

# The Influence of Attention on the Detection of the List Length Effect in Recognition Memory

**Angela Kinnell (angela.kinnell@gmail.com)**

School of Psychology, Level 4, Hughes Building, North Terrace  
University of Adelaide, South Australia, 5005.

**Simon Dennis (simon.dennis@gmail.com)**

Department of Psychology, Ohio State University,  
1835 Neil Avenue, Columbus, OH, 43210 USA

## Abstract

The list length effect in recognition memory has been the subject of recent debate. Many studies have identified the effect, however Dennis and Humphreys (2001) argued that previous list length effect findings were the result of a failure to control for four potential confounds. The list length effect can be used to discriminate between item and context noise models of recognition memory. Item noise models predict the effect, while context noise models do not. In this paper, the role of attention on the detection of the list length effect is explored. The attention task at study was manipulated; participants either rated the pleasantness of study items or read the words only. In addition, the design was either retroactive or proactive. The results suggest that it is the proactive design in which the list length effect is evident. When the retroactive design is used in conjunction with the pleasantness rating task, there is the most even performance across list lengths and a nonsignificant effect of list length. This is consistent with context noise models of recognition.

**Keywords:** recognition; list length; memory models.

The source of interference in recognition memory has been the subject of recent debate. Some researchers have argued that interference to the memory trace for a particular item comes about as a consequence of the other items on the study list (e.g. Cary & Reder, 2003; Gillund & Shiffrin, 1984; Gronlund & Elam, 1994; Murdock, 1982). At the other end of the spectrum, Dennis and Humphreys (2001) have argued that interference comes from all of the previous contexts in which that particular item has been seen. Alternatively, interference could arise through a combination of the two (Criss & Shiffrin, 2004). Arising from these different approaches are a number of different mathematical models of recognition memory designed to fit several well-documented recognition findings. There has recently been some question as to whether one of these findings, the list length effect, is a real effect or is the result of one or more confounds (Dennis & Humphreys, 2001). The list length effect refers to the finding that recognition performance is superior for shorter rather than longer lists. The existence, or non-existence, of the list length effect can be used to help differentiate between these models of recognition memory.

## Item Noise Models

The item noise approach is based upon the idea that it is the other items on the study list that interfere with one's ability to recognize a test probe. There are numerous mathematical models of recognition memory that define interference as arising from other list items. These include global matching models (GMMs) such as the Theory of Distributed Associative Memory (TODAM; Gronlund & Elam, 1994; Murdock, 1982), Minerva II (Hintzman, 1986), the Matrix model (Pike, 1984) and Search of Associative Memory (SAM; Gillund & Shiffrin, 1984) as well as the Retrieving Effectively from Memory model (REM; Shiffrin & Steyvers, 1997) and the Subjective Likelihood Model (SLiM; McClelland & Chappell, 1998).

In these models a reconstruction of the study list context is used, often implicitly, to retrieve the items that appeared on the study list from memory. These are then compared to the test probe, with the level of activation integrated across all items. If this level of activation is above a certain threshold, the individual will respond "yes" indicating that they recognize the word from the study list. If this activation fails to reach the threshold, a "no" response will ensue. A longer study list means that more items must be matched to the cue resulting in greater interference as each new item introduces more variability. Thus, a list length effect is predicted by item noise models (Clark & Gronlund, 1996).

## Context Noise Models

Alternatively, interference could arise from the other contexts in which a word has been encountered in the past and any interference from other items is negligible (Dennis & Humphreys, 2001). Context noise models are much fewer in number than item noise models and include the Bind Cue Decide Model of Episodic Memory (BCDMEM; Dennis & Humphreys, 2001) and the model of Anderson and Bower (1972). In BCDMEM, the test probe is used to cue retrieval of previous contexts in which that word has been encountered. These contexts are then compared to the reinstated study context. The stronger the global match between these, the higher the level of activation with the result being a "yes" response. Failure to reach the activation threshold will result in a "no" response. Other study list items do not affect

performance as their memory traces are not activated at any point in the retrieval process.

It is also possible that neither item noise nor context noise alone can account for interference in recognition, but rather it may be that a combination of the two is involved (e.g. Criss and Shiffrin, 2004).

