Psychological Review
2005, Vol. 112, No. 1, 3–42

0033-295X/05/$12.00 DOI: 10.1037/0033-295X.112.1.3

The Demise of Short-Term Memory Revisited: Empirical and Computational Investigations of Recency Effects

Eddy J. Davelaar
Birkbeck College

Yonatan Goshen-Gottstein and Amir Ashkenazi
Tel-Aviv University

Henk J. Haarmann
University of Maryland (College Park)

Marius Usher
Birkbeck College

In the single-store model of memory, the enhanced recall for the last items in a free-recall task (i.e., the recency effect) is understood to reflect a general property of memory rather than a separate short-term store. This interpretation is supported by the finding of a long-term recency effect under conditions that eliminate the contribution from the short-term store. In this article, evidence is reviewed showing that recency effects in the short and long terms have different properties, and it is suggested that 2 memory components are needed to account for the recency effects: an episodic contextual system with changing context and an activation-based short-term memory buffer that drives the encoding of item–context associations. A neurocomputational model based on these 2 components is shown to account for previously observed dissociations and to make novel predictions, which are confirmed in a set of experiments.

In recent years, the memory literature has seen an increased interest in theoretical accounts of serial-position effects in list memory (J. R. Anderson, Bothell, Lebiere, & Matessa, 1998; Haarmann & Usher, 2001; Howard & Kahana, 1999, 2002; Nairne, Neath, Serra, & Byun, 1997; Tan & Ward, 2000; Ward, 2002). This body of work, partially motivated by the controversy over the need to assume a short-term buffer in accounting for data in list memory, has centered on two versions of the free-recall paradigm: the immediate and the continuous-distractor free recall task (also known as the through-list distractor procedure).

In immediate free recall, participants are presented with a sequence of items and, after presentation of the final item, are required to report all items in any order. Compared with middle list items, the final few (or *recency*) items are reported with a higher probability. This finding has been called the *recency effect* (which in this article is referred to as *short-term recency*). The original explanation of the short-term recency effect was that at the start of the recall phase, the final few items reside in a capacity-limited short-term buffer, from which the items can be reported immediately. Subsequent items are then subject to a slower, probabilistic retrieval process from a long-term store (Atkinson & Shiffrin, 1968; Glanzer, 1972; Waugh & Norman, 1965).

This dual-store approach to human memory gained much support as several manipulations were identified that differentially affected recall performance for recency and prerecency (middle list) items (for a review, see Glanzer, 1972; for a critical discussion, see Wickelgren, 1973). For example, short-term recency was eliminated when participants were engaged in a distractor task (e.g., counting backward) after list presentation for as little as 15 s (e.g., Glanzer & Cunitz, 1966) or when participants were instructed to start the recall with items from the beginning of the list (Dalezman, 1976). The recall probability of prerecency items was not affected by these manipulations but was negatively affected by increasing the list length or presentation rate (Glanzer & Cunitz, 1966; Murdock, 1962; Raymond, 1969) and was prone to being affected by proactive interference (Craik & Birtwistle, 1971) and brain damage in amnesia (Baddeley & Warrington, 1970). These latter manipulations did not affect recency items.

Eddy J. Davelaar and Marius Usher, School of Psychology, Birkbeck College, University of London, London; Yonatan Goshen-Gottstein and Amir Ashkenazi, Department of Psychology, Tel-Aviv University, Ramat Aviv, Israel; Henk J. Haarmann, Center for the Advanced Study of Language, University of Maryland (College Park).

Parts of this work have been presented at the 41st (November 2000, New Orleans, LA) and 43rd (November 2002, Kansas City, KS) Annual Meetings of the Psychonomic Society, the 19th Annual Meeting of the Cognitive Section of the British Psychological Society (September 2003, Reading, England), the 5th International Conference on Cognitive Modelling (April 2003, Bamberg, Germany), and the 25th Annual Meeting of the Cognitive Science Society (July 2003, Boston). Eddy J. Davelaar is supported by the Economic and Social Research Council (Grant T026271312), and Yonatan Goshen-Gottstein is supported by the Israel Science Foundation (Grant 894-01).

We thank Mark Chappell for a critical reading of the manuscript. We also thank Amit Shahar for help in running Experiment 1. We are especially indebted to Michael J. Kahana, whose stimulating discussions inspired us to pursue this line of research.

Correspondence concerning this article should be addressed to Eddy J. Davelaar or Marius Usher, School of Psychology, Birkbeck College, University of London, Malet Street, London WC1E 7HX, England, or to Yonatan Goshen-Gottstein, Department of Psychology, Tel-Aviv University, Ramat Aviv 69978, Israel. E-mail: e.davelaar@bbk.ac.uk, m.usher@bbk.ac.uk, or goshen@freud.tau.ac.il

Later on, however, the dual-store approach to short-term recency was challenged when recency was obtained in situations in which the last items or events in memory should have been eliminated from the short-term buffer, both in real-life situations (Baddeley & Hitch, 1977; da Costa Pinto & Baddeley, 1991; Hitch & Ferguson, 1991; Sehulster, 1989) and in the experimental laboratory. In particular, long-term recency effects were observed in continuous-distractor free recall (Bjork & Whitten, 1974, Tzeng, 1973), which is identical to immediate free recall with the exception of including a distractor task before and after every item in the list. According to the dual-store approach, the distractor task following list presentation displaces the last list items, which are presumed to reside in the short-term buffer. Hence, a recency effect is not expected to be found. Given that a recency effect is found in continuous-distractor free recall, it follows that this *long-term recency effect* has a different source, most probably related to mechanisms of retrieval from the long-term store (but for a dual-store interpretation of the long-term recency effect, see Koppenaal & Glanzer, 1990; for a rebuttal of such an interpretation, see Neath, 1993a; Thapar & Greene, 1993).

The now-standard account of long-term recency is based on encoding and retrieval processes within a single memory store. This position can be understood to account for recency by assuming that the recall probability of an item is a function of the (global or local) distinctiveness of that item along a temporal (or perhaps positional) dimension (Crowder, 1976; Glenberg & Swanson, 1986; Nairne et al., 1997; Neath, 1993b). Operationally, the discriminability of an item can be defined as a function of the ratio between the temporal distance between two items (the interpresentation interval; IPI) and the temporal distance between the final item and the recall phase (the retention interval; RI). Indeed, Glenberg and colleagues (Glenberg, Bradley, Kraus, & Renzaglia, 1983) have shown that the logarithm of the ratio between IPI and RI predicts the slope of the best-fitting linear function over the last three serial positions (but see Nairne et al., 1997, who found with a constant IPI:RI ratio that the slope decreases with increases in distractor interval).

A more mechanistic account of both short-term and long-term recency effects within the single-store framework is based on the assumption that during the encoding phase, the episodic context changes and gets associated with currently presented items. At retrieval, the recall probability of an item is a function of the similarity between the test context and the context that was associated with that item during study (Dennis & Humphreys, 2001; Glenberg et al., 1980, 1983; Howard & Kahana, 1999, 2002; Mensink & Raaijmakers, 1988; see also Estes, 1955, 1997; Murdock, 1972). Recently, much progress has been obtained within the framework of models that show how contextual retrieval can mediate list memory (Howard & Kahana, 1999, 2002), providing a detailed account of both serial-position functions and contiguity effects (i.e., effects of *lag recency*, a measure based on the conditional probability for successive outputs; see the next section) in both immediate and continuous-distractor free recall. Remarkably, as these single-store theories provide a unifying account for the recency effects in both immediate and continuous-distractor free recall and their absence in delayed free recall (here and elsewhere, *delayed free recall* refers to recall of items following a distractor task; e.g., Glanzer & Cunitz, 1966), some theorists have gone so far as to proclaim the "demise of short-term memory" (Crowder, 1982; but for criticism, see Healy & McNamara, 1996; Raaijmakers, 1993).

In principle, the dual-store account of recency in immediate free recall is not inconsistent with a contextual retrieval account of long-term recency, as the two effects could be the products of different mechanisms. Nevertheless, as advocated by Crowder (1993), the principle of parsimony may require that "the burden of evidence should be with those who say these two, similar recency effects are caused by different mechanisms" (p. 143). In this article, we attempt to meet this challenge.

Because of the commitment to a single mechanism, single-store models tend to predict that experimental manipulations have equivalent effects on short- and long-term recency. Indeed, it has even been suggested that associations exist between the two tasks under the manipulation of variables such as semantic similarity, word frequency, and list length (Greene, 1986a; Greene & Crowder, 1984). Although these associations support the unitary view, there are at least four reasons to interpret them with caution. First, it is not at all clear that all dual-store models must predict a dissociation on all these variables (see Simulation 2).[1] Second, these associations constitute null effects, which are difficult to prove. Third, the variables have never been manipulated within a single study that examined both tasks. Therefore, differences in methodologies (design, material, or procedure) may have introduced confounds that masked possible dissociations. Fourth, contrary to Greene and Crowder's (Crowder, 1993; Greene, 1986a; Greene & Crowder, 1984) claims, over the years a number of dissociations between short- and long-term recency effects have been uncovered, as we describe below.

Although it is true that both immediate and continuous-distractor free recall reveal recency effects, in many studies the level of recall for the last few items is larger in immediate than in continuous-distractor free recall with a constant IPI:RI ratio (e.g., Howard & Kahana, 1999; Nairne et al., 1997; Poltrock & MacLeod, 1977). Furthermore, unlike long-term recency effects, short-term recency effects are sensitive to output order (Dalezman, 1976; Whitten, 1978). Moreover, short-term effects alone are insensitive to damage to the medial–temporal lobe (Carlesimo, Marfia, Loasses, & Caltagirone, 1996; see also the next section). Yet another dissociation is found when a final free recall task is required after a series of study lists. In immediate recall, a negative recency effect (lower recall for recency items compared with prerecency items) is found (Craik, 1970). Such an effect is absent in continuous-distractor free recall (e.g., Bjork & Whitten, 1974). Finally, even though recently developed contextual retrieval theories account for contiguity effects (measured by conditional response probabilities for successive recalls; Kahana, 1996), these

---

[1] The idea that lexical and semantic variables affect only the long-term memory component is a conclusion of a specific interpretation of the original dual-store model (Atkinson & Shiffrin, 1968, 1971), which viewed the short-term buffer as purely phonological and thus unaffected by lexical–semantic variables. The more general dual-store model ("the modal model"; Murdock, 1967) was not committed to this assumption, the buffer being seen as central to conscious thoughts and thus necessarily having a lexical–semantic content. More recently, the existence of lexical–semantic content within the buffer has been explicitly suggested in neuropsychological studies (R. C. Martin et al., 1994; Romani & Martin, 1999; see also Haarmann & Usher, 2001, and the General Discussion).

theories are silent to the observation that these effects differ between immediate and continuous-distractor free recall for the first few output positions (see the next section; see also Howard & Kahana, 1999; Kahana, 1996).

In this article, we argue that now, after a number of powerful single-store theories have taken into account the objections against dual-store models (see Crowder, 1982; Greene, 1986b; Howard & Kahana, 1999) and have highlighted the important contribution of contextual retrieval in list memory (Glenberg & Swanson, 1986; Howard & Kahana, 1999, 2002), it may be time to explore whether a theory that combines a contextual retrieval component with an additional short-term store component might provide an even more comprehensive account of serial-position effects in list memory. In particular, such a combined *context–activation theory* might be able to explain the dissociations that have been observed between short- and long-term recency effects.

A combined theory including two components (i.e., a short-term buffer and a changing episodic context) has previously been developed by Mensink and Raaijmakers (1988) but has not been used to account for data in the continuous-distractor task. In this article, we present a computational context–activation model based on similar components to account for serial-position effects in list memory, focusing on the dissociations between short- and long-term recency as well as on lag-recency effects (which are described in the next section). More important, the model predicts a novel dissociation between short- and long-term recency, on the basis of a manipulation of proactive interference.

Although similar to the Mensink and Raaijmakers (1988) model, our model implements the short-term buffer in terms of activation levels rather than through the use of a box metaphor with a fixed number of slots. In this model, items are removed from (being deactivated in) the buffer because of mutual inhibition with newly entered items. This implementation allowed us to address the internal dynamics of the buffer, thereby leading to a second prediction involving a shift from recency to primacy with an increase in the presentation rate. The predictions of a dissociation on the basis of proactive interference and of a shift from recency to primacy were tested in two experiments.

The model we present here applies to a number of tasks that measure item information in list memory. Among them are primarily the free-recall task and its related versions (immediate, delayed, and continuous-distractor recall), all of which have played a central role in earlier investigations. In addition, as we show, the model applies to the cued-recall paradigm (Waugh, 1970; Waugh & Norman, 1965). Although we did not attempt here to model in detail control processes and semantic effects, we indicate how the model can address such processes (see the Utility of the Dynamic Buffer section). In the General Discussion, we examine how the model can be extended beyond accuracy data to address response latencies. At the outset we note that our model does not address serial-order recall, which we see as involving additional processes that are not part of the current model, such as rehearsal and phonological encoding.

The remainder of this article is organized as follows. First, we outline some relevant data that we believe require a dual-store explanation. We then present our combined context–activation model and discuss its account of recency and lag-recency effects in immediate and continuous-distractor free recall and of the dissociations discussed above. Subsequently, we examine the model's predictions and report the experimental tests, and finally, we explore the properties of the activation buffer and its function in memory control.

Critical Data

There is a large database of findings that can inform a theory of list memory. Here, we focus on those that are relevant (and perhaps critical) to the debate regarding the need for postulating a second component (e.g., a short-term store or an activation component) above and beyond a contextual retrieval component (e.g., episodic long-term memory). These effects involve dissociations between immediate and continuous-distractor free recall that have been documented over the years, some of which may not have received enough attention in the memory literature. The effects include output-order effects on serial-position functions (Dalezman, 1976; Whitten, 1978), dissociations due to amnesia (Carlesimo et al., 1996), negative recency effects (Craik, 1970), and output-position effects on lag recency (Howard & Kahana, 1999; Kahana, 1996). Additional dissociations have also been reported, raising further challenges for single-store models (for a review, see Cowan, 1995; for supporting neuroimaging data, see Talmi, Grady, Goshen-Gottstein, & Moscovitch, in press). However, we do not focus on these additional dissociations, as they involve serial-order recall and modality effects, which are beyond the scope of the current study.

Dissociation I: Directed Output Order

In both immediate and continuous-distractor free recall, participants are free to recall the items in any order. However, in immediate free recall, participants typically recall items from the end of the list before reporting other items (Dalezman, 1976; Howard & Kahana, 1999; Kahana, 1996; Nilsson, Wright, & Murdock, 1975). This pattern is not always found in continuous-distractor free recall (Bjork & Whitten, 1974; but see Howard & Kahana, 1999). Of importance, when instructions are used to manipulate the order of recall—starting with items from the end (*end first*) or beginning (*beginning first*) of the list—a dissociation is found between immediate and continuous-distractor free recall (see Figure 8 with model simulations). In immediate free recall, short-term recency is present under end-first instructions (and does not differ from standard immediate free recall) but is absent under beginning-first instructions (Dalezman, 1976). In contrast, long-term recency is present both under end-first instructions and under beginning-first instructions (Whitten, 1978; and under both sets of instructions, does not differ from performance in standard continuous-distractor free recall; cf. Bjork & Whitten, 1974). This dissociation led one of the original discoverers of the long-term recency effect to suggest that "it seems most reasonable to search for different explanations for short-term and long-term recency effects" (Whitten, 1978, p. 690). We embrace this suggestion.

Dissociation II: Amnesic Syndrome

A neuropsychological dissociation strengthens the conclusion that short- and long-term recency rely on different cognitive processes (Carlesimo et al., 1996). Carlesimo and colleagues (Carlesimo et al., 1996) showed that the absolute immediate free recall performance of the last three (i.e., most recent) serial positions of a 10-word list did not differ between amnesic patients and healthy control participants (see also Baddeley & Warrington, 1970; Capitani, Della Sala, Logie, & Spinnler, 1992). However, performance for prerecency positions in immediate free recall and for all serial positions, including recency positions, in continuous-distractor free recall was lower for the patients compared with the control participants (see Figure 7 with model simulations).

Dissociation III: Negative Recency

A dissociation in performance between immediate and continuous-distractor free recall that has been ignored in the literature is the negative recency effect (i.e., worse recall performance for recency compared with prerecency items) in final free recall. In immediate free recall, if participants are given an unexpected final free recall task at the end of the experiment and are asked to report words from all the lists they previously studied, worse memory is found for the items that occupied the last positions in the original lists. The worse memory for the last items is labeled the "negative recency effect" (Craik, 1970). The finding of negative recency has been replicated many times and with different immediate memory paradigms (e.g., R. L. Cohen, 1970; Craik, Gardiner, & Watkins, 1970; Engle, 1974; Madigan & McCabe, 1971).

A common dual-store interpretation for the negative recency effect is that recency items are in the buffer for a relatively shorter duration than are prerecency items and, therefore, have less time to be episodically encoded. In the final recall test, only the episodic traces contribute to performance, and thus the recency items are at a relative disadvantage and are more poorly recollected. In contrast, a test of final free recall after a series of continuous-distractor trials does not produce negative recency (e.g., Bjork & Whitten, 1974; Glenberg et al., 1980; Koppenaal & Glanzer, 1990; Tzeng, 1973; Whitten, 1978).

Dissociation IV: Interaction of Task and Output Position on Lag Recency

Kahana (1996) showed that recall transitions follow a robust contiguity (lag-recency) pattern. That is, the probability of recalling item j immediately after recalling item i was larger when the words (i, j) were more contiguous, that is, when the lag, $|i - j|$, between the presentations of both items in the study sequence was smaller. In addition, this *lag-recency effect* was found to be asymmetric, such that forward transitions were more likely than backward transitions (i.e., after recall of item i, item $i + 1$ was more likely to be recalled than item $i - 1$). The presence of a lag-recency effect in immediate, delayed, and continuous-distractor free recall motivated Howard and Kahana (1999, 2002) to develop a single-store model that accounts for all these effects (but see Kahana, 1996, who used a short-term buffer to interpret lag-recency effects in immediate free recall).

However, despite the impressive success of Howard and Kahana's (1999, 2002) theory in accounting for data patterns in immediate and continuous-distractor free recall, one aspect of the data on lag recency has not yet been explained. In delayed and continuous-distractor free recall, the lag-recency effect is independent of the output position. In immediate free recall, however, the asymmetry is stronger for the first few recall transitions (Kahana, 1996; Kahana, Howard, Zaromb, & Wingfield, 2002), suggesting a different underlying mechanism. Specifically, in immediate free recall, participants typically start with an item that was presented two or three positions before the end of the list, and then recall proceeds in the forward direction (Kahana, 1996; Laming, 1999). This interaction between task, output position, and lag recency can be explained if one assumes a short-term buffer from which the initial few items in immediate free recall are retrieved in the order in which they entered the buffer (with the oldest item being retrieved first; Davelaar, 2003; Howard & Kahana, 1999; Kahana, 1996).

In summary, we believe that to account for the different dissociations between recency and prerecency items, the existence of a short-term buffer must be assumed. According to this assumption, items that reside in the buffer at the end of the encoding phase are negatively affected by a beginning-first recall, are unaffected in amnesic patients, are less episodically encoded (leading to patterns of negative recency in final recall), and are reported in a predominantly forward manner.

Expected New Dissociation: Proactive Interference

As mentioned in the introduction, several variables, such as list length, presentation rate, word frequency, semantic similarity, and proactive interference affect immediate free recall performance of prerecency but not of recency positions (Craik & Birtwistle, 1971; Glanzer, 1972). Although some of these variables have been reported to have similar effects in continuous-distractor free recall (Greene, 1986a; Greene & Crowder, 1984), we focus here on the manipulation of proactive interference, whose effect has not yet been investigated in continuous-distractor free recall and is theoretically predicted to dissociate the two tasks.

Proactive interference is the observation of a negative correlation between recall performance and the number of preceding trials. Proactive interference is especially large when the items of previous and current trials belong to the same category (Wickens, 1970). The effect of proactive interference in immediate free recall was demonstrated in a study by Craik and Birtwistle (1971) in which participants performed five trials with 15 words per list. All 75 words came from the same semantic category (e.g., animal names). The results showed that recall probability for the items in the list became lower as the lists progressed. Of importance, the recall of the last 6 items in the list (but not of earlier items) was unaffected by proactive interference. Thus, the manipulation of proactive interference dissociated memory for prerecency and recency items. This dissociation was explained in terms of the last items being in the short-term buffer, thereby rendering them immune to proactive interference.

Craik and Birtwistle (1971) approximated the contributions of retrieval from the short- and the long-term stores and found that proactive interference affected only the retrieval from the long-term store. This is consistent with the views that proactive interference is due to competition from related items in previous trials (which are encoded in episodic memory) on the retrieval of items in the current trial (e.g., Wixted & Rohrer, 1993) and that proactive interference does not affect the capacity-limited short-term buffer, as it is found for supra- but not for subspan lists (Halford, Maybery, & Bain, 1988). Because in continuous-distractor free recall all items are retrieved from the long-term store, we expect that all serial positions will show a decrease in recall performance due to proactive interference. This prediction and the aforementioned dissociations are explicitly demonstrated in the model we present in the next section.

Model

We show that a dual-store model that includes not only an episodic long-term memory component but also an activation-based buffer component can account for the critical data described above. To do so, we combine an activation-based buffer with the simplest implementation of an episodic long-term memory component that captures some well-established findings (e.g., long-term recency, asymmetric lag recency). Therefore, although our modeling of the activation buffer is very detailed, the buffer effects being our focus, the episodic component is a simplified implementation. Although combining these two elements is sufficient to account for the critical data, future studies may require combining the activation buffer with a more sophisticated episodic long-term memory model (e.g., the temporal context model [TCM]; Howard & Kahana, 2002).

General Assumptions

The model consists of two components. The first component corresponds to a lexical–semantic long-term memory system in which activated representations constitute an activation-based short-term buffer. This is in line with the neurophysiological (Miller, Erickson, & Desimone, 1996), neuropsychological (R. C. Martin, Shelton, & Yaffee, 1994), neuroimaging (Gabrieli, Poldrack, & Desmond, 1998), and neurocomputational (Durstewitz, Seamans, & Sejnowski, 2000) studies suggesting that the left lateral prefrontal cortex underlies an activation-based semantic buffer. The second component corresponds to an episodic contextual system in which each unit represents a different episodic context. This is in line with studies suggesting that the medial–temporal lobes are involved in the encoding and retrieval of episodic memories (Marr, 1971; Scoville & Milner, 1957). Connections between the two components correspond to a matrix of episodic memory traces. As in previous neural network models of memory (Becker & Lim, 2003; Chappell & Humphreys, 1994), we assume that list memory is mediated by the interaction between a lexical–semantic system and an episodic contextual system. In our model, we conceive of the episodic long-term memory system in terms of the episodic contextual system and the matrix of connections between these two components. This matrix can be viewed as a simplification of a hippocampal system that encodes conjunctions between context and items (O'Reilly & Rudy, 2001).

