Semantic structure and episodic memory

Marc W. Howard
Department of Psychology
Syracuse University

Kelly M. Addis
Department of Psychology
Indiana University

Bing Jing
Department of Psychology
Syracuse University

Michael J. Kahana
Department of Psychology
University of Pennsylvania

Invited chapter for *LSA: A Road Toward Meaning*. In press September 8, 2005

Abstract

In remembering a list of words, subjects' order of recall can reveal the influence of both semantic and temporal associations among items. In this chapter, we examine how well measures of semantic relatedness (e.g., Landauer & Dumais, 1997; Steyvers, Shiffrin, & Nelson, 2004) predict the order of subject's recalls. Analysis of recall transitions reveal that subtle variations in semantic relatedness strongly influence memory retrieval. Contrary to the view that temporal and semantic similarity strictly compete as retrieval cues, the data reveal that these two factors are separately modifiable, at least under certain conditions. These findings are not easily reconciled within current models of episodic and semantic memory.

A central function of episodic memory is to form and utilize associations between items experienced at nearby times. In addition to these newly–formed episodic associations, subjects enter the laboratory with a great deal of knowledge about verbal stimuli. Studying the relation between episodic and pre–existing, or semantic, associations can help shed light on the processes that lead to episodic retrieval. One prominent view is that episodic memory and semantic memory are separate memory systems (Tulving, 1983, 2002), and

---

The authors wish to acknowledge support from NIH grants MH069938 to MWH and MH55687 to MJK. Mark Steyvers generously provided the WAS vectors used here. We thank both Mark Steyvers and Simon Dennis for useful discussions concerning this work.

that semantic and episodic cues compete during memory retrieval. This competition would predict a reciprocal relation between the efficacy of episodic and semantic cues in predicting episodic retrieval.

For decades the difficulty in measuring the complex network of pre–existing associations among words handicapped researchers seeking to understand the relation between episodic and semantic memory. Until very recently, researchers were forced to rely on subjective judgments in measuring the effects of semantic similarity on episodic recall (e.g., Romney, Brewer, & Batchelder, 1993; Schwartz & Humphreys, 1973). Unfortunately, the combinatorics of directly measuring relations among tens of thousands of words renders this approach extremely difficult to accomplish for verbal learning experiments that use random lists of words. The recent development of computational methods to estimate semantic similarity has created new opportunities for examining the interaction between semantic and episodic associations. In this chapter, we report results derived from computational estimates of semantic similarity, in particular latent semantic analysis (LSA, Landauer & Dumais, 1997), which estimates semantic similarity by extracting information about the contexts in which words appear, coupled with conditional analyses of recall transitions.

Conditional analyses of transitions in free recall

LSA can be used to measure the effect of semantic associations on episodic recall (Howard & Kahana, 2002b). We will describe the ability of conditional measures of semantic and temporal factors to illuminate the complexity of learning. We will also review another computational method for assessing word similarity, the word association space (WAS, Steyvers et al., 2004) derived from free association norms and compare its properties to those of LSA. We start by introducing methods for conditional analyses of transitions in free recall.

Conditional analyses of temporal factors using the lag–CRP

In free recall, subjects recall as many items from a list as possible without experimenter–imposed constraints on the order of recall. By observing the transitions from one recall to the next as the subject searches through his memory of the list, we can learn about the structure of memory. In this chapter we study two classes of variables that affect recall transitions: semantic similarity and temporal proximity. Kahana (1996) developed a measure, the *conditional response probability* as a function of lag, or lag–CRP, to describe the effect of temporal proximity on episodic recall transitions. Temporal proximity between two items in a list can be measured by lag, the difference in their serial positions. The lag–CRP measures the probability of recall transitions of various temporal lags.

Recall transitions measured with the lag–CRP show evidence for two effects: contiguity and asymmetry. Contiguity means that recall transitions between nearby items in the list are more likely than recall transitions between distant items in the list; while asymmetry means that forward recall transitions are more likely than backward recall transitions. Both of these properties can be seen illustrated for a wide variety of data in Figure 1. The ubiquity of contiguity and asymmetry (Kahana, 1996; Howard & Kahana, 1999; Kahana & Caplan, 2002; Kahana, Howard, Zaromb, & Wingfield, 2002) suggests that they are a very general property of episodic memory for items learned in series. Because it characterizes the fundamental nature of temporal associations, the lag–CRP has also proven to be an important tool in developing models of free recall memory (Howard & Kahana, 2002a; Howard, Fotedar, Datey, & Hasselmo, 2005; Howard, Wingfield, & Kahana, In press).

In calculating the lag–CRP, a probability is estimated for recall transitions of each possible lag. In estimating these event probabilities, we divide the number of times the event occurs by the number of times the event could have occurred. A concrete example should help to illustrate this process. Consider the list ABSENCE HOLLOW PUPIL RIVER DARLING CAMPAIGN HELMET. Let's suppose that a subject recalls the words "RIVER, CAMPAIGN, DARLING," in that order. The first pair of recalls, RIVER–CAMPAIGN is associated with a lag of +2 because CAMPAIGN was two positions after RIVER in the list. The numerator for lag +2 would be incremented. The denominators for lags −3 to +3 would all be incremented. For the second recalled pair, CAMPAIGN–DARLING, the numerator for −1 would be incremented because DARLING was presented one position before campaign. The denominators for lags −5 to +1 would be incremented, with the exception of lag −2, which would have been an erroneous response because RIVER was already recalled. In averaging over retrievals, lists, and subjects, we arrive at an approximation of the conditional probability of recalling items at that lag.

Howard and Kahana (1999) used the lag–CRP to measure temporally–defined associations between items in continuous–distractor free recall (CDFR). In CDFR, a distractor task intervenes between each item presentation. The duration of the inter–item distractor task is referred to as the inter–presentation interval (IPI). A distractor task also follows the last item in the list prior to the recall test. The duration of the interval following the last item is referred to as the retention interval (RI). Howard and Kahana (1999, Experiment 2) showed that despite changes in the IPI ranging from 0 to 16 s there was no significant change in the shape of the lag–CRP curves. This finding is not simply a consequence of reduced attention given to the distractor; the 16 s RI was enough to severely disrupt the recency effect in the 0 s IPI condition. Initially, the finding that the lag–CRP persists across a delay long enough to disrupt the recency effect seems paradoxical; as the absolute strength of temporal connections between items decreases, there is little or no effect on the lag–CRP. This paradox, however, is only apparent. The lag–CRP is a relative measure that determines the probability of recalling an item at a particular lag, given that *some* recall transition is made. As the IPI increases, the overall number of items recalled decreases, but the relative probability of making recall transitions to various lags is unaffected. Howard and Kahana (1999) showed that these data were inconsistent with a description based on the Raaijmakers and Shiffrin (1980) Search of Associative Memory (SAM) model, the dominant model of serial position effects in free recall at the time.