### **The List Length Effect**

The existence of the list length effect in recognition has been very well documented (e.g. Bowles & Glanzer, 1983; Cary & Reder, 2003; Gronlund & Elam, 1994) and as a result its existence has been somewhat ubiquitously accepted in the literature on recognition memory.

Nevertheless, a number of published studies have reported nonsignificant effects of list length (e.g. Jang & Huber, 2008; Murnane & Shiffrin, 1991; Schulman, 1974). Dennis and Humphreys (2001) argued that previous studies which had identified the list length effect had failed to control for four possible confounds; retention interval; attention; displaced rehearsal and contextual reinstatement. When they controlled for these confounds, they found no significant difference in recognition performance between a 24 word and a 72 word list. Controlling for the potential confounds seems to have had an impact on the results. Each of these confounds will now be discussed.

**Retention Interval** Retention interval is defined as the amount of time elapsed between the presentation of a target at study and again at test. The shorter retention interval for short list items would favor performance on that list and could result in a list length effect finding.

Retention interval can be controlled using either a retroactive or proactive design. In the retroactive design, a period of filler activity follows the short list such that the duration of the short list and filler is equal to that of the long list. In this design it is the first words of the long list (the same number as in the short list) that are tested. The proactive design is the converse of this with filler activity preceding the short study list and the last words of the long list being tested.

**Displaced Rehearsal** Displaced Rehearsal becomes an issue when retention interval is controlled and only some long list items are tested while all short list items are tested. In this case, any rehearsal of short list items will be beneficial to performance while there is no such guarantee with rehearsal of long list items. In addition, the filler following the short list in the retroactive design provides an opportunity to rehearse short list words, while in the long list words are continually presented.

Displaced rehearsal can be controlled by ensuring that the filler activity is engaging and/or making the recognition test incidental.

**Contextual Reinstatement** Reinstating the study context at test is important in both item and context noise models of recognition memory. The more accurate this reinstated study context is, the better the recognition performance is likely to be. Context varies with the passing of time with more scope for variability in the long list which would negatively impact performance.

In the retroactive design, the filler activity following the short list before the test list can act to encourage participants to reinstate the study context. When the test list immediately follows the long list, contextual reinstatement is not encouraged.

Including an extended period of filler activity between study and test for both long and short lists can encourage contextual reinstatement in both length conditions. Dennis, Lee and Kinnell (2008) found a nonsignificant effect of list length when contextual reinstatement was controlled in this way and a list length effect when no control was implemented. A Bayesian analysis of the same data favored the error-only model in both conditions. On this basis, it appears that while controlling for contextual reinstatement is important, it is not the most telling of the potential confounds.

**Attention** The final potential confound of the list length effect is attention. It is likely that participants will tire over the course of the long list to a greater extent than in the short list and pay less attention to the items. This is more problematic in the proactive design (Cary & Reder, 2003; Dennis & Humphreys, 2001; Underwood, 1978) when it is the final items of the long list that are tested and compared to performance on the short list.

Having participants perform an encoding task that requires a response during study can help control for lapses in attention as it ensures that all items will have been processed to some level, regardless of fatigue. Since there may be no way to completely eliminate attentional lapses in the proactive condition, using the retroactive condition may also control for attention.

### **The List Length Effect Revisited**

In 2003, Cary and Reder carried out a partial replication of Dennis and Humphreys' (2001) Experiment 1. Their third experiment included controls for the four potential confounds listed above. However, under these conditions, Cary and Reder (2003) identified a list length effect. There were a number of differences between the studies of Cary and Reder (2003) and Dennis and Humphreys (2001) that should be noted.

The first difference was that Cary and Reder's experiment involved a 1:4 (20:80 words) list length ratio while Dennis and Humphreys' study had a 1:3 (24:72 words) ratio meaning it is more likely that a length effect would be identified in the former experiment. Secondly, Cary and Reder included a two minute period of filler activity as a control for contextual reinstatement while Dennis and Humphrey had eight minutes of filler. The

study of Dennis et al. (2008) suggests that two minutes may not be sufficient to encourage contextual reinstatement after both lists. The two experiments also differed in the way that the results were analyzed. Cary and Reder collapsed results from both the retroactive and proactive design conditions together for analysis. Given that the length effect is more likely in the proactive condition due to lapses in attention, this may have influenced the results. The aim of the present experiment is to investigate the role of attention on the detection of the list length effect in recognition memory using a variation on the methods of Dennis and Humphreys (2001) and Cary and Reder (2003).

