Syracuse University

SURFACE

Theses - ALL

8-25-2017

THE SOURCE OF LIST STRENGTH EFFECT WITHIN THE RETRIEVING EFFECTIVELY FROM MEMORY FRAMEWORK: THE LEVEL OF COMPETITION

YANXIU CHEN
Syracuse University

Abstract

To study episodic memory in a laboratory, we study interference within a list of stimuli. With list strength paradigm, we study how such interference is affected by how well stimuli are encoded, and the encoding strength of other items in the list. A stimulus can be weak or strong, and it can be in a pure list, composed of all weak or all strong stimuli, or a mixed list, composed of both weak and strong stimuli. A list strength effect (LSE) refers to the interaction between stimulus strength and list type. In free recall, where the cue used at test is only context, we have consistently observed a LSE. Yet, in cued recall, where the cue is made of item, we have consistently observed a null LSE. Thus, we attributed the source of LSE to the type of cue used at retrieval. Based on REM, this framing is potentially misleading. It is not the type of cue (context or item) that is critical, but the level of competition. Typical experiments have confounded these two factors because they have manipulated item to create a low level of competition within a list, while context to create a high level of competition. Therefore, in this study, we manipulated the level of competition (between-subject) and the type of cue (within-subject) simultaneously on the list strength paradigm. Data shows LSE was determined by the level of competition, not the type of cue probing memory at test. Fitting REM to the data confirms this statement. Nevertheless, whether memory was cued with context or item affected recall performance differently.

THE SOURCE OF LIST STRENGTH EFFECT WITHIN THE RETRIEVING EFFECTIVELY FROM MEMORY FRAMEWORK: THE LEVEL OF COMPETITION

by

Yanxiu Chen

B.A., Applied Psychology, Guangzhou Chinese Medicine University, 2015

Thesis

Submitted in partial fulfilment of the requirements towards a Master of Science in Experimental Psychology

Syracuse University

August, 2017

Table of Contents

List of Tables

List of Figures

Introduction

Episodic memory refers to memory of the experiences in our lives. The encoding, storage, and retrieval of these experiences define us. Thus, it’s important to study what and how manipulations help and harm these processes. Episodic memories have item and context information (McGaugh, 1932; Davachi, 2006; Rugg et al., 2012). For example, in a memory of giving an academic talk, item information is the content of this talk, whereas context information includes the size of the conference room where the talk happened, the lighting of the room, etc. In a lab, we typically investigate episodic memory by having people study words (item information) in a list in an experimental booth (context information).

In a list strength paradigm, we manipulate stimulus strength and list type. Stimulus strength can be weak or strong, and is manipulated through repetition, study time, or depth of processing (Craik & Lockhart, 1972; Hintzman, 1974; Malmberg & Shiffrin, 2005; Wilson & Criss, 2017). List type can be pure or mixed in strength. A pure list consists of stimuli with the same strength, whereas a mixed list consists of stimuli of different strengths. From such manipulations, we have three lists, pure weak, pure strong, and mixed list. We also have four kinds of stimuli in terms of list and strength, pure weak, pure strong, mixed weak, and mixed strong stimuli. With this design, we can investigate whether memory of stimuli with the same strength in a pure list is as well as in that in a mixed list. Specifically, we compare memory of mixed strong verses pure strong stimuli, and mixed weak verses pure weak stimuli to ask whether memory for a given item is affected by the strength of the other items encoded in the same context.

Studying the list strength paradigm is theoretically important because different results are observed when applying this paradigm in different tasks. In free recall, we ask participants to generate words from a studied list. Consistently, researchers have found that memory of mixed strong is better than that of pure strong, whereas memory of pure weak is better than that of mixed weak (Tulving and Hastie, 1972; Wixted, Ghadisha, & Vera, 1997; Malmberg & Shiffrin, 2005). We term this pattern of data a (positive) list strength effect (LSE). One explanation is that items compete to be remembered and strong stimuli outcompete weak stimuli in that process. Many memory models (global matching models, Humphreys, Pike, Bain, & Tehan, 1989), including the initial version of the Search of Associative Memory model (SAM, Raaijmakers & Shiffrin, 1981; Gillund & Shiffrin, 1984), and the Theory of Distributed Associative Memory model (Murdock, 1982), predict a positive list strength effect, not only in free recall, but also cued recall.

However, in cued recall, observations contradict predictions of these models. In a typical cued recall experiment, participants study pairs of words and one word is presented as a cue to retrieve the item it was studied with. We have found memory of mixed strong is as good as that of pure strong, and memory of pure weak is as bad as that of mixed weak (Ratcliff, Clark and Shiffrin, 1990; Shiffrin, Ratcliff, Murnane, & Nobel, 1993; Murdock & Kahana, 1993a, 1993b; Wilson & Criss, 2017). We term this pattern of results a null LSE. To make sense of these findings, Shiffrin, Ratcliff, & Clark (1990) added a differentiation mechanism to SAM. Later, Shiffrin & Steyvers (1997) developed the Retrieving Effectively from Memory model (REM) with differentiation as its core mechanism. Differentiation refers to the idea that, the more a stimulus is studied, the easier for participants to separate it from other studied stimuli. Based on differentiation, participants can remember strong stimuli better than weak stimuli because of their enhanced clarity resulting from strengthening. This produces similar memory performances between mixed strong and pure strong, as well as between mixed weak and pure weak, or, a null LSE.

Essentially, REM can predict both a positive LSE and a null LSE depending on the cue provided and the degree of competition that cue elicits in the different lists. Every stimulus has one trace, including context and item encoded and stored in episodic memory (Murdock, 1982; Murnane, Malmberg & Phelps, 1999; Criss & Shiffrin, 2004b). The consequence of strengthening a stimulus is that this ensemble becomes less confusable compared to others (i.e., differentiation, see Criss, 2006; Criss & Koop, 2015; Criss & McClelland, 2006; Murnane & Shiffrin, 1991a, 1991b). At retrieval, sampling is initiated by a cue, which is compared to the corresponding part of every stored ensemble. A single trace is sampled and the probably of sampling is in proportion to how well it matches the cue. Therefore, when this cue is shared by more than one stimuli, there is a higher level of competition at sampling. In this case, mixed strong ensembles outcompete mixed weak ones, and have a higher chance of getting sampled. This results in a positive LSE. Free recall is an example where the cue (list context) is shared by all stimuli on the list in a typical experimental design. However, when this cue is uniquely associated with one stimulus, there is a low level of competition at sampling. In this case, little competition between mixed strong and mixed weak ensembles exist. This alone would result in a null LSE. In short, based on REM, the level of competition is the source of LSE. Cued recall is a mixed situation because every cue word was studied with a single target word but the context is shared. There is a null LSE in cued recall (Wilson & Criss, 2017).

However, this competition-based explanation is not the explanation for the null LSE in cued recall reported in the literature. Instead, the seminal LSE papers (i.e., Ratcliff, Clark & Shiffrin, 1990; Shiffrin, Ratcliff and Clark, 1990) claim that the difference between free recall and cued recall is the amount of weight placed on context as a cue. Free recall uses context as the only cue and a positive LSE is observed. Cued recall places less emphasis on context and more on the item cue, and a null LSE is observed. Note that in typical studies, cue presented at retrieval is always confounded with the level of competition it induced. Take studying and testing on a list of word pairs as an example. Typically, this list is designed in a way that they are presented in the same room with the same colors, font sizes, and locations of the screen, in other words the context is shared for all items (Ratcliff, Clark and Shiffrin, 1990; Murnane Phelps, 1993, 1994 & 1995). In free recall, context is the only cue. In cued recall, the context is again shared but every word pair has a single cue and target and the cue is both a word from the pair and context. Therefore, cue type and level of competition is confounded. In free recall, where context is the one and only cue, a high level of competition yields a positive LSE. In cued recall, where the item is part of the cue, a lower level of competition yields a null LSE.

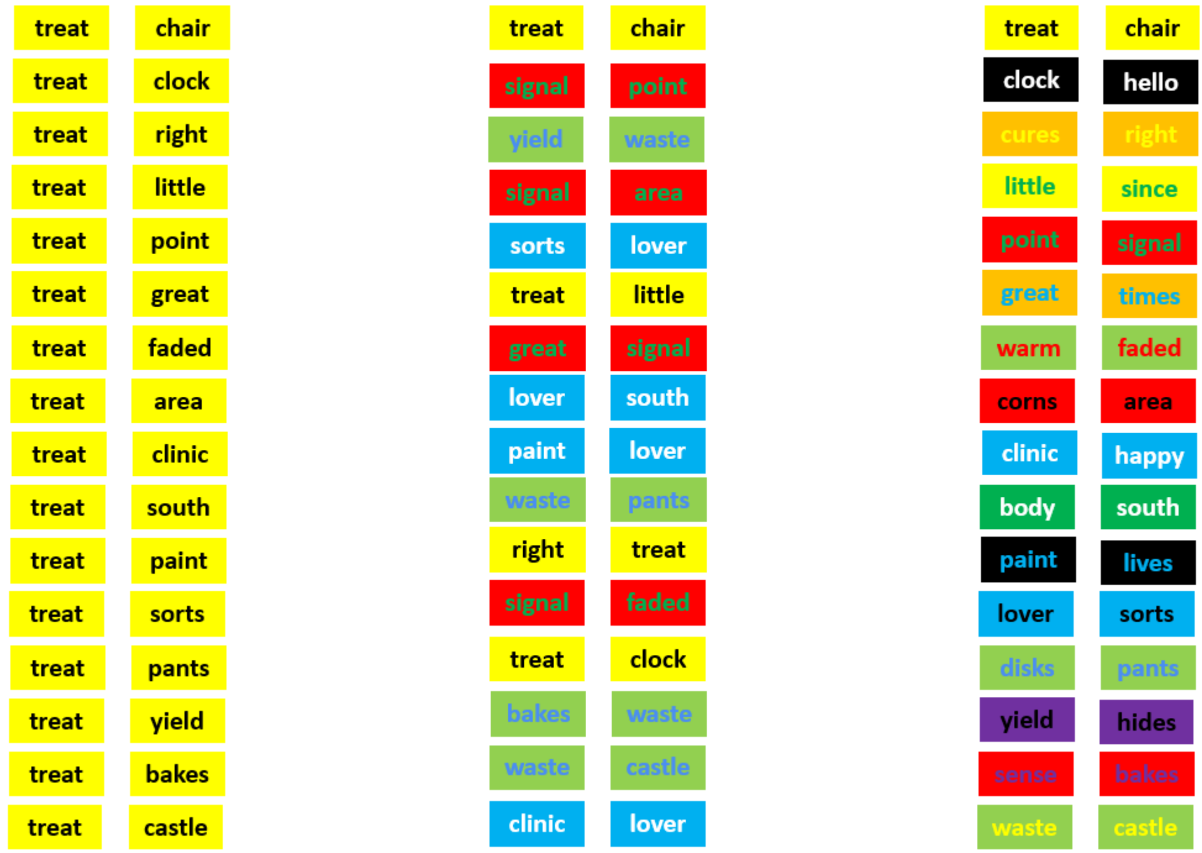
In this project, we evaluate whether cueing with item or context, and whether the level of competition is critical to the magnitude and direction of the LSE. We manipulated the level of competition at 3 levels - high, middle, or low, by varying the number of cues shared by a list of words. We manipulated the type of cue, context or item, by either presenting a context cue or an item cue at test. We expected to find the magnitude and direction of observed LSEs change across levels of competition, but not across types of cue used at test.