Conditional analyses of semantic factors using LSA

The basic approach of analyzing recall transitions in the lag–CRP can be generalized from lag to any relevant stimulus dimension. The LSA–CRP(Howard & Kahana, 2002b) measures the effect of LSA $\cos\theta_{ij}$ on individual recall transitions. Howard and Kahana (2002b) examined the LSA–CRP for a continuous–distractor free recall study (Howard & Kahana, 1999). They found that the LSA–CRP for high values of $\cos\theta_{ij}$ was about twice as large as that for lower values of $\cos\theta_{ij}$—LSA had a highly significant effect on recall transitions in free recall. Surprisingly, they also found that as the length of the inter–item distractor increased, presumably weakening the strength of temporal associations, the

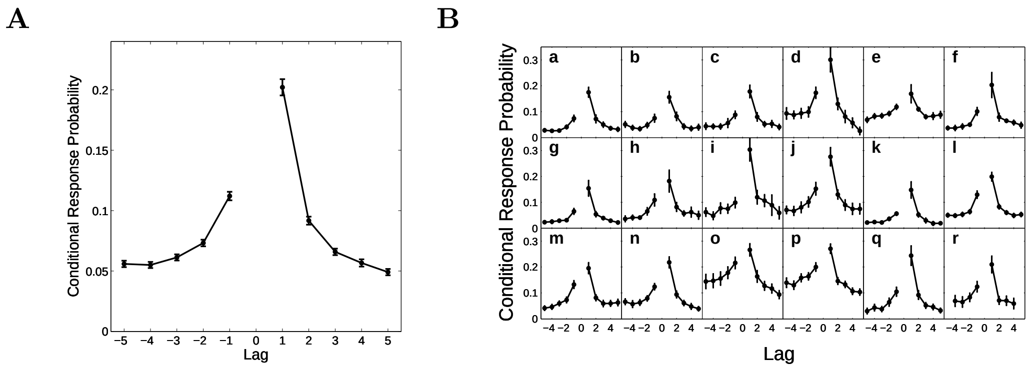

Figure 1. **Temporally–defined associations in free recall.** Each panel shows the probability of recalling a word from serial position $i+\text{lag}$ immediately following recall of serial position i—that is, the conditional–response probability (CRP) as a function of lag. **a.** Lag–CRP averaged across 18 different experiments. **b.** Lag–CRP curves from the following studies: a. Murdock (1962) (LL 20, 2 s). b. Murdock (1962) (LL 30, 1 s). c. Murdock and Okada (1970). d. Kahana et al. (2002) (Exp. 1). e. Howard and Kahana (1999) (Exp. 2). f. Murdock (1962) (LL 20, 1 s). g. Murdock (1962) (LL 40, 1 s). h. Murdock and Metcalfe (1978) (LL 20, 5 s/item). i. Howard and Kahana (1999) (Exp. 1, delayed). j. Kahana et al (2002) (Exp. 2). k. Roberts (1972). l. Zaromb et al. (submitted, Exp. 1) m. Zaromb et al. (submitted, Exp. 2) n. Thapar et al. (unpublished) o. Kimball and Bjork (2002). p. Kimball, Bjork & Bjork (2001). q. Kahana & Howard (2005, massed condition). r. Kahana et al. (2005).

effect of $\cos\theta_{ij}$ on recall transitions decreased. This finding suggests a deep relation between episodic and semantic associations.

In calculating the lag–CRP we estimated a probability for transitions to each possible lag. Lag is a discrete variable that only takes on certain values. To perform a similar analysis with LSA $\cos\theta_{ij}$, a continuous variable, we must first choose some way to discretize $\cos\theta_{ij}$. Howard and Kahana (2002b) took the distribution of observed $\cos\theta_{ij}$ values for the pairs within the pool of words used in this experiment (see Figure 5**a**) and formed 100 bins with equal numbers of members. Figure 2**a** shows typical pairs in these bins and their corresponding $\cos\theta_{ij}$. For each pair of recalled words, there is some $\cos\theta_{ij}$ and some corresponding $\cos\theta_{ij}$ bin between the just–recalled word and each other available word in the list. As with the lag–CRP, the LSA–CRP estimates the probability of making recall transitions that fall in each $\cos\theta_{ij}$ bin.

The following example illustrates the calculation of the LSA–CRP. Suppose the subject studied the list ABSENCE HOLLOW PUPIL RIVER DARLING CAMPAIGN HELMET and recalled the words RIVER, CAMPAIGN and DARLING, in sequence. For the first pair of recalled items, we find that the $\cos\theta_{ij}$ between the RIVER and CAMPAIGN falls into bin 41. We therefore increment both the numerator and denominator associated with bin 41. In addition to recording information about the observed event, we also need to keep track of the other possible events that could have been observed at that recall transition. If we have 100 $\cos\theta_{ij}$ bins and a list with 12 items, not all of the bins could have been observed on any particular recall transition. Accordingly, we calculate the $\cos\theta_{ij}$ bin between RIVER and all the potentially–recalled words in the list and increment the corresponding

a

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
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

b

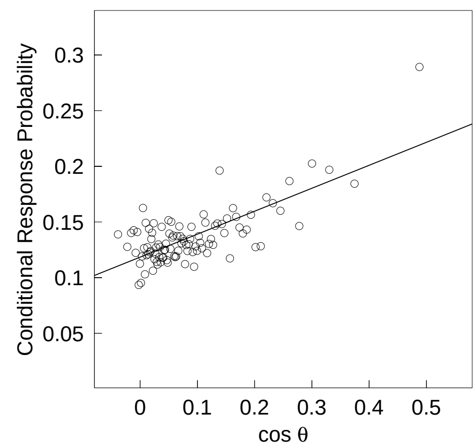

Figure 2. **Semantic associations in free recall. a.** Word pairs drawn from selected $\cos\theta_{ij}$ bin. Only very high bins predominantly contain pairs with obvious semantic relations. **b.** The LSA–CRP shows the probability of successively recalling words from different $\cos\theta_{ij}$ bins. Each pair of words in the word pool used in the experiment has a value of $\cos\theta_{ij}$ associated with it. This distribution was divided into 100 bins containing equal numbers of pairs, so that each pair was associated with a bin. Each time a word was recalled, each potentially–recalled word has a similarity to the just–recalled word and is thus associated with a bin. The left panel shows probability of recall as a function of the average $\cos\theta_{ij}$ in each bin. Data originally reported in Howard and Kahana (1999).

denominators. For instance, because RIVER–ABSENCE falls into bin 65, we increment the denominator associated with bin 65. Because RIVER–HOLLOW falls into bin 53, we increment the denominator associated with bin 53, and so on. We then move on to the next pair of recalled items, CAMPAIGN–DARLING. This recall would be analyzed in the same manner as RIVER–CAMPAIGN, with the exception that CAMPAIGN–RIVER would be excluded from the denominator because the already–recalled item RIVER would have been an erroneous response at that output position.