## The Experiment

### Method

**Participants** Participants were 160 Psychology students from the University of Adelaide. Each received either course credit or a payment of \$12 in exchange for their participation. All gave informed consent.

**Design** This experiment had a  $2 \times 2 \times 2 \times 2$  factorial design with the factors being list length (short or long), word frequency (low or high), attention task (pleasantness rating or read only) and design (retroactive or proactive). List length and word frequency were within subjects factors while attention task and design were between subjects manipulations. The word frequency manipulation was included as a check of the power of the experimental design.

**Materials** The stimuli for this study were 140 five and six letter words from the Sydney Morning Herald Word Database (Dennis, 1995). Half of the words were of high frequency (100-200 occurrences per million) and half were low frequency (1-4 occurrences per million). All lists had the same number of five and six letter, and high and low frequency words. All words were randomly assigned to lists with no participant seeing the same word twice, except for targets.

**Procedure** Participants were first given an overview of the study and allowed to practice the computerized sliding tile puzzle used as the filler task.

Participants studied one short (20 word) and one long (80 word) list, the same list lengths as in Cary and Reder's (2003) study. Each study word appeared for 3000ms. Test lists were made up of 20 targets and 20 distractors. All lists had half high frequency and half low frequency words. All words were presented in lower-case letters in the center of a computer screen.

Participants were split equally into two attention task conditions. In the pleasantness rating condition, participants were asked to rate the pleasantness of each word on the study list on a six point Likert scale (1: least

pleasant, 6: most pleasant) by clicking the appropriate button while that word was being displayed on screen. Participants were told that if they missed rating one of the words within the 3000ms they should rate the next word instead. In the read only task condition, participants simply read the words of the study list as they appeared on the screen. No response was required.

Within each condition, the design of the lists was either retroactive or proactive. Participants were again divided equally into these conditions. In the retroactive design, the short list was followed by a three minute period of sliding tile puzzle filler and the first 20 words of the long list were included as targets at test. In the proactive design, there was three minutes of puzzle filler before the beginning of the short list and the last 20 words of the long list were tested.

Participants were given 15 seconds notice before the onset of the test list which was in the form of the yes/no recognition paradigm. Each word was presented in the middle of the screen above two response buttons marked "yes" and "no". Participants were instructed to respond "yes" if they recognized the word from the study list and to respond "no" if they did not recognize that word by clicking on the appropriate button. The test list was self paced and a response was recorded for each test word. The targets were either the entire study list (short list), the first 20 words of the long study list (retroactive design) or the last 20 words of the long list (proactive design).

Contextual reinstatement was encouraged following both short and long lists with an eight minute period of sliding tile puzzle filler activity before each test list.

The experiment was counterbalanced for order, within each condition half of the participants began with the short list and the other half began with the long list.

### Results

The results of this experiment were analyzed using both a standard repeated measures analysis of variance (ANOVA) as well as the Bayesian analysis developed by Dennis et al. (2008; see this paper for further details). This method allows comparison of two competing models on a given set of data. With regard to the list length effect, we are interested in whether there is a systematic difference in discriminability in the short list compared with the long list. The two models are the error-only model (no systematic difference) and the error-plus-effect model. Inferences are made for each subject as to which model better reflects their performance. Results are reported as a pair of values indicating the posterior probability that at least 90% of subjects conform to the error-only or the error-plus-effect model, respectively. Note it is possible that neither of these results occur if, for instance, half of the subjects follow an error-only model and half follow an error-plus-effect model.

A major benefit of this method of analysis is that it allows for the accumulation of evidence in favor of the null hypothesis. This is important in the present case

where the failure to identify a significant effect of list length is also theoretically interesting.