Following two other process models of free-recall memory (ACT-R, J. R. Anderson et al., 1998; search of associative memory [SAM], Raaijmakers & Shiffrin, 1980, 1981), we assume that the buffer implements item-based rather than time-based forgetting (Glanzer, Gianutsos, & Dubin, 1969). In particular, items are lost from the buffer by being (probabilistically) displaced by incoming items. This is mediated by a mechanism based on recurrent self-excitation and lateral inhibition (see Grossberg, 1978, for previous explorations of these mechanisms in the domain of short-term memory).

In our model, we also assume that items residing in the capacity-limited buffer are subject to encoding in episodic memory. Episodic encoding is supported by Hebbian learning mechanisms, in which the connections between active buffer units and active context units increase. Episodic encoding is proportional to the integral of the above-threshold stimulus activation (this is analogous to the use of a threshold in the domain of visual information acquisition; Busey & Loftus, 1994).

During the recall phase, we assume that all items in the buffer are unloaded,[2] after which an elaborate search through episodic memory ensues. As in SAM, episodic retrieval involves two stages. In the first, the context is used to select items for retrieval, and in the second, the selected item is recovered. Intuitively, the recovery of items can be seen as probabilistically retrieving the phonological motor program.

As in models using contextual retrieval (Burgess & Hitch, 1999; Dennis & Humphreys, 2001; Henson, 1998; Howard & Kahana, 2002; Mensink & Raaijmakers, 1988), we assume that the context in which items are encoded changes during list presentation as well as during retrieval. For simplicity, we use a localistic one-dimensional representation of episodic context (Burgess & Hitch, 1999) in which nearby units correspond to similar episodic contexts.

Consistent with previous studies (G. D. A. Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1999; Henson, 1998; Metcalfe & Murdock, 1981; Shiffrin & Cook, 1978), we assume that, because of enhanced attention (Murdock, 1960; Neath, 1993b; Shiffrin, 1970), the contexts at the end and the beginning of list presentation are accessible during the retrieval phase to drive further recall. The assumption that the start context is accessible during retrieval allows us to account for primacy effects in continuous-distractor free recall. Reports of no primacy effects in continuous-distractor free recall (e.g., Howard & Kahana, 1999; but see our Experiment 1) can be explained within our model in terms of diminished access to the start context (due to attentional effects in tasks in which an additional semantic judgment is required for every word at encoding). Further support for the special access to the start context is revealed in the output-order data from Murdock and Okada (1970; see also Metcalfe & Murdock, 1981, and the General Discussion), who showed that a break occurs in the recall process, after which recall proceeds with items from the beginning of the list.

The model can be applied to both cued- and free-recall paradigms. In cued recall, the external cue is used to probe memory. In free recall, in which no external retrieval cue is provided, the model undergoes a search through context space, retrieving the contexts to use as retrieval cues for the list items.

Overview of Structure and Processes

Structurally, the model consists of two sets of interconnected components with localistic representations. In the lexical–semantic component, each unit corresponds to a different semantic chunk (e.g., "cat," "red-nosed reindeer," and "3 + 4 = 7"). In the episodic component, each unit corresponds to a distinct episodic context (see Figure 1). Each lexical–semantic unit has four different types of connections. First, each lexical–semantic unit has a

---

[2] The term *unloading* is used to distinguish between the fast output of buffer items and the slower process of retrieval from episodic memory. (See Appendix B for the unloading dynamics.)

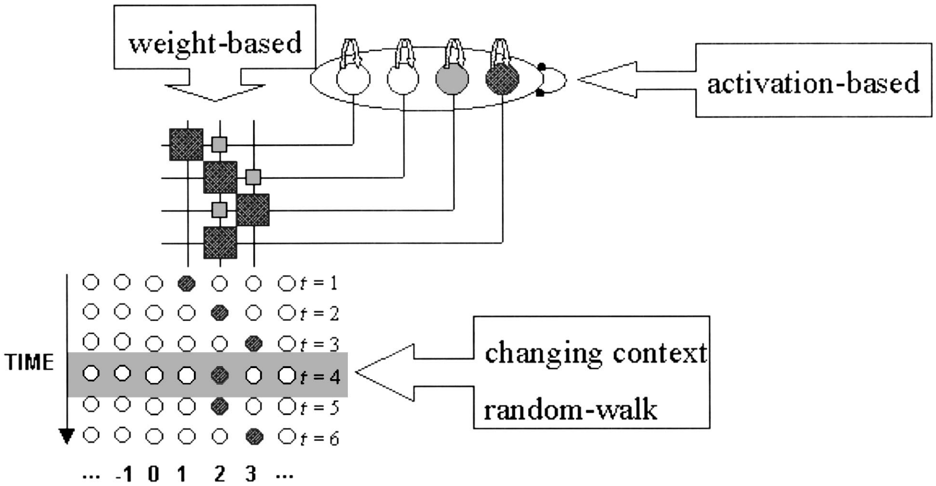

Figure 1. Architecture of the context–activation model. The activation-based short-term buffer (implemented as a recurrent network of localistic units) is illustrated by the ellipse (top), and a context representation (implemented as a linear arrangement of localistic units) is illustrated as it evolves in time (bottom left). The two systems are connected by a matrix of (positive) weights (with no upper boundary). All units are represented as circles; black-filled circles are units that are highly active, and gray-filled circles are units that are less active. Buffer units have a continuous activation level, whereas context units have a binary (on/off) activation level. The strengths between the context and buffer units are depicted as squares of various sizes (the larger the square the stronger the connection). With every presentation of an item, the context moves left or right according to a random walk. The matrix of connection weights between the activation-based buffer and the context representation forms the episodic memory. This figure shows the state of affairs when Item 4 is active in the buffer (with a capacity of two items) and Context Unit 2 is active. The episodic memory matrix contains strong (large squares) and weak (small squares) connections, which are used during episodic retrieval.

self-excitatory connection back to itself, which leads it to recycle some of its activation and maintain it after stimulus offset. Second, lexical–semantic units inhibit each other, causing them to compete for activation and resulting in displacement from the buffer when too many units are active at the same time (i.e., when the capacity of the system is surpassed). Together, the lexical–semantic units and the self-excitatory and inhibitory connections implement a capacity-limited activation-based buffer. Third, each lexical–semantic unit is weakly connected with its semantic associates, forming a localistic semantic network. Fourth, each lexical–semantic unit receives input from the contextual representation. Episodic learning involves generating a set of connection weights between the contextual and the lexical–semantic representations.

Processing in the model involves (a) sequentially activating items (corresponding to a memory list) in the short-term buffer; (b) changing the context according to a random walk process; (c) encoding items in the episodic long-term memory system (i.e., changing the weights between lexical–semantic and context units) on the basis of activation of lexical–semantic representations and the variable context; and (d) recalling items from the activation-based short-term buffer and the episodic long-term memory system.

The evolution of the connection matrix that satisfies a–c is illustrated in Figure 2, for the encoding of four sequentially activated list items (and for a buffer with a capacity of two items, used here for illustration purposes). When Item 1 is presented ($1 < t < 500$), the corresponding lexical–semantic unit becomes active. Context Unit 1 is also active during this interval. This results in strengthening of the connection between Item 1 and Context Unit 1, as depicted by the large square. Next ($501 < t < 1{,}000$), Item 2 becomes active and partially inhibits Item 1 (which is now at a lower level of activation; see model implementation), as shown by the gray-filled circle. At this time, Context Unit 2 is active, and therefore, the connection between Item 2 and Context Unit 2 is strengthened. Note that Context Unit 2 is also connected with Item 1, although the increase in strength is smaller, as depicted by the smaller square. When Item 3 is presented ($1{,}001 < t < 1{,}500$), Item 1 is displaced from the buffer. In this example, Context Unit 3 is active and is connected to Item 3 and to Item 2, albeit with a lower strength to the latter. Finally, when Item 4 is presented ($1{,}501 < t < 2{,}000$), Context Unit 2 becomes activated again and is thus associated with Item 4 and, to a lesser extent, with Item 3. The resulting connection matrix corresponds to that in Figure 1.

Implementation

Here we present the implementation procedure of the model (see Appendix A for additional details and simulation protocol).

Buffer component. The buffer layer has a lexicon of N units, of which only a small number are used to simulate a trial in a list-memory experiment. Systematic explorations (Davelaar, 2003) have shown that the size of the lexicon does not alter the buffer dynamics. Equation 1 implements the assumptions regarding the activation dynamics of units in the buffer layer. Each unit has a self-recurrent excitatory connection of strength α_1, which recycles some of its activation back to itself, permitting activation maintenance (i.e., short-term retention) and counteracting time-based activation decay (with proportion $1 - \lambda$). In addition, each unit competes with every other unit via global inhibition (i.e., inhibits every other unit) of strength β,

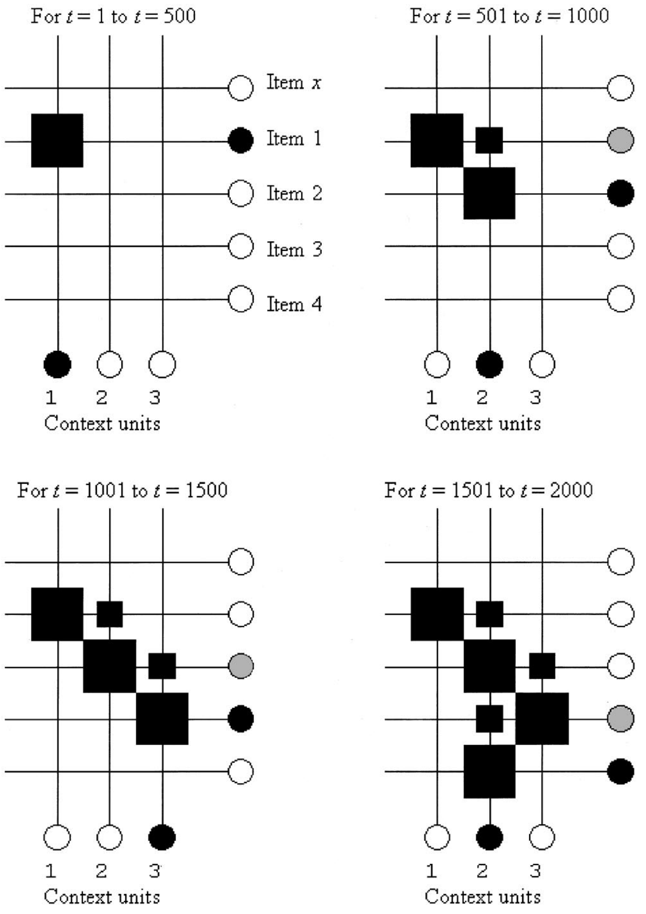

Figure 2. Evolution of the matrix of episodic connection strengths during the presentation of the first four items. The lexical–semantic units are presented vertically on the right, and the context units are presented horizontally below the matrix. Item *x* corresponds to a lexical–semantic item that is not part of an experimental list but is part of the lexicon. Activation levels are depicted by the darkness of the circles. Square sizes depict the magnitudes of the connection strength between item and context units. Each panel corresponds to a time interval in which a new word is presented.

giving rise to capacity limitations and to displacement from the buffer (when this capacity is exceeded). The units are assumed to be activated through bottom-up sensory stimulation. The activation output of each unit is a function of its previous activation (after losing some of it because of temporal decay) combined with self-recurrent excitatory input (which overrides this decay), global inhibitory input it receives from other buffer units, excitatory input from semantic associates, sensory input, and small stochastic fluctuations. The activations of all units ($i = 1$–N, $N >$ list length) are updated in parallel according to Equation 1, where the activation, $x_i(t)$, at time t depends on the activation in the previous time step, $x_i(t-1)$; the recurrent self-excitation, $\alpha_1 \mathrm{F}(x_i(t))$; the global inhibition, $\beta \Sigma \mathrm{F}(x_j(t))$; the excitation from semantic associates, $\alpha_2 \mathrm{F}(x_i(t))$, the sensory input, $\mathrm{I}_i(t)$; and zero mean Gaussian noise, ξ, with standard deviation σ and the decay parameter λ ($0 < \lambda < 1$):

$$x_i(t+1) = \lambda x_i(t) + (1-\lambda)[\alpha_1 \mathrm{F}(x_i(t)) - \beta \sum \mathrm{F}(x_j(t)) + \alpha_2 \mathrm{F}(x_i(t)) + \mathrm{I}_i(t) + \xi]. \quad (1)$$

Because we mainly focus on memory for lists of unrelated words, $\alpha_2 = 0$ in most of our simulations, except in the Utility of the Dynamic Buffer section, where we address semantic effects. The output activation function $\mathrm{F}(x) = x/(1 + x)$ (for $x > 0$, 0 otherwise) is threshold linear at low input (J. A. Anderson, Silverstein, Ritz, & Jones, 1977; Usher & McClelland, 2001) and includes a saturation at high input (for discussion, see Usher & Davelaar, 2002). This activation function is also used in standard neurocomputational textbooks (see, e.g., O'Reilly & Munakata, 2000, pp. 46–49, where it is labeled XX1). In the simulations, each time step or iteration corresponds to a small, constant time interval (in the order of milliseconds).

Contextual system. The context layer consists of a linear arrangement of units, which are indexed by integers . . . , −2, −1, 0, 1, 2, . . . (see Figure 1). In contrast to the buffer layer, in which several units can be active simultaneously, only one context unit can be active at each time step. A trial begins with one active context unit that gets associated with the start signal (e.g., *GET READY*). On the presentation of each new item, the context may or may not change, according to a random walk: At each time step the context moves one unit rightward, one unit leftward, or remains at the same position with probabilities P^+, P^-, and $1 - \mathrm{P}^- - \mathrm{P}^+$, respectively (where $\mathrm{P}^+ + \mathrm{P}^- < 1$). To account for the asymmetry of the lag recency, we introduce a bias (or drift) in the random walk,[3] $\mathrm{P}^+ > \mathrm{P}^-$.

Encoding. During encoding, the active context unit, c_j, gets associated with all those buffer units (but not the distractor units) that are active above a threshold φ_1 (e.g., $\varphi_1 = .20$). The context unit c_j becomes associated with the active buffer unit i by modifying the strength of the connection between them, in accordance with a learning rate parameter, ε, and in proportion to the above-threshold level of activation of the buffer unit. Thus, episodic encoding is supported by Hebbian learning mechanisms, in which the connections between active buffer units and active context units increase. Specifically, the episodic trace strength $W_{i,j}$ of the connection is updated according to Equation 2, which is integrated across time until the time of retrieval (i.e., the presentation of the recall prompt):

$$\Delta W_{ij} = \varepsilon(\max[0, \mathrm{F}(x_i) - \varphi_1]). \quad (2)$$

We assume (not modeled explicitly) that the recall prompt increases the retrieval inhibition (see the Utility of the Dynamic Buffer section and Appendix B) and, therefore, speeds up the deactivation of the buffer items.[4] The distractor task is simulated by activating a sequence of nonlist distractor units (see Simulation 1) that are part of the lexicon (and thus compete) but are not subject to episodic learning.

---

[3] The assumption of a drift was the simplest assumption we could use to account for the lag-recency asymmetry in continuous-distractor free recall. A more complex assumption to account for the effect was used by Howard and Kahana (2002), who introduced a second type of context (a preexperimental context) that is always retrieved with the item. Because the lag-recency effect was not the main focus of our investigation and because we saw the two assumptions as functionally equivalent, we decided to rely on the simplest mechanism in our simulations. Critically, notwithstanding the drift, our model was able to capture the critical data and dissociations. That is, the basic mechanisms of our model were not hurt, as they might have been, by the introduction of a drift into the random walk.

[4] This is consistent with other computational models that address separately the processes of encoding and retrieval (Hasselmo & Wyble, 1997; O'Reilly & McClelland, 1994).

Retrieval. In free recall, the retrieval processes are driven by both the activation buffer and the internal context representation activating the items that are associated with it, in proportion to the episodic connection weights (W_i). The items are then probabilistically selected for verbal report. To approximate this process, we assume (J. R. Anderson et al., 1998; Raaijmakers & Shiffrin, 1980, 1981) that active items are immediately reported first. If several items are active at the time of test, these items are reported in order of their episodic strengths to the current context, such that the one with the strongest episodic connection to the current context is retrieved first, followed by the item with the next-strongest connection to the current context, and so on. This rule is a simplification of a more complex retrieval process in which the inhibition in the buffer is increased (see the Utility of the Dynamic Buffer section) and the buffer interacts with the contextual system (see Appendix B). After retrieval of the items from the activation buffer, a more elaborate episodic retrieval process ensues, as described below.

The time steps at retrieval are not the same as the time steps at encoding but are more course grained and are referred to here as *retrieval attempts* (that extend over a number of time steps). (We do not attempt to model the detailed dynamics of the retrieval, but see Appendix B.) For each retrieval attempt, we follow the implementation of two previous process models of free recall. As in the ACT-R model and in SAM, episodic contextual retrieval proceeds in two stages: selection and recovery. The equations describing selection and recovery are identical to those of the ACT-R model. Specifically, the probability, P^{sel}, of selecting a particular item, *i,* is a noisy competitive process, based on the relative strength between item *i* and the active context unit compared with all other strength values of items that are associated with the active context unit. This is approximated by a Luce choice rule (Luce, 1959) with ρ as the selection noise:

$$P_i^{sel} = \frac{\exp(W_i/\rho)}{\Sigma \exp(W_j/\rho)}. \tag{3}$$

When an item is selected, a recovery process follows, determining whether the selected item is retrieved. The probability P^{rec} of recovering the selected item is a sigmoidal function of the episodic strength in which the strength W_i is compared with a recovery threshold φ_2 with retrieval noise, τ.

$$P_i^{rec} = \frac{1}{1 + \exp(\varphi_2 - W_i)/\tau}. \tag{4}$$

Note that although a recovered item cannot be selected again for overt report, it continues to compete in the selection probability and thereby affect the recall of subsequent items. This can lead to reselection of a previously retrieved item, which does not produce an overt report. We label an attempt that does not produce an output a *silent event.*

During the retrieval stage, the random walk of the active context continues. The total duration of the retrieval phase is fixed as a certain number of 2*k* retrieval attempts. The context may change with every retrieval attempt, according to the same probabilities P^+ and P^- as those during encoding. For the first *k* attempts, the context continues to change from where it ended after list presentation. Each attempt may result in the retrieval of an item or may elicit a silent event (with the silent event reflecting selection of an extra-list item, reselection of a retrieved item, or recovery failure). After *k* attempts, the context unit that was active at the beginning of the study trial is reactivated, and the context continues to change for an additional *k* retrieval attempts. For the sake of simplicity, it is assumed that the start context is always retrieved during the recall phase. The retrieval phase is terminated after a fixed number of 2*k* retrieval attempts have been made. If another list is presented for study, its first item is associated with the context unit that was active at the end of the retrieval phase of the previous list (after the 2*k* retrieval attempts).

In cued recall, probing memory with the cue is implemented by selecting the activated item that is consistent with the cue (see Usher & Davelaar, 2002, for a neurophysiological model of this selection process). We applied the cue only to the activated items in cued recall, consistent with the experimental procedure we modeled that involves a time deadline (see Experiment 2).

Model Behavior

Here we present the model's behavior in accounting for serial position in list memory. First, we present the buffer dynamics and its contributions to immediate free recall. Then we present two examples of the full model being applied to simulating a full encoding–retrieval trial in immediate and in continuous-distractor free recall.

Buffer dynamics. In the simulations presented in this article, a trial in a list-memory experiment is modeled by successively presenting input to a number (i.e., the list length) of units, each for a number of iterations (related to the presentation time). When a buffer unit is active above threshold, it is said that the item represented by this unit is in active memory.[5] Because of the self-excitation, units can remain active after stimulus offset (but with a decrease in activation due to the cessation of bottom-up sensory input), and several units in the buffer layer may be active simultaneously. When new units are activated, this increases the global inhibition, which in turn affects all units that do not receive sensory input. Ultimately, this results in the (probabilistic) displacement of the unit with the lowest level of activation from the buffer. However, when no new units are activated, the self-recurrent excitation causes active units to remain active, even when they no longer receive sensory input (Haarmann & Usher, 2001).

Figure 3 illustrates these activation dynamics for the presentation of a list of 12 words. It shows the activation trajectories of 12 buffer units that were activated sequentially for 500 iterations each (the sensory input, I, set to be equal to .33, switches from unit to unit every 500 iterations). The iteration number is given on the abscissa, whereas the output activation level, F(x), is set on the ordinate. Notice that items are deactivated in the buffer because of their displacement by new items and not because of passive decay. If an unfilled delay exists between the last item and the recall prompt (the interval from t = 6,000 to t = 8,000), the activation of the last item is maintained (see also Appendix B).

Although the buffer layer in Figure 3 contains 24 units, only 12 received sensory input. For the sake of clarity, the noise level for this simulation alone was set to 0. The values for the self-recurrency (α_1 = 2.0) and global inhibition (β = .20) were arbitrarily chosen within boundaries on the basis of previous investigations (Davelaar, 2003). At each iteration, all units were updated, not only the unit that

---

[5] Our model is formulated within a *dual-trace* framework, in which above-threshold activation constitutes the contents of the short-term store. This is in contrast to *dual-weight* models (e.g., Burgess & Hitch, 1999) that relate the short-term buffer to fast-decaying weights (for the taxonomy, see Levy & Bairaktaris, 1995).

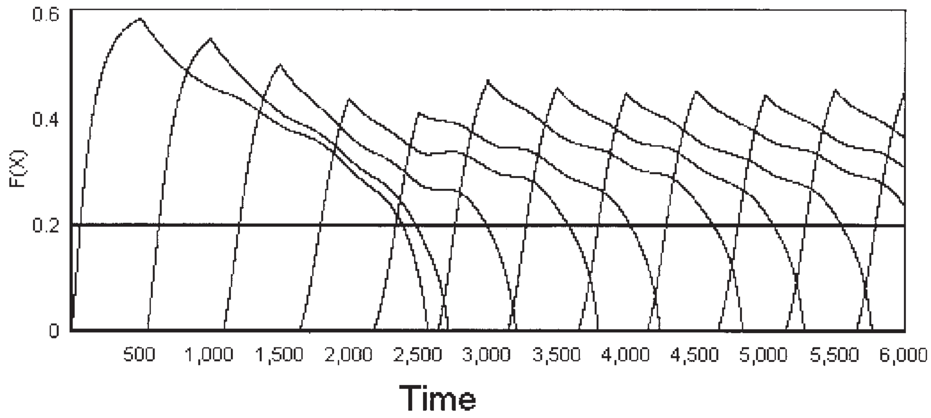

Figure 3. Activation trajectories of 12 sequentially activated buffer units, up to the moment when (in immediate free recall) a recall prompt is provided. The number of time steps is set on the abscissa, whereas the output activation value, F(*x*), is set on the ordinate. All units active above a certain memory threshold value (e.g., .2) are assumed to be accessible for subsequent recall from the buffer.

received input. The first item in the list reaches a level of activation that is higher than that of subsequent items. This is because the first item enters an empty buffer and does not have to overcome the inhibition of already active items.