Figure 2**b** shows the LSA–CRP calculated for data from Experiment 2 of Howard and Kahana (1999). To quantify the general trend of the relation between recall transitions and LSA $\cos\theta_{ij}$, Howard and Kahana (2002b) calculated a regression line for the recall probability across the 100 $\cos\theta_{ij}$ bins for each subject. The line in Figure 2**b** represents the average (across subjects) regression. According to this regression, the conditional probability of recalling an available item with a very high $\cos\theta_{ij}$ is about twice that of an available item with a $\cos\theta_{ij}$ near zero. Even after excluding the twenty highest $\cos\theta_{ij}$ bins, the regression of recall probability on $\cos\theta_{ij}$ remained significant. This illustrates LSA's ability to capture relatively subtle semantic relations and the relevance of these relations for episodic recall.

Insight into the relation between episodic and semantic cues

The foregoing subsections have described a common framework based on conditional analyses of recall transitions to assess the influence of temporal and semantic factors on recall order in the free recall task. Armed with these methods, we can begin to ask questions about the relation between episodic and semantic cues on episodic retrieval. One possibility is that episodic and semantic memory rely on distinct memory systems (Tulving, 1983, 2002). In this case one might expect that at retrieval, subjects rely on some combination of semantic and episodic cues. In this case, episodic and semantic factors on retrieval would be inversely related to each other.

Using the LSA–CRP to measure the effect of semantic similarity on recall transitions using LSA, Howard and Kahana (2002b) examined how the effect of semantic similarity is modulated by temporal variables in two sets of analyses of the continuous distractor free recall data from Experiment 2 of Howard and Kahana (1999). In the first of these analyses, Howard and Kahana (2002b) calculated an LSA–CRP separately for transitions at each of several values of |lag|, collapsing over forward and backward transitions. This enabled them to look at how the effect of semantic similarity on output order interacted with the effect of temporal distance. Howard and Kahana (2002b) found that in delayed free recall, with the IPI set to zero, there was a larger effect of LSA $\cos\theta_{ij}$ on retrieval transitions when the words in question were also presented at small values of |lag|. That is, when the IPI was zero, there was a larger effect of semantic similarity between words that shared a strong temporal relation. However, in the three conditions in which the IPI was non–zero, the interaction was not different from zero and was significantly smaller than the effect observed with an IPI of zero. Even an IPI as short as 2 s was sufficient to disrupt the interaction. This finding is consistent with the idea that it is necessary to discover the semantic relations between words during encoding (Glanzer, 1969) and that it is easier to co–activate words that were presented close together in time. The presence of an interitem distractor, even a brief one, would presumably be sufficient to disrupt these active encoding processes.

The foregoing analysis did not support the hypothesis that episodic and semantic memory are reciprocally related components of memory retrieval in an episodic memory task. In addition to this analysis, Howard and Kahana (2002b) calculated an LSA–CRP separately for each condition of Experiment 2 of Howard and Kahana (1999). These conditions vary on the value of IPI, ranging from 0 s to 16 s. Increasing the IPI would be expected to result in a decrease in the strength of temporally–defined associations between items. Indeed, the overall number of recalled items drops dramatically as the IPI increases. If episodic and semantic associations compete with each other to determine recall order, then one might expect the decrease in the strength of temporal associations to be accompanied by an increase in the effect of semantic similarity, as measured by the LSA–CRP. In fact, Howard and Kahana (2002b) found exactly the opposite—as the IPI increased from 0 to 16 s, the slope of the regression line relating the CRP to LSA $\cos\theta_{ij}$ decreased. This finding enables us to reject the hypothesis that episodic and semantic memory systems compete for resources in episodic retrieval. Rather, they appear to support each other such that semantic cues have the largest effect on retrieval when episodic cues are also strong.

Conditional LSA analyses with learning

To examine the relation between semantic and temporal factors during learning, we analyze changes in semantic and temporal organization over learning trials from a study of temporal retrieval effects reported by Klein, Addis, and Kahana (2005). In the FR–Varied condition, subjects performed free recall on a list of words that was presented in a different random order on each of five study–test cycles. In this condition, the temporal associations between words are changing from trial to trial as presentation order changes. The repetition of the words over trials may allow a general increase in the strength of temporal associations, but this increase would be diffuse, as competing associations are formed on each successive trial. In the FR–Constant condition, the list was repeated in the same order on each cycle. In this condition, the strength of temporal cues should also increase with learning. Moreover, one might expect that the relative strength of the temporal associations would remain fixed over learning due to the fact that the list is presented in a consistent order each time. These two conditions should place different pressures on temporal and semantic factors as learning progresses.

Results

Figure 3**a** shows lag–CRPs for the FR–Constant (top) and FR–Varied (bottom) conditions. Subjects showed increased use of temporal associations across trials in the FR–Constant condition, but decreased use of temporal associations across trials in the FR–Varied condition. Figure 3**b** shows LSA–CRPs, calculated with 100 $\cos\theta_{ij}$ bins for the same data. In the FR–Varied condition, the LSA–CRP appears to grow more pronounced over learning trials, whereas the LSA–CRP appears largely unchanged across trials in the FR–Constant condition.

To assess these visual impressions more quantitatively, we computed summary statistics to describe the effect of temporal and semantic factors on their respective CRPs. To derive a summary statistic for temporal factors, we fit the power function CRP $= A\text{lag}^{-B}$ to the forward ($1 \leq \text{lag} \leq 5$) and backward ($-5 \leq \text{lag} \leq -1$) components of the lag–CRP (Kahana et al., 2002). We took the (across–subjects) average exponent B of the power function fits as a measure of the modulation of the lag–CRP by lag. We took the average slope of a linear regression of LSA–CRP to $\cos\theta_{ij}$ for each bin fit to each individual subject's data as measure of $\cos\theta_{ij}$'s effect on recall transitions.