**List Length Overall** A  $2 \times 2 \times 2 \times 2$  (length x frequency x task x design) repeated measures ANOVA yielded a nonsignificant effect of list length on  $d'$  ( $F(1,156) = 2.71, p = .1$ ) and the hit rate ( $F(1,156) = 1.21, p = .27$ ). However, there was a statistically significant effect of list length on the false alarm rate ( $F(1,156) = 11.01, p = .001, \eta_p^2 = .07$ ).

For comparison with the results of Cary and Reder (2003) a  $2 \times 2 \times 2$  (length x frequency x design) repeated measures ANOVA was carried out for the pleasantness task condition. Analysis revealed a nonsignificant interaction between list length and design on  $d'$  ( $F(1,78) = 2.09, p = .15$ ), the hit rate ( $F(1,78) = 3.70, p = .06$ ) and the false alarm rate ( $F(1,78) = .02, p = .89$ ). Cary and Reder (2003) obtained the same result and on that basis collapsed the retroactive and proactive conditions. In the present analysis, the conditions will remain separated.

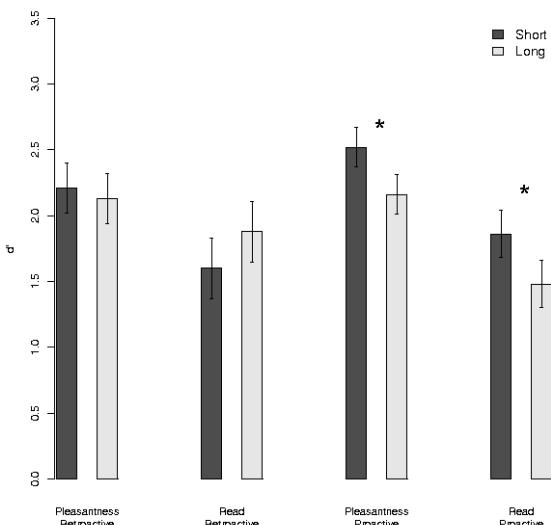


Figure 1:  $d'$  values for each of the four attention conditions. Bars represent 95% within subjects confidence intervals.

**Pleasantness Retroactive Condition** In the pleasantness retroactive condition, repeated measures ANOVAs found a nonsignificant effect of list length on  $d'$  ( $F(1,39) = .38, p = .54$ , see Figure 1 for short and long list  $d'$  in each condition) and the hit rate ( $F(1,39) = .21, p = .21$ ) while the effect on the false alarm rate was marginally significant ( $F(1,39) = 3.95, p = .054$ ). Similarly, the Bayesian analysis of the  $d'$  values found in favor of the error-only model (.81, .01).

**Read Only Retroactive Condition** Repeated measures ANOVAs in the read only retroactive condition yielded nonsignificant effects of list length on  $d'$  ( $F(1,39) = 3.06, p = .09$ ) and the false alarm rate ( $F(1,39) = .60, p = .44$ ).

There was, however, a statistically significant effect of list length on the hit rate ( $F(1,39) = 9.95, p = .003, \eta_p^2 = .20$ ). It should be noted, however, that in this condition, performance on the long list was superior to that of the short list, meaning that this result is significant in the opposite direction to that previously identified in the literature. Bayesian analysis of the  $d'$  values again found in favor of the error-only model (.78, .06).

**Pleasantness Proactive Condition** In the pleasantness proactive condition, repeated measures ANOVAs yielded statistically significant effects of list length on both  $d'$  ( $F(1,39) = 11.55, p = .002, \eta_p^2 = .23$ ) and false alarm rate ( $F(1,39) = 6.72, p = .013, \eta_p^2 = .15$ ). There was no significant effect on the hit rate ( $F(1,39) = 2.42, p = .13$ ). In contrast to the ANOVA  $d'$  results, the Bayesian analysis found in favor of the error-only model (.68, .13). Thus, while the ANOVA suggested a positive list length effect, the Bayesian analysis favored the error-only model, suggesting that it may be a minority of participants driving the effect in the ANOVA.

**Read Only Proactive Condition** A repeated measures ANOVA in the read only proactive condition yielded a significant effect of list length on  $d'$  ( $F(1,39) = 8.26, p < .001, \eta_p^2 = .17$ ). There was, however, a nonsignificant effect of list length on both the hit rate ( $F(1,39) = 2.40, p = .13$ ) and the false alarm rate ( $F(1,39) = 3.65, p = .06$ ), although the false alarm rate was close to significance. The Bayesian analysis of  $d'$  values was ambiguous for this condition (0.46, 0.27).