A simplified account of the contribution of the active memory component to serial-position functions can be obtained by assuming that at the end of the trial (at $t = 6{,}000$), all items active above a threshold, φ_1, are unloaded. Given that new items do not enter the buffer, displacement of activated items is negligible (but see Appendix B for a more detailed account of the retrieval dynamics). As a result, almost all items that are active above the buffer threshold at recall remain in that state during the time that items are unloaded. The unloading of items from the buffer can apply to both cued and free recall. For the former (see Experiment 2), an item is reported only if it is active and its identity fits with the cue (e.g., "cat" in response to the cue *animal*). For free recall (see Experiment 1), we assume that all the items above the activation threshold are reported.[6] In the example, only the last four activated units are active above the buffer threshold ($\varphi_1 = .20$; horizontal line)[7] at the time of test ($t = 6{,}000$).

Figure 4 presents the serial-position function of the proportion of 1,000 simulation runs that an item is active above threshold at time of test ($t = 6{,}000$), with a moderate level of noise ($\sigma = 1$). The function shows that only the units that were activated last are still active above threshold; that is, the function displays a clear recency effect. The capacity of the system can be estimated by summing the proportions of the serial-position function. In Figure 4, this capacity is 3.93, which is consistent with the value of 4 ± 1 argued in a recent review to be the better estimate of the capacity of short-term memory (Cowan, 2001) and was also obtained in previous studies (Haarmann, Davelaar, & Usher, 2003; Haarmann & Usher, 2001).

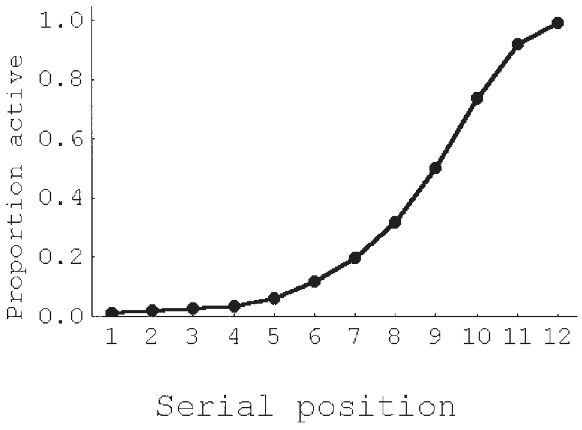

Figure 4. Serial-position function showing the proportion of simulation runs that an item presented at a given input position is still active above threshold at the end of the sequence.

Example trial: Immediate free recall (full model). The following examples (top of next page) illustrate the full process of encoding (in bold) and retrieval (nonbold) phases for an immediate (IFR) and a continuous-distractor (CD) free recall trial. Consider first an immediate free recall trial with nine items. The time arrow goes from left to right.

The top row represents the external events, such as presentation of the start (S) and the end (E) cues, the list items (the digits in bold), and the retrieved items (the digits in italics), and includes the silent moments in the retrieval phase (x = no item selected, y = recovery failure, and z = reselection of item). The second row represents the number of the context unit that is active at that time. The change of context is independent of the items that are encoded or retrieved and continues throughout the retrieval phase. At the beginning of the retrieval phase, Context Unit 4 is active (associated with E), and Items 7, 8, and 9 are in active memory, as denoted by the underline beneath those items.

---

[6] A detailed discussion on the control processes involved at retrieval in cued and free recall is provided in the Utility of the Dynamic Buffer section and Appendix B.

[7] This threshold is used to prevent noisy activation from affecting performance. The model is not sensitive to the precise value chosen for this threshold (e.g., any value between .1 and .2 is sufficient; see Figure 3) because of the abrupt deactivation process (for analysis, see Usher & Davelaar, 2002).

		Encoding											Retrieval														
IFR	Item:	**S**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**E**	*8*	*9*	*7*	*6*	z	z	x	x	x	*1*	*2*	y	*5*	*4*	z
	Context:	**0**	**0**	**1**	**2**	**1**	**2**	**3**	**3**	**4**	**4**	**4**	_	_	_	3	3	4	5	5	6	0	1	1	2	1	2

		Encoding													Retrieval											
CD	Item:	**S**	**D**	**1**	**D**	**2**	**D**	**3**	**D**	**4**	**D**	**5**	**D**	**E**	5	x	x	x	x	x	x	1	x	3	y	4
	Context:	**0**	**0**	**1**	**2**	**1**	**2**	**3**	**3**	**4**	**4**	**4**	**3**	**3**	4	5	5	6	5	6	0	1	1	2	3	4

At the beginning of the recall stage, Items 7, 8, and 9 are unloaded (context does not change here). The order of output is Item 8, Item 9, and then Item 7 (see Appendix B for a detailed discussion). Next, the context changes to Unit 3. At this stage, Items 6 and 7 start to compete for selection or retrieval. In this example, Item 6 is selected and recovered. Next, Context Unit 3 remains active but leads to reselection of Item 6 (or 7), leading to a silent event z. The context changes to 4, but again no output is made, as all associated items have been reported. A further change in the context results in a context that is not associated with any items. After five unsuccessful attempts, the start context is retrieved, and the context changes again (for $k = 6$ attempts). Item 1 is retrieved, followed by Item 2, a reselection of Item 2, retrieval of Item 5, retrieval of Item 4, and a reselection of Item 2. Note that if another list were presented, its first items would be associated with Context Unit 2, which is the unit that is active at the end of the retrieval phase.

Example trial: Continuous-distractor free recall (full model). The next example illustrates the encoding and retrieval in a continuous-distractor free recall trial of five items. The letter *D* indicates a nonlist distractor item. For the sake of simplifying the illustration, we assumed in this example a single distractor item during the IPI and the RI. However, this was not the case in our simulations, in which we typically used a number of distractors per interval.

At the time of retrieval, Items 1 and 2 are associated with Context Unit 1, Item 3 with Context Unit 3, and Item 4 with Context Unit 4. The idea that no item resides in active memory is represented by the fact that no items are underlined. The context changes to Unit 4, and Items 4 and 5 compete for output, whereby Item 5 (probabilistically) wins the competition. After five unsuccessful attempts in which no items are retrieved because of lack of association with the active context units, the start context is used, and Items 1, 3, and 4 are retrieved.

Simulations

Although the complete model has many parameters, all are fixed at values based on previous explorations[8] (Davelaar, 2003; Davelaar & Usher, 2002; Usher & Cohen, 1999). Even though methodological differences across experiments may warrant some variation in parameter values (which may be needed for precise quantitative fits), all simulations reported in this article were conducted with the same set of parameters to provide a stringent test of the model's ability to account for the qualitative patterns of the critical data. See Appendix A for details on the parameter values and details on the procedure used in all simulations.

Simulating the Basic Data

In the following simulations, we show that the two-component context–activation model can account for the basic effects in list memory described above. Critically, although simulations of the activation component have been shown to work (Haarmann & Usher, 2001; Usher & Cohen, 1999), here we combine the buffer with a changing context component and ask whether their combined contribution, together with an episodic weight matrix, resembles empirical behavior.

Simulation 1A: Serial-position functions. The model simulated 1,000 trial runs of immediate, delayed, and continuous-distractor free recall. The simulated list was of length 12. For simplification, we model the distractor interval as a sequence of 12 distractor items that are the same as other items except that they are not part of the list. We assume that these items are not retrieved because they belong, for example, to a different category. Each unit (list or distractor item) was activated for 500 iterations. The buffer layer contained a total of 40 units, all of which were updated in parallel at every time step (this included the 12 list items, the 12 distractor items, and 16 nonlist/nondistractor items). As shown in Figure 5, the model produces serial-position functions for immediate, delayed (Glanzer & Cunitz, 1966), and continuous-distractor free recall that are in agreement with those described in the literature.

Examination of Figure 5 reveals a primacy effect in immediate, delayed, and continuous-distractor free recall. The buffer (as well as the retrieval of the start context, which can serve as a retrieval cue) mediates the primacy effect in immediate and delayed free recall. In these tasks, the first item enters an empty buffer, which has a low level of inhibition. The item can therefore reach a high level of activation and can stay active longer than subsequent items (see the simulation in Figure 3). This leads to a stronger episodic trace (see Equation 2) for the first item compared with middle list items (which enter an occupied buffer with a moderate amount of inhibition). During retrieval, the first item has an advantage over middle list items, leading to the primacy effect. The primacy effect in immediate and delayed free recall is further enhanced by the use of the start context during retrieval. In contrast, the primacy effect in continuous-distractor free recall is due only to the retrieval of

---

[8] The previous explorations have led us to a set of parameter values that capture the qualitative aspects of the data from several memory paradigms such as free and cued recall as well as the Brown–Peterson task (J. Brown, 1958; Peterson & Peterson, 1959).

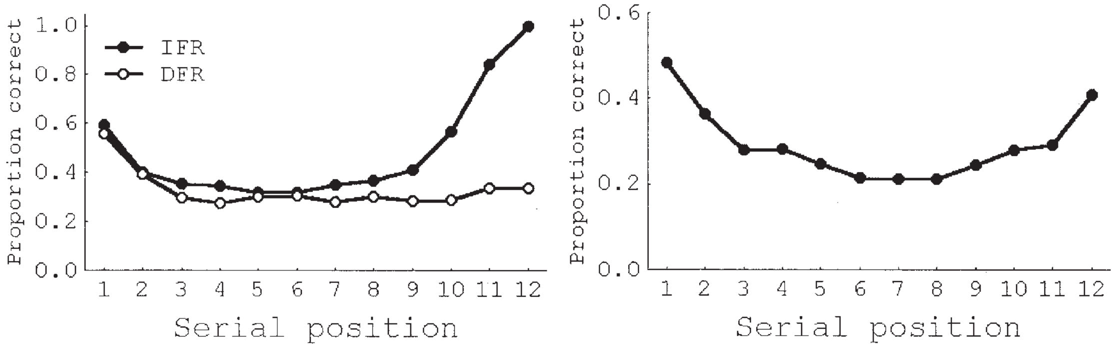

Figure 5. Left: Simulated serial-position functions for immediate free recall (IFR) and delayed free recall (DFR). Right: Simulated serial-position function for continuous-distractor free recall.

the early list items after the retrieval of the start context,[9] because the first list item enters a buffer that is already filled with distractor items and can, therefore, reach only a moderate level of activation.

More important to our present concerns, Figure 5 reveals that the distractor interval eliminates the recency effect in delayed free recall (Figure 5, left), yet the recency effect is restored in continuous-distractor free recall. The large recency effect in immediate free recall is due only to the contribution of the short-term buffer to retrieval of these items (see the simulation in Figure 3), as the episodic component yields only a negative recency contribution (see Figure 10 below). In delayed free recall (Glanzer & Cunitz, 1966), the active items are displaced by the distractor items that become activated in the postlist distractor interval, thereby eliminating a possible contribution from the buffer. Although in continuous-distractor free recall none of the items are in active memory at retrieval, a recency effect is still obtained. The long-term recency effect occurs because some recency items are more likely to have been associated with the end context than are prerecency items. Note that the recency effect that is found in immediate free recall is larger than that found in the continuous-distractor task, which mimics the standard empirical finding (Howard & Kahana, 1999; Poltrock & MacLeod, 1977). The larger effect in immediate free recall can be understood as emerging from the errorless unloading of items from the short-term buffer. This contrasts with the recency effect in continuous-distractor free recall, which is primarily mediated by the reinstatement of the encoding context in episodic memory and is error prone.

Simulation 1B: Contiguity effects. Next, we examined the contiguity effects (lag recency: the greater probability for retrieving items from nearby than from remote serial positions) for the simulation runs of delayed and continuous-distractor free recall in the previous simulation (see Figure 6). We calculated the lag-recency functions across output positions as described in Howard and Kahana (1999). Because lag recency in immediate free recall changes with output position, we postpone its presentation to Simulation 6.

As can be seen, the model produces lag-recency functions with a forward bias for delayed and continuous-distractor free recall. In the model, lag recency occurs because when context n leads to the retrieval of an item, the next item that is recalled is likely to be associated with context $n - 1$, n, or $n + 1$. The asymmetry, however, is a direct consequence of the bias in the random walk (i.e., the greater probability for a contextual change in the forward than in the backward direction).

Simulation 2: List-length effects. Greene (1986a) reported that list length did not dissociate performance in immediate free recall and in the continuous-distractor task. He found this result to argue against the notion of a dual-store model mediating the recency effect in free recall. However, our dual-store model does not predict that any variable will necessarily dissociate recency effects in immediate and continuous-distractor free recall. To see whether list length would produce a dissociation in our dual-store model, we ran 1,000 trial runs of immediate and continuous-distractor free recall with lists of length 20. All other parameters were held constant. The results are shown in Figure 7, together with the serial-position function obtained in Simulation 1A for list lengths of 12.

The model shows the general list-length effect in both tasks, with the proportion of items recalled being higher for short than for long lists (12% difference for immediate and 8% difference for continuous-distractor free recall). More important, the results reveal that list length has a qualitatively similar effect on the serial-position function of immediate and continuous-distractor free recall. In both tasks, prerecency but not recency items are affected by the list-length manipulation.

The model accounts for this association as follows. In immediate free recall, the last items are in the activation buffer, from which they are reported with almost no error. However, the prerecency items are retrieved through a competitive retrieval process in episodic memory. As more items compete for episodic retrieval,

---

[9] Note that the model produces primacy without recourse to a rehearsal mechanism. Although there are different interpretations of rehearsal, the typical one involves a loop of deactivation and activation of an item. Adding a mechanism that reactivates displaced buffer items would increase the primacy effect further and allow a consideration of the data obtained with the overt rehearsal paradigm (Brodie & Murdock, 1977; Rundus & Atkinson, 1970; Ward, 2002). Nevertheless, when measures are taken that are assumed to eliminate the use of a rehearsal strategy, small primacy effects are still found (Baddeley & Hitch, 1974, 1977; Howard & Kahana, 1999).

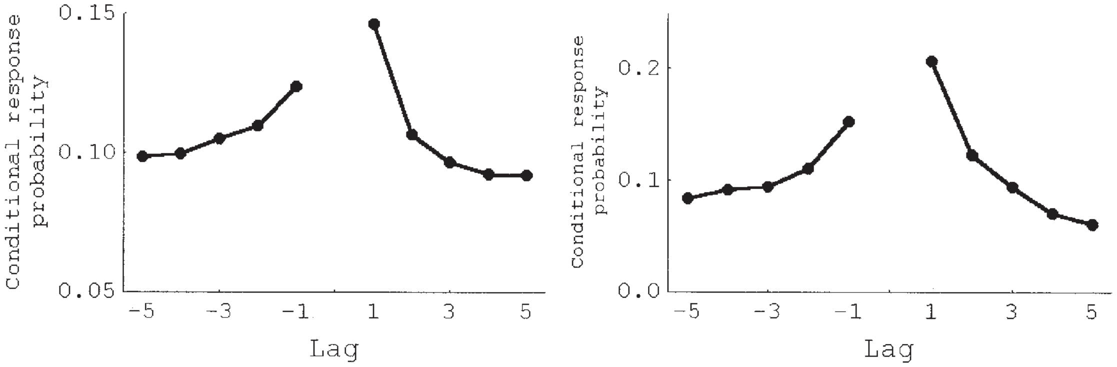

Figure 6. Lag-recency functions for delayed free recall (left) and continuous-distractor free recall (right).

the probability that an item will be selected decreases (see Equation 3) and, thus, the recall performance with the increase in list length decreases. In continuous-distractor free recall, all items are retrieved from episodic memory, but the context units that are associated with the last items are most likely to become activated during retrieval (as retrieval starts with the end context), thereby minimizing the effect of list length. The context units that are associated with middle list items, however, are visited less often and, in longer lists, are associated with more list items. Hence, middle list items are retrieved with lower probability in long than in short lists.

The finding that variables like list length have qualitatively similar effects on serial-position functions in immediate and continuous-distractor free recall led Greene (1986a) to argue that a single mechanism underlies performance in the two tasks. However, as mentioned in the introduction, it is not clear that a dual-store model needs to predict a dissociation between the two tasks for all variables. The context–activation model, which is a dual-store model, predicts the association found with list length. For recency items, both the activation buffer and a retrieval mechanism that uses the end context predict that list length should not affect performance. For prerecency items, the model predicts that list length should affect the weight-based recall in both immediate and continuous-distractor free recall.

In summary, the model accounts for both short- and long-term recency effects and the elimination of recency in delayed free recall. In addition, the model explains primacy effects without recourse to a rehearsal mechanism. Moreover, the model suggests a different explanation for primacy effects in immediate than in continuous-distractor free recall. Finally, the model proposes different mechanisms underlying short- and long-term recency, despite the association between the two tasks with a list-length manipulation.

Simulating the Critical Data

The previous simulations revealed (among other things) that the context–activation model accounts for short- and long-term recency in different ways. Short-term recency is mainly due to retrieval from the buffer, whereas long-term recency (and long-term primacy) is exclusively due to the episodic encoding–retrieval mechanism that operates on a changing context representation. Given that the underlying mechanisms behind the two recency effects may be so different, these differences may account for the dissociations between short- and long-term recency discussed in the Critical Data section. We now apply the model to these dissociations.

Simulation 3: Directed output order. When participants start their recall protocol with items from the beginning of the list, long-term recency remains, but short-term recency is no longer found (Dalezman, 1976; Whitten, 1978). To simulate the effect of output order, we conducted 1,000 simulations for immediate and

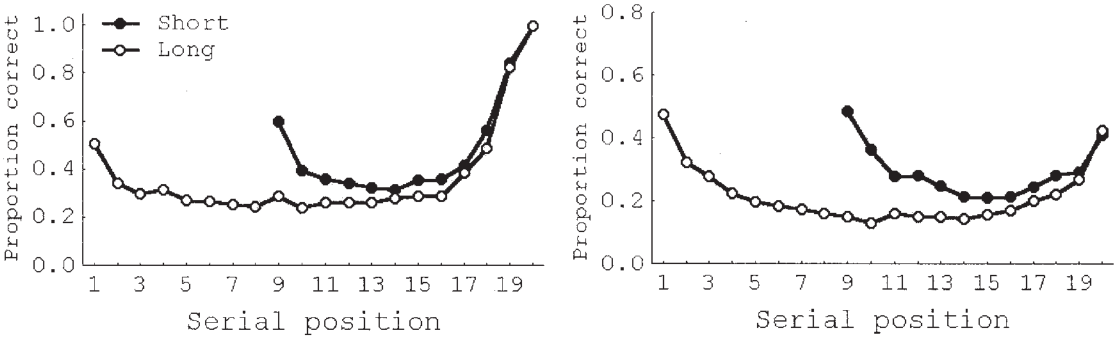

Figure 7. List-length effects in immediate free recall (left) and continuous-distractor free recall (right) for list lengths of 12 (short) and 20 (long).

continuous-distractor free recall. In the *end-first condition,* the end context was used to drive recall, as occurs when no instructions are given concerning output order. For immediate free recall, this means that the active items were reported first and were followed by episodic retrieval. After *k* attempts from the end context, the context changes to the start context and drifts for another *k* attempts. In the *beginning-first condition,* we assumed that participants use the start context to start recall from the beginning of the list and that the retrieved items displace the current contents of the buffer (see Figure B3 in Appendix B). We further assume for the beginning-first condition that two additional contextual updates are inserted after the final item or distractor unit (consistent with data showing a longer duration before the first item is retrieved; see Laming, 1999, for a reanalysis of Murdock & Okada, 1970). During this period, episodic traces are still formed. Retrieval is then driven initially from the start context for *k* attempts, followed by a further *k* attempts from the end context (which is the context associated with the two contextual updates after the final item or distractor unit).

As can be seen in Figure 8, the model parallels the empirical data. The short-term recency effect is absent in the beginning-first condition because items in the activation buffer are displaced by reported items. Still, these items are retrieved from episodic memory later in the recall protocol (in the second set of *k* retrieval attempts), as can be seen by the level of recall being the same as that for middle list items. The additional episodic encoding before the retrieval of the first items overcomes an otherwise negative recency effect. In continuous-distractor free recall, directed output order does not have a major impact on the recency effect.

Simulation 4: Neuropsychological dissociation. Whereas output order affects only recency items in immediate free recall, the opposite effect was found with amnesic patients. Carlesimo and colleagues (Carlesimo et al., 1996) showed that compared with matched control participants, amnesic patients exhibit lower per-

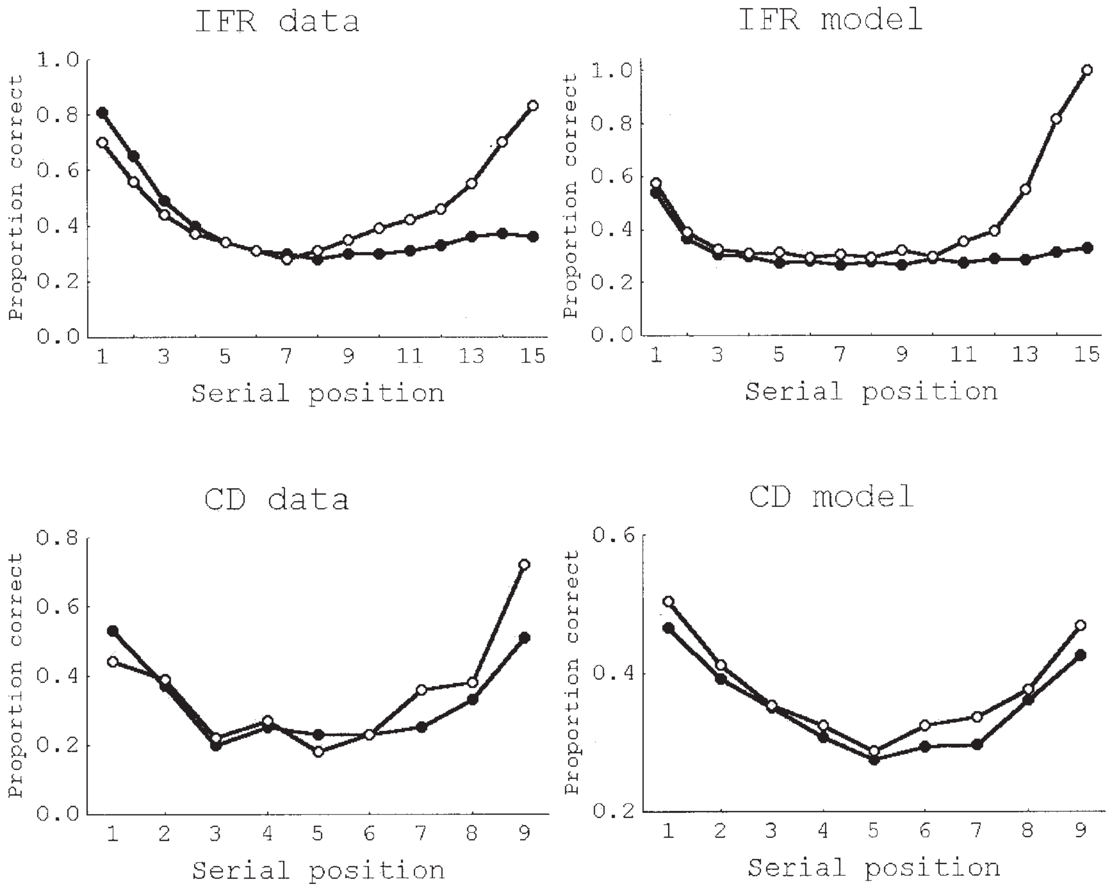

Figure 8. Data from Dalezman (1976; top left) and Whitten (1978; bottom left) and model simulations (top right and bottom right) for the effect of instructed output order on immediate free recall (IFR; top) and continuous-distractor free recall (CD; bottom). Lines with open circles represent the end-first condition, and lines with filled circles represent the beginning-first condition. The top left panel is adapted from “Effects of Output Order on Immediate, Delayed, and Final Recall Performance,” by J. J. Dalezman, 1976, *Journal of Experimental Psychology: Human Learning and Memory, 2,* p. 599. Copyright 1976 by the American Psychological Association. The bottom left panel is adapted from “Output Interference and Long-Term Serial Position Effects,” by W. B. Whitten, 1978, *Journal of Experimental Psychology: Human Learning and Memory, 4,* p. 688. Copyright 1978 by the American Psychological Association.

formance levels for all positions in continuous-distractor free recall but for only prerecency positions in immediate free recall.