Figure 3**c** shows forward and backward lag exponents averaged across subjects for each condition. A 2 (direction) $\times$ 2 (condition) $\times$ 5 (trial) repeated measures analysis of variance (ANOVA) showed main effects of direction [$(F(1,11) = 14.98, MSe = 0.86, p < .005$], condition [$F(1,11) = 69.13, MSe = 0.90, p < .001$], and trial [$F(4,44) = 12.17, MSe = 0.09, p < .001$]. There were also significant interactions between direction and condition [$F(1,11) = 17.39, MSe = 0.39, p < .005$] and, critically, between condition and trial [$F(4,44) = 23.68, MSe = 0.15, p < .001$]. Neither the interaction between direction and trial [$(F(4,44) = 1.07, n.s.$] nor the three–way interaction between direction, condition and trial [$F(4,44) = 1.33, n.s.$] approached significance. When considering just the FR–Constant condition, there was a significant effect of direction [$F(1,11) = 18.56, MSe = 1.03, p < .005$], and a significant effect of trial [$F(4,44) = 3.41, MSe = 0.13, p < .02$], but no interaction between direction and trial [$F(4,44) = 1.05, n.s.$]. The FR–Varied condition

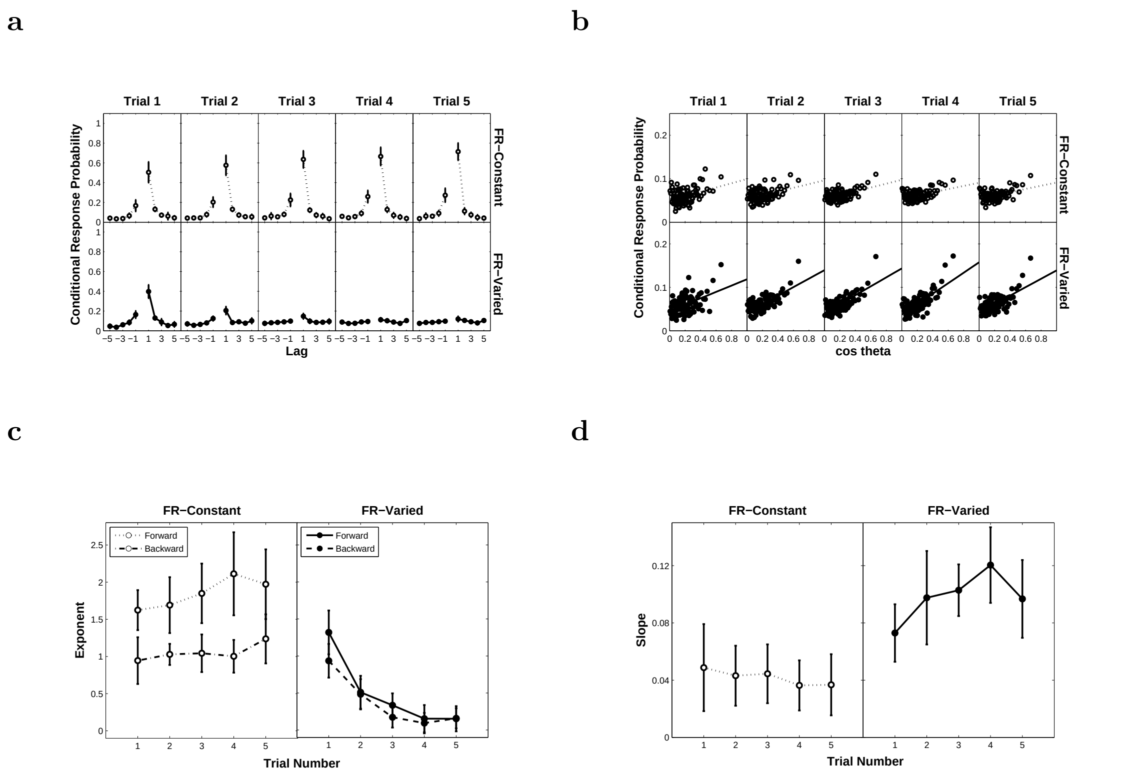

Figure 3. **Changes in temporal and semantic associations with learning.** Each panel shows data from two experimental conditions. In the FR–Constant condition, the word list was presented in the same order on each learning trial. In the FR–Varied condition, the list was presented in a new random order on each learning trial. In all cases, error bars reflect 95% confidence intervals. **a.** Temporally–defined associations remain largely unchanged with learning in the FR–Constant condition, whereas they rapidly diminish in the FR–Varied condition. **b.** Semantic associations remain relatively constant across learning trials in the FR–Constant condition, whereas they increase markedly in the FR–Varied condition. **c.** The exponents of power function fits to the forward and backward components of the lag–CRPs shown in **a** demonstrate a modest increase in the effect of temporal associations in the FR–Constant condition, but a dramatic decrease in the FR–Varied condition. **d.** Regression slopes of the LSA–CRP (**b**) are relatively constant across trials in the FR–Constant condition. In contrast, there is an increase in the slope over trials in the FR–Varied condition. Data from Klein, Addis and Kahana (2005).

showed a significant effect of trial [$F(4, 44) = 36.91, MSe = 0.11, p < .001$], but no effect of direction [$F(1, 11) = 2.20, n.s.$].

These analyses indicate that, whereas the effect of temporal factors on retrieval—as measured by the lag–CRP with lag calculated relative to the most recent list presentation—decreases with learning in the FR–Varied condition, it increases with learning when the lists are repeated in a constant order. In the FR–Constant condition, although the strength of temporal associations can be assumed to increase across trials, the lack of a significant interaction between direction and trial indicates that the asymmetry in temporally–defined associations does not change over trials. This finding is particularly challenging for theoretical accounts of the lag–CRP: although the discrepancy between the lag–CRP for adjacent lags and remote lags grew over time, the discrepancy between the lag–CRP for forward and backward recall transitions did not.

