The Word Frequency and List Length Effects on Cued Recall

Research Thesis

Presented in partial fulfillment of the requirements for graduation

with research distinction in Psychology in the undergraduate

colleges of The Ohio State University

By

Rui Cao

The Ohio State University

May 2012

Project Advisor: Dr. Simon Dennis, Department of Psychology

Abstract

The list length effect is a phenomenon in which performance improves when the number of studied items decreases. This effect is present in most memory tasks. However, Dennis and Humphreys (2001) showed that the list length effect can be eliminated in recognition memory tasks if controls for retention interval, attention, rehearsal and contextual reinstatement are employed. In some unpublished data we have also found a null-list length effect in cued recall experiments, when it is presented in an experiment set that contains cued recall, associated recognition, and single item recognition. Therefore we would like to single out the cued recall experiment to see if we can replicate this finding. Word frequency effects (high frequency advantages on recall and low frequency advantages on recognition) represent another disassociation between recognition and recall. We are also interested in the word frequency effect in cued recall and examine it separately for cue and target items. In the experiment, we manipulated both the list length and word frequency effects with a filler task between study and test. We found that there is a small difference between list length but not statistically significant. There are also low frequency advantage for cues and high frequency advantage for targets. These findings place critical constraints on viable models of cued recall.

Introduction

People usually do not differentiate the memory tasks they are performing in their daily life. For example, when you start working at a new office, there may be a moment when you try to remember if the guy who just entered the elevator is someone from your office. After a while you may be able to name a list of people in the office when asked. Furthermore, when you do recall a name, you might also recall his or her child because you saw them during the “bring your kids to work” day. Even though all these events seem so natural to you and you do not feel like you need more than the word “remember” to describe them, they are different memory tasks that are studied separately. The first one is a single item recognition memory test, where you need to judge if an item and a context are associated. The second one is a free recall memory test where you need to recall items associated with the given context. And the last one is cued recall memory test where you use one item to recall another item that is associated with the given item.

Whether various memory tests can be accomplished by similar retrieval processes is still an open question. The differences between two memory tests, single item recognition and free recall have especially drawn people’s attention. In the present study, we focused on differences between free recall and recognition on the word frequency effect and list length effect, and examine how the two factors affect cued recall, where participants were given an item cue (like in single item recognition) and are required to produce the target (like in free recall) that was paired with a given cue during the study phase (Criss, Aue and Smith 2011).

Word frequency effect

One of the most confusing disassociations between recognition and recall is the word frequency effect. Word frequency refers to how often a particular word is used in our daily life. Previous studies on free recall have shown that people perform recall tasks better when asked to recall pure high frequency word lists versus pure low frequency word lists (e. g. Deese, 1960; Postman, 1970; Ward, Woodward, Stevens, and Stinson, 2003). However, in the mixed word list condition (a list that contains both high frequency words and low frequency words mixed together in random order), things get complicated: some experiments report a reduced high frequency advantage; some report a low frequency advantage while some report that there is no difference (Ward, et al., 2003). In general, the absence of high frequency word advantage in mixed list conditions is called the mixed list paradox (Gillund & Shiffrin, 1984). On the other hand, there is a reliable low frequency advantage in recognition memory test cross pure list and mixed list conditions (e.g. Glanzer and Adams, 1985; Balota and Neely, 1980).

There are several potential explanations for the high frequency advantage in recall processes. One simple and straight forward explanation follows. Because we see high frequency words a lot in our daily life, they are more ready in our mind and thus more retrievable (Madan, Glaholt & Caplan, 2010). This simple explanation can deal with the high frequency advantage in pure lists. However, it fails to predict the absence of a high frequency word advantage in the mixed list condition. Some argue that in the mixed list condition, low frequency words stand out among high frequency words and draw more attention during study, which leads to better encoding processes that help to eliminate the high frequency advantage. (e.g. May & Tryk, 1970).