We modeled the amnesia deficit by assuming a partial disconnection between the contextual and the lexical–semantic systems. Specifically, connections between context and lexical units were set to 0 for 50% of the context units, reflecting hippocampal damage (see Appendix A for simulation details).

Figure 9 presents the model's results averaged over 1,000 simulations. The model shows a good qualitative correspondence with a single parameter modification (the 50% damage to the context connections). As in the data, the last few serial positions in immediate free recall are unaffected by the "lesioning" of the context system, whereas prerecency positions in immediate free recall show lower performance levels. As in the data, in continuous-distractor free recall, we find a parallel drop in performance compared with that of the "nonlesioned" model. This is entirely because, with the exception of the recency items in immediate free recall, all items are retrieved from episodic memory.

Simulation 5: Negative recency. Negative recency in final free recall is observed for immediate free recall (e.g., Craik, 1970) but not for continuous-distractor free recall (Bjork & Whitten, 1974; Glenberg et al., 1980; Koppenaal & Glanzer, 1990; Whitten, 1978; Tzeng, 1973). It is generally assumed that recall performance in the final free recall test is based primarily on the traces in episodic memory. In the model, we examined this by measuring the episodic strengths of the list items in immediate and continuous-distractor free recall. Whereas in the continuous-distractor task the strength values are independent of serial position (as each item is preceded and followed by a distractor interval), in immediate free recall, there is a small one-item primacy effect in the episodic strengths, due to the first item entering an empty buffer and reaching particularly high levels of activation. This provides an additional contribution to primacy besides that due to the retrieval of the start-cue context. More important, unlike in the continuous-distractor task, in immediate free recall, the strength values of items in the last serial positions are smaller than those of other

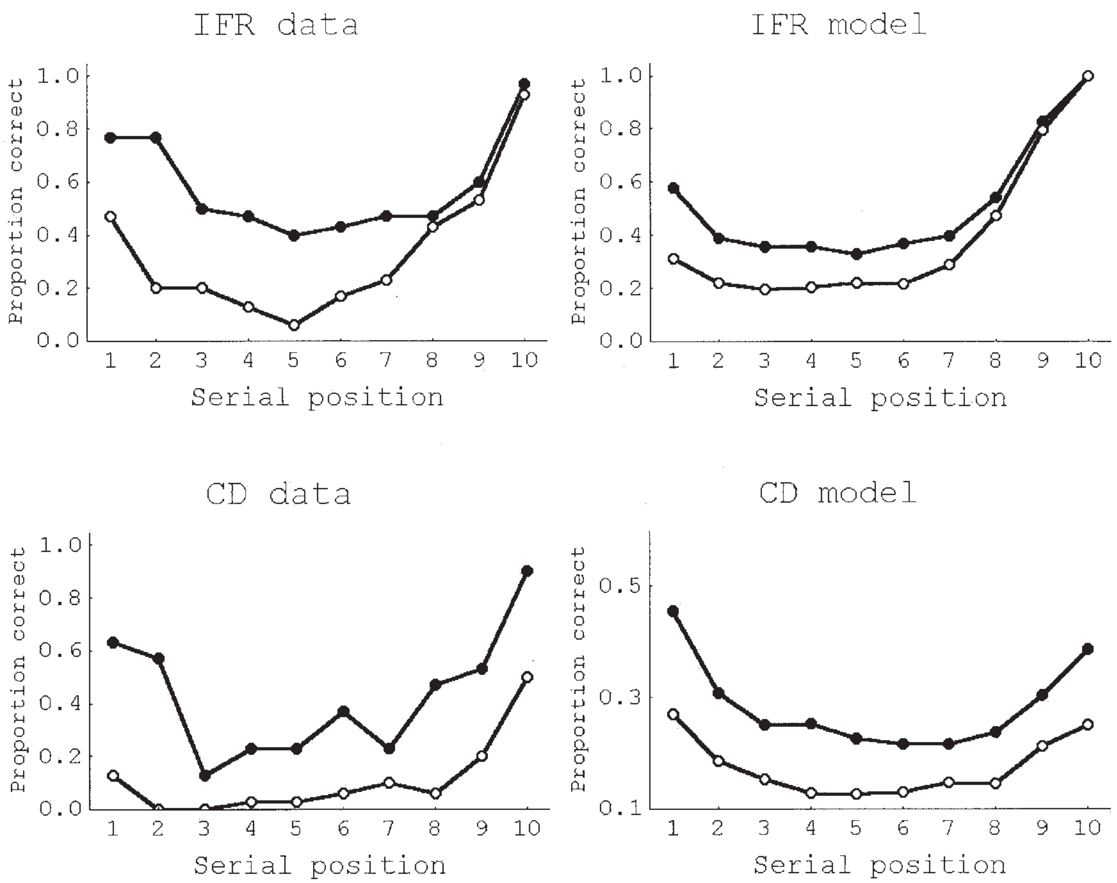

Figure 9. Data from Carlesimo et al. (1996; top left and bottom left) and model simulations (top right and bottom right) for the recall performance of amnesic patients (lines with open circles) compared with a control group (lines with filled circles). Top: Immediate free recall (IFR). Bottom: Continuous-distractor free recall (CD). The top left and bottom left panels are reprinted from *Neuropsychologia, 34,* G. A. Carlesimo, G. A. Marfia, A. Loasses, and C. Caltagirone, "Recency Effect in Anterograde Amnesia: Evidence for Distinct Memory Stores Underlying Enhanced Retrieval of Terminal Items in Immediate and Delayed Recall Paradigms," pp. 177–184, Copyright 1996, with permission from Elsevier.

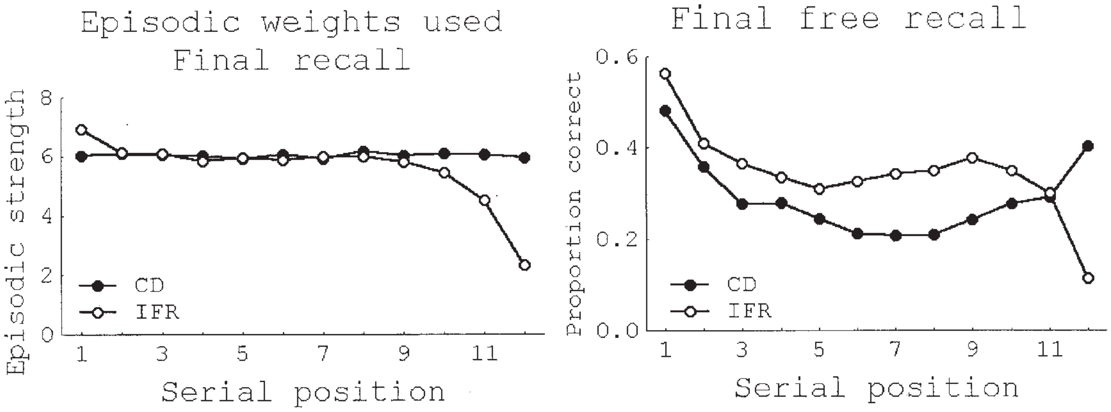

Figure 10. Left: Values of episodic strengths as a function of serial position in immediate free recall (IFR) and continuous-distractor free recall (CD). Right: Serial-position functions for final free recall after a trial of IFR or CD.

items, demonstrating the negative recency dissociation (see Figure 10, left). This is due to the assumption that encoding continues only until the recall prompt (e.g., t = 6,000 in Figure 3). Because in a final free recall test only these episodic weights are available to drive retrieval (as the buffer does not contain list items), this leads to a negative recency effect after a series of trials with immediate free recall. Figure 10 (right) shows serial-position functions of trials in which only the episodic strengths are used for retrieval. Consistent with the data, negative recency is obtained in a final free recall test after immediate free recall but not after continuous-distractor free recall.

Simulation 6: Interaction between output position, task, and lag recency. The asymmetry in lag recency—that is, the greater probability for retrieving items from nearby than from remote serial positions—is more pronounced in the first few than in later output positions in immediate free recall but not in continuous-distractor free recall. Figure 11 presents lag-recency functions for both immediate and continuous-distractor free recall (of 12-item lists). These were computed (as in Howard & Kahana, 1999) across all output positions (lines with filled circles). The lag-recency function computed for the first two reported words alone (first-recall transition) is also given (lines with open circles). The asymmetry is stronger for immediate free recall than for continuous-distractor free recall.

The model captures the interaction between the tasks, output position, and lag recency. In immediate free recall, the asymmetry is much stronger than in continuous-distractor free recall. This is due to the interaction between the buffer and the episodic system in immediate free recall; for the first few items (the buffer items), the order of report is from strong to weak episodic strength (see Appendix B for justification). Because of the negative recency in episodic strengths (see Figure 10), this results in an increase in the forward bias, which makes the asymmetry of lag recency larger between the first two output positions. As opposed to this, in continuous-distractor free recall, in which no negative recency exists (and no items are reported from the buffer), the forward bias in lag recency does not change for the first output positions. This is seen in the similar lag-recency functions for the first-recall transition and across all output positions. This analysis is consistent with Kahana's (Howard & Kahana, 1999; Kahana, 1996) suggestion that a short-term buffer underlies the interaction between output position and lag recency.[10]

Although the main reason for using a changing context component was to obtain long-term recency effects, the lag-recency results are consistent with the TCM framework (Howard & Kahana, 2002), suggesting that a changing context is a parsimonious way to address both long-term recency and lag-recency effects. The addition of the activation buffer, however, helps to further account for the interaction between lag recency and output order.

Discussion of Dissociation Simulations

So far, the model has been able to account for the dissociations between immediate and continuous-distractor free recall. The existence of these dissociations weakens the argument that short- and long-term recency effects can be explained through a single mechanism. We now summarize our model's account of these dissociations.

The first dissociation is the absence of short- but not long-term recency when participants are instructed to start their recall with items from the beginning of the list. The model accounts for this by assuming that in immediate free recall, retrieval of items from the beginning of the list displaces the items in the activation buffer, thereby eliminating the short-term recency effect. As the buffer is not involved in retrieval in the continuous-distractor task, directed output order does not affect long-term recency.

---

[10] We should note that the model overestimates the asymmetry in lag recency in immediate free recall for the first-recall transition. This is due to the assumption that the reporting order for items in the buffer is always in the order of item context strength. Through the introduction of noise to this retrieval process, the asymmetry can be weakened (although the dissociation would remain). The simulation therefore presents an illustration of the maximum possible asymmetry in lag recency for immediate free recall.

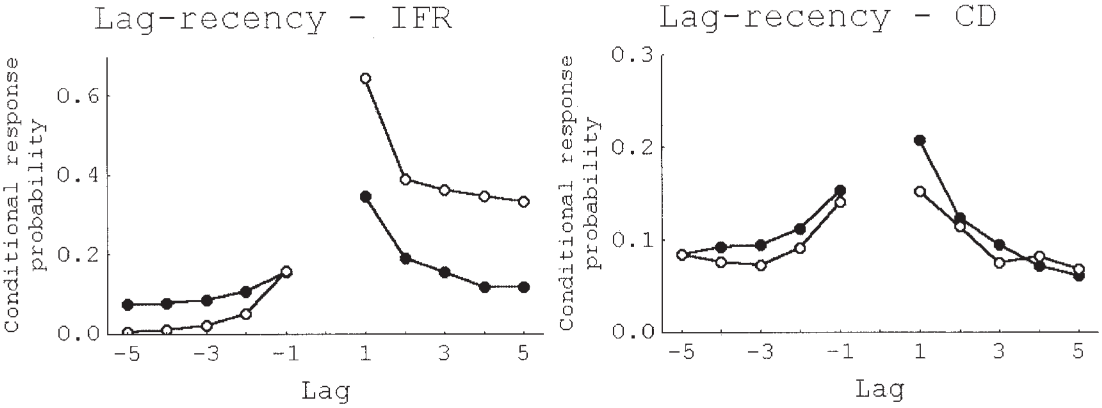

Figure 11. Lag-recency functions for immediate free recall (IFR; left) and continuous-distractor free recall (CD; right). Lines with solid circles represent lag-recency functions computed over all output positions, and lines with open circles represent lag-recency functions of the first-recall transition.

The second dissociation is the decreased recall performance for recency items in continuous-distractor free recall but not in immediate free recall in amnesic patients. The model explains these results in terms of the difficulties that amnesic patients have in episodic memory processes like encoding and retrieval while having an intact short-term buffer (Baddeley & Warrington, 1970). In continuous-distractor free recall, all items are retrieved from episodic memory and are, therefore, affected in amnesic patients. In immediate free recall, however, only the prerecency items are affected, as they are the ones that are retrieved from episodic memory, whereas the recency items reside in the intact short-term buffer.

The third dissociation between immediate and continuous-distractor free recall is that in a final free recall task, a negative recency effect is found for immediate free recall but not for continuous-distractor free recall. To account for this finding, the model suggests that in final free recall, retrieval relies on the strength of the episodic traces that have been laid down during study. Because in continuous-distractor free recall the strengths of all traces are equal, a negative recency effect is not found. In immediate free recall, however, the strengths decrease toward the end of the list, leading to negative recency. Moreover, in immediate free recall, the contextual contribution to recency is not sufficiently strong to override the negative recency profile in the strengths of final list items.

The fourth dissociation is that lag-recency functions differ across the first output positions in immediate but not in continuous-distractor free recall. The model captures this dissociation by assuming that items in the buffer in immediate free recall are reported in a predominantly forward manner according to their episodic strengths. However, in the continuous-distractor task the buffer does not play any role and all traces are of equal strength. Therefore, the model does not predict any such interaction between output position and lag recency.

The reported results form a double dissociation between short- and long-term recency, for which our model provides a parsimonious account by assuming a critical contribution made by a short-term buffer. Indeed, these dissociations, together with the dissociation that is described in Experiment 1, meet the challenge set forth by Broadbent (1971), who pointed out that

> In general, one must be aware of concluding that the appearance in short-term memory of an effect known from longer-term studies is evidence for identity of the two situations. . . . Only success or failure of attempts to show differences between the two situations is of interest in distinguishing the theories. (pp. 342–343)

Given that our context–activation model has accounted for the critical data, we now turn to describe two further predictions that rely on the postulated existence of a short-term buffer.

Predictions of the Model

Proactive Interference

The neuropsychological dissociation between short- and long-term recency was localized at the episodic component that is used to retrieve prerecency items. The buffer component is postulated to be intact, and therefore, the short-term recency effect is spared. As discussed in the introduction, another effect that is present for prerecency but not for recency items in immediate free recall is proactive interference (Craik & Birtwistle, 1971). Dual-store theories can account for this finding by assuming that, as in the amnesic syndrome, proactive interference affects the retrieval from the long-term store. As such, the short-term recency effect, which is due to unloading from the short-term buffer, should be unaffected by proactive interference. With regard to continuous-distractor free recall, all items are retrieved from the long-term store, and so proactive interference was predicted to occur at all serial positions, including recency positions.

We simulated 1,000 pairs of trials in immediate and continuous-distractor free recall. In each pair, the start context for the second trial was the context that was active at the end of the retrieval of the first trial. There are two sources for proactive interference in the model. First, items in the two lists can be associated with the same context unit (because the random walks overlap), and the item–context associations formed during encoding of List 1 items

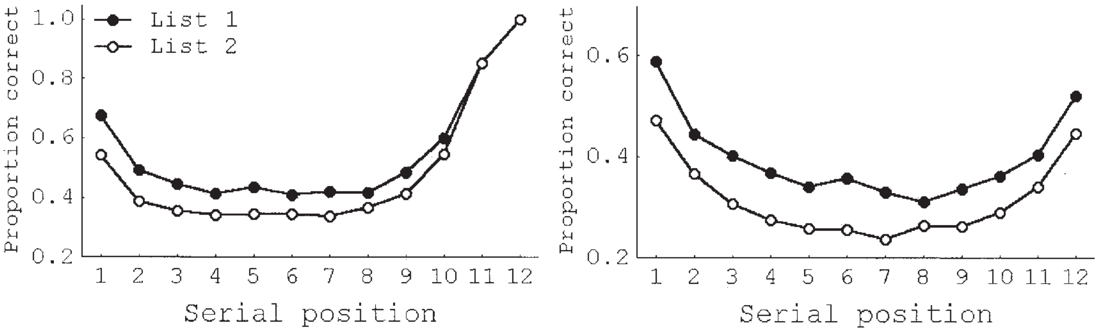

Figure 12. Model predictions for the presence of proactive interference in immediate free recall (left) and continuous-distractor free recall (right). The serial-position function of List 2 is compared with that of List 1 and shows that the proactive interference manipulation affects recency positions only in the continuous-distractor task.

were maintained during the encoding of List 2 items. As a result, during episodic retrieval of List 2 items the context units used to retrieve List 2 items may also retrieve those items from List 1 with which they are associated. Second, in typical proactive interference experiments, all items are drawn from the same semantic category. To simulate this, during the selection and recovery stages of retrieval,[11] we gave all items an additional semantic input (the contribution was chosen to be about one third of the typical episodic trace strength). Both sources lead List 1 items to intrude during retrieval of List 2 items (see Appendix A for the simulation protocol).

As can be seen in Figure 12, in immediate free recall, the model exhibits proactive interference for prerecency positions but not for recency positions. In contrast, in continuous-distractor free recall, the model shows proactive interference at all positions including the recency positions. The model, therefore, predicts another dissociation between short- and long-term recency, which forms a conceptual analogue to the neuropsychological dissociation above, in that both dissociations are based on a variable that affects retrieval from the long-term store.

A related prediction of the model concerns the conditions under which the dissociation as a function of proactive interference should emerge. Our model predicts that the dissociation with proactive interference should not be found if recall begins with items from the beginning of the list (the beginning-first condition) but should be found if recall begins with items at the end of the list. The reason for this is that when recall starts with items from the beginning of the list, recency items are recalled only relatively late in the recall phase and hence, if in fact recalled, are retrieved only from episodic memory. Figure 13 shows simulation results of the effect of proactive interference on the immediate recall of two lists when recall starts with items presented at the beginning of the list (all parameters were the same as those in the previous simulations). Critically, the recency effect is absent in this beginning-first condition (which is in accordance with the analysis under Simulation 3). That proactive interference is not present in immediate free recall in the end-first condition, even for recency items, places an important constraint on the conditions for which a dissociation between short- and long-term recency is observed with proactive interference—that is, this dissociation is observed only when recency items are recalled first.

Presentation Rate

When exploring the parameter space of the activation-based buffer, we noticed that the model predicts an interaction between presentation rate and serial position: a shift from recency to primacy with an increase in presentation rate. Here, we describe the mechanism that we argue to be responsible for this shift. To illustrate this shift, we ran 1,000 simulation trials with lists of 12 items for four different presentation durations, in which presentation duration corresponded with the number of iterations that an item representation receives sensory input. We kept the other parameters the same as in the other simulations and disabled the episodic component in order to detect the pure contribution of the buffer. As the episodic component contributes very little at fast presentation rates, the results (for fast presentation rates) do not change when the full model (buffer plus episodic component) is used. However, at slow presentation rates, the episodic component is expected to add a baseline to recall performance that is independent of the shift from recency to primacy. Here we focus only on the buffer prediction involving this shift.

Figure 14 shows serial-position functions of the probability that an item is in active memory at time of test for the different presentation rates. This may correspond to a test of cued recall (if the cue uniquely specifies the item) or to a hypothetical test of free recall in which all active items could be reported. What is immediately striking is that with an increase in presentation rate, the recency profile turns into a primacy profile. Note that this primacy profile is not due to episodic encoding or retrieval (as in immediate free recall under a slow presentation rate) but has its source solely within the short-term buffer.

The mechanism responsible for the shift from recency to primacy in the model can be understood as follows. Activated representations corresponding to items presented in the memory list compete with each other because of the global inhibition in the buffer. Therefore, because it takes time to build up activation, the maximum activation level that each item can reach depends on the presentation duration. With a slow presentation rate (see Figure 15, top), each item can

---

[11] Because all list items belong to the same semantic category, we decided to add a general category bias rather than include interitem associations (α_2), as they are effectively equivalent.

overcome the inhibition of preceding active items, reaching a higher level of activation, and eventually displace earlier items from the buffer. With a faster presentation rate (see Figure 15, bottom), however, less time is available for each item to reach the level of activation of the preceding items, leading to a relative disadvantage for later items compared with preceding items. Therefore, early items are not displaced by later items, leading to a primacy profile.

The shift from recency to primacy as function of the presentation rate is in stark contrast to the predictions of a whole family of models that view recency in list memory as a characteristic of the retrieval process based on temporal discriminability (Crowder, 1976; Neath, 1993b; Tan & Ward, 2000; Ward, 2002). Although these models may show some sensitivity to presentation rate, they do not predict such a dramatic shift. In fact, the default assumption of temporal discriminability models is that recall is determined by the temporal scale–invariance ratio rule (Crowder, 1976; Neath, 1993b). Accordingly, recall is a function of the ratio of the durations of the IPI and the RI at the end of the list, which are unaffected by presentation rate. The recency-to-primacy shift also contrasts with buffer models that use a first-in, first-out (knock-out model; Kahana, 1996; Philips, Shiffrin, & Atkinson, 1967) or random (Kahana, 1996; Raaijmakers & Shiffrin, 1980, 1981) displacement process. In these models, the buffer is insensitive to presentation rate and always contains the most recently presented items (for further discussion of these issues, see the Utility of the Dynamic Buffer section).

