The Null List Strength Effect in Recognition Memory: Environmental Statistics and Connectionist Accounts

Simon Dennis
Department of Psychology
The University of Queensland
Australia, 4072
mav@psych.psy.uq.oz.au

Abstract

In recognition paradigms, increasing the number of occurrences or presentation time in a study list of some words improves performance on these words (*the item strength effect*), but does not affect the performance on other words (*null list strength effect*). In contrast, adding new items results in a deterioration of performance on the other words (*list length effect*). Taken together these results place strong constraints on models of recognition memory. To explain these data an account based on optimisation to the environment is presented. A summary is given of environmental analyses which suggest that (1) the likelihood of recurrence of a word within a context increases as the number of occurrences increases; (2) the repetition rates of other words in a context has no significant effect on the recurrence probability of a word; and (3) the recurrence probability of a word drops as a function of the number of words since the last occurrence of that word. A training set which reflected these constraints was constructed and presented to an optimising connectionist network which was designed to extract recurrence statistics (the Hebbian Recurrent Network). The resultant model is able to model all three of the effects outlined above.

Introduction

The effect on memory of study lists that contain items of different strengths has been of interest for more than a century (Ebbinghaus, 1964; Thorndike, 1965; Tulving & Hastie, 1972; Ratcliff, Clark, & Shiffrin, 1990). The strength of an item is manipulated either by increasing the number of times it occurs in the study list or by increasing the length of time for which it is studied. Items of greater strength are generally retrieved more accurately than weak items (Ebbinghaus, 1964; Thorndike, 1965). Of more recent interest is what happens to performance on the weak items as a consequence of strengthening other items within the list. Table outlines the pattern of accuracy results for the weak items within different list types for recognition, cued recall and free recall.

For free recall, Tulving and Hastie (1972) demonstrated that performance on unstrengthened items dropped significantly in comparison to the AB and ABC control conditions. For recognition, however, strengthening an item does not degrade performance significantly on the non-strengthened items (known as the *null list strength effect,* Shiffrin, Ratcliff, & Clark, 1990; Ratcliff et al., 1990; Yonelinas, Hockley, & Murdock, 1992). This result contradicted the major mathematical memory models all of which predicted that performance on A in the ABB condition should be as poor if not worse than that in the ABC condition (Ratcliff et al., 1990).

Since the initial experimental work on the null list strength effect, a number of researchers have attempted to account for the data. Shiffrin et al. (1990) made two modifications to the SAM model. The first was to assert that when an item was strengthened, either by additional study time or additional study presentations, the same representation (or image) in memory was affected. Secondly, strengthening an item not only increased its context-to-item association, but also decreased its similarity to all other items. This process is called *differentiation* and provides an explanation of the null list strength effect.

The model by Chappell and Humphreys (in press), while implemented in a connectionist architecture, relies in part on a similar principle to SAM when explaining the null list strength effect. When an item is strengthened, context-to-central weights are updated (probabilistically). While the representations in the Chappell and Humphreys (in press) model are distributed rather than local, the effect is to increase the context-to-item association. At the same time, the weights of a central autoassociator are increased and a global inhibition is learned. The net effect of these manipulations is to make strengthened items more differentiated. Hence, the model accounts for the null list strength effect and demonstrates the difficulty in distinguishing between distributed and nondistributed models of human memory.

Heathcote (1994) has also used a connectionist model called Episodic ART to account for the list strength effect. The model relies on separating the explanatory mechanisms of the list length and list strength to different layers of the network. Decoupling these mechanisms allows the pattern of results to be modelled without altering the similarity structure of the representational space of the input words.

Another way of eliminating the list strength effect is to assume that pre-experimental items are also present in memory. This approach was used by Murdock and Kahana (1992) in their extension of the TODAM model. The effect of adding additional items is to swamp with noise any difference between the mean familiarity of an item that occurred with unstrengthened items and an item that occurred with strengthened items. While such a method can decrease the list strength effect to negligible levels, it also undermines the prediction of the list length effect and, hence, is not sufficient to explain the pattern of results (Chappell & Humphreys, in press). This example illustrates the importance of modelling the entire pattern of recognition results. It is the building of a model that captures the list strength effect while also accounting for the item strength and list length effects that has proven difficult.