The analysis of average LSA slopes confirmed our qualitative impressions. A 2 (condition) $\times$ 5 (trial) repeated measures ANOVA showed a significant main effect of condition [$F(1, 11) = 24.4, MSe = .004, p < .001$], no main effect of trial [$F(4, 44) = 1.98, n.s.$], and a significant interaction [$F(4, 44) = 4.23, MSe = 0.001, p < .01$]. This interaction was driven by an increase in slope with trial in the FR–Varied condition [$F(4, 44) = 5.728, MSe = 0.001, p < .005$]. There was no effect of trial on the LSA–CRP slope in the FR–Constant condition [$F(4, 44) < 1, n.s.$].

Discussion

The changing influence of semantic and episodic cues across learning trials, as described above, appears inconsistent with the hypothesis that semantic and episodic cues strictly compete during retrieval. In the FR–Constant condition, it seems reasonable to assume that the strength of temporal associations among list items increases with learning. Even though the list was presented in a consistent order on each trial, temporal factors did not come to completely dominate recall transitions. For instance, the lag–CRP from the FR–Constant condition did not come to resemble the type of lag–CRP function one would expect from serial recall, in which the vast majority of list items are recalled at a lag of +1 (Klein et al., 2005). Even as the effect of lag on retrieval increased over learning trials, the regression slope of the LSA–CRP did not decrease significantly, as one would expect if there were an inverse relation between episodic and semantic factors. Even more puzzling, the pattern we observed previously—increased semantic effects when temporal cues were strong (Howard & Kahana, 2002b)—was also not observed here: as temporal cues increased in strength with learning, there was no corresponding increase in the effect of semantic similarity on retrieval.

In the FR–Varied condition, the effect of semantic similarity increased over learning trials while the lag–CRP flattened. At first glance it appears that these results from the FR–Varied condition provide straightforward support for the hypothesis of an inverse relation between temporal and semantic factors on retrieval, but on closer examination it is not so obvious that this is the lesson provided by these data. Although the lag–CRP decreased over learning trials in the FR–Varied condition, it should be noted that lag was calculated based on the most recent list presentation. It might be more accurate to say that when the list was presented in multiple orders, there is a decrease over trials in the relative importance of the most recent presentation order to specifying the temporal associations between list items. It

Table 1: **The complex relation between temporal and semantic factors in free recall.** The column labeled "Increasing IPI," where IPI is an abbreviation for inter–presentation interval, refers to results reported by Howard and Kahana (2002b). CDFR stands for continuous–distractor free recall. The columns labeled "FR–Varied" and "FR–Constant," where FR stands for free recall, refer to results originally reported in Klein et al., (2005) and here. A table entry with a plus sign indicates that the experimental manipulation heading the column resulted in an increase in the dependent measure labeling the row. A minus sign indicates a decrease in the dependent measure while an equal sign indicates no effect.

Manipulation	Increasing IPI	Learning Trials	
Condition	CDFR	FR–Varied	FR–Constant
Recall probability	−	+	+
Temporal associations	=	−	+
Semantic associations	−	+	=

should also be noted that the increase in the effect of semantic similarity with learning in the FR–Varied condition confirms that the increase in subjective organization (Tulving, 1966) observed with learning is accompanied by an increase in the effects of semantic similarity on retrieval transitions (Schwartz & Humphreys, 1973).

Although we can reject the hypothesis that temporal and semantic factors exhibit a simple inverse relation during episodic retrieval, the hypothesis that the effect of semantic similarity is greatest when temporal cues are strongest was also not supported by the results of the FR–Constant condition. As summarized in Table 1, increasing the length of the IPI in continuous–distractor free recall reduced both overall recall probability and subjects' reliance on semantic associations (Howard & Kahana, 2002b). The same manipulation did not significantly alter subjects' reliance on temporally–defined associations in choosing items to recall (Howard & Kahana, 1999, 2002b). Here we showed that when words were learned with a variable order of presentation across trials, probability of recall increased along with the effect of semantic relatedness on memory retrieval, while the effect of temporal proximity decreased (Figure 3, solid). However, when lists were learned in a consistent order, although probability of recall still increased, the effect of temporal factors increased while the effect of LSA $\cos\theta_{ij}$ on output order remained constant (Figure 3, dashed). Taken together, these three sets of findings illustrate a more complex relation between temporal and semantic factors in episodic recall than has previously been appreciated. Our ability to describe these processes is largely a consequence of the methodological advantage of being able to assess semantic similarity between large numbers of pairs of words without the time or expense of subjective ratings. In the next section, we discuss the potential for other computational methods of assessing semantic similarity to measure output order in free recall.

WAS and LSA

The preceding work demonstrates that LSA, especially when combined with conditional measures of memory retrieval, provides a valuable tool in the experimental study of memory. This, of course, does not preclude the possibility that other computational measures may also provide useful experimental tools for studying the effects of semantic structure on episodic memory performance. In this section, we discuss semantic associations measured with the word association space (WAS, Steyvers et al., 2004), another computational method that has successfully been used to describe the effects of semantic factors on episodic memory performance.

WAS: Word association space

LSA constructs a semantic representation from two main components. First, information about word co–occurrence in naturally–occurring text is extracted. Second, this information is provided as input to a singular value decomposition (SVD) step that performs dimensional reduction. Recent computational measures developed by Steyvers and colleagues (Griffiths & Steyvers, 2002, 2003; Steyvers et al., 2004) have explored variations on both of these major aspects of LSA. Although we will not conduct conditional analyses with it here, the Topics model (Griffiths & Steyvers, 2002, 2003, this volume) warrants some discussion here. The Topics model attempts to capture the semantic relations between words by hypothesizing that a body of text, such as the TASA corpus often used to construct LSA spaces, is generated from a set of discrete topics. These topics function something like latent variables. The conditional probabilities of observing a particular word in a particular topic, as well as the probability that each context contains a particular topic are then adjusted to best predict the observed corpus. The Topics model, like LSA, takes in information about word co–occurrence but uses different mathematical techniques to estimate relations between words. In contrast, WAS (Steyvers et al., 2004) uses mathematical techniques similar to those used in LSA but with a very different source of input information.

WAS starts with an associative matrix constructed from free association norms collected at the University of South Florida (USF Norms, Nelson, McEvoy, & Schreiber, 2004). These norms have provided a valuable experimental tool in evaluating the role of semantic factors on item recognition (Nelson, Zhang, & McKinney, 2001), cued recall (Nelson & Zhang, 2000), and extralist cued recall (Nelson, Bennett, Gee, Schreiber, & McKinney, 1993). In particular, the USF norms have proven useful in revealing the importance of indirect word associations in guiding episodic recall (Nelson, Bennett, & Leibert, 1997; Nelson, McKinney, Gee, & Janczura, 1998).