Experimental Tests

The first experiment addressed the model's prediction of a dissociation between short- and long-term recency with proactive interference. The second experiment focused on the prediction of an interaction between presentation rate and serial position. As discussed below, both predictions were confirmed.

Experiment 1: Proactive Interference

As discussed in the introduction, proactive interference has been found to affect prerecency items only in immediate free recall (Craik & Birtwistle, 1971). One interpretation of this result is that proactive interference affects only items that are retrieved from episodic memory. Specifically, the retrieval of prerecency List 2 items is negatively affected by intrusions of List 1 items, both of which are retrieved from episodic memory. The recency List 2 items, however, are immune to proactive interference, because these items are unloaded from a short-term buffer that is not affected by proactive interference. According to this interpretation, in continuous-distractor free recall, in which all items are retrieved from the episodic system, proactive interference should affect performance at all serial positions, even for recency items. This was indeed the prediction of our model. Hence, a dissociation is predicted between short- and long-term recency.

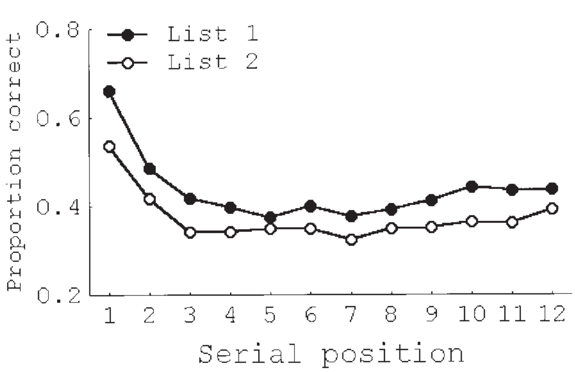

Figure 13. Model predictions of proactive interference in immediate free recall under beginning-first instruction. Note that the recency effect is eliminated and that proactive interference occurs even at recency positions.

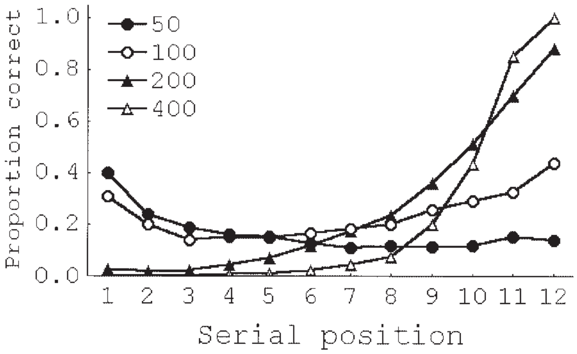

Figure 14. Model predictions for the effect of presentation rate on the serial-position profile for four rates (measured by the number of iterations per item). Profiles represent the probability that an item at that position is active above threshold at test.

Although we found no explanation of Craik and Birtwistle's (1971) findings by single-store theorists, proponents of such a view might argue for a different interpretation of the proactive interference dissociation between recency and prerecency items in immediate free recall, leading in turn to a different prediction concerning the effect of proactive interference in the continuous-distractor task. According to this hypothetical interpretation, in immediate free recall, recency items may be immune to proactive interference not because they are unloaded from a short-term buffer but because of their high level of temporal distinctiveness (Bjork & Whitten, 1974; Crowder, 1976; Glenberg et al., 1983). One should then expect no dissociation with proactive interference between immediate free recall and the continuous-distractor task because in both tasks, recency items are temporally distinct.

It is unclear at the moment whether such a temporal distinctiveness interpretation can indeed be supported by computational models of temporal context, because it is not a priori obvious that the higher overlap between the retrieval context and the encoding context of recency items would render these items immune to proactive interference. Indeed, our particular context–activation model, in which the temporal context component is responsible for long-term recency, predicts that recency items should not be immune to proactive interference in the continuous-distractor task.[12] Still, we do not want to exclude the possibility that a different model of contextual retrieval could be formulated that would show immunity from proactive interference at recency in both immedi-

---

[12] In our model, the random walk of the context and the switch to the start context during retrieval lead to contextual overlap between items in consecutive lists. This in turn leads to intrusions of List 1 items during retrieval of List 2 items.

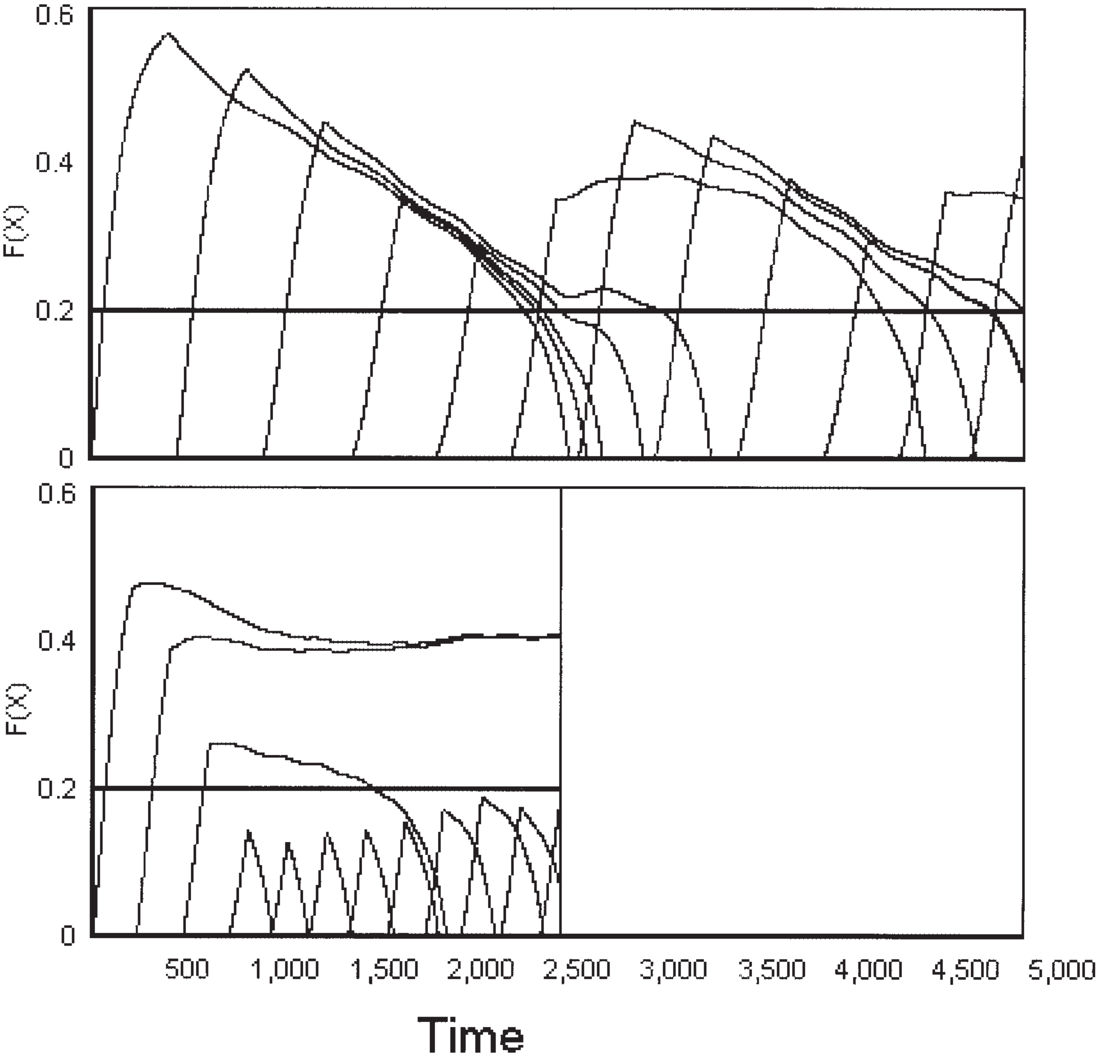

Figure 15. Activation trajectories for two presentation durations. Top: Slow presentation rate (400 iterations per item). Bottom: Fast presentation rate (200 iterations per item). F(*x*) = output activation value.

ate free recall and the continuous-distractor task without relying on a short-term buffer. Therefore, our manipulation of proactive interference (in two tasks within the same experiment) is important, not only for verifying the validity of our model but also to constrain future models of list memory that address the debate regarding the existence of a short-term buffer in immediate free recall.

In summary, the purpose of Experiment 1 was to determine whether temporal context is a general mechanism underlying both short- and long-term recency effects that can render recency items immune to proactive interference. An association between short- and long-term recency (in terms of the effect of proactive interference) would, therefore, support a single-store model of memory. Alternatively, temporal context may provide an incomplete account for recency effects, and a short-term store needs to be added to account for dissociations in recency effects in immediate and continuous-distractor free recall. A dissociation between the two tasks (in terms of the effect of proactive interference) would, therefore, support a dual-store model of memory and, in particular, a context–activation type model.

In a pilot study, Goshen-Gottstein, Ashkenazi, and Usher (2000) manipulated the task (immediate vs. continuous-distractor free recall) between participants and used an IPI:RI ratio of 10:30 in continuous-distractor free recall. A significant triple interaction (Task × List × Position) was obtained (experimental design and data graphs are available from http://freud.tau.ac.il/~goshen/data.htm), with proactive interference affecting only prerecency items in immediate free recall but affecting both prerecency and recency items in continuous-distractor free recall. In the pilot study, the number of possible confounds in design, materials, and procedure was kept to a minimum through the use of both tasks in a single study. Still, the IPI:RI ratio was close to unity in immediate free recall but was smaller than unity in the continuous-distractor task, and this confound may have mediated the dissociation. Also, it is possible that despite the random allocation procedure, participants who were allocated to the continuous-distractor task were more susceptible to proactive interference than were those who were allocated to the immediate free recall task (for research on individual differences in the susceptibility to proactive interference, see Kane & Engle, 2000). To overcome the earlier criticisms, in the current experiment we used a within-subjects design and set the IPI equal to the RI in the two tasks. Following the typical procedure of proactive interference studies (e.g., Craik & Birtwistle, 1971), we used lists of words chosen from the same semantic categories in consecutive trials, as this manipulation was shown to increase proactive interference.

Method

Participants. A total of 31 Tel-Aviv University undergraduates, ages 22–28 with normal or corrected-to-normal vision, participated in the experiment for course credit.

Design and materials. The design crossed the within-subject factors task (immediate and continuous-distractor free recall), position (1–12), and list (first, second). The two tasks were presented in separate blocks, with the order of the blocks counterbalanced across participants.

There were 10 pairs of critical lists (for a total of 20 lists), 5 pairs for each task. One additional pair of practice lists preceded each of the blocks. All lists contained 12 words. To maximize proactive interference between the first and second lists, we created each pair of lists such that it contained words from the same semantic category. To minimize interference across list pairs, we ensured that the semantic categories of each pair were unique and differed not only from other list pairs but also from the practice-list categories (i.e., release from proactive interference; Wickens, Born, & Allen, 1963). The two practice trials contained words from different semantic categories to acquaint participants with changes in the semantic category.

The 10 pairs of lists were separated into two different sets, with each set containing 5 different pairs of critical lists. The two sets were counterbalanced across participants so that each set appeared an equal number of times in immediate and continuous-distractor free recall. The order of the lists (i.e., List 1, List 2) within the pairs was also counterbalanced, such that across participants each list appeared an equal number of times as the first list and as the second list in both tasks. The words within each list were randomized for each participant. All the words were recorded in a male voice and were judged by two judges for clarity. The volume of the auditory presentation was kept constant during the entire experiment (measured at 45–55 dB).

For the continuous-distractor task, the distractor activity consisted of solving mathematical problems for 15 s following the presentation of each of the words as well as prior to the presentation of the first word (i.e., IPI = RI = 15 s). The problems consisted of the addition or subtraction of two single-digit numbers (e.g., $3 + 4 =$) that were displayed on the computer monitor, for which the result was always a positive value between 1 and 9. The numbers were presented in 48-point font.

Procedure. The instructions of the immediate free recall task and the continuous-distractor task were presented on the monitor and read aloud by the experimenter prior to administration of the corresponding task. For both tasks, participants were told that they would hear a number of lists of words presented by the computer. Prior to hearing any list, they were told which category the items belonged to (e.g., "The following list includes names of vegetables").

For the continuous-distractor task, participants were told that immediately after the presentation of each word (and prior to the presentation of the first word), mathematical problems would be displayed on the monitor, one immediately after the other. The position of the problem alternated between 2.5 and 3.0 cm from the top of the screen (and was centered horizontally). Participants were asked to read aloud the problems and to state the solutions (verbally and by typing on the numerical keypad) to as many of the problems as possible. Each mathematical problem was displayed separately and remained on the screen either until the answer was typed or until the next word was delivered, whichever came first. It was emphasized that the mathematical task and the memory task were both important and that for the mathematical task, speed and accuracy were equally important.

After the final word (immediate free recall) or problem (the RI of the continuous-distractor free recall), the word *recall* appeared on the computer screen, prompting the participant to write down as many of the words that he or she could recall within 1 min on a blank page given by the experimenter (one page per list). Participants were not informed of the practice lists, and as far as they knew, all of the lists were test lists.

During the retrieval interval, a small clock that appeared on the lower side of the monitor showed the remaining time for the retrieval interval. At the end of each retrieval interval, the experimenter took the paper from the participant, and the following category name was announced (e.g., "The following list includes names of vegetables"). Immediately afterward, the participant pressed a key to start presentation of the next list.

Results

The simulations indicated that the dissociation between short- and long-term recency would be largest when recall starts with items from the end of the list. As such, only those trials (i.e., pairs of lists) in which participants started with items from the second half of the list on both lists (henceforth, *useful trials*) were included in the analysis. This procedure led to differences in the number of useful trials that participants contributed in each of the tasks. To overcome disproportionate contributions, we weighted every participant according to the lowest number of useful trials between the two tasks. For example, a participant who contributed four out of five useful immediate free recall trials and three out of five useful continuous-distractor free recall trials was weighted at 3.

The participants in the continuous-distractor group performed almost at ceiling on solving the math problems. Figure 16 presents descriptive serial-position curves, which, for the purpose of presentation, were smoothed by averaging each score with the preceding and following scores (in the actual analysis, the raw data were analyzed). Examination of Figure 16 reveals that the overall level of performance on the second list was lower than that on the first, in both tasks, establishing that we were successful in inducing proactive interference. In addition, only the recency positions in immediate free recall were unaffected by proactive interference.

The weighted contributions were submitted to a 2 (task) × 2 (list) × 12 (position) mixed analysis of variance (ANOVA). This yielded significant main effects of task, $F(1, 63) = 23.35$, $MSE = 0.173$, $p < .001$; list, $F(1, 63) = 76.65$, $MSE = 0.064$, $p < .001$; and position, $F(11, 693) = 68.88$, $MSE = 0.054$, $p < .001$. Of the interactions, only the interaction between task and list was marginally significant, $F(1, 63) = 3.86$, $MSE = 0.050$, $p < .055$, whereas all other interactions were significant: Task × Position, $F(11, 693) = 3.77$, $MSE = 0.054$, $p < .001$; List × Position, $F(11, 693) = 2.15$, $MSE = 0.056$, $p < .05$; and Task × List × Position, $F(11, 693) = 2.71$, $MSE = 0.056$, $p < .001$.

When the averaged recall performance for the middle (Serial Positions 5–8) and end (Serial Positions 9–12) clusters were submitted to the same analysis, all main effects remained significant: task, $F(1, 63) = 25.86$, $MSE = 0.026$, $p < .001$; list, $F(1, 63) = 39.18$, $MSE = 0.016$, $p < .001$; and cluster, $F(1, 63) = 279.96$, $MSE = 0.028$, $p < .001$. Of the two-way interactions, only

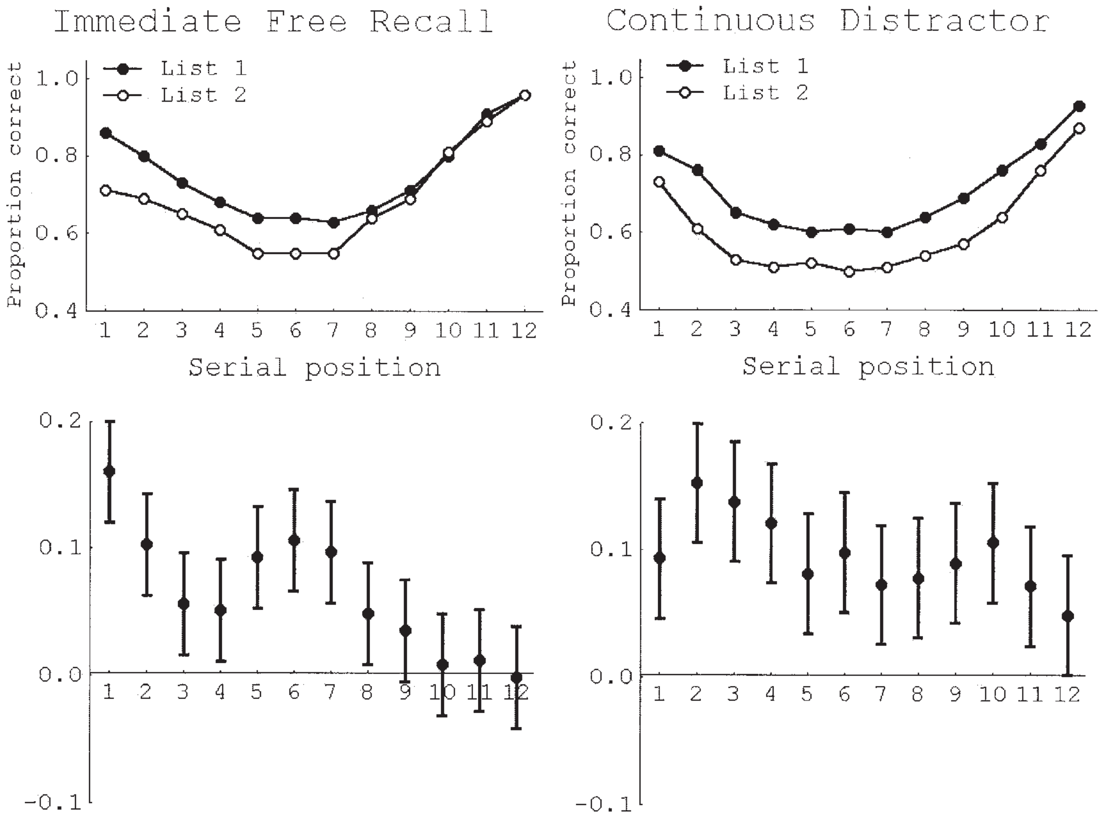

Figure 16. Results of Experiment 1: Serial-position functions of the first and second lists of a pair. Left: Immediate free recall. Right: Continuous-distractor free recall. Differences in performance between the second and first lists are plotted in the bottom graphs. Error bars represent 95% confidence intervals for within-subject designs (see Loftus & Masson, 1994).

the interaction between list and cluster, *F*(1, 63) = 4.99, *MSE* = 0.014, $p < .05$, was significant. Most important, the three-way interaction reached significance, *F*(1, 63) = 6.99, *MSE* = 0.014, $p < .05$.

To understand the nature of this triple interaction, we conducted separate analyses for immediate and continuous-distractor free recall. These analyses revealed that the triple interaction was due to the presence of a significant two-way interaction between list and cluster in immediate free recall, *F*(1, 63) = 15.13, *MSE* = 0.011, $p < .001$, but not in continuous-distractor free recall, *F*(1, 63) < 1, *MSE* = 0.017, *ns.* In immediate free recall, there was an effect of list at middle but not end positions, whereas in continuous-distractor recall, there was an effect of list at both middle and end positions (see the confidence intervals in Figure 16). Together, these results establish that the proactive interference manipulation affected all positions in continuous-distractor free recall but only prerecency positions in immediate free recall.

Discussion

In this experiment, we obtained short- and long-term recency effects in immediate and continuous-distractor free recall, respectively. Most important, the results indicate that short- but not long-term recency is immune to proactive interference. This is consistent with the view that long-term recency depends on retrieval from episodic memory, which is affected by manipulations such as proactive interference. In contrast, short-term recency is thought to be predominantly dependent on retrieval from a short-term buffer and is, therefore, not affected by the proactive interference manipulation.

It is important to point out that the triple interaction was significant when those trials were included in the analysis in which the participant started the recall phase with items from the second half of the list. When we included all trials in which the participants started with items from the beginning of the list, the triple interaction was not significant in both the analyses of the full (2 × 2 × 12 ANOVA), *F*(11, 330) < 1, *MSE* = 0.033, *ns,* and averaged (2 × 2 × 2 ANOVA), *F*(1, 22) = 1.10, *MSE* = 0.013, *ns,* data. Indeed, our computational model had predicted that there would be a dissociation only if items from the second list half were reported first. That this constraint was borne out by the data provides additional support for the validity of the model and demonstrates its usefulness in informing experimental design.

The model accounts for the dissociation in terms of the reliance of long-term recency on retrieval from episodic memory, a memory system that is susceptible to proactive interference. On the other hand, in immediate free recall (in which recall starts with items from the end of the list), recency items are unloaded from an activation-based short-term buffer, and therefore, it is only these items that are immune to proactive interference. As such, any dual-store model could account for the dissociation when both contextual retrieval and unloading from a buffer are built into the design. Still, our model is unique in that it details the nature of the contribution of the short-term buffer as well as the contextual system and provides an exact account of how these two components contribute in their unique way to short- and long-term recency effects. Nonetheless, in Experiment 2 we focus on a manipulation that affects the fine balance between excitation and inhibition within the activation-based buffer. This manipulation would provide support for our view that the short-term buffer itself needs consideration beyond that of a single parameter for buffer capacity (e.g., Raaijmakers & Shiffrin, 1980; see also Appendix C). We thereby address the mechanisms of the buffer and its defining property: capacity limitations.

Experiment 2: Presentation Rate

Experiment 2 was designed to test the context–activation model's prediction that a shift from recency to primacy would occur when the presentation duration for items is shortened. This prediction is particularly important because existing single- and dual-store theories do not predict this shift in the profile; they predict a recency function for all presentation durations. As discussed earlier, this shift is mediated by the dynamics of the global inhibition and self-excitation in the activation buffer. Although previous research has reported shifts from recency to primacy induced by increasing the duration of the RI (e.g., Neath & Crowder, 1990), the effect of presentation rate on the amount of recency and primacy has not yet been investigated.

As we tried to estimate the dynamics of the activation buffer with regard to serial position of items in the list, we chose an experimental setup that maximizes the contribution of the buffer while minimizing the contribution of episodic memory and, at the same time, provides a clean version of serial-position information. To this end, and following the results of Experiment 1, we set the proactive interference to a relative high level by using a small word pool with replacement. Furthermore, we also imposed a deadline for recall that penalized slow episodic retrieval processes (Waugh, 1970).