Another more popular explanation for the word frequency effect in recall is about the associations between list items. One particular model that successfully predicts the word frequency effect in recall using that assumption is the Search of Associative Memory model (SAM; Raaijmakers & Shiffrin, 1981). According to that model, memory is about forming a set of associations between items in the list and between items and the context. Context is a mental representation of the studied list that enables participants to focus retrieval on the study list. Participants begin recall by using the context as a cue, but then use subsequently recalled words as cues along with the context. This may explain the high frequency word advantage in recall because the strong preexisting associations between the high frequency words will enable participants to better cue subsequent words in recall (Gillund & Shiffrin, 1984). However, the ability to activate the connections among high frequency words is limited in the mixed list condition. Since words have to be near each other in the study list in order to activate the inter-item associations, mixed lists are less likely to have high frequency words adjacent to each other due to the presence of the low frequency words. In other words, it is easier for two high frequency words to activate their pre-experiment connection if they appear close to each other temporally and the chance is relatively lower in the mixed list condition. For example, if you see the high frequency word A followed by several low frequency words in a mixed list, by the time you see the next high frequency word B, the word A is no longer activate. This makes it difficult to active the pre-experiment connection between A and B compared to the situation where you see A followed by B immediately as in pure high frequency lists.

For recognition, it is a different case: people generally show a robust low frequency word advantage (Glanzer & Adams, 1985). There are many models that try to capture the word frequency effect in recognition (e.g. Glanzer & Adams, 1990; Shiffrin & Steyvers, 1997; Dennis & Humphreys, 2001). However, the underlying resource (item noise vs. context noise) for such an effect in recognition is still under debate. The item noise refers to interferences from other studied items and the context noise refers to interferences from pre-experiment contexts that the target item has appeared in (Kinnell & Dennis, 2011). The item noise approach (e.g. Retrieving Effectively from Memory model, REM; Shiffrin & Steyvers, 1997) would argue that low frequency words benefit from more distinguished features (for example, less commonly used letters). The less overlapping features with other items in the study list as well as the distracters in test result a low false alarm rate and more confident hit rate for low frequency words. And this advantage will not be affected by mixed list condition.

There are fewer models that apply the context noise approach. One successful theoretical model of recognition memory applying context noise approach is the bind-cue-decide model (BCDMEM; Dennis & Humphreys, 2001). Based on the model, during the test phase of a recognition memory test, we use the given item to form a retrieved context and compare it with the reinstated context. We give a “yes” response if the retrieved context and the reinstated context are similar and a “no” response if they are very different. High frequency targets are less likely to be recognized because they have been experienced in more contexts, which lead to a noisier retrieved context and make it less likely to match the reinstated context. For example, when you are making the judgment about if you have seen the word “chair” in the previous study list, you might hesitate because you are not sure if the familiar feeling about the word comes from the memory about the studied list, or the fact that you just saw a chair in the morning and thus activated the word in your mind then. Unlike recall, the mixed lists condition do not affect the low frequency advantage because the influence from the other items is negligible; that is to say since the main interference comes from the word itself, other words in the list will not impact participants’ performance for a particular word.

List length effect

Another difference between free recall and recognition is the list length effect. The list length effect refers to the phenomenon where participants’ performance decreases as the study list gets longer. There is a robust list length effect in free recall: adding more items to the study list decreases the proportion of words recalled (Raaijmakers & Shiffrin, 1981). Most models of free recall predict that effect and argue that this is due to more competition between items in longer lists. For example, the SAM model argues that when participants start to sample words from memory and determine if they are the targets, the probability of sampling an item is lower for the longer list. Although more samples are made with longer lists, the sampling effect is more powerful (Gillund & Shiffrin, 1984). It was believed that there was also a list length effect in recognition (e.g. Nobel & Shiffrin, 2001). This could be explained by the item noise approach, which assumes that the main interference in recognition memory comes from other items in the list for recognition (Gillund & Shiffrin, 1984). However, there is a series of papers demonstrating that there is no list length effect when confounding variables are controlled (e.g. Dennis & Humphreys, 2001; Criss & Shiffrin, 2004a; Dennis, Lee & Kinnel, 2008).