Experiment

Design

The experiment was a 2*3*2*2 design. We manipulated the type of cue presented at test (context/ item), and the level of competition at test (high/middle/low), in a list (pure/ mixed) strength (weak/ strong) paradigm. Level of competition was a between-subject variable. All other variables were within-subject variables, resulting in 6 study-distractor-test cycles per participant.

We attempted to reduce interference across list by presenting each list in a unique location on the monitor. Each list was randomly assigned without replacement to one of six positions (top left, top center, top right, bottom left, bottom center, and bottom right) on the screen where all word pairs from this list were presented for study. Each study list included 16 pairs of words presented in a colored font and color-filled box for 1.5 seconds (see Figure 1 for examples). The participants engaged in an encoding task where they judged the degree of association between the two items using a scale of 1 to 5 (1 being no association at all, 5 being highly associated), typed their response, and then hit ‘enter’. The participants could take as long they need to respond. A blank screen separated word pairs for an interval of 0.1 second following the response.



High Conditions **Middle Conditions** **Low Conditions**

Figure 1: Examples of study lists in high, middle, and low conditions.

One independent variable is the level of competition at test. We varied, on every study list, how many cues are shared at 3 levels ranging from all items having a shared cue to no items having a shared cue as illustrated in Figure 1. In the high conditions, all target words were assigned the same cue word, and were presented under the same context. In the middle conditions, a quarter of the word pairs within a list were assigned the same cue word and were presented in the same context. In the low conditions, all target words have a unique cue word and context. The high, middle, and low conditions should produce, respectively, a high, middle, and low level of competition at test.

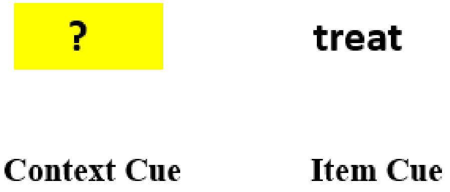


Figure 2: An example of context and item cue presentations at test.

Another independent variable was whether the cue presented at test was context or item (see Figure 2 for an example). An item cue was a word which is typical in studies of cued recall. A context cue was the color combination in which the target word was presented. We make the common assumption that semantic, phonological features of words are item features, and background color and word color are contexts (e.g., Murnane & Phelps, 1993, 1994 & 1995). We randomly selected items from a pool of 800 high frequency words of letter length 4 to 11. We used the RGB triplet to specify colors. In MATLAB, the RGB triplet is a three-element row vector, whose elements range from 0 to 1 and specify the intensities of the red, green, and blue components of the color. Though there exists infinite number of colors, to make sure colors are easy to identify for participants, we used either 0 or 1 for any of three elements. Also, since in a typical experiment design, white background color and black word color are used, we exclude the combination of white background color with any word color. This gave us 49 combinations of background colors and word colors.

The third and fourth independent variables make up the list strength paradigm. A study list can be pure or mixed with respect to stimulus strength and a stimulus strength can be weak or strong. In a pure list, all stimuli had the same strength, while in a mixed list, half of the stimuli were weak, the other half were strong. Stimulus strength was manipulated by the number of repetitions. In the pure weak list, 16 word pairs were presented once. In the pure strong list, 16 word pairs were presented four times, with a full set of pairs presented before any repeat (Ratcliff, Clark & Shiffrin, 1990; Malmberg & Shiffrin, 2005). For each of the four presentations, both left-right order and serial order of the 16 stimuli were randomized anew for each cycle of repetition. In the mixed list, 8 strong word pairs were presented three times, with both their left-right order and serial order randomized. Then all word pairs were randomly intermixed and presented once. This construction was to ensure that, across all blocks, the lag between the final study presentation and the test position is the same.

Procedure

Prior to study participants were informed that this was a memory task and that they'd be presented with either words or colors as retrieval cues. After studying each list, participants entered the distractor stage, which lasted for 1 minute. The participants were asked to complete a series of mathematical problems. During the distractor task, two single-digit numbers were randomly chosen and presented for 1.5 second. After the presentation of each number pair, the participants were given as much time as they need to add these two digits, and enter their response.

Retrieval was cued with either a context or an item and participants were instructed to recall all words studied with this cue. The cue and response box were centered on the screen. Responses were recorded as participants type and press enter after each word. Participants clicked a box labeled 'finished' when they have finished recalling. A timer indicated how much time had elapsed and the finished button was only active after 48, 16, and 4 seconds in, respectively, high, middle, and low conditions (or 4 seconds per word associated with the cue).

The order of list, order of items within a list, screen location, and context-word and word-word pairing were randomly assigned for each participant. The study was conducted in Matlab with Psychtoolbox (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007).

Measurement

All responses were recorded. The primary interest is accuracy. An accurate response is be a word studied with the cue presented at test. However, for any list, the number of potential correct responses under context cueing is larger than that under item cueing. For instance, in the middle condition the number of potential correct responses is 5 when cued with context and 4 when cued with item (e.g., the item cue is removed from the set of potential answers), see Figure 1. In addition, cue words in the high and middle conditions repeat multiple times. For instance, in the example middle condition in Figure 1, cue word 'treat' is presented four times. To create a fair comparison, here, we considered only target words in the calculation of accuracy with a common denominator of 16. Later we consider alternative approaches.

In addition, to directly measure the magnitude and direction of a LSE, we report Difference of Differences scores (***DoD***). In Equation 1, ***MS***, ***MW***, ***PS***, and ***PW*** stand for, respectively, accuracies of mixed strong, mixed weak, pure strong, and pure weak of a given list. Every ***DoD*** score is between -2 and 2. A ***DoD*** score bigger than 0 indicates a positive list strength effect. A ***DoD*** score less than 0 indicates a negative list strength effect. And a ***DoD*** score close to 0 indicates a null list strength effect.

DoD = (MS-MW) – (PS-PW) (E1)

This study was pre-registered at AsPredicted https://aspredicted.org/blind.php/?x=gb6yx4, see Appendix A for the text.

Simulation-based Predictions

Before collecting any data, we generated predictions from REM (see Figure 3), a version in which we consider context and item as being identical in nature, and encoded equally well by participants. Details regarding implementation are illustrated later. Corresponding to the competition-based explanation, the magnitude and direction of observed LSEs change across levels of competition, but not across types of cue used at test.

Analysis Plan

The analyses are those that we pre-registered (https://aspredicted.org/blind.php/?x=gb6yx4) (see Appendix A).

We planned to conduct a 2 (pure/ mixed) *2 (weak/ strong) *2 (context cue/ item cue) *3 (high/ middle/ low) analysis of variance (ANOVA) on accuracy first. From this analysis, we simply intend to evaluate whether the strengthening manipulation was effective. We expected to find better memory for strong than weak stimuli.

Then, for our primary purpose, we planned to calculate ***DoD*** scores of high-context, middle-context, low-context, high-item, middle-item, and low-item conditions, and conduct a 2 (context cue/ item cue) *3 (high/ middle/ low) ANOVA on ***DoD*** scores. Based on pre-experiment predictions from REM (see Figure 3), we expected to find the DoD decrease from the high to middle to low conditions, but not across context cue and item cue conditions.

