resources for processing any other information, be it other items, context, location, order, etc. This difference allows SAC to explain why the mixed-list paradox appears in recognition memory (Malmberg & Murnane, 2002) or source memory (current experiments) in which a single item is probed and order is irrelevant. The item-order hypothesis would struggle to explain why we see a list-composition effect in the current experiments, because disrupting order information should not have detrimental effects on memory for single probes. Aside from this difference, SAC is a computational implementation of this idea, while the item-order hypothesis was never implemented formally in a computational model. Given the general similarities, SAC could be considered a computational generalization of the item-order hypothesis.

## References

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), 7–7. https://doi.org/10.1167/9.10.7

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Clark, S. E. (1992). Word frequency effects in associative and item recognition. *Memory & Cognition*, *20*(3), 231–243. https://doi.org/10.3758/BF03199660

Criss, A. H., Aue, W. R., & Smith, L. (2011). The effects of word frequency and context variability in cued recall. *Journal of Memory and Language*, *64*(2), 119–132. https://doi.org/10.1016/j.jml.2010.10.001

Criss, A. H., & McClelland, J. L. (2006). Differentiating the differentiation models: A comparison of the retrieving effectively from memory model (REM) and the subjective likelihood model (SLiM). *Journal of Memory and Language*, *55*(4), 447–460.

Davis, O. C., Geller, A. S., Rizzuto, D. S., & Kahana, M. J. (2008). Temporal associative processes revealed by intrusions in paired-associate recall. *Psychonomic Bulletin & Review*, *15*(1), 64–69. https://doi.org/10.3758/PBR.15.1.64

Deese, J. (1960). Frequency of usage and number of words in free recall: The role of association. *Psychological Reports*, *7*(2), 337–344.

DeLosh, E., & Mcdaniel, M. (1996). The role of order information in free recall: Application to the word-frequency effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1136–1146. https://doi.org/10.1037/0278-7393.22.5.1136

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*(2), 452.

DeWitt, M. R., Knight, J. B., Hicks, J. L., & Ball, B. H. (2012). The effects of prior knowledge on the encoding of episodic contextual details. *Psychonomic Bulletin & Review*, *19*(2), 251–257. https://doi.org/10.3758/s13423-011-0196-4

Diana, R. A., & Reder, L. M. (2006). The low-frequency encoding disadvantage: Word frequency affects processing demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 805.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*(1), 1.

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 5.

Glanzer, M., & Bowles, N. (1976). Analysis of the word-frequency effect in recognition memory. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(1), 21–31. https://doi.org/10.1037/0278-7393.2.1.21

Glanzer, M., Hilford, A., & Kim, K. (2004). Six Regularities of Source Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1176–1195. https://doi.org/10.1037/0278-7393.30.6.1176

Gregg, V., Montgomery, D. C., & Castano, D. (1980). Recall of common and uncommon words from pure and mixed lists. *Journal of Verbal Learning and Verbal Behavior*, *19*(2), 240–245.

Guitard, D., Miller, L. M., Neath, I., & Roodenrys, S. (2019). Does contextual diversity affect serial recall? *Journal of Cognitive Psychology*, *31*(4), 379–396. https://doi.org/10.1080/20445911.2019.1626401

Healey, M. K., Long, N. M., & Kahana, M. J. (2018). Contiguity in episodic memory. *Psychonomic Bulletin & Review*, 1–22.

Healey, M. K., Long, N. M., & Kahana, M. J. (2019). Contiguity in episodic memory. *Psychonomic Bulletin & Review*, *26*(3), 699–720.

Hicks, J. L., Marsh, R. L., & Cook, G. I. (2005). An Observation on the Role of Context Variability in Free Recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(5), 1160–1164. https://doi.org/10.1037/0278-7393.31.5.1160

Hitch, G. J., Flude, B., & Burgess, N. (2009). Slave to the rhythm: Experimental tests of a model for verbal short-term memory and long-term sequence learning. *Journal of Memory and Language*, *61*(1), 97–111.

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology-Learning Memory and Cognition*, *23*(5), 1217–1232.

Hulme, C., Stuart, G., Brown, G. D. A., & Morin, C. (2003). High- and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial recall. *Journal of Memory and Language*, *49*(4), 500–518. https://doi.org/10.1016/S0749-596X(03)00096-2

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, *24*(1), 103–109.

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience*, *17*(3), 347–356. https://doi.org/10.1038/nn.3655

Madan, C. R., Glaholt, M. G., & Caplan, J. B. (2010). The influence of item properties on association-memory. *Journal of Memory and Language*, *63*(1), 46–63. https://doi.org/10.1016/j.jml.2010.03.001

Maddox, W. T., & Estes, W. K. (1997). Direct and indirect stimulus-frequency effects in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(3), 539.

Malmberg, K. J., & Murnane, K. (2002). List composition and the word-frequency effect for recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 616–630. https://doi.org/10.1037/0278-7393.28.4.616

Malmberg, K. J., & Nelson, T. O. (2003). The word frequency effect for recognition memory and the elevated-attention hypothesis. *Memory & Cognition*, *31*(1), 35–43.

Mandler, G., Goodman, G. O., & Wilkes-Gibbs, D. L. (1982). The word-frequency paradox in recognition. *Memory & Cognition*, *10*(1), 33–42. https://doi.org/10.3758/BF03197623

Marsh, R. L., Cook, G. I., & Hicks, J. L. (2006). The effect of context variability on source memory. *Memory & Cognition*, *34*(8), 1578–1586. https://doi.org/10.3758/BF03195921

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

Nelson, D. L., & McEvoy, C. L. (2000). What is this thing called frequency? *Memory & Cognition*, *28*(4), 509–522. https://doi.org/10.3758/BF03201241

Osth, A. F., & Dennis, S. (2015). Sources of Interference in Item and Associative Recognition Memory. *Psychological Review*. cmedm. http://kenli.nbu.bg:2048/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=cmedm&AN=25730734&site=eds-live

Osth, A. F., & Fox, J. (2019). Are associations formed across pairs? A test of learning by temporal contiguity in associative recognition. *Psychonomic Bulletin & Review*, *26*(5), 1650–1656. https://doi.org/10.3758/s13423-019-01616-7

Osth, A. F., Fox, J., McKague, M., Heathcote, A., & Dennis, S. (2018). The list strength effect in source memory: Data and a global matching model. *Journal of Memory and Language*, *103*, 91–113.

Ozubko, J. D., & Joordens, S. (2007). The mixed truth about frequency effects on free recall: Effects of study list composition. *Psychonomic Bulletin & Review*, *14*(5), 871–876. https://doi.org/10.3758/BF03194114

Page, M. P., Cumming, N., Norris, D., McNeil, A. M., & Hitch, G. J. (2013). Repetition-spacing and item-overlap effects in the Hebb repetition task. *Journal of Memory and Language*, *69*(4), 506–526.

Parmentier, F. B. R., Comesana, M., & Soares, A. P. (2017). Disentangling the effects of word frequency and contextual diversity on serial recall performance. *Quarterly Journal of Experimental Psychology*, *70*(1), 1–17. https://doi.org/10.1080/17470218.2015.1105268

Popov, V., Marevic, I., Rummel, J., & Reder, L. M. (2019). Forgetting Is a Feature, Not a Bug: Intentionally Forgetting Some Things Helps Us Remember Others by Freeing Up Working Memory Resources. *Psychological Science*, *30*(9), 1303–1317. https://doi.org/10.1177/0956797619859531

Popov, V., & Reder, L. (2020a). *Frequency effects in recognition and recall*. PsyArXiv. https://doi.org/10.31234/osf.io/xb8es

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

Reder, L. M., Donavos, D. K., & Erickson, M. A. (2002). Perceptual match effects in direct tests of memory: The role of contextual fan. *Memory & Cognition*, *30*(2), 312–323.

Reder, L. M., Liu, X. L., Keinath, A., & Popov, V. (2016). Building knowledge requires bricks, not sand: The critical role of familiar constituents in learning. *Psychonomic Bulletin & Review*, *23*(1), 271–277. https://doi.org/10.3758/s13423-015-0889-1

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember–know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(2), 294.

Reder, L. M., Paynter, C., Diana, R. A., Ngiam, J., & Dickison, D. (2007). Experience is a Double-Edged Sword: A Computational Model of The Encoding/Retrieval Trade-Off With Familiarity. In *Psychology of Learning and Motivation* (Vol. 48, pp. 271–312). Elsevier. http://linkinghub.elsevier.com/retrieve/pii/S0079742107480070

Serra, M., & Nairne, J. S. (1993). Design controversies and the generation effect: Support for an item-order hypothesis. *Memory & Cognition*, *21*(1), 34–40.

Smith, P. L., & Little, D. R. (2018). Small is beautiful: In defense of the small-N design. *Psychonomic Bulletin & Review*, *25*(6), 2083–2101. https://doi.org/10.3758/s13423-018-1451-8

Steyvers, M., & Malmberg, K. J. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 760–766. https://doi.org/10.1037/0278-7393.29.5.760

van Heuven, W. J. B., Mandera, P., Keuleers, E., & Brysbaert, M. (2014). SUBTLEX-UK: A new and improved word frequency database for British English. *The Quarterly Journal of Experimental Psychology*, *67*(6), 1176–1190. https://doi.org/10.1080/17470218.2013.850521

Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 11–11.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235. https://doi.org/10.1038/nature06860

## Appendix A: Mixture model comparisons

***Table A1:*** Relative AIC values for each mixture model fit for each participant. In each row, 0 represents the best fitting mixture model. The value in each cell reflects the difference between the fit of the model in the corresponding column relative to the best fitting model for that participant.

- **Model 1** = A single parameter von Mises distribution with no guessing or misbinding parameters
- **Model 2** = A two parameter mixture model of a central von Mises distribution, and a guessing uniform distribution (Zhang & Luck, 2008)
- **Model 3** = A two parameter mixture model of a central von M distribution plus four misbinding vM distributions but no uniform guessing
- **Model 4** = A three parameter mixture model with a central von Mises distribution, four misbinding von Mises distributions centered on non-target locations, and a uniform guessing distribution (Bays et al., 2009)
- **Model 5** = A four parameter mixture model similar to Model 4, but with separate precision parameters for correct and for misbinding responses

| Participant | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|
| 1 | 500.4 | 74.3 | **0** | 2 | 1.1 |
| 2 | 144.6 | 24.8 | 15.3 | **0** | 2 |
| 3 | 300.2 | 34.4 | 11 | **0** | 2 |
| 4 | 356.5 | 14.7 | 4.2 | **0** | 2 |
| 5 | 470.8 | 28.1 | 29.4 | **0** | 2 |
| 6 | 457.7 | 74.8 | 11 | **0** | 1.9 |
| 7 | 386.6 | 34.8 | **0** | 0.2 | 0.2 |
| 8 | 328.3 | 59.4 | 4.1 | 0.7 | **0** |
| 9 | 243.1 | 20.2 | 16 | **0** | 0.5 |
| 10 | 386.3 | 88.7 | 20.4 | **0** | 1.3 |
| 11 | 357.9 | 15.8 | 14.1 | **0** | 2 |
| 12 | 557.6 | 94.2 | 17.5 | 1.9 | **0** |
| 13 | 292.5 | 45 | 30.7 | **0** | 0.6 |
| 14 | 313.5 | 7.3 | 4.9 | **0** | 1.6 |
| 15 | 353.1 | 45.2 | 22 | 2.3 | **0** |
| 16 | 215.1 | 32.4 | 6.1 | **0** | 1.9 |
| 17 | 468.4 | 35.8 | 4.8 | 0.2 | **0** |
| 18 | 420.2 | 26.1 | 13.1 | **0** | 1.8 |
| 19 | 135.8 | 9.6 | 22.7 | **0** | 1.6 |
| 20 | 360.7 | 12.1 | 4.5 | **0** | 2 |
| 21 | 281.6 | 72.5 | 29.6 | **0** | 0.7 |
| 22 | 333.1 | 27.9 | 1.3 | **0** | 1.9 |
| 23 | 472.5 | 25 | 5.1 | 0.9 | **0** |

| Participant | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|
| 24 | 586.5 | 25.9 | 1.5 | **0** | 0.4 |
| 25 | 489.1 | 37 | **0** | 2 | 3.6 |
| 26 | 320.9 | 41.5 | 5.6 | **0** | 1.9 |
| 27 | 423.8 | 32.4 | 6.5 | **0** | 1.7 |
| 28 | 348.1 | 36.7 | **0** | 0.2 | 1.8 |
| 29 | 478.3 | 40 | **0** | 2 | 3.6 |
| 30 | 215.6 | 27 | 23.2 | **0** | 1.7 |
| 31 | 491.2 | 58.9 | **0** | 1.2 | 2.2 |
| 32 | 137.4 | 29.6 | 5.4 | **0** | 1.9 |
| 33 | 265.4 | 33.6 | **0** | 0.3 | 1.1 |
| 34 | 307.1 | 9.3 | 21.7 | **0** | 2 |
| 35 | 262.4 | 28.8 | 59.5 | **0** | 1.8 |
| 36 | 300.3 | 79.3 | 34.4 | **0** | 1 |
| 37 | 345.1 | 30.3 | 20.7 | 2.1 | **0** |
| 38 | 446.8 | 24.1 | 18.3 | **0** | 1.9 |
| 39 | 181.7 | 50.4 | 22 | **0** | 0.1 |
| 40 | 495.2 | 89.3 | 1.6 | **0** | 0.2 |
| 41 | 256.5 | 60.1 | 13.9 | **0** | 2 |
| 42 | 451.2 | 54.8 | 3.7 | **0** | **0** |
| 43 | 240.9 | 49.6 | 1.5 | 1.5 | **0** |
| 44 | 157.4 | 7.6 | 7.9 | **0** | 2 |
| 45 | 160.7 | 4.1 | 26.3 | **0** | 1.4 |
| 46 | 514.8 | 55 | 7.8 | **0** | 0.1 |
| 47 | 252.8 | 28.1 | 9.7 | **0** | 1.9 |
| 48 | 428.4 | 98 | 3.5 | **0** | 1.2 |
| 49 | 332.9 | 9 | 1.4 | **0** | 1.9 |
| 50 | 429.7 | 58.1 | 38.9 | **0** | 1.2 |
| 51 | 133.6 | 43 | 25.3 | **0** | 0.2 |
| 52 | 455.5 | 69.9 | 2.2 | 3.9 | **0** |
| 53 | 140.5 | 19.5 | 4 | 1.2 | **0** |
| 54 | 276.5 | 7.1 | 1 | **0** | 1.9 |
| 55 | 293.7 | 49.7 | **0** | 1.9 | 1.9 |
| 56 | 310.4 | 34.6 | 2.7 | 0.2 | **0** |
| 57 | 254.6 | 20.1 | 5.2 | **0** | 1.6 |
| 58 | 255.4 | 11.7 | 5 | 0.9 | **0** |
| 59 | 112 | 6.4 | 19.3 | **0** | 1.7 |
| 60 | 421.1 | 35.3 | **0** | 1.8 | 1.6 |
| 61 | 572.5 | 126.9 | 2.2 | 0.1 | **0** |
| 62 | 515.8 | 52.7 | 9.2 | 1.8 | **0** |
| 63 | 383.4 | 39.9 | 28 | **0** | 1.6 |
| 64 | 336 | 102.8 | 0.1 | 1.8 | **0** |
| 65 | 377.8 | 47.6 | 12.8 | 0.2 | **0** |

| Participant | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---|---|---|---|---|---|
| 66 | 206.1 | 30.5 | **0** | 1.5 | 3.5 |
| 67 | 250.3 | 47.5 | 10.6 | **0** | 0.9 |
| 68 | 378.3 | 132.5 | 29.8 | **0** | 0.6 |
| 69 | 215.9 | 15.3 | 11.9 | **0** | 0.7 |
| 70 | 409.1 | 53.5 | 2.4 | **0** | 1.3 |
| 71 | 421.9 | 38.8 | 11.2 | **0** | 2 |
| 72 | 406.2 | 25.9 | 10.7 | 0.8 | **0** |
| 73 | 215.7 | 46.2 | 0.5 | **0** | 1.9 |
| 74 | 293.4 | 32.5 | 11.4 | **0** | 0.5 |
| 75 | 438.2 | 56.4 | 7 | 3.7 | **0** |
| 76 | 176.1 | 17.1 | 14.7 | 1.6 | **0** |
| 77 | 300.2 | 52 | 37.6 | **0** | 1.6 |
| 78 | 375.2 | 67.9 | 8.1 | **0** | 2 |
| 79 | 352.4 | 39.5 | 8.1 | **0** | 0.8 |
| 80 | 374.4 | 32.4 | 1.1 | **0** | 0.7 |
| 81 | 423.9 | 38.2 | **0** | **0** | 1.3 |
| 82 | 296.6 | 34.5 | 5.1 | 0.9 | **0** |
| 83 | 247.5 | 17.5 | 24.7 | **0** | 1.6 |
| 84 | 367.2 | 12.6 | **0** | 1.7 | 3.5 |
| 85 | 88.3 | **0** | 5.5 | **0** | 2 |
| 86 | 284.8 | 18.1 | **0** | 1.8 | 3.2 |
| 87 | 347.7 | 43.4 | 37.7 | **0** | 0.3 |
| 88 | 496.8 | 44 | **0** | 2 | 3.9 |
| 89 | 285.6 | 33.3 | 38.7 | **0** | 0.6 |
| 90 | 357.6 | 10.4 | 49.5 | **0** | 2 |
| 91 | 465.2 | 66.6 | 18.1 | 1.6 | **0** |
| 92 | 330.9 | 29.8 | 39.3 | **0** | 1.9 |
| 93 | 409.7 | 45.3 | 21 | 3.7 | **0** |
| 94 | 196.1 | 25.2 | 14 | **0** | 1.6 |
| 95 | 244.9 | 27.3 | 5.8 | **0** | 1.6 |
| 96 | 260.7 | 72.9 | **0** | 0.2 | 2.1 |
| 97 | 208.8 | 12.2 | 5.4 | 0.1 | **0** |
| 98 | 252 | 33.7 | 1.9 | **0** | 1.9 |
| 99 | 247.5 | 10.3 | 6 | **0** | **0** |
| 100 | 397.5 | 71.3 | 6.2 | **0** | 0.7 |
| 101 | 325.4 | 46.8 | 13.5 | **0** | 1.9 |
| 102 | 271.4 | 24.3 | 22.5 | **0** | 1.5 |
| 103 | 283.5 | 21.1 | 0.9 | **0** | 2 |
| 104 | 485.1 | 18.9 | 1.5 | 1.7 | **0** |
| 105 | 418.6 | 56 | 9.2 | **0** | 1.6 |
| 106 | 426.8 | 77.5 | 6.2 | **0** | 2 |
| Proportion best fitting | 0 % | 1 % | 13 % | **64 %** | 22 % |