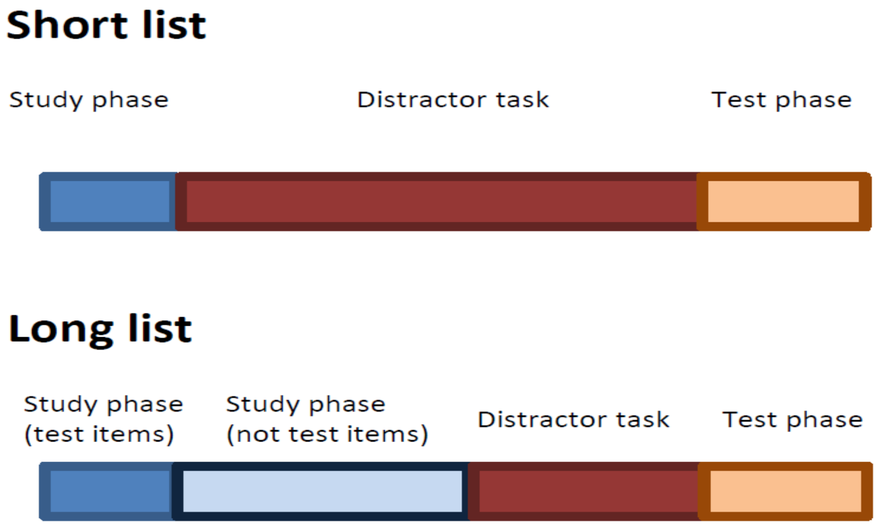

Figure 1. The framework describes the procedure to properly control the confounded variables: retention interval, attention and displaced rehearsal. In the framework, the long list contains two parts: the first part contains as many items as the length of short list and will be tested later in the test phase; the second part will not be tested to control the attention variable. The distractor task will fill the time gap between study phase and test phase, and it is interesting enough to stop participants from rehearsing (to control the displaced rehearsal) and balance the different retention intervals between the short list and the long list by testing only the first part of the long list. The current study applies this framework.

Kinnel and Dennis (2010) argue that long lists are treated “unfairly” compared to short lists because long lists require more time to view. The paper argues that with proper control over the experiment (shown in Figure 1), the list length effect could be eliminated. One consequence of the extra time for the long list is a longer retention interval, meaning that the time gap between being studied and being tested is longer for items in long lists. Memory loss could potentially happen during the long retention interval in the recognition memory test. Attention is another possible confounded variable. It is likely that participants become tired over the course of long lists and pay less attention to each item. At last, displaced rehearsal (rehearse items that will not be tested later) will arise when you control the above two variables by applying a filler task between study and test (control retention interval) and only test the beginning part of the long list (same length as the short list to control the attention). Because all of the items in the short list will be tested while only part of the long list will be tested, participants will benefit from any rehearsal of the short list items; such benefit is not necessarily true for the long lists. The way to control the rehearsal is to ensure that the filler task is interesting enough to stop participants from rehearsing. The empirical data shows that by properly controlling these confounded variables, one can eliminate the list length effect on recognition memory test (Dennis & Humphreys, 2001; Dennis, Lee & Kinnel, 2008; Kinnell & Dennis, 2010).

This null list-length effect can be accommodated by the BCDMEM model which proposes that the recognition decision is based on whether or not the list context is included in the probe item’s previous context as we described earlier.

Because the main interference comes from the item itself, adding more items to a study list does not affect this process (Dennis & Humphreys, 2001).

As discussed above, there are several successful models that could capture the characteristics of recognition memory test and free recall memory test separately. However, because of those differences, there is no comprehensive memory model for episodic memory that could count for both recognition and free recall (Criss, Aue and Smith, 2011). The purpose of this particular paper is to contribute to building a model that can explain various memory tests. Therefore, we examine both the word frequency and the list length effect on cued recall, during which participants are offered an item as cue like recognition memory test and are also required to generate an item like other recall tests.

Cued recall

In cued recall, participants are typically presented with lists of paired words; later one of the paired words (the cue) is given and participants are asked to provide the other one (the target). Here we assume that participants form a three-way binding between the paired items as well as the study context. Later during the test phase, participants will use both the given word and the context as cues to retrieve the target. As we mentioned earlier, cued recall shares properties with both free-recall (need to produce the target words) and single item recognition (use given items as cue). However, most of the models and theories discussed above have not been directly applied to cued recall. The underling retrieval process of cued recall is still unclear.