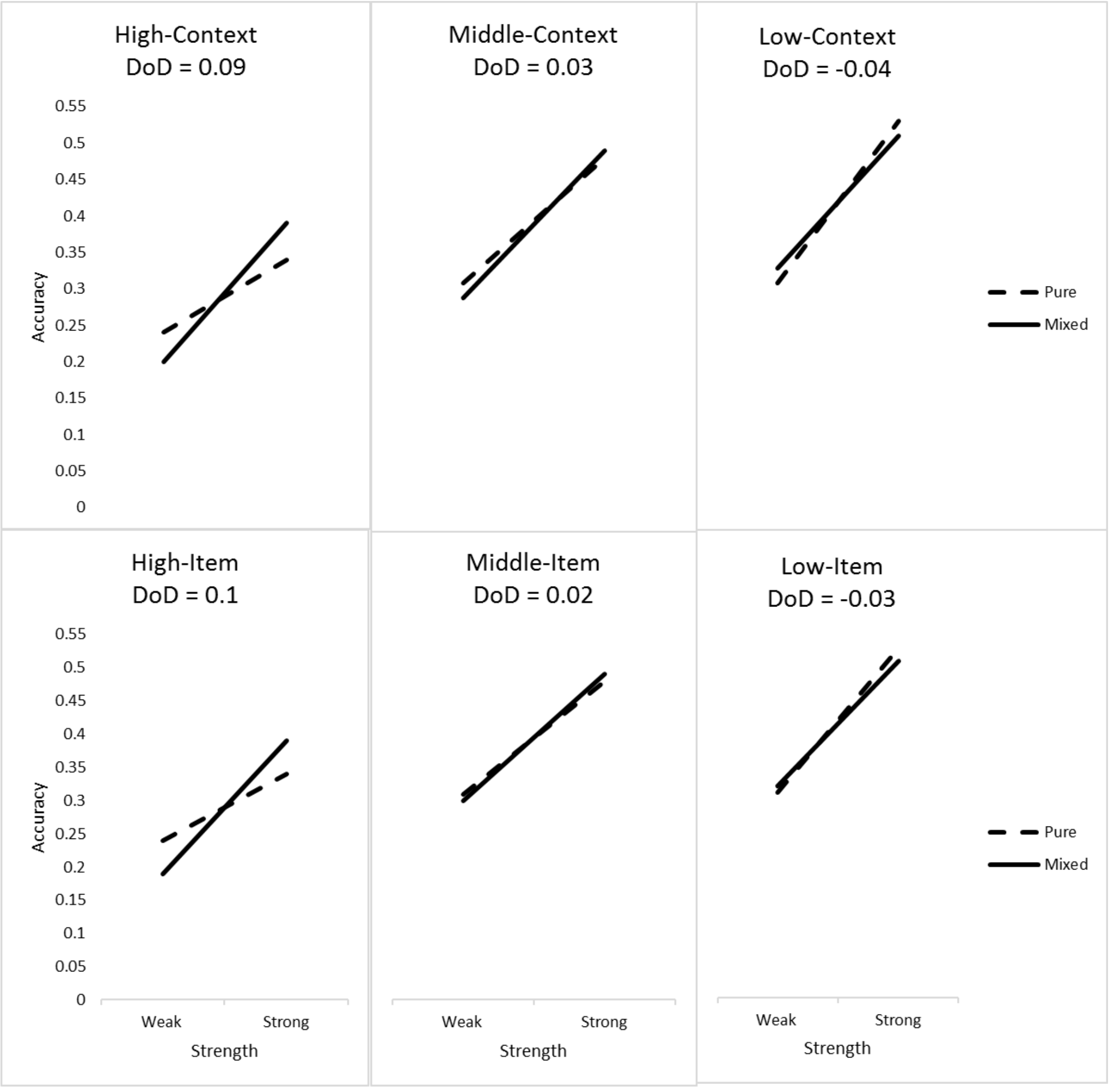


Figure 3: Simulation-based predictions of accuracies and Difference of Differences scores (***DoD***) of all conditions.

We employ both frequentist and Bayesian ANOVAs on accuracy and ***DoD*** scores. From the latter, with the Bayes Factor ($\boldsymbol{BF_{10}}$), we report a continuous value of the evidence favoring one model (a model with an effect) over another one (a null model), rather than drawing dichotomous inference from the data (see Etz & Vandekerckhove, in press, Morey, 2015). For instance, a $\boldsymbol{BF_{10}}$ =100 indicate the data are 100 times more likely from the model of effect compared to the model with no effect (see Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010). Analyses were conducted in JASP (Love et al, 2015; JASP Team, 2016) with default priors.

Finally, we fit REM to the data. Specifically, we planned to estimate REM parameters from the data, and use these parameter values to generate post-experiment predictions, to which we can compare the data.

Subjects

We recruited 180 participants in total, with 60 participants each for shared, middle, and unique conditions. The sample size was based on our previous studies of the list strength effect in cued recall (Wilson & Criss, 2017). Due to the manipulation of colors in our experiment, all participants reported that they could detect differences among colors and were not color-blind. Data from 4 participants of the middle, and 13 of the unique condition were excluded because they gave no correct response in any condition.

Results and Discussion

Data Analysis

Based on preregistration, we calculated accuracies (see Table 1, Figure 4), and ***DoD*** scores (see Figure 4, Figure 5) of all conditions. Then, we conducted 2 (pure/ mixed) *2 (weak/ strong) *2 (context cue/ item cue) *3 (high/ middle/ low) repeated measure ANOVA on accuracies of all conditions. As expected, a main effect from stimulus strength was found, F(1,159) = 220.887, $p < 0.001$, BF_{10} = 1.812e + 35.

Table 1

Means and standard errors of the mean of accuracies by level of competition, type of cue, list type, and stimulus strength

		List Type			
Levels of Competition	Type of Cue	Pure		Mixed	
		Stimulus Strength		Stimulus Strength	
		Weak	Strong	Weak	Strong
High	Context	0.20 (0.02)	0.31 (0.03)	0.17 (0.02)	0.42 (0.03)
	Item	0.23 (0.03)	0.38 (0.03)	0.16 (0.02)	0.41 (0.04)
Middle	Context	0.06(0.01)	0.14 (0.03)	0.09 (0.02)	0.18 (0.03)
	Item	0.15 (0.02)	0.32 (0.03)	0.20 (0.02)	0.35 (0.03)
Low	Context	0.013 (0.01)	0.056 (0.01)	0.016(0.01)	0.072 (0.02)
	Item	0.28 (0.03)	0.48 (0.05)	0.24 (0.034	0.42 (0.05)

For our primary purpose, we conducted a mixed ANOVA on ***DoD*** scores, with level of competition (high/ middle/ low) as a between-subject factor and cue as within-subject factor (context / item).

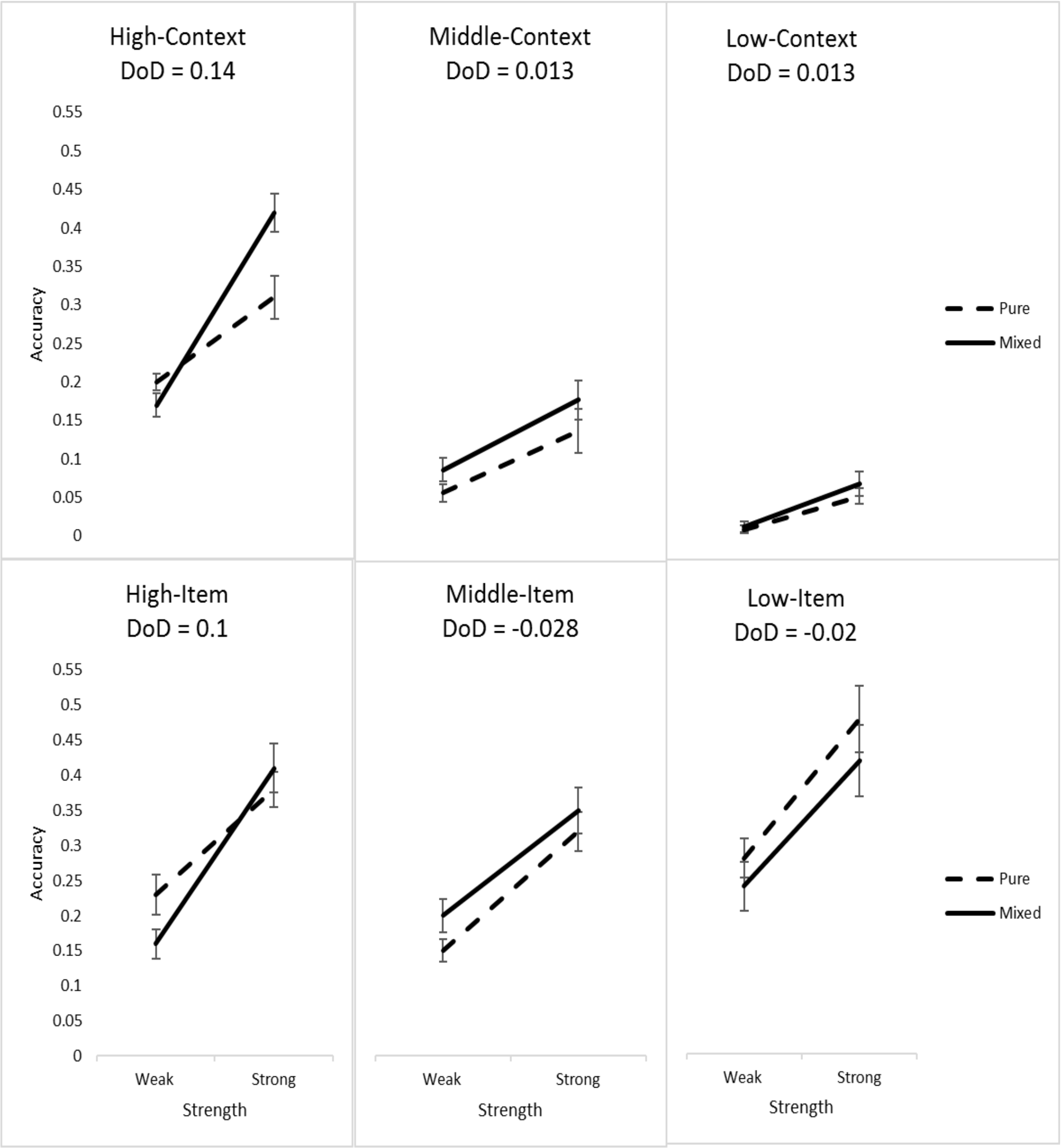


Figure 4: Data of accuracies and Difference of Differences scores (***DoD***) of all conditions.

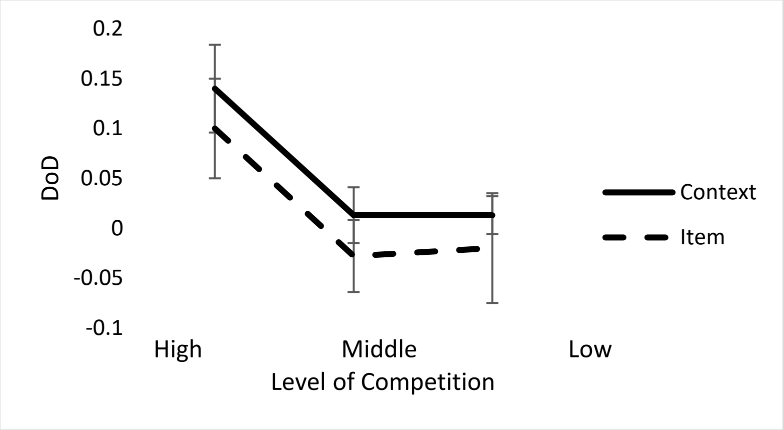


Figure 5: Difference of Differences scores (***DoD***) of high, middle, and low context conditions, as well as high, middle, and low item conditions.

As expected, the ANOVAs on ***DoD*** scores revealed no main effect from cue type, F (1,159) = 1.635, p = 0.203, BF_{10} = 0.287, whereas a main effect from the level of competition was found, F (2, 171) = 5.752, p = 0.004, BF_{10} = 6.118. No interaction between cue type and levels of competition was found, F (2, 159) = 0.011, p = 0.989, BF_{10} = 0.064.