An additional change in procedure was made in this experiment because the serial-position functions in tasks such as free recall are affected by factors such as output interference and recall strategies (which are difficult to model; but see Gronlund & Shiffrin, 1986). Factors such as output interference and recall strategies may obscure the memory availability of items from different serial positions at the moment of recall. To reduce the influence of such factors, we used cued recall as the retrieval task. In this task, a single serial position is probed per list, eliminating output interference and constraining the recall strategy.

Method

Participants. Twenty undergraduates (age range = 19–30 years) from the University of London, London, participated in the experiment in exchange for £5 (U.S.$9). All participants were native speakers of English, were right-handed, and had normal or corrected-to-normal vision.

Design and materials. The experiment conformed to a 4 × 6 within-subject design, with presentation rate (100, 200, 400, and 800 ms) and serial position as independent variables. Recall accuracy was measured as function of serial position.

Twenty-four words taken from six different semantic categories formed a word pool from which the lists were constructed with replacement. Every list had one word from each category. The words, the probed position, and the probed category were not repeated on consecutive trials. The presentation of the trials at the different rates was blocked such that in each block, all serial positions were probed five times, with the probed position and category randomly varied across trials. The order of the presentation rate conditions was counterbalanced across participants. Each participant completed 30 trials at every presentation rate.

Procedure. Participants were given instructions on the screen as well as verbally by the experimenter. Participants were shown the category names and exemplars for 1 min before the practice trials. The experiment had a total of 16 practice trials plus 120 experimental trials. Before each block, 4 practice trials were given at the presentation rate of that block. A trial started with a fixation stimulus (+++, for 1 s) accompanied by an alerting beep, which was followed by the words of the trial, presented one at the time at one of the fixed durations of 100, 200, 400, or 800 ms. After the last word, a category name was presented, prompting the participant to verbally recall the item that had been presented in the list belonging to the cued category within 1.5 s (a second beep was presented after this time). The experimenter recorded the verbal response. The participant initiated the next trial by pressing the space bar.

Results

Experiment 2 tested the prediction of the context–activation model by varying the presentation duration through four levels. The model predicts that to obtain the shift from recency to primacy, the rate of presentation should be very fast. To quantify the primacy–recency gradient, we calculated a *primacy–recency index* (PR_{index}) from the sum of the multiplications of the probability of correct recall $P(i)$ at position i, with the position number normalized for the sum of probabilities and the list length L, as in

$$PR_{index} = \frac{\Sigma P(i) \cdot i}{(L + 1)\Sigma P(j)} . \tag{5}$$

This index, which has a value between 0 and 1, indicates the relative degree of recency and primacy. Values larger than .5 correspond to greater recency compared with primacy, and the reverse is true for values smaller than .5.[13]

Two statistical analyses were conducted. In the first, recall probabilities were entered as function of serial position and presentation rate. The second tested the PR_{index} as a function of the presentation rate, which is a more direct test of the hypothesis regarding whether the amount of primacy and recency is affected by presentation rate.

The results are shown in Figure 17. As the main focus of the experiment was on the two extreme rate conditions, the two middle rates are combined for the purpose of presentation. The analysis, however, was performed on the full factorial design. Figure 17

---

[13] The exact boundaries of the index are $1/(L + 1)$ and $L/(L + 1)$. The denominator assures that when there is as much primacy as there is recency, the index is .5. To see this, let us assume a constant (flat) serial-position function. One can easily check from Equation 5 that the PR_{index} equals .5, corresponding to a serial-position function with neither recency nor primacy. Similarly, PR_{index} = .5 for every serial-position function that is symmetrical around the middle point $[(L + 1)/2]$.

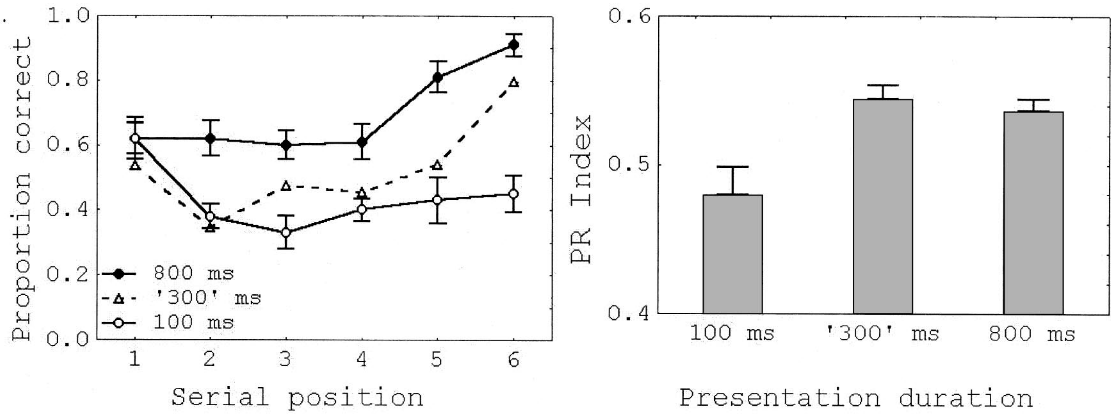

Figure 17. Results of Experiment 2: Presentation rate effects in category-cued recall. Presentation rates were 100, 200, 400, and 800 ms; the middle rates are combined ('300') in the figure. Left: Effect of presentation rate on the serial-position function. Right: Effect of presentation rate on the primacy–recency (PR) index. Bars represent standard errors.

shows the serial-position functions (left) and the PR_{index} (right) as function of presentation rate. It can be seen that the total recall performance decreased with the increase in presentation rate. More important, the slowest condition showed recency, whereas the fastest condition showed a shift to primacy. The middle presentation conditions fell approximately between the two extremes. The PR_{index} shows an abrupt drop, indicating a shift toward primacy at the fastest presentation condition.

A 4 (rate) × 6 (position) repeated measures ANOVA revealed a main effect of rate, $F(3, 57) = 47.82$, $MSE = 0.034$, $p < .001$; a main effect of position, $F(5, 95) = 16.71$, $MSE = 0.061$, $p < .001$; and the predicted interaction between rate and position, $F(15, 285) = 3.38$, $MSE = 0.046$, $p < .001$. Adjusted t tests (significance level at $.05/8 = .006$) revealed that when the first (for primacy) and last (for recency) items were compared with the average of the two middle items, only the 800-, 400-, and 200-ms conditions showed recency, $t(19) = 5.75$, $p < .001$, $t(19) = 5.45$, $p < .001$, and $t(19) = 7.60$, $p < .001$, respectively, whereas only the fastest (100-ms) condition showed primacy, $t(19) = 3.42$, $p < .005$. A second ANOVA conducted on the PR_{index} revealed a significant decrease with increase in presentation rate, $F(3, 57) = 7.20$, $MSE = 0.003$, $p < .001$, which was due to the abrupt drop between the 200- and 100-ms conditions, $t(19) = 3.94$, $p < .001$.

Discussion

The data demonstrated that with an increase in presentation rate, the serial-position function changes from one with a recency profile to one with a primacy profile. This supports our activation-based model, which explains this effect in terms of the dynamics within the short-term buffer. However, a possible alternative explanation might be that the primacy effect in the fastest condition was due to forward masking of all words except the first word (which had no premask). To rule out this interpretation, we ran a control experiment to test this masking hypothesis. We tested 4 participants on 576 trials (96 trials per serial position) with lists that were presented at a rate of 100 ms per word (the critical condition). In addition, the list was both pre- and postmasked (i.e., a row of six ampersands was presented for 100 ms before the first item and 100 ms after the last item in the list). Contrary to the masking hypothesis, a clear primacy effect was obtained, extending over two serial positions. Compared with the average of the two middle positions, recall was better for Positions 1, $t(3) = 4.54$, $p < .05$, and 2, $t(3) = 6.31$, $p < .01$.

The activation-based buffer model predicts that with the increase in presentation rate, the serial-position function shifts from recency to primacy. The results presented here and, in particular, the contrast between the two extreme presentation rate conditions challenge theories that maintain that all memory phenomena can be accounted for in terms of retrieval from a single recency-based memory system that conforms to a ratio-rule-type principle (Crowder, 1976; Neath, 1993b). Without auxiliary assumptions, such theories do not predict a shift from recency to primacy with an increase in presentation rate.[14]

Our alternative account of recency, which is based on the activation buffer, is able to explain these effects and provides important insight into the nature of the displacement process that takes place in the short-term buffer. This process is dynamic and is dependent on presentation rate, which affects the effective level of competition between activated representations. Further properties of the dynamic buffer are investigated in the next section.

Utility of the Dynamic Buffer: The Control of Memory

In the previous sections of this article, we have presented evidence that suggests a role for a dynamic buffer in recall performance. However, it is difficult to imagine that such a buffer

---

[14] A single-store recency-based theory extended with a refractory encoding limitation (i.e., after an item is encoded, a refractory time needs to be available before a new item can be encoded) may predict a decrease in performance after the first item (although masking is factored out). Such a theory, however, should predict that the refractory effect is maximal at Item 2, which is not consistent with our data. Further investigations could further contrast alternative explanations for the recency to primacy shift.

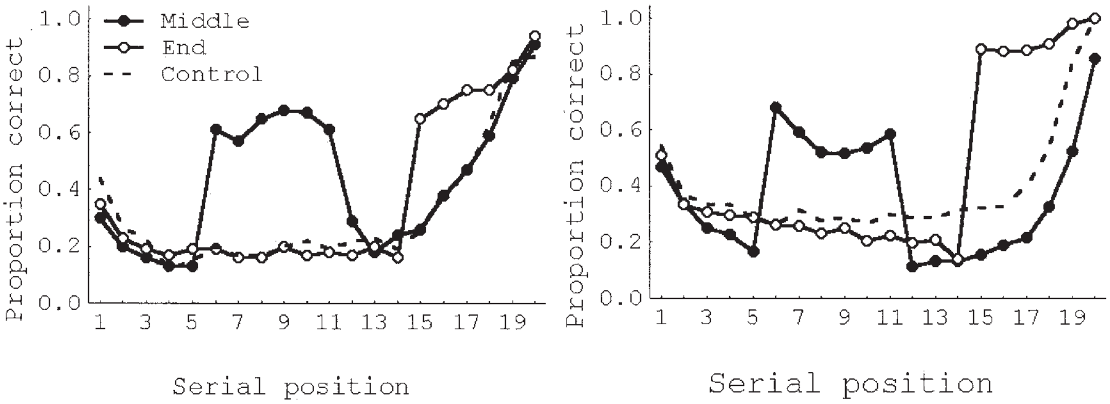

Figure 18. Comparison between data on the effects of semantic clustering in immediate free recall and the simulation. Left: Data from Craik and Levy (1970). Adapted from "Semantic and Acoustic Information in Primary Memory," by F. I. M. Craik and B. A. Levy, 1970, *Journal of Experimental Psychology, 86,* p. 80. Copyright 1970 by the American Psychological Association. Right: Model simulation. In the simulation, the associates have interconnections of strength .14. Standard values were used for all other parameters.

evolved only to enable humans to unload the most recent items in recall tasks. Indeed, there are more rigid ways in which a buffer could have been designed, in particular, with a fixed capacity (of *n* slots) and with a first-in, first-out (knock-out model; Kahana, 1996; Philips et al., 1967) or with a random displacement rule (Kahana, 1996; Raaijmakers & Shiffrin, 1980, 1981). In the following sections, we suggest possible advantages of having self-excitatory and competitive interactions dynamically control the content of the buffer. As these simulations are meant to explore more theoretical issues related to the properties of the buffer and its function, readers whose interest focuses on accounting for recency data may prefer to go directly to the General Discussion.

Content Can Attenuate the Displacement Process: The Role of Semantics

In rigid buffers, the semantic content of items does not influence the probability that the item will be displaced from the buffer. However, if one assumes that semantically related items are interconnected within the dynamic buffer (our buffer is the activated part of a lexical–semantic memory), then, as we show here, these items will remain longer in the buffer, even at the expense of the more recent item (if that recent item is itself unrelated). That is, displacement from our dynamic buffer may display intelligent properties that not only favor the maintenance of the more recent items but also factor in the semantic content of items. This idea is contrary to the traditional view that the buffer represents exclusively phonological information (Baddeley, 1972; but see R. C. Martin, 2003, for a revaluation of this view). Note that whereas an exclusively phonological buffer would be rather limited in its cognitive utility,[15] a buffer that is sensitive to semantic variables is likely to support a wider range of cognitive functions, such as language comprehension (Haarmann, Cameron, & Ruchkin, 2002, 2003; Haarmann, Davelaar, & Usher, 2003; Jackendoff, 2002; R. C. Martin & Romani, 1994), reasoning (Haarmann, Davelaar, & Usher, 2003), and contextual processing (Haarmann, Ashling, Davelaar, & Usher, in press). With regard to memory, we now demonstrate that the dynamic properties of the buffer can lead to semantic effects in immediate free recall that have previously been thought of as stemming from long-term memory.

A well-known finding of semantic effects in immediate free recall was obtained in a study by Craik and Levy (1970). In this study, participants were tested with lists of 20 words presented under three semantic clustering conditions (Craik & Levy, 1970). In the control condition, all the words in the list were unrelated. In the end condition, there was a cluster of 6 semantically related words at the end of the list (Positions 15–20). In the middle condition, there was a cluster of 6 semantically related words in the middle of the list (Positions 6–11). The results revealed that in both the middle and the end conditions, the cluster words were recalled better than the corresponding words in the control condition (see Figure 18, left).

Craik and Levy (1970) interpreted these semantic clustering effects as emerging from long-term memory, by estimating the separate contributions of short- and long-term memory using the procedure derived by Waugh and Norman (1965). However, the application of this procedure required the assumption that semantically related middle list items are not maintained in the buffer, an assumption that is questionable for a dynamic buffer (e.g., Davelaar & Usher, 2003; Watkins, 1974).[16]

In our dynamic buffer, semantically related items can prevent each other from being displaced by strengthening the activation of

---

[15] Besides its involvement in serial recall, the phonological loop has also been suggested to be instrumental in learning new language (Baddeley, Gathercole, & Papagno, 1998).

[16] The main assumption of the Waugh and Norman (1965) procedure is that items from the middle positions are retrieved from episodic long-term memory alone. Thus, Craik and Levy (1970) used the recall performance of the cluster of middle list items (in the middle condition) to estimate the contribution from long-term memory to the cluster in the end condition. Using this assumption, Craik and Levy concluded that the contribution from short-term memory actually decreased under the semantic clustering manipulation. Although their assumption seemed logical at the time, we are now in a position to make use of an explicit computational model to show that the main assumption is questionable for a dynamic buffer.

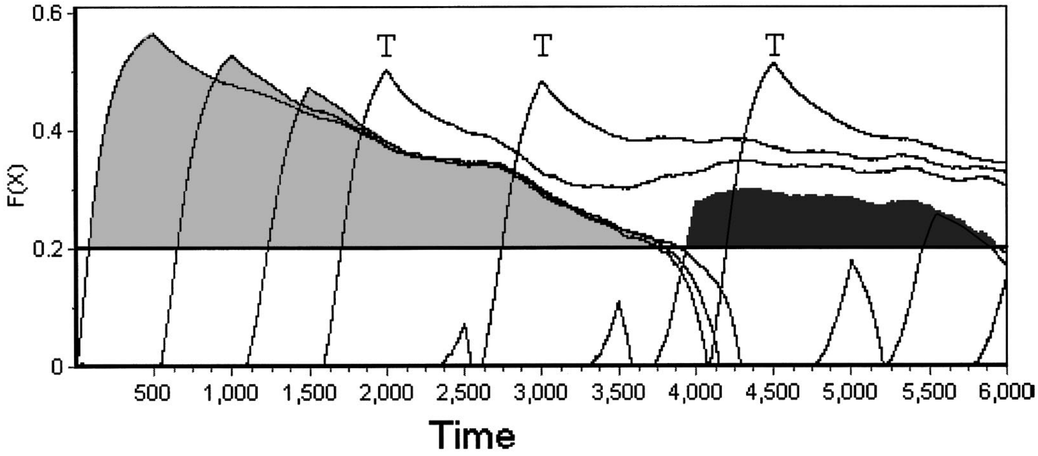

Figure 19. Activation trajectories for a list of 12 items (Items 4, 6, and 9 are the targets). Targets receive an attentionally controlled input of .33, whereas distractors receive an input of .24. Note that only the targets are active at the end of the sequence. F(x) = output activation value; T = target.

one another (Haarmann & Usher, 2001; Usher & Cohen, 1999; for detailed analysis, see Davelaar, 2003). To examine the context–activation model's prediction for the semantic clustering effect, one must set α_2 in Equation 1 to represent associative (excitatory) connections between items; for simplicity, we use only two levels of association strengths (unrelated: $\alpha_2 = 0.00$; related: $\alpha_2 = .14$). The simulation result is presented in Figure 18 (right).

The model reproduces most of the patterns observed in the data. Moreover, it demonstrates that the elevated recall of middle list positions occurs because these items have a lower probability of being displaced from the buffer. Finally, the longer maintenance time of the associates in the buffer leads to stronger episodic connections to the context layer. The enhanced episodic contribution is thus also mediated by the buffer. In summary, our buffer displays intelligent properties that allow it to keep items active on the basis of not only the time they entered the buffer but also their semantic content.

Attentional Control of the Buffer Dynamics: Selective Updating and Selection

In the previous section, we demonstrated how a dynamic activation buffer can make use of the associative structure of the learning material to achieve a type of control over the items it maintains. In this section, we examine two other types of memory control processes that further extend the cognitive utility of the buffer. The first is selective updating, which involves the ability to maintain items that are evaluated as particularly significant (even though they may not be the most recent ones) and update the buffer content depending on the relative significance of incoming items. The second is adapting the buffer parameters to the requirements of the task (e.g., encoding vs. retrieval).

There are many situations that require flexible (selective) updating of information so that only a prespecified subset of items (targets) is designated as important for performance while other items (distractors) are to be ignored (e.g., remembering only the digits within a sequence of digits and letters). Such situations require that, following a fast categorization stage, a selective attention mechanism is recruited to boost the input of the targets (or to attenuate the input of the distractors) into the activation buffer.

In previous work, researchers have explored how such a selective boosting of input can be realized by neuromodulatory brain mechanisms (Usher, Cohen, Servan-Schreiber, Rajkowski, & Aston-Jones, 1999; see also Braver & Cohen, 2000) that control the buffer. Here, we assume that the input to the activation buffer can change in proportion to the importance of the item (akin to gain modulation of input in the cognitive control model of Braver & Cohen, 2000). Figure 19 illustrates the activation trajectories for a memory list of 12 items, out of which only Items 4, 6, and 9 are designated as targets. To simplify, we assume two levels of input modulation, with distractors receiving a weak input of $I_n = .24$ and targets receiving a stronger input of $I_t = .33$.

Examination of the trajectories reveals that the dynamic buffer behaves intelligently in that it can selectively maintain some items (the targets), even at the expense of more recent ones. This is due to the targets receiving larger modulated input, which helps them reach higher levels of activation, whereas distractors either are displaced from the buffer or never succeeded in entering it (as they do not receive enough input to overcome the inhibition from the targets).

A more complex type of selective updating is needed if one is presented with a sequence of objects from which only the largest objects have to be reported (Palladino, Cornoldi, De Beni, & Pazzaglia, 2001). In such a task, what is considered a target at one time in the sequence can later turn out to be a distractor. This situation can also be modeled via the simulation in Figure 19 (assuming that the input to large items is higher than the input to small items as a result of a fast categorization process, which was not modeled explicitly). We can observe that small items (1–3) are maintained in the buffer until larger items (4 and 6) are presented; until that moment, Items 1–3 are rightfully considered *initial targets.* A valuable utility of dynamic buffer is its ability to deactivate initial targets to enable storage of subsequent items that turn out to be the true targets (e.g., Items 4, 6, and 9).

However, having occupied the buffer carries with it a hidden cost. The initial targets are predicted to be encoded more strongly than other distractors (i.e., the small items) that never occupied the buffer because they were presented after the real targets (see the gray and black areas in Figure 19, which are proportional to the episodic trace strengths of early and late distractors, respectively). Therefore, the initial targets are predicted to cause more intrusions, as recently reported by Palladino et al. (2001).

The second utility of the activation buffer, which allows it to control information, is the ability to adapt its processing parameters to the task and process demands. To illustrate the need for this, we consider the difference between encoding and retrieval in free recall or the difference between retrieval in free recall and in cued recall. During encoding, it is advantageous to maintain a larger number of items in the activation buffer (the longer the item is in the buffer the stronger the episodic learning). In contrast, at retrieval, a more restrictive capacity is advantageous to implement selection for output. Similarly, in free recall, a reasonable strategy to optimize performance may be to maintain all the active items in the buffer until they are reported (see Appendix C). In contrast, in cued recall, a selection needs to be made immediately after the probe presentation, and so a more restrictive strategy may be optimal. We propose that a neuromodulatory control of the buffer parameters can help to optimize task performance by modulating the self-excitation and global inhibition (via dopamine and norepinephrine; e.g., O'Reilly, Braver, & Cohen, 1999; Usher et al., 1999).

To illustrate how the control of the parameters adapts the buffer to the requirements of the task, we show in Figure 20 a simulation exploring a way to implement retrieval in a category-cued recall task (see Experiment 2). Without modulation of the buffer parameters and without a probe-related input, the model is able to retain three items at the end of the list presentation (see Figure 12, top) but is unable to select the relevant one. If we assume that the category probe sends a small amount of activation ($I = .10$) to its exemplar, which in this example is the fifth item (see Figure 12, middle),[17] we see that despite an increase in activation for the target item, the model is unable to make a clear selection. Modulation of the global inhibition, however, can facilitate the selection process. When the global inhibition is increased (from .15 during encoding to .45 during retrieval, $t > 3{,}000$), a correct selection of the target item is made (see Figure 12, bottom).

Together, the simulations shown in this section suggest that the dynamic activation buffer transcends the role of a temporary store. Rather, this buffer (in interaction with attentional and neuromodulatory systems) may play an active role in memory control, which is an essential function of working memory (Atkinson & Shiffrin, 1968, 1971; Baddeley, 1986; Baddeley & Hitch, 1974). In particular, when the memory list includes related words or when the task specifies only some items as targets, the buffer exhibits intelligent behavior that involves a selective type of updating. In addition, the modulation of the parameters can switch the buffer function from maintenance at encoding to selection at retrieval.

General Discussion

In this article, we have presented a neurocomputational model of list memory that includes two components, a changing context/episodic system and a capacity-limited activation buffer. This context–activation model has been shown to account for a host of results, which we now summarize.

First, the model accounts for serial-position functions in immediate free recall and in the continuous-distractor task. In particular, the changing context enables the model to account for long-term recency effects, because the context when retrieval begins is more similar (in terms of proximity within the context layer) to the context when the last list items were encoded. The changing context also allows the model to account for the general pattern of lag-recency effects. These effects arise because items that are studied in close proximity to each other are encoded with context units that are proximal and tend to follow each other. This is our simplified way to account for contextual similarity (Howard & Kahana, 2002) within a localist framework.