The accounts outlined above rely on the memory mechanism to generate each of the effects. Another possibility is that it is the way in which these mechanisms interact with the environment which leads to the observed results. That is, in the course of their pre-experimental experience subjects come in contact with a large sample of words both written and spoken. The distribution of these words is far from uniform with a number of factors including syntax, semantics and context playing determining roles. If the subjects memory system is optimised to these statistics we would expect performance in the laboratory to reflect the predispositions which the subjects have acquired from their everyday experience. In the next section, the environmental statistics relevant to each of the phenomena are characterised.

Environmental Analysis

Item Strength

Anderson and Schooler (1991) examined a number of sources (*New York Times*, parental speech, electronic mail) to determine the probability that an item will occur given its frequency, recency and pattern of prior exposures. They found that the probability that an item will recur is linearly related to the frequency of the item (Anderson & Schooler, 1991) and that the constants of proportionality were near one in all three sources. It might be expected then that a mechanism optimised to the environment would show better performance on frequently presented items, as is the case.

One potential problem with the analysis which Anderson and Schooler (1991) undertook is that they did not control for the effect of general word frequency. In the laboratory it is the frequency within the study context that is manipulated. The linear relationship between occurrence frequency and recurrence rate, which Anderson and Schooler (1991) found, may have been a consequence of the general word frequency of the item. Yet performance generally decreases as general word frequency increases (see Dennis, 1993, for a discussion of the word frequency efffect).

To determine if frequency within the current context had an effect independent of general word frequency, Dennis (1993) took articles from the Minnesota Daily and assessed situational frequency independently of general word frequency. The results suggest that the probability of recurrence does increase with situational frequency.

List Length

In addition to frequency information, Anderson and Schooler (1991) studied the effect of retention interval. In this case, they plotted the probability of an item recurring as a function of the time since its last occurrence. They found that recurrence probability showed a similar negatively accelerating curve as seen in the laboratory when interval is manipulated. Furthermore, when both axes were subjected to a log transform a linear relationship emerged indicating that the original curve was a power function with an exponent around 0.75.

However, the effects of the interval between occurrences of an item may be a consequence of both the number of items in the context and the degree to which context has changed (c.f. Gillund & Shiffrin, 1984, analysis of context shifts and test delays in the experimental situation). The statistics collected by Anderson and Schooler (1991) were amalgamated over time by context. Three general contexts, namely, the *New York Times*, a database of children's verbal interactions, and mail messages received by John Anderson were analysed separately. While there may still have been some contextual drift within each broad context, the methodology would have minimised it. Further, in each of the analyses interval information was condensed using 100 item windows. Again this would work to minimise the amount of contextual drift. Hence, it is assumed that the number of unique items in a context makes a major contribution to the decrease in recurrence probability with delay. Consequently, a memory system optimised to these statistics might be expected to show a list length effect.

List Strength

The last of the environmental statistics to be considered relates directly to the null list strength effect. In environmental terms the effect can be translated to the rate of recurrence of a target word as a consequence of the rate of repetition of other words in the context. Dennis (1993) was unable to find any such effect in the Minnesota sample even when the item strength statistics were significant. The possibility arises then that the null list strength occurs because the memory system is optimised to these statistics.

Table 1: Accuracy results for nonrepeated items in different list types.

Test Type	AB/ABB	AB/ABC	ABB/ABC
Recognition	$=_{(<)}^{*}$	$>$	$>^{*}$
Cued Recall	$=_{(>)}^{*}$	$>$	$>^{*}$
Free Recall	$>$	$>$	$<$

Note. Asterisks indicate the results that were not predicted by the mathematical memory models. $=_{(<)}$ means equal or perhaps slightly worse. $=_{(>)}$ means equal or perhaps slightly better. AB is a short list without repeated items. ABB is a long list with repeated items. ABC is a long list without repeated items. This table is compiled from results by Tulving & Hastie (1972) and Ratcliff, Clark & Shiffrin (1990).

The Null List Strength Effect and the Hebbian Recurrent Network

In the previous section, the statistics relevant to the effect were summarised and it was postulated that in conjunction with an appropriate optimising memory system, they could account for the phenomenon. In this section, the Hebbian Recurrent Network (HRN, Dennis, 1993; Dennis & Wiles, 1993) is proposed as such a memory system. The HRN is trained with a data set constrained by the environmental statistics and its performance is then compared against that found in the experimental setting.

The conditions for the list strength simulations most closely resembled the recognition conditions of Ratcliff et al.'s (1990) experiment six. In this experiment, it was shown that the strength, length and null list strength effects could be obtained under the same conditions. The current simulations attempt to model these findings by constructing a training set which embodies the statistics outlined above, and training the HRN on this training set. There are three primary hypotheses:

Item Strength Effect: Items which are presented multiple times should be recognised more accurately than items presented once.

Length Effect: Items from lists containing many unique items should be recognised less accurately than items from lists with few unique items.

Null List Strength Effect: The strength of list items other than the test item should not affect performance on the test item.

Method

Figure 1 shows the architecture which was used. The backpropagation algorithm with a weight decay term (see Moody, 1992) was used to train the feedforward weights and the following Hebbian rule was used on the connections from the hidden units to the context units:

$$\Delta w_{ij} = a_i a_j - \lambda w_{ij} \quad (1)$$

where w_{ij} is the weight from unit i to unit j, Δw_{ij} is that change made to this weight, a_i is the activation of unit i, a_j is the activation of unit j and λ is the memory decay rate. The network was trained with a learning rate of 0.05, a memory decay rate of 0.2 and a weight decay of 0.000002. Initial weights were selected from a uniform distribution between -1 and 1 and then updated after each pattern presentation. Twenty simulations were run for 1500 epochs by which time the performance had reached asymptote. Initial weights, and training and test sets (containing 1000 sequences each) were generated for each simulation.

The first step in constructing each of the training sequences was to decide on the study list type. There were four list types (i.e. pure weak, mixed, pure strong and length). Pure weak lists consisted of two items presented once each. Mixed lists consisted of two items. One was presented once and the other twice. Pure strong lists also contained two items both of which were presented twice.

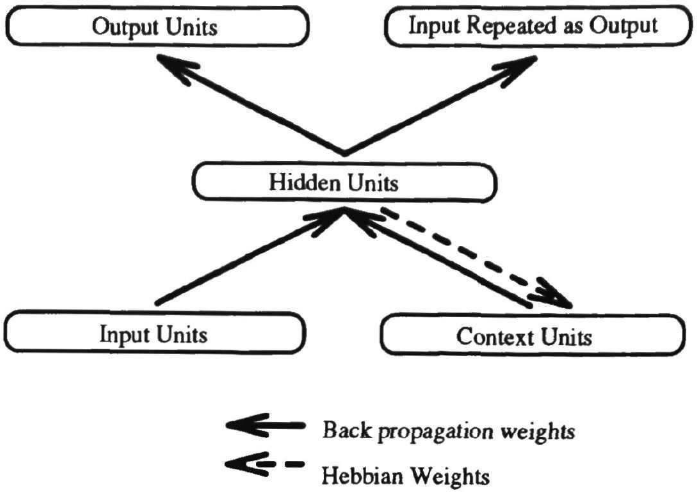

Figure 1: The HRN architecture for the list strength simulations. The architecture is a Hebbian Recurrent Network (HRN). The input vectors consisted of seventeen components. All items were locally encoded. Sixteen of the input units represented words while one was a "pause" signal which indicated that no input was being presented at that time. The "pause" input was presented when the decision whether the test item was in the study list was being made. Of the outputs 17 were the inputs repeated and the other three coded for "blank", "yes" and "no". Either "Yes" or "No" was expected when the recognition decision was made and the "Blank" pattern was expected at all other times.