In summary, from analysis, the critical findings match our prior expectations. (1) The detection of a list strength effect depends on the level of competition at test, which is consistent with pre-experiment predictions generated from REM (Figure 3) (Shiffrin & Steyvers, 1997). (2) A list strength effect doesn't depend on whether the presented cue at test is context or item, which contradicts assertions in prior studies (Ratcliff, Clark & Shiffrin, 1990; Shiffrin, Ratcliff & Clark, 1990), but is consistent with REM. Overall, these findings confirm our hypothesis that the source of list strength effect is not the type cue used at test, but the level of competition.

However, two observations are inconsistent with our expectations. First, the middle and low conditions are almost identical in terms of ***DoD***. Second, accuracy when cued by context is poor for the middle and low conditions. Since they are not our primary interests in this project and did not qualitatively alter our experimental expectations, we cease discussing them for now. Nevertheless, we provided our countered explanations and solutions in Discussion.

Finally, to qualitatively account for the pattern of the data, we fit REM with our data. We estimated REM parameters from data of some conditions, and generated post-experiment predictions.

Model Fitting

REM was designed and has been used to tackle conventional memory paradigms, such as free recall and cued recall. In this project, the high-context is traditional free recall and the low-item condition is traditional cued recall. Therefore, we estimated all parameters from the high-context and/or low-item conditions and used those parameters to fit the remaining conditions.

In this segment, we briefly review a computational description of REM, how REM was implemented in previous related research, and how we implemented in ours.

Representation

In REM, episodic memory is composed of individual memory traces. Every memory trace is represented as a vector of feature values. Features can be item features or context features. Item features refer to the semantic aspects of this memory trace, and context features refer to the surrounding context of the event. The number of features was fixed at 20 per item and 20 for context (as in Shiffrin & Steyvers, 1997). Feature values are drawn randomly from a geometric distribution with a parameter $\boldsymbol{g}$.

$$\boldsymbol{P(v=i)=(1-g)^{i-1}g, where\ i=1, \dots \dots\ \infty\ \ (E2)}$$

Equation 2 is used to generate features and shows the probability that any feature value $\boldsymbol{v}$ in a stimulus is assigned the value $\boldsymbol{i}$. Every feature value of every stimulus, context or item, is generated independently based on Equation 1. One way to understand parameter $\boldsymbol{g}$ is that it represents environmental frequency – features with small values are common and features with large values are uncommon. Or, broadly speaking, the nature of the stimulus. For instance, with a high $\boldsymbol{g}$ value of item, we can generate stimuli representing high frequency words, whereas with a small $\boldsymbol{g}$ value of item, we can generate stimulus representing low frequency words. In addition, $\boldsymbol{g_{sys}}$ is the system's long-term estimate of $\boldsymbol{g}$ used in the decision rule. The $\boldsymbol{g}$ from which a stimulus is generated may match or may differ from $\boldsymbol{g_{sys.}}$

When implementing REM, researchers have set geometric parameter of context ($\boldsymbol{g_c}$) and that of item ($\boldsymbol{g_i}$) to be the same, as we did for pre-experiment simulations in this project for simplicity reason, although there is reason to assume different values for item and context features (see REM4, Shiffrin & Stevyers, 1997). In our model fitting, for geometric parameter of item features, we gave $\boldsymbol{g_i}$ the value of 0.45 to simulate the usage of high frequency words in our experiment and $\boldsymbol{g_{sys}}$ was set as 0.4 (following Shiffrin & Stevyers, 1997). We estimated $\boldsymbol{g_c}$ for context because we used color and have no prior knowledge about what that parameter value should be.

Storage and Encoding

In REM, a memory trace is most likely imperfect. Every feature value of every memory trace can be stored with the wrong or correct value or it can be not stored (represented by the value 0 in the simulations). When an item is presented for the first time, every feature value is stored with some probability $\boldsymbol{u^*}$ and encoded independently. For those that are stored, every feature value can be correctly encoded with a probability of $\boldsymbol{c}$. If a feature value is incorrectly encoded, it takes a random value chosen from the geometric distribution with the parameter $\boldsymbol{g_{sys}}$. It's possible that this randomly assigned feature value matches to the actual value by chance. Once a feature is assigned a value, correct or not, it is not updated with further learning during the event. In the full model, repetitions result in storing some of the remaining feature values that are zero. For simplicity, we use different **u*** values for strong and weak conditions rather than implement updating, or multiple number of storage attempts as in Malmberg & Shiffrin (2005).

In our modeling fitting, we fixed $\boldsymbol{c}$ as 0.7 (Malmberg & Shiffrin, 2005), whereas we estimated the probabilities of storing a feature, in weak and strong stimulus strength manipulations, of context, $\boldsymbol{u_{wc}}$, and $\boldsymbol{u_{sc}}$, and item, $\boldsymbol{u_{wi}}$, and $\boldsymbol{u_{si}}$, from the data. It's important to emphasize that, to constrain the model fits as much as possible, we estimated the context encoding strength parameters from the high-context condition (i.e., free recall) and the item encoding strength parameters from the low-item conditions (i.e., cued recall).

Retrieval

Based on REM, retrieval is always initiated by probing memory with a cue. The nature of the cue depends on the retrieval task. In this project, under context cuing, the cue is a vector of context features, and, under item cuing, the cue is a vector of item features.

When a cue, $\boldsymbol{q}$, is presented, it is compared to every stored memory trace indexed by $\boldsymbol{j}$ and based on Equation 3, a likelihood ratio $\boldsymbol{\lambda}$ is calculated indicating the match between $\boldsymbol{q}$ and $\boldsymbol{j}$. The likelihood ratio is computed in the spirit of the probability of this cue and this memory trace match given this memory trace was constructed from this cue, $\boldsymbol{P}(\boldsymbol{Match}|\boldsymbol{Y_{qj}})$, over the probability of this cue and this memory trace match given this memory trace was not constructed from this cue, $\boldsymbol{P(Match|N_{qj})}$. In the equation 3, $\boldsymbol{n_{qj}}$ is the number of nonzero feature values that mismatch, $\boldsymbol{n_{ijm}}$ is the number of matching occurrences when the feature has the value $\boldsymbol{i}$.

$$\boldsymbol{\lambda_j = \frac{P(Match|Y_{qj})}{P(Match|N_{qj})} = (1-c)^{n_{ij}} \prod_{\forall i}^{\infty} \left[\frac{c+(1-c)g_{sys}(1-g_{sys})^{i-1}}{g_{sys}(1-g_{sys})^{i-1}}\right]^{n_{ijm}} \quad (E3)}$$

In cued recall and free recall, likelihood ratios are used to compute the probability of sampling a single memory trace, based on Equation 4.

$$\boldsymbol{P(S_j|q) = \frac{\lambda_j^y}{\sum \lambda_k^y} \ (E4)}$$

where $\boldsymbol{P(S_j|q)}$ is the probability of sampling memory trace $\boldsymbol{j}$ when presented cue $\boldsymbol{q}$, $\boldsymbol{\lambda_j}$ is the likelihood ratio of memory trace $\boldsymbol{j}$ computed from Equation 3, $\boldsymbol{y}$ is the scaling parameter, and $\sum\boldsymbol{\lambda_k}^y$ is the sum of likelihood ratio across all $\boldsymbol{k}$ memory traces after scaling. The sampling probability of one memory trace is positively correlated to the relative size of the likelihood ratio of this memory trace among all traces. The full model includes a threshold for sampling. Specifically, the likelihood ratio of a memory trace needs to exceed a threshold value, $\boldsymbol{\phi}$, to be sampled, and bestowed the chances of being recovered. Typically, the system attempts to recover every trace, no matter how poor the match (e.g., $\boldsymbol{\phi} = 0$). However, we will explore the necessity of this parameter for conditions where memory accuracy is very poor in the Appendix B.

Following sampling, we have recovery of a memory trace. After a memory trace has been sampled, the probability of recovery (i.e., reporting the contents of the memory trace) is calculated based on Equation 5.

$$\boldsymbol{P(R_j|q) = \rho^{\tau}}\ (\boldsymbol{E5})$$

where $\boldsymbol{P(R_j|q)}$ is the probability of recovering a sampled memory trace, $\boldsymbol{\rho}$ is the proportion of item features from the to-be-retrieved target that are stored correctly, and $\boldsymbol{\tau}$ is a scaling parameter. Finally, in recall, multiple attempts at retrieval are made with the same cue and a parameter is needed to decide when to stop searching. Search continues until the number of output failures hits a limit ($\boldsymbol{K_{max}}$). Output failures include (1) failing to exceed the sampling threshold, (2) recovery failure, (3) successful recovery of a response already given, (4) successful sampling of a response that underwent recovery failure(s) for the same cue. Parameters $\boldsymbol{y}$, and $\boldsymbol{\tau}$ are fixed as, respectively, 0.2, 0.5 (following Malmberg & Shiffrin, 2005). Parameter $\boldsymbol{K_{max}}$ is also fixed so as to make it unimportant. We set $\boldsymbol{K_{max}}$ equal to the number of possible correct responses, that is, the values of $\boldsymbol{K_{max}}$ of high-context, middle-context, low-context, high-item, middle-item, and low-item are, respectively, 17, 5, 2, 16, 4, and 1.

To reiterate our approach, we fixed as many parameters as possible to standard values. We first estimated item-related parameters, $\boldsymbol{u_{wi}}$ and $\boldsymbol{u_{si}}$, from the low-item condition. Then we fed those along with the fixed parameters to the high-context cue condition and estimated context-related parameters, $\boldsymbol{g_c}$, $\boldsymbol{u_{wc}}$, and $\boldsymbol{u_{sc}}$. Parameters were estimated in R studio, via Maximum Likelihood (ML) estimation (Akaike, 1973; Eliason, 1993). ML estimation allows us to identify one specific model (i.e., REM) with a combination of parameters, which has the maximum probability of observing the data. We used $\boldsymbol{G^2}$, also known as log-likelihood ratio (Cochran, 1952), as ML estimator (see Equation 6), calculating the difference between data ($\boldsymbol{D}$) and model prediction ($\boldsymbol{P}$). This gave us a completed set of REM parameters (see Table 2).

$$\boldsymbol{G^2 = 2 * \sum (D * (\log D - \log P))}\ (\boldsymbol{E6})$$

Table 2 summarizes all parameters.