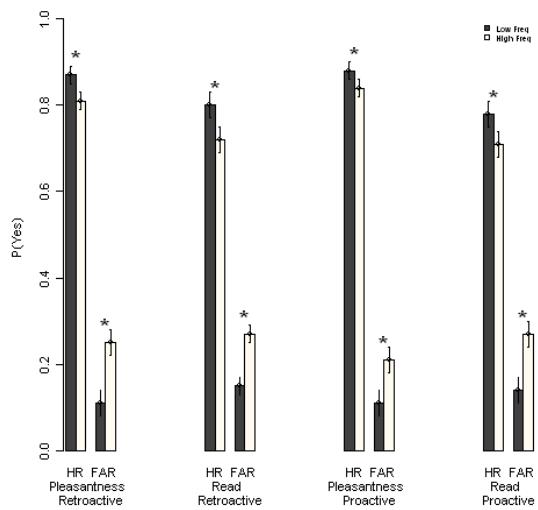


Figure 2: A significant word frequency effect was identified in both hit and false alarm rates in each condition. Bars represent 95% within subjects confidence intervals.

**Word Frequency** A  $2 \times 2 \times 2 \times 2$  (length x frequency x task x design) repeated measures ANOVA yielded a

significant effect of word frequency on  $d'$  ( $F(1,156) = 232.79, p < .001, \eta_p^2 = .60$ ) in the overall data. Planned comparisons were carried out on the word frequency data in each of the four conditions, collapsing across list length. There was a significant word frequency effect for  $d'$ , hit rates and false alarm rates in each of the four conditions using both the standard ANOVA analysis and the Bayesian analysis (see Figure 2 for hit and false alarm rate word frequency data). These findings suggest that the power of this experiment was not so poor that we were unable to detect effects of any kind.

## General Discussion

The experiment reported here aimed to explore the possible confounding effect of attention on the detection of the list length effect. Results indicated that attention does play a significant role. Most importantly, a list length effect was identified when the proactive condition was used as a control for retention interval but there was no significant effect when the retroactive design was used. This is not unexpected given that the proactive design involves the last 20 words of an 80 word list being tested as targets and this performance is compared to that of a 20 word short list. In this case, the amount of attention paid to each block of words is likely to differ and give rise to the list length effect. In the retroactive design, however, all targets in both test lists were presented at the beginning of both the long and short list study blocks where there should be no differences in attention.

In their third experiment, Cary and Reder (2003) included both retroactive and proactive designs. In their analysis, the data were collapsed across these conditions into one, based on the finding of a nonsignificant interaction between list length and study design. The results of experiment 1 suggest that collapsing the data in this way may have been problematic. We also found a nonsignificant interaction between list length and design and when we collapsed across the study design condition as Cary and Reder (2003) did, we also identified a positive list length effect when the pleasantness rating task was used at study. However, as has already been noted, a positive list length effect was identified in the proactive design only. This finding suggests that the nonsignificance of the interaction effect should not be used as justification for collapsing across conditions and that this may have altered the interpretation of Cary and Reder's (2003) findings. On the basis of the present experiment, it appears that the design of the experiment is important and that it is the proactive condition which drives the effect.

The results also indicated that the task (pleasantness rating vs. read only) used to control attention at study does not have a large influence on the list length effect finding. Nevertheless, the Retroactive Pleasantness condition, that is the most controlled condition, had the smallest effect size. Interestingly, the Retroactive Read condition had the first case, to our knowledge, of a reverse list length effect

with long list performance superior to that of the short list (although this effect was only marginally significant).

The current results support the assertion by Dennis and Humphreys (2001) that the history of positive list length effect findings may have been influenced by a failure to control for four possible confounds; retention interval, attention, displaced rehearsal and contextual reinstatement. When controls for these confounds are implemented, the effect of list length is nonsignificant. More specifically, using the retroactive design to control for retention interval, including a pleasantness rating, or similar, task at study in an attempt to hold the attention of participants and encourage them to process all words and the inclusion of an extended (eight minute) period of filler activity before the test list of both long and short lists leads to more equivalent performance across list lengths. Of course, the implementation of these controls may never be perfect. However, when all controls are in place, as the results of the present experiment has shown, there is no significant effect of list length in recognition memory.