Using the notion of two different sources dominating short- and long-term recency, we could also accommodate the larger recency that is typically found in immediate free recall. The larger effect in immediate free recall can be understood as emerging from the errorless unloading of items from the short-term buffer. This contrasts with the recency effect in the continuous-distractor task, which is primarily mediated by the reinstatement of the encoding context in episodic memory and is error prone.

Second, the context–activation model provides a coherent explanation for four dissociations reported in the literature between recall patterns in immediate free recall and in the continuous-distractor task to which, we believe, insufficient attention has been paid. The first dissociation is the absence of recency in immediate, but not in continuous-distractor, free recall when participants are instructed to start with items from the beginning of the list (Dalezman, 1976; Whitten, 1978). The second dissociation is the neuropsychological dissociation in amnesia showing a decrease in performance for recency items in the continuous-distractor task but not in immediate free recall (Carlesimo et al., 1996). The third dissociation is the appearance of negative recency in final free recall in the immediate, but not in the continuous-distractor, paradigm. The fourth dissociation is that lag recency changes with output position in immediate free recall but not in continuous-distractor free recall. All these dissociations are accounted for as the result of the contribution of the short-term buffer to the recency effect in immediate free recall but not in the continuous-distractor task, as described in the *Discussion of Dissociation Simulations* section.

Third and most important, the context–activation model gave rise to two novel predictions that critically depended on the presence of an activation-based buffer. The first prediction was that the performance for recency items is immune to proactive interference in immediate, but not in continuous-distractor, free recall, even when the ratio between IPI and RI in the tasks is preserved. Our model predicts the dissociation with proactive interference. The long-term recency is purely due to contextual retrieval-based mechanisms that are susceptible to proactive

---

[17] This magnitude of the input was chosen so that it would not be large enough to reactivate items that had been displaced from the buffer yet would be sufficiently large to elevate the activation of the exemplar. The probe-related input is thus thought to select among active items (nonactive items require additional episodic retrieval processes to be selected; cf. Diller, Nobel, & Shiffrin, 2001).

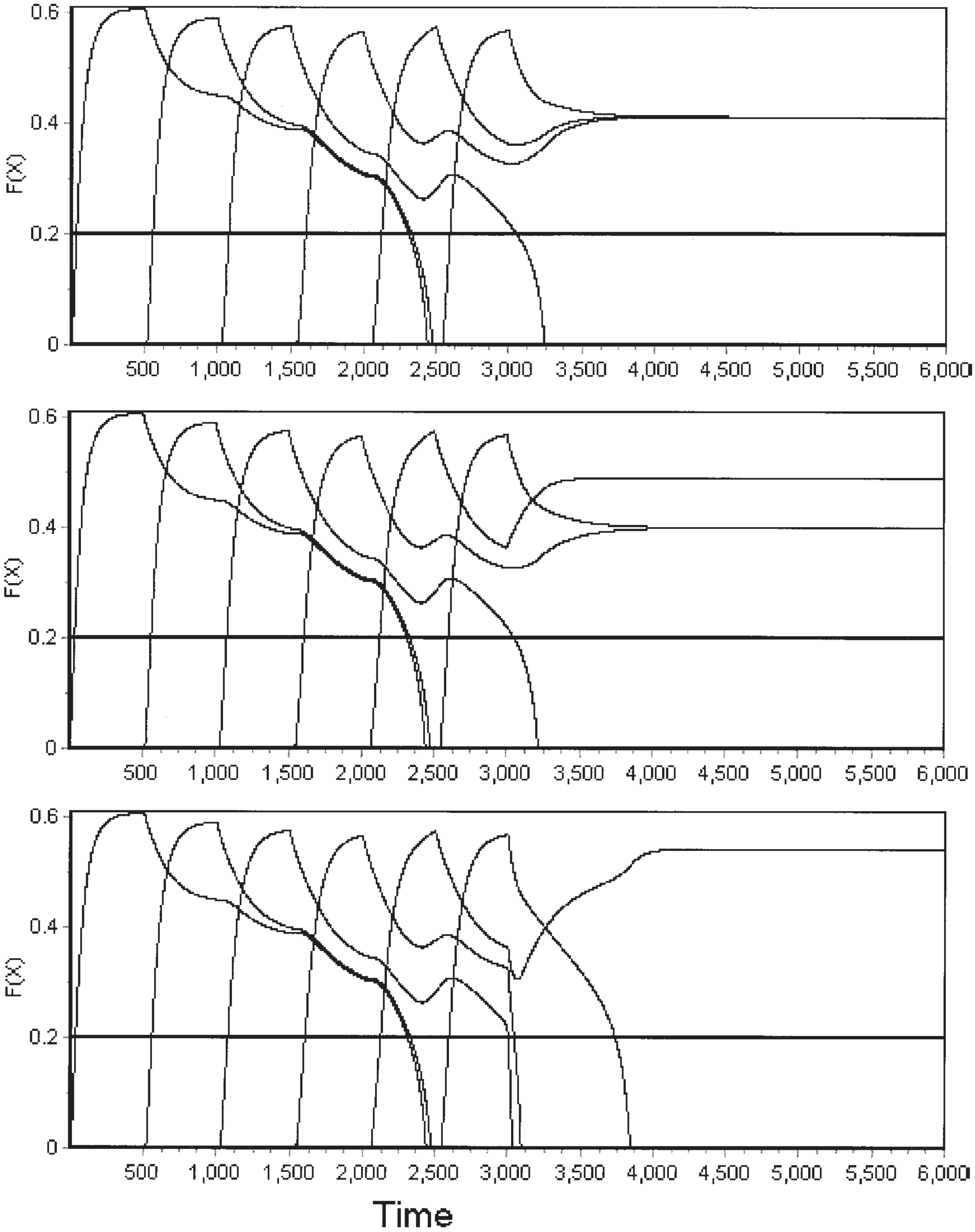

Figure 20. Activation trajectories for a category-cued recall task with a six-word list. Top: Inhibition remains constant ($\beta = .15$) during encoding and retrieval. The probe does not activate its exemplar. Middle: Constant inhibition and the probe activate the target (Item 5) with I = .10. Bottom: An increase in inhibition from .15 to .45 and probe activation lead to correct selection of the target item (Item 5). F(x) = output activation value.

interference. In contrast, the model's short-term recency effect is predominantly the result of unloading of the contents of the activation buffer, which are immune to proactive interference.

Experiment 1 showed that proactive interference dissociated short-term recency from long-term recency, thereby confirming the first prediction of our model. Unlike previous studies that reported associations between immediate free recall and the continuous-distractor task in different experiments across different labs (Greene, 1986a; Greene & Crowder, 1984), our study was the first to examine the two tasks within a single experiment using the same materials, design, and procedure, thereby avoiding potential confounds.

The model's second prediction involved a shift from recency to primacy with an increase in presentation rate. This prediction arose from the internal dynamics of the activation buffer. The nature of the displacement process in the buffer depends on a fine balance between excitation and inhibition. At slow presentation rates, an incoming item accumulates enough activation for self-support and enough activation to overcome the inhibition from previously presented items. Therefore, a displacement process can ensue, during which early items in the buffer are more likely to be displaced than newer ones. At fast presentation rates, however, incoming items do not accumulate enough activation to overcome the competition from previous items in the buffer. In other words, at fast presentation rates, new items are less likely to enter the buffer, thereby leading to a primacy effect. This prediction was confirmed in Experiment 2. In the following sections, we address a number of key implications arising from the model, discuss the nature of the activation buffer and its extension to response latencies, and compare the model with other prevalent theories.

The Need for a Buffer System: Single- Versus Dual-Store Theories

The results have a series of implications in the wider debate between single and dual theories of list memory. First, the four dissociations between immediate and continuous-distractor free recall (directed output order, amnesic syndrome, lag-recency/output-position interaction, and negative recency in final recall) are in stark contradiction with earlier claims of single-store theorists (Crowder, 1982; Greene, 1986a, 1992) that experimental manipulations have equivalent effects on short- and long-term recency. As we have shown, our dual-store model provides a natural account for these dissociations.

Second, the context–activation model demonstrates that a single-store retrieval mechanism based on contextual change does not necessarily account for the proactive interference dissociation between short- and long-term recency with equal IPI:RI ratios, whereas a dual-store account can. In the model, when we disabled the retrieval from the activation buffer, we found parallel serial-position functions for the first and second trials in immediate as well as in continuous-distractor free recall. Thus, the contextual component alone does not account for the immunity to proactive interference at recency in immediate free recall that was found in our Experiment 1 and in Craik and Birtwistle (1971). Although none of the existing models of contextual encoding have yet been used to account for proactive interference on serial-position functions, one cannot dismiss the idea that a more complex single-store model might still be able to account for the immunity to proactive interference at recency in immediate free recall. The challenge for such a model, however, would be to account simultaneously for the absence of proactive interference at recency in immediate free recall and its presence in the continuous-distractor task.

Third, the shift from recency to primacy with presentation rate (Experiment 2), which was predicted by the dynamics of the activation buffer, poses a problem for current single-store models (e.g., Glenberg et al., 1983; Howard & Kahana, 2002; Tan & Ward, 2000; Ward, 2002). These models view recency as a generic property of memory retrieval due to the enhanced similarity between encoding and retrieval contexts of the last list items (and they do not address the probability for episodic encoding). Recently, such a theory was proposed by Ward and colleagues (Tan & Ward, 2000; Ward, 2002), who have shown that a recency-based model that includes a mechanism of overt rehearsal can also account for primacy effects and lexicality effects in immediate free recall. However, this approach is silent with respect to the presence of primacy effects in continuous-distractor free recall, whose raison d'être is the elimination of rehearsal. More important, such models will have difficulty explaining (without auxiliary assumptions) how the mere increase in the presentation rate, which in fact precludes rehearsal, makes the recall at primacy positions better than that at recency positions.

Fourth, a dynamic activation buffer not only is useful in list-memory tasks but also allows the system to flexibly allocate its cognitive resources to wider domains of information processing. The buffer exhibits sensitivity to semantic organization, such that associates that are presented in close temporal proximity are maintained longer than unrelated items, even when the latter are more recent. Moreover, the dynamic nature of the buffer provides the system with a means to control the type of incoming information and to adapt to task requirements (encoding vs. retrieval). Together, the storage function and the relation to cognitive control make the dynamic activation buffer a credible candidate for being central in a general working-memory system, as originally suggested in the Atkinson and Shiffrin (1968, 1971) model.

Although it may be possible to construct single-store models that account for dissociations between immediate and continuous-distractor free recall, it is the consideration of a large data set (as well as theoretical considerations on memory control) that informs the debate on whether it is necessary to postulate a short-term buffer. This data set includes not only dissociations with experimental variables (directed output order, negative recency in final recall, proactive interference) but also neuropsychological dissociations (amnesic) and detailed information on recall transitions (e.g., lag recency) as well as presentation rate effects. It is this rich data set that suggests to us that the concept of a short-term buffer has an explanatory value in the study of basic effects (e.g., recency effects) in list memory (see Cowan, 1995, Section 4.2, for additional arguments against the sufficiency of the single-store accounts). We believe that, taken together, the presentation rate effect and the set of dissociations we reviewed and reported here support a dual-activation/context-type theory of memory and provide a challenge to the opponent single-store account. To rephrase comments by Mark Twain,[18] we suggest "it seems that reports of the demise of short-term memory may have been much exaggerated."

The Nature of the Buffer and Insights From Neuroscience

One of the textbook objections to the dual-store memory models is based on neuropsychological data of the short-term memory syndrome (Shallice & Warrington, 1970). This syndrome involves

---

[18] Mark Twain reacted to a report about his own death (Twain, 1940).

the finding of preserved performance in recall from long-term memory despite a deficit in short-term memory tasks (a defective buffer), seemingly contrary to the bottleneck assumption of the dual-store model (i.e., the assumption that information must pass a short-term buffer before entering a long-term repository). Two patients, K.F. (Shallice & Warrington, 1970; Warrington & Shallice, 1969) and P.V. (Baddeley, Papagno, & Vallar, 1988; Basso, Spinnler, Vallar, & Zanobio, 1982), have illustrated the short-term memory syndrome, both showing reduced digit span but spared long-term learning (see also patient I.R. reported by Belleville, Caza, & Peretz, 2003). However, the argument that the selective short-term memory impairment refutes the bottleneck assumption of the dual-store model is valid only if one further assumes that short-term memory cannot be fractionated according to the representational code.

Recently, Martin and colleagues (R. C. Martin et al., 1994; Romani & Martin, 1999) proposed an alternative account for these findings. On the basis of additional evidence from patients with brain damage (e.g., E.A. and A.B.), they argued for a separation between a phonological short-term memory component that mediates phonological long-term learning and a semantic short-term memory component (i.e., a lexical–semantic buffer) that mediates semantic long-term learning (Romani & Martin, 1999). They reasoned that patients like P.V. and K.F., although impaired in phonological short-term memory (which explains their reduced digit span), may not have been impaired in semantic short-term memory (which was not tested for P.V. and K.F. but was confirmed for their patient E.A.). This reasoning would explain their patient's preserved recall in long-term memory tasks requiring semantic encoding.

Consistent with the dual-store model, these patients are impaired in the encoding of new phonological forms in long-term memory (new language learning; Baddeley et al., 1988). Moreover, Romani and Martin (1999) correctly predicted the existence of patients with the opposing pattern of deficit, such as their patient A.B., who was impaired in semantic short-term memory but not in phonological short-term memory and showed intact phonological but not semantic long-term learning. This revalidates the dual-store bottleneck assumption with the constraint that it holds for each representational code, separately.[19]

The existence of a semantic buffer is consistent with a series of cognitive studies that have demonstrated semantic effects in immediate memory (Haarmann & Usher, 2001; Raser, 1972; Shulman, 1970). Moreover, a semantic working memory system within the left prefrontal cortex has been proposed on the basis of neuroimaging studies, which have indicated a role for this system in the control of semantic retrieval (Gabrieli et al., 1998; Wagner, Pare-Blagoev, Clark, & Poldrack, 2001; but see Fletcher & Henson, 2001, for a wider review). We should emphasize that we do not exclude the existence of a phonological buffer, but we instead support the view of two interacting short-term buffers that maintain different linguistic codes (see, e.g., Forde & Humphreys, 2002; Knott, Patterson, & Hodges, 1997; N. Martin & Saffran, 1997; Shulman, 1971). In this work, we have focused on the semantic buffer, as we believe that this plays a larger role in item recall as opposed to serial recall, and we assume that participants can strategically allocate their attention to the buffer system that will maximize task performance (cf. Shallice, 1975).

Unlike a decay-based phonological buffer, a semantic working memory system with limited capacity and with sensitivity to control processes (see the Utility of the Dynamic Buffer section) may play a larger role in complex cognition. This is illustrated by a series of computational studies that used an activation buffer in a variety of domains, such as text comprehension (Haarmann, Just, & Carpenter, 1997; Just & Carpenter, 1992), problem solving (J. R. Anderson & Lebiere, 1998), and contextual processing (J. D. Cohen & Servan-Schreiber, 1992). Moreover, correlational studies provide additional support for a role of a semantic buffer in these domains (Haarmann et al., in press; Haarmann, Davelaar, & Usher, 2003) and suggest that a reduction in the buffer capacity explains age-related decline in context processing (as in the AX version of the continuous performance task; Braver et al., 2001). Indeed, the need for an additional nonphonological verbal buffer prompted Baddeley (2000) to suggest a fourth component to the original (Baddeley & Hitch, 1974) working memory theory, named *the episodic buffer*. Because the activated part of the semantic memory also provides episodic information, further research is needed to examine whether Baddeley's (2000) episodic buffer is different from the lexical–semantic buffer discussed here.

Extending to Response Latencies

We have presented a model that relies on a minimal set of assumptions to account for accuracy data in immediate and continuous-distractor free recall and in category-cued recall. In most of the experiments conducted in this field, the data involve accuracy as a function of serial position in the list. There are a few studies, however, that have also examined recall latencies and interresponse times (Murdock, 1972; Murdock & Okada, 1970; Waugh, 1970; for review, see Wixted & Rohrer, 1993). Although, in its present form, our model does not predict response latencies and interretrieval times, we believe that it provides a natural framework for doing so. Response latencies can be simulated (see Appendix B) in a dynamical model in which the cue (external or internal) sends input to the item units and the model's parameters (excitation and inhibition) are modulated. This results in a race of the units' activation toward a response threshold. The time to reach threshold can then serve as a measure for response latencies.

To obtain a more complete account of response latencies (and interretrieval times) in immediate free recall, one needs to integrate the dynamics of the retrieval process with the temporal dynamics of the contextual change. Such an enterprise requires additional assumptions and, thus, should be the subject of future investigation. Nevertheless, we would like to briefly suggest how a dual-store (context–activation) model can account for response latency data, challenging a recent analysis of such data in free recall by

---

[19] Recently, Belleville et al. (2003) tested patient I.R. on a host of short- and long-term memory tasks while preventing confounding with linguistic code. These authors found that patient I.R. has a deficit in phonological and not semantic processing, supporting the criticism made by Romani and Martin (1999). However, the claim of Belleville and colleagues that "there appears to be no need to postulate an independent short-term store" (p. 700) is articulated against the view that the verbal short-term store is exclusively phonological. However, their data are silent with regard to the view of the existence of a lexical–semantic short-term store.

Laming (1999), who concluded that “a separate short-term store is no better than make-believe”(p. 425; see also Ward, 2002).

Laming (1999) reanalyzed the latency data from an immediate free recall study by Murdock and Okada (1970). His conclusion against a separate short-term store was based on the fact that when examining the latency of first recall, one finds similar latencies for trials that start with recency and nonrecency items. This contradicts the assumption he made on behalf of dual-store theories that items in short-term memory have faster first-recall latencies than do items retrieved from episodic long-term memory. This assumption, however, is questionable because first-recall items from nonrecency positions may still be residing in the short-term buffer (see Figure 4, showing a nonzero probability for middle list items to be in the buffer). Moreover, as we discuss in Appendix B, when a middle list item resides in the buffer, this item is likely to be reported first.

A more critical test of the existence of the short-term store based on latency data involves a probed-recall procedure that avoids output-order competition. As we have shown in the *Attentional Control of the Buffer Dynamics* section (see Figure 20), a selective cue can immediately retrieve items in the activation buffer, whereas nonbuffer items need a slower contextual support for retrieval. Because recency items reside most often in the buffer when the retrieval cue is presented, they are predicted to have shorter latencies. This was precisely the result found by Waugh (1970), who examined the distribution of recall latencies in probed recall as function of serial position.

Furthermore, interresponse latency data from Murdock and Okada (1970) seem to suggest additional support for such a model. As reported by Murdock (1972; Metcalfe & Murdock, 1981), a discontinuity in interresponse times is found after the report of a first chunk of three to four recency items and before the retrieval of earlier items starts. Specifically, the first four interresponse times for a sequence of 17, 18, 19, 20, and n (out of a 20-word list; n being any other nonrecency item) are 426, 639, 752, and then a significantly slower interresponse time of 2,830 ms. Critically, n in the above sequence most often corresponds to the first list item. This can be explained in terms of the time needed for the transition from the end to the start context (Metcalfe & Murdock, 1981). If the discontinuity exists for other (e.g., 17, 18, 19, 20, and $n = 15$) items that are too remote from the start context, an account based on retrieval dynamics of the buffer can provide a plausible explanation. In particular, recency items, which are still active, are retrieved closely together and before nonactive items (see Figure B1, showing a discontinuity between the retrieval of active and nonactive items).

Relations to Other Memory Models and Theories

The context–activation theory we present in this article shares important properties with a number of leading memory models. First, as in previous neural network models (Becker & Lim, 2003; Chappell & Humphreys, 1994), we assume that list memory is mediated by the combined function of a contextual and a lexical–semantic system. Whereas in these models the lexical–semantic system is a winner-take-all neural network, in our model the lexical–semantic system has a larger capacity. Second, the idea that long-term memory representations are used during list presentation for episodic encoding has also been used in other recent models of list memory, such as ACT-R (J. R. Anderson et al., 1998) and serial-order-in-a-box (SOB; Farrell & Lewandowsky, 2002), in domains as diverse as free recall and serial recall. Our model shares with SOB and ACT-R the idea that attractor-type lexical traces are used to convey episodic information. Whereas in SOB and ACT-R this is done by strengthening existing representations, our model uses activation and connections to context to distinguish items in the current trial from items in previous trials. Third, our model is closely related to computational models that have simulated serial-position functions in free recall: SAM (Mensink & Raaijmakers, 1988; Raaijmakers & Shiffrin, 1980), ACT-R (J. R. Anderson et al., 1998), and TCM (Howard & Kahana, 2002). We briefly highlight the similarities and differences between the context–activation model and these previous models. As in SAM and ACT-R, we assume the existence of a short-term memory buffer that drives the encoding into the episodic system, and we follow these models closely in the way in which the two-stage retrieval process operates.

During encoding, the buffer actively maintains a limited number of items, while above-threshold activation of items leads to the formation of episodic traces. At retrieval, items in the activation buffer are readily available for output. Our activation model differs from SAM mainly with regard to the dynamics of the displacement from the buffer. Whereas SAM assumes a random displacement governed by a capacity parameter, in our model the displacement is governed by a dynamic activation process whose characteristics depend conjointly on the presentation rate and order (see Appendix C for a detailed comparison between the traditional and activation-based buffer models). Another property of the SAM buffer is its support of associative encoding between items in the buffer. Although this principle is consistent with our model (Hebbian learning could create associative connections between coactive items), we have not implemented it in this version of our model in which we attempted to aim for simplicity. Further data, however, could motivate the inclusion of such a component. In some sense, the context–activation model can be seen as a neural implementation of the SAM framework, in which the buffer is made dynamic and is conceptualized in terms of activation and in which the context representation is made variable.

The view of an activation-based short-term memory buffer is not new and has been featured in many models and theories of memory (J. R. Anderson, 1972; Atkinson & Shiffrin, 1968; Broadbent, 1957; Cowan, 1988, 1999; Hebb, 1949; James, 1890; Just & Carpenter, 1992; D. A. Norman, 1968; Shiffrin, 1976). For example, Atkinson and Shiffrin (1971) proposed that “one might consider the short-term store simply as being a temporary activation of some portion of the long-term store” (p. 83). In contrast, Baddeley (1996) criticized the view that the short-term store is the temporary activation of the long-term store, in that “such a view is so general as to be theoretically sterile, unless an attempt is made to specify in detail the processes involved” (p. 22). In this article, we have made such an attempt by developing an explicit computational model in which the short-term memory buffer is the activated part of a lexical–semantic representation that is distinct from the episodic system.

Likewise, the distinction between episodic memory and active memory is not new and corresponds to the distinction between synaptic weights and neural reverberation that has been proposed by Hebb (1949; see also James, 1890).

