Psychological Review
2013, Vol. 120, No. 1, 155–189

0033-295X/13/$12.00 DOI: 10.1037/a0030851

A Buffer Model of Memory Encoding and Temporal Correlations in Retrieval

Melissa Lehman
Purdue University and University of South Florida

Kenneth J. Malmberg
University of South Florida

Atkinson and Shiffrin's (1968) dual-store model of memory includes structural aspects of memory along with control processes. The rehearsal buffer is a process by which items are kept in mind and long-term episodic traces are formed. The model has been both influential and controversial. Here, we describe a novel variant of Atkinson and Shiffrin's buffer model within the framework of the retrieving effectively from memory theory (REM; Shiffrin & Steyvers, 1997) that accounts for findings previously thought to be difficult for such models to explain. This model assumes a limited-capacity buffer where information is stored about items, along with information about associations between items and between items and the context in which they are studied. The strength of association between items and context is limited by the number of items simultaneously occupying the buffer (Lehman & Malmberg, 2009). The contents of the buffer are managed by complementary processes of rehearsal and compartmentalization (Lehman & Malmberg, 2011). New findings that directly test a priori predictions of the model are reported, including serial position effects and conditional and first recall probabilities in immediate and delayed free recall, in a continuous distractor paradigm, and in experiments using list-length manipulations of single-item and paired-item study lists.

Keywords: episodic memory, context, memory models, compartmentalization

Supplemental materials: http://dx.doi.org/10.1037/a0030851.supp

Atkinson and Shiffrin's (1968) theory of memory was built on the architectures envisioned by James (1890) and Broadbent (1958) to include formal descriptions of cognitive control that exemplified the new possibilities of the cognitive revolution in psychology. Their theory proposed *buffer processes* that managed information to suit the goals of the subject, and they showed that these processes could play a crucial role in understanding many phenomena that characterize human memory. Since their proposal, some have questioned whether buffer processing is necessary to account for these phenomena or even whether it is a useful theoretical construct (e.g., Cowan, 1997; Craik & Lockhart, 1972; Crowder, 1982; Howard & Kahana, 2002; Sederberg, Howard, & Kahana, 2008; Usher, Davelaar, Haarmann, & Goshen-Gottstein, 2008). The critique often focuses on the proposed encoding operations, recency effects, and a structural distinction between long-term and short-term memory. The structural debate, in particular, has encouraged a fundamental mischaracterization of the theory. By implying that the buffer is a vessel for the temporary storage of information or a set of activated episodic representations, it actually obscures the critical assumption. The buffer is independent of whatever mnemonic structures one might propose; it is a control process that manipulates information to serve task demands and the goals of the subject.

Our focus in this article is on how goals and task demands directly influence patterns of mnemonic behavior otherwise thought to be invariant. Teasing apart the contributions to episodic memory of structural versus control processes is a challenge. Although we cannot directly observe the buffer process, we can observe how the behavior of the subject reflects control operations that are specified by the theory with the use of formal modeling. For instance, others have demonstrated that the various patterns of recency effects often used as a basis to reject the theory are perfectly consistent with a buffer model (Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005; Healy & McNamara, 1996). In this article, we extend these observations and link them within our framework to other phenomena thought to challenge the buffer model. Our buffer model is developed within the retrieving effectively from memory theory (REM; Shiffrin & Steyvers, 1997). Building on developments in the theoretical framework that have been made over the past several years (Gillund & Shiffrin, 1984; Lehman & Malmberg, 2009; Malmberg & Shiffrin, 2005; Raaijmakers & Shiffrin, 1980, 1981), we address recency effects, temporal correlations, output order, individual differences, and maintenance rehearsal within a modern, comprehensive process-level framework of human memory. In addition, we present new findings collected from experiments that were

This article was published Online First December 10, 2012.

Melissa Lehman, Department of Psychological Sciences, Purdue University, and Department of Psychology, University of South Florida; Kenneth J. Malmberg, Department of Psychology, University of South Florida.

This research served as partial fulfillment of Melissa Lehman's doctoral dissertation at the University of South Florida. We thank her committee members, Mark Goldman, Cathy McEvoy, Jon Rottenberg, and Joe Vandello, along with Doug Nelson, for their helpful feedback.

Correspondence concerning this article should be addressed to Melissa Lehman, Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907. E-mail: lehman8@purdue.edu

specifically designed to directly test the present buffer model and that pose challenges for those theories that do not posit a role for cognitive control in the encoding and retrieval of information from memory.

The Scope of the Present Investigations

In this article, we describe a buffer model in the REM framework (Lehman & Malmberg, 2009; Malmberg & Shiffrin, 2005; Shiffrin & Steyvers, 1997) that accounts for several challenging findings. As such, the model was not developed to account for these phenomena. Rather, the immediately pertinent predictions of the model are derived from a set of general assumptions that happens to allow the framework to address several outstanding issues and generate some provocative novel predictions.

In the prior investigations, we found it useful to revisit the role of associative information in episodic memory (cf. search of associative memory; SAM; Raaijmakers & Shiffrin, 1980), and we have become increasingly cognizant of the importance of a detailed analysis of the time course of encoding of different types of episodic information (Criss & Malmberg, 2008; Malmberg & Nelson, 2003; Malmberg & Shiffrin, 2005). The result is a somewhat more realistic buffer model of encoding that accounts for several findings that were previously beyond the scope of the theory. It is also within this framework that we address several outstanding issues and derive predictions concerning the effect of the buffer on the operation of human memory.

We begin with a discussion of the Lehman and Malmberg (2009) buffer model, including a description of the basic REM model for free recall and recognition (Malmberg & Shiffrin, 2005; Shiffrin & Steyvers, 1997). Next, we examine the model's predictions related to maintenance rehearsal, primacy and recency effects, the continuous distractor paradigm, chunking operations, and list-length manipulations. We also discuss reaction time data and how the predictions of the model apply to the temporal dynamics of free recall. Finally, the model is compared to other single-store and dual-store models that have been used to account for data in related paradigms. In conclusion, the evidence that we present indicates that a buffer model can account for several findings that many have used to reject such models and that the buffer model predicts several new findings that neither a single-store nor a dual-store model can account for without the incorporation of flexible control processes, such as a rehearsal buffer.

A Buffer Model of Encoding and Temporal Correlations in Retrieval

The Buffer Model Briefly

The buffer model was initially developed to account for the memorial costs and benefits that occur due to the context change associated with directed forgetting (Lehman & Malmberg, 2009). Our model incorporates ideas developed by Atkinson and Shiffrin (1968), Raaijmakers and Shiffrin (1980), Mensink and Raaijmakers (1989), Shiffrin and Steyvers (1997), and Malmberg and Shiffrin (2005) into a single, unified model. The next section describes the formal details of the model. There are two critical components that distinguish this model from prior models: (a) information can be maintained or dropped by the buffer in order to achieve a reasonable or otherwise desirable encoding strategy (Lehman & Malmberg, 2011), and (b) information about items, context, and associative information is stored according to different time courses (Malmberg & Shiffrin, 2005; Raaijmakers & Shiffrin, 1980).

REM

Representation. REM assumes that memory consists of lexical/semantic images, which contain features representing knowledge about items and all the contexts in which the items have been encountered (Shiffrin & Steyvers, 1997). During the study of a list of items, a separate episodic image is stored for each item; the image consists of one or more associations, including the association between the studied item and the study context. Additionally, associations may be stored between the studied items and other studied items. Item-to-context and item-to-item associations play a critical role in the model. The composition of the episodic image is determined by the operations of a limited-capacity rehearsal buffer (Lehman & Malmberg, 2009; Raaijmakers & Shiffrin, 1981; see also Atkinson & Shiffrin, 1968; Malmberg & Shiffrin, 2005).

When a study list consists of a random sample of words from the language and when the words are presented in succession, item features change randomly from trial to trial. In contrast to the rapidly changing item features, changes in context reflect significant changes in tasks, goals, setting, and so on, and therefore the context on trial n is likely to be very similar to the context on trial $n - 1$ (e.g., Annis, Malmberg, Criss, & Shiffrin, 2012; Klein, Shiffrin, & Criss, 2007). Because the goals and activities of the subject change when the task changes, a more significant change in context occurs between the study phase and the testing phase of the experiment (Lehman & Malmberg, 2009).

Rehearsal versus compartmentalization. The contents of the buffer, including both items and context, are managed to serve specific objectives (Atkinson & Shiffrin, 1968). This requires identifying the items that are to be processed by the buffer, identifying the operations performed on them, and determining when and which items should no longer be the focus of processing. Thus, when one is modeling buffer operations, key questions concern the number or items to be processed and initiation and halting of the processing, as the representations of the events depend critically on how these questions are answered.

Maintaining an item so that it is available for processing is referred to as *rehearsal*. Atkinson and Shiffrin (1968) assumed that when a list of items is being studied, newly encountered items are subject to rehearsal. The number of items to be simultaneously rehearsed depended on the task and goals of the subject, but given a sufficiently long list of items to study, the capacity of the buffer was reached at some point. According to Atkinson and Shiffrin's buffer model, items were randomly selected to be "knocked out" of the buffer. Lehman and Malmberg (2009, 2011), on the other hand, speculated that specific items may also be intentionally dropped from the buffer when they are no longer needed to perform a certain task and that contextual elements may be managed to aid or impede access to long-term images. We have referred to the control process responsible for the intentional change in the contents of the buffer as *compartmentalization*. Whereas rehearsal seeks to maintain information in the buffer, compartmentalization seeks to change it. In this sense, compartmentalization and rehearsal are complementary control processes.

One task in which compartmentalization is revealed is intentional forgetting, which is often examined by having subjects study multiple lists. After a critical list, they may or may not receive an instruction to forget it. The instruction to forget imparts a significant amount of forgetting of the items from the critical list (Bjork, LaBerge, & Legrand, 1968). However, this forgetting is viewed as a consequence of compartmentalization, not as a goal to forget the occurrence of the items on the critical lists. The goal is to remove items from the buffer in order to free limited resources by rendering interfering traces less accessible than they would otherwise be (Lehman & Malmberg, 2011).

Accordingly, Lehman and Malmberg (2009) assumed that the instruction to forget encourages the subject to direct his or her attention away from the prior items and mentally change the context present prior to the instruction to forget.[1] The switch to different contextual features causes mismatches between the context stored in the "to-be-forgotten" traces and context that will be stored on subsequent trials (Sahakyan & Kelley, 2002). Thus, removing to-be-forgotten items from the buffer and shifting the context to the to-be-encoded words causes the forgetting of the to-be-forgotten items, enhances memory for the to-be-remembered items, and enhances list discrimination (Lehman & Malmberg, 2009). Although accounting for intentional forgetting is not the focus of the present investigation, the idea that the buffer controls via rehearsal and compartmentalization both what information is rehearsed or encoded and what information is not rehearsed or encoded is the critical assumption for several tests of the model that we present later.

The time course of encoding. It has been become increasingly apparent that the information comprised by episodic memory traces differs both in what it represents and in the time course of encoding. For instance, Malmberg and Nelson (2003) proposed that early during the encoding of a word, lexical features are encoded, perhaps as a by-product of lexical access, and it is the storage of these lower level features that produces the word-frequency mirror effect for recognition memory (Glanzer & Adams, 1985; see also Criss & Malmberg, 2008).[2] Once the meaning of the word is retrieved from lexical/semantic memory, encoding focuses on higher level features. During this later phase of encoding, attention may be turned to other items in the buffer and/or words associated in long-term memory with the current study word, whose meanings have been retrieved to create a richer episodic trace.

The analyses of Malmberg and Shiffrin (2005) suggest that under common conditions, context is stored in perhaps the first 1 to 2 seconds subsequent to the item presentation, whereas item features may be stored over a longer time course. Thus, increasing the amount of time an item is studied beyond 2 seconds does not increase the amount of context stored. Context storage is enhanced, on the other hand, by spaced repetitions, as contextual features are accumulated in a trace stored on a prior study trial. This "one-shot" hypothesis explains several dissociations between implicit and explicit memory that are related to conditions under which items are studied, why context-dependent recognition memory increases with increases in spaced but not massed repetitions (Murnane & Phelps, 1995), and findings from our lab that spaced but not massed repetitions enhance list discrimination (Malmberg, 2012).[3]

Together, the early-phase encoding hypothesis and the one-shot hypothesis suggest a model in which context features and low-level item features are encoded relatively quickly subsequent to the presentation of the study item, whereas high-level item features belonging to the current study items and other items in the buffer are encoded for a more extended period of time. A more detailed account of the encoding and the buffer is required to capture these assumptions about the time course of encoding. Like Atkinson and Shiffrin (1968), Lehman and Malmberg (2009) proposed that as items are presented, they enter and leave the buffer. While they are attended to, they reside in the buffer, and information may be encoded about them in one or more episodic traces. The extent of encoding for a given item on a given trial depends on the number of items concurrently being rehearsed, which is influenced by task demands, the nature of the stimuli, and so on. For instance, Atkinson and Shiffrin's analyses indicated that the buffer capacity varied according to task demands ranging from two to five items. This suggests that the capacity of the buffer may not always be fully utilized and/or that different encoding operations impose different demands on its capacity.

What is encoded is determined by the goals of the subject. When studying a list, we assume, subjects engage a range of control processes that are available for the encoding of new traces, such as creating sentences out of the items and using interactive imagery. During this phase of encoding, various types of information may be encoded depending on the nature of the task, the goals of the subject, and prior experience. In addition, associations are created after the items have been identified and their features and context have been encoded.

The particular strategy employed to create these associations likely reflects the subject's past experiences and goals, and a long literature indicates that the encoding operations lead to nonrandom sequences of output during free recall (Bousfield & Puff, 1964; Tulving & Pearlstone, 1966). Of particular interest, a growing amount of evidence indicates that items from adjacent study positions are more likely to be recalled in adjacent output positions than items from more distant serial positions (Howard & Kahana, 1999; Kahana, 1996; but see Farrell & Lewandowsky, 2008). This is referred to as the lag-recency effect (Howard & Kahana, 2002). It suggests that interitem associations among items residing in the buffer are created (Raaijmakers & Shiffrin, 1980). These items may be items that remained in the buffer as the result of earlier

[1] Some of the mechanisms underlying intentional forgetting appear to be similar to those that induce context-dependent forgetting. For instance, memory is impaired when testing occurs in a context other than the one in place during study (Godden & Baddeley, 1975), and Sahakyan and Kelley (2002) found that a change in context has the same effect on memory for the critical list that the instruction to forget does. Moreover, reinstating the context encoded during the study of the critical list eliminates the deficit caused by the instruction to forget.

[2] The assumption that lower level lexical features are the focus of an initial phase of encoding explains, among other things, why the word-frequency effect does not increase after several hundred milliseconds of study and why probing memory with nonlexical features does not produce a mirror effect (Criss & Malmberg, 2008; Malmberg & Nelson, 2003).

[3] The method and results of this experiment are available from KJM on request.

study trials, items that are retrieved from long-term images representing past study trials, or pre-experimental knowledge.

The orderly fashion in which items are output from memory is captured in REM by the storage and retrieval of interitem associations (Xu & Malmberg, 2007). Although SAM assumed a prominent role for interitem associations created among items occupying the buffer contemporaneously (Raaijmakers & Shiffrin, 1980), we take advantage of REM's multidimensional representations concerning the nature of capacity limitations. In particular, we assume that subjects associate items residing in the buffer by storing traces that contain features from both the current item and previously studied items. Thus, although some encoding resources are dedicated to encoding the currently presented item, contextual features are also being stored, along with some features of the previously presented item (and perhaps earlier items, depending on the buffer size). The concatenation of item, context, and associative vectors produces two types of associations in a single trace. There is the item-to-context association, which indicates that the current item occurred in the study context, and the interitem association, which represents the co-occurrence of two items in the buffer.

The model further assumes that encoding capacity is split between the storage of item, context, and interitem associative information. There is an upper limit on the amount of active encoding that may take place in a finite amount of time. When resources are plentiful, this limit is not a major issue, and a great deal of interitem associative information may be stored. However, when resources are taxed, fewer, weaker associative traces are the result. For example, consider the second study trial where now two buffer items compete for encoding resources. Most of the resources are initially spent encoding item *n,* the second item on the list, and an item-to-context association. The remainder of the encoding resources is divvied up between the storage of interitem associative information and context-to-item associations among the items currently being rehearsed. When multiple items are associated to the context, as takes place when more than one item is in the buffer, the context-to-item associations that are created are relatively weak. Thus, items studied in the earliest serial positions have stronger item-to-context associations, and they are therefore more likely to be recalled when a context cue is used to probe memory (see Sederberg et al., 2006). As a result, the model predicts primacy in the serial position curve and predicts that the first item output in delayed free recall is most likely to be first item studied (Lehman & Malmberg, 2009). The limited capacity to encode item-to-context associations is a critical assumption of our buffer model.

Retrieval–free recall. Retrieval in free recall is conceived of as a series of sampling and recovering operations in REM (Lehman & Malmberg, 2009; Malmberg & Shiffrin, 2005; Shiffrin & Steyvers, 1997). The likelihood of sampling a trace is determined by the match between the features in that trace and the features in the cue used to probe memory. In free recall, subjects use a context cue to initiate sampling. The content of context cue determines from what part of a list the subject begins recalling, and this may be influenced by task demands, the retention interval, and so on. For a task where subjects study multiple lists and then must recall from a specified list, an ideal cue is one that is specific to that list insofar as it would maximally reduce interference from the nontarget list in memory. For instance, if the lists consist of exemplars drawn from mutually exclusive categories, the category information, if used as a retrieval cue, helps to reduce interference from the other lists (Lehman & Malmberg, 2011). However, in many experimental conditions, the available contextual cues are far more impoverished, and often the best the subject can do is to use temporal contextual information to probe memory. Temporal context is information that is correlated within a list and even between lists. To the extent that the temporal context is correlated among traces or lists, interference from nontarget traces will take place and free recall will suffer (Lehman & Malmberg, 2009).

A key question concerns how contextual cues are reinstated, especially given the wealth of evidence that subjects often initially retrieve the first item studied (e.g., Lehman & Malmberg, 2009). It is obvious that the better temporal context is encoded, the easier it will be to reinstate at the time of retrieval. According to the buffer model, the strength with which context is encoded depends on the number of items being rehearsed. Hence, temporal context features associated to items studied near the beginning of the list are most easily reinstated. Under typical encoding conditions, this gives a competitive advantage to the first item on the list, and therefore it most likely to be the first item recalled.

It also seems obvious that tasks that elicit more context change (e.g., a directed forgetting task) make it more difficult to reinstate context features. Therefore, we have assumed that the cue will consist of larger proportions of current "test" context features as the amount of context change between study and test increases (Lehman & Malmberg, 2009). These assumptions also play an important role in our modeling of increases in the retention interval in some of the experiments that we report here. However, at this time, we do not specify a formal model of context reinstatement.

In order to account for buffer operations during retrieval, we assume that immediate free recall is usually initiated by sampling only from the buffer, using only the most recent context as a cue, and that the probability of sampling from the buffer increases as it becomes more difficult to reinstate the context features from the beginning of the list. After sampling from the buffer occurs (the number of attempts to sample from the buffer is equal to the size of the buffer), retrieval continues by sampling from the entire list. Regardless of whether samples are taken from the buffer or long-term traces, when an item is successfully retrieved, the next cue used to probe memory consists of the previously used context cue, updated with some context and item features stored in the retrieved trace. Thus, a co-rehearsed item will be most likely to be sampled next, as it will share some of the item information and have similar context features.

Formal Model

Here we describe how to implement the aforementioned assumptions in a REM framework suitable for conducting simulations. Those not immediately interested in this aspect of the theory may skip to the next section.

Items, Context, and Associations

Lexical/semantic images are vectors of w geometrically distributed feature values representing knowledge about items and all the contexts in which they have been encountered. For words, the item knowledge consists of orthographic, phonological, and meaning features. The single-parameter (g) geometric distribution is convenient to use. However, the key assumption is that features vary in the rate in which they occur in the environment. As g increases, the mean of the geometric distribution decreases and lesser feature values become more common. The assumption that rare words are composed of more uncommon features than common words produces a mirror effect for recognition memory (Malmberg, Steyvers, Stephens, & Shiffrin, 2002). However, the geometric distribution does not play an important role in the present modeling, and g is set to .4, which is its default value (Shiffrin & Steyvers, 1997).

For each item on the list, a separate episodic image is stored that consists of one or more associations. At least one of the associations in each image is between a studied item and the study context. More complex images are formed by appending additional item vectors to the image representing the items occupying the buffer (Malmberg & Xu, 2007; Shiffrin & Steyvers, 1997; Xu & Malmberg, 2007). We assume that many of these items are from the study list, and therefore the buffer operations determine which study items form new long-term associations. Extralist items retrieved from long-term memory may also be associated to study items in long-term images, particularly if the orienting task emphasizes relating the study items to prior knowledge (Lehman & Malmberg, 2009, 2011; Shiffrin & Atkinson, 1969). To keep things simple, however, we assume that only items from the current study list make up the associations encoded in long-term images.

To create associations between two representations, A and B, vectors representing A and B are concatenated in the long-term image representing their co-occurrence. For instance, the occurrence of an item in a particular context is represented by appending an episodic item vector to an episodic context vector (Shiffrin & Steyvers, 1997). The item vectors consist of the item features obtained from the corresponding lexical/semantic images of the studied items. The context vectors consist of the context features associated with the study list. Associations among items can be similarly formed by appending their item vectors (Malmberg, 2008; Malmberg & Xu, 2007; Xu & Malmberg, 2007). The configuration of the episodic image and the relative strengths of the vectors constituting the content of an image are determined by the operations of the rehearsal buffer specified below. Unless noted otherwise below, we assume in our simulations that the capacity of the rehearsal buffer is three items, because this facilitates computations and is generally consistent with prior assumptions (Atkinson & Shiffrin, 1968; Lehman & Malmberg, 2009; Raaijmakers & Shiffrin, 1980).

The Dynamics of Context

Context consists of a vector of w features. The feature values vary in their base rate according to the geometric distribution (Malmberg & Shiffrin, 2005; Shiffrin & Steyvers, 1997). An open question is whether the context vector should be based on the same geometric distribution of feature values as the item vectors. One could imagine finding oneself studying familiar words in a novel situation, consisting of relatively unfamiliar context features, for instance. However, there is no obvious reason why we should be presently concerned with such matters; therefore, we assume that item and context features vary according to the same geometric distribution, but that context features change at a slow pace while the goals of an individual are consistent. For the present, we therefore assume that context features change from trial to trial with a probability of β_w within a list and with a probability of β_b between the study and test lists, where $\beta_w < \beta_b$.