The list length effect on cued recall is not well studied in literature. A few experiments have reported a list length effect (e.g. Nobel & Shiffrin, 2001). Nobel and Shiffrin (2001) tried to explain the list length effect in cued recall by applying a retrieval process from the REM model using the item noise approach. The general idea is to calculate a likelihood ratio by comparing the item cue to every studied pair that was stored and choose the most related pair to produce targets. However, those experiments failed to apply paradigms that could control the confounding variables such as retention interval and attention. On the other hand, Chapman and Dennis (unpublished data) found a null list-length effect in cued recall when controlling the length of the retention interval, rehearsals and attention. The null list-length result can be achieved by assuming that participants primarily use the item cue to retrieve the target and use the context cue to reduce but not eliminate the interference from pre-experiment memories. In this way, the main noise during retrieval comes from the pre-experiment memories associated with the item cue and the interference from other studied items is negligible.

In this paper, we will also assess the word frequency effect on the target and the cue separately. We predict that there will be a low frequency cue advantage because the low frequency cues are likely to be associated with fewer items. Therefore there are fewer pre-existing associated items to compete during the retrieval process with a low frequency cue compared to that for a high frequency cue. Also, there should be an advantage for targets for the high retrievalbility in our memory of high frequent words

(Madan, et al, 2010). Although, the simple high availability explanation failed to account for free-recall (Hulme, Stuart, Brown & Morin, 2001), Madan et al. (2010) argues that this is the main source of high frequency target advantage in cued recall. Such results will help us eliminate the gap between recognition memory and recall memory and help us to build a unified memory model.

Method

Participants

Participants were 99 undergraduate students in an introductory psychology course at the Ohio State University. Participants enrolled in the experiment to fulfill a course requirement. 29 subjects were dropped due to their low performance (we dropped participants if they recall less than one correctly for each list). The reason that we lost so many subjects might be because we employed no encoding task and used a long retention interval, which may have weakened the context cue and led to a lot of proactive interference from pre-experimental memories.

Material

The stimuli for the experiment were lists of 12 or 48 paired words randomly drawn from a pool of 5-, 6-, and 7-letter low frequency words (LF 1-4 Google count per million) and high frequency words(HF100-200 Google count per million). The short lists (12 pairs per list) were composed of 24 words, and the long lists (48 pairs per list) were composed of 96 words. The stimuli for the filler tasks were playing card games.

Design

The within-subjects factors of the experiment were the length of the study list, the word frequency of the cue and the word frequency of the target. There were two conditions of the list length (12 pairs and 48 pairs) and 4 combined conditions of each pair (HF cue with HF target; HF cue with LF target; LF cue with HF target; LF cue with LF target). The dependent measure was the accuracy when the participants tried to produce the target words based on provided cue words.

Procedure

The experiment consisted of two blocks (one for the short list condition and one for the long list condition; order was counterbalanced) each containing a study phase, a filler task, and a test phase, in that order. The length of the study phase is dependent on the list length condition (66 s for the short list and 264 s for the long list). There were no words repeated between the short lists and the long lists for each participant. The filler task was 120 s for the long list condition and 120s plus the time gap between the study phases of short list and long list (that is 318s for the short condition). The test phase is self-paced.

During the study phase, pairs of items were presented for 5s without any encoding tasks, followed by a clear screen for 0.5s. Participants were instructed that the item on the left of each pair would be the cue, and the one on the right would be the to-be-generated item (the target). There were filler tasks of a card game between the study phases and the test phases where participants were instructed to make certain responses to the card presented on the screen based on certain rules to earn points (for example, pressing the ENTER key when seeing two cards from the same suit in a row will give them one point). The length of the study phase and the test phase varied based on the list length condition as we described above. The stimuli were presented via OpenSesame software.

On each trial of the test phase, the cue word was presented on the left of the screen, and participants were instructed to type the word that was paired with the given cue in the study phase. Participants were allowed to give no answer by pressing the enter key. All 12 pairs in the short list and the first 12 pairs in the long list were tested, each containing equal number of all four frequency pair combinations in random order. Spelling errors and typographical errors were forgiven and scored appropriately later during data analysis.