The calculation that results in the WAS space operates on the normed probability A_{ij} of responding with word j when probed with word i. Steyvers et al. (2004) started with a symmetric measure of associative strength generated from the free association matrix:

$$S^{(1)}_{ij} := A_{ij} + A_{ji}.$$

They supplemented this with a measure that included indirect associations, which prior studies have shown affect extralist cued recall performance (Nelson et al., 1997; Nelson & McEvoy, 2000):

$$S^{(2)}_{ij} := S^{(1)}_{ij} + \sum_k S^{(1)}_{ik} S^{(1)}_{kj}.$$

The associative matrix, $S^{(1)}$ or $S^{(2)}$, is then decomposed using singular value decomposition yielding a representation with reduced dimensionality. Steyvers et al. (2004) also evaluated a version of WAS that performed multidimensional scaling (MDS) on a path length derived from the free association norms. Because the SVD solution typically outperformed, or showed equivalent performance to that of the MDS solution, and because of the formal similarity of the SVD solution to LSA, we will only examine WAS–SVD here, and will refer to it simply as WAS.

Scalar measures of WAS and memory performance measures

In order to illustrate the utility of WAS as an experimental tool, Steyvers et al. (2004) examined the correlation between WAS $\cos\theta_{ij}$ and judgments of remembered semantic similarity between a probe word and a study list, extralist cued recall performance, and extralist intrusions in free recall. These correlations were compared to correlations calculated using $S_{ij}^{(1)}$ and $S_{ij}^{(2)}$, as defined above, as well as LSA $\cos\theta_{ij}$. In all three cases, WAS $\cos\theta_{ij}$ showed a stronger correlation with performance than LSA $\cos\theta_{ij}$. We will discuss the recall results of Steyvers et al. (2004) in more detail here.

In extra–list cued recall, a list is presented for study (e.g., Nelson et al., 1998; Tehan & Humphreys, 1996). An associate of one of the list words is then presented as a cue for that target list word. Steyvers et al. (2004) took a large set of cue–target pairs and correlated the various scalar semantic measures with the probability that each cue evoked its target across lists and subjects. They found that WAS showed a slightly higher correlation than the raw associative strengths. For instance, the correlation for $S^{(2)}$ was approximately .5, whereas the correlation for the $\cos\theta_{ij}$ from the vector space calculated from $S^{(2)}$ retaining 400 dimensions was .55. However, both of these numbers were considerably higher than the correlation observed for LSA $\cos\theta_{ij}$, which was about .3 (right panel of Fig. 2 Steyvers et al., 2004).

The free recall data examined by Steyvers et al. (2004) was originally reported by Deese (1959), who found that lists with a particular construction yield high rates of intrusion of particular items. For instance, when presented with the list DREAM PILLOW BED TIRED ... subjects were as likely to recall the word "SLEEP" as they were to recall words that were actually presented. Steyvers et al. (2004) correlated the probability of intruding a non–presented target item with the average similarity of that item to the presented list items using the various scalar measures mentioned previously. Although they found that the correlation between WAS $\cos\theta_{ij}$ and the probability of an intrusion was much higher than that of LSA $\cos\theta_{ij}$, the correlation for both was much lower than that for the raw free association measures $S^{(1)}$ and $S^{(2)}$. This superiority is perhaps not surprising insofar as Deese (1959) constructed these lists using free association norms. In order to shed further light on the differences between WAS and LSA as measures of the effect of semantic similarity on episodic retrieval, we will calculate conditional analyses for LSA and WAS for the free recall data on which the LSA–CRP was developed (Howard & Kahana, 1999, 2002b).

Conditional analyses with WAS

We were provided with a set of 400–dimensional WAS vectors derived from $S^{(2)}$ for the 422–word subset of the Toronto Noun Pool for which free association norms are available. We used these WAS vectors to calculate $\cos\theta_{ij}$ for each pair of items in the subset and calculated a WAS–CRP precisely analogous to the LSA–CRP described above for the data from Experiment 2 of Howard and Kahana (1999). Missing pairs were ignored in these analyses. To ensure that any discrepancies between these and prior analyses were not a consequence of the missing words we recalculated the LSA–CRP with only the 422 words for which WAS vectors were available. As an exploratory method, we also calculated a CRP using the inner product rather than $\cos\theta_{ij}$ for each LSA and WAS. All of these analyses were calculated using both 10 and 100 bins. Although the results were comparable, it is our opinion that the results with 100 bins are more illuminating, so we will focus on these here.

Both LSA and WAS showed a definite increase in the CRP with increasing semantic similarity for both $\cos\theta_{ij}$ and inner product (see Figure 4). All of these measures of semantic similarity hold some merit for describing the effect of semantic similarity on episodic recall. The average of the LSA $\cos\theta_{ij}$ regressions yielded a slope of $0.21 \pm .03$ (95% confidence interval). The average of the WAS $\cos\theta_{ij}$ regressions yielded a slope of $0.31 \pm .03$ (95% confidence interval). Although the average slope was considerably larger for WAS $\cos\theta_{ij}$ than for LSA $\cos\theta_{ij}$, it is extremely difficult to interpret this result because LSA and WAS distribute $\cos\theta_{ij}$ very differently across pairs. This can be seen informally by comparing the density of points along the $\cos\theta_{ij}$ axis in Figure 4**a** vs **c**. Figure 4**c**, representing the WAS $\cos\theta_{ij}$ CRP, shows a relatively wide gap between the highest $\cos\theta_{ij}$ and the rest of the distribution. This is because bin 100 contains a wider range of WAS $\cos\theta_{ij}$ values. The difference in the distribution of these variables makes it difficult to directly compare the efficacy of LSA and WAS as measures of the effect of semantic factors on episodic recall. Further, as we shall see shortly, the various measures under consideration here have different sensitivities to word frequency. We will discuss the effect that the shape of the distributions and sensitivity to word frequency have on the suitability of both WAS and LSA for various applications.

To help clarify these issues, Figure 5 shows distributions of LSA and WAS $\cos\theta_{ij}$ and inner product on log–log coordinates. In generating the distributions, the word pool was first split into high– and low–frequency halves. The distribution of pairs composed of two high–frequency words are shown in grey; the distribution of pairs composed of two low–frequency words are shown in black. From these distributions we can infer several properties of these measures. First, LSA $\cos\theta_{ij}$ is distributed differently than the other measures, showing a broad peak and a curved tail in log–log coordinates. The other distributions (with the exception of the LSA inner product high–frequency distribution) show an approximately linear tail on log–log coordinates, suggesting a power law tail to the distribution.

