Journal of Memory and Language **46,** 85–98 (2002)
doi:10.1006/jmla.2001.2798, available online at http://www.academicpress.com on IDEAL®

When Does Semantic Similarity Help Episodic Retrieval?

Marc W. Howard and Michael J. Kahana

Volen Center for Complex Systems, Brandeis University

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. Using estimates of semantic similarity provided by latent semantic analysis (LSA; Landauer & Dumais, 1997), we simultaneously assessed the effects of temporal and semantic proximity on output order in delayed and continuous-distractor free recall of random word lists. These analyses revealed that subtle variations in semantic similarity have large effects on recall transitions in delayed free recall. Further, these effects decrease as the duration of the interitem distractor (IPI) was increased from 2–16 s. In contrast, the effect of temporal proximity on recall transitions did not change with increasing IPI. This dissociation in the effects of interitem distraction on semantic and temporal similarity effects presents a new challenge for models of free recall and episodic memory retrieval.

Key Words: semantic similarity; episodic retrieval; free recall; latent semantic analysis.

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 relation between the study times of list items (Howard & Kahana, 1999; Kahana, 1996; Serra & Nairne, 2000). Semantic factors, in contrast, refer to preexisting relations among to-be-remembered items. 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, nor have studies of temporal factors 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, researchers typically use categorized word lists to study 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, in studying temporal factors, researchers typically use lists of randomly arranged concrete words. Even in such "unre-

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 and Tom Landauer and Douglas Nelson for reading an early draft of the manuscript.

Address correspondence and reprint requests to Michael Kahana, Volen National Center for Complex Systems, MS 013, Brandeis University, Waltham, MA 02254-9110. E-mail: kahana@brandeis.edu.

0749-596X/02 $35.00

lated" 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 episodic factors in retrieval.

This paper unifies these two approaches by examining the interacting roles of semantic and temporal factors in free recall of random word lists. Because free recall experiments typically utilize large pools of words (to avoid the need for within-session repetition of items), we use *latent semantic analysis* (Landauer & Dumais, 1997) to derive a measure of semantic similarity for all possible pairs of words used in our studies.

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, interresponse times (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 preexperimental (semantic) associations among list items thus exert a powerful influence on both output order and latency.

To examine the influence of preexperimental (semantic) associations on retrieval, we must first determine 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 previously documented effects of semantic organization on recall of categorized word lists are entirely a consequence of mediation via category names (Tulving & Pearlstone, 1966). To analyze the effects of semantic organization when subjects recall lists of words that lack a categorical structure requires a fine-grained measure of semantic similarity. One way to construct such a measure is to collect similarity ratings on all possible pairs of words used in a given experiment and then use those ratings to build a metric model of the representational space. The similarity ratings used in this approach can be obtained from 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). Latent semantic analysis (LSA; Landauer & Dumais, 1997) provides an important alternative to the subjective similarity approach. Instead of deriving interitem similarities from subjective responses, LSA uses the natural tendency of words with similar meanings to occur in similar contexts as the basis for deriving a statistical model of a semantic similarity space. LSA can thus provide a measure of the similarity among any pair of words in the English language. One goal of this paper is to examine how LSA fares in predicting output order effects in free recall. Armed with this technique we also endeavor to address the larger question of how semantic and temporal factors interact in predicting recall performance.

Temporal Factors in Free Recall: The Lag-Recency Effect

Temporally defined interitem associations exert a strong influence on output order in free recall (Kahana, 1996). These associations are inferred from subjects' tendency to successively recall items from nearby list positions. Given that a subject has just recalled an item from serial position i, and that the next recall is from serial position j, Kahana (1996) plotted the relationship between recall probability and the lag (separation, in items) between i and j. This measure, the *conditional response probability as a function of lag*, or lag-CRP, defines the distribution of successive recalls as a function of lag and thus provides a measure of contiguity effects in free recall.

Figure 1A shows a typical lag-CRP function, based on data from Murdock and Okada (1970). Positive values of lag $= (j - i)$ 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 contained the subsequence "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.

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 also shows a marked asymmetry: forward recalls are much more likely than backward recalls. The lag-CRP provides a convenient measure of episodically formed interitem associations. These effects have been replicated in numerous free recall studies varying in modality of item presentation, rate of item presentation, and list length (Kahana, 1996). Surprisingly, Howard and Kahana (1999) found that a demanding interitem distractor task did not affect subjects' tendency to form associations among successively studied items. This result, reflecting a scale invariance of associative memory, is fundamentally incompatible with models that assume that interitem association arises from either temporal contiguity or the cooccurrence of items in a working memory buffer (e.g., Raaijmakers & Shiffrin, 1980). This is, however, compatible with notions of temporal coding in which retrieval of an item recovers the temporal context that obtained when the item was studied, which in turn activates items that share that temporal context (Howard & Kahana, 2001).

Analyses of IRTs also reveal associative processes in free recall. In studies where IRTs are recorded between each successive response, one can observe shorter IRTs between successive recall of items that were studied in nearby input positions. We refer to IRTs conditional on recall transitions as conditional response latencies. Figure 1B plots conditional response latency as a function of lag (lag-CRL) for data from Murdock and Okada (1970).

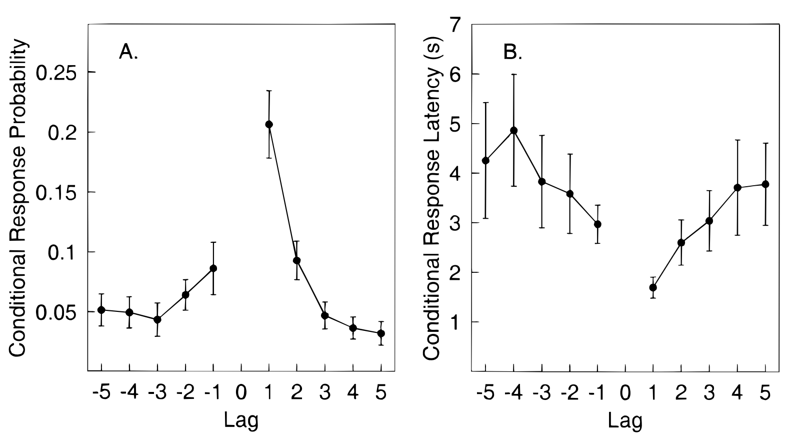

FIG. 1. The lag recency effect. (A) The conditional response probability as a function of lag. (B) The conditional response latency as a function of lag. Both measures show an advantage for recalls to nearby serial positions and an asymmetry favoring forward recalls. Data is from Murdock and Okada (1970). Both measures were calculated excluding the first three output positions to eliminate the recency effect. Error bars are 95% confidence intervals calculated according to the method of Loftus and Masson (1994).

Latent Semantic Analysis as a Measure of Semantic Similarity

In this paper we use a fine-grained measure of preexperimental 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 body of 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 large corpus of naturally occurring text (e.g., an encyclopedia). Let M denote the number of paragraphs in the corpus. Let N denote the number of unique words in the corpus. Using the M paragraphs in the corpus, LSA proceeds as follows. First, define a matrix, T, whose elements, *T(word, paragraph)* record the number of times each word occurs in each paragraph. Each row in T is thus an M-dimensional vector representing a given word's probability of occurring in each paragraph.[1]

Latent semantic analysis is based on the observation that words that are similar in meaning tend to occur in the same paragraphs. Therefore, the M-dimensional rows 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. However, instead of calculating that cosine directly, it is advantageous, to first express the N-dimensional column vectors in terms of a smaller number D of orthogonal N-vectors. Singular value decomposition is used to eliminate correlation between the N dimensions. This leaves an $N \times D$ matrix where D is much smaller than M.The similarities are then calculated as the cosine of the angle between each pair of D-vectors in this reduced matrix.

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.[2] This can be measured by taking the cosine of the angle, θ, between the vectors. In the analyses reported in this paper, we used LSA vectors of 300 dimensions drawn from the TASA-All[3] space and weighted by the singular values.

ANALYSIS 1: LSA cos θ 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 as measured by LSA cos θ. This new measure, the *LSA-CRP*, provides an assay of the effect of preexperimental associations on retrieval.

Calculating the LSA-CRP

Given that a participant recalls two words in succession, we expect those words to come, on average, from nearby points in LSA space. To assess this effect, we can plot the probability of a transition as a function of cos θ between the successively recalled words. To do

[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] 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 high cos θ 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.

[3] See http://128.138.223.70/spaces.html.

so, however, we must first discretize cos θ by dividing the range of cos θ's (-1 to 1) into bins of some small fixed interval. We can then calculate the probability of successively recalling words whose semantic similarity, as measured by cos θ, falls within a given bin. It turns out that the distribution of cos θ values across words chosen from the Toronto noun pool (Friendly, Franklin, Hoffman, & Rubin, 1982) is highly skewed, with the mode very close to zero and the mean at about 0.09. To correct for the *a priori* distribution of cos θ, we transform cos θ for each pair into a percentile score, thus dividing the cos θ distribution into 100 bins containing an equal number of pairs. 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), although in some cases this requires a bit of imagination (e.g., FOREHEAD–PONY). As we move to lower values of cos θ, the bins become more closely spaced and the pairs in the bins become less and less obviously related.

TABLE 1

Examples of Word Pairs Taken from Different Bins of cos θ

Bin	Pair	cos θ
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

Note. The pairs in the table were chosen quasi-randomly. Only very high bins contain predominantly pairs with obvious semantic relationships between them.

The simple approach we have just outlined runs into a minor complication. Specifically, consider what would happen in an imaginary free recall experiment in which subjects' recall transitions are exclusively guided by semantic similarity as measured by LSA cos θ. That is, after recalling the first word, they proceed by selecting the remaining word that has the highest cos θ to the just-recalled word. Further assume that subjects successfully recall every list item. Under these ideal conditions, the probability of recalling a word in the very highest cos θ 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, there would not generally be an item from bin 100 available for every recall, and we would get a probability of 0.1.[4] 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 does not 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 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.

[4] This is, of course, only true in the expectation.

Experimental Methods

To assess the effect of LSA cos θ on the recall transitions, we reanalyzed data from Howard and Kahana (1999, Experiment 2). Here we briefly review the experimental methods of that experiment. Over 10 sessions, each of 16 subjects 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. 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-s arithmetic distractor task prior to attempting free recall. The recall period was fixed at 60 s.

A key variable in Howard and Kahana's 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 the between-item distractor activity, the interpresentation interval (IPI), was 0 (standard delayed free recall), 2, 8, or 16 s continuous distractor free recall.) This section examines whether LSA cos θ predicts recall transitions and IRTs in free recall of "unrelated" word lists. In subsequent sections we examine how this effect changes with varying IPI.

Results

Conditional response probability. Figure 2A shows the LSA-CRP plotted as a function of the mean cos θ value of each bin. The data used to generate this figure were collapsed over all IPI conditions. As can be seen, subjects are more likely to make transitions to words with a higher value of cos θ relative to the just-recalled word. A linear regression of LSA-CRP(bin) to cos $\theta(bin)$ was performed for each subject. The mean slope, 0.21 $\pm$ 0.02 was reliably different from zero, $t(15) = 9.1$, $p < .001$. The intercept was 0.11.

LSA cos θ has a substantial effect on recall transitions in free recall: going from cos $\theta = 0$ to cos $\theta \simeq 1$, the probability of a recall transition increases by about a factor of two. The correlation between LSA-CRP and cos θ was .75. The significant relationship between cos θ 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 $\theta > 0.141$), the slope was reduced to 0.06 $\pm$ 0.03, but it remained significantly different from zero, $t(15) = 2.24$, $p < .05$. This means that the LSA-CRP is sensitive to subtle variations in semantic similarity (compare Table 1).

Conditional response latency. LSA cos θ also affects IRTs in free recall. Figure 2B plots mean IRT as a function of the bin of the cos θ distribution. IRTs are shorter when the successively recalled words are similar (i.e., have high cos θ). The mean of the slopes of a linear regression, -5 ± 1*s*, 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 θ 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

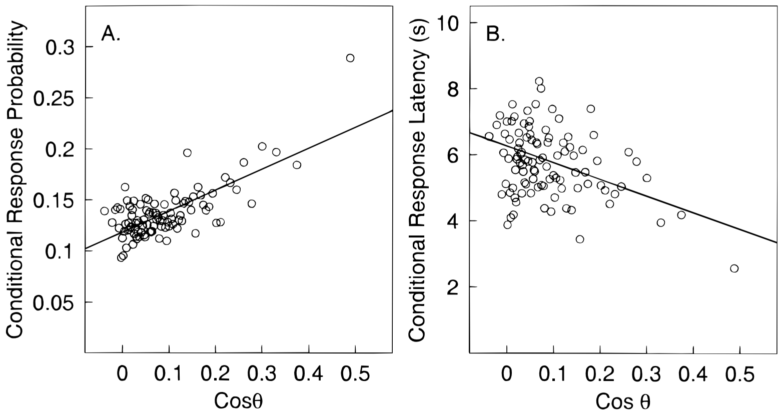

FIG. 2. LSA predicts output order and response latency in single-trial free recall. Data are from Experiment 2 of Howard and Kahana (1999). The distribution of LSA cos *θ* was divided into 100 bins with an equal number of pairs (see Table 1). (A) The conditional response probability as a function of mean LSA cos *θ* for each bin. (B) The conditional response latency as a function of mean LSA cos *θ* for each bin. The lines in each figure represent the average fit of a regression applied separately to each subject's data (see text for details).

about twice as likely to be recalled in succession as very low similarity pairs. Further, LSA cos *θ* affects latency in free recall. High-similarity pairs are recalled on average several *seconds* faster than low-similarity pairs.

ANALYSIS 2: THE EFFECT OF TEMPORAL AND 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 3 illustrates the lag-recency effect for several levels of the IPI in Experiment 2 of Howard and Kahana (1999). As can be seen, the lag-recency effect was not reduced by increasing the IPI from 0 to 16 s. 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 each list, was sufficient to eliminate the end-of-list recency effect.

Insofar as the lag-recency effect is insensitive to the absolute delay between list items (Fig. 3), 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 produces episodic associations (as postulated by Atkinson & Shiffrin, 1968; Glanzer, 1972; Raaijmakers & Shiffrin, 1980), one should see a lag-recency effect because nearby items spend more time together in short-term memory. However, because a long interitem distractor should disrupt short-term memory, the scale-invariance of the lag-recency effect requires an alternative hypothesis. To explain this persistence we proposed that the associative process evident in the lag-recency effect reflects mediation by a gradually changing contextual representation (Howard & Kahana, 2001).

Although several studies have manipulated semantic variables in continuous-distractor free recall (Greene & Crowder, 1984; Greene, 1986; Gregg, Montgomery, & Castaño, 1980) it remains unknown whether semantic similarity in-

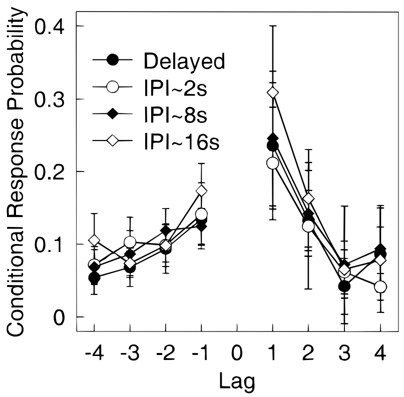

FIG. 3. The approximate scale invariance of the lag-recency effect. Shown is the conditional response probability (CRP) function for each of the four conditions of Experiment 2 of Howard & Kahana (1999). The fact that each of these curves is peaked around 0 illustrates the lag-recency effect. The advantage 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 interpresentation interval (IPI), which was varied between 0 (the curve labeled "Delayed") and 16 s. During this IPI, subjects performed a demanding arithmetic task. Despite 16 s of interitem distractor, the CRP is largely unaffected. Error bars are 95% confidence intervals.

fluences output order when the presentation of list items is separated by a demanding distractor task. Here, we directly examine whether a manipulation of the temporal relations among list items affects the impact of semantic similarity on retrieval.

Results

Figure 4A plots the slope of the LSA-CRP for the four IPI conditions of Experiment 2 of Howard and Kahana (1999). LSA cos θ had a significant effect on output order, as evidenced by a significant positive slope of the LSA-CRP, in each of the four experimental conditions ($p < .01$ for each condition). The interesting result, as shown in the figure, is that the LSA-CRP slope decreased with increasing IPI. To assess the significance of this effect, we regressed the slope of the LSA-CRP on the duration of the IPI (IPI was 0 for delayed free recall, 2 for IPI = 2, and so on). This regression revealed a significant negative slope of -0.008 ± 0.002, $t(15) = 3.4$, $p < .01$. Thus, the temporal proximity of the list items influences the effect of semantic proximity on retrieval transitions: when items were separated by a long distractor, the retrieval transitions were less likely to be driven by semantic relations. In contrast, this manipulation of IPI had virtually no effect on the lag-recency effect, as shown in Fig. 4B.

Discussion

If temporal and semantic effects on retrieval are functionally distinct, then increasing the temporal separation of list items ought to reduce temporally driven recall transitions without affecting semantically driven recall transitions. In this case, the lag-CRP should decrease, and the LSA-CRP should remain constant. Our data (Fig. 4) show exactly the opposite pattern. Increasing the temporal separation among the list items had virtually no effect on participants' tendency to recall items from nearby list positions successively. But this manipulation of a temporal factor did influence participants' use of semantic similarity as a retrieval cue: The LSA-CRP, which was significantly greater than zero for all values of IPI, decreased systematically as the IPI was increased. This means that semantic similarity has a smaller effect on recall transitions as the temporal separation of list items increases.

Howard and Kahana (1999) interpreted the scale-invariance in the lag-recency effect, as measured by the lag-CRP, as being analogous to the scale invariance in the recency effect, as demonstrated by studies of long-term recency (Glenberg et al., 1980; Nairne et al., 1997). Both recency and lag-recency effects can be understood in terms of a competitive retrieval mechanism. If a randomly varying temporal context is associated with each list item, then presentation of context at time of test will differentially activate each list item, with recent items being "stronger" than older items. But if retrieval is competitive, recency will only be sensitive to the relative temporal separation among list items. Similarly, if recalling an item retrieves the temporal context associated with it during encoding, then items studied at nearby time points will be activated, thus producing the lag-

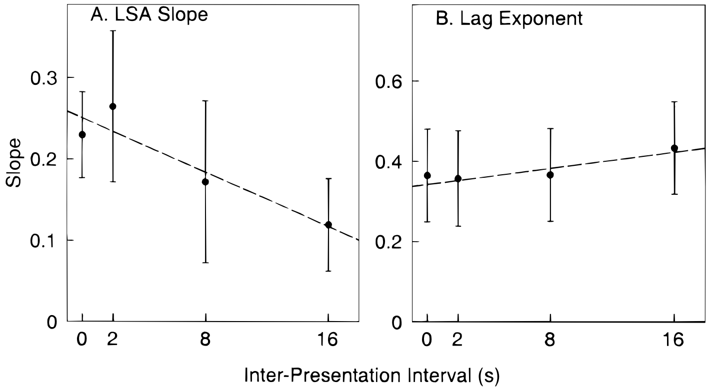

FIG. 4. Summary of results across conditions of Experiment 2 of Howard and Kahana (1999). These conditions varied in the duration of the distractor activity between presentation of the list items, the interpresentation interval (IPI). (A) The LSA slope (see the line in Fig. 2A) measures the overall effect of semantic factors on recall transitions. As the IPI increases, LSA slope gradually decreases. For this graph, LSA slope was calculated collapsed over output position. Error bars reflect 95% confidence intervals calculated according to the procedure of Loftus and Masson (1994). The dashed line is the mean linear regression, Slope $= (-0.008 \pm 0.002)$IPI $+ 0.25 \pm 0.03$. (B) An analogous figure for the effect of IPI on the lag-CRP effect. Power functions of the form CRP $= A|\text{lag}|^{-B}$ were fit to each lobe of the lag-CRP for each subject. Shown is the average value of B. The dashed line represents the mean linear regression Exponent $= (0.005 \pm 0.004)$IPI $+ 0.34 \pm 0.05$. The error bars represent 95% confidence intervals calculated according to the method of Loftus and Masson (1994).

recency effect. Here too, competitive retrieval will ensure that it is the relative proximity rather than absolute proximity that determines the lag-recency effect.

But, if competitive retrieval explains the approximate scale invariance in lag-recency (Fig. 4B), what explains the decrease in the influence of semantic similarity as temporal separation is increased? The following analysis considers one possible explanation.

ANALYSIS 3: TESTS OF AN ENCODING EXPLANATION OF THE LSA-CRP

According to the classic two-store model of human memory (e.g., Atkinson & Shiffrin, 1968), semantic relationships influence recall from LTS but exert no effect on recall from STS (see Glanzer, 1972, for a review). Similarly, it was hypothesized 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. Because items drop out of STS as new items are encoded, neighboring items spend more time together in STS. This could account for the LSA-CRP and its attenutation by a demanding interitem distractor (see Fig. 4A). The longer the distractor, the lower the probability of two items occupying the buffer together. Because co-occupancy increases the strength of the *semantic* association, increasing IPI should decrease the slope of the LSA CRP. If this process were the sole source of semantic effects on output order, then a sufficiently long interitem distractor ought to eliminate the effect of semantic similarity on output order.

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 be in STS if it were the previous item than if it were presented earlier in the list. If such organizational processes in STS are important in driving the effect of cos θ on recall transitions, then there should be a bigger effect for words that were close together in the list than for words that were far apart in the list. A prediction of this account, 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.

This STS-based explanation is one specific instance of a broad class of encoding-based accounts. One can readily substitute working memory or study-phase retrieval for STS. Any organizational process that requires the co-activation of the to-be-organized items, where activation is a function of recency, can predict the qualitative pattern shown in Fig 4A.

Method

We calculated CRP as a joint function of lag and LSA bin for each subject. To calculate these probabilities, we kept track of a matrix of numerators and denominators corresponding to LSA-bin and lag. We collapsed lag into |lag| and collapsed |lag|*s* ≥ 9 into a single bin to increase statistical power. We then calculated the LSA-CRP separately for each value of |lag|. Our interest was in whether the LSA-slope is greater for small values of |lag| than for large values of |lag|.

Because subjects make more recall transitions at short |lag|*s* than at long |lag|*s* (see Fig. 3), and because we were interested in the relative effect of cos θ, 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 effect of LSA decreases as |lag| increases.

Results

Delayed free recall (IPI $=$ 0 s) showed a significant interaction slope (-0.5 ± 0.1, $t(15) = 5.1$, $p < 0.001$). 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 (-0.1 ± 0.1, $t(15) < 1.0$). 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. The pairwise comparisons between delayed free recall and each of the continuous-distractor conditions were significant ($p < .05$). None of the comparisons between the continuous-distractor conditions approached significance.

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 θ on the probability of recall transition for both conditions (Fig. 4A), there was a significant interaction slope for the delayed condition, and not for the IPI $=$ 2 s condition. Further, the interaction slopes from these two conditions differed from each other significantly. The effect of LSA on output order is apparently dissociable from the interaction slope. As a consequence, we can conclude that encoding processes, of the type proposed by Glanzer (1972), are not the sole cause of the LSA-CRP. Further, such encoding processes cannot explain the attenuation of the LSA-CRP with increasing IPI.

The elimination of the interaction slope with the inclusion of an interitem distractor of any duration suggests that the effect is a consequence of active rehearsal processes that are easily disrupted by an interitem 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 bridges the two main approaches to the study of free recall. One approach examines the effects of semantic factors. The other approach examines the effects of temporal factors. We 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.

Analyzing a large free-recall data set reported in Howard and Kahana (1999), we found that LSA cos θ had a substantial effect on both output order and interresponse times. We measured this effect by computing the conditional response probability and latency as a function of LSA cos θ (the LSA-CRP). We found that words with very high LSA cos θ to the just-recalled word were about twice as likely to be recalled as words with low cos θ (see Fig. 2A). Further, the IRT associated with pairs of words with a very high cos θ was several *seconds* faster than that associated with pairs of words with low cos θ (see Fig. 2B).

The key variable in Howard and Kahana's study was the duration of an arithmetic distractor task between the presentation of successive list items. Across four conditions, the duration of the between-item distractor activity, the interpresentation interval (IPI), was 0 (standard delayed free recall), 2, 8, or 16 s.

With this design we were able to examine how varying IPI influenced participants' use of semantic similarity in guiding retrieval. Although semantic similarity predicted recall transitions for all levels of the IPI, the slope of the LSA-CRP declined significantly from a mean of 0.26 in the IPI = 2 condition to 0.12 in the IPI = 16 condition (see Fig. 4A). There was little or no difference between the IPI = 0 and IPI = 2 conditions, $t(15) < 1.0$. Clearly, increasing the duration of an interitem distractor reduces the influence of semantic similarity on output order.

This result is surprising when one considers that the same manipulation (varying the IPI in continuous-distraction free recall) did not influence participants' use of temporal proximity in guiding recall. This can be seen in the lag-CRP functions shown in Fig. 3 and the analysis of the lag-CRP exponents in Fig. 4B. Howard and Kahana (1999; see also, Howard & Kahana, 2001) interpreted this scale invariance in the utilization of temporal proximity in recall as evidence for contextual retrieval coupled with a competitive retrieval mechanism. Competitive retrieval will ensure that it is the relative proximity rather than absolute proximity that determines the lag-CRP.

Retrieval Transitions in Free Recall

Most recent work on formal models of episodic memory (e.g., Chappell & Humphreys, 1994; Dennis & Humphreys, 2001; 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 accurately describe the serial position curve without describing the basic properties of memory retrieval reflected in the transitions. Conversely, a description of the transitions[5] nearly guarantees a description of the serial position curve. The CRP simplifies the task of the theorist.

The lag-CRP described in Kahana (1996) has proved to be a valuable tool in distinguishing

[5] 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).

between different models of the episodic component of free recall (Howard & Kahana, 1999, 2001). The LSA-CRP (Figs. 2 and 4A) may prove to do the same for the study of semantic factors. Although we have shown that LSA is a useful measure, we make no claim as to the relative usefulness of LSA as compared to free association norms, subjective judgments of similarity, or any other measure of semantic similarity. Indeed, an analogue of the LSA-CRP could be calculated for any other measure of semantic similarity.

Explanations Based on the Operation of Short-Term Memory

If separating items by a difficult distractor task does not affect the utilization of temporal associations, why should this manipulation influence the utilization of semantic associations? One possible explanation is offered by theories of short-term memory (Atkinson & Shiffrin, 1968). Glanzer (1972) proposed that the semantic association between two studied items in long-term memory is strengthened when they are rehearsed together in a limited-capacity short-term store (STS). This account of semantic processing in free recall predicted the finding that semantic associations differentially affect recall of prerecency items (Glanzer & Schwartz, 1971), and that recall of semantic associates is enhanced when they appear in nearby list position (Glanzer, 1969).

According to this view, subjects "discover" the weak semantic relations among items that appear in nearby positions in our random word lists. These associations, in turn, influence output order and IRTs, as shown in Fig. 2. Because the interitem distractor disrupts rehearsal, increasing the IPI should substantially reduce subjects' ability to encode these associations. This could explain the decrease in the LSA-CRP with increasing IPI (Fig. 4B).

But this account also predicts that semantic similarity should have a bigger effect on recall transitions for adjacent list items. In the IPI = 0 s (delayed free recall) condition this is exactly what we we found. The LSA-CRP was much steeper for word pairs from adjacent list positions.[6] However, this effect proved to be fragile; though it was highly significant in the IPI = 0 s condition, there was no trend in this direction for any of the three continuous distractor conditions (IPIs of 2, 8, and 16 s). This suggests that the decrease in the LSA-CRP across these conditions is a consequence of some other process. In addition, we know that this class of STS-based accounts fails to explain the approximate scale-invariance of the lag-recency effect (see Figs. 3 and 4B, also Howard & Kahana, 1999). In the next subsection, we work toward an alternative account of the principal finding of this paper, namely, the decrease in the efficacy of semantic similarity as the temporal separation of the list is increased.

Episodic Associations Are Contextually Mediated

Suppose that different memory cues interact during retrieval, such that the effectiveness of one type of cue is enhanced by the presence of another strong cue. This is the case, for example, in 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. So, if item *j* has just been recalled, then the probability of sampling item *i* for recall could be given by

$$P_s(i \mid j) = \frac{S_{Ci} S_{ij}}{\sum_k S_{Ck} S_{kj}},$$

[6] It is not necessary to postulate some form of STS to explain this result. 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.

where S_{ij} is the associative strength between items *i* and *j* in long-term memory and S_{Ci} is the associative strength between context (*C*) and item *i*.

The interitem 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 interitem associative strengths, then each entry of the matrix would represent a sum of one term for preexisting associations and another term for newly formed associations. In this case, increasing the strength of the newly formed associations, as a consequence of temporal contiguity (small IPI), should *diminish* the relative importance of preexisting semantic associations. The present result, showing that the effect of semantic similarity on recall transitions *decreases* with increasing IPI (see Fig. 4A), argues against this view. If, however, the episodic associations reflected in the lag-recency effect are not a consequence of changes in the item-to-item strength S_{ij}, but rather are mediated by the contextual cue, S_{Ci}, then this difficulty is eliminated.

If the lag-CRP and the LSA-CRP arise from different cue strengths, then there is at least the possibility of explaining the effect of IPI on the LSA-CRP (Fig. 4A) within the context of SAM. If the lag-CRP is driven by variation in S_{Ci}, and the LSA-CRP is driven by variation in S_{ij}, then the contribution of semantic similarity is mediated by context cue strengths in the sampling equation—if S_{Ci} is large, then variation in S_{ij} will have a bigger effect on sampling, and eventually recall, than if S_{Ci} is small. It is obviously an open question as to whether such a scheme could simultaneously describe the persistence of the lag-recency effect (Fig. 3), the decrease in LSA-slope with increasing IPI (Fig. 4A), and the lack of an interaction slope for all non-zero values of IPI (Analysis 3).

Conclusions

We have shown that a temporal manipulation of list structure exerts a strong influence on the role of semantic similarity in episodic retrieval. When list items are presented without interruption, semantic similarity has its greatest effect. As the list items become separated by longer and longer intervals, the effect of semantic similarity becomes progressively weaker. This is not because rehearsal of nearby items in a short-term store facilitates discovery of subtle semantic relations. Although such an effect is probably at work in ordinary free recall, it is not present under conditions of continuous distraction. A more plausible account may involve the interaction of semantic and temporal cues at retrieval.

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, pp. 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.

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

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 in 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, pp. 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.

Glenberg, A. M., Bradley, M. M., Stevenson, J. A., Kraus, T. A., Tkachuk, M. J., & Gretz, A. L. (1980). A two-process account of long-term serial position effects.

Journal of Experimental Psychology: Human Learning and Memory, **6,** 355–369.

Greene, R. L. (1986). A common basis for recency effects in immediate and delayed recall. *Journal of Experimental Psychology: Learning, Memory and 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,** 240–245.

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

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. (2001). A distributed representation of temporal context. *Journal of Mathematical Psychology*, in press.

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 & Review*, **7,** 516–521.

Laming, D. (1999). Testing the idea of distinct storage mechanisms in memory. *International Journal of Psychology*, **34,** 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.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, **1,** 476–490.

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). Inter-response 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, pp. 207–262). New York: Academic Press.

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. (2000). Part-set cuing of order information: Implications for associative theories of serial order memory. *Memory & Cognition*, **28,** 847–855.

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

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

(Received July 30, 2000)
(Revision received January 16, 2001)