Encoding

Upon the presentation of the first list item, features are copied to an episodic trace from a lexical/semantic trace and the context vector with the probability,

$$c\lfloor 1-\left(1-u_x^*\right)^t\rfloor,$$

where u_x^* is the probability of storing a feature given t attempts to do so and c is probability of copying that feature correctly (Lehman & Malmberg, 2009, 2011; Malmberg & Shiffrin, 2005; Shiffrin & Steyvers, 1997).[4] If a feature is stored but is copied incorrectly, a feature is drawn randomly from the geometric distribution identified by the g parameter. If a feature is not encoded, it takes the value 0. The probabilities of copying item and context features differ, as the item is assumed to be the focus of attention and the context is not; u_i^* is the probability of storing an item feature, and u_c^* is the probability of storing a context feature. Thus, the result of encoding is a noisy, incomplete representation of an event.

Interitem associations are concatenated vectors of item features (Shiffrin & Steyvers, 1997; Xu & Malmberg, 2007). For each item, n, subsequent to the first item, item n is encoded as discussed above. The remaining buffered items are stored in the trace representing trial n as associative vectors composed of item features, concatenated to the item vector. For example, although some encoding resources are dedicated to encoding item n's features, contextual features are also being stored along with some features from item $n-1$ (and perhaps also features from items $n-2$, $n-3$, etc., depending on the buffer size). Because of the limited nature of encoding resources, however, encoding is split between the storage of different information. When multiple items are associated to the context, the context-to-item associations are weaker than when only one item is associated to the context. This is accomplished by reducing the u_x^* parameter for context features such that $u_c^* < u_{c1}^*$, where the latter term is the probability of encoding a context feature for the first list item. Some of the buffer capacity is also spent encoding interitem associations. This is represented by appending to the item-n trace, a trace relatively weakly representing the item $n-1$, and it is implemented by reducing the u_x^* value for item $n-1$ (referred to as u_a^*). Upon the presentation of the third item, this process repeats. Upon presentation of the fourth item, one of the first three items is randomly selected and dropped from the buffer, and the encoding cycle begins anew. As a result of this process, the length of the vector

[4] This is an oversimplification of encoding that is sufficient for present purposes but may not be appropriate for modeling spaced items repetitions (cf. Malmberg, Criss, Gangwani, & Shiffrin, 2012). When items are repeated, REM assumes that information usually is accumulated in a prior trace; otherwise, a separate trace is formed (Criss, 2006; Malmberg, Holden, & Shiffrin, 2004; Shiffrin & Steyvers, 1997). The propensity for trace accumulation is directly related to differentiation, and additional assumptions are required to specify this propensity in more realistic models.

depends on the number of items in the buffer; in the current model, the length of the vector was set to 8, so the trace would contain eight context features, eight item features, and eight co-rehearsed item features for each other item in the buffer.

Retrieval

Retrieval in free recall occurs via a series of sampling and recovery attempts. Sampling it is governed by a Luce choice rule. The rule assumes that the probability of sampling a given trace, j, is a positive function of the match of trace j to the retrieval cue and a negative function of the match of other $N - 1$ traces to retrieval cue,

$$P(j|Q) = \frac{\lambda_j}{\sum_{k=1}^{N} \lambda_k}, \tag{1}$$

where λ_j is a likelihood ratio computed for each trace,

$$\lambda_j = (1 - c)^{n_{jq}} \prod_{i=1}^{\infty} \left[\frac{c + (1 - c)g(1 - g)^{i-1}}{g(1 - g)^{i-1}} \right]^{n_{ijm}} \tag{2}$$

and where n_{jq} is the number of mismatching features in the jth concatenated trace and n_{ijm} is the number of features in the jth concatenated trace that match the features in the retrieval cue. Sampling occurs with replacement (Raaijmakers & Shiffrin, 1981), although the model is not allowed to output the same item twice (Lehman & Malmberg, 2009).

For free recall, subjects use a context cue to begin sampling. The nature of this cue depends on the task. Because context is most strongly associated to the first few items on the list, we assume that the cue will consist of a combination of the current context and some proportion (represented by the γ_I parameter) of reinstated context features from the beginning of the list. In immediate free recall from a single list, the combined context cue consists of current context features and of γ_I reinstated features from the beginning of the list. Sampling may occur from either the set of items currently being rehearsed or the set of long-term traces stored during study. The probability of sampling from the buffer increases as it becomes more difficult to reinstate context features from the beginning of the list, so the probability of initiating recall by sampling from the buffer is equal to $1 - \gamma_I$. After sampling from the buffer occurs, retrieval continues by sampling from the entire list in the same fashion. For delayed free recall, we assume that no list items are being rehearsed at the time free recall is initiated after a sufficiently long retention interval filled with a distracting activity. The effect of delay is represented in the model by the storage of additional traces after the study list, which are generated in the same manner as list items. This is an overly simplified model of what occurs during the retention interval, as the distractor activities are likely very different from what occurs during study (cf. Annis et al., 2012). However, for present purposes, the additional considerations would only complicate matters. For delayed free recall, the probe is used to retrieve long-term episodic traces only. Thus, whereas in immediate free recall, sampling may occur from the set of items that is being rehearsed, in delayed free recall, sampling occurs from the entire list. As in immediate free recall, the cue used to probe memory consists of a mixture of test context features and γ_I reinstated beginning-of-list context features; however, γ_I is reduced after the delay.

Once a trace is sampled, recovery of its contents is attempted. Because the contents are only a noisy, incomplete representation of a study event, the contents of some traces are more likely to be recovered than others. The recovery probability is a positive function of the number of features in the sampled trace that match the retrieval cue, x,

$$\frac{1}{1 + e^{-x+b}},$$

where b is a scaling parameter. After an item is output, the next cue used to probe memory consists of the same combined context cue with a portion, γ, of the context features replaced by those from the last item recalled. Additionally, some proportion, γ_m, of the item information from the last item recalled is also used in the cue. If no item is output, the original combined context cue is used for the next probe of memory. The sample-and-recovery process repeats κ times

Maintenance Rehearsal

With these assumptions in place, we can illustrate how this model accounts for a number of findings beginning with the results of a widely cited experiment conducted by Craik and Watkins (1973; Experiment 1) on the effect of maintenance rehearsal on free recall. Subjects were instructed to study lists of words, of which there were multiple critical words identified by their first letter on each; only the final critical word was to be reported. Craik and Watkins assumed that subjects would rehearse the most recent critical word until the next critical word was encountered or the end of the list was reached. This is consistent with Atkinson and Shiffrin's assumption that the retrieval of words is facilitated by rehearsal if one may be reasonably confident that it may be imminently needed, and this is what Craik and Watkins referred to as maintenance rehearsal. The amount of maintenance rehearsal of the critical words was varied by varying the number of study trials, or lag, between critical words. For example, if critical word n was presented, followed by two noncritical words, followed by another critical word, critical word n would be represented as having a lag of two, and it was assumed that more maintenance rehearsal would occur for items with a long lag than for items with a shorter lag. In addition, the presentation rate was varied between lists from .5 s to 2.0 s per word, and thus the amount of maintenance rehearsal was also directly related to the presentation rate. At the end of the experiment, memory was tested via free recall for all the items studied on the several lists during the experiment. This provided the data shown in Figure 1.

There were three key findings from Craik and Watkins's (1973) experiment. Critical words that were reported at the end of the lists were better recalled than critical words that were not, increases in study time improved recall for both types of critical words, but lag had only a small effect on final recall. Of these findings, the last one is commonly cited as being disconfirmation of the buffer model and is attributed to Craik and Watkins's (1973) conclusion that "maintenance rehearsal had no effect on later recall" (p. 605). Although Craik and Watkins's conclusion has become part of the conventional wisdom, it is not supported by the data. Note that subjects were assumed to be engaged in maintenance rehearsal of a critical item throughout the course of the study lists. If Craik and

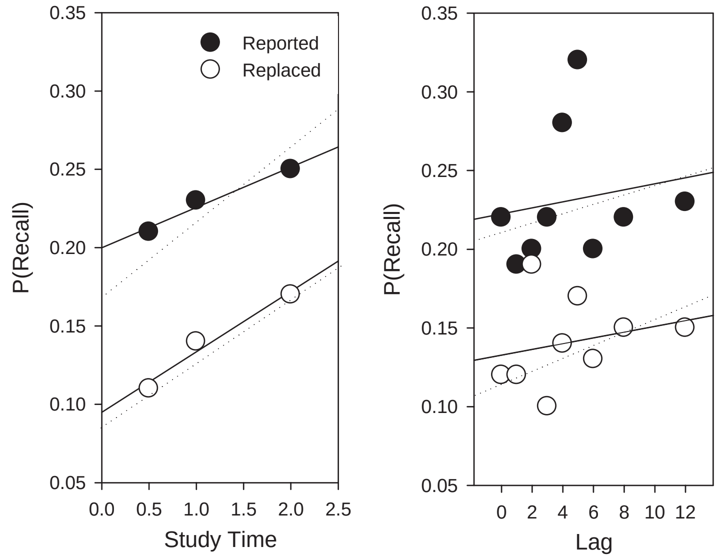


Figure 1. Data from Craik and Watkins (1973) and model predictions from the Lehman–Malmberg model. Regression lines based on Craik and Watkins's data are shown as solid lines; model predictions are represented by dotted lines. Reported items are those critical words reported at the end of the list; replaced items are those that were replaced by another critical word and never reported.

Watkins's conclusion that maintenance rehearsal has no effect on recall was correct, one would expect that the ability to recall any item was negligible. However, recall is well above chance and increases with increases in study time. It might be that the conclusions were meant to be limited to the effect of lag on recall. However, the regressions shown in Figure 1 indicate that recall of the critical words increased slightly with increases in lag, just as predicted by Atkinson and Shiffrin (1968):

> When the subject is concentrating on rehearsal, the information transferred would be in a relatively weak state and easily subject to interference. On the other hand, the subject may divert his effort from rehearsal to various encoding operations which will increase the strength of the stored information. (p. 115)

Thus, it appears that the distinction between maintenance and elaborative rehearsal used to criticize or even reject buffer models was, in fact, originally described by Atkinson and Shiffrin and incorporated into their original framework. Moreover, manipulations of rehearsal time such as those used by Craik and Watkins (1973) have been shown to enhance recognition memory (Glenberg, Smith, & Green, 1977; see also Darley & Glass, 1975; Nelson, 1977).

In summary, prior findings suggest that increasing the amount of maintenance rehearsal is in fact sufficient for additional episodic encoding to take place, and our REM buffer model provides an excellent account. When modeling encoding within the framework of the buffer model, it is important to analyze the task and goals of the subject. Craik and Watkins's task required subjects to identify a single critical word and report it at the end of a list. Because only one word has to be reported, the buffer model assumes that the buffer size is one; all buffer resources are devoted to rehearsing the critical word until the next critical word on the list is encountered or the end of the list is reached. Noncritical words are not rehearsed, and for sake of simplicity we assume that nothing is stored—although this certainly is an overly strong assumption, as the images stored on noncritical trials are presumably very weak—but for now, we will keep the model, which has four free parameters, as simple as possible. All the parameter values of the model are listed in Table 1. The majority are "default" REM values that have been used to model a large number of phenomena (e.g., Criss & Shiffrin, 2005; Malmberg, Holden, & Shiffrin, 2004; Malmberg & Shiffrin, 2005; Malmberg & Xu, 2007; Malmberg, Zeelenberg, & Shiffrin, 2004; Shiffrin & Steyvers, 1997). The remaining parameter values are those associated with the buffer and were used by Lehman and Malmberg (2009). Fixing all these parameters to those that account for other findings leaves u^*, a, and l as free encoding parameters for the simulation of Craik and Watkins's experiment.

When the first critical word is encountered, it enters the buffer and an item-to-context association is created. Both item and context features are stored with a $u^* = .1$. We assume that the number of attempts at storing these features, t, varies linearly with lag, j, scaled by a free parameter, a: $t = aj$. Whereas most item features are randomly generated for each item on the list, we assume that when subjects are rehearsing items beginning with a critical letter, some proportion, l, of the item features represents that critical letter and is shared across all rehearsed items from the list. When the next critical item is encountered, the prior item is compartmentalized from the buffer, and resources are devoted to rehearsing the new critical item. j represents the lag between the presentations of two critical words; hence, encoding strength increases with lag.

When the end of the list is reached, the item occupying the buffer is retrieved. It is inevitably the final critical word on the list. REM assumes that repetitions produce additional storage of features, usually in a trace stored on a prior presentation (Malmberg, Holden, & Shiffrin, 2004; Shiffrin & Steyvers, 1997). The accu-

Table 1
REM Parameter Values to Generate Lehman–Malmberg Model Predictions for Immediate Free Recall

Parameter	Value	Description
g	.4	Environmental base rate (standard value)
w	8	Number of item and context features
c	.8	Probability of correctly storing a feature
u^*_i	.5	Probability of storing an item feature
u^*_c	.3	Probability of storing a context feature
u^*_a	.3	Probability of copying a co-rehearsed item's feature
u^*_{c1}	.4	Probability of storing a context feature for first item on a list
t	3	Number of storage attempts
κ	10	Stopping rule parameter
β_b	.5	Probability of change for context features between lists
β_w	.2	Probability of change for context features within list
γ_1	.4	Probability of reinstating context features from beginning of list
γ_m	.4	Probability of reinstating item features from retrieved item
γ	.4	Probability of reinstating context features from retrieved item
b	5	Scaling parameter for recovery
l	.2	Proportion of item features shared among all list items

Note. For the Craik and Watkins (1973) simulations, a single u^* value of .1 was used for both item and context information, and l (.2) of the w item features representing the shared first letter remained the same for all critical items on the list. Additionally, the value of $t = a^*lag$, where a for reported words = 10 and a for unreported words = 2. For recognition, $\gamma_1 = .7$. For delayed testing, $\gamma_1 = .2$, and 10 additional items are stored after the list. For the continuous distractor task, $\gamma_1 = .3$. For pairs, u^*_a for items within a pair = .5; u^*_a for items in different pairs = 0; u^*_{c1} for the first item in a pair = .6; between pairs, $\beta_w = .3$. For the list-length model, $\gamma_1 = (List\ Length)^{-1}$. The stopping rule is as follows: Recall terminates when the number of sampling attempts has reached ($List\ Length^*\kappa$). REM = retrieving effectively from memory.

mulation of context features plays an important role in free recall (Malmberg & Shiffrin, 2005), and recent findings suggest that retrieval enhances memory more than does additional study (Karpicke & Roediger, 2008). Thus, a is assumed to be incremented as the result of retrieval. It is therefore greater for reported critical items than for unreported critical items (10 vs. 2), reflecting the additional encoding of the item-to-context association as the result of retrieval.

The data from Craik and Watkins's experiment were generated during a final phase, in which free recall of all of the items presented during the first phase of the experiment was attempted. Prior to this test we assumed that the features representing the end-of-list context changed at a rate of .2 (the β_w parameter) and that memory was probed with a retrieval cue consisting of a combination of the features representing the test context and those l features representing the first letter of the critical words. Both the lag and study time model predictions were subject to a weighted least squares objective goodness-of-fit test (Busemeyer & Diederich, 2010). Figure 1 shows that the buffer model provides an excellent fit to the data, $\chi^2(15) = 4.06 < 25.00$. In order to assess whether the buffer model produced a better fit to the data than the model without a buffer, we generated model predictions for both the model described and the model without the buffer. The latter was identical to the buffer model, except that t, the number of storage attempts (representing rehearsals in the buffer), did not vary with lag; thus, the nonbuffer model had one fewer free parameter. We compared goodness of fit using the likelihood ratio test (Glover & Dixon, 2004), according to which the buffer model produced a significantly greater fit than the nested model without the buffer, $\chi^2(1) = 7.77 > 3.84$.

In a second experiment conducted by Craik and Watkins (1973), subjects recalled the last four items immediately after the list or after 20 additional seconds to rehearse the items. Despite the last four items receiving a greater number of rehearsals in the delayed condition, recall of these items was equivalent across the delayed and immediate conditions. Craik and Watkins concluded that the effect of rehearsal on memory was minimal. In another experiment conducted by Wixted and McDowell (1989), the rehearsal time was allocated to the subject after the first third of the list, after the second third, or at the end of the list. They found that the extra rehearsal was beneficial only when it was provided in the beginning or middle of the list and not at the end. These results suggest to some that additional rehearsal has no impact on the storage of long-term memory traces, and they have been used to reject prior versions of the buffer model. However, buffer models in fact predict that additional rehearsal would have no effect on immediate recall, as recall would be initiated by retrieving the items that were being rehearsed. Increases in the amount of time devoted to rehearsing an item strengthen the encoding of a long-term episodic trace, but additional rehearsal does not affect the representation of the items being rehearsed. Hence, the benefits of the improved encoding of the long-term trace are achieved only when access to the long-term traces is required to perform the free recall task. For instance, a different outcome would occur if the 20 s of additional rehearsal were followed by a filled retention interval. In that case, retrieval from long-term episodic traces would be required, and it would likely be less accurate. We have much more to say about the differences between immediate and delayed free recall in subsequent sections of this article.

Serial Position Effects

Primacy

Few topics have received more attention from memory researchers than the form of the serial position curve for free recall. It has fascinated and frustrated legions because it has important implications for understanding the relationship among the brain, mind, and behavior; yet, after more than 125 years it still defies a coherent explanation. Ebbinghaus's (1885) initial investigations established scientifically what is intuitively obvious; recent events are more likely to be remembered than more distant events. It is no wonder that early research on serial position effects focused on primacy; memory for the initial few items of a long list is better than memory for more recent items (e.g., Robinson & Brown, 1926; Welch & Burnett, 1924). Tulving (2008) has even gone so far as to declare that primacy is a fundamental law of memory: "Of two sequential events the second one tends to be retained less well" (p. 32). Thus, the classical issue is how to reconcile the fact that memory diminishes over time, but items encountered early in sequence are remembered better than items encountered later.

Like other models (Anderson & Bower, 1972; Dennis & Humphreys, 2001; Howard & Kahana, 2002; Mensink & Raaijmakers, 1989), the model we describe above predicts long-term forgetting on the assumption that context changes over time; hence, the learning context becomes more difficult to reinstate at the time of retrieval. In addition, forgetting increases with increases in the number of intervening events (i.e., lag), as the items encountered more recently interfere with retrieval of items encountered further in the past (Criss & Shiffrin, 2005). There is a lesser consensus on an explanation for primacy. The original buffer model explanation is that items from early serial positions are rehearsed longer than items from later serial positions and are therefore more likely to be encoded in a long-term trace (Atkinson & Shiffrin, 1968). Reports that rehearsal is positively related to primacy under intentional learning conditions are consistent with this model (Rundus, 1971; Tan & Ward, 2000; Ward & Tan, 2004), as are reports that incidental learning conditions, which do not require or encourage rehearsal, drastically diminish primacy (Marshall & Werder, 1972). However, other findings indicate that primacy is affected by factors that occur after the list is studied and hence have no impact on rehearsal. For instance, primacy effects are drastically diminished when memory is tested via recognition (Murdock & Anderson, 1975). More directly related to free recall is the finding that large decreases in primacy occur when subjects are given directed forgetting instructions at the end of the list (Lehman & Malmberg, 2009; cf. Bjork, LaBerge, & Legrand, 1968). Because the instruction to remember or forget occurs only after the first list is studied, one would assume that the rehearsal strategies are the same in both conditions; thus, rehearsal is necessary but not sufficient to induce primacy. A key question is what is behind the unraveling of the effects of rehearsal vis-à-vis primacy.

Like the Atkinson and Shiffrin (1968) model, the present buffer model predicts primacy as the result of a capacity limitation. The more items occupying the buffer, the weaker the long-term memory encodings that represent those items will be. In the Atkinson and Shiffrin model, rehearsal of a given item decreases as the number of items in the buffer approaches its capacity, decreasing the probability it may be retrieved from a long-term trace. The present model also assumes that the capacity of the buffer is limited, but it makes more specific assumptions concerning the structure of long-term traces and how they are encoded. As described in detail above, long-term traces consist of associations between items and between items and context. The strength of the context-to-item association stored for a given item is negatively related to the number of buffered items (cf. Davelaar et al., 2005). Thus, the reason why future retrieval becomes less likely as more items are rehearsed is that there is a capacity limitation on the number of items that may be simultaneously associated to the study context.

Context cues play an important role in retrieval during free recall (cf. Malmberg & Shiffrin, 2005; Raaijmakers & Shiffrin, 1980). Context is most strongly associated to items from the earliest serial positions, as relatively few items are being rehearsed. As a result, context cues will be more likely to retrieve traces of items from early serial positions, especially the first item on the study list. This is due to the sampling advantage for those items, as they are most strongly associated to the reinstated study context. Thus, primacy is the result of the interaction between the encoding of long-term traces, particularly the item-to-context associations, and the cues used to retrieve those traces.

Note that the nature of the retrieval cues is critically important. Under typical laboratory conditions, in which the study list consists of a randomly sampled set of items, a context cue may be all that is available to probe memory on the initial attempt of delayed free recall. Because context is assumed to change relatively slowly over the course of the study list, items from early serial positions will not only be more likely to be retrieved but will also tend to be output first because they are most strongly associated to the context. We return to the topic of output order shortly. For now, what is important is that access to the information contained in a trace is achieved as a consequence of retrieval, and this item information is used as a retrieval cue on subsequent retrievals (cf. Howard & Kahana, 1999, 2002; Raaijmakers & Shiffrin, 1980).

The probe consisting of the context and the item features retrieved from a trace will tend to produce a sampling advantage for traces of items that were rehearsed with the item that was just retrieved. Hence, retrieval of a given item is heavily influenced by the retrieval of items from earlier serial positions. When retrieval fails given a set of cues, context is again used to probe memory, and the sequence begins anew. On these later retrieval attempts, it is likely that items from the earliest serial positions have already been output, and thus (due to sampling with replacement) the sampling advantage for items from early serial positions produces output interference (Criss, Malmberg, & Shiffrin, 2011; Roediger, 1974; Tulving & Arbuckle, 1966).