These differences in the shape of the distributions make it extremely difficult to interpret analyses that summarize the effect of semantic similarity with a single scalar measure. An advantage of the conditional approach used here is that it enables us to examine the properties of these measures without obscuring differences in the shape of the distribution that would skew, for instance, average $\cos\theta_{ij}$ of adjacent recalls.

LSA $\cos\theta_{ij}$ is also distinctive from the other distributions by virtue of its sensitivity to word frequency. Whereas the other distributions are shifted to the right for the high–frequency pairs (grey line), the low–frequency pairs actually have a higher average $\cos\theta_{ij}$ than the high–frequency pairs. This pattern is reversed for the other measures, especially prominently for the LSA inner product. This suggests that variation in LSA $\cos\theta_{ij}$ is anti–correlated with the normative frequency of the words for which $\cos\theta_{ij}$ is being calculated.

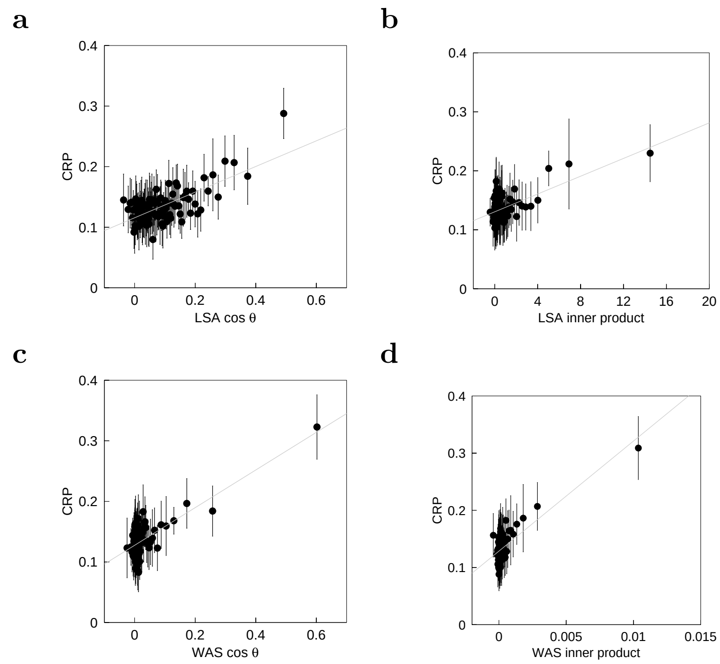

Figure 4. **Semantic associations measured using LSA and WAS. a.** CRP computed using the $\cos\theta_{ij}$ calculated from LSA. **b.** CRP computed using the inner product calculated from LSA. **c.** CRP computed using the $\cos\theta_{ij}$ calculated from WAS. **d.** CRP computed using the inner product calculated from WAS. Note that the scales on the x–axes in **b** and **d** are not directly comparable. The error bars are 95% confidence intervals.

This analysis also suggests that the WAS $\cos\theta_{ij}$ and inner product of a pair are weakly correlated with the word frequency of the members of the pair and that LSA inner product should be strongly correlated with the word frequency of the members of the pair. Each of these measures provides some degree of sensitivity to word frequency. Depending on what one wants to measure in a particular experiment, sensitivity to word frequency may or may not be a desirable property. It is perhaps worth noting that the weighting function employed in LSA specifically tries to minimize the importance of high frequency words. However the fact that the LSA inner product is strongly correlated with word frequency but does not show a CRP vastly different from the other measures suggests that the CRPs of the other measures are not strongly influenced by word frequency.

Some final thoughts

We have discussed ways to assess the roles of newly–formed episodic and pre–existing semantic associations in determining episodic recall performance. We briefly reviewed conditional analyses of temporal factors in free recall and the extension of these measures to semantic factors using LSA $\cos\theta_{ij}$. These conditional analyses described a complex relation between episodic and temporal factors in free recall (Table 1). We reviewed the finding that the effect of semantic factors on recall decreased as the duration of the delay between list items, or IPI, was increased (Howard & Kahana, 2002b). We examined the effect of learn-