Results

Word frequency effect

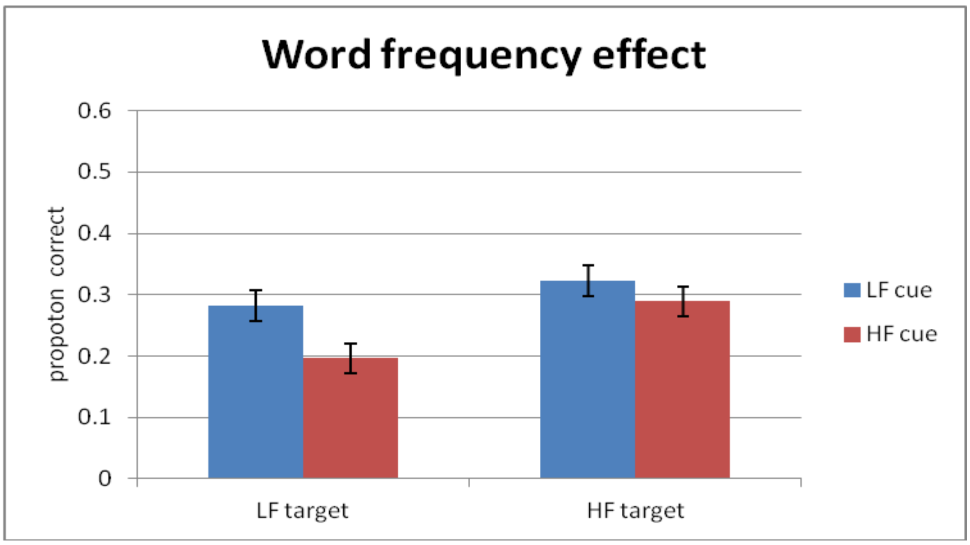

Figure 2. The word frequency effect. LF is low frequency words and HF is high frequency words.

A 2x2x2 (list length x Word Frequency of cues x Word frequency of targets) ANOVA was performed on the proportion of correct responses. As shown in Figure 2, we found a low frequency word advantage for cues and a high frequency word advantage for targets. The mean and SE for each cue x target combination is shown in the Table 1. The mean of proportion correct for low frequency cue was 0.302 with SE = 0.018 and the mean for high frequency cues was 0.243 with SE=0.016. The difference was 0.060, and the effect was significant: F (1, 69) = 9.309, p<0.01. The mean proportion correct for low frequency targets was 0.239 with SE= 0.016 and the mean for high frequency target is 0.306 with SE = 0.018. The difference was 0.067 and the effect was also statistically significant: F (1, 69) = 11.65, p<0.01. None of the interactions were significant.

List length effect

Paired Samples Statistics

	Mean	N	Std. Deviation	Std. Error Mean
Long list	.260	70	.195	.023
Short list	.279	70	.210	.025

Table 1 shows the probability of correctly recalling the target item in the experiment.

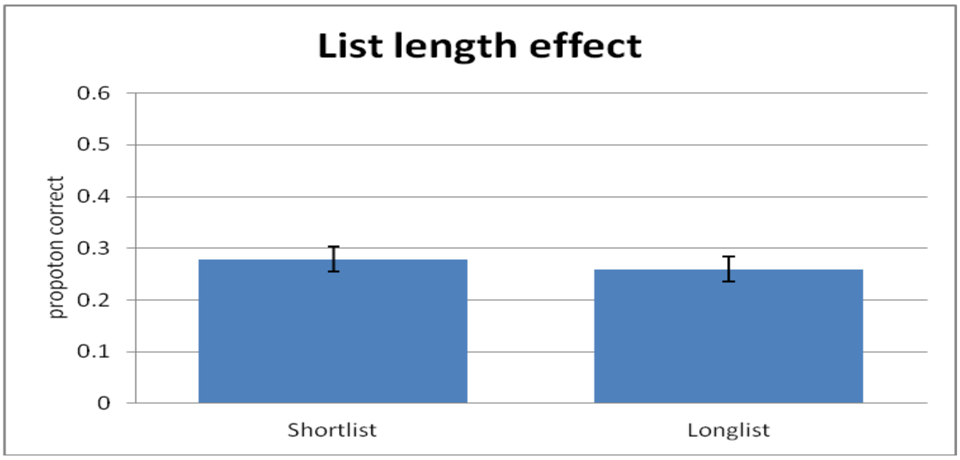

Figure 3 list length effect results.The mean for short list was 0.279 with SE = 0.025, and the mean for long list was 0.260 with SE=0.023.