The result of this sequence of sampling with different cues is primacy. Figure 2 shows how primacy in delayed free recall is affected by the number of buffered items. For one function, the buffer size is assumed to be three items, which is the default value we assume for most of the simulations of intentional learning reported in this article, and the next function shows the result of decreasing the buffer size to one item. Reducing the size of the

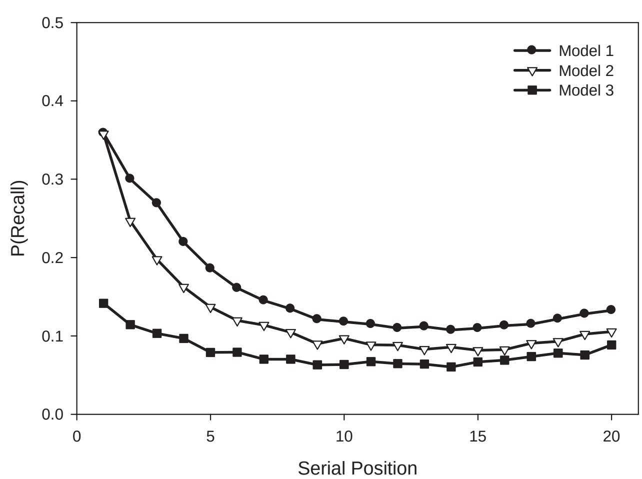


Figure 2. Model predictions for primacy effects in free recall. Shown are three models of delayed free recall. In Model 1, the buffer size is 3. In Model 2, the buffer size is 1. In Model 3, the buffer size is 1, u^*_{cl} = u^*_c = .2, and γ = .06. All other parameter values are the same as those used in the delayed free recall model, shown in Table 1.

buffer to one item simulates a model with a very limited capacity buffer. Although primacy is reduced, encoding still occurs, but it results in less richly encoded traces. This is how the model accounts for the formation of long-term memories in situations where buffer operations are disrupted, for example, for individuals with working memory impairments (e.g., patients P.V., Baddeley & Wilson, 1988; and K.F., Shallice & Warringon, 1970).[5] The other function is the result of decreasing the number of buffered items from three to one and reducing the probability of encoding context from $u^* = .3$ to $u^* = .2$. This is intended to model the conditions of incidental learning, in which there is no need for the subject to maintain an item across study trials once the incidental orienting task is completed (e.g., Nelson, 1977). Primacy is reduced because similar numbers of items occupy the buffer across all serial positions, and therefore the strength of the item-to-context association is similar across serial positions. In addition, context is more weakly encoded, as items are unlikely to be actively maintained for enough time for context to be completely bound to the item (cf. Malmberg & Shiffrin, 2005). Because context is more weakly encoded, it logically follows that beginning-of-the-list context is also more difficult to reinstate and the sampling advantage for the items from early serial positions is severely reduced. This is represented in the model by a reduced γ_I value for incidentally encoded lists.

The reduction of primacy, however, is not only observed under incidental encoding conditions. Primacy is virtually eliminated when memory is tested via a recognition procedure. However, to the best of our knowledge, only rarely have models of free recall been constrained by such findings (cf. Murdock & Anderson, 1975). Figure 3 shows the serial position curve for two-alternative forced-choice recognition we obtained in a recent experiment replicating Murdock and Anderson but using longer lists, like those typically used to test recognition memory (the method and results are described in the Appendix; Experiment 1). Again, the primacy effect is very small. However, in contrast to the distortion

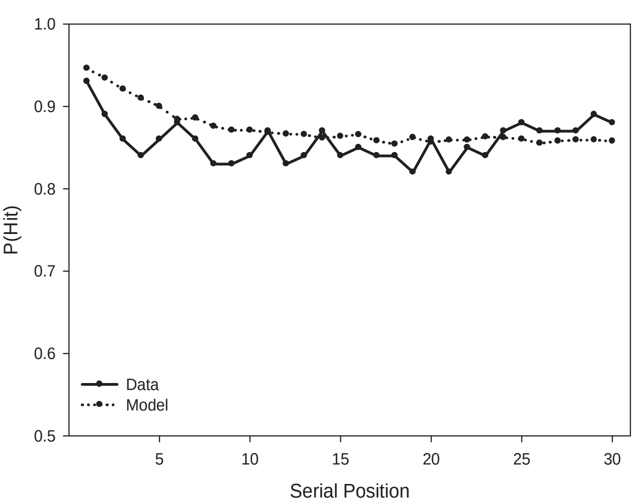


Figure 3. Data and model predictions for two-alternative forced choice recognition (Experiment 1). The probability of storing an item feature $(u^*_i) = .5$, and the probability of storing a context feature $(u^*_c) = .3$; all other parameter values are the same as those used in the immediate free recall model, shown in Table 1.

of the primacy effect obtained from incidental encoding, this disruption in the primacy effect is not due to changes in encoding strategies; the buffer size is assumed to be three. Rather, recognition is based on the familiarity of the test item, which is generated by a global-matching process that compares the retrieval cue to the traces stored during study and computes the average likelihood ratio for the cue, or the *odds* (Shiffrin & Steyvers, 1997). The retrieval cue used in the present simulations for recognition consists of item features from the presented word along with the current context features. For two-alternative forced-choice recognition, the odds are computed for the target and foil, and the item selected is the one that produced the greatest odds (Criss et al., 2011; Malmberg & Murnane, 2002). Primacy is very small for recognition because the item features in the retrieval cue are those that are diagnostic for the recognition decision, as the context features used to probe memory are shared by the target and the foil. Because the strength of the item features does not vary with serial position, very little primacy is expected for recognition.

Recency

The flip side of primacy is recency. Most generally, buffer models assume that the immediate free recall advantage observed for the last several items on the study list is the result of outputting rehearsed items when the instruction to recall is given (Atkinson & Shiffrin, 1968). During a filled retention interval, rehearsal does not occur, making end-of-list items increasingly less likely to be in a readily available state. This accounts for the decrease in recency with increases in the duration of a filled retention interval (Glanzer & Cunitz, 1966). Other versions assume that the level of activation of a trace is positively related to recency, which provides a strong retrieval advantage for the most recently studied items (Cowan, 1997; Davelaar et al., 2005; Shiffrin, 1975). Both assumptions are consistent with the finding that the recency effect is greatly reduced when testing occurs after a sufficiently long retention interval (Postman & Phillips, 1964). The major difference between these two models is that the original buffer model strongly emphasizes the role of the buffer as an active control process, specifying how memory is affected by the goals of the subject vis-à-vis task demands, whereas the activation model focuses on describing changes in the state of memory that result from factors largely out of control of the subject.

The present model is a blend of the two approaches. The buffer is actively engaged in processing end-of-list items and encoding long-term traces, whose activations at the time of retrieval are influenced by the cues used to probe memory. Hence, recency is dependent on the cues used to probe memory; under some conditions it is advantageous to initially retrieve the most recently studied items, whereas at other times it is advantageous to initially retrieve items from the beginning of the list. We hypothesized that there are individual differences in the retrieval strategies used by subjects during immediate free recall; some subjects may have a preference for initially retrieving end-of-list items, whereas others may prefer to initially retrieve beginning-of-list items. For in-

[5] The current model is used to account for the encoding and retrieval of verbal information; as with the Atkinson and Shiffrin (1968) model, it is assumed that visual information is processed via a separate buffer process, allowing the model to account for dissociations between memory for audio-verbal-linguistic and visual information.

stance, subjects with lower capacities for rehearsing items may be more inclined to initiate retrieval from the beginning of the list; relatively few items from the end of the list may be readily available, and their retrieval would cause output interference when these subjects subsequently attempt to recall items from the beginning of the list. Recently, for instance, Unsworth, Brewer, and Spillers (2011) classified subjects as primacy oriented, recency oriented, or both (using a combination of the two strategies) on a delayed free recall task. They showed that the group that used a combination of primacy and recency strategies performed better overall and had higher average working memory capacity.

To assess individual difference in the immediate free recall, we analyzed the data by first binning the data from the subjects in the top 25% of overall recall and those in the lowest 25% (the method and results are described in the Appendix; Experiment 2). Eight lists of words were studied. The top panel of Figure 4 plots overall recall as function of list. The top performers not only recalled more items but also improved over the course of testing, whereas the bottom performers did not. The middle panel shows that the serial position functions for top and bottom performers are similar in shape, with the top performers exhibiting more recency than the bottom performers, and their recency extends to earlier serial positions. Indeed, the lower panel shows that top performers were more likely than bottom performers to initiate retrieval from the end of the list, whereas the bottom performers were more likely than the top performers to initiate retrieval at the beginning of the list.[6]

The variability in recency observed among samples from a healthy population of young adult university students is difficult to reconcile with models that assume that recency reflects a structural organization of long-term memory traces that dictates enhanced access for end-of-list items based only on the similarity between test context and end-of-list context (e.g., Howard & Kahana, 2002). It is also striking to note that those who were best at immediate free recall improved their performance over the course of the experiment, whereas the worst performers did not. The individual differences combined with the practice effect suggest within the current framework that the top performers learned to adopt a "sure-thing" strategy of retrieval from the end of the list in order to enhance overall free recall performance, implicating the role of control processes in producing the patterns of recency we observed.

The original buffer models specified the dynamics of the buffer and consequently the probabilities that certain traces would be retrieved, but they made no assumptions concerning how items are retrieved. To account for recency in immediate free recall, they simply assumed that end-of-list items were most likely to be retrieved because these items were currently being rehearsed (Atkinson & Shiffrin, 1968). This is, of course, somewhat of an unsatisfactory account of recency for at least two reasons. First, it is largely descriptive, as the mechanisms by which end-of-list items are more likely to be available are not detailed. Second, subsequent findings demonstrated that recency is not ubiquitous, and the classical buffer account was not extended to account for them.

According to the present model, the buffer is actively engaged in processing items during encoding, including end-of-list items, whose activations at the time of retrieval are influenced by the cues used to probe memory (cf. Davelaar et al., 2005). Fluctuations in recency that are observed across different testing procedures arise from the use of different cues to probe memory. For immediate free recall, the test context is used to initially probe memory. Those items most recently rehearsed will be more likely to be sampled and output early during recall, as the context stored in their traces will tend to be most similar to the context at test. It is important to note that in order to account for recency we need not state that traces of items at the end of the list are in a privileged state, somehow separate from those stored earlier. It is sufficient to state that they are more strongly associated to the current testing context than those studied earlier. This association is a direct consequence of similarity of cue used to probe memory and the contents of the memory traces. Likewise, we need not state that traces of items at the end of the list are not in a privileged state, for they may be. The only necessary assertion is that variability in recency arises from control processes influencing the construction of retrieval cues at test. For instance, if one is instructed to recall from the beginning or from the end of the list it is not unreasonable that the obedient subject would do so (Ebbinghaus, 1885; Li & Lewandowsky, 1995). In fact, we have shown above that there is variability in these tendencies between subjects that must be explained by any viable model. The tendency to initiate recall from the beginning of the list is likewise pronounced in delayed free recall; hence, it follows that the context features consist of a mixture of features representing the test context and the context stored during study. Because items from early serial positions are most strongly associated to the study context, they tend to be sampled and output early during recall.

We have shown that there are individual differences relating overall recall performance to primacy, recency, and the order in which items are retrieved from memory. Although some models do not offer an obvious account for individual differences in the temporal relationships in free recall data (Howard & Kahana, 2002), we propose that these differences are the result of differences in the manner in which the buffer is utilized during retrieval (i.e., some subjects choose to sample from the buffer and some do not).

Continuous Distractor Task

Whereas recency is often eliminated in delayed free recall, recency is observed when a filled delay is interpolated between each study item (Bjork & Whitten, 1974). This procedure is known as the *continuous-distractor procedure*, and the effect is referred to as *long-term recency*. Proponents of single-store models have stated that long-term recency effects in a continuous distractor paradigm are troubling for dual-store models because they assume that recency effects are due to the presence of items in a short-term buffer at the time of test (Crowder, 1989; Howard & Kahana, 1999). For example, Crowder (1989) stated, "The traditional association of the recency effect in free recall with some transient memory has now been discredited by the work of Bjork and Whitten (1974)" (p. 274).

However, we noted above that a buffer model need not assume that traces of end-of-list items are in any way more or less transient than any other traces in order to account for recency, and it is

[6] These data are consistent with findings that serial recall performance is better for backward serial recall than for forward serial recall for long lists (i.e., greater than eight items; Anderson, Bothell, Lebiere, & Matessa, 1998).

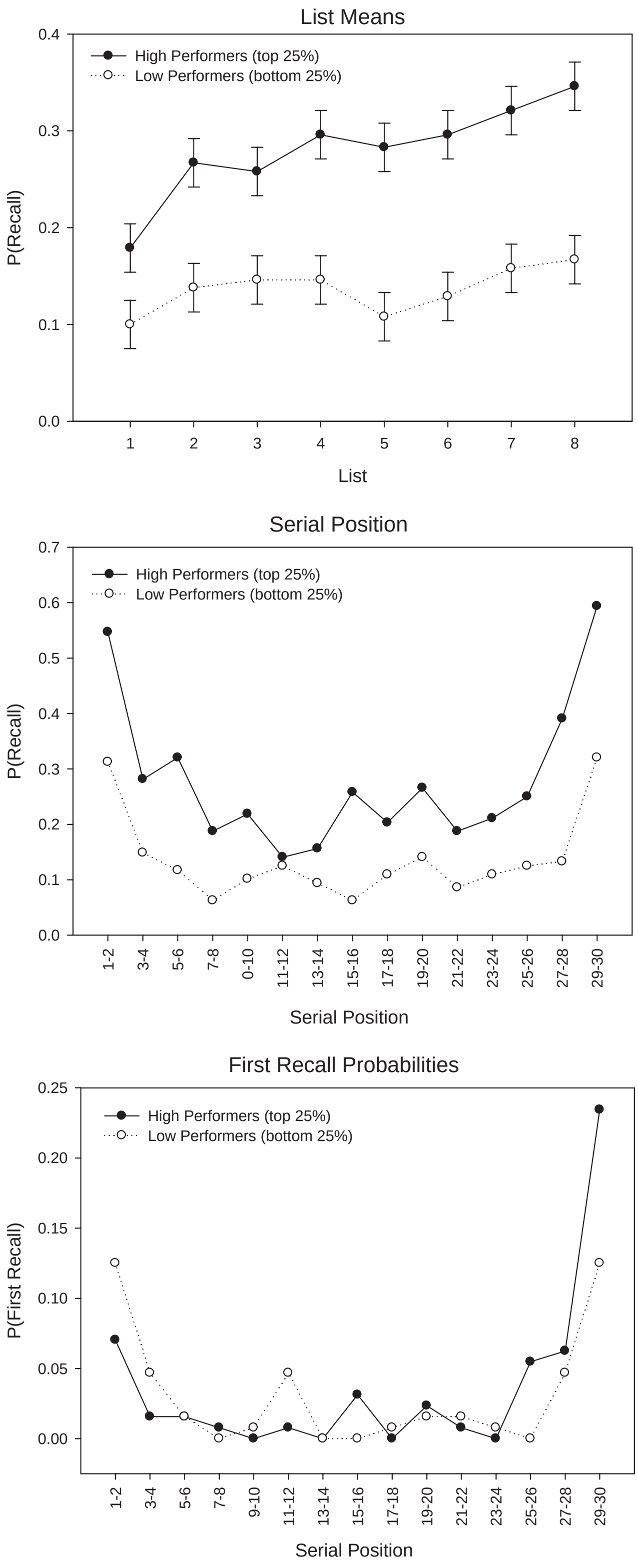


Figure 4. Mean recall data, serial position effects, and first recall probabilities for high performers and low performers (Experiment 2). Error bars represent standard error.

obvious that all recency effects need not result from the same mechanism, as was originally noted by Bjork and Whitten (1974). That is, the superficial form of the serial position data does not specify a unique model any more than an increase in overall recall uniquely specifies a model. Moreover, the superficial similarity between the recency observed in immediate free recall and long-term recency is more apparent than it is real, because long-term recency is attenuated (e.g., Howard & Kahana, 1999) and immediate and long-term recency are differentially affected by various experimental tasks (Cowan, Wood, & Borne, 1994; Davelaar, Haarmann, Goshen-Gottstein, & Usher, 2006; Glenberg, 1984; Whitten, 1978). Nevertheless, we assert that the similarity between delayed free recall and continuous distractor free recall procedures begs a comprehensive account that is also constrained by findings concerning the order in which items are recalled.

According to the present model, long-term recency is the result of the use of a context cue largely reflective of the context present at the end of the list; hence, the recency obtained from the continuous distractor procedure. This assumption is not trivial; it has implications for the order in which items are recalled. To wit, Lehman and Malmberg (2009) concluded that the primacy in the first recall probabilities (FRPs) for delayed free recall was consistent with a model in which reinstated study context was at least partially used to initially probe memory. On the assumption that the first item studied tends to be most strongly associated with reinstated context, it will tend to be output first. Although their findings are consistent with this prediction, the situation is different when a continuous distraction procedure is used. The retention intervals commonly used in the continuous distractor procedure are relatively short compared to the retention interval used in most delayed free recall experiments. Because retention intervals are very short, many continuous distractor experiments simply blur the distinction between immediate and delayed recall (cf. Glanzer & Cunitz, 1966; Howard & Kahana, 1999); hence, they have only a moderate impact on the recency. Under these conditions it is more productive to initially retrieve items that were rehearsed a few seconds ago, as the test context is so similar to that which was just stored. Thus, the prediction is that recency should be observed both in the serial position curve and in the FRPs when a continuous distraction procedure is used. We report a test of this prediction shortly.

To more concretely specify an account of the continuous distractor task, we generated a model in which a short distractor task is presented after each item on the list, including the last item. The continuous distractor task is represented in the model by reducing the buffer capacity from three items to two items under the assumption that it will be more difficult to rehearse items while completing the distractor task between items. Additionally, because there is a short distractor task for both immediate and delayed free recall after a continuous distractor task, the model assumes that memory is probed in a fashion similar to that in standard delay conditions, where the likelihood of reinstating features from the beginning of the list (γ_I) is reduced due to the increased context change that has occurred throughout the list as a result of the continuous distractor task (parameter values are listed in Table 1). Thus, the model predicts reduced overall performance in an immediate continuous distractor conditions compared to an immediate control condition. In addition, we modeled the effect of increasing the retention interval following a continuous distractor task. Although a delay of 10 s (a time commonly used in the continuous distractor task) may not be sufficient to eliminate rehearsal of end-of-list items (Glanzer & Cunitz, 1966; Howard and Kahana, 1999), a longer delay should be more effective. The prediction of the buffer model was that end-of-list context features would be less similar to the current context features after the long retention interval than after a short retention interval (Davelaar et al., 2005; Howard & Kahana, 2002), and the model thus predicts that long-term recency will be eliminated after a sufficiently long delay. Moreover, the context stored for items in the initial serial positions is stored more strongly, according to the present model (recall that the probability of storing contextual features, u_c^*, is greater for the beginning of the list). Therefore, the context cue used to probe memory after a long delay would consist of a mixture of context features present at test and those mentally reinstated from the beginning of the list. As a result, initial probes with such a context cue produce primacy in the FRPs as the result of the enhanced item-to-context association stored for the first item studied. Additionally, as the result of the storage of information from co-rehearsed items that occurs during encoding, along with the incorporation of context features associated with recalled items into the memory cue during retrieval, the model predicts an asymmetry in the lag-recency effect, in which transitions are more likely in the forward direction than in the backward direction (Kahana, 1996).[7]

To test the model's predictions, we conducted an experiment examining the continuous distractor task in immediate and delayed free recall (the method and results are presented in the Appendix; Experiment 3). This experiment examined both standard and continuous distractor free recall in both immediate and delayed free recall conditions. For delayed free recall, we chose a 5-min retention interval to decrease the likelihood that items would be covertly rehearsed at the time when free recall was initiated. To the best of our knowledge, such a long retention interval, meant to prevent rehearsal prior to delayed free recall in both the standard and continuous distraction conditions, has never been included in prior experiments.

We also took steps to reduce rehearsal during the interpolated periods of distraction. Subjective reports obtained from pilot studies in which a math task was used in the periods of continuous distraction indicated that subjects were easily rehearsing items; thus, the distractor task may not sufficiently reduce the size of the rehearsal buffer. Rather than using a math task as the interpolated activity, we required subjects to provide rhymes for irrelevant, extralist words on the assumption that the verbal nature of this task would make it more difficult for subjects to rehearse the list items during the distractor period (Gardiner, Thompson, & Maskarinec, 1974).[8]

Figure 5 shows the results and a fit of the model to the data from the continuous distractor experiment. The predictions were derived a priori; only the fitting of the model was post hoc. The model predicted all of the effects in a continuous distractor task, including primacy in all conditions, the lag-recency effect (along with the asymmetry in lag-recency), and the long-term recency effect for immediate free recall in the continuous distractor condition. The later finding rebuts the claims made that dual-store models cannot predict long-term recency. Importantly, the model's predictions are consistent with the attenuation of long-term recency in the continuous distractor condition (the difference between the probability of recalling the last item on the list is and the probability of recalling the middle item on the list is about 40% in the immediate control condition and about 10% in the immediate continuous distractor condition). Additionally, consistent with the data, the model predicts the elimination of recency after a long delay in both the control and continuous distractor conditions. This new finding challenges models that predict long-term recency on the assumption that the context cue at test matches the end-of-list traces between the beginning-of-list traces (cf. Howard & Kahana, 2002).

Note that the majority of parameter values used to fit these data were those described earlier and those used by Lehman and Malmberg (2009). Thus, the fit was contingent on assumptions made concerning the nature of the distractor task and the use of different retrieval cues in delayed versus immediate free recall. One key assumption is that the buffer size is reduced from three to two items during the continuous distractor task. The other key assumption is that different retrieval cues were initially used to probe memory. In the immediate tests of memory, the test context was used to probe memory; in the delayed testing conditions, a mixture of test context and reinstated study context features was used. The manner in which these context cues used to probe memory were generated is described above; the cues differed in that the probability of reinstating beginning of list context features, γ_I, is greater for immediate recall, and additional traces are generated after the delay (parameter values are listed in Table 1). Thus, on the basis of these assumptions about the controlled use of the buffer, the model is able to predict findings that have been argued to be troublesome for buffer models, in addition to unique findings of no long-term recency after a delay.

Chunking

A key role for the buffer is to specify the items to be rehearsed, which in turn determines what is encoded in long-term traces. To this point, we have assumed that when a list of single items is studied, items are rehearsed sequentially. A long, uninterrupted associative chain of traces is formed, with each trace "linked" to the prior trace via the similarity of the context stored within and items that are associated when co-rehearsed. Thus, earlier versions of the Lehman–Malmberg model did not include chunking operations (Miller, 1956). Nevertheless, we hypothesized that when pairs of items are studied, subjects may instead focus on rehearsing the pairs, dropping the pair from the prior trial via the compartmentalization process when presented with the next pair. The result is a set of chunks in memory consisting of item–item associations and item–context associations that are related via the overlap in similarity that is assumed between the context features stored in each chunk. Thus, the primary differences between the models of item and pair encoding are that the buffer size is reduced to two when pairs are studied and that once a new pair is presented the prior pair is no longer rehearsed. Taken together, the model

[7] The current model does not predict the decline in lag-recency (+1 transitions) across the first four output positions shown by Howard and Kahana (1999) and predicted by some models (Farrell, 2010), but our data do not show this either.

[8] We thank Doug Nelson for this suggestion.