## Implications for models of recognition memory

The absence of the list length effect in recognition memory presents some problems for existing mathematical models of memory and item noise models in particular. These models have been designed to accommodate and predict a positive list length effect finding. Conversely, the null list length effect finding lends support to context noise models of recognition which do not predict such an effect. The specific implications for several well known models of recognition memory will now be discussed.

**Global matching models** The GMMs all predict a list length effect in recognition memory. While these models differ from each other, they all predict the effect in a similar way. The test probe cues the retrieval of all study list items from memory which are then compared to the test probe. The results of these comparisons are summed to produce a global level of activation which is compared against a criterion in order to decide whether to produce a yes or a no response. While the effect of increasing list length on the means of the signal and distractor distributions is equivalent, the inclusion of more items results in an increase in the variance of these distributions, performance drops and a list length effect is predicted (Clark & Gronlund, 1996).

The present nonsignificant list length effect findings are problematic for the GMMs. The global matching process necessarily involves all retrieved study list items in the decision process and as such a list length effect is predicted under all conditions. Furthermore, it seems doubtful that these models can be easily modified to capture the null list length effect finding. If one reduces the amount of interference in these models to the insignificant levels required to capture the current results then the models would no longer be able to account for

the increases in the probability of saying "yes" to items presented at test that are similar to study items. The primary purposes of the introduction of the global matching assumption are undermined.

#### **Bind Cue Decide Model of Episodic Memory**

BCDMEM (Dennis and Humphreys, 2001) is a context noise model of recognition memory. The test word is the cue and all previous contexts in which that word has been encountered are retrieved from memory. If one of the retrieved contexts matches the reinstated study context a 'yes' response will result. The greater the number of contexts in which an item has been seen, the greater the interference and the poorer the recognition performance. This happens regardless of the length of the study list as other list items are not considered during retrieval. Thus, BCDMEM does not predict a list length effect and is consistent with the results of the experiments presented here.

### **Conclusions**

In summary, the results presented here suggest that there is no significant effect of list length in recognition memory when potential confounds are controlled. This finding is consistent with context noise models of recognition memory. This nonsignificant effect of list length, however, challenges the majority of existing mathematical models of recognition memory which assume that interference is a result of other list items. Even if one is unwilling to accept that there is no item interference at all in recognition for words, the current experiments demonstrate that the contribution of this kind of interference is extremely small. Certainly, item noise does not deserve to command the central role that it has in mathematical models of recognition over the past three decades.

### **References**

Anderson, J.R., & Bower, G.H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, 79, 97-123.

Bowles, N. L., & Glanzer, M. (1983). An analysis of interference in recognition memory. *Memory & Cognition*, 11, 307-315.

Cary, M., & Reder, L. M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *Journal of Memory and Language*, 49, 231-248.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, 3, 37-60.

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

Dennis, S. (1995). *The Sydney Morning Herald word database*. Available: <http://psy.uq.edu.au/CogPsych/Noetica>.

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

Gronlund, S. D., & Elam, L. E. (1994). List-length effect: Recognition accuracy and variance of underlying distributions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 1355-1369.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review*, 93, 411-428.

Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 112-127.

Knowlton, B. J., & Squire, L. R. (1995). Remembering and knowing: Two different expressions of declarative memory. *Journal of Experimental Psychology: General*, 21, 699-710.

McClelland, J.L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective likelihood approach to the effects of experience in recognition memory. *Psychological Review*, 105, 724-760.

Murdock, B. B., Jr. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, 89, 609-626.

Murnane, K., & Shiffrin, R. M. (1991). Interference and the representation of events in memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 855-874.

Pike, R. (1984). A comparison of convolution and matrix distributed memory systems. *Psychological Review*, 91, 281-294.

Schulman, A. I. (1974). The declining course of recognition memory. *Memory & Cognition*, 2, 14-18.

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

Underwood, B. J. (1978). Recognition memory as a function of length of study list. *Bulletin of the Psychonomic Society*, 12, 89-91.