Activation-based processes underlying cognitive performance have been investigated in a large body of neurophysiological (e.g., Goldman-Rakic, 1992; Miller et al., 1996), neuroimaging (Gabrieli et al., 1998), and neurocomputational (e.g., Compte, Brunel, Goldman-Rakic, & Wang, 2000; Durstewitz et al., 2000) studies. The view that activation-based processes limit cognitive ability is featured in some prominent models of high-level cognition, such as ACT-R (J. R. Anderson, Reder, & Lebiere, 1996; Daily, Lovett, & Reder, 2001; Lovett, Reder, & Lebiere, 1999) and the collaborative activation-based processing system (CAPS; Just & Carpenter, 1992). The use of activation-based mechanisms to account for working memory tasks has also recently occurred in neurocomputational models (Braver & Cohen, 2000; Carpenter & Grossberg, 1993; Davelaar & Usher, 2002; Grossberg, 1978; Haarmann & Usher, 2001; Taylor & Taylor, 2000; Usher & Cohen, 1999). The most important aspect of this literature is that capacity limitation is a dynamic result of competitive neural mechanisms instead of a fixed parameter (e.g., source activation in ACT-R, buffer size in SAM).

In addition to sharing properties with activation-based models, our model also shares properties with previous models of list memory that include a changing context (Dennis & Humphreys, 2001; Glenberg et al., 1983; Howard & Kahana, 2002; Mensink & Raaijmakers, 1988). We assume that items presented in close temporal proximity are more likely to be encoded in a similar context and that the similarity between contexts at encoding and retrieval relates to recall performance. The representation of context in our model includes a simplified, localistic implementation of context. Other models, in contrast, rely on distributed context (Dennis & Humphreys, 2001; Howard & Kahana, 2002; Mensink & Raaijmakers, 1988; K. A. Norman & O'Reilly, 2003). Furthermore, some models, but not ours, make distinctions between experimental and pre-experimental contexts with separate retrieval properties (Howard & Kahana, 2002). While noting these differences, it is important to remember that our main focus was not the details of the contextual encoding per se but rather the understanding of the mechanisms that distinguish memory recall in immediate versus delayed tasks such as the continuous-distractor task. As such, we preferred to choose the simplest possible scheme of contextual encoding and examine to what extent such a scheme, together with the activation buffer, could reproduce behavioral patterns in list memory. Therefore, we view our implementation not as capturing the entire range of complexity of contextual encoding but only as a simple scheme that surprisingly seems to account for much of the variability in the literature when added to the activation buffer.

Whereas our model was able to account for many of the behavioral patterns in the literature, there is one feature of the model that may suggest the need for a more complex representation of context. Although we have obtained robust long-term recency effects, these effects have a lower magnitude than do the effects that are sometimes found in the literature. Although some studies have reported small long-term recency effects (e.g., Greene, 1986a; Greene & Crowder, 1984; Poltrock & MacLeod, 1977), other studies (including ours) have reported long-term recency effects of a larger magnitude than our model predicts. The reason for our smaller long-term recency effect is probably the use of a localistic, instead of a distributed, context representation. With a distributed (e.g., Howard & Kahana, 2002) or an extended window-type (Burgess & Hitch, 1999) context, the changing context at retrieval always shares some features with the context in which the last items were encoded and is, therefore, likely to trigger their retrieval in most of the trials. In our model, the localistic context makes retrieval dependent on the random walk revisiting the same context unit. Therefore, although the encoding context is revisited with a higher probability for the last list items, still, in many of the trials, the retrieval context and the encoding context are completely different (i.e., are represented by different, even if neighboring, context units), and retrieval fails entirely. Although we believe that all our central predictions do not depend on such distributed representations, initial explorations with an extended window-type context (Burgess & Hitch, 1999) confirmed the above analysis, and future work should investigate the additional contribution of a distributed context in accounting more precisely for various data patterns.

Concluding Remarks

We have presented a context–activation theory of list memory that accounts for data in free- and cued-recall paradigms by assuming different sources underlying short- and long-term recency effects. We suggest that besides a contextual retrieval process, an activation-based short-term buffer is necessary. The inclusion of the buffer not only allows an account of the observed dissociations but also provides a parsimonious explanation for the shift from recency to primacy with an increase in presentation rate and is well consistent with neuropsychological data.

One final thought may be relevant regarding the role the principle of parsimony has played in the debate between the single- and dual-store theories of memory. On the face of it, a single-store theory is more parsimonious than a dual-store theory, which assumes two entities instead of one. However, the situation may reverse if the dual-store theory is framed in terms of well-accepted distinctions between synaptic and activation-based memory processes, because single-store models will require much more complex processes (e.g., several components of context) to account for the same data. Although a single-store account may be possible, we believe that now both parsimony and the data favor the dual-store approach. To rephrase Crowder's (1993) remark, we suggest "the burden of evidence lies now with those who say that *all* recency effects are caused by a *single* mechanism."

References

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review, 84,* 413–451.

Anderson, J. R. (1972). FRAN: A simulation model of free recall. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation: Vol. 5. Advances in research and theory* (pp. 315–378). New York: Academic Press.

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

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Mahwah, NJ: Erlbaum.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology, 30,* 221–256.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Vol. 2. Advances in research and theory* (pp. 89–195).New York: Academic Press.

Atkinson, R. C., & Shiffrin, R. M. (1971). The control of short-term memory. *Scientific American, 225,* 82–90.

Baddeley, A. D. (1972). Retrieval rules and semantic coding in short-term memory. *Psychological Bulletin, 78,* 379–385.

Baddeley, A. D. (1986). *Working memory.* Oxford, England: Oxford University Press.

Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49*(A), 5–28.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423.

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105,* 158–173.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Vol. 6. Advances in research and theory* (pp. 47–90). New York: Academic Press.

Baddeley, A. D., & Hitch, G. J. (1977). Recency re-examined. In S. Dornic (Ed.), *Attention and performance VI* (pp. 647–667). Hillsdale, NJ: Erlbaum.

Baddeley, A. D., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language, 27,* 586–595.

Baddeley, A. D., & Warrington, E. K. (1970). Amnesia and the distinction between long- and short-term memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 176–189.

Basso, A., Spinnler, H., Vallar, G., & Zanobio, E. (1982). Left hemisphere damage and selective impairment of auditory verbal short-term memory: A case study. *Neuropsychologia, 20,* 263–274.

Becker, S., & Lim, J. (2003). A computational model of prefrontal control in free recall: Strategic memory use in the California Verbal Learning Task. *Journal of Cognitive Neuroscience, 15,* 821–832.

Belleville, S., Caza, N., & Peretz, I. (2003). A neuropsychological argument for a processing view of memory. *Journal of Memory and Language, 48,* 686–703.

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

Braver, T. S., Barch, D. M., Keys, B. A., Carter, C. S., Cohen, J. D., Kaye, J. A., et al. (2001). Context processing in older adults: Evidence for a theory relating cognitive control to neurobiology in healthy aging. *Journal of Experimental Psychology: General, 130,* 746–763.

Braver, T. S., & Cohen, J. D. (2000). On the control of control: The role of dopamine in regulating prefrontal function and working memory. In S. Monsell & J. Driver (Eds.), *Attention and performance XVIII: Control of cognitive processes* (pp. 713–737). Cambridge, MA: MIT Press.

Broadbent, D. E. (1957). A mechanical model for human attention and immediate memory. *Psychological Review, 64,* 205–215.

Broadbent, D. E. (1971). *Decision and stress.* London: Academic Press.

Brodie, D. A., & Murdock, B. B. (1977). Effect of presentation time on nominal and functional serial-position curves of free recall. *Journal of Verbal Learning and Verbal Behavior, 16,* 185–200.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107,* 127–181.

Brown, J. (1958). Some tests of the decay theory of immediate memory. *Quarterly Journal of Experimental Psychology, 10,* 12–21.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review, 106,* 551–581.

Busey, T. A., & Loftus, G. R. (1994). Sensory and cognitive components of visual information acquisition. *Psychological Review, 101,* 446–469.

Capitani, E., Della Sala, S., Logie, R. H., & Spinnler, H. (1992). Recency, primacy, and memory: Reappraising and standardising the serial position curve. *Cortex, 28,* 315–342.

Carlesimo, G. A., Marfia, G. A., Loasses, A., & Caltagirone, C. (1996). Recency effect in anterograde amnesia: Evidence for distinct memory stores underlying enhanced retrieval of terminal items in immediate and delayed recall paradigms. *Neuropsychologia, 34,* 177–184.

Carpenter, G. A., & Grossberg, S. (1993). Normal and amnesic learning, recognition, and memory by a neural model of cortico-hippocampal interactions. *Trends in Neurosciences, 16,* 131–137.

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

Cohen, J. D., & Servan-Schreiber, D. (1992). Context, cortex, and dopamine: A connectionist approach to behavior and biology in schizophrenia. *Psychological Review, 99,* 45–77.

Cohen, R. L. (1970). Recency effects in long-term recall and recognition. *Journal of Verbal Learning and Verbal Behavior, 9,* 672–678.

Compte, A., Brunel, N., Goldman-Rakic, P. S., & Wang, X.-J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex, 10,* 910–923.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104,* 163–191.

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

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62–101). Cambridge, England: Cambridge University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–185.

Craik, F. I. M. (1970). The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior, 9,* 143–148.

Craik, F. I. M., & Birtwistle, J. (1971). Proactive inhibition in free recall. *Journal of Experimental Psychology, 91,* 120–123.

Craik, F. I. M., Gardiner, J. M., & Watkins, M. J. (1970). Further evidence for a negative recency effect in free recall. *Journal of Verbal Learning and Verbal Behavior, 9,* 554–560.

Craik, F. I. M., & Levy, B. A. (1970). Semantic and acoustic information in primary memory. *Journal of Experimental Psychology, 86,* 77–82.

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.

Crowder, R. G. (1993). Short-term memory: Where do we stand? *Memory & Cognition, 21,* 142–145.

da Costa Pinto, A., & Baddeley, A. D. (1991). Where did you park your car? Analysis of a naturalistic long-term recency effect. *European Journal of Cognitive Psychology, 3,* 297–313.

Daily, L. Z., Lovett, M. C., & Reder, L. M. (2001). Modeling individual differences in working memory performance: A source activation account. *Cognitive Science, 25,* 315–353.

Dalezman, J. J. (1976). Effects of output order on immediate, delayed, and final recall performance. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 597–608.

Davelaar, E. J. (2003). *Active memory: Its nature and its role in mnemonic and linguistic behaviour.* Unpublished doctoral dissertation, University of London, London.

Davelaar, E. J., & Usher, M. (2002). An activation-based theory of immediate item memory. In J. A. Bullinaria & W. Lowe (Eds.), *Connectionist models of cognition and perception: Proceedings of the seventh neural computation and psychology workshop* (pp. 118–130). Singapore: World Scientific.

Davelaar, E. J., & Usher, M. (2003). Semantic clustering in free recall: Evidence for a semantic buffer? In F. Detje, D. Dorner, & H. Schaub, (Eds.), *The logic of cognitive systems. Proceedings of the 5th International Conference on Cognitive Modeling* (pp. 239–240). Bamberg, Germany: Universitats-Verlag.

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

Diller, D. E., Nobel, P. A., & Shiffrin, R. M. (2001). An ARC-REM model for accuracy and response time in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 414–435.

Durstewitz, D., Seamans, J. K., & Sejnowski, T. J. (2000). Neurocomputational models of working memory. *Nature Neuroscience, 3*(Suppl.), 1184–1191.

Engle, R. W. (1974). Negative recency in delayed recognition. *Journal of Verbal Learning and Verbal Behavior, 13,* 209–216.

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

Estes, W. K. (1997). Processes of memory loss, recovery, and distortion. *Psychological Review, 104,* 148–169.

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review, 9,* 59–79.

Fletcher, P. C., & Henson, R. N. A. (2001). Frontal lobes and human memory: Insights from functional neuroimaging. *Brain, 124,* 849–881.

Forde, E. M. E., & Humphreys, G. W. (2002). The role of semantic knowledge in short-term memory. *Neurocase, 8,* 13–27.

Gabrieli, J. D. E., Poldrack, R. A., & Desmond, J. E. (1998). The role of left prefrontal cortex in language and memory. *Proceedings of the National Academy of Sciences, U.S.A., 95,* 906–913.

Glanzer, M. (1972). Storage mechanisms in recall. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation: Vol. 5* (pp. 129–193). New York: Academic Press.

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior, 5,* 351–360.

Glanzer, M., Gianutsos, R., & Dubin, S. (1969). The removal of items from short-term storage. *Journal of Verbal Learning and Verbal Behavior, 13,* 114–131.

Glenberg, A. M., Bradley, M. M., Kraus, T. A., & Renzaglia, G. J. (1983). Studies of the long-term recency effect: Support for a contextually guided retrieval hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 231–255.

Glenberg, A. M., Bradley, M. M., Stevenson, J. A., Kraus, T. A., Tkachuk, M. J., & Gretz, A. L. (1980). A two-process account of long-term serial position effects. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 355–369.

Glenberg, A. M., & Swanson, N. C. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 3–15.

Goldman-Rakic, P. S. (1992). Working memory and the mind. *Scientific American, 267,* 111–117.

Goshen-Gottstein, Y., Ashkenazi, A., & Usher, M. (2000, November). *Reports of the demise of short-term memory are much exaggerated.* Paper presented at the 41st Annual Meeting of the Psychonomic Society, New Orleans, LA.

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

Greene, R. L. (1986b). Sources of recency effects in free recall. *Psychological Bulletin, 99,* 221–228.

Greene, R. L. (1992). *Human memory: Paradigms and paradoxes.* Hillsdale, NJ: Erlbaum.

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

Gronlund, S. D., & Shiffrin, R. M. (1986). Retrieval strategies in recall of natural categories and categorized lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12,* 550–561.

Grossberg, S. (1978). A theory of human memory: Self-organization and performance of sensory–motor codes, maps, and plans. In R. Rosen & F. Snell (Eds.), *Progress in theoretical biology* (Vol. 5, pp. 233–374). New York: Academic Press.

Haarmann, H. J., Ashling, G. E., Davelaar, E. J., & Usher, M. (in press). Age-related declines in context maintenance and semantic short-term memory. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*.

Haarmann, H. J., Cameron, K. A., & Ruchkin, D. S. (2002). Long-range neural synchronization mediates on-line sentence processing: EEG coherence evidence from filler-gap constructions. *Psychophysiology,* 39, 820–825.

Haarmann, H. J., Cameron, K. A., & Ruchkin, D. S. (2003). Short-term semantic retention during on-line sentence comprehension: Brain potential evidence from filler-gap constructions. *Cognitive Brain Research,* 15, 178–190.

Haarmann, H. J., Davelaar, E. J., & Usher, M. (2003). Individual differences in semantic short-term memory capacity and reading comprehension. *Journal of Memory and Language, 48,* 320–345.

Haarmann, H. J., Just, M. A., & Carpenter, P. A. (1997). Aphasic sentence comprehension as a resource deficit: A computational approach. *Brain and Language, 59,* 76–120.

Haarmann, H. J., & Usher, M. (2001). Maintenance of semantic information in capacity-limited item short-term memory. *Psychonomic Bulletin & Review, 8,* 568–578.

Halford, G. S., Maybery, M. T., & Bain, J. D. (1988). Set-size effects in primary memory: An age-related capacity limitation? *Memory & Cognition, 16,* 480–487.

Hasselmo, M. E., & Wyble, B. P. (1997). Simulation of the effects of scopolamine on free recall and recognition in a network model of the hippocampus. *Behavioral Brain Research, 89,* 1–34.

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

Hebb, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.

Henson, R. N. A. (1998). Short-term memory for serial order: The start–end model. *Cognitive Psychology, 36,* 73–137.

Hitch, G. J., & Ferguson, J. (1991). Prospective memory for future intentions: Some comparisons with memory for past events. *European Journal of Cognitive Psychology, 3,* 285–295.

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

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

Humphreys, M. S., Bain, J. D., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review, 96,* 208–233.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution.* Oxford, England: Oxford University Press.

James, W. (1890). *Principles of psychology.* New York: Holt.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99,* 122–149.

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

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

Kane, M. J., & Engle, R. W. (2000). Working memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 336–358.

Knott, R., Patterson, K., & Hodges, J. R. (1997). Lexical and semantic binding effects in short-term memory: Evidence from semantic dementia. *Cognitive Neuropsychology, 14,* 1165–1216.

Koppenaal, L., & Glanzer, M. (1990). An examination of the continuous distractor task and the "long-term recency effect." *Memory & Cognition, 18,* 183–195.

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

Levy, J. P., & Bairaktaris, D. (1995). Connectionist dual-weight architectures. *Language and Cognitive Processes, 10,* 265–283.

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

Lovett, M. C., Reder, L. M., & Lebiere, C. (1999). Modelling working memory in a unified architecture: An ACT-R perspective. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 135–182). Cambridge, England: Cambridge University Press.

Luce, R. D. (1959). *Individual choice behavior.* New York: Wiley.

Madigan, S. A., & McCabe, L. (1971). Perfect recall and total forgetting: A problem for models of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 10,* 101–106.

Marr, D. (1971). Simple memory: A theory for archicortex. *Philosophical Transactions of the Royal Society of London, 262,* 23–81.

Martin, N., & Saffran, E. M. (1997). Language and auditory–verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology, 14,* 641–682.

Martin, R. C. (2003). Language processing: Functional organization and neuroanatomical basis. *Annual Review of Psychology, 54,* 55–89.

Martin, R. C., & Romani, C. (1994). Verbal working memory and sentence comprehension: A multi-components view. *Neuropsychology, 9,* 506–523.

Martin, R. C., Shelton, J. R., & Yaffee, L. S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language, 33,* 83–111.

Mensink, G.-J., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review, 95,* 434–455.

Metcalfe, J., & Murdock, B. B. (1981). An encoding and retrieval model in single-trial free recall. *Journal of Verbal Learning and Verbal Behavior, 20,* 161–189.

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *Journal of Neuroscience, 16,* 5154–5167.

Murdock, B. B., Jr. (1960). The distinctiveness of stimuli. *Psychological Review, 67,* 16–31.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Verbal Learning and Verbal Behavior, 64,* 482–488.

Murdock, B. B. (1967). Recent development in short-term memory. *British Journal of Psychology, 58,* 421–433.

Murdock, B. B. (1972). Short-term memory. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation: Vol. 5* (pp. 67–127). New York: Academic Press.

Murdock, B. B., & Okada, R. (1970). Interresponse times in single-trial free recall. *Journal of Experimental Psychology, 86,* 263–267.

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

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

Neath, I. (1993b). Distinctiveness and serial position effects in recognition. *Memory & Cognition, 21,* 689–698.

Neath, I., & Crowder, R. G. (1990). Schedules of presentation and temporal distinctiveness in human memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 316–327.

Nilsson, L.-G., Wright, E., & Murdock, B. B. (1975). The effects of visual presentation method on single-trial free recall. *Memory & Cognition, 3,* 427–433.

Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review, 75,* 522–536.

Norman, K. A., & O'Reilly, R. C. (2003). Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. *Psychological Review, 110,* 611–646.

O'Reilly, R. C., Braver, T. S., & Cohen, J. D. (1999). A biologically-based computational model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 375–411). Cambridge, England: Cambridge University Press.

O'Reilly, R. C., & McClelland, J. L. (1994). Hippocampal conjunctive encoding, storage, and recall: Avoiding a tradeoff. *Hippocampus, 6,* 661–682.

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain.* Cambridge, MA: MIT Press.

O'Reilly, R. C., & Rudy, J. W. (2001). Conjunctive representations in learning and memory: Principles of cortical and hippocampal function. *Psychological Review, 108,* 311–345.

Palladino, P., Cornoldi, C., De Beni, R., & Pazzaglia, F. (2001). Working memory and updating processes in reading comprehension. *Memory & Cognition, 29,* 344–354.

Peterson, L., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58,* 193–198.

Philips, J. L., Shiffrin, R. M., & Atkinson, R. C. (1967). Effects of list length on short-term memory. *Journal of Verbal Learning and Verbal Behavior, 6,* 303–311.

Poltrock, S. E., & MacLeod, C. M. (1977). Primacy and recency in the continuous distractor paradigm. *Journal of Experimental Psychology: Human Learning and Memory, 3,* 560–571.

Raaijmakers, J. G. W. (1993). The story of the two-store model of memory: Past criticisms, current status, and future directions. In D. E. Meyer & S. Kornblum (Eds.), *Attention and performance XIV: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 467–488). Cambridge, MA: MIT Press.

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

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

Raser, G. A. (1972). Recoding of semantic and acoustic information in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 11,* 692–697.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108.

Raymond, B. J. (1969). Short-term storage and long-term storage in free recall. *Journal of Verbal Learning and Verbal Behavior, 8,* 567–574.

Romani, C., & Martin, R. C. (1999). A deficit in the short-term retention of lexical–semantic information: Forgetting words but remembering a story. *Journal of Experimental Psychology: General, 128,* 56–77.

Rundus, D., & Atkinson, R. C. (1970). Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior, 9,* 99–105.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery and Psychiatry, 20,* 11–21.

Sehulster, J. R. (1989). Content and temporal structure of autobiographical knowledge: Remembering twenty-five seasons at the Metropolitan Opera. *Memory & Cognition, 17,* 59–606.

Shallice, T. (1975). On the contents of primary memory. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 269–280). London: Academic Press.

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: A neuropsychological study. *Quarterly Journal of Experimental Psychology, 24,* 30–40.

Shiffrin, R. M. (1970). Memory search. In D. A. Norman (Ed.), *Models of human memory* (pp. 375–447). New York: Academic Press.

Shiffrin, R. M. (1976). Capacity limitations in information processing, attention and memory. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes: Vol. 4. Attention and memory* (pp. 177–236). Hillsdale, NJ: Erlbaum.

Shiffrin, R. M., & Cook, J. R. (1978). Short-term forgetting of item and order information. *Journal of Verbal Learning and Verbal Behavior, 17,* 189–218.

Shulman, H. G. (1970). Encoding and retention of semantic and phonemic information in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 499–508.

Shulman, H. G. (1971). Similarity effects in short-term memory. *Psychological Bulletin, 73,* 399–415.

Talmi, D., Grady, C. L., Goshen-Gottstein, Y., & Moscovitch, M. (in press). Neuroimaging the serial position curve: An fMRI test of the single-process vs. the dual process (STM-LTM) model. *Psychological Science*.

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.

Taylor, J. G., & Taylor, N. R. (2000). Analysis of recurrent cortico-basal ganglia-thalamic loops for working memory. *Biological Cybernetics, 82,* 415–432.

Thapar, A., & Greene, R. L. (1993). Evidence against a short-term store account of long-term recency effects. *Memory & Cognition, 21,* 329–337.

Twain, M. (1940). The report of my death. In B. De Voto (Ed.), *Mark Twain in eruption* (pp. 252–253). New York: Harper.

Tzeng, O. J. (1973). Positive recency effect in a delayed free recall. *Journal of Verbal Learