When Does Semantic Similarity Help Episodic Retrieval?

Marc W. Howard Michael J. Kahana
Volen Center for Complex Systems
Brandeis University

In Press, *Journal of Memory and Language*

Free recall illustrates the spontaneous organization of memory. This organization comes in two forms: the temporal organization of the list and the semantic relations among list items. With estimates of semantic similarity provided by latent semantic analysis (LSA, T. K. Landauer and S. Dumais, 1997), we simultaneously assessed the effects of temporal and semantic factors on output order in delayed and continuous-distractor free recall of random word lists. These analyses revealed that LSA-based estimates of semantic similarity predict recall transitions in both delayed and continuous-distractor free recall. Furthermore, increasing the duration of the inter-item distractor from 2-16 seconds reduced the effect of semantic similarity on recall transitions. Subsequent analyses argue against an encoding-based explanation of this phenomenon. This result, coupled with the approximate scale invariance of the lag-recency effect (M. W. Howard and M. J. Kahana, 1999) present a strong challenge for models of free recall and episodic memory retrieval.

Recent years have seen a resurgence of interest in the dynamics of retrieval in free recall (Howard & Kahana, 1999; Kahana, 1996; Kahana & Wingfield, 2000; Rohrer & Wixted, 1994; Romney, Brewer, & Batchelder, 1993; Wingfield, Lindfield, & Kahana, 1998). In studying the memory processes involved in free recall, researchers have typically employed one of two general approaches. In the first approach, the experimenter manipulates what we will refer to as temporal factors; in the second approach the experimenter manipulates semantic factors. Temporal factors describe the structure of the learning episode and its effect on memory performance. This would include for example, recency, the relation between the time an item is studied and the time of its test (Nairne, Neath, Serra, & Byun, 1997; Neath, 1993), and contiguity, the relations between the study times of list items (Howard & Kahana, 1999; Kahana, 1996; Serra & Nairne, in press). Semantic factors, in contrast, refer to pre-existing relations among to-be-remembered stimuli. Studies focusing on semantic factors typically employ word association norms or categorized word lists to examine the effect of semantic structure on the dynamics of retrieval (e.g. Bousfeld, 1953; Glanzer, Koppenaal, & Nelson, 1972; Kahana & Wingfield, 2000; Pollio, Richards, & Lucas, 1969; Tulving & Pearlstone, 1966). Analyses of category clustering, response bursting, and subjective organization all fall within this general approach. Although these two approaches are by no means mutually exclusive, studies of semantic factors have not simultaneously examined temporal factors. Conversely, studies of temporal factors have not simultaneously examined semantic factors.

This omission poses a problem—failure to understand the joint effects of semantic and temporal factors limits our understanding of both. For instance, categorized word lists are often used to examine the role of semantic factors on episodic retrieval. Yet, in these tasks, performance depends substantially on whether the categories are presented in blocked or randomized fashion—a manipulation of temporal organization (e.g., D'Agostino, 1969). Conversely, when one studies temporal factors, one tries to assemble random lists of words that are as unrelated as possible. Even in such unrelated word lists semantic factors can influence organization in retrieval (Schwartz & Humphreys, 1973; Tulving, 1962). Failure to take into account semantic structure in random word lists thus limits our understanding of the role of 'purely' episodic factors in retrieval.

This paper attempts to unify these two approaches by examining the interacting roles of semantic and temporal factors in free recall of random word lists. Because experiments on free recall of random word lists typically utilize large pools of words (to avoid the need for within-session repetition of items), we need a method that can provide data on semantic similarity for a very large number of word pairs. Therefore, we employ a measure of similarity derived from the application of Landauer's latent semantic analysis (LSA, Landauer & Dumais, 1997) to the words used in our experiments.

The authors acknowledge support from NIH research grants MH55687 and AG15852. We are grateful to Tom Landauer and Darrell Laham for kindly providing us with the raw LSA vectors used in this paper's analyses. Correspondence concerning this article should be addressed to Michael Kahana, Volen National Center for Complex Systems, MS 013, Brandeis University, Waltham, MA 02254-9110. Electronic mail may be sent via Internet to kahana@brandeis.edu.

Semantic Factors in Free Recall

In free recall of categorized word lists, subjects tend to recall words from the same natural category together, even when presentation order is randomized. In addition, IRTs to words in the same category as the just-recalled word are faster than those to words from a different category, a phenomenon known as response bursting (Patterson, Meltzer, & Mandler, 1971; Pollio et al., 1969; Wingfield et al., 1998). The pre-experimental (semantic) associations among list items thus exert a powerful influence on both output order and latency.

To examine the influence of pre-experimental (semantic) associations on retrieval, we must first have a way of determining the associative strengths among list items. In studies of categorized free recall, any two words can be thought of as having an associative strength of 1.0 if they come from the same category, or an associative strength of 0.0 if they come from different categories. There is obvious benefit in extending this analysis to lists that do not fall into natural categories. For instance, it could be that the effect of semantic similarity is limited to mediation via category names (Tulving & Pearlstone, 1966). To analyze the effects of semantic structure when subjects recall lists of random (non-categorized) words requires a fine-grained measure of semantic similarity. Such a measure can be derived from subjective similarity ratings obtained for all possible pairs of words used in a given experiment. These ratings can be obtained using triadic judgments (e.g. Romney et al., 1993), card sorting (e.g. Schwartz & Humphreys, 1973), or free association (e.g. Bousfeld, 1953; Nelson, Schreiber, & McEvoy, 1992; Nelson, McKinney, Gee, & Janczura, 1998). The proximity matrix produced using any of these methods can then be used to generate various hypotheses about the underlying associative structure of the materials. LSA, the alternative we use here, has not previously been applied to free recall.

Temporal Factors in Free Recall: The Lag-Recency Effect

Output order in free recall is affected by episodically-formed inter-item associations. These associations are inferred from subjects' tendency to successively recall items from nearby list positions. Kahana (1996) introduced a measure of this tendency. Conditional on a subject successively recalling two list items, the *conditional response probability* characterizes some property of these transitions.

To measure the effect of temporal factors on these recall transitions, we can define the variable *lag*. Given that the subject has just recalled an item from serial position i, and that the next recall is from serial position j, *lag* is defined as $j-i$. Positive values of *lag* correspond to forward recalls; negative values of *lag* correspond to backward recalls. Large absolute values of *lag* correspond to words spaced widely in the list; small absolute values correspond to words close together in the list. For example, if the list had contained the sub–sequence "ABSENCE HOLLOW PUPIL" and a subject recalled "HOLLOW" then "PUPIL", the recall of "PUPIL" would have a lag of +1. If, instead, the subject recalled "HOLLOW" then "ABSENCE", the recall of "ABSENCE" would be associated with a lag of -1. In this case, the subject is 'moving backward in the list'. "ABSENCE" followed by "PUPIL" would yield a lag of +2.

The *Lag-CRP*, measures the distribution of successive recalls as a function of *lag*, and thus provides a measure of contiguity effects in free recall.

As expected, the Lag-CRP reveals that successively recalled items are more likely to come from nearby serial positions than from remote serial positions. We refer to this property as the *lag-recency effect* (Howard & Kahana, 1999; Kahana, 1996). This effect shows a marked asymmetry: forward recalls are much more likely than backward recalls (Howard & Kahana, 1999; Kahana, 1996). The Lag-CRP provides a convenient measure of episodically formed inter-item associations.

Figure 1 provides examples of several Lag-CRP's calculated from data reported in Howard and Kahana (1999). Because the current work reports secondary analyses on data previously reported as Experiment 2 of Howard and Kahana (1999), we here provide a brief summary of the methods. Over 10 sessions, each of sixteen participants studied and attempted free recall of 150 different lists, each consisting of 12 words sampled at random from the noun subset of the Toronto word pool (Friendly, Franklin, Hoffman, & Rubin, 1982). For each trial, words were presented visually at a rate of 1 word / 1.2 s. To minimize rehearsal, participants were required to perform a semantic orienting task (judging each presented word as either concrete or abstract). Participants were then given a 16 second arithmetic distractor task prior to attempting free recall. The recall period was fixed at 60 s.

The key variable in this study was the duration of an arithmetic distractor task between the presentation of successive list items (this task was identical to the end-of-list distractor task). Across four conditions, the duration of between-item distractor activity, the interpresentation interval (IPI) was 0 s (No IPI, standard delayed free recall), 2 s, 8 s or 16 s (Longest IPI, continuous-distractor). The key result of Howard and Kahana (1999) was that the CRP functions for the no-IPI (IPI=0 s) and the longest-IPI (IPI= 16 s) conditions were highly similar, in contrast to the predictions of an account of the lag-recency effect based associations mediated by a limited capacity short-term store (e.g., the search of associative memory, Raaijmakers & Shiffrin, 1980).

Latent Semantic Analysis as a Measure of Semantic Similarity

In this paper we use a fine-grained measure of pre-experimental associations between words derived from LSA (Landauer & Dumais, 1997). LSA is particularly useful for the present application because it can provide an estimate of similarity for each of the pairs in our experimental pool without additional data collection.

LSA is based on the assumption that words that are similar in meaning tend to be used in similar contexts. If this is so, then the statistical properties of words in a large of body of

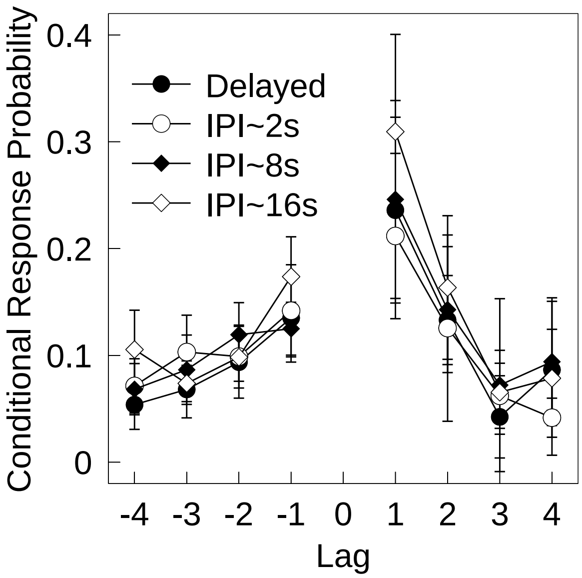

Figure 1. The approximate scale invariance of the lag-recency effect. Shown is the conditional response probability (CRP) function for each of the four conditions of Experiment 2 of Howard & Kahana (1999). The fact that each of these curves is peaked around 0 illustrates the lag-recency effect. The advantage for positive values (e.g., compare lags of $+1$ with -1) indicates that forward recall transitions are more likely than backward recall transitions. The parameter in this figure is the inter-presentation interval (IPI), which was varied between 0 seconds (the curve labeled 'Delayed') and 16 seconds. During this IPI, subject performed a demanding arithmetic task. Despite 16 seconds of inter-item distractor, the CRP is largely unaffected. Error bars are 95% confidence intervals.

naturally-occurring text should tell us something about their relationships. LSA provides a technique for evaluating the relationships between words based on such bodies of text.

Consider a corpus of naturally-occurring text, say, an encyclopedia, which contains *N* unique words (*N* will typically be on the order of tens of thousands). Define 'context' as a paragraph. Using the *M* contexts in the corpus, LSA proceeds as follows. First, form the $N \times M$ matrix that records how often each word occurs in each context.[1] The *M*-dimensional vectors (rows in our $N \times M$ matrix) representing related words will be more similar to one another than those representing unrelated words. That similarity is measured by the cosine of the angle between each pair of vectors. Although one could, in principle, compute similarities among words by measuring the correlation between their *M*-dimensional vectors, this is suboptimal because the dimensions of *M* will be highly correlated.[2] Using singular value decomposition, an algorithm similar to factor analysis, we can transform the $N \times M$ co-occurrence matrix into an $N \times D$ matrix, where *D* is a set of orthogonal (uncorrelated) dimensions much smaller than the original *M* contexts. We can then re-express the words as vectors of dimension *D*, and recompute the similarities. Landauer and Dumais (1997) found that LSA performs best using about $D = 300$ dimensions. The extent that two vectors "point" to the same region in LSA-space provides an operationally-defined measure of the semantic similarity of the corresponding words.[3] This can be measured by taking the cosine of the angle between the vectors, $\cos\theta$. In the analyses reported in this paper, we used LSA vectors of 300 dimensions drawn from the TASA-All[4] space and weighted by the singular values.

Analysis 1: LSA $\cos\theta$ Predicts Output Order in Free Recall

The CRP as a function of lag measures the effect of episodically-formed associations on retrieval (Kahana, 1996; Howard & Kahana, 1999). Here we generalize this method, reporting the CRP as a function of semantic similarity. This new measure, the *LSA-CRP*, will enable us to measure the effect of pre-experimental associations on retrieval. The primary goal of these first analyses is to demonstrate the effectiveness of this tool. We do so by demonstrating that LSA predicts output order and latency in free recall. Because subsequent analyses will use measures derived from the LSA-CRP, this also provides an opportunity to become comfortable with this measure and its properties.

Calculating the LSA-CRP

Ideally, what we would like to do is plot the probability of a recall transition as a function of the LSA $\cos\theta$ between the successively recalled words. Because $\cos\theta$ is a continuous measure (like age or height) it is useful to discretize $\cos\theta$ values by assigning them to different bins for the purposes of making figures, and to preserve the analogy with the Lag-CRP. One way to do this is to divide up the range of $\cos\theta$s (-1 to 1) into bins of some small fixed interval. For each pair of words recalled in succession, we would see what bin their $\cos\theta$ falls into. To compute the probability of recalling a word with a particular $\cos\theta$ to the just-recalled word, we simply increment the numerator for that bin, and increment the denominator for all the bins. If the set of possible recalls

[1] Landauer and Dumais (1997) actually log transform the entries in this matrix, but this is not necessary to get the main idea. For a more thorough treatment and discussion see (Landauer & Dumais, 1997; Landauer, Foltz, & Laham, 1998).

[2] That is, the *M* basis vectors form an over-complete set.

[3] One might argue that LSA does not capture what we ordinarily think of as semantic similarity. For instance, words that occur in similar contexts may have a quite high $\cos\theta$ despite bearing no resemblance to each other (e.g., TELESCOPE-STAR). Much the same criticism could be made of free association norms. In this paper, our primary interest is in comparing and contrasting the effects on memory retrieval of relatively permanent, structural relationships between words with the effects of the transient structure of the learning episode. LSA, although not measuring semantic similarity on the basis of objects' properties, does measure characteristics of the permanent structural relationships between words, and is probably highly correlated with any measure of 'pure semantics' one might choose to construct. We will not concern ourselves further with these distinctions, but simply treat LSA as an operationally defined measure of semantic similarity.

[4] http://128.138.223.70/spaces.html

Table 1
For selected $\cos\theta$ *bins in the Toronto Noun Pool, this table shows pairs of words in that bin, and the* $\cos\theta$ *between them. The pairs in the table were chosen quasi-randomly. Only very high bins contain predominantly pairs with obvious semantic relationships between them.*

Bin	Pair	$\cos\theta$
1	FAILURE-SPIDER	-0.077
⋮		
10	RECEIPT-LIQUID	0.003
⋮		
30	MUSIC-BARGAIN	0.031
⋮		
50	WINDOW-DISTANCE	0.062
⋮		
70	NUMBER-JOURNAL	0.109
⋮		
75	DIAMOND-IRON	0.122
⋮		
80	OYSTER-COUPLE	0.144
⋮		
85	BUBBLE-MOMENT	0.168
⋮		
90	BUTCHER-DINNER	0.203
91	PONY-FOREHEAD	0.209
92	AUTUMN-COLOR	0.219
93	SUBJECT-RESEARCH	0.232
94	WRINKLE-LEATHER	0.247
95	CRYSTAL-SILVER	0.257
96	WATER-ANCHOR	0.278
97	MAJOR-PROJECT	0.299
98	FURY-BULLET	0.328
99	FINGER-BUTTON	0.360
100	SUCCESS-FAILURE	0.549

at a given output position is evenly distributed among the different bins, then this simple approach would work fine. There are, however two problems.

The first problem with this approach is that the distribution of $\cos\theta$ across the interval -1 to 1 is non-uniform. It turns out that LSA $\cos\theta$, for the pool of words used in our experiments, has a skewed distribution, with the mode very close to zero and the mean at about .09. To correct for the *a priori* distribution of $\cos\theta$, we transform $\cos\theta$ for each pair into a percentile score, thus dividing the $\cos\theta$ distribution into 100 bins containing an equal number of pairs. To illustrate this, and in an attempt to provide some intuition into $\cos\theta$, Table 1 shows representative pairs that fall into the different percentile bins. Only bin 100 contains exclusively what most observers would consider obvious semantic associates. Pairs in bins 90-99 generally have some basis for a semantic relationship (e.g., WATER-ANCHOR from bin 96), although in some cases this requires a bit of imagination (e.g. FOREHEAD-PONY from bin 91). As we move to lower values of $\cos\theta$, the bins become more closely spaced and the pairs in the bins become less and less obviously related.

A second problem stems from the fact that lists used in free recall experiments, and in particular the experiments reported here, are relatively small compared to the resolution with which we would like to measure $\cos\theta$. Specifically, consider what would happen in an imaginary free recall experiment in which the subjects remember all of the items in each list and use semantic similarity exclusively to organize their output. That is, they choose the first word randomly, and then proceed by selecting the remaining word that has the highest similarity to the just-recalled word. Further assume that LSA provides a perfect measure of the subjects' idea of semantic similarity in this imaginary experiment. Under these ideal conditions, the probability of recalling a word in the very highest $\cos\theta$ bin should be 1.0 if the probabilities have been calculated correctly. However, if we simply incremented the denominator for each of our 100 bins at each recall attempt the result would not be a probability of 1.0 for the highest bin, but rather a probability given by the number of potential recalls divided by the number of bins. For example, if we had a list length of 11, we would get a probability of 0.1. [5]

An analogous problem arises in calculating the Lag-CRP. For instance, if the first word in the list is recalled, and then another word, it doesn't make sense to increment the denominators for the backward recalls ($lag < 0$). Because the just-recalled word is the first on the list, there is no possibility of a valid backward recall. As a consequence, we only increment the denominators corresponding to the lags of *possible* recalls in calculating the Lag-CRP. Similarly, in calculating the measure used here, the LSA-CRP, we increment the denominators for the bins associated with the set of available words from the list. A word is considered 'available' if: (a) The word was in the present list, and (b) the word has not previously been recalled on this trial. Each denominator may be incremented no more than once per word recalled.

Results

Accuracy. Figure 2A shows the LSA-CRP plotted as a function of the mean $\cos\theta$ value of each bin. The data used to generate this figure were collapsed over all conditions of Experiment 2 in Howard and Kahana (1999). As can be seen, subjects are more likely to make transitions to words with a higher value of $\cos\theta$ relative to the just-recalled word. A linear regression of LSA-CRP(bin) to $\overline{\cos\theta}(bin)$ was performed for each subject. The mean slope, $.21 \pm .02$ was reliably different from zero, $t(15) = 9.1,\ p < .001$. The intercept was .11. LSA $\cos\theta$ has a large effect on recall transitions in free recall: going from $\cos\theta = 0$ to $\cos\theta \simeq 1$, the probability of

[5] This is, of course, only exactly true in the expectation. The imaginary experiment also collected an infinite amount of data.

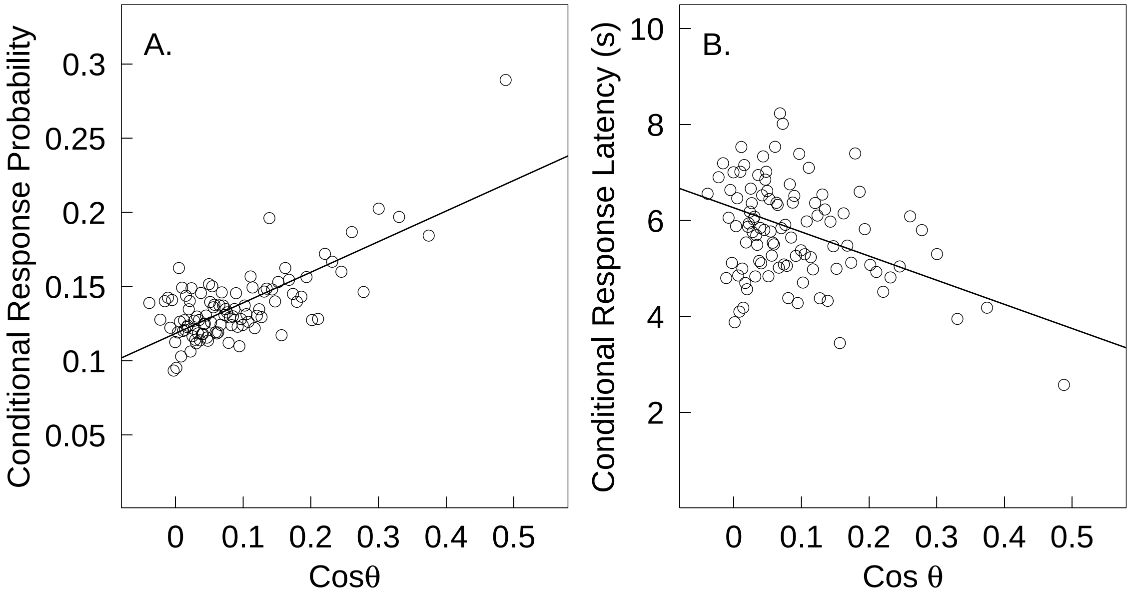

Figure 2. LSA predicts output order and response latency in single trial free recall. Data is from Experiment 2 of Howard and Kahana (1999). The distribution of LSA $\cos\theta$ was divided into 100 bins with an equal number of pairs (see Table 1). A. The conditional response probability as a function of mean LSA $\cos\theta$ for each bin. A linear regression was performed for each subject. The line in the figure was generated by taking the mean slope and intercept. B. The conditional response latency as a function of mean LSA $\cos\theta$ for each bin. A linear regression was performed for each subject. The line in the figure was generated by taking the mean slope and intercept.

a recall transition increases by about a factor of two. The correlation between LSA-CRP and $\cos\theta$ was .75.

The significant relationship between $\cos\theta$ and the CRP is not just a consequence of the highest bins, which contain predominantly pairs with an obvious semantic relationship. We recalculated the slope of the regression excluding upper portions of the distribution in 5-bin increments. When we excluded the 20 highest bins (excluding all pairs with $\cos\theta >$ .141), the slope was reduced to $.06 \pm .03$, but remained significantly different from zero, $t(15) = 2.24,\ p < .05$. This means that the LSA-CRP is sensitive to rather subtle variations in semantic similarity (compare Table 1).

Latency. LSA $\cos\theta$ also affects IRTs in free recall. Figure 2B plots mean IRT as a function of the bin of the $\cos\theta$ distribution. IRTs are shorter when the successively recalled words are similar (i.e. have high $\cos\theta$). The mean of the slopes of a linear regression, $-5 \pm 1s$, was significantly different from zero, $t(15) = 3.85,\ p < .01$. The correlation was -.37. It is well-known that IRTs increase with output position in single-trial free recall (Murdock & Okada, 1970; Rohrer & Wixted, 1994). To ensure that the effect of $\cos\theta$ on IRTs is not confounded with output position, we examined IRTs for just the first pair of words that subjects recalled. Even under these conditions, a regression revealed a significant negative slope of -6.3 s which was also significantly different from zero, $t(15) = 2.2,\ p < .05$. This constitutes a parametric relationship between semantic similarity and latency in free recall. Although there are a great many studies showing an effect of category membership on latency in free recall, this is, to our knowledge the first report of an effect of such subtle gradations of meaning on latency in free recall.

Discussion

The similarity of LSA vectors corresponding to pairs of words, as measured by the cosine of the angles between these vectors, is highly predictive of output order in free recall. Using a novel measure of this effect, the LSA-CRP, we demonstrated that very high similarity pairs are about twice as likely to be recalled in succession as very low similarity pairs. Further, LSA $\cos\theta$ affects latency in free recall. High-similarity pairs are recalled on average several *seconds* faster than low-similarity pairs.

Analysis 2: The Effect of Semantic Similarity on Output Order in Continuous-distractor Free Recall

Howard and Kahana (1999) examined the lag-recency effect in the continuous distractor paradigm (Bjork & Whitten, 1974). In this paradigm, list items, rather than being presented one after another, are separated by a period of distractor activity (e.g., mental arithmetic). Manipulating the length of this distractor (the IPI) alters the absolute time between list items while preserving the relative spacing of the list.

Figure 1 illustrates the lag-recency effect for several levels of the IPI in the study of Howard and Kahana (1999). As can be seen in the figure, the lag-recency effect was not reduced by increasing the IPI between 0 seconds (standard delayed free recall) and 16 seconds. Although this amount of distractor activity had essentially no impact on the lag-recency effect, the same amount of distractor activity, presented at the end of the list, was sufficient to eliminate the end-of-list recency effect (Howard & Kahana, 1999).

Insofar as the lag-recency effect is insensitive to the absolute delay between list items (Figure 1), it can be said to exhibit a scale-invariance with respect to time. Prior to this discovery, the lag-recency effect was interpreted as evidence for associations formed in short-term memory (Kahana, 1996). If short-term memory is the locus for episodically-formed associations (as postulated by Atkinson & Shiffrin, 1968; Glanzer, 1972; Raaijmakers & Shiffrin, 1980), then this would predict the lag-recency effect because nearby items presumably spend more time together in short-term memory than remote items. However, because a long inter-item distractor should disrupt short-term memory, the scale-invariance of the lag-recency effect requires an alternative hypothesis.

Because the continuous distractor paradigm provides a means of dramatically varying the temporal relations among list items, this paradigm provides a suitable testing ground for the analysis of temporally defined associations. This manipulation of the temporal relations among list items should not change the pre-experimental (semantic) associations among these items. However, it is possible that such a temporal manipulation modulates the efficacy of pre-experimental associations as retrieval cues.

The effects of semantic variables in continuous-distractor free recall have not been well-studied (but see Greene & Crowder, 1984; Greene, 1986; Gregg, Montgomery, & Castaño, 1980). In particular, there are no reports of the effect of semantic similarity on output order in continuous-distractor free recall. In this paper we directly examine whether a manipulation of the temporal relations among list items affects the impact of pre-experimental associations on retrieval. This question, concerning the interaction between arguably the two most basic processes in human memory, remains unanswered. Here we show that as the temporal separation of items in a list is increased, the effect of pre-experimental associations on output order decreases.

Method

Using the LSA-CRP, we examine the effect of semantic similarity on retrieval transitions as the length of the IPI is varied. As in the previous analysis, we re-analyze data from Howard and Kahana (1999). Because recall probability decreased with increasing IPI (in our study and in most other studies using the continuous distractor paradigm) we restricted our analysis to the first recall transition except as noted. This would eliminate any possible confound caused by the differences in recall performance across the IPI conditions.

Results

Figure 3A plots the slope of the LSA-CRP for the four IPI conditions of Experiment 2 of Howard and Kahana (1999). Among the continuous-distractor conditions, $\cos\theta$ has the largest effect on recall transitions in the IPI = 2 (shortest inter-item distractor) condition, and the smallest effect in the IPI = 16 (longest inter-item distractor) condition. The effect of $\cos\theta$ on retrieval transitions in delayed free recall (IPI = 0) was comparable to the IPI = 2 continuous-distractor condition. $\cos\theta$ had a significant effect on output order, as evidenced by a significant positive slope of the LSA-CRP, in delayed free recall (IPI = 0) and in two of the three continuous-distractor conditions (IPI = 2 s and IPI = 8 s). In the longest IPI condition, the slope of the LSA-CRP ($.06 \pm .04$) did not quite reach statistical significant $t(15) = 1.46$, however, when data from all output positions were entered in the analysis, the slope of the LSA-CRP was significant for all four conditions ($.12 \pm .03$ for the IPI = 16 s condition, $t(15) = 4.1, p < .001$), and decreased monotonically with IPI for the continuous-distractor conditions (IPI$\geq$2).

To further assess the significance of the effect of IPI, we performed a regression of LSA-slope to IPI, where IPI was 0 for delayed free recall, 2 for IPI=2, and so on. This regression revealed a significant negative slope of $-.014 \pm .004, t(15) = -3.3, p < .01$. The effect of LSA on output order decreases as the delay between items becomes larger.

Discussion

The slope of the LSA-CRP, which was significant, or nearly so, for all values of IPI, decreased systematically as the IPI was increased. This means that pre-experimental associations (as measured by LSA $\cos\theta$) have a smaller effect on recall transitions as the temporal separation of list items increases. This systematic change stands in stark contrast to the near-invariance of the Lag-CRP across the same manipulation (see Figure 1).

There are two broad classes of explanations that can account for this effect of IPI on the efficacy of pre-experimental associations at retrieval. In one class, semantic and episodic cues combine at retrieval in such a way that the effectiveness of semantic cues depends on the absolute strength of episodic cues. Because the absolute strength of episodic cues presumably decreases with IPI, the effectiveness of semantic cues will decrease as well.[6] The other class are encoding accounts. According to these accounts, the utilization of semantic cues requires organizational processes during study. For instance, rehearsal could be required to discover the semantic relationships between list items (e.g., Glanzer, 1972). Such rehearsal processes are presumably increasingly disrupted as the inter-item distractors become longer. The following analysis looks

[6] Howard and Kahana (1999, Submitted) did not interpret the scale invariance of the Lag-CRP (see Figure 1) as evidence that temporally-defined associative strength is unaffected by increases in the IPI. Rather, we took the approximate scale invariance as an analogue of the long-term recency effect, which was interpreted as a consequence of a competitive retrieval mechanism. As a consequence, the relative, rather than the absolute spacing of the list is the primary factor determining the existence of a recency effect. Similarly, we assumed in our modeling work on the lag-recency effect (Howard & Kahana, 1999, Submitted), that the "strength" of association between list items was decreased in continuous distractor free recall, but that the competitive retrieval process led to approximate scale invariance because the relative "strength" was about the same.

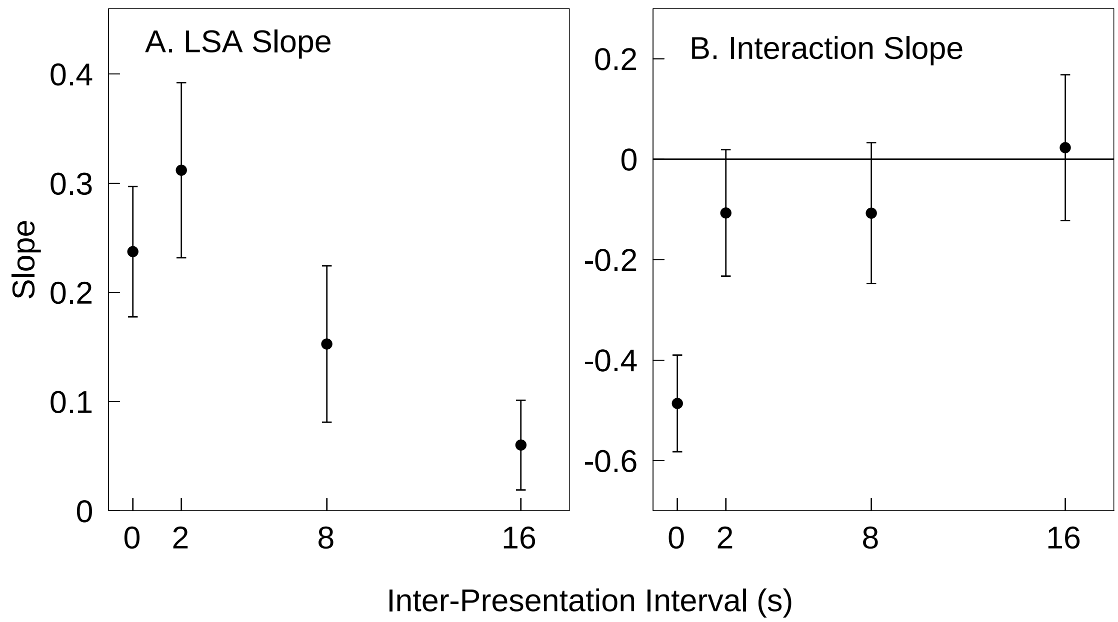

Figure 3. Summary of results across conditions of Experiment 2 of Howard and Kahana (1999). These conditions varied in the duration of distractor activity between presentation of the list items, the inter-presentation interval (IPI). Panel A. The LSA slope (see the line in Figure 2 A) measures the overall effect of semantic factors on recall transitions. As the IPI increases, LSA slope gradually decreases. Panel B. The interaction slope gives the slope of normalized LSA slope against the absolute value of lag. Negative values mean that the effect of LSA on recall transitions is relatively higher when the absolute value of lag is small. This property only holds in delayed free recall when the IPI is zero. The error bars represent the standard error of the mean.

for a characteristic effect predicted by such encoding accounts.

Analysis 3: Tests of an Encoding Explanation of the LSA-CRP

According to the multi-store models of free recall that were prominent in the late 1960s and early 1970s (Atkinson & Shiffrin, 1968; Glanzer, 1972), semantic relationships were believed to have an effect on recall from LTS, but no effect on recall from STS (see Glanzer, 1972, for a review). Similarly, it was proposed that co-occurrence in STS is necessary to discover and utilize the semantic relationships between words (Glanzer, 1969). This view held that if semantically similar words were simultaneously in STS, their association would be more efficiently stored in LTS. If one assumes that STS is increasingly disrupted by longer distractor intervals, one can explain the results shown in Figure 3A.

The STS-based explanation is one specific instance of a broad class of encoding-based accounts that could be invoked to describe the results of the preceding analysis, and indeed, to explain the finding that semantic variables affect free recall in the first place. All of these accounts make a very specific prediction. During study, the accessibility of prior items to STS[7]should show a recency effect. According to the STS-based explanation it is necessary to actively 'discover' the semantic relationship between ABSENCE and HOLLOW during study to fully exploit that relationship during retrieval. Both words must be simultaneously present in STS for this process to work. If ABSENCE has just been presented, HOLLOW is more likely to also be in STS if it was the previous item presented than if it was presented six words back. If such organizational processes in STS are important in driving the effect of $\cos\theta$ on recall transitions, then there should be a bigger effect for words that were close together in the list than for words that are far apart in the list. A prediction of these accounts, then, is that there should be a more prominent LSA-CRP for words that are nearby in the list than for words that are far apart in the list.

Method

We again analyzed Experiment 2 of Howard and Kahana (1999). As we found that our results did not track with probability of recall, we have collapsed over output positions for the purposes of these analyses.

We calculated CRP as a joint function of lag and LSA bin for each subject. To do this, we kept track of a matrix of numerators and denominators corresponding to LSA-bin and lag. To increase our statistical power, we collapsed lag into $|lag|$ and collapsed $|lag|$s$\geq$ 9 into a single bin. This resulted in a 100 ($\cos\theta$ bin) $\times$ 9 $|lag|$ matrix of numerators and denominators. We then calculated the LSA-CRP separately for

[7] Here we have in mind the Atkinson and Shiffrin (1968) buffer model, with any reasonable drop-out rule. The reader will probably note, however, that this argument applies equally well if one substitutes working memory or study-phase retrieval for STS. Any study-phase organizational process that requires the co-activation of the to-be-organized items, where activation is a function of recency, makes the same broad prediction.

each value of $|lag|$.

Our interest was in whether the *relative* effect of $\cos\theta$ is bigger for small values of $|lag|$ than for large values of $|lag|$. For each of the 9 values of $|lag|$, we calculated the slope of CRP to $\cos\theta$. We will refer to this as the LSA-slope.

Because subjects make more recall transitions at short $|lag|$s than at long $|lag|$s (see Figure 1), and because we were interested in the relative effect of $\cos\theta$, we divided the LSA-slope by the average CRP for each $|lag|$. To assess whether this normalized LSA-slope was greater for small $|lag|$s than for large $|lag|$s, we regressed the normalized LSA-slope to $|lag|$. Insofar as this "interaction slope" is less than zero, we will have demonstrated that the is effect of LSA decreases with short $|lag|$.

Results

The obtained interaction slopes for each condition, with 95% confidence intervals are presented in Figure 3B. Delayed free recall (IPI=0 s) showed a significant interaction slope. The slope was not significantly different from zero for any of the continuous-distractor conditions, nor when all three continuous-distractor conditions were pooled together ($-.11 \pm .11$, $|t(15)| < 1.0$). The pair-wise comparisons between delayed free recall and each of the continuous-distractor conditions was significant, $t(15) = -2.3, p < .05$, $t(15) = -2.5, p < .025$, and $t(15) = -2.5,\ p < .05$, respectively. None of the comparisons between the continuous-distractor conditions approached significance.

These analyses suggest that there is a discontinuity between the IPI=0 s delayed free recall condition and the IPI=2 s continuous distractor condition (Figure 3B). In delayed free recall, there was a larger relative effect of semantic similarity at small values of $|lag|$. There was no reliable evidence for such an effect in continuous-distractor free recall.

Discussion

The comparison of delayed free recall (IPI=0 s) with the IPI=2 s condition is particularly informative. Although we found a strong effect of $\cos\theta$ on the probability of recall transition for both condition, there was a strong effect for the delayed condition, and no effect for the IPI=2 s condition. The effect of LSA on output order is apparently dissociable from the interaction slope. As a consequence, we can conclude that encoding processes are not the sole cause of the LSA-CRP. Further, encoding processes are an unlikely explanation of the change of the LSA-CRP with increasing IPI.

The elimination of the interaction slope with the inclusion of an inter-item distractor of any duration suggests that the effect is a consequence of active rehearsal processes that are easily disrupted by an inter-item distractor. This makes sense if one carefully considers the timing of the experimental trials in Howard and Kahana (1999). In each condition, the words were presented on the screen for 1.2 s. During this time, the subject had to perform a judgment of concreteness on the to-be-remembered word. In the continuous-distractor conditions (IPI>0), each word was preceded by an arithmetic distractor, requiring the subject to spend the first portion of the 1.2 s switching from the arithmetic task to the orienting task, thus leaving little time for active rehearsal.

General Discussion

This paper has attempted to bridge the two main approaches to the study of free recall. One approach examines the effects of semantic factors on free recall. The other approach examines the effects of temporal factors on free recall. We have used latent semantic analysis (LSA, Landauer & Dumais, 1997) to operationalize a measure of semantic similarity. Using LSA, we were able to simultaneously measure semantic and temporal influences on output order in the free recall of randomly assembled word lists.

The first novel finding of this report is that LSA predicts output order and latency in free recall (see Figure 2). Words with very high LSA $\cos\theta$ to the just-recalled word were about twice as likely to be recalled as words with low $\cos\theta$. Further, the IRT associated with pairs of words with a very high $\cos\theta$ was several *seconds* faster than that associated with pairs of words with low $\cos\theta$. This finding is not particularly surprising, except insofar as it validates LSA-based measures of semantic similarity and our method, the LSA-CRP.

The second novel finding reported here is that the effects just described apply to both continuous-distractor and delayed free recall. This is the first report of a direct effect of semantic similarity on output order in the continuous-distractor paradigm. Because the demanding distractors interpolated between study items are likely to disrupt rehearsal in the continuous-distractor paradigm, this finding argues against the idea that inter-item rehearsal is required for the utilization of semantic cues.

A third novel finding relates to the way in which the effectiveness of semantic retrieval cues changes as the temporal separation between items increases (see Figure 3A). An active encoding account, in which rehearsal processes are increasingly disrupted with increasing IPI, can account for this pattern of results. This type of explanation predicts that the LSA-CRP should show a bigger effect for pairs that are close together in the list than for pairs that are far apart in the list. Analysis 3 failed to find such an effect for the continuous-distractor conditions, suggesting that the decrease with IPI is a property of memory retrieval. This finding is potentially of great importance, enabling models of free recall to be tested simultaneously against episodic and semantic associative processes.

Two Processes at Work

We have shown two basic differences between delayed (IPI=0) and continuous-distractor (IPI=2-16 s) free recall.

1. In delayed free recall, semantic similarity had a much larger effect on retrieval transitions than in continuous-distractor free recall (see Figure 3A). Furthermore, within continuous-distractor free recall, the effect of semantic similarity, as measured by the LSA-slope, decreased monotonically across IPIs of 2, 8, and 16 s.

2. In delayed free recall, semantic similarity exerts a more powerful effect on retrieval transitions for words that appeared in nearby serial positions than for words that appeared far apart in the list. This effect, however, is extremely fragile; an inter-item distractor of just two seconds is sufficient to disrupt the process giving rise to this interaction slope (see Figure 3B).

We interpret these two effects as evidence for two processes at work. We hypothesize that the first effect is a retrieval phenomenon, as discussed below. We interpret the second effect, as seen in the 'interaction slope' (Figure 3B), as a consequence of encoding differences across IPI conditions.

The striking interaction slope at IPI=0s which is eliminated for all IPIs$\geq$2 s, suggests that the interaction is a consequence of an active encoding strategy that is disrupted by any inter-item distractor. This is reminiscent of the postulate that discovery of inter-item associations in short-term memory (STM) is necessary for efficient storage of these associations in long-term memory (Glanzer, 1972). This account of semantic processing in free recall successfully accounted for the finding that semantic associations differentially affect recall of pre-recency items (Glanzer & Schwartz, 1971), and that recall of semantic associates is enhanced when they appear in nearby list positions (Glanzer, 1969).

Within the framework of the SAM model (Raaijmakers & Shiffrin, 1981), suppose that the associative strength between items that co-occur in short-term store (STS) is increased by an amount that is proportional to the semantic relatedness of the items, and that STS is emptied by any intervening distractor activity. In this case, one should see a greater effect of semantic similarity in delayed *vs* continuous-distractor free recall, as we observed (Figure 3A). This explanation of the pattern of results in Figure 3A hinges a special semantic advantage for items that appeared in nearby list positions. This advantage for adjacent items was indeed seen in delayed, but not continuous-distractor free recall. This explanation, however, is extremely unlikely to explain the effect of semantic similarity in general, nor the decrement in the effectiveness of semantic cues in the longest-IPI condition. If such an explanation were to be the only basis for an effect of semantic factors on output order in free recall, then we would expect to see LSA-slope track with interaction slope. A cursory examination of Figure 3 shows that this is not the case.

It is not necessary, however, to postulate some form of STS to explain the interaction slope. Suppose that during study, each presented word causes subjects to think of related words from the study list. Further suppose that this study-phase retrieval selects items *via* a rule that exhibits a recency effect, and favors words similar to the just-presented word. Because this makes the retrieved words effectively closer in the list to the presented words that prompted them, this should increase the associative strength between nearby list items that are similar in meaning. In this way, existing semantic relationships for recent words can be "amplified" by the study-phase retrieval, making it possible to explain the contiguity effect on the LSA-slope.

Constraints on Theories of Free Recall

This work, in conjunction with our prior work on the lag-recency effect present something of a puzzle. As the IPI is increased, the LSA-CRP decreases (becomes flatter), whereas the Lag-CRP remains largely unaffected (see Figure 1). This pair of results places strong constraints on theories of memory retrieval in free recall, and episodic memory in general.

At present there is one published model of memory that has been applied to associative effects in continuous-distractor free recall and is capable of explaining the persistence of the lag-recency effect with the inclusion of a filled IPI. We discuss this model and how it might simultaneously account for these two pieces of data.

Retrieved variable-context SAM (Howard & Kahana, 1999). The retrieved variable-context model (Howard & Kahana, 1999) extended prior work by Mensink and Raaijmakers (1988), which in turn provided an extension to the search of associative memory model (SAM Raaijmakers & Shiffrin, 1980). In SAM, the probability of sampling an item for recall is a function of the multiplicative strength of the various cues available to the subject. In free recall, the cues are typically taken to include the just-recalled item and context. If item i is the just-recalled item, then the probability of sampling item j is given by:

$$P_S(j|i) = \frac{S_{ij}S_{Cj}}{\sum_k S_{ik}S_{Ck}},$$

where S_{ij} is the associative strength between item i and item j and S_{Cj} is the cue strength of context on item j.

Previous work applying SAM to free recall had assumed that the lag-recency effect (see Figure 1) was a consequence of episodically-formed associations in STS (Kahana, 1996). That is, the lag-recency effect was seen as a consequence of changes in the strength of direct inter-item associations, S_{ij}. These changes were caused by the formation of associations in STS during list presentation. The finding of a lag-recency effect in continuous-distractor free recall was problematic for this explanation; presumably the interpolation of a demanding inter-item distractor would clear STS prior to the presentation of each item. An alternative to STS and direct inter-item cue strengths was required to explain the persistence of the lag-recency effect in continuous-distractor free recall.

The alternative we pursued was to search for an explanation based on properties of the context cue, S_{Cj}. Previous work had already elaborated the context cue from that used in the original SAM model. Rather than a fixed list context cue, as used in (Raaijmakers & Shiffrin, 1980), Mensink and Raaijmakers (1988) used variable context (e.g. Estes, 1955), to explain interference effects in paired-associate learning. The key property of variable context is that elements of context fluctuate. As a consequence, the state of context changes gradually over time. To explain the persistence of the lag-recency effect, as well as its approximate scale invariance, we proposed that the lag-recency effect was a consequence of *retrieved* variable context. That is, we assumed that when an item is recalled, the state of context is reset to the state that obtained when that item was presented. Because context changes gradually, this retrieved context will be a better cue for words that are nearby in the list, leading to a lag-recency effect.

The approximate scale invariance with the increase of IPI follows from the fact that the sampling equation is competitive; the probability of sampling item j depends not only on the strengths of the cues on item j, but also on the strengths of the cues to all the other items in the list. As a result, the key factor determining sampling of an item is not its absolute cue strength, but its strength relative to the other items in the list. Because increasing the IPI diminishes the strength of all context cue strengths similarly, the relative advantage for nearby items is preserved.

Within SAM, the inter-item associative strength matrix is the logical source of semantic similarity effects in free recall. If the LSA-CRP and the Lag-CRP are both driven by inter-item associative strength, then each entry of the matrix would represent a sum of one term for pre-existing associations and another term for newly-formed associations. In this case, decreasing the strength of the newly-formed associations, which one would expect as a consequence of increasing the IPI, should *enhance* the relative importance of pre-existing, semantic, associations. The present result, showing that the effect of semantic similarity on recall transitions *decreases* with increasing IPI (see Figure 3A), argues against this view. If, however, retrieved variable context, and not direct inter-item associations, mediates the lag-recency effect (as in Howard & Kahana, 1999), then contextually-mediated associative cue strength and semantic cue strength combine multiplicatively in the sampling equation. Because the contribution of semantic similarity is mediated by context cue strengths, it is reasonable that the effect of semantic similarity should decrease as absolute context cue strength decreases.

Methodological Issues in Free Recall

Most recent work on formal models of episodic memory (e.g. Chappell & Humphreys, 1994; Murdock, 1993; Shiffrin & Steyvers, 1997) has focused on tasks where the cue is clearly defined by the experimenter (e.g., cued recall and recognition). In free recall, subjects generate a series of responses, each serving as a cue for the next. The practical problem with modeling output order in free recall is that it forces the theorist to consider more complex models. If the 'strength' of an item changes as recall progresses, the researcher has to sum the items' strengths over all possible retrieval paths, weighted by the probability of each path—which is, of course, a function of the items' changing strengths!

The CRP analysis captures the transitions from word to word in the output protocol. This makes it ideally suited for modeling free recall. Other statistics, such as the serial position curve, measure the end-product of many sequentially applied transitions. A model can describe the serial position curve without describing the basic properties of memory retrieval reflected in the transitions. Conversely, a description of the transitions[8] nearly guarantees a description of the serial position curve. The CRP simplifies the task of the theorist. The probability of observing a given sequence of recalls is the joint CRP of each transition it contains. If one can account for the various CRPs in a principled way, the job is greatly simplified, if not completely solved.

The Lag-CRP described in Kahana (1996) has proved to be a valuable tool in distinguishing between different models of the episodic component of free recall (Howard & Kahana, 1999, Submitted). The LSA-CRP (Figures 2 and 3) may prove to do the same for the study of semantic factors. Although we have shown here that LSA is a useful measure, we make no claim one way or another as to the relative usefulness of LSA as compared to free association norms, subjective judgments of similarity or whatever other measure of semantic similarity one may favor. Indeed, the LSA-CRP could be trivially re-calculated for any fine-grained measure of semantic similarity one might care to consider.

The knowledge that LSA provides a useful (if not optimal) measure of the effect of semantic structure on episodic memory performance, coupled with the extremely large numbers of words for which LSA vectors are available, make it an attractive technique for such tasks as assembling experimental lists that have some *a priori* structure. The methodological utility of LSA in the verbal learning laboratory could turn out to be very important.

The analysis of the effect of semantic factors on episodic tasks has experienced a resurgence of interest in recent years (Nelson et al., 1992, 1998). Clearly, a complete description of free recall, and episodic memory in general, must account for the effects of semantic cues and their interaction with temporal cues in an organic way. The practical utility of LSA as a measure of semantic similarity will make it a useful tool in this endeavor.

References

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, p. 89-105). New York: Academic Press.

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology, 6*, 173-189.

Bousfeld, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology, 49*, 229-240.

Chappell, M., & Humphreys, M. (1994). An autoassociative neural network for sparse representations: Analysis and application to models of recognition and cued recall. *Psychological Review, 101*, 103-128.

D'Agostino, P. R. (1969). The blocked-random effect in recall and recognition. *Journal of Verbal Learning and Verbal Behavior, 8*, 815-820.

[8] This includes the transition to the first word recalled. In the case of the serial position curve, this is measured by the probability of first recall (Hogan, 1975; Howard & Kahana, 1999; Laming, 1999).

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological Review*, *62*, 145-154.

Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982). The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods and Instrumentation*, *14*, 375-399.

Glanzer, M. (1969). Distance between related words if free recall : Trace of the STS. *Journal of Verbal Learning and Verbal Behavior*, *8*, 105-111.

Glanzer, M. (1972). Storage mechanisms in recall. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 5, p. 129-193). New York: Academic Press.

Glanzer, M., Koppenaal, L., & Nelson, R. (1972). Effects of relations between words on short-term storage and long-term storage. *Journal of Verbal Learning and Verbal Behavior*, *11*, 403-416.

Glanzer, M., & Schwartz, A. (1971). Mnemonic structure in free recall: Differential effects on STS and LTS. *Journal of Verbal Learning and Verbal Behavior*, *10*, 194-198.

Greene, R. L. (1986). A common basis for recency effects in immediate and delayed recall. *Journal of Experimental Psychology : Learning, Memory & Cognition*, *12*, 413-418.

Greene, R. L., & Crowder, R. G. (1984). Effects of semantic similarity on long-term recency. *American Journal of Psychology*, *97*, 441-449.

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

Hogan, R. M. (1975). Interitem encoding and directed search in free recall. *Memory & Cognition*, *3*(2), 197-209.

Howard, M., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *25*, 923-941.

Howard, M. W., & Kahana, M. J. (Submitted). Distributed representation of temporal context. *Journal of Mathematical Psychology*.

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

Kahana, M. J., & Wingfield, A. W. (2000). A functional relation between learning and organization in free recall. *Psychonomic Bulletin and Review*.

Laming, D. (1999). Testing the idea of distinct storage mechanisms in memory. *International Journal of Psychology*, *34*(5/6), 419-426.

Landauer, T. K., & Dumais, S. T. (1997). Solution to Plato's problem : The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211-240.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. *Discourse Processes*, *25*, 259-284.

Mensink, G.-J. M., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, *95*, 434-55.

Murdock, B. B. (1993). TODAM2: A model for the storage and retrieval of item, associative, and serial-order information. *Psychological Review*, *100*, 183-203.

Murdock, B. B., & Okada, R. (1970). Interresponse times in single-trial free recall. *Journal of Verbal Learning and Verbal Behavior*, *86*, 263-267.

Nairne, J. S., Neath, I., Serra, M., & Byun, E. (1997). Positional distinctiveness and the ratio rule in free recall. *Journal of Memory and Language*, *37*, 155-166.

Neath, I. (1993). Contextual and distinctive processes and the serial position function. *Journal of Memory and Language*, *32*, 820-840.

Nelson, D. L., McKinney, V. M., Gee, N. R., & Janczura, G. A. (1998). Interpreting the influence of implicitly activated memories on recall and recognition. *Psychological Review*, *105*, 299-324.

Nelson, D. L., Schreiber, T. A., & McEvoy, C. L. (1992). Processing implicit and explicit representations. *Psychological Review*, *95*, 322-348.

Patterson, K. E., Meltzer, R. H., & Mandler, G. (1971). Interresponse times in categorized free recall. *Journal of Verbal Learning and Verbal Behavior*, *10*, 417-426.

Pollio, H. R., Richards, S., & Lucas, R. (1969). Temporal properties of category recall. *Journal of Verbal Learning and Verbal Behavior*, *8*, 529-536.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 14, p. 207-262). New York: Academic Press.

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

Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition*, *22*, 511-524.

Romney, A. K., Brewer, D. D., & Batchelder, W. H. (1993). Predicting clustering from semantic structure. *Psychological Science*, *4*, 28-34.

Schwartz, R. M., & Humphreys, M. S. (1973). Similarity judgements and free recall of unrelated words. *Journal of Experimental Psychology*, *101*, 10-13.

Serra, M., & Nairne, J. S. (in press). Part-set cuing of order information: Implications for associative theories of serial order memory. *Memory & Cognition*.

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

Tulving, E. (1962). Subjective organization in free recall of "unrelated" words. *Psychological Review*, *69*(4), 344-354.

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, *5*, 381-391.

Wingfield, A., Lindfield, K. C., & Kahana, M. J. (1998). Adult age differences in the temporal characteristics of category free recall. *Psychology and Aging*, *13*, 256-266.