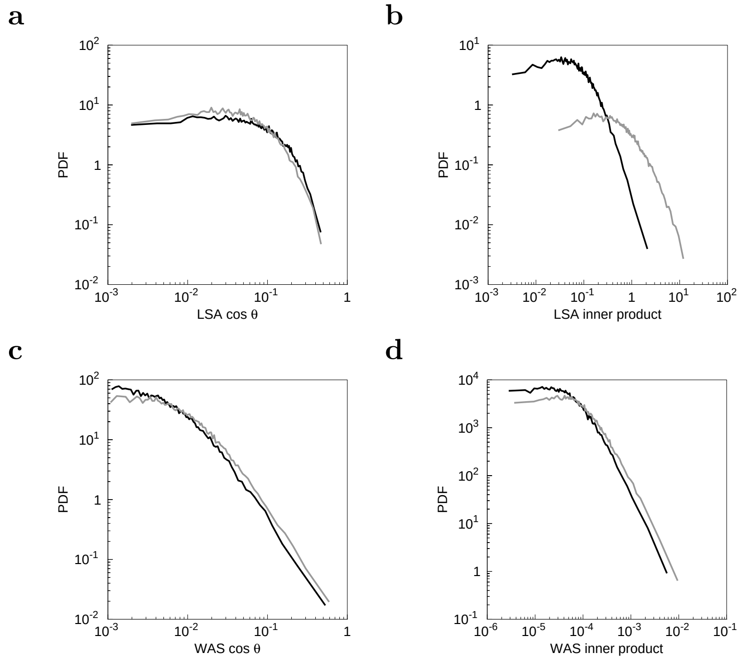

Figure 5. **The effect of word frequency on the distribution of LSA and WAS similarity.** A 422–word subset of the Toronto Noun Pool was split into a high–frequency and low–frequency subset by a median split on Kučera–Francis word frequency. This figure shows the distribution of similarity scores derived from LSA or WAS in log–log coordinates. The black lines show distributions for pairs taken from the low–frequency subset; the grey lines show the distribution for pairs from the high–frequency subset. **a.** LSA $\cos\theta_{ij}$ is lower for high–frequency pairs than for low–frequency pairs. This suggests that with respect to the angular distribution of vectors, high–frequency words are not clustered but are distributed more like corners of a box. **b.** LSA inner product. In LSA, vector length is correlated ($r \simeq .6$) with Kučera–Francis word frequency. As a consequence, LSA inner product is much greater for high–frequency pairs than for low–frequency pairs. **c.** WAS $\cos\theta_{ij}$. In contrast to the LSA $\cos\theta_{ij}$ distributions, the WAS $\cos\theta_{ij}$ distributions contain many more very low–similarity pairs. Also in contrast to LSA $\cos\theta_{ij}$, WAS $\cos\theta_{ij}$ is higher for high–frequency pairs than low frequency pairs. This suggests that in WAS high–frequency words are clustered in a central region of the space. **d.** WAS inner product. Like the LSA inner product, the WAS inner product shows a separation between high frequency pairs and low frequency pairs such that high–frequency pairs have a higher inner product. These distributions are not as widely separated as those for LSA inner product.

ing on temporal and semantic factors using two presentation schedules. When words were repeated in a random order across learning trials, the effect of temporal factors—at least with reference to the most–recently presented list—decreased dramatically, while the effect of semantic factors, as illustrated by the LSA–CRP, increased dramatically. This finding is consistent with previous results on the relation between temporal and semantic factors during learning (Schwartz & Humphreys, 1973). However, when the items were presented in a consistent order on each study–test cycle, the effect of temporal factors increased, while the effect of semantic factors remained approximately constant. These findings present something of a puzzle for models of temporal and semantic factors in episodic recall.

Do episodic and semantic associations result from a common source? If they do, retrieved temporal context would be a strong candidate. LSA and the Topics model both ultimately describe a word in terms of the contexts in which it appears. Contextual retrieval is also a key feature of two recent models of episodic memory. Dennis and Humphreys (2001) used retrieved context as the central mechanism for a model of episodic recognition memory. This model predicted that words that occur in many contexts should be harder to recognize than words that occur in few contexts, with word frequency controlled. This prediction has been confirmed (Steyvers & Malmberg, 2003). The temporal context model (TCM, Howard & Kahana, 2002a; Howard et al., 2005) exploits contextual retrieval to describe newly–formed associations in episodic recall. Because TCM describes context as a combination of patterns retrieved by items, it may be possible to describe something not too dissimilar to the contextual mixing performed by both LSA and the Topics model. Kwantes (in press) has recently built a model that captures some aspects of semantic meaning from a framework built on MINERVA, a model that was developed to describe episodic memory performance. A description of the interactions between temporal and semantic factors in episodic recall will be an essential step in developing a unified model of episodic and semantic learning.

References

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*, 17-22.

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

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

Griffiths, T., & Steyvers, M. (2002). A probabilistic approach to semantic representation. *Proceedings of the 24th Annual Conference of the Cognitive Science Society.*

Griffiths, T., & Steyvers, M. (2003). Prediction and semantic association. *Advances in Neural Information Processing Systems*, *15*.

Howard, M. W., Fotedar, M. S., Datey, A. V., & Hasselmo, M. E. (2005). The temporal context model in spatial navigation and relational learning: Toward a common explanation of medial temporal lobe function across domains. *Psychological Review*, *112*(1), 75-116.

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

Howard, M. W., & Kahana, M. J. (2002b). When does semantic similarity help episodic retrieval? *Journal of Memory and Language*, *46*, 85-98.

Howard, M. W., Wingfield, A., & Kahana, M. J. (In press). Aging and contextual binding: Modeling recency and lag-recency effects with the temporal context model. *Psychonomic Bulletin & Review*.

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

Kahana, M. J., & Caplan, J. B. (2002). Associative asymmetry in probed recall of serial lists. *Memory & Cognition*, *30*, 841-849.

Kahana, M. J., Dolan, E. D., Sauder, C. L., & Wingfield, A. (2005). Intrusions in episodic recall: Age differences in editing of overt responses. *Journal of Gerontology: Psychological Sciences*, *60*, 92-97.

Kahana, M. J., & Howard, M. W. (2005). Spacing and lag effects in free recall of pure lists. *Psychonomic Bulletin & Review*, *12*, 159-164.

Kahana, M. J., Howard, M. W., Zaromb, F., & Wingfield, A. (2002). Age dissociates recency and lag-recency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 530-540.

Klein, K. A., Addis, K. M., & Kahana, M. J. (2005). A comparative analysis of serial and free recall. *Memory & Cognition*, *33*(5).

Kwantes, P. J. (in press). Using context to build semantics. *Psychonomic Bulletin & Review*.

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.

Nelson, D. L., Bennett, D. J., Gee, N. R., Schreiber, T. A., & McKinney, V. M. (1993). Implicit memory: effects of network size and interconnectivity on cued recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19(4)*, 747-764.

Nelson, D. L., Bennett, D. J., & Leibert, T. W. (1997). One step is not enough : Making better use of association norms to predict cued recall. *Memory & Cognition*, *25*, 785-796.

Nelson, D. L., & McEvoy, C. L. (2000). What is this thing called frequency? *Memory & Cognition*, *28*(4), 509-522.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The university of south florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments and Computers*, *36(3)*, 402-407.

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., & Zhang, N. (2000). The ties that bind what is known to the recall of what is new. *Psychonomic Bulletin and Review*, *7(4)*, 604-617.

Nelson, D. L., Zhang, N., & McKinney, V. M. (2001). The ties that bind what is known to the recognition of what is new. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27(5)*, 1147-1159.

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.

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 judgments and free recall of unrelated words. *Journal of Experimental Psychology*, *101*, 10-15.

Steyvers, M., & Malmberg, K. J. (2003). The effect of normative context variability on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(5), 760-6.

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2004). Word association spaces for predicting semantic similarity effects in episodic memory. In A. F. Healy (Ed.), *Cognitive psychology and its applications: Festschrift in honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer.* Washington, DC: American Psychological Association.

Tehan, G., & Humphreys, M. S. (1996). Cuing effects in short-term recall. *Memory & Cognition*, *24*(6), 719-32.

Tulving, E. (1966). Subjective organization and effects of repetition in multi-trial free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *5 (2)*, 193-197.

Tulving, E. (1983). *Elements of episodic memory.* New York: Oxford.

Tulving, E. (2002). Episodic memory: from mind to brain. *Annual Review of Psychology*, *53*, 1-25.