The length condition contained eight items each presented once. Items were chosen with equal probability and the order of the study items was randomised. The test item was chosen with probabilities which depended upon condition. Table 2 provides examples of each of the conditions and Dennis (1993) provides a quantitative outline and justification of the training set statistics.

Results

Figure 2 shows the learning curves (i.e. d' as a function of the training epoch) for each of the conditions. A two way analysis of variance was conducted. Strong items were found to be better recognised than weak items, $F(1, 19) = 238.83$, $p < 0.001$, as illustrated by strong-to-weak ratios of 1.88 and 1.92 for pure and mixed lists respectively. There was no significant difference between pure

List Type	Condition	Study List	Test Item
Pure Weak	Pure Weak	GB	G
Mixed	Mixed Weak	CFC	F
	Mixed Strong	GGE	G
Pure Strong	Pure Strong	DJDJ	J
Length	Length	HAIDBGLE	I

Table 2: Example target lists from the list strength training set.

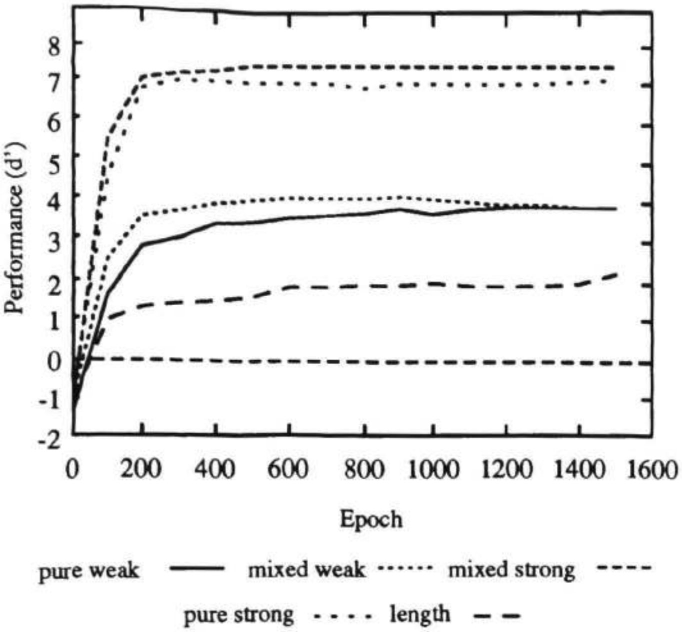

Figure 2: Training epoch verses performance (D') of the HRN for each of the list strength conditions. The strong conditions show a better performance than the weak conditions which show better performance than the length condition.

and mixed lists, $F(1,19) = 0.46$, $p = 0.50$. Further, there was no significant list strength effect as assessed by the interaction term, $F(1,19) = 0.36$, $p = 0.55$.

In addition to considering the interaction of the pure/mixed and strength factors, the analysis of the list strength effect has included the calculation of a ratio of ratios measure given by the following formula (Shiffrin et al., 1990; Ratcliff et al., 1990):

$$R_r = \frac{d'(mixed\ strong)/d'(mixed\ weak)}{d'(pure\ strong)/d'(pure\ weak)} \tag{2}$$

If weak items are not impaired by strengthening other items in the list (and conversely, if strong items do not improve by weakening other items in the list), the ratio of ratios measure will be near one.

Figure 3 shows the list strength ratio as a function of training epoch. Before training, the ratio of ratios is well above one. At 100 epochs the list strength ratio is below one since the performance on pure weak items lags behind that on the mixed weak items. As training progresses it rises to approximately one where it remains until 1500 epochs where it is 1.02.