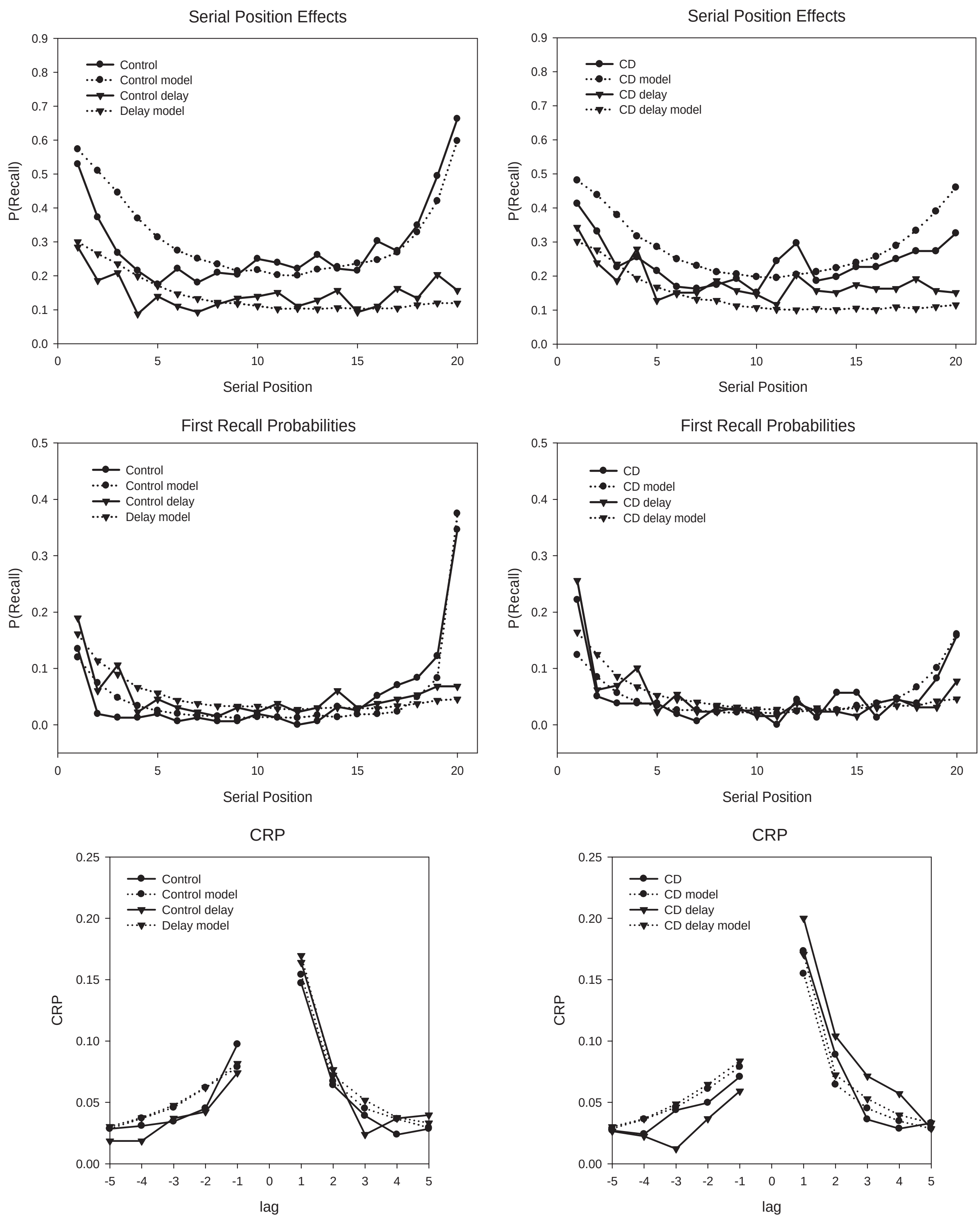


Figure 5. Continuous distractor and control data and model predictions (Experiment 3). Control refers to immediate free recall after standard list presentation, and CD refers to immediate free recall after continuous distractor list presentation. Control delay refers to delayed free recall after standard list presentation, and CD delay refers to delayed free recall after continuous distractor list presentation. CRP = conditional recall probability.

assumes that each chunk consists of two item–context associations and one item–item association.[9] We implemented chunking into the model in the following manner:

Assumption 1: Associations are made between items in the same chunk. If items are from different chunks, the associations between them are minimal. In the model, we made the stronger assumption that associations between chunks are nil ($u_a^* = 0$). During encoding, both members of a pair are stored in a single trace. The staggered presentation used in our experiment allowed us to assign an ordinal value to each member of pair (described in the Appendix). For the first item in a pair, item information and item–context associative information are stored according to the same encoding process described in prior sections. When the second item in a pair enters the buffer, item information corresponding to that item and associative information from the first item in the pair are stored. The context is more strongly associated with the first item in the pair than the second item in the pair, because the number of the rehearsed items has increased and the first item in the pair is associated with the second item. For simplicity, however, we assume in the current model that the second item in a pair is not associated with the context. Accordingly, u_a^* for pairs $> u_a^*$ for single items, and u_c^* for the first item in a pair $> u_c^*$ for the second item in a pair (which is currently set to zero).

Assumption 2: Chunks are recalled in an all-or-none manner (cf. Johnson, 1970). Recall from chunks begins with the first item; the other items are maintained in short-term memory while the first item is being retrieved. Because this is an immediate free recall task, during retrieval, the current test context is initially used as a cue. When an item that is the first member of a pair is successfully retrieved, the next item to be sampled is the second item from that pair. This is similar to an "unpacking" process—once the chunk is successfully accessed, the entire chunk can be retrieved (Johnson, 1970). The recovery process occurs as it does in the single-item model. After a pair is sampled, the current test context cue is again used to probe memory.

As before, qualitative predictions were generated a priori, and aside from those listed above, the model uses the same parameter values for pairs as for single items (see Table 1). The model predictions are shown as dotted lines in Figure 6. The model makes three notable predictions. First, as shown in the middle graph, FRPs differ for single and paired items. For single items, the item on the list that is most likely to be output first is the most recently studied item. However, for paired items, the item that is most likely to be output first is the penultimate item, or the first item from the last pair. This prediction is derived from Assumption 1 described above. Next, the model predicts that pairs are much more likely than single items to make a +1 lag transition (representing transitioning within a pair). This is the result of the chunks created by the compartmentalization process. Finally, the model predicts a zigzag pattern in the serial position curve for paired items, in which spikes occur throughout the list, representing an increased likelihood of recalling the first item in a pair (Davelaar et al., 2006). Anderson, Bothell, Lebiere, and Matessa (1998) showed similar spikes in reaction time data for the first item in a group when list items were visually grouped into units of three (although as theirs was a study of serial recall, output order effects were not of interest). These findings are consistent with the assumption that context cues must be reinstated in order to access the first item in a pair, after which the chunk is "unpacked" (Johnson, 1970), and the second item is sampled automatically.

To test these predictions, we conducted an experiment using single- and paired-item study lists (the method is described in the Appendix; Experiment 4). In the paired-item study condition, subjects were instructed form sentences in their mind for each pair. We assumed a priori that this would lead to a strong interitem association between members of a pair and that when the next pair was presented, the prior pair would be removed from rehearsal via compartmentalization. As a result, no interitem association would be created between members of different pairs, and this would have a profound effect on the order in which items were output.[10]

There was not a significant difference in the proportion of words recalled from single or paired word lists, $F(1, 38) = 2.49$, $MSE = .037$, $p = .123$. This is consistent with the assumption that the vast majority of the parameters in the model are shared between encoding conditions. The top panel of Figure 6 shows significant primacy and recency for both singles and pairs study conditions. There was a significant effect of serial position, $F(29, 1102) = 25.02$, $p < .001$, but there was no significant serial position by condition (single or paired list) interaction, $F(29, 1102) = 1.33$, $p = .117$. Thus, the form of the serial position function was unaffected by the switch from single items to pairs between conditions. However, the key predictions concern the order in which items are recalled. For the FRPs, there was a significant effect of serial position, $F(29, 1102) = 37.16$, $MSE = .007$, $p < .001$, and a significant serial position by condition interaction, $F(29, 1102) = 19.01$, $p < .001$. As shown in the middle panel of Figure 6, subjects in the singles condition were most likely to initiate recall with the last item on the list, whereas subjects in the pairs condition were most likely to initiate recall with the penultimate item on the list (the first item in the last pair). We also

[9] Although many features of classic chunking theories already exist in the model, our nomenclature may differ. For instance, Johnson (1970) described chunks as "items or information sets which are stored within the same memory code" (p. 172), where a memory code is a mental representation of information that was learned, analogous to a *trace* in our model. As with memory traces in the REM model, codes are representations of information that are distinct from the information itself (lexical/semantic traces in REM can include errors or missing information). Johnson also describes *recoding*, the process of learning a code for a chunk, and *decoding*, the process of translating the code into the information it represents, which roughly correspond to the encoding and retrieval processes in REM.

[10] The model of the chunking process reflects control processes implemented by the subjects in order to accomplish the task, which in this case was to remember pairs of items. Pilot work showed subjects tended to rehearse pairs of items as chunks, regardless of instructions; however, instructing subjects to create a sentence with pairs of words results in less noisy data. This highlights an important point regarding the use of the buffer during encoding: The buffer is used to accomplish a specific task; when pairs are presented, it is convenient for the subject to form chunks rather than chains, as when a stream of single items is presented. This assumption distinguishes buffer models from other models.

measured conditional recall probabilities (CRPs), which refers to the probabilities of successively outputting items of a given lag during recall, where lag refers to the distance between the serial position of item *n* and item $n - 1$ (Kahana, 1996). For CRPs, there was a significant effect of lag, $F(59, 2242) = 37.92$, $MSE = .001$, $p < .001$, and a significant lag by condition interaction, $F(59, 2242) = 8.67$, $p < .001$. The bottom panel of Figure 6 shows more +1 transitions in the paired list condition (within-pair transitions) than in the single list condition.[11]

Thus, as predicted by the model, recall patterns differ for single-item and paired-item study lists, as revealed by serial positions (SPs), FRPs, and CRPs. The model accurately fits the data for both single-item and paired-item study lists. As predicted by the model, for single items, the first item output is most likely to be the last item on the list, and for pairs, the first item output is most likely to be the penultimate item on the list (the second item from the last pair). The model predicts a zigzag effect in the SP curves; however, there appears to be too much noise in the data to detect these zigzag patterns. Like the model, the data show a greater likelihood of making a +1 transition for pairs than for single items. Thus, the model is capturing all of the features that distinguish recall patterns in single-item lists from those of paired-item lists, including data generated to test a priori predictions of the model and data from paradigms that have been said to be troublesome for buffer models, including tasks manipulating maintenance rehearsal and the continuous distractor task.

Are End-of-List Traces in a Privileged State?

Finally, we turn to an issue to which we have alluded, one that has vexed the field for decades. Are the traces of items that compose the recency portion of the immediate free recall serial position curve in some privileged state that makes them more accessible than those traces representing items from earlier serial positions? Although the nature—or even the existence—of such states has been subject of persistent speculation (Anderson, 2000; Cowan, 1995; Crowder, 1976; James, 1890; Nairne, 1990; a portion of the 1993 volume of *Memory & Cognition* was dedicated to this issue), the issue within the current framework boils down to whether recency items compete with nonrecency items during retrieval. If so, the number of items composing the study list should affect various aspects of retrieval, especially the FRPs. To this point, the issue of concern to so many has not been important vis-à-vis the ability of the model to predict the data. We note, however, that we may address the issue by comparing versions of the current model that posit recency traces as being only part of the set of long-term traces stored during study to those that state recency items need not compete with long-term traces for retrieval.

All versions of the current model predict a decrease in immediate free recall performance with an increase in list length. In fact such findings are common, and similar magnitude recency effects have been observed in lists that exceed the size of the buffer (Murdock, 1962; see also Ward, Tan, & Grenfell-Essam, 2010). A critical prediction, however, concerns the FRPs. In the previously described experiment, we observed differential first recall probabilities for single- versus paired-item lists. For single-item lists, the last item on the list is the most likely item to be recalled first, whereas for paired-item lists, the second to last item on the list (or first item of the last pair) is most likely to be recalled first. If these patterns are due to retrieval of privileged traces from a rehearsal buffer, additional items studied should not influence the items that are currently present in the buffer at time of test, and the same first recall patterns should be seen regardless of list length. If, however, the memory system does not include a rehearsal buffer, first recall probabilities should be influenced by list length because the most recently studied items would suffer different amounts of interference in lists of different lengths. That is, items from the end of list should become increasingly less likely to be initially output as the length of the study list increases, as long as the length of the list exceeds the capacity of the buffer.

For quantitative prediction it is necessary to specify the parameters of the model that will be affected by list length. First, it seems possible that a greater number of attempts at sampling and recovery of items will be made for longer list lengths than shorter list lengths (but see Harbison, Dougherty, Davelaar, & Fayyad, 2009). Thus, the number of sampling and recovery attempts = ($\mathit{List\ Length} * \chi$). Next, we assume that as list length increases (and context changes), it will be more difficult to reinstate context features from the beginning of the list, and as it becomes more difficult to reinstate these features, the probability of sampling from the buffer becomes more likely. Thus, the probability of reinstating beginning-of-list context features is also a function of list length, $\gamma_1 = (\mathit{List\ Length})^{-1}$, and the probability of initially sampling from the buffer ($1 - \gamma_1$) will increase as it becomes harder to reinstate the beginning-of-list context features.

The model predictions for FRPs and SPs are shown in the right panels of Figure 7 for single items and of Figure 8 for paired items, and model predictions for CRPs are shown in Figure 9. The model predicts differential patterns of FRPs and CRPs for single versus paired items. Importantly, it predicts that patterns of FRPs should not change with list length (for all list lengths greater than four). Additionally, it makes clear predictions about zigzag effects in the SP curves. For paired items, we see a zigzag pattern in both SP and FRP curves. Although the penultimate item is most likely to be recalled first, items from earlier in the list are sometimes recalled first; however, the first item in a pair is always more likely to be recalled first than the second item in a pair. These effects are consistent for all list lengths (greater than four).

A second test of the model involves reaction times in free recall. As suggested by Davelaar et al. (2005), if retrieval in immediate free recall begins by sampling items from the buffer, response time to output the first item during recall should not be influenced by other items on the list that are not present in the buffer. Thus, a model that includes retrieval from the buffer predicts that time delay to output the first item during recall should not be affected by list length, whereas a model that does not include retrieval from the buffer would predict that a memory search should include all

[11] The apparent relative flatness in the CRP curves may be due to multiple factors. First, the scaling of Figure 6 makes the curve appear much flatter than that in Figure 5. Additionally, the probability of making nearby transitions is low with large list lengths (30 items), and when list length is shorter, this probability is higher (e.g., with the 20-item list length in Figure 5). These data are consistent with the shape of the CRP curves in Kahana's (1996) analyses, whereas others have shown much steeper curves (Ward et al., 2010). It is not clear which differences in procedures produce such large differences between these sets of data.

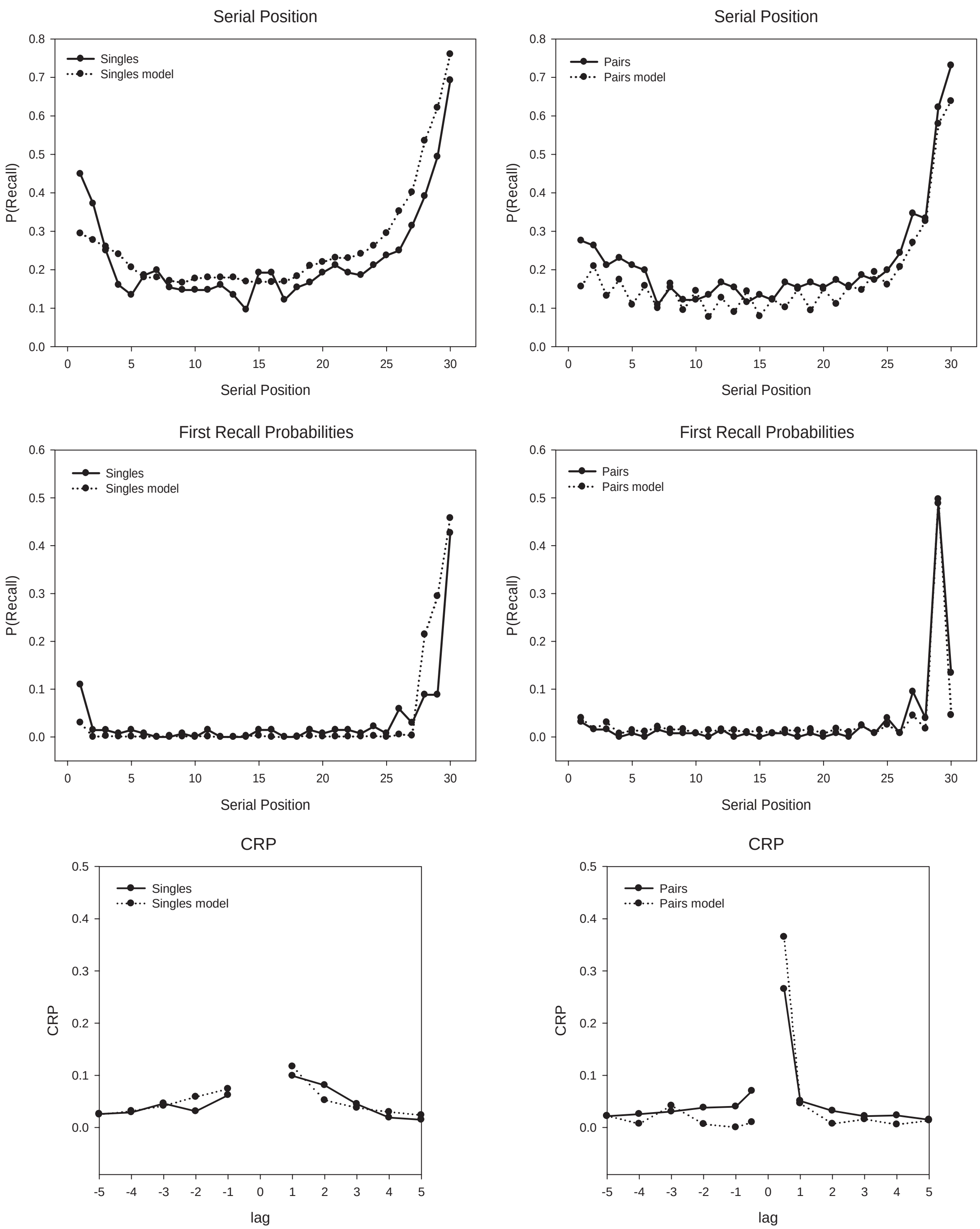


Figure 6. Single versus pairs data and model predictions (Experiment 4). CRP = conditional recall probability. For pairs, a transition within a pair is represented by the point between lag 0 and lag 1 (or 0 and –1). A lag of 1 indicates moving from the last item in one pair to the first item in the subsequently presented pair (and the reverse for a lag of –1). All other transitions represent transitions to individual items in different pairs.

items rather than just items that exist in a privileged state; thus, response times should be longer to output the first item from a long list than a short list. Although the Lehman–Malmberg model is not a model of reaction time, these predictions are consistent with those of other models of reaction time (Davelaar et al., 2005). The prediction is based on the assumption that reaction times are a function of the probability of sampling a given item compared to the probability of sampling all items in a retrieval set; the proba-

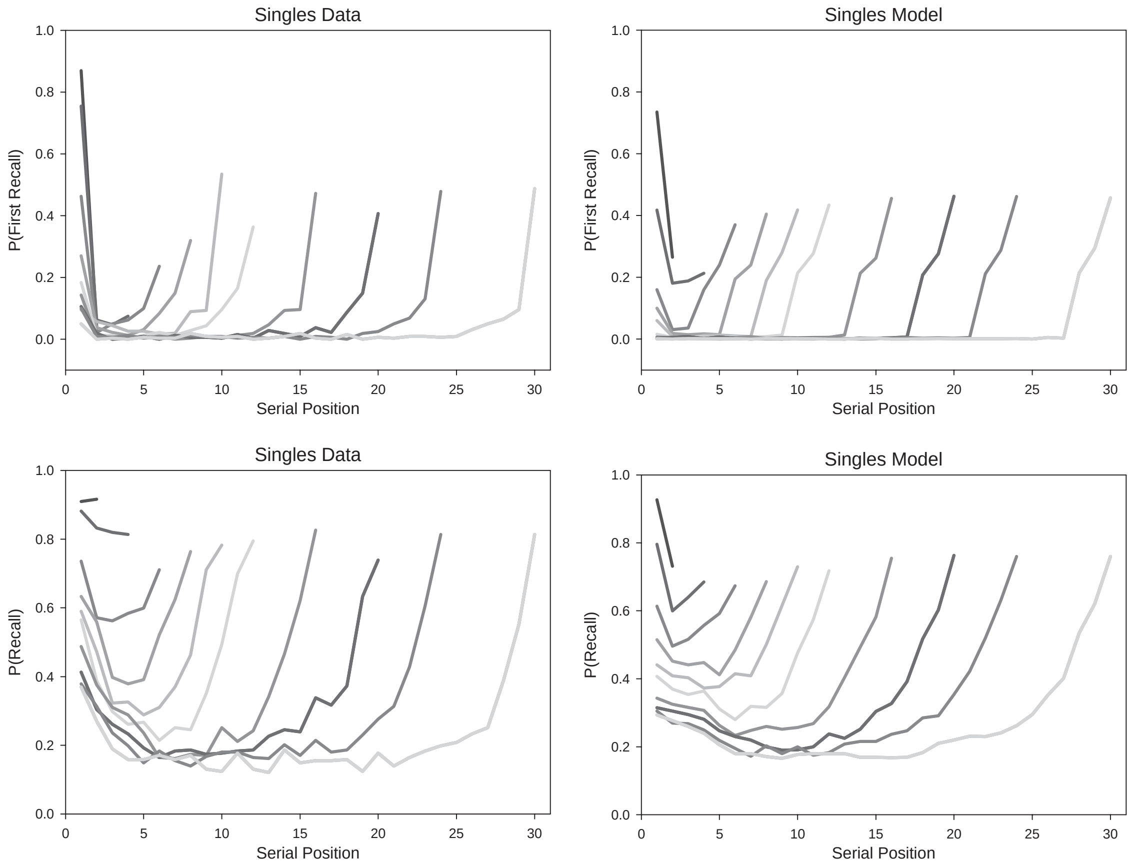


Figure 7. Single-item list-length data and model predictions for first recall probabilities and serial position effects (Experiment 5). First recall probabilities are shown in the top row, and serial position effects are shown in the bottom row. Data are shown in the left column, and model predictions are shown in the right column. Each line represents a given list length (e.g., the line ending at serial position 30 represents data from a list length of 30 items).

bility of sampling a given item will be greater when there are fewer items in the retrieval set, as is the case with sampling from a buffer, than when there are more items in a retrieval set. Thus, if the first attempt at retrieval is restricted to the items in a buffer, the probability of sampling a given item in the buffer would not be affected by list length, and reaction time for first item recalled should be consistent across list lengths. If, however, retrieval does not occur from a buffer, the likelihood of sampling a given item on the initial attempt to retrieve will decrease with increases in list length, predicting a greater reaction time for items from longer lists.