Table 2

REM parameters and their values

Parameter Name	Parameter Meaning	Parameter Value	Fixed or Estimated
$\boldsymbol{N_f}$	Number of features per context/item	20	Fixed
$\boldsymbol{g_i}$	Geometric parameter to generate item features	0.45	Fixed
$\boldsymbol{g_c}$	Geometric parameter to generate context features	0.45	Estimated from high-context
$\boldsymbol{u_{wi}}$	Probability of storing an item feature value for weak stimulus	0.38	Estimated from low-item
$\boldsymbol{u_{si}}$	Probability of storing an item feature value for strong stimulus	0.56	Estimated from low-item
$\boldsymbol{u_{wc}}$	Probability of storing a context feature value for weak stimulus	0.12	Estimated from high-context
$\boldsymbol{u_{sc}}$	Probability of storing a context feature value for strong stimulus	0.48	Estimated from high-context
$\boldsymbol{c}$	Probability of encoding a feature value for any stimulus	0.7	Fixed
$\boldsymbol{g_{sys}}$	Geometric parameter when calculating likelihood ratio	0.4	Fixed
$\boldsymbol{y}$	Scaling parameter of likelihood ratio	0.2	Fixed
$\boldsymbol{\tau}$	Scaling parameter of proportion when calculating probability of recovery	0.5	Fixed
$\boldsymbol{K_{max}}$	Number of failures before retrieval stops	Number of possible correct responses	Fixed but varied among conditions
$\boldsymbol{\phi}$	Sampling threshold	0	Fixed

We then used them to generate post-experiment predictions, with 1000 simulated subjects (see Figure 6 and Figure 7). Overall, the patterns of data were qualitatively accounted for by REM. Based on ***DoD*** scores of Figure 6 and Figure 7, it's clear that the magnitude and direction of observed LSEs change across levels of competition, not the type of cue, as do the data. This confirms out hypothesis based on REM, that the level of competition is the source of LSE, but the type of cue used at retrieval task might not be the causal factor. In these fits we did not attempt to account for the similarity of the DoD values for the middle and low conditions nor did we attempt to account for the poor performance for context cuing especially in the middle and low conditions.

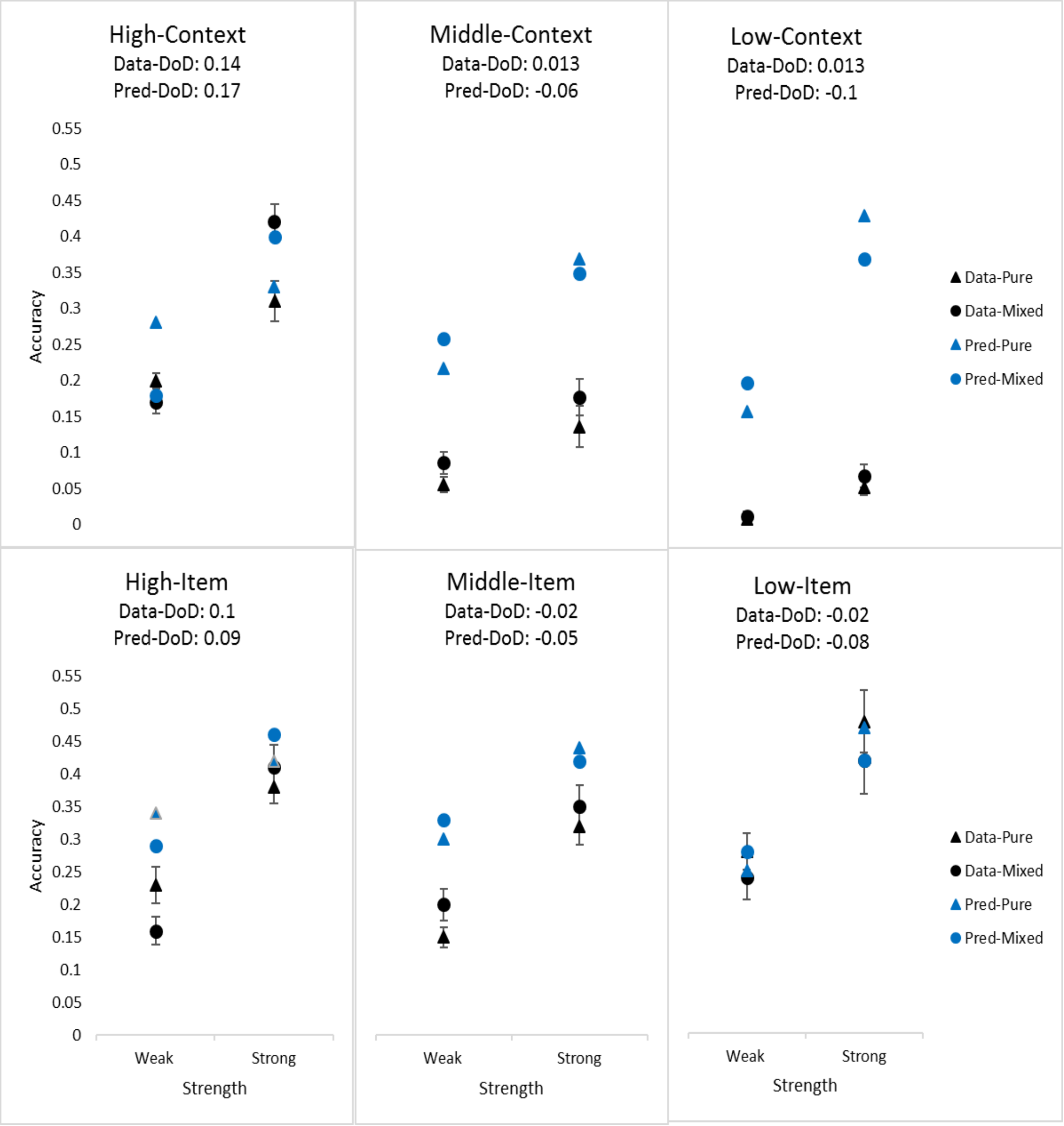


Figure 6: Post-experiment predictions and data of accuracies and Difference of Differences scores (***DoD***) of all conditions.

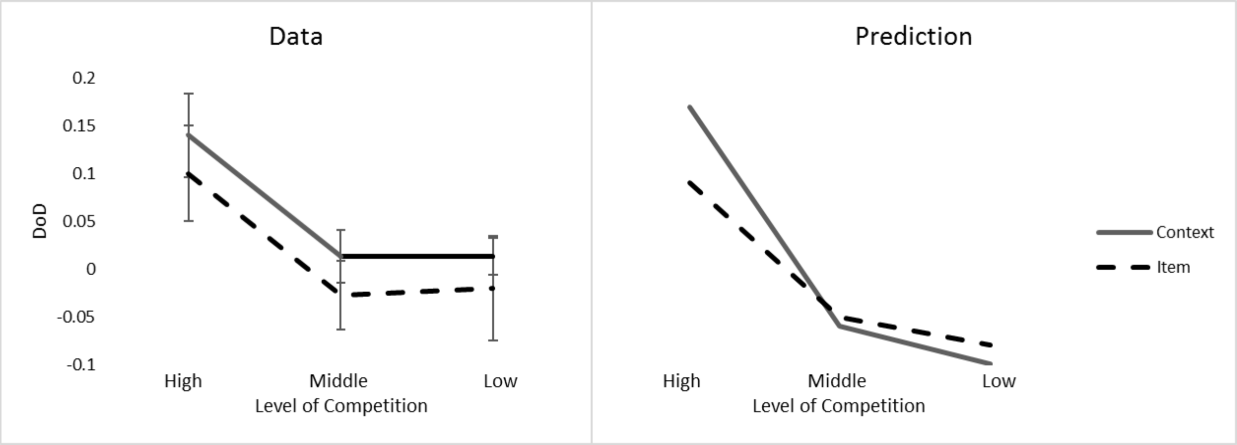


Figure 7: Post-experiment predictions (right) and data (left) of Difference of Differences scores (***DoD***) of high, middle, and low context conditions, as well as high, middle, and low item conditions.

General Discussion

In this project, we observed the magnitude and direction of LSEs changed across levels of competition, but not across types of cue used at test. Specifically, ***DoD*** scores higher than 0 (positive LSEs) were found in the high conditions, for both context and item cuing, while ***DoD*** scores close to 0 (null or slight negative LSEs) were found in the middle and low conditions, for both context and item cuing. This pattern is consistent with REM's mechanisms. The pattern of the data qualitatively matches REM's predictions generated before (Figure 3) and after (Figure 6 and Figure 7) the experiments.

One possible reason the field has attributed LSE to cue type or retrieval task (Ratcliff, Clark and Shiffrin, 1990; Shiffrin, Ratcliff, & Clark, 1990; Malmberg & Shiffrin, 2005) is because we don't manipulate or vary context in most experiments, especially in a list strength paradigm. It is not that context is ignored in theoretical development, in fact the opposite is true. Models have implemented how context changes slowly within a study list (Howard & Kahana, 2002; Mensink & Raaijmakers, 1988; Shiffrin & Steyvers, 1997) and there have been empirical and theoretical studies of context-related effects on memory (Smith, 1979; Murnane & Phelps, 1993, 1994, 1995; Murnane, Phelps, & Malmberg, 1999; Park, Arndt, & Reder, 2006; Lehman & Malmberg, 2009). However, in a standard list strength paradigm, we have not considered the possibility of context change. In REM, retrieval is triggered by the comparison between stored memory traces and a presented cue. When context remains the same and items vary, comparing the context cue to contexts stored in memory traces produces a high level of competition among memory traces. Strong contexts in memory in a mixed list outcompete weak ones leading to a positive LSE in free recall. In contrast, item cues produce a low level of competition among traces leading to a null LSE in cued recall. In short, having shared context and unique items confounds the type of cue and level of competition. The experiment reported here breaks that confounding and shows that the level of competition is critical to the direction and magnitude of the LSE.

We observed two unexpected findings. First, we observed very low accuracy in the middle-context and low-context conditions. Second, we did not see a gradual change from positive to negative LSE. In the following part of Discussion, we propose explanations and potential solutions.