Analyses of variance comparing the length condition against the two weak conditions were conducted. In both cases the performance in the length condition was significantly poorer than in the weak conditions (pure weak, $F(1,19) = 8.32$, $p < 0.01$; mixed weak, $F(1,19) = 17.14$, $p < 0.001$).

Discussion

As outlined in the results section the item strength and list length results were reproduced by the model while

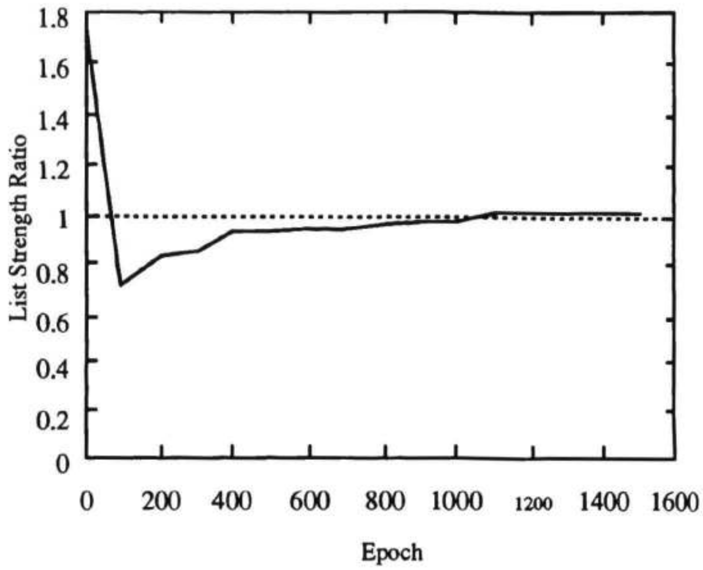

Figure 3: Training epoch verses the list strength ratio of the HRN. The list strength ratio rises gradually to near one indicating that there is no list strength effect.

no list strength effect was observed. Hence, the three hypotheses were upheld in accordance with human experimental data (Ratcliff et al., 1990).

In each case, these constraints reflect the findings of the environmental analysis and hence, the suggestion is that the human memory system demonstrates the pattern of performance results that it does because it is optimised to the environmental frequencies of recurrence.

Acknowledgements

This paper was supported by an Australian Postgraduate Research Award to the author. I would also like to acknowledge the supervision of Dr. Janet Wiles.

References

Anderson, J. R., & Schooler, L. J. (1991). Reflections of the environment in memory. *Psychological Science, 2*(6), 396–408.

Chappell, M., & Humphreys, M. S. (in press). Analysis of a neural network with application to human memory modelling. *Journal of Experimental Psychology.*

Dennis, S. (1993). *The Integration of Learning into Models of Human Memory.* Ph.D. thesis, The University of Queensland.

Dennis, S., & Wiles, J. (1993). Introducing learning into models of human memory: The hebbian recurrent network. In *Program of the Fifteenth Annual Conference of the Cognitive Science Society*, pp. 394–399. Lawrence Erlbaum Associates.

Ebbinghaus, H. (1964). *Memory: A Contributon to Experimental Psychology.* New York, NY: Dover Publications. Translated by H A Ruger and C E Bussenius. Original written in 1885.

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

Moody, J. (1992). The effective number of parameters. In J. E. Moody, S. J. Hanson, & R. P. Lippmann (Eds.), *Advances in Neural Information Systems 4*, pp. 847–854. San Mateo, CA: Morgan Kaufmann Publishers.

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

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

Thorndike, E. L. (1965). *Animal Intelligence: Experimental Studies.* New York, NY: Hafner Publishing Company. Original published in 1911 by Teachers College, Columbia University.

Tulving, E., & Hastie, E. (1972). Inhibition effects of intralist repetition in free recall. *Journal of Experimental Psychology, 92*(3), 297–304.

Yonelinas, A. P., Hockley, W. E., & Murdock, B. B. (1992). Tests of the list-strength effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*(2), 345–355.