The list length results are shown in Figure 2. The mean proportion correct for the short list was 0.279 with SE = 0.025 and the mean for long list was 0.260 with SE=0.023. The difference was small (0.019) and statistically insignificant ($F(1, 69) = 0.405$, $p=0.527$). The absence of list length effect was consistent with our hypothesis.

Discussion

The experiment demonstrates that the list length effect (12 vs. 48 pairs) can be eliminated with control over confounded variables like attention and retention interval. The experiment has also shown that word frequency has an effect on both the cue and the target separately. Performance is better for the low frequency cues and high frequency targets.

Word frequency effect

There is a high frequency advantage for the target and a low frequency advantage for the cue. The high frequency target advantage is consistent with previous studies (e.g. Criss et al, 2011). The effect is generally explained by the assumption that high frequency words are more accessible in our memory system and thus more easily retrieved (Criss et al., 2011; Mandan et al, 2010). Although such an assumption is not sufficient to explaining the word frequency effect in free recall, it seems to be the main reason for the high frequency target advantage.

Another explanation we can apply from free recall is the inter-item association hypothesis. This hypothesis assumes that high frequency words are better at forming inter-item association within the frequency and is recognized as the standard explanation for word frequency effect in free recall. However, the inter-item association assumptions would predict a better performance for HH pairs than for LH pairs, which is not supported by the current experiment as well as some other experiments (e.g. Criss et al., 2011). The ambiguous explanation for the high frequency target advantage indicates that more effort is required to understand the retrieval process in cued recall.

One interesting result in the current study is the presence of a low frequency cue advantage, which was not observed in the previous studies (e.g. Criss et al., 2011; Madan et al., 2010). The word frequency effect on cues is rather ambiguous in the literature: Criss et al. (2011) reported inconsistent results across several experiments and Gillund and Shiffrin (1984) reported the differences to be small. However, the general patterns reported in these data set indicate that HH pairs are best recalled, LL pairs are worst and HL, LH are somewhere in between, which leads to a null word frequency effect for the cue or sometimes a high frequency cue advantage. One potential explanation that might account for the different results on the cue is the strength of the context cue. The general argument is that if the context cue is weakened by the absence of an encoding task and/or the long retention interval, the interferences from pre-experiment associations (which is largely determined by how often you see the words in your daily life, or word frequency) is increased. The low frequency words then benefit because they have fewer pre-experiment associations and serve as better cues compared to high frequency words.

The lack of an encoding task generally leads to a weaker connection both between the studied pair and between the pairs and the context (Gillund & Shiffrin, 1984). The weakened context cue reduces its ability to eliminate pre-experiment associations while the weakened associations between items will also lead to a more interference from pre-experiment memory. This assumption could explain the different patterns we found compared the results of some previous studies (e.g. Criss et al., 2010). The fact that participants perform worse in our experiment compare to previous studies may indicate that the less sufficient context cue has weakened participants' performance by introduce more interference from the pre-experiment memories.

Another explanation for the weak context cue comes from the long distractor task in the current study. The previous studies which failed to find a low frequency cue advantage shared a relatively short distractor task between study and test phase (various from about 30s to 60s). In current study, the distractor task for the long list condition was 120s and that for the short list condition was over 5mins. The ability to reinstate context may decrease as the length of the retention interval increases (Dennis & Humphreys, 2001). Such a long retention interval might reduce participants' ability to reinstate context and thus lead to a less effective retrieval of the memory of target pairs as well as less inhibited pre-experiment memory. Such an effect will also result in more interference from pre-experiment associations of the cue; thus the low frequency cue advantage will be present.

Criss et al (2010) offer another potential explanation for the lack of a low frequency advantage. The authors argue that it takes effort and attention to encode low frequency words (e.g. Criss & Melmberg, 2008). In previous studies, the encoding tasks focused on forming the association between the paired items and thus led participants' attention to that direction, which reduced the extra attention that the low frequency words required during encoding. Therefore, the low frequency advantage that is usually present in single item recognition is absent under the encoding task condition. However, in the present study, there is no encoding task to eliminate the low frequency advantage.