We suspected that one plausible reason accuracy was so low in the low and middle context conditions might be that context and item information are different in nature and might have different stimulus properties (e.g., different ***g*** values). We have considered context and item as two primary sources of information (Clark & Shiffrin, 1987, 1992) and sources of interference (Criss & Shiffrin, 2004a; Criss et al., 2011; Dennis & Humphreys, 2001) in episodic memory. Although item and context do not have clearly defined boundaries (Malmberg & Shiffrin, 2005), psychologists have been studying the distinction between context and item in memory for a long time (e.g., McGaugh, 1932; Craik, Luo & Sakuta, 2010; Rugg et al., 2012; Wang, Yonelinas & Ranganath, 2013). Malmberg and Shiffrin (2005) defined item information as the semantic meaning of the studied and tested stimuli, whereas context information as everything else happening during the encounter of the stimuli. Participants might also distribute attention to context and item differently (e.g., different ***u***). For one, our encoding task focused on encoding item information of the stimuli, such as the scale of association between two words in a word pair (Murnane & Phelps, 1993, 1994, 1995; Murnane, Phelps, & Malmberg, 1999). Also, context information, by its nature, is not the center of attention, and is potentially encoded differently from item. For instance, various strengthening techniques can make participants remember a stimulus better, but effects relying on better encoding of the context are only present when this stimulus is strengthened via repetition in a spaced fashion (Experiment 1 vs experiment 2 in Murnane & Phelps, 1995; Malmberg & Shiffrin, 2005). Therefore, perhaps it was an oversimplification to assume that context and item would behave similarly in the model and participants' mind. Allowing ***g*** and ***u*** to vary did not account for the pattern of data. To try to establish the causal mechanism for low performance in the two context cuing conditions, we conducted further model analysis in Appendix B. In the end, we hypothesize that items are easier to bind to items than items are to bind to color context. A mechanism for this is outside the scope of REM as currently implemented.

In retrospect, perhaps selecting color as context was not an ideal choice. We did so following (Murnane & Phelps, 1993, 1994, 1995) who showed a bias to say "old" in recognition memory when context, defined by color, matched at study and test. However, other studies using pictures as context showed more compelling changes in accuracy when context matched in recognition memory (Murnane, Phelps, and Malmberg, 1999). Therefore, one potential angle is to replicate our studies with pictures instead of colors as context. Using pictures as contexts could make participants differentiate contexts better. This will hopefully avoid low accuracies in context conditions.

As for the second finding, we label the condition middle but of course there are multiple different ‘middles’ between the high and low conditions which are the most extreme manipulations possible (all targets share a cue and no targets share a cue). To follow up we conducted one more experiment, where every list shared two cues, producing another mid-level of competition between the high and low. Data are provided in Appendix C. Together the 4 conditions show a gradual change from a positive LSE in the high condition to a null LSE in the low condition.

In summary, we found the source of LSE is the level of competition at retrieval, consistent with REM, but not the type of cue used at test, as what we have misunderstood. We also found context has more complicated effect on recall performance than item, which we consider evidence to distinguish context and item in memory studies and memory models.

Appendices

Appendix A. Preregistration document

Source of LSE within REM, Syracuse, October 2016. (#1736)

Created: 10/26/2016 06:12 PM (PT)

Shared: 04/16/2017 06:04 PM (PT)

1) What's the main question being asked or hypothesis being tested in this study?

Previous work has evaluated the list strength effect as a function of retrieval task. When context is the only cue (e.g., free recall), a positive list strength effect has been observed but when an item and context serve as a cue (e.g., cued recall), a null list strength effect has been observed. Based on the REM framework, we think this framing is misleading. It is not the cue type or retrieval task that is critical. Instead, the magnitude and direction of the strength effect depends on the level of competition, rather than the type of retrieval cue (item or context). For this project, we define level of competition as the number of words that were associated with the cue during encoding.

2) Describe the key dependent variable(s) specifying how they will be measured.

The key dependent variable is the list strength effect. We measure list strength effect as a difference of differences (see below) for each participant and each condition. DoD= (MS-MW) - (PS-PW), where MS, MW, PS, and PW stands for the measured accuracies of, respectively, mixed strong, mixed weak, pure strong, and pure weak items.

3) How many and which conditions will participants be assigned to?

Participants will complete a full list strength paradigm (weak list, strong list, mixed list) for each condition. In a weak list, all word pairs are presented once. In a strong list, all word pairs are presented four times. In a mixed list, half of the word pairs are presented once, and the other half are presented four times. There are 6 conditions in total, in a mixed-design. Type of cue provided at test (context or item) is manipulated within-subject and number of cues shared by a list (shared/ middle/ unique) is manipulated between-subject. A cue is deemed as an item cue when we present one of words from the word pair with word color black and background color white. A cue is deemed as a context cue when we show a combination of word color and background color on the screen. In shared conditions, all words are assigned the same cue word, and are presented under the same context. In middle conditions, each quarter of word pairs within a list are assigned one cue word and are presented in the same context. In unique conditions, all words have their exclusive cue words and context. Therefore, each participant will complete 6 blocks of study-distractor-test.

4) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

We will conduct a 2 (pure/ mixed) *2 (weak/ strong) *2 (item cue/ context cue) *3 (shared/ middle/ unique) ANOVA on accuracy for archival purposes. Here, we are simply evaluating whether the strengthening manipulation was effective. Strong pairs should be better remembered than weak pairs. The primary analysis of interest is a 2 (content cue/ context cue) *3 (shared/ middle/ unique) ANOVA on DoD scores. Based on model simulations with REM, we have no specific prediction related to the type of cue, but to the number of cues shared by a list. For simulations, every stimulus has 60 features in total, with 20 context features and 40 content features. Every feature of a weak stimulus is stored with a probability of 0.2, and that of strong stimulus with 0.4. A context cue is made of the 20 context features. A content cue is made of some proportion of context features, and 20 content features. In both cueing conditions, DoD should be positive for the shared condition and decrease in magnitude (toward 0, a null effect, or even to a negative value which is a negative list strength effect) for the middle and unique conditions, in that order.

5) Any secondary analyses?

6) How many observations will be collected or what will determine sample size? No need to justify decision, but be precise about exactly how the number will be determined.

We plan to recruit approximately 180 participants in total, with approximately 60 participants each for the shared, middle, and unique conditions. This sample size has been used in previous research studying list strength paradigm. We run between 1-10 subjects at a time, so we will stop data collection when we are approximately within the range of plus or minus 10 subjects.

7) Anything else you would like to pre-register? (e.g., data exclusions, variables collected for exploratory purposes, unusual analyses planned?)

Due to the manipulation of colors in our experiment, participants must be in healthy condition to detect differences among colors, and must not be color-blind. If any participant has an accuracy of zero on any list, their data will not be included into data analysis. Intrusions and corrects are defined with strict criterion. We programmed in a way such that every word participants respond can only be counted as a correct when it is exactly the same as it is presented at study, otherwise it is counted as an intrusion. For every participant, we also plan to measure the time taken to produce every response at test.

8) Have any data been collected for this study already?

No, no data have been collected for this study yet

Verify authenticity: http://aspredicted.org/blind.php/?x=gb6yx4

Appendix B. Further Analysis and Model Fitting to Explain Observed Low Accuracy

As stated, the accuracy of the middle-context and low-context conditions were low. Here we elaborate by further exploring the data and model.

One possibility is that color contexts cues were not an effective probe for memory of an item. To explore this, we looked at the number of output failures per cue. An output failure refers to a refusal to produce any response for a given cue. Providing that our suspicion were true, we would see the number of output failures increase as the number of context cues increases, and also for context compared to item cuing. Note that we expected to see the number of output failures increases as level of competition goes from high, middle, to low, in that increasing the number of cues implies increased chances for participants to give up responding.

Figure 8 displays the number of output failures of high-, middle-, and low-context conditions, as well as high-, middle-, and low-item conditions.

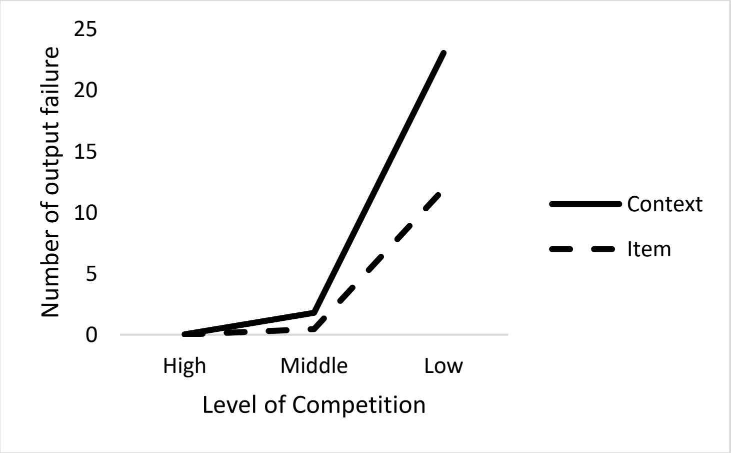


Figure 8: Number of output failures of high, middle, and low context conditions, as well as high, middle, and low item conditions. An output failure means that no response was provided.

We conducted a 3 (high/ middle/ low) * 2 (context/ item) ANOVA and Bayesian ANOVA. A main effect of cue type was detected, F(1,159) = 32.53, p < 0.001, BF_{10} = 1389.414. A main effect of level of competition was also detected, F(2,159) = 200.7, p < 0.001, BF_{10} = 1.121e+38. An interaction between cue type and the level of competition was detected, F(2,159) = 21.70, p < 0.001, BF_{10} = 26238095.2. This shows that as the level of competition decreases, the number of output failures increases under both context cuing and item cuing, but more severely under context cuing. This is consistent with the idea that color contexts area poor cue to retrieve memory for an item.

Furthermore, as we seek to account for the data mechanistically via a memory model, we adjusted the elementary version of REM described and implemented in the main text. That is, we try to find the mechanism that underlies the pattern of data we observed. We vary the threshold for sampling, $\boldsymbol{\phi_m}$ of middle-context condition and $\boldsymbol{\phi_l}$ of low-context condition. We remind you that, previously, we set $\boldsymbol{\phi}$ to be 0 in all conditions, meaning that a memory trace was always sampled no matter how well the cue matched memory. Allowing $\boldsymbol{\phi}$ to change is in middle-context and low-context conditions is one way to implement the idea that context is a poor cue that may not succeed. With all the other parameters the same as fixed and estimated, we estimated $\boldsymbol{\phi_m}$ and $\boldsymbol{\phi_l}$, respectively, from middle-context and low-context condition. Then we generated predictions with 1000 simulated subjects, to see if 1) low accuracy is successfully predicted in middle-context and low-context conditions, and 2) DoDs between predictions and data are still a qualitative match.

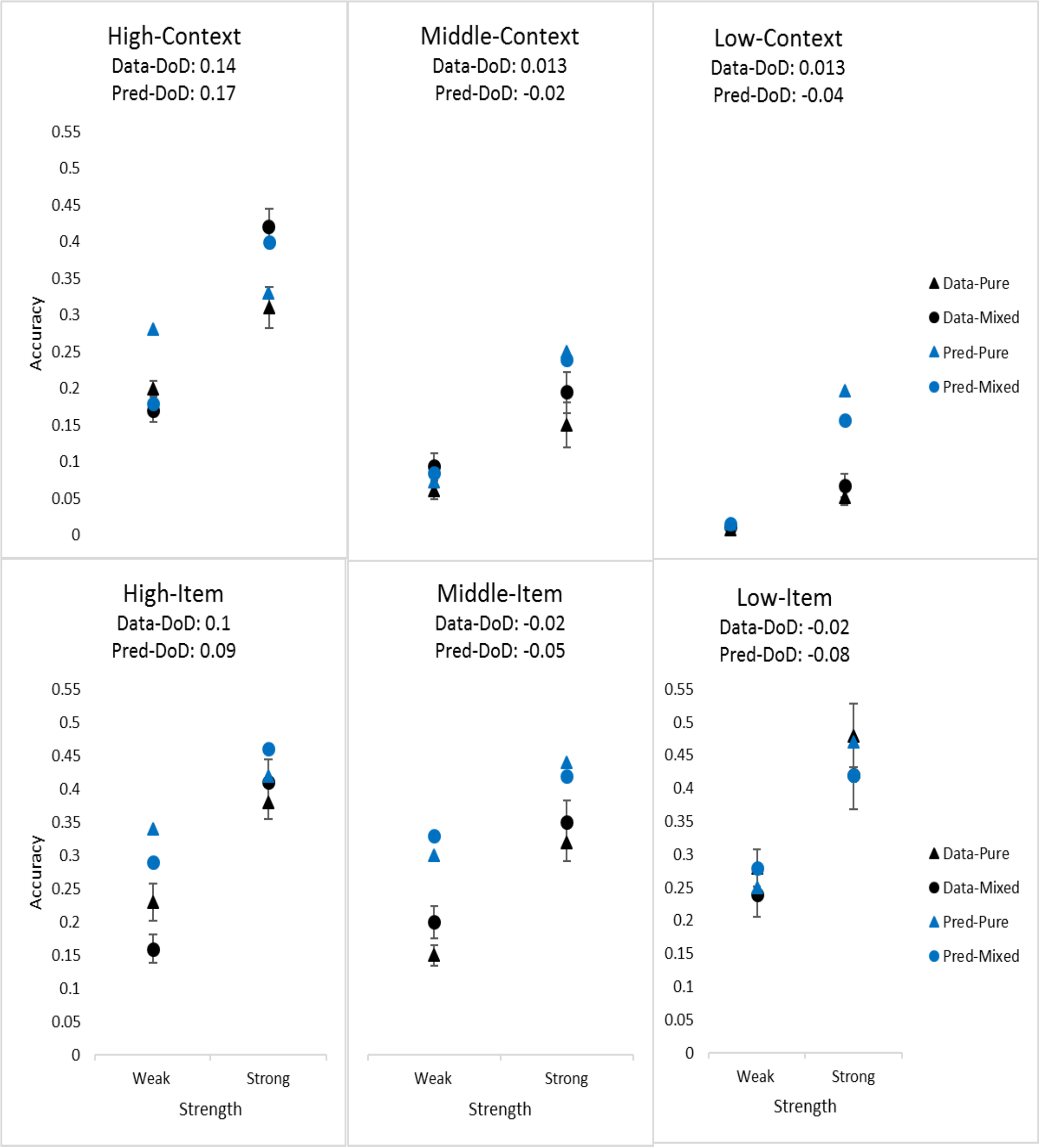


Figure 9: Post-experiment predictions and data of accuracies and Difference of Differences scores (***DoD***) of all conditions.

We estimated $\boldsymbol{\phi_m}$ and $\boldsymbol{\phi_I}$, respectively, to be 1.28 and 2.25. Predictions and data are provided in Figure 9 and Figure 10. Based on Figure 9, low accuracies were successfully predicted in the middle-context and low-context conditions. In addition, based on Figure 10, we can see this version of REM again predicted LSEs, or DoDs changing across the level of competition, not between the type of cue. This confirms our primary theory of this project, the source of LSE is the level of competition.

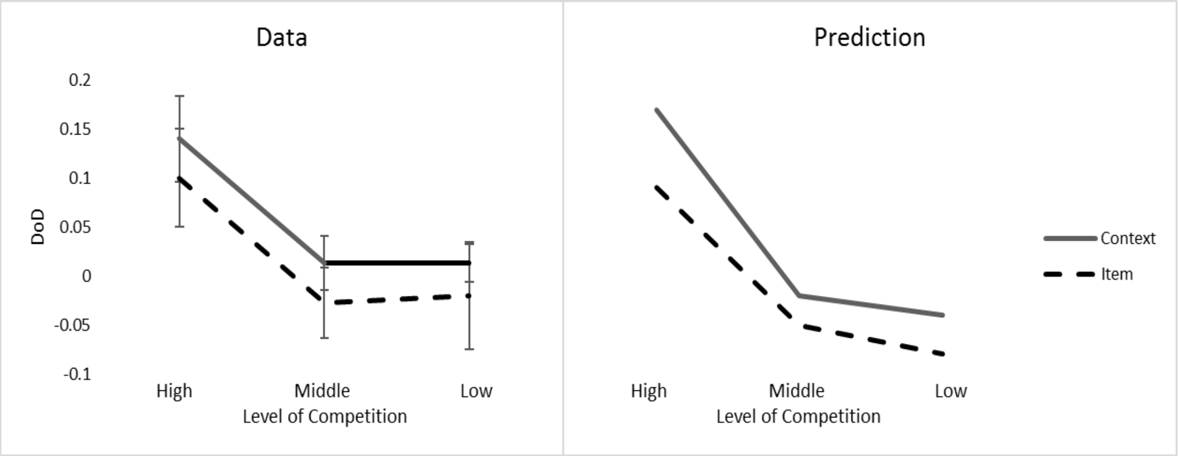


Figure 10: Post-experiment predictions and data of Difference of Differences scores (***DoD***) of high, middle, and low context conditions, as well as high, middle, and low item conditions.

Appendix C. Data and Predictions from the Middle2 Conditions

As mentioned before, we unexpectedly failed to observe a gradual change from positive to null, to negative LSE across our 3 conditions. Therefore, we conducted one more experiment, where every list shared two cues, producing a level of competition between the high and middle conditions. This better reflects the full range between a unique for every item and a shared cue.

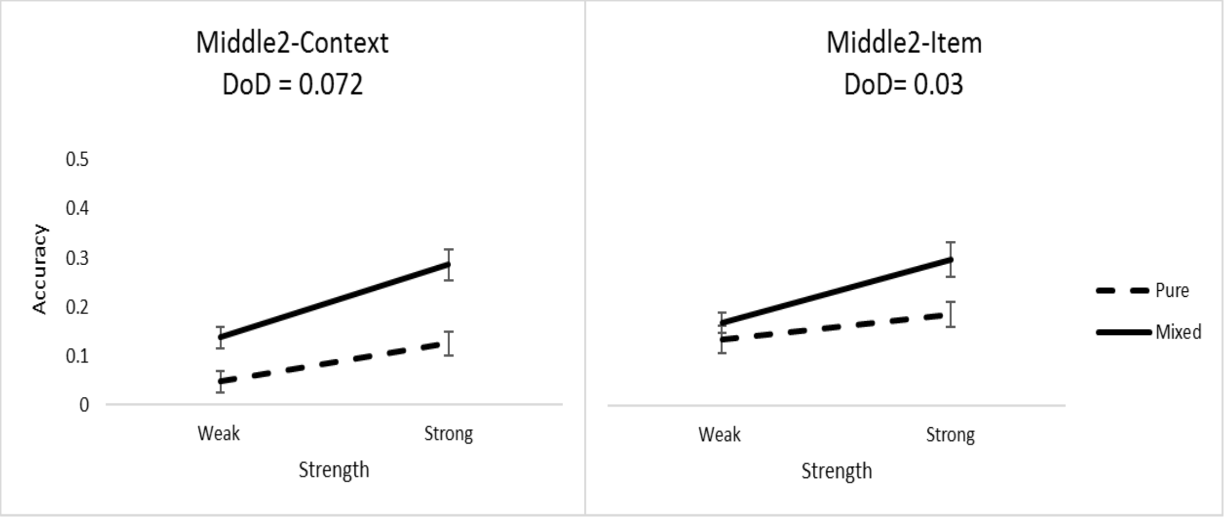


Figure 11: Accuracies and Difference of Differences scores (***DoD***) of middle2 conditions.

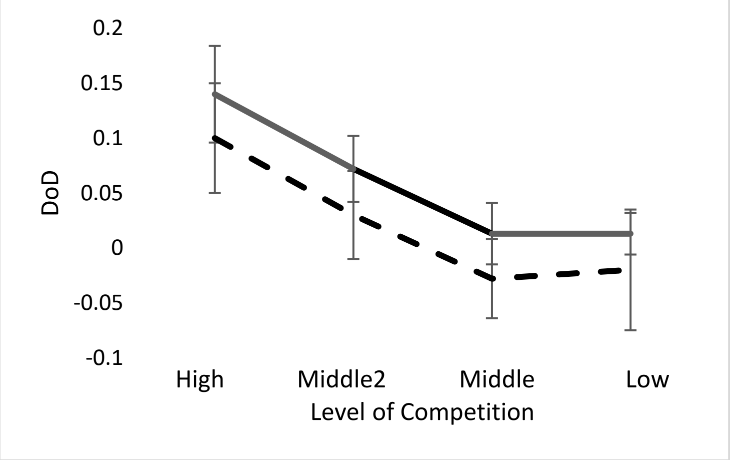


Figure 12: Predictions and data of Difference of Differences scores (***DoD***) of all conditions, generated without $\boldsymbol{col}$, $\boldsymbol{\phi_m}$, and $\boldsymbol{\phi_l}$.

We recruited 63 participants in total. The experimental design is identical except there are 2 cues possible. Here we only provide descriptive data (see Figure 11 and 12). First, as in the other studies, LSEs didn’t differ much between context cuing and item cuing. Furthermore, based on Figure 11, the observed LSEs in the middle2 conditions, under context cuing and item cuing, were around null to slightly positive. Last but not least, based on Figure 12, we can see the full change of LSEs across the levels of competition. This confirms our hypothesis that the level of competition is the source of LSE, not the type of cue used at test.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. Pages 267{281 of: Petrov, B. N., and Csaki, F. (eds), *Second International Symposium on Information Theory*. Budapest: Akademiai Kiado.

Brainard, D. H. (1997) The Psychophysics Toolbox, Spatial Vision 10:433-436.

Clark, S. E., & Shiffrin, R. M. (1987). Recognition of multiple-item probes. *Memory & Cognition*, 15(5), 367–378.

Clark, S. E., & Shiffrin, R. M. (1992). Cuing effects and associative information in recognition memory. *Memory & Cognition*, 20(5), 580–598.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684.

Craik, F. I. M., Luo, L., & Sakuta, Y. (2010). Effects of aging and divided attention on memory for items and their contexts. *Psychology and Aging*, 25, 968–979.

Criss, A. H. (2006). The consequences of differentiation in episodic memory: Similarity and the strength based mirror effect. *Journal of Memory and Language*, 55(4), 461–478.

Criss, A. H. (2009). The distribution of subjective memory strength: list strength and response bias. *Cognitive Psychology*, 59(4), 297–319. doi: 10.1016/j.cogpsych.2009.07.003

Criss, A. H. (2010). Differentiation and response bias in episodic memory: evidence from reaction time distributions., *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(2), 484–499.

Criss, A. H., & Koop, G. J. (2015). Differentiation in episodic memory. In J. Raaijmakers, A. H. Criss, R. Goldstone, R. Nosofsky, & M. Steyvers (Eds.), *Cognitive modeling in perception and memory: A festschrift for Richard M. Shiffrin* (pp. 112–125). Psychology Press.

Criss, A. H., Malmberg, K. J., & Shiffrin, R. M. (2011). Output interference in recognition memory. *Journal of Memory and Language*, 64, 316–326.

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, 447–460

Criss, A.H. & Shiffrin, R.M. (2004a). Context noise and item noise jointly determine recognition memory: A comment on Dennis & Humphreys (2001). *Psychological Review, 111*(3), 800-807.

Criss, A.H., & Shiffrin, R.M. (2004b). Pairs do not suffer interference from other types of pairs or single items in associative recognition. *Memory & Cognition 32*(8), 1284-1297.

Cochran, W. G. (1952). The χ 2 test of goodness of fit. *Annals of Mathematical Statistics*, 25, 315–345.

Davelaar, E.J., Haarmann, H.J., Goshen-Gottstein, Y., & Usher, M. (2006). Semantic similarity dissociates short- from long-term recency effects: Testing a neurocomputational model of list memory. *Memory & Cognition, 34*(2), 323-334.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, 108(2), 452–478.

Diller, D. E., Nobel, P. A., & Shiffrin, R. M. (2001). An ARC–REM Model for Accuracy and Response Time in Recognition and Recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 27(2), 414-435.

Eliason, S. R. (1993). Maximum likelihood estimation: Logic and practice. *Quantitative applications in the social sciences*. Newbury Park, CA: Sage.

Etz, A. & Vandekerckhove, J. (in press). Introduction to Bayesian Inference for Psychology. *Psychonomic Bulletin & Review*

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, 91, 1-67.

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, 13(1), 8–20.

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), 5–16.

Grider, R. L . & Malmberg, K. J. (2008). Discriminating Between Changes in Bias and Changes in Accuracy for Recognition Memory of Emotional Stimuli. *Memory & Cognition*, 36(5), 933-946.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 77–99). Hillsdale, NJ: Erlbaum.

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

Humphreys, M.S., Pike, R., Bain, J.D., & Tehan, G. (1989). Global matching: A comparison of the SAM, Minerva II, Matrix, and TODAM models. *Journal of Mathematical Psychology*, *33*, 36-67.

JASP Team (2016). JASP (Version 0.7.5.5) [Computer software]

Kleiner M, Brainard D, Pelli D, 2007, "What's new in Psychtoolbox-3?" Perception 36 ECVP Abstract Supplement.

Lehman, M., & Malmberg, K. J. (2009). A global theory of remembering and forgetting from multiple lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 35(4), 970–988.

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., Ly, A., Gronau, Q. F., Smira, M., Epskamp, S., Matzke, D., Wild, A., Rouder, J. N., Morey, R. D. & Wagenmakers, E.-J. (2015). JASP (Version 0.7) [Computer software].

Malmberg, K.J., & Shiffrin, R.M. (2005). The "one-shot" hypothesis for context storage. Journal of Experimental Psychology: *Learning, Memory, and Cognition*, 31(2), 322-336.

McGeoch JA. (1932). Forgetting and the law of disuse. *Psychology Review*. 39:352–70

Mensink, G-J. M., Raaijmakers, J.G. (1989). A model for contextual fluctuation. *Journal of Mathematical Psychology*, 33(2), 172-186.

Morey, R. (2015). http://bayesfactor.blogspot.co.uk/2015/01/on-verbal-categories-forinterpretationhtml?m=1

Murdock, B.B., (1982). A Theory for the Storage and Retrieval of Item and Context Information. *Psychological Review 89*(6), 609-626.

Murdock, B.B., & Kahana, M.J. (1993). Analysis of the list-strength effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(3), 689-697.

Murnane, K., & Phelps, M. P. (1993). A global activation approach to the effect of changes in environmental context on recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 882-894.

Murnane, K., & Phelps, M. P. (1994). When does a different environmental context make a difference in recognition? A global activation model. *Memory & Cognition*, 22, 584-590.

Murnane, K., & Phelps, M. P. (1995). Effects of changes in relative cue strength on context-dependent recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 158-172.

Murnane, K., Phelps, M.P., & Malmberg, K.J. (1999). Context-dependent recognition memory: the ICE theory. *Journal of Experimental Psychology: General*, 128, 403-415.

Park, H., Arndt, J.D., & Reder, L.M. (2006). A contextual interference account of distinctiveness effects in recognition. *Memory & Cognition, 34*(4), 743-751.

Pelli, D. G. (1997) The VideoToolbox software for visual psychophysics: Transforming numbers into movies, Spatial Vision 10:437-442.

Ratcliff, R., Clark, S.E., & Shiffrin, R.M. (1990). List strength effect I: Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(2), 163-178.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, 88, 93–134.

Rugg, M. D., Vilberg, K. L., Mattson, J. T., Yu, S. S., Johnson, J. D., & Suzuki, M. (2012). Item memory, context memory and the hippocampus: fMRI evidence. *Neuropsychologia*, 50, 3070–3079.

Shiffrin, R.M., Ratcliff, R.C., & Clark, S.E. (1990). List strength effect II: Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16(2), 179-195.

Shiffrin, R. M., Ratcliff, R. C., Murnane, K., & Nobel, P. (1993). TODAM and the list strength and list-length effects: Comment on Murdock and Kahana (1993). *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 19(6), 145–149.

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

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 460-471.

Starns, J. J., White, C. N., & Ratcliff, R. (2010). A direct test of the differentiation mechanism: REM, BCDMEM, and the strength-based mirror effect in recognition memory. *Journal of Memory and Language*, 63(1), 18–34.

Tulving, E., & Hastie, R. (1972). Inhibition effects of intralist repetition in free recall. *Journal of Experimental Psychology,* 92, 297- 304.

Wang, W. C., Yonelinas, A. P., & Ranganath, C. (2013). Dissociable neural correlates of item and context retrieval in the medial temporal lobes. *Behavioural Brain Research*, 254, 102–107.

Wilson, J.H., & Criss, A.H. (2017). The list strength effect in cued recall. *Journal of Memory and Language*, 95, 78-88.

Wixted, J. T., Ghadisha, H., & Vera, R. (1997). Recall latency following pure- and mixed-strength lists: A direct test of the relative strength model of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 523–538.

SHARON CHEN

August 2017

Education

B.A. Applied Psychology, Department of Psychology, Guangzhou University of Chinese Medicine (GZUCM), 2011-2015.

Graduate Student. Cognition, Brain, and Behavior Program, Syracuse University, 2015 – present.

Conference Presentations

Chen, S & Criss, A.H. (April 2015). The Investigation of Context in Memory Retrieval: The Elicitation of Positive and Negative List Strength Effect in Cued Recall. CBB Seminar of Syracuse University.

Chen. S & Criss, A.H. (July 2016). Weighting of Item versus Context in Memory Retrieval: List Strength Effect(s) in Cued Recall. Poster Presentation at Computational and Mathematical Modeling of Cognition 2016, Dobbiaco, Italy.

Chen. S & Criss, A.H. (July 2016). Weighting of Item versus Context in Memory Retrieval: List Strength Effect(s) in Cued Recall. Speed Talk at Computational and Mathematical Modeling of Cognition 2016, Dobbiaco, Italy.

Chen. S & Criss, A.H. (November 2016). Investigation of the Source of List Strength Effect within The Retrieving Effectively from Memory Framework. CBB Seminar of Syracuse University.

Chen. S, Wilson, J.H, & Criss, A.H. (November 2016). Investigation of the Source of List Strength Effect within The Retrieving Effectively from Memory (REM) Framework: Types of Cue versus Levels of Competition. Poster Presentation at Psychonomic Society's 57th Annual Meeting, Boston.

Chen. S & Criss, A.H. (February 2017). Investigation of the Source of List Strength Effect within The Retrieving Effectively from Memory Framework. CBB Seminar of Syracuse University.

Research Experience

Graduate Student, Memory Modeling Lab

Fall 2015 – Present

PI: Amy H. Criss, Ph.D., Syracuse University

Teaching Experience

Teaching Assistant:

Introduction to Psychology

Cognitive Psychology Lab