To test these predictions, we conducted an experiment utilizing single-item and paired-item study lists, where list length was also manipulated (cf. Ward et al., 2010), and examined SPs, FRPs, and CRPs, in addition to reaction time to output the first item recalled for each list (the method is described in the Appendix; Experiment 5).[12] As shown in the left panels of Figure 7, there is a shift from primacy toward recency as list length increases for both first recall probabilities and serial position effects; for longer lists, subjects are more likely to begin recall with the last item on the list and more likely to recall items from the end of the list than the beginning of the list. The left panels of Figure 8 reveal that the single-item and pair functions relating list length to FRPs are quite different. Whereas subjects are likely to begin recall with the last item on the list for long lists of single items, subjects are most likely to begin recall with the penultimate item on the list for paired items, $t(158) = 7.64$, $SE = 0.04$, $p < .001$. This pattern is consistent for all list lengths greater than four. Additionally, for all list lengths we see a zigzag pattern in the FRPs; items that are the first member of a pair are more likely to be recalled first than items that are the second member of a pair (for which the probability of first recall is almost zero). This was confirmed by a chi-square test comparing the likelihood of first recalling the first member of a pair versus the second member of a pair for a list of 24 items, excluding the first and last pairs on the list, $\chi^2(1) = 16.71 > 3.84$. The zigzag pattern does not occur for single items, $\chi^2(1) = 3.00 < 3.84$. Another difference between the patterns of FRPs is shown the bottom panel of Figure 8; throughout the list, when the first item from a pair is recalled, the second item from that pair is

[12] Due to the nature of the experiment, we sometimes focus on qualitative patterns visible in the data rather than quantitative statistical comparisons. For instance, the examination of a serial position by list length interaction is not possible, as each list length has a different number of serial position points. Fortunately, the results are clear.

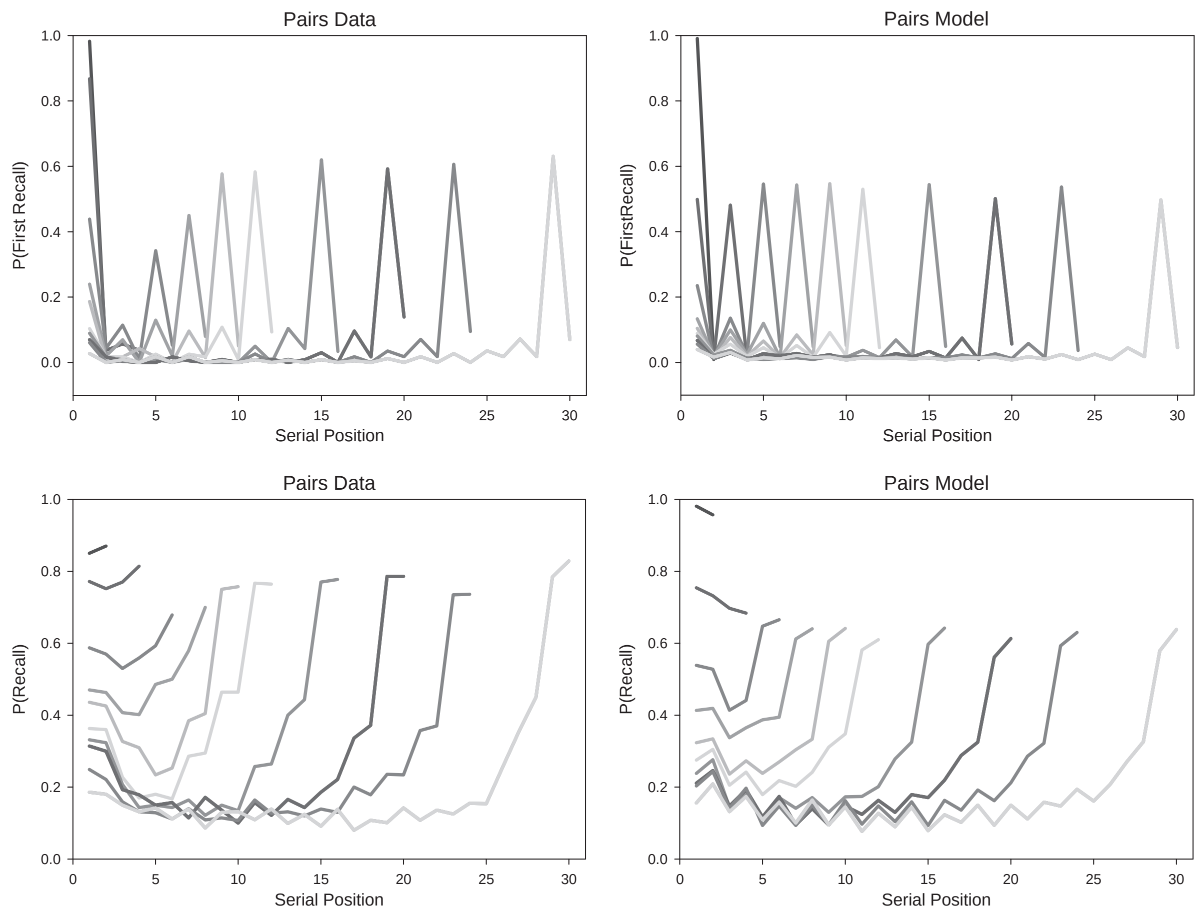


Figure 8. Paired-item list-length data and model predictions for first recall probabilities and serial position effects (Experiment 5). First recall probabilities are shown in the top row, and serial position effects are shown in the bottom row. Data are shown in the left column, and model predictions are shown in the right column. Each line represents a given list length (e.g., the line ending at serial position 30 represents data from a list length of 30 items).

recalled with almost equal probability, and this is obviously not the case for adjacently studied single items.

The differences in single-item and pair study are also observed in lag recency. Figure 9 shows CRPs for two list lengths, one short (six items) and one long (24 items). For both the short list and the long list, we see the typical patterns in the CRPs: a greater likelihood to transition to a nearby serial position and an asymmetry in that recall is more likely to move in the forward direction. For both list lengths, there is a greater likelihood of moving forward one item within a pair than moving forward one item for single items (ps $< .05$).

Finally, we examined reaction time to output the first item recalled. This was measured from the time the test began until the subject pushed *Enter* to submit the word. One subject whose mean response times were greater than three standard deviations from the mean was removed from these analyses. As shown in Figure 10, reaction time to output the first item recalled was consistent across all list lengths. Reaction times did not differ between single- and paired-item lists, and there was no effect of list length on reaction time or no interaction of condition and list length (ps $>$ .05). These data give support to the theory that initial retrieval occurs from a privileged buffer state in immediate free recall.

Because subjects could decide where to begin recall for each list, it is possible to take a closer look at the effects of this decision on time to output the first item during recall. If items in the buffer are protected from interference, increasing list length should not affect response time to retrieve the first item when recall begins with an item in the buffer. However, increases in list length should produce interference, increasing response time, when recall begins at the beginning of the list. Thus, we compared the relationship between list length and response time for trials in which recall was initiated from the buffer (one of the last three items on the list) versus trials in which recall was initiated from the beginning of the list (one of the first three items on the list). Although the time to output the first item was positively correlated with list length when recall began at the beginning of the list ($r = .72$, $p = .03$), there was no correlation when recall began at the end of the list ($r = -.10$, $p = .81$).

In general, the model does a good job of fitting the data. It predicts the correct patterns of SPs and FRPs across list lengths.

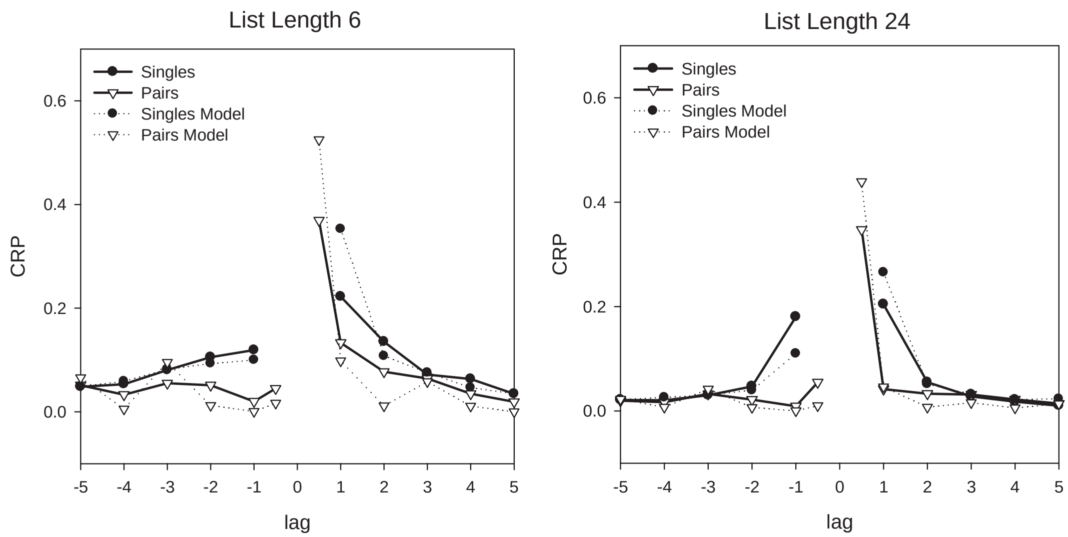


Figure 9. Conditional response probability (CRP) data and model prediction for single- and paired-item study lists of lengths 6 and 24 (Experiment 5). For pairs, a transition within a pair is represented by the point between lag 0 and lag 1 (or 0 and −1). A lag of 1 indicates moving from the last item in one pair to the first item in the subsequently presented pair (and the reverse for a lag of −1). All other transitions represent transitions to individual items in different pairs.

The model also produces good qualitative fits for the CRPs, where lag +1 transitions are more likely for pairs then for single items. Quantitatively, the model overpredicts such transitions, especially for pairs. This is mainly due to the simplifying assumptions that we made related to chunking processes: Context is stored for only the first item in a pair, and if the first item in a pair is retrieved, the next item sampled will always be the second item from a pair. Introducing some variation in both of these components, such that context is weakly stored for the second item in a pair and other items are allowed to be sampled after the first item in a pair is retrieved, would likely lead to more accurate quantitative fits. However, at this time, we are more interested in whether the model is able to produce the qualitative patterns of data in these conditions, which it does quite well. One key pattern accounted for by the model is the strong tendency to first output the penultimate item on paired lists (the first item in the last pair) and the last item on single lists. Such differential rehearsal patterns created by differing experimental designs have created challenges in other models (Laming, 2010).

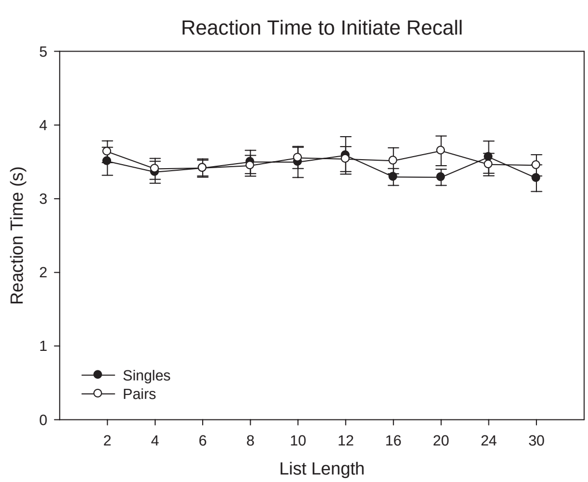


Figure 10. Reaction time data for single- and paired-item study lists of each length (Experiment 5). Error bars represent standard error.

It should be noted that although recall patterns were similar for all list lengths greater than four, these patterns differed for short lists (i.e., lists less than four items long). This finding is consistent with prior work showing that for short lists, subjects typically begin recall with items at the beginning of the list rather than the end of the list, as they do with long lists (Ward et al., 2010). As our intention is to examine predictions regarding the privileged state of items present in the buffer, we are more interested in model predictions at longer list lengths. The model fits the data less well for lists of lengths two and four, because it includes no assumptions about different retrieval processes used for short lists. Additional assumptions would be necessary in order to fit the data for short lists. For example, one might assume that when the capacity of the buffer has not yet been reached, subjects rely solely on reinstating the beginning-of-list context as a cue rather than relying somewhat on current context.

General Discussion

Lehman and Malmberg (2009, 2011) developed a model devised within the REM framework, which is a direct descendant of SAM (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1980) and the Atkinson and Shiffrin theories of memory (Atkinson & Shiffrin, 1968; Shiffrin & Atkinson, 1969); the theory evolved in response to new findings and new applications. However, all the models developed within this framework over the years share a set of core assumptions. One foundational assumption is that episodic memory is affected by both the structural properties of memory and control processes. The structural properties consist of memory stores and processes of encoding and retrieval from them. Control processes serve transient goals of the subject to perform a specific task. The result is a comprehensive integrated account of dozens of mnemonic phenomena. However, there were several noteworthy findings that persisted in remaining outside the scope of prior models. Indeed, some of these findings were taken by many as being a challenge to the role of the control processes, specifically the rehearsal buffer, in producing them.

The present challenge was to account for these findings within the framework of the Lehman–Malmberg model and to test several a piori predictions of the model that was originally developed to account for remembering and forgetting of multiple lists (Lehman & Malmberg, 2009). The primary focus of the present investigations was the model of free recall to provide accounts of the effects of different types of rehearsal, the fluctuations in recency observed between tasks and between individuals, chunking, and its effect on the dynamics of retrieval. Last, we presented the results of a free recall experiment that manipulated list length, finding that increasing list length decreased overall accuracy but increasing the number of items studied did not affect the probabilities of first recall or their latencies. The latter finding is difficult to reconcile with single-store models or models that do not assume that items are in relatively accessible state while they are being rehearsed (e.g., context maintenance and retrieval, or CMR; Polyn, Norman, & Kahana, 2009). Thus, we have documented the sufficiency of the dual-store model and, to the extent that the single-store framework is challenged, we have also documented its necessity.

Capacity Limitations

A key assumption of the Lehman–Malmberg model concerns the nature of capacity limitations. Prior models have often been characterized as having a capacity limitation specified in terms of the maximum number of items that may be simultaneously rehearsed. Although there is undoubtedly a ceiling on this number, the buffer capacity as viewed by Atkinson and Shiffrin (1968) is dynamic. The number of items that are simultaneously in the process of being rehearsed depends upon factors such as the goals of the subject, task demands, and the nature of the stimuli. In this sense, the capacity of the buffer is regulated or controlled in order to serve the needs of the subject. Our model adopts this position as well. However, we also assume that the number of items that are simultaneously rehearsed affects the extent to which items are encoded and bound to the context in which they are encountered. We specifically assume that as the number of items being rehearsed increases, the amount of context stored with each one decreases. That is, there is a limited capacity for associating items to the learning context, and this capacity therefore must be distributed in a fashion that serves the goal of subject.

One way to regulate the degree to which items are bound to the encoding context is by varying the number of items to be rehearsed. Decreasing this number increases the strength of association of items to the context, on average. A notable exception concerns the first item on the list, which according to the present model is always bound most strongly to the context as long as the goal of the subject is to rehearse more than one item at a time. Thus, as tasks or situations during encoding change, the extent to which items may be bound to the context may vary. This assumption is related to the one-shot hypothesis insofar as both assume that the time course of encoding of different types of information is different, with context being encoded relatively quickly and automatically and item information being encoded over longer time frames and at times in controlled fashion (Malmberg & Shiffrin, 2005).

Maintenance and Elaborative Rehearsal

Elaborative rehearsal, often described as rehearsal that requires the processing of semantic or associative information (Craik & Lockhart, 1972), is in actuality conceptualized rather vaguely, insofar as it is said to produce a deeper level of processing (Nelson, 1977). We have assumed that the additional processing and attention paid to semantic features lead to the encoding of additional item features compared to maintenance rehearsal but do not result in the encoding of additional context features (Malmberg & Shiffrin, 2005), which allows the REM framework to account for a variety of dissociations between implicit and explicit memory tasks. We have also assumed that maintenance rehearsal primarily leads to the encoding of item information, which has relatively small effects on free recall and larger effects on recognition performance (Glenberg et al., 1977). From a mechanistic perspective, during maintenance rehearsal the buffer size is set to one item because there is a need to maintain only a single item in memory, but when more than one item is rehearsed an interitem association is created between them (Raaijmakers & Shiffrin, 1980). Under elaborative encoding conditions, for instance, more time in the buffer produces more opportunities to store associative information and other types of elaborative information, such as extralist items retrieved from lexical/semantic memory, which may be incorporated into the rich memory traces (Atkinson & Shiffrin, 1968). Thus, as the storage of elaborative/associative information has marked effects on free recall, more time spent on elaborative processing in the buffer will produce larger increases in free recall performance than maintenance rehearsal does.

Compartmentalization

Prior buffer models assumed various rules for when items were dropped from the buffer. Typically, these items were randomly selected or the item that had resided in the buffer the longest was dropped or "knocked out." This model is the one that we adopted to describe the rehearsal process involved in studying a list of single, unrelated items. The result of this model is a series of long-term traces consisting of multiple item–item associations. This can be analogically visualized as a long chain of episodic traces—although this is not strictly the case as there are no links per se—that are related temporally via their context representations and via the similarity of the items that they comprise. This representation of the list is a direct reflection of the control process invoked during study.

Changes in the buffer operations in response to task demands or changes in the stimuli will affect the list representation in memory. For instance, in order to account for the costs and benefits of list-method-directed forgetting, Lehman and Malmberg (2011) proposed a compartmentalization process that actively drops items from the buffer in favor of items that may serve task demands.[13] The dropping of items from the buffer affects the nature of the long-term traces that are stored. For instance, we assumed that, during the study of pairs of items, the capacity of the buffer would be reduced by virtue of the compartmentalizing of each pair. The resulting long-term trace consists of one item-to-item association and two item-to-context associations, with strength of the item-to-context association being greater for the item of the pair that was initially the focus of attention. Hence, the list is represented as a series of chunks related temporally via the similarity of the contextual presentations.

The operation of the buffer is closely related to the construct of attention in psychological science; due to the limited capacity to process information when performing many tasks, attention is required for the selection of information to which one may allocate one's processing efforts. The notion that there is a control process that is complementary to rehearsal may be novel in the memory literature, but such ideas have been proposed in the visual attention literature by Posner and Petersen (1990). To wit, they proposed that to attend to a new item, one must first disengage attention, shift attention to the visual location of the new item, and then engage attention. The processes of engaging and disengaging visual attention may be analogous to rehearsal and compartmentalization. The first allows for the maintenance of information in a ready state to use in order to complete a goal, and the latter allows one to disengage from a prior item in order to engage a new item.

The Encoding of Context and Its Use as a Retrieval Cue

Akey assumption of all models of memory is that what is retrieved from memory depends on the similarity between the retrieval cue used to probe memory and what is stored in memory. The more similar the cue is to the contents of a given trace, the more likely that trace will be retrieved. Conversely, the more similar the cue is to other traces in memory, the less likely a given trace will be retrieved. Minimizing the interference from irrelevant memory traces is the goal of constructing effective retrieval cues. For most laboratory experiments, there are several relevant traces representing the items encountered on a study list, but for free recall very little information is provided to the subject that he or she may use to construct a retrieval cue. One might suspect, therefore, that a context cue representing the current set and the setting of the subject is used as an initial probe, but this almost certainly too simplistic. Subjects are quite good at retrieving items from prior study lists without generating too many intrusions from more recently studied lists (e.g., Shiffrin, 1970), indicating that the subject must be able to reinstate context features associated with prior lists. Moreover, a probe with the current test context in many cases makes no sense, if the goal is to minimize the amount of interference from irrelevant memory traces. For instance, in Experiment 3, we used a 5-min retention interval filled with the presentation of a "How It's Made" video. If the goal was to retrieve information from memory about "How It's Made," it would make somewhat more sense to probe with the current test context, but the free recall task in fact instructed subjects to skip over that information and recall what preceded the retention interval.

Another challenge for the assumption that the current test context is used as an initial probe of memory is that the patterns of FRP (first recall probabilities) observed are decidedly nonrandom. The most interesting—and challenging—patterns of FRPs are found in delayed free recall, where subjects usually initiate retrieval by recalling the item from the first serial position. This finding is predicted by the buffer model on the assumption that context encoding is directly related to the number of items currently being rehearsed. Hence, a free parameter directly related to the difference in the strength with which context is stored over the course of list is the capacity of the buffer. The processes of compartmentalization are also related to the strength of context encoding insofar as when the current contents of the buffer are eliminated and a change of mental context is the result, the strength with which context is bound to the next item to be rehearsed is increased. These assumptions allow the model to account for the effect-intentional forgetting on the FRPs (Lehman & Malmberg, 2009).

Over the course of the list, we also assume, context changes at a relatively slow rate. This allows the model to account for the bidirectional lag-recency effects, especially over very short lags. However, the change in context over the course of the lists provides a new challenge; for delayed free recall, how do you account for the primacy in FRPs? Although the model predicts that context will be encoded most strongly for items at the beginning of the list, the change in context provides an advantage for the last items on the list when the test context is used to probe memory.

Therefore, to address both logical flaws in the assumption that current test context is used as an initial probe during free recall and the data that suggest that it is wrong, Lehman and Malmberg (2009) assumed that memory is probed with a mixture of current test context and the context that was stored most strongly during study. We assume that the current test context plays some role in the probe of memory, because context effects are commonly reported for free recall (e.g., Godden & Baddeley, 1975), and there is also evidence the reinstatement of context can be beneficial (Sahakyan & Kelley, 2002; Smith, 1979). Moreover, the context from the beginning of the list is mostly likely to be retrieved, as it was encoded when the lowest number of items was being rehearsed. This compound cue, composed of current and beginning-of-list context, allows the model to account for a variety of findings, including primacy in the FRPs and lag-recency. In the model, γ_I is the proportion of features from the beginning of a list

[13] To develop an account of intentional forgetting, we experimented with several models, and we identified the key assumption that a cognitive control process in the form of a buffer was necessary in order to account for the effects of directed forgetting on recall rates, serial position data, and first recall probabilities (see the Appendix in Lehman & Malmberg, 2009, for a description of various models that were found to be insufficient to account for intentional forgetting). Critically, a simple change in mental context was not sufficient to explain the full range of effects associated with the instruction to forget. Although a context change produces the costs and benefits of directed forgetting, it was not sufficient for producing the primacy and recency patterns seen in the data. Hence, a mechanism for producing the change in mental context was proposed in the form of compartmentalization, which leads one to drop all items that are extraneous to the task at hand from the rehearsal buffer.

that is reinstated to be used to initially probe memory, which we assume is negatively related to the amount of context change that has occurred, and γ_I is allowed to vary freely in the model for most of the simulations reported (see also Lehman & Malmberg, 2009, 2011). In other cases, such as the list-length simulations, we directly related γ_I to the amount of context change that has occurred between study and test. The latter approach is preferred, of course, but at this time we have no modeling for making direct measures of the amount of context change caused by other factors (e.g., instruction to forget) independently from the model.

Context Reinstatement and Individual Differences in List Memory

The buffer model does a decent job accounting for various patterns of interference and forgetting, in addition to the order in which items are retrieved from memory. Additional insight into memory task performance is provided elsewhere with extensions of the basic buffer model to account for tasks and variables (Lehman & Malmberg, 2009, 2011; see also Malmberg, Lehman, & Sahakyan, 2006; Malmberg & Shiffrin, 2005). Consider the overwhelming tendency of subjects to retrieve specific memories representing the first item on the study list during delayed free recall even at very long retention intervals (see Figure 5). Primacy in first recall probabilities requires the mental time traveler to land on a dime in memory space. To do so, the model assumes, subjects perform delayed free recall probe memory with a mixture of the test context and context features that are most strongly encoded during study; because context is most strongly stored with the initial item, the context features encoded during the earliest study positions are most likely to be reinstated. The control processes needed to reinstate the context features used in an initial probe of memory are not well understood, but using the duffer model we have made some headway.

The relationship between primacy and context reinstatement. Although recency, rehearsal, and different forms of encoding have been extensively investigated (Craik & Lockhart, 1972; Glanzer & Cunitz, 1966; Nelson, 1977; Rundus, 1971; Slamecka & Graf, 1978; Tan & Ward, 2000), less effort has been made to understand primacy and how access to memory is controlled (but see Tan & Ward, 2000). However, the relationship between primacy in the first recall probabilities and the control of memory search is quite informative for models of context reinstatement, because the item that is output first during free recall is determined in large part by the context cue used to initially probe memory.

According to the buffer model, primacy in both the serial position function and the first recall probabilities is the result of a relatively strong association between the primacy items and context and the use of these context features in an initial probe of memory during delayed free recall. Items being rehearsed or encoded are bound to the current context, and the strength of item–context associations is constrained by the system's limited capacity to store information; as the number of buffered items increases, the strength with which the items are associated to the context decreases. This assumption produces primacy for two reasons. First, because context is stored most strongly during the primacy portion of the study list, those context features present during the primacy portion of the list are most easily reinstated. Second, when memory is probed with reinstated context features, primacy items have a retrieval advantage over other items because the context to which they were bound better matches the context features used to probe memory.

A key question is just how those context features are reinstated in first place in order to create the initial retrieval cue. Indeed, this may be the $64,000 question for episodic memory research. However, the models are elusive because the context reinstatement is a complex process that involves the induction of generalizations from recent experience, an interaction of general knowledge about certain events with prior knowledge about the world, and structural aspects of episodic memory including the nature of temporal context.

We find that studying primacy in relatively complex tasks is useful when modeling the reinstatement of context. For instance, multiple-list memory tasks are different than single-list memory tasks insofar as there are at least two relevant contexts in which targets may have been encountered, although retrieval should be confined to one of them when following task demands. Multilist experiments are ideal for examining context reinstatement, as context must be used to localize memory access in order to retrieve from a specific list (and perhaps edit inappropriate items from being output; cf. Lehman & Malmberg, 2009). To see why, first consider the experiment in which we observed a relationship between individual differences in the form of the serial position function and the overall level of immediate free recall performance (see Figure 4). The subjects in that experiment completed eight study–test cycles, and overall performance actually increased over the course of the experiment despite the potential buildup of proactive interference during the course of the experiment. In a similar recognition memory experiment, memory did not improve with increases in the number of study–test cycles; it actually got worse (Lehman, 2011). Thus, it is not that proactive interference is not a potential problem for the subject completing multilist memory task. Rather, subjects are very good at forming an initial cue to probe memory. The quality of the retrieval cues generated by subjects is also reflected in very low intrusion rates, even though by the end of the experiment there are several traces containing relatively strongly encoded context features representing the initial items on prior lists.

To account for the surprising lack of proactive interference in immediate free recall, one might assume that control processes operating during retrieval evolved over the course of the experiment in order to increase overall performance. Indeed, our buffer model was designed to account for multiple-list memory tasks, including free recall (Lehman & Malmberg, 2009, 2011).

One clue about the sort of context features that are reinstated and used to probe memory comes from the relatively high performance on multilist memory tasks (cf. Shiffrin, 1970). A retrieval cue is used that likely contains information representing temporal relationships between lists. Temporal context has been assumed in prior models to change relatively slowly within list and more quickly between lists (Malmberg & Shiffrin, 2005; Mensink & Raaijmakers, 1989), although a certain overlap in temporal context features is a reasonable assumption. In fact, the assumption of a between-list correlation between the temporal context information predicted the observed and very diagnostic pattern of costs and benefits associated with intentional forgetting, implicating the primacy portion of the serial position curves as the main contributor (Lehman & Malmberg, 2009).

Temporal context. The issue of how context is reinstated is also closely related to the question, what is context? Or, more specifically, what does temporal context represent? The obvious answer is that temporal context represents the passage of time per se, but this definition is probably too restrictive; from a mathematical perspective, temporal context is simply a representation of some aspect of experience that is correlated among a set of memory traces (cf. Estes, 1955). Temporal context may represent aspects of our environment, the goals of the subject, the task performed by the subject, or any internal state that is stored in several traces.

Effective performance of a multiple-list memory task requires the retrieval of temporal context in order to generate the initial cue with which to probe memory. According to our buffer model, it is unnecessary for the subject to intentionally retrieve the beginning-of-list temporal context features; these may be retrieved simply because they are most strongly or perhaps most frequently encoded and hence are most characteristic of the target list and therefore discovered by an inductive process. This suggests that at certain times temporal context comprising information about the typical item on a study list may be reinstated and used to probe memory. In one multiple-list study, for instance, subjects studied categorized lists. In one condition, we provided the category-context cue that was uncorrelated between lists, whereas in the other condition, we provided the list number as the retrieval cue. As predicted by the buffer model, proactive interference was eliminated when the category cue was used to probe memory because the temporal context used to probe memory was correlated only within a list and not between lists (Lehman & Malmberg, 2011).

Thus, a variety of context cues may be reinstated, and the cue used can have profound impacts on recall performance. And just as general knowledge inducted through experience about the compositions of different memory lists may be used to cue memory when performing a multiple-list free recall task, knowledge about the way the world works may also influence the reinstatement of the context used to retrieve items from memory. For example, individuals might be asked to report in any order what they had for dinner last Friday. This is a free recall task, and the first step is to generate the appropriate retrieval cue, which may be compared to a problem that must be solved; when the menu is reported, the problem has been solved. Consider how K. Malmberg might accomplish this task. He might know by looking at the date on his computer that it is September, that classes just started, and recall that last weekend was Labor Day weekend. His family typically goes to the beach for Labor Day weekend, and this information could be used to further probe memory to isolate the relevant traces in memory. This year his grandparents called to say they were arriving late and that they were to meet at Angelo's in Panacea, Florida. This probably means that seafood was on the menu, as he knows that Angelo's is a seafood joint. Indeed, he and his wife shared a seafood platter of oysters, gulf shrimp, and grouper, and the family drank beer and had a Greek salad served with potato salad.

A similar process may be invoked when an individual is asked to retrieve from a particular list in the lab. He may know that he was in an experiment, which consisted of several study lists, each separated by a short math task, and he may use this knowledge of the structure of recent experience to locate specific traces in memory; the more distant the lists, the more difficult it is to arrive at the appropriate retrieval cue.

Individual differences. Because individuals differ in past experience, it is quite obvious that the information that they use to construct retrieval cues and probe memory may be responsible for variability in performance between individuals. The author's experience is different than the reader's, and although the goal of reporting what one had for dinner last Friday is the same, the paths and even strategies for retrieving this information may be quite different. Of course, many of these individual differences are limited by the constrained nature of very simple memory experiments, but individual differences in memory also reflect the integrity of the structural aspects of memory and the efficiency of the strategies used to control encoding and retrieval. For instance, Unsworth, Engle, and colleagues have documented associations that exist between an individual's working memory capacity and the ability to retrieve from memory over the course of several minutes (Unsworth et al., 2011; Unsworth & Engle, 2007). This important finding, in our view, suggests a model that assumes that, in the absence of queries consisting of explicit information or information specific to a single prior event, retrieval reflects a mixture of general knowledge the subject has about how the world typically behaves based on his own individual experiences and that the ability of the subject to manage this information is related to capacity of working memory; those with more working memory capacity will generally be more successful at retrieval from long-term memory because they are better able to manage the construction of a series of retrieval cues with which to probe memory. One might object to this reasoning based on our finding that subjects who are better at free recall are more likely than subjects with lower overall free recall performance to retrieve from buffered items before primacy items (see Figure 4). However, it makes sense for those who have difficulty creating or managing retrieval cues to begin with primacy items, as the construction of retrieval cues is going to be easiest in the absence of output interference caused by reporting the buffered items.

Comparing the Lehman–Malmberg Model to Other Models

A role for control processes. A key feature of the current model is that it is formally related to a variety of other models developed within the REM and SAM frameworks. It is therefore potentially useful in a formal manner when attempting to account for the performance of tasks such as cued recall, associative recognition, plurality discrimination, and free association. Although they are not formally related to the theory of distributed associative memory (TODAM; Murdock, 1997), both models provide mechanisms by which relatively complex memory traces may be created. A key difference between the current model and TODAM is that we related these memory structures to specific tasks, stimuli, and goals of the subject, thereby identifying them as the product of cognitive control processes. Although the structural components of our model and Anderson et al.'s (1998) adaptive character of thought–rational (ACT-R) model are very different, ACT-R's chunking model and our model are similar insofar as they are both buffer models and ACT-R would likely make similar predictions to the current model regarding chunking. According to

ACT-R, chunks are stored by creating associative links between items. As links in the ACT-R model can be differentially weighted, differentially adjusting the link weights for items within a chunk and items from different chunks would likely produce patterns similar to those seen here. Hence, although the nature of the representations that we propose is distinctive of the REM framework, other models that incorporate similar buffer operations may also be useful in understanding how subjects regulate encoding activities in order to achieve their goals vis-à-vis task demands.

Davelaar and colleagues (Davelaar et al., 2005, 2006) developed a dual-store neurocomputational model of memory that includes both short-term and long-term memory components inspired by neurophysiological processes. This short-term buffer and a long-term memory store are linked to processes of activation and inhibition in the brain. The Davelaar et al. (2005) model, like ACT-R, conceives of retrieval as based on a trace's activation level; traces currently in the buffer are all assumed to be activated, whereas traces not in the buffer are retrieved based on episodic context matching. Retrieval from the short-term buffer is based on current activation levels. An item from the buffer is displaced when it is deactivated via inhibition from other items that have entered the buffer (see also Grossberg & Pearson, 2008). In this sense, displacement from the buffer is a structural component of memory; it could be adjusted within a single task to allow for control over the contents of the buffer, such as intentional displacement. Such an assumption to allow for intentional control over the contents of the buffer would only enhance the ability of the Davelaar et al. model.

On the other hand, a clear difference between the Davelaar model and our model concerns the need for an inhibition of traces in order to drop them from the buffer once they have served their purpose. Our model does not assume that inhibition of a trace is necessary in order to allocate resources to rehearsing another item. Although our data do not provoke us to make a strong argument against activation models like the Davelaar et al. model, we believe our model provides more consistency with past work. For example, it is possible to maintain a small number of items in a rehearsal buffer while processing unrelated sentences (Baddeley & Hitch, 1974). Such findings are consistent with the idea that the rehearsal buffer is a flexible control process that can be adapted to a task and are challenging to models that conceive of the buffer as a structural short-term memory store. Although Davelaar et al. (2005) did not propose a model of control processes, they did address the ways in which such processes may contribute to the model. For example, they suggest that items can be selectively maintained or inhibited in their buffer. However, although Davelaar et al. (2006) showed that semantic relatedness between items can contribute to mutual activation in the buffer, such activation levels are influenced automatically by pre-experimental associations between items and are not under the control of an individual. Further, at this time, their model has not been used to account for rehearsal processes during encoding and serial order information that results from the use of the buffer as a control process during encoding.

In contrast to buffer models, the temporal context model (TCM) assumes no role for control processes in recency effects or contiguity effects (Howard & Kahana, 2002). For instance, fluctuations in recency are solely due to changes in context that occur between study and test. Accordingly, context changes continuously during study such that context stored on one trial is positively correlated with context stored on the prior trial. Because the context associated with items at the end of a sequence is more similar to the context used to probe memory during retrieval, the end-of-list items are more likely to be recalled than items studied earlier. Increases in the length of the retention interval decrease recency on the assumption that the context used to probe memory becomes less similar to the context stored during study. This is similar to what we have proposed, but unlike the Lehman–Malmberg model, TCM always predicts recency in probabilities of first recall. However, whereas some observations indicate that the first item output in free recall tends to be the last one studied, other observations indicate that the first item output in delayed free recall tends to be the first item studied (Lehman & Malmberg, 2009). Thus, the order in which items are recalled is variable. This observation is provocative insofar as it suggests the possibility that situational factors and indeed even individual differences might be sources of this variability, which is what we observed in the experiment described above. Currently, there is no mechanism by which TCM predicts the primacy in first recall probabilities.

The debate over long-term recency. A fundamental test of TCM is its account of long-term recency. Critical long-term recency data are obtained from the free recall following continuous distraction (Bjork & Whitten, 1974), and on this note Howard and Kahana (2002) concluded,

> Recency and contiguity [i.e., lag recency] reflect basic memory processes that probably do not depend on a short-term store. This of course does not preclude an influence of STS [short-term store] (or something very much like it) on recency or formation of associations in situations where active maintenance of information takes place. (p. 290)

The critical assumptions are that during periods of distraction, the subject does not covertly rehearse items from the prior study trial and that as a result they are dropped from the buffer. However, extant data do not support these assumptions. For instance, there is a host of observations that indicate that the effectiveness of distractor tasks in eliminating active rehearsal is limited (for a review, see Gardiner et al., 1974). The major implication of this is that buffer operations are likely to be attenuated but still involved in the processing of items during periods of continuous distraction. This is consistent with the assumption of the present model in which the encoding of associative information is attenuated but not eliminated by continuous distraction.

Despite the ability of dual-store models to account for long-term recency, modeling long-term recency has been controversial. Davelaar et al. (2005) showed how a buffer model predicts long-term and lag-recency effects, and they described dissociations between short-term and long-term recency that were inconsistent with the single-store TCM. Sederberg et al. (2008), however, described an elegant single-process model, referred to as TCM-A, which accounts for these dissociations on the assumptions that the current context state is driven by studied items and that the current contextual state is maintained until new items are added, in which case the older features must decay. Usher et al. (2008) argued that TCM-A in fact was a buffer model, as the limited-capacity nature of the maintained contextual representation is analogous to a fixed-capacity buffer where items are displaced to make room for new items. Finally, Kahana, Sederberg, and Howard (2008) argued that TCM-A is quite distinct from classic dual-process buffer models, which is not necessarily to say that TCM-A is not a buffer model or that its account of long-term recency does not hinge on the maintained contextual representation.[14]

Modeling the buffer. This debate over long-term recency highlights an important distinction between single-process and dual-process models of free recall and in fact may be recast as a debate over how to model buffer operations and representations. Indeed, we favor such an interpretation, especially insofar as it highlights a distinction between the Lehman–Malmberg model and prior buffer models. Traditionally, dual-store models were also dual-process models, in which retrieval occurred via different processes for short-term and long-term memory stores. Whereas a single-store model utilizes only a single-retrieval process, prior dual-store models have assumed that items in short-term memory are "dumped" during retrieval, in an automatic fashion, and items in long-term memory are retrieved via another process, such as sampling and recovery (Atkinson & Shiffrin, 1968). TCM assumes that retrieval occurs via a single context-driven retrieval process, whereas Davelaar et al.'s model is a dual-process model: First, the buffer is dumped, then the retrieval process continues for the remaining items on the list based on contextual matching.

Although proponents of single-store models favor the more parsimonious single-process approach to retrieval, a buffer does not preclude the simplicity of a single retrieval process. For instance, although TCM-A is a single-process model, and even though the limited-capacity, actively maintained contextual representation system is not called a buffer, it contains many of the essential features of a dual-store buffer model. To wit, "This pre-experimental context representation is then added to the current state of temporal context, which must first decay *to make room* [emphasis added] for the newly inserted item" (Sederberg et al., 2008, p. 897). Thus, TCM-A may be viewed as a dual-store single-process buffer model if we exchange the buffer for the current state of temporal context. The upshot is that the current debate over long-term recency is as much about how to model the short-term buffer as it is about anything else.

Like TCM-A, the Lehman–Malmberg model diverges from classical buffer models in that it uses a single retrieval process that was originally described by Raaijmakers and Shiffrin (1980). During retrieval, items being rehearsed by the buffer are assumed to be in a privileged state, such that they are more easily accessible and resistant to interference from other list items. Retrieval occurs via a sampling and recovery processes, determined by the match between a trace and the cue used to probe memory, both for items present in the buffer and for other list items (see Nairne, 1990, for a different but related account of sampling in primary memory). However, the present model is different from TCM-A, as it includes flexible control processes used during both encoding and retrieval. During retrieval, control over the composition of the retrieval cue is maintained in order to most efficiently retrieve from the buffer and/or the long-term store. For instance, we have shown above that subjects who perform at the high level of free recall disproportionately tend to initially retrieve from the buffer during immediate free recall. Those in the worst performing group, much like those who perform delayed free recall, tend to initiate retrieval at the beginning of the list, which is accomplished by probing memory with a cue comprising mentally reinstated context, not the test context. The control processes of rehearsal and compartmentalization further provide for the ability of the subject to adapt the episodic representation that is encoded to task demands and subjective goals. Thus, our model is unique in the role it assumes for control processes. It is also able to account for the data, like the retrieval dynamics of single versus paired study and the effect of list length on first recall probabilities, for which a buffer is necessary to explain not only how items are retrieved but also how items are encoded.

Chunking. As we have described, the current model is similar to TODAM and ACT-R in its ability to encode organized representations that reflect the buffer operations during encoding. The context maintenance and retrieval model (Polyn et al., 2009), based in the TCM framework, also shares many similarities with the current model, including a fluctuating context representation and recall that is driven by contextual states associated with retrieved items. CMR accounts for both temporal and semantic clustering during free recall by assuming that retrieval of an item retrieves the episodic context state associated with studying that item, along with all context states ever associated with that item, and that semantically related items will have occurred in similar pre-experimental contexts. Thus, subsequently recalled items may be items that were studied in nearby temporal proximity or items that occurred proximally in pre-experimental contexts. Similar predictions are derived from our Lehman–Malmberg model. We have shown how the current model accounts for the clustering during recall when pairs of items are studied. However, it may not be necessary to assume that semantic clustering during retrieval, like that observed by Bousfield (1953), is the result of the retrieval of experimental contexts. The use of context-plus-item cues comprising the item information associated with a retrieved item produces clustering during output on the assumption that semantically related items are more similar than semantically unrelated items (Criss & Shiffrin, 2004).

Like TCM, CMR does not include a user-controlled buffer. Nonetheless, it could potentially account for the paired-item data presented here because it assumes that contextual shifts driven by task-related disruptions of cognitive events isolate groups of items. Thus, it may predict the chunking that occurs with the study of paired items. However, a key challenge for CMR is to explain why the first item of pair is invariably recalled before the second item of a pair. According to the present model, compartmentalization in the buffer leads to enhanced encoding of item and context features associated with the first item of the pair, and this enhanced

[14] We have implemented in our model ideas that have been proposed but not implemented in previous models. First, the model elucidates some of the encoding and retrieval processes involved in producing various recall phenomena not accounted for by TCM. For example, as suggested by Sederberg et al. (2008), we explored the mechanisms involved in producing the primacy effect, an issue that has largely been ignored in the buffer debate (but see also Laming, 2010; Tan & Ward, 2000). Not only does our model provide consistent fits to primacy effects under a variety of differing experimental conditions, but it also predicts changes in the magnitude of the primacy effect associated with context manipulations, such as directed forgetting. This is due specifically to the flexible use of the buffer to distribute encoding resources. Further, our model predicts a release of long-term recency when a sufficiently long post-list distractor task is used in a list with a continuous distractor task, and other models have been shown to predict long-term recency whenever a continuous distractor task is used. Thus, although TCM and its variants have been able to account for a large number of data, it is not clear how it could account for the key data patterns presented here that require the flexible buffer.

encoding provides a sampling advantage to it. According to CMR, there is no role for cognitive control in the encoding of chunks resulting from disruptions of cognitive events, and there is no internal structure to a chunk other than it is isolated from other traces.

List-length effects. We argued that the recency invariance associated with changes in list length is difficult to explain with models that do not provide for some sort of privileged state that buffers end-of-list items from the interference of items from earlier serial positions. This does not imply that there is no list-length effect for free recall; rather, the list-length effect for free recall may not be deemed sufficient to discriminate between dual-store and single-store models (Murdock, 1962). For instance, Polyn et al. (2009) showed that CMR is able to account for the list-length effect on the assumption that the final items on a list are insulated from the retrieval competition experienced by other items on the list by the support of the post-list context cue. On the other hand, one might argue that assuming that recent items are insulated from retrieval competition is tantamount to assuming that they are in a privileged state. Moreover, CMR predicts that the last list item will always tend to be recalled first, regardless of list length; thus, this item is truly insulated from the competition produced by other list items. However, we found that the last item was recalled first only 80% of the time. If this item does not receive protection from competition of other items via its perfect match to the retrieval context, it is not clear how CMR would predict that this item is insulated from interference without assuming that it exists in some kind of privileged state. Further, we previously found that buffer processes were necessary in order to account for overall recall and serial position data in directed forgetting (Lehman & Malmberg, 2009), and it would be challenging to account for such effects without the buffer. Thus, the effect of list length on the recency portion of the serial position curve for immediate free recall may not be sufficient for discriminating between single-store and dual-store models, but the effect of list-length manipulations on the first recall probabilities is, and it favors the dual-store model.

Free and serial recall. Whereas in free recall, subjects are asked to recall as many items as they can from a list in any order, a serial recall task requires them to output the items in a specified order. Ward et al. (2010) suggested that under similar encoding conditions, the same mechanisms underlie free and serial recall, and recent efforts have been made to account for both free recall and serial recall via similar mechanisms (Farrell, 2012; Grossberg & Pearson, 2008).

For instance, Farrell (2012) has developed a model of both free and serial recall that shares important assumptions with the Lehman–Malmberg model (2009, 2011; see also Malmberg & Shiffrin, 2005, and Xu & Malmberg, 2007). The most salient convergence between the models is the assumption that list items are organized into higher level units that we refer to as chunks and Farrell refers to as groups, and these higher level units are associated to contextual elements. This assumption is not new, of course (e.g., Miller, 1956; Raaijmakers & Shiffrin, 1980), but prior accounts of serial position functions, first recall probabilities, and lag recency did not propose a role for chunks. This similarity between the models provides reasonable approximations to the effects of specific variables on overall recall performance and to predictions of the order in which items are recalled. It is quite natural, therefore, to relate the models and findings discussed by Farrell to those discussed here.

An account of output order is of course the goal of any model of serial recall, but it is interesting to note that the same mechanisms also seem important to an accurate account for free recall and item recognition (Lehman & Malmberg, 2009). The Lehman–Malmberg model was originally designed to account for long-term list memory, particularly list-method-directed forgetting effects on free recall and item recognition, and then by extension a variety of findings said to challenge buffer models. It is compelling that these basic mechanisms were developed independently to account for very different empirical phenomena. It is also noteworthy that these models are implemented in very different architectures. This indicates that the specific modeling assumptions and parameterizations are not so important for predicting any single finding, as both models when sufficiently explored and extended are likely to be able to account for similar patterns human behavior; rather, the convergence on the assumption that memories are constructed in an organized manner centered on the relationship between items and their associations to context is paramount.

Although there is overlap in some of the basic assumptions of these models, a few points of divergence are relevant to the current discussion. Perhaps the most obvious distinction is that the Farrell model is a single-store model and ours is a dual-store model:

> However, there is no separate buffer or form of memory, and the distinction between short-term memory and long-term memory is purely quantitative: If a group is recalled immediately after its presentation, the context will be intact, and activity-filled delays will force a retrieval of the group context that will generally give an incomplete representation of the group context. One contribution of this paper is to outline a common account of both short-term or "primary" memory and episodic memory without assuming a separate short-term buffer or store. (Farrell, 2012, p. 227)

It is important to accurately characterize the buffer as a process. The buffer controls the encoding of long-term traces and the access to pertinent information, which may or may not occur in a form different from a long-term trace. According to our model, however, we assume that the buffer operates on episodic representation distinct from the long-term traces that are encoded.

The lack of buffer in Farrell's model impacts the manner in which items are encoded and retrieved. Farrell assumes that within a group, the strength of associating items with the group context and the within-group position is determined via a novelty-based primacy gradient (see also Grossberg & Pearson, 2008). The key assumption underlying the primacy gradient is that items encountered earlier in a sequence are encoded more strongly than later items. This is implemented, as in our model, by linking items at the beginning more strongly to the study context. The primary difference is that in our model the "primacy gradient" is derived from the specification of the buffer operations, and in the Farrell model it is an assumption implemented mathematically.

In our view, one advantage of the buffer model is that prediction may be derived from its specifications. For instance, we assume that contextual encoding is influenced by the number of items in the buffer (leading to storage of more contextual information for items at the beginning of a list or the beginning of a chunk), and this is affected by intentional rehearsal and compartmentalization in response to task demands and encoding conditions. As shown in

Figure 2, for example, this allows us to account for the effect of incidental encoding and maintenance rehearsal on the assumption that under these conditions the encoding of interitem associations is halted by reducing the buffer size to 1, which negatively affects all retrieval of long-term traces, and by reducing the strength of the item-to-context associations. The main effect is to a reduction in primacy, which is often observed (Marshall & Werder, 1972).

The lack of influence from cognitive control processes on memory encoding and retrieval (i.e., the reliance on structural aspects of memory) also becomes evident in subtle differences in the organization of memory. For instance, the manner in which items are grouped is a random process in the Farrell model. However, like other models that claim to not use buffers or control strategies to account for the free recall, Farrell's model treads very close to making the critical assumption of buffer models in order to account for the differences in free recall performance observed when singles versus pairs are studied. Namely, with a change from single-item study to, say, the study of pairs, Farrell's system creates different representations that reflect the change. However, without appealing to a mechanism of cognitive control, it is not obvious how the system changes in reaction to different input structures or why it does. Perhaps the Farrell model assumes that input itself is structured, and rather than the model creating the mnemonic structures that reflect the processing of the input, the model simply represents what is input. A strong prediction seems to be that subjects would always group pairs, even when task does not require memory for them.

In the present model, we assume that the buffer size reflects prior experience, the goals of the subject, task demands, and encoding conditions, all of which affect the organization of the long-term traces encoded. It should be quite obvious that these factors may change, introducing what would appear to be random influences on the organization of memory, but what separates our model from Farrell's model is our emphasis on understanding how these factors influence memory. What is a key to our model is the assumption that subjects actively rehearse items when they are needed to perform a task and compartmentalize the contents of memory when they are no longer needed (Lehman & Malmberg, 2011). Indeed, control processes seem critical for understanding a variety of influences on human memory ranging from levels of processing manipulations to intentional forgetting. For example, the items themselves are no more novel in forget conditions than in remember conditions, but we see differences in primacy across the two conditions. Additionally, the Lehman–Malmberg (2009) model was developed in a framework that has been used to account for over 40 years' worth of data, and it includes a complete model of recognition memory, in addition to free recall. Although others have hinted at the possibility of extending their models to recognition (cf. Farrell, 2012), most of those discussed here have not been applied to recognition. Thus, an important goal for model comparison is to extend the models to account for these other memory tasks.

Extending the Model

This model is used to generate average predictions for data that are collapsed over many subjects. In research on cognition, we typically look for group differences in the effects of our independent variables and fit the model to data that represent average performance. One of the short-term goals for this model is to begin to account for individual differences in cognitive processes. It was the analyses of the serial position patterns of high and low performers that motivated our account of performance in the other tasks described in this paper. For example, the finding that high and low performers initiated recall in different parts of the list suggested that subjects may use different cues to initiate recall depending on the types of delays involved in a task. In the simulations described, the tendency to initiate recall with a cue that corresponds to the beginning of the list was determined by a single parameter (γ_I), which could be manipulated in order to account for individual differences. Such manipulations would allow the model to account for data described by Unsworth et al. (2011) related to similar individual differences in serial position patterns.

Examinations of individual differences may also be informative in understanding why some individuals experience certain cognitive patterns and others do not (Widaman, 2008; see also McDowd & Hoffman, 2008). The current model may be extended to account for patterns of cognition seen in special populations and can allow us to evaluate theories about what leads to these patterns. Further, the model can be used to generate testable predictions related to such patterns. One area of research to which this model may be applied is in the study of cognitive deficits in depression.

There is a large body of research related to cognitive deficits in depression, much of which suggests that depressed individuals show impairments in various cognitive processes, including concentration, attention, and memory (American Psychiatric Association, 2000 [*DSM–IV–TR*]; Burt, Zembar, & Niederehe, 1995; Christopher & MacDonald, 2005; Cohen, Weingartner, Smallberg, Pickard, & Murphy, 1982; Kalska, Punamäki, Mäkinen-Pelli, & Saarinen, 1999), or that they have negative information-processing biases (Blaney, 1986; Bradley & Mathews, 1983; Matt, Vázquez, & Campbell, 1992).

Recent interest has developed in intentional forgetting processes in depression, driven by the suggestion that intentional forgetting is related to inhibitory processes and that inhibitory processes are impaired in individuals with depression (Johnson, 2007; Joormann, Yoon, & Zetsche, 2007; Power, Dalgleish, Claudio, Tata, & Kentish, 2000). If this is true, depressed individuals would be expected to show deficits on intentional forgetting tasks, particularly when negative materials, toward which depressed individuals may be negatively biased to attend, are involved.

Lehman and Malmberg (2009) proposed the context differentiation model for intentional forgetting; the model provides a formal account of the process commonly referred to as inhibition, which causes the effects of intentional forgetting. Lehman and Malmberg (2011) proposed that impairments seen on directed forgetting tasks in depressed populations may be the result of a failure to use context differentiation in order to compartmentalize information that is not relevant to the task at hand. This problem may also manifest in the symptoms of depression, such as ruminating on depressive thoughts when these thoughts are not presently useful.

The Lehman–Malmberg model can be utilized to make formal predictions about what will occur in various tasks for individuals who are not able to use context differentiation in order to complete the compartmentalization process. These predictions can then be tested in order to evaluate the viability of this theory. For example, we can generate predictions based on simulations using trials where contextual differentiation is used and compare these to those for trials where contextual differentiation is not possible and trials where contextual differentiation is forced (e.g., through the use of changes in environmental context). We can then evaluate these predictions by generating data from depressed subjects and comparing these data to our model's prediction (a task that is currently under way).

One future direction for this model is to take on these issues from a neural perspective (Cowan, 1995; Widaman, 2008). Dissociations in neuropsychology are often cited as evidence of separate short-term and long-term memory stores (Davelaar et al., 2005). However, Crowder (1989) suggests that one type of amnesia may be due to a specific type of coding deficit in which the relations of items to their temporal contexts are not properly coded. A model that is able to account for biological measures in addition to behavioral measures would be useful in evaluating the dual-store issue. Additionally, it would allow us to make a greater variety of predictions with the model. For example, depressed individuals show reduced activity in the prefrontal cortex (Henriques & Davidson, 1991), an area that is associated with various cognitive processes (Posner, 1992). Developing a comprehensive model of both brain and behavioral processes would allow us to further explore the applications of this model in special populations. One advantage of the Davelaar et al. (2005) model is that it incorporates features of the neural processes of excitation and inhibition in order to generate activation levels in memory (see also the TCM-A model, Sederberg et al., 2008, which too relates encoding to neural activity). Future work will attempt to accomplish similar goals in the framework of the Lehman–Malmberg model. Perhaps the assumptions of these models can be combined in a way that provides a model with the advantages of the Lehman–Malmberg model and models of neural processes in order to account for a much larger variety of data than can currently be handled by either model alone.

Conclusion

In sum, the Lehman–Malmberg model successfully accounts for a variety of data in multiple episodic memory paradigms. It has been shown to make accurate predictions related to both intentional and unintentional forgetting, maintenance rehearsal, continuous distractor tasks, and chunking operations. The key characteristics that allow the model to fit this wide array of data are the operations of the buffer as a process during both encoding and retrieval. At present, our findings suggest that control processes, which we conceptualize as a rehearsal buffer in the Atkinson and Shiffrin (1968) tradition, are a necessary component of models of episodic memory.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Anderson, J. R. (2000). *Learning and memory* (2nd ed.). New York, NY: Wiley.

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language, 38,* 341–380. doi:10.1006/jmla.1997.2553

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79,* 97–123. doi:10.1037/h0033773

Annis, J., Malmberg, K. J., Criss, A. H., & Shiffrin, R. M. (2012). *Temporal and task-context-dependent recognition testing*. Manuscript submitted for publication.

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* (pp. 89–195). New York, NY: Academic Press.

Baddeley, A. D., & Hitch, G. J. L. (1974). Working memory. In G. A. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York, NY: Academic Press.

Baddeley, A., & Wilson, B. (1988). Comprehension and working memory: A single case neuropsychological study. *Journal of Memory and Language, 27,* 479–498. doi:10.1016/0749-596X(88)90021-6

Bjork, R. A., LaBerge, D., & Legrand, R. (1968). The modification of short-term memory through instructions to forget. *Psychonomic Science, 10,* 55–56.

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology, 6,* 173–189. doi:10.1016/0010-0285(74)90009-7

Blaney, P. H. (1986). Affect and memory: A review. *Psychological Bulletin, 99,* 229–246. doi:10.1037/0033-2909.99.2.229

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology, 49,* 229–240. doi:10.1080/00221309.1953.9710088

Bousfield, W. A., & Puff, C. R. (1964). Clustering as a function of response dominance. *Journal of Experimental Psychology, 67,* 76–79. doi:10.1037/h0040012

Bradley, B., & Mathews, A. (1983). Negative self-schemata in clinical depression. *British Journal of Clinical Psychology, 22,* 173–181. doi:10.1111/j.2044-8260.1983.tb00598.x

Broadbent, D. E. (1958). *Perception and comunication*. New York, NY: Academic Press.

Burt, D. B., Zembar, M. J., & Niederehe, G. (1995). Depression and memory impairment: A meta-analysis of the association, its pattern, and specificity. *Psychological Bulletin, 117,* 285–305. doi:10.1037/0033-2909.117.2.285

Busemeyer, J. R., & Diederich, A. (2010). *Cognitive modeling*. Thousand Oaks, CA: Sage.

Christopher, G., & MacDonald, J. (2005). The impact of clinical depression on working memory. *Cognitive Neuropsychiatry, 10,* 379–399. doi:10.1080/13546800444000128

Cohen, R. M., Weingartner, H., Smallberg, S. A., Pickard, D., & Murphy, D. L. (1982). Effort and cognition in depression. *Archives of General Psychiatry, 39,* 593–597. doi:10.1001/archpsyc.1982.04290050061012

Cowan, N. (1995). *Attention and memory: An integrated framework*. New York, NY: Oxford University Press.

Cowan, N. (1997). The development of working memory. In N. Cowan (Ed.), *The development of memory in childhood* (pp. 163–199). Hove, England: Psychology Press.

Cowan, N., Wood, N. L., & Borne, D. N. (1994). Reconfirmation of the short-term storage concept. *Psychological Science, 5,* 103–106. doi:10.1111/j.1467-9280.1994.tb00639.x

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11,* 671–684. doi:10.1016/S0022-5371(72)80001-X

Craik, F. I., & Watkins, M. J. (1973). The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 12,* 599–607. doi:10.1016/S0022-5371(73)80039-8

Criss, A. H. (2006). The consequences of differentiation in episodic memory: Similarity and the strength based mirror effect. *Journal of Memory and Language, 55,* 461–478. doi:10.1016/j.jml.2006.08.003

Criss, A. H., & Malmberg, K. J. (2008). Evidence in support of the elevated attention hypothesis of recognition memory. *Journal of Memory and Language, 59,* 331–345. doi:10.1016/j.jml.2008.05.002

Criss, A. H., Malmberg, K. J., & Shiffrin, R. M. (2011). Output interference in recognition memory. *Journal of Memory and Language, 64,* 316–326. doi:10.1016/j.jml.2011.02.003

Criss, A. H., & Shiffrin, R. M. (2004). Context noise and item noise jointly determine recognition memory: A comment on Dennis & Humphreys (2001). *Psychological Review, 111,* 800–807. doi:10.1037/0033-295X.111.3.800

Criss, A. H., & Shiffrin, R. M. (2005). List discrimination and representation in associative recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 1199–1212. doi:10.1037/0278-7393.31.6.1199

Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.

Crowder, R. G. (1982). The demise of short-term memory. *Acta Psychologica, 50,* 291–323. doi:10.1016/0001-6918(82)90044-0

Crowder, R. G. (1989). Modularity and dissociations in memory systems. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 271–294). Hillsdale, NJ: Erlbaum.

Darley, C. F., & Glass, A. L. (1975). Effects of rehearsal and serial list position on recall. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 453–458. doi:10.1037/0278-7393.1.4.453

Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review, 112,* 3–42. doi:10.1037/0033-295X.112.1.3

Davelaar, E. J., Haarmann, H. J., Goshen-Gottstein, Y., & Usher, M. (2006). Semantic similarity dissociates short- from long-term recency effects: Testing a neurocomputational model of list memory. *Memory & Cognition, 34,* 323–334. doi:10.3758/BF03193410

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic recognition memory. *Psychological Review, 108,* 452–478. doi:10.1037/0033-295X.108.2.452

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology*. New York, NY: Teachers College, Columbia University.

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

Farrell, S. (2010). Dissociating conditional recency in immediate and delayed free recall: A challenge for unitary models of recency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 324–347. doi:10.1037/a0018042

Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review, 119,* 223–271. doi:10.1037/a0027371

Farrell, S., & Lewandowsky, S. (2008). Empirical and theoretical limits on lag recency in free recall. *Psychonomic Bulletin & Review, 15,* 1236–1250. doi:10.3758/PBR.15.6.1236

Francis, W. N., & Kucera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston, MA: Houghton Mifflin.

Gardiner, J. M., Thompson, C. P., & Maskarinec, A. S. (1974). Negative recency in initial free recall. *Journal of Experimental Psychology, 103,* 71–78. doi:10.1037/h0036829

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91,* 1–67. doi:10.1037/0033-295X.91.1.1

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition, 13,* 8–20. doi:10.3758/BF03198438

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior, 5,* 351–360. doi:10.1016/S0022-5371(66)80044-0

Glenberg, A. M. (1984). A retrieval account of the long-term modality effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 16–31. doi:10.1037/0278-7393.10.1.16

Glenberg, A., Smith, S. M., & Green, C. (1977). Type I rehearsal: Maintenance and more. *Journal of Verbal Learning and Verbal Behavior, 16,* 339–352. doi:10.1016/S0022-5371(77)80055-8

Glover, S., & Dixon, P. (2004). Likelihood ratios: A simple, intuitive statistic for empirical psychologists. *Psychonomic Bulletin & Review, 11,* 791–806. doi:10.3758/BF03196706

Godden, D., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and under water. *British Journal of Psychology, 66,* 325–331. doi:10.1111/j.2044-8295.1975.tb01468.x

Grossberg, S., & Pearson, L. R. (2008). Laminar cortical dynamics of cognitive and motor working memory, sequence learning and performance: Toward a unified theory of how the cerebral cortex works. *Psychological Review, 115,* 677–732. doi:10.1037/a0012618

Harbison, J. I., Dougherty, M. R., Davelaar, E. J., & Fayyad, B. (2009). On the lawfulness of the decision to terminate memory search. *Cognition, 111,* 416–421. doi:10.1016/j.cognition.2009.03.002

Healy, A. F., & McNamara, D. S. (1996). Verbal learning and memory: Does the modal model still work? *Annual Review of Psychology, 47,* 143–172. doi:10.1146/annurev.psych.47.1.143

Henriques, J. B., & Davidson, R. J. (1991). Left frontal hypoactivation in depression. *Journal of Abnormal Psychology, 100,* 535–545. doi:10.1037/0021-843X.100.4.535

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. doi:10.1037/0278-7393.25.4.923

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46,* 269–299. doi:10.1006/jmps.2001.1388

James, W. (1890). *The principles of psychology*. New York, NY: Holt.

Johnson, N. F. (1970). The role of chunking and organization in the process of recall. *The psychology of learning and motivation: Advances in research and theory, 4,* 171–247.

Johnson, S. L. (2007). Cognitive inhibition across psychopathologies. *Applied and Preventive Psychology, 12,* 97–98. doi:10.1016/j.appsy.2007.11.001

Joormann, J., Yoon, K. L., & Zetsche, U. (2007). Cognitive inhibition in depression. *Applied and Preventive Psychology, 12,* 128–139. doi:10.1016/j.appsy.2007.09.002

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition, 24,* 103–109. doi:10.3758/BF03197276

Kahana, M. J., Sederberg, P. B., & Howard, M. W. (2008). Putting short-term memory into context: Reply to Usher and colleagues. *Psychological Review, 115,* 1119–1126. doi:10.1037/a0013724

Kalska, H., Punamäki, R., Mäkinen-Pelli, T., & Saarinen, M. (1999). Memory and metamemory functioning among depressed patients. *Applied Neuropsychology, 6,* 96–107. doi:10.1207/s15324826an0602_5

Karpicke, J. D., & Roediger, H. L. (2008, February 15). The critical importance of retrieval for learning. *Science, 319,* 966–968. doi:10.1126/science.1152408

Klein, K. A., Shiffrin, R. M., & Criss, A. H. (2007). Putting context in context. In J. S. Nairne (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III* (pp. 171–189). New York, NY: Psychology Press.

Laming, D. (2010). Serial position curves in free recall. *Psychological Review, 117,* 93–133. doi:10.1037/a0017839

Lehman, M. (2011). [Multilist recognition memory]. Unpublished raw data.

Lehman, M., & Malmberg, K. J. (2009). A global theory of remembering and forgetting from multiple lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 970–988. doi:10.1037/a0015728

Lehman, M., & Malmberg, K. J. (2011). Overcoming the effects of intentional forgetting. *Memory & Cognition, 39,* 335–347. doi:10.3758/s13421-010-0025-4

Li, S. C., & Lewandowsky, S. (1995). Forward and backward recall: Different retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 837–847. doi:10.1037/0278-7393.21.4.837

Malmberg, K. J. (2008). Towards an understanding of individual differences in episodic memory: Modeling the dynamics of recognition memory. In A. Benjamin & B. Ross (Eds.), *The psychology of learning and motivation: Skill and strategy in memory use* (Vol. 48, pp. 313–349). London, England: Academic Press.

Malmberg, K. J. (2012). *Spacing effects in list discrimination*. Unpublished manuscript.

Malmberg, K. J., Criss, A. H., Gangwani, T. H., & Shiffrin, R. M. (2012). Overcoming the negative consequences of interference that results from recognition memory testing. *Psychological Science, 23,* 115–119. doi:10.1177/0956797611430692

Malmberg, K. J., Holden, J. E., & Shiffrin, R. M. (2004). Modeling the effects of repetitions, similarity, and normative word frequency on old–new recognition and judgments of frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 319–331. doi:10.1037/0278-7393.30.2.319

Malmberg, K., Lehman, M., & Sahakyan, L. (2006). On the cost and benefit of taking it out of context: Modeling the inhibition associated with directed forgetting. In *Proceedings of the 28th Meeting of the Cognitive Science Society* (pp. 549–554). Hillsdale, NJ: Erlbaum.

Malmberg, K. J., & Murnane, K. (2002). List composition and the word-frequency effect for recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 616–630. doi:10.1037/0278-7393.28.4.616

Malmberg, K. J., & Nelson, T. O. (2003). The word-frequency effect for recognition memory and the elevated-attention hypothesis. *Memory & Cognition, 31,* 35–43. doi:10.3758/BF03196080

Malmberg, K. J., & Shiffrin, R. M. (2005). The "one-shot" hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 322–336. doi:10.1037/0278-7393.31.2.322

Malmberg, K. J., Steyvers, M., Stephens, J. D., & Shiffrin, R. M. (2002). Feature-frequency effects in recognition memory. *Memory & Cognition, 30,* 607–613. doi:10.3758/BF03194962

Malmberg, K. J., & Xu, J. (2007). On the flexibility and on the fallibility of associative memory. *Memory & Cognition, 35,* 545–556. doi:10.3758/BF03193293

Malmberg, K. J., Zeelenberg, R., & Shiffrin, R. M. (2004). Turning up the noise or turning down the volume? On the nature of the impairment of episodic recognition memory by midazolam. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 540–549. doi:10.1037/0278-7393.30.2.540

Marshall, P. H., & Werder, P. R. (1972). The effects of the elimination of rehearsal on primacy and recency. *Journal of Verbal Learning and Verbal Behavior, 11,* 649–653. doi:10.1016/S0022-5371(72)80049-5

Matt, G. E., Vázquez, C., & Campbell, W. K. (1992). Mood-congruent recall of affectively toned stimuli: A meta-analytic review. *Clinical Psychology Review, 12,* 227–255. doi:10.1016/0272-7358(92)90116-P

McDowd, J. M., & Hoffman, L. (2008). Challenges in attention: Measures, methods, and applications. In S. M. Hofer & D. F. Alwin (Eds.), *Handbook of cognitive aging: Interdisciplinary perspectives* (pp. 122–133). Los Angeles, CA: Sage.

Mensink, G. J. M., & Raaijmakers, J. G. W. (1989). A model for contextual fluctuation. *Journal of Mathematical Psychology, 33,* 172–186. doi:10.1016/0022-2496(89)90029-1

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97. doi:10.1037/h0043158

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64,* 482–488. doi:10.1037/h0045106

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review, 104,* 839–862. doi:10.1037/0033-295X.104.4.839

Murdock, B. B., & Anderson, R. E. (1975). Encoding, storage and retrieval of item information. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 145–194). Hillsdale, NJ: Erlbaum.

Murnane, K., & Phelps, M. P. (1995). Effects of changes in relative cue strength on context-dependent recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 158–172. doi:10.1037/0278-7393.21.1.158

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18,* 251–269. doi:10.3758/BF03213879

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (1998). *The University of South Florida word association, rhyme, and word fragment norms*. Retrieved from http://www.usf.edu/FreeAssociation/

Nelson, T. O. (1977). Repetition and depth of processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 151–171. doi:10.1016/S0022-5371(77)80044-3

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review, 116,* 129–156. doi:10.1037/a0014420

Posner, M. I. (1992). Attention as a cognitive and neural system. *Current Directions in Psychological Science, 1,* 11–14. doi:10.1111/1467-8721.ep10767759

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13,* 25–42. doi:10.1146/annurev.ne.13.030190.000325

Postman, L., & Phillips, L. W. (1964). The effects of variable contexts on the acquisition and retention of paired associates. *American Journal of Psychology, 77,* 64–74. doi:10.2307/1419272

Power, M. J., Dalgleish, T., Claudio, V., Tata, P., & Kentish, J. (2000). The directed forgetting task: Application to emotionally valent material. *Journal of Affective Disorders, 57,* 147–157. doi:10.1016/S0165-0327(99)00084-1

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search in 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, NY: Academic Press.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134. doi:10.1037/0033-295X.88.2.93

Robinson, E. S., & Brown, M. A. (1926). Effect of serial position on memorization. *American Journal of Psychology, 37,* 538–552. doi:10.2307/1414914

Roediger, H. L. (1974). Inhibiting effects of recall. *Memory & Cognition, 2,* 261–269. doi:10.3758/BF03208993

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology, 89,* 63–77. doi:10.1037/h0031185

Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 1064–1072. doi:10.1037/0278-7393.28.6.1064

Sederberg, P. B., Gauthier, L. V., Terushkin, V., Miller, J. F., Barnathan, J. A., & Kahana, M. J. (2006). Oscillatory correlates of the primacy effect in episodic memory. *NeuroImage, 32,* 1422–1431. doi:10.1016/j.neuroimage.2006.04.223

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review, 115,* 893–912. doi:10.1037/a0013396

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal* *of Experimental Psychology, 22,* 261–273. doi:10.1080/00335557043000203

Shiffrin, R. M. (1970). Forgetting, trace erosion or retrieval failure? *Science, 168,* 1601–1603. doi:10.1126/science.168.3939.1601

Shiffrin, R. M. (1975). The locus and role of attention. In P. M. A. Rabbit & S. Dornic (Eds.), *Attention and performance V* (pp. 168–193). New York, NY: Academic Press.

Shiffrin, R. M., & Atkinson, R. C. (1969). Storage and retrieval processes in long-term memory. *Psychological Review, 76,* 179–193. doi:10.1037/h0027277

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—Retrieving effectively from memory. *Psychonomic Bulletin & Review, 4,* 145–166. doi:10.3758/BF03209391

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 592–604. doi:10.1037/0278-7393.4.6.592

Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 460–471. doi:10.1037/0278-7393.5.5.460

Tan, L., & Ward, G. (2000). A recency-based account of primacy effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1589–1625. doi:10.1037/0278-7393.26.6.1589

Tulving, E. (2008). On the law of primacy. In M. A. Gluck, J. R. Anderson, & S. M. Kosslyn (Eds.), *Memory and mind: A festschrift for Gordon H. Bower* (pp. 31–48). Mahwah, NJ: Erlbaum.

Tulving, E., & Arbuckle, T. Y. (1966). Input and output interference in short-term associative memory. *Journal of Experimental Psychology, 72,* 145–150. doi:10.1037/h0023344

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5,* 381–391. doi:10.1016/S0022-5371(66)80048-8

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2011). Inter- and intra-individual variation in immediate free recall: An examination of serial position functions and recall initiation strategies. *Memory, 19,* 67–82. doi:10.1080/09658211.2010.535658

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114,* 104–132. doi:10.1037/0033-295X.114.1.104

Usher, M., Davelaar, E. J., Haarmann, H. J., & Goshen-Gottstein, Y. (2008). Short-term memory after all: Comment on Sederberg, Howard, and Kahana (2008). *Psychological Review, 115,* 1108–1118. doi:10.1037/a0013725

Ward, G., & Tan, L. (2004). The effect of the length of to-be-remembered lists and intervening lists on free recall: A reexamination using overt rehearsal. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 1196–1210. doi:10.1037/0278-7393.30.6.1196

Ward, G., Tan, L., & Grenfell-Essam, R. (2010). Examining the relationship between free recall and immediate serial recall: The effects of list length and output order. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1207–1241. doi:10.1037/a0020122

Welch, G. B., & Burnett, C. T. (1924). Is primacy a factor in association-formation? *American Journal of Psychology, 35,* 396–401. doi:10.2307/1414018

Whitten, W. B. (1978). Output interference and long-term serial position effects. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 685–692. doi:10.1037/0278-7393.4.6.685

Widaman, K. F. (2008). Integrative perspectives on cognitive aging: Measurement and modeling with mixtures of psychological and biological variables. In S. M. Hofer & D. F. Alwin (Eds.), *Handbook of cognitive aging: Interdisciplinary perspectives* (pp. 50–68). Los Angeles, CA: Sage.

Wixted, J. T., & McDowell, J. J. (1989). Contributions to the functional analysis of single-trial free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 685–697. doi:10.1037/0278-7393.15.4.685

Xu, J., & Malmberg, K. J. (2007). Modeling the effects of verbal and nonverbal pair strength on associative recognition. *Memory & Cognition, 35,* 526–544. doi:10.3758/BF03193292

(*Appendix follows*)

Appendix

Methods and Results for Experiments 1 Through 5

Experiment 1: Recognition Memory Serial Position

Method

Subjects, materials, and procedure. Subjects were 41 undergraduate psychology students at the University of South Florida who participated in exchange for course credit. For each subject, eight word lists were created, each consisting of 30 randomly related concrete nouns (between 20 and 50 occurrences per million; Francis & Kucera, 1982). The entire experiment was presented on a computer in an individual subject booth. At the beginning of the experiment, subjects were told that they would be studying multiple lists of words. The words appeared on the screen one at a time for 1 s. Immediately after each list was presented, a two-alternative forced-choice test was given for that list. The test consisted of all of the words from the study list in addition to 30 foils selected from the same pool. Subjects were shown a target word on one side of the screen (left or right, chosen randomly) and a foil on the other side of the screen, and they were asked to indicate which word was studied. After the test, they were given their percentage score for the list and told to try to improve their score for the next list. They then completed a 30-s math task before beginning the next list.

Results

According to a one-way within-subjects analysis of variance (ANOVA), there was a significant effect of serial position, $F(29, 1160) = 1.75$, $MSE = .02$, $p = .01$. The primacy effect is very small, and there is no hint of recency (the hit rate for the first item did not differ significantly from the hit rate for the second item on the list, $p > .05$, but was significantly higher than the hit rate for the item in the middle of the list [serial position 15], $t(40) = 2.76$, $SD = 0.19$, $p = .01$; the hit rate for the last item did not differ from that of any other items on the list, $ps > .05$).

Experiment 2: Immediate Free Recall Individual Differences

Method

Subjects, materials, and procedure. Subjects were 32 undergraduate psychology students at the University of South Florida who participated in exchange for course credit. For each subject, eight word lists were created, each consisting of 30 randomly related concrete nouns (between 20 and 50 occurrences per million; Francis & Kucera, 1982). At the beginning of the experiment, subjects were told that they would be studying multiple lists of words on which they would later be tested. Words appeared on the screen one at a time for 1 s. After the study list was presented, an immediate free recall test was given. Sixty seconds were allowed for recall. After each test, subjects were given their percentage score for the list and told to try to improve their score for the next list. For the analyses, subjects were divided according to the total number of words recalled in the experiment. The eight subjects who made up the top 25% were selected as "high performers," and the eight subjects who made up the bottom 25% were selected as "low performers."

Results

According to a one-way between-subjects ANOVA, high performers recalled significantly more words than low performers, $F(1, 14) = 9.30$, $MSE = .002$, $p = .001$. According to a two-way mixed ANOVA, high performers were significantly more likely to initiate recall at the end of the list than low performers, $F(29, 406) = 1.43$, $MSE = .03$, $p = .01$. Further, high performers exhibited greater recency effects than low performers, as measured by the probability of recalling the last item on the list, $t(14) = 2.07$, $SD = 0.20$, $p = .03$.

Experiment 3: Immediate and Delayed Continuous Distractor Free Recall

Method

Subjects and materials. Subjects were 86 undergraduate psychology students at the University of South Florida who participated in exchange for course credit. For each subject, eight word lists were created, each consisting of 20 randomly related concrete nouns (between 20 and 50 occurrences per million; Francis & Kucera, 1982). Additionally, four lists of rhyme words were created. Each rhyme list consisted of 20 monosyllabic words with a rhyme-set size of at least 12 (Nelson, McEvoy, & Schreiber, 1998). The experiment was presented on a computer in an individual subject booth, and the rhyme lists were printed in paper booklets.

Procedure. The four conditions were manipulated within subjects in blocks, where two lists were presented in each condition. The order of the conditions was counterbalanced. At the beginning of the experiment, subjects were told that they would be studying multiple lists of words, and the instructions for each list would appear before the list. For all conditions, words appeared on the screen one at a time for 1 s. For all conditions, 60 s were allowed for recall. After each test, subjects were given their percentage score for the list and told to try to improve their score for the next list.

In the *immediate control* condition, words appeared on the screen one at a time with a .5-s interstimulus interval (ISI). Immediately after the list was presented, a free recall test was given for that list, in which subjects were instructed to enter all of the words they remembered from that list onto the screen.

(*Appendix continues*)

In the *delayed control* condition, words appeared on the screen one at a time with a .5-s ISI. After the list was presented, subjects completed a 5-min distractor task. During the task, they watched a 4.5-min "How It's Made" video, with a 30-s quiz afterward (to assure that they were attending to the video). After the distractor task, they completed the same free recall task as described in the immediate control condition.

In the *immediate continuous distractor* condition, subjects were to alternate between memorizing a word on the screen and writing rhymes for a different word in the printed booklet. Before beginning the continuous distractor condition, subjects read explicit instructions detailing the procedure, followed by a quiz to be sure they understood the procedure. To encourage participation, they were told that they needed to reach a certain number of rhyming words in order to complete the experiment. They then saw a demonstration and completed a two-word practice list, after which the experimenter checked to be sure that they were attempting to memorize only the words on the screen and write rhymes for only the words in the booklet during the practice trial. Once they correctly completed the task, they began the study list. Words appeared on the screen one at a time, with a 10-s delay after each word. During this delay, the ***** symbol appeared on the screen, alerting subjects that they should now turn to their rhyme booklets and begin creating rhymes for the next word in the booklet. After 10 s, a tone alerted them to look back at the screen for the next word. This repeated throughout the list, so that after each studied word, they had to provide rhyming words for a word in the booklet (including after the last item on the list). Subjects then completed the same free recall task as described in the immediate control condition.

In the *delayed continuous distractor* condition, the procedure was the same as in the immediate continuous distractor condition, except that subjects completed the same 5-min distractor task after study as in the delayed control condition, followed by the same free recall task described above.

Results

In the control conditions, recall was significantly greater in the immediate condition than in the delayed condition, $t(85) = 11.445$, $SD = 0.12$, $p < .001$. Serial position analyses revealed a serial position by condition interaction, $F(28, 2380) = 5.06$, $MSE = .194$, $p < .001$. As shown in Figure 5, both primacy and recency are present in the immediate testing condition. In the delayed condition, the recency effect was eliminated. All differences were significant at $\alpha = .05$. A significant serial position by condition interaction is also present in first recall probabilities, $F(28, 2380) = 8.15$, $MSE = .046$, $p < .001$. Figure 5 shows that in the immediate condition, subjects were most likely to initiate recall with the last item on the list, whereas in the delayed condition, they were more likely to begin recall at the beginning of the list. For conditional recall probabilities, there was not a significant lag by condition interaction, $F(37, 3145) = 1.59$, $MSE = .024$, $p = .11$, as shown in Figure 5. Because of the marginal p value, planned comparisons were conducted, revealing no significant differences between the two conditions at any lag.

In the continuous distractor conditions, recall was significantly greater in the immediate condition than in the delayed condition, $t(85) = 4.604$, $SD = 0.12$, $p < .001$. Serial position analyses revealed no significant serial position by condition interaction, $F(28, 2380) = 1.45$, $MSE = .145$ $p = .10$. Although the interaction was not significant, planned comparisons revealed differences in recency (the last item on the list.), but no other differences were significant, as shown in Figure 5. A significant serial position by condition interaction is also present in first recall probabilities, $F(28, 2380) = 1.68$, $MSE = .046$, $p = .03$. Again, planned comparisons revealed differences in first recall probability for the last item on the list but for no other serial positions. For conditional recall probabilities, there was not a significant lag by condition interaction, $F(37, 3145) = 1.76$, $MSE = .025$, $p = .07$, as shown in Figure 5. Because of the marginal p value, planned comparisons were conducted, revealing no significant differences between the two conditions at any lag.

Experiment 4: Single- and Paired-Item Study Lists

Method

Subjects and materials. Subjects were 39 undergraduate psychology students at the University of South Florida who participated in exchange for course credit. For each subject, eight word lists were created, each consisting of 30 randomly related concrete nouns (between 20 and 50 occurrences per million; Francis & Kucera, 1982). The entire experiment was presented on a computer in an individual subject booth.

Procedure. At the beginning of the experiment, subjects were told that they would be studying multiple lists of words. All subjects studied four lists of single words in one block and four lists of paired words in another block. The blocks were counterbalanced, with the instructions given at the beginning of the block. Instructions for the single lists informed subjects that they would be shown the words one at a time, and they should try to create a sentence in order to memorize each word. Instructions for the paired lists informed subjects that they would be seeing pairs of words, and they should try to create a sentence in order to memorize both words. These instructions were chosen in order to try to standardize strategies used by subjects. To be sure that the results from this study were not an artifact of the instructions, we conducted another experiment utilizing the same procedures but with no instructions given to subjects regarding how to memorize the words. No differences were present in the results of the two experiments.

For the single-word lists, the words appeared on the screen one at a time. For half of the lists, the words remained on the screen for 1 s with an ISI of 375 ms. For the other half of the lists, the words remained on the screen for 875 ms with an ISI of 500 ms. This was done so that half of the lists would have a study time equal to that of the words in the paired lists, and the other half would have a total study list time equal to that of the words in the paired lists. As there were no differences in the results between these different study times, these data were collapsed across study times and were not further analyzed.

(*Appendix continues*)

For the paired-word lists, the words appeared on the screen in a staggered fashion in order to maintain a temporal order to the words. The first word in a pair appeared on the screen and remained. After 250 ms, the second word of the pair appeared on the screen adjacent to the first word. After 1.75 s, the first word disappeared from the screen, so that the second word from the pair remained alone on the screen. After 250 ms, the second word also disappeared from the screen. After an ISI of 500 ms, the process continued for the next pair. This staggered presentation was used so that the words would appear in pairs but would maintain a temporal order like that of the single-word lists.

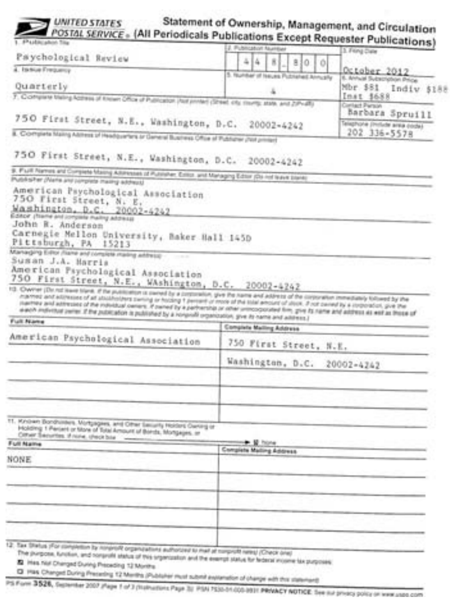
Immediately after each list was presented, a free recall test was given for that list, in which subjects were instructed to enter all of the words they remembered from that list onto the screen. They were given 60 s to do this. After the test, they were given their percentage score for the list and told to try to improve their score for the next list. They then completed a 30-s math task, which required addition of pairs of two-digit numbers, before beginning the next list.

Experiment 5: Single- and Paired-Item Study Lists With List-Length Manipulations

Method

Subjects, materials, and procedure. Subjects were 176 undergraduate psychology students at the University of South Florida who participated in exchange for course credit (as all variables were manipulated within subject, it was necessary to collect data from many subjects in order to eliminate noise and see clear serial position effects). For each subject, two word lists of each list length (2, 4, 6, 8, 10, 12, 16, 20, 24, and 30 items) were created for each study condition (single-item lists and paired-item lists), each consisting of randomly related concrete nouns (between 20 and 50 occurrences per million; Francis & Kucera, 1982). Thus, for each subject, a total of 40 word lists was created, half for the single-item study list condition and half for the paired-item study list condition. Item presentation occurred in the same manner described in Experiment 4 for single and paired lists, with similar study times. In order to reduce the effects of fatigue, we ran the experiment in two sessions, a week apart, so that subjects completed only 20 study–test cycles in each session. In one session, subjects completed all study–test cycles for paired-item lists, and in the other they completed all study–test cycles for single-item lists (with the order of single and paired conditions counterbalanced). In each study–test cycle, the lists were randomly presented, such that the length of each new list was not predictable, and subjects were not told in advance the length of each list.

Received September 19, 2011
Revision received September 17, 2012
Accepted October 1, 2012 ■



UNITED STATES POSTAL SERVICE® Statement of Ownership, Management, and Circulation (All Periodicals Publications Except Requester Publications)

1. Publication Title: Psychological Review
2. Publication Number: 4 4 8 - 8 0 0
3. Filing Date: October 2012
4. Issue Frequency: Quarterly
5. Number of Issues Published Annually: 4
6. Annual Subscription Price: Mbr $81 Indiv $188 Inst $688
7. Complete Mailing Address of Known Office of Publication (Not printer): 750 First Street, N.E., Washington, D.C. 20002-4242
Contact Person: Barbara Spruill
Telephone (Include area code): 202 336-5578
8. Complete Mailing Address of Headquarters or General Business Office of Publisher (Not printer): 750 First Street, N.E., Washington, D.C. 20002-4242
9. Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor (Do not leave blank)
Publisher (Name and complete mailing address): American Psychological Association, 750 First Street, N. E., Washington, D.C. 20002-4242
Editor (Name and complete mailing address): John R. Anderson, Carnegie Mellon University, Baker Hall 145D, Pittsburgh, PA 15213
Managing Editor (Name and complete mailing address): Susan J.A. Harris, American Psychological Association, 750 First Street, N.E., WAshington, D.C. 20002-4242
10. Owner

Full Name	Complete Mailing Address
American Psychological Association	750 First Street, N.E.
	Washington, D.C. 20002-4242

11. Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities. If none, check box ☒ None

Full Name	Complete Mailing Address
NONE	

12. Tax Status (For completion by nonprofit organizations authorized to mail at nonprofit rates) (Check one)
The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes:
☒ Has Not Changed During Preceding 12 Months
☐ Has Changed During Preceding 12 Months (Publisher must submit explanation of change with this statement)

PS Form 3526, September 2007 (Page 1 of 3) PRIVACY NOTICE: See our privacy policy on www.usps.com

13. Publication Title: Psychological Review
14. Issue Date for Circulation Data Below: July 2012

15. Extent and Nature of Circulation		Average No. Copies Each Issue During Preceding 12 Months	No. Copies of Single Issue Published Nearest to Filing Date
a. Total Number of Copies (Net press run)		1700	1550
b. Paid Circulation (By Mail and Outside the Mail)	(1) Mailed Outside-County Paid Subscriptions Stated on PS Form 3541 (Include paid distribution above nominal rate, advertiser's proof copies, and exchange copies)	1030	925
	(2) Mailed In-County Paid Subscriptions Stated on PS Form 3541 (Include paid distribution above nominal rate, advertiser's proof copies, and exchange copies)		
	(3) Paid Distribution Outside the Mails Including Sales Through Dealers and Carriers, Street Vendors, Counter Sales, and Other Paid Distribution Outside USPS®	388	385
	(4) Paid Distribution by Other Classes of Mail Through the USPS (e.g. First-Class Mail®)		
c. Total Paid Distribution (Sum of 15b (1), (2), (3), and (4))		1418	1310
d. Free or Nominal Rate Distribution (By Mail and Outside the Mail)	(1) Free or Nominal Rate Outside-County Copies included on PS Form 3541		
	(2) Free or Nominal Rate In-County Copies Included on PS Form 3541		
	(3) Free or Nominal Rate Copies Mailed at Other Classes Through the USPS (e.g. First-Class Mail)		
	(4) Free or Nominal Rate Distribution Outside the Mail (Carriers or other means)	16	16
e. Total Free or Nominal Rate Distribution (Sum of 15d (1), (2), (3) and (4))		16	16
f. Total Distribution (Sum of 15c and 15e)		1434	1326
g. Copies not Distributed (See Instructions to Publishers #4 (page #3))		266	224
h. Total (Sum of 15f and g)		1700	1550
i. Percent Paid (15c divided by 15f times 100)		99%	99%

16. Publication of Statement of Ownership
☒ If the publication is a general publication, publication of this statement is required. Will be printed in the January 2013 issue of this publication.
☐ Publication not required.

17. Signature and Title of Editor, Publisher, Business Manager, or Owner: Barbara Spruill, Director, Service Center Operations
Date: 10/3/2012

I certify that all information furnished on this form is true and complete. I understand that anyone who furnishes false or misleading information on this form or who omits material or information requested on the form may be subject to criminal sanctions (including fines and imprisonment) and/or civil sanctions (including civil penalties).

PS Form 3526, September 2007 (Page 2 of 3)