List length effect

The null list length effect was consistent with our assumption that the main interference comes from the pre-associations of the item cue. During the retrieval process, participants primarily use the given item to retrieve the target and use the context cue to reduce pre-experiment associations. Therefore, adding more pairs to the list will not influence the retrieval process for each individual pair. It would be hard for an item-noise approach to predict such an effect.

It was a novel finding because it is generally assumed that there is a list length effect because of the recall-like process in cued recall (e.g. Raaijmakers & Shiffrin, 1981). The reason why there is list length effect reported in some of the previous studies (e.g. Raaijmakers & Shiffrin, 1981; Nobel & Shiffrin, 2001) may be that they fail to eliminate the confounded variables that we discussed earlier. For example, in

Nobel and Shiffrin (2001), the distractor task lasts 26s for both short and long list conditions, which potentially may introduce a longer retention interval for the long list condition and thus reduce the performance. This null list-length finding indicates that the item noise from other pairs in the list might be negligible and needs to be considered when building models for cued recall.

On the other hand, the item noise approach could potentially argue that there is still a list length effect in cued recall. The reason that it is not observed in the current experiment is because the weak context cue failed to eliminate the interference from the pre-experiment memories. Under the competition of the large pre-experiment memories, the difference between 12 pairs and 48 pairs is difficult to be observed. The encoding task could potentially help to bring back the list length effect.

Because the presence of an encoding task may play a key role in explaining the different results of the present study and previous studies, we plan to further investigate by adding an encoding task to the current paradigm. We predict that the encoding task may strengthen the context cue and reduce the interference from previous memory and thus reduce the low frequency cue advantage. Whether we can observe a list length effect with encoding task may help us clarify the role of item noise in cued recall.

Conclusion

In summary, the experiment presented in this article suggests that there is no list length effect and word frequency effects exist on both the cue and the target in cued recall. The high frequency target advantage confirms previous findings and provides some evidence that high frequency words are easier to retrieve. The null list-length effect and low frequency cue advantage are inconsistent with the majority of previous findings on cued recall. However, the results are consistent with some other experiments that manipulated list length with control over confounded variables and question the traditional belief in the item-noise interference in cued recall. The low frequency cue advantage is consistent with our predictions and raises some interesting questions regarding the interactions between the context cue and the item cue in cued recall. How the encoding task and retention interval influence the context cues also requires further exploration. All the findings presented in the paper need further examination, but the results provide important constraints on a unified model of memory.

References

Clark, S., & Burchett, R. (1994). Word frequency and list composition effects in associative recognition and recall. *Memory & Cognition, 1994, 22(1), 55-62*

Criss, A., Aue, W., Smith, L. (2011). The effects of word frequency and context variability in cued recall. *Journal of Memory and Language, 64, 119-132*

Criss, A., & Malmberg, K. (2008). Evidence in favor of the early-phase elevated-attention hypothesis: The effects of letter frequency and object frequency. *Journal of Memory and Language, 59, 331–345.*

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

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

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

Dennis, S., Lee, M. D., & Kinnell, A. (2008). Bayesian analysis of recognition memory: The case of the list length effect. *Journal of Memory and Language*, *59*, 361-376.

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**, 8-20.

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

Madan, C., Glaholt, M., & Caplan, J. (2010). The influence of item properties on association-memory. *Journal of Memory and Language,63, 46–63.*

May, R., & Tryk, E.(1970). Word sequence, word frequency and free recall. *Canadian Journal of Psychology,24(5), 299-304*

Nobel, P., & Shiffrin, R. (2001). Retrieval process in recognition and cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, Vol 27(2), Mar 2001, 384-413*

Postman, L. (1970). Effects of word frequency on acquisition and retention under conditions of free recall learning. *Quarterly Journal of Experimental Psychology, 22,* 185–195.

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

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

Ward, G., Woodward, G., Stevens, A. & Stinson, C. (2003). Using over rehearsal to explain word frequency effects in free recall. *Journal of Experimental Psychology: Learning, Memory and cognition, 29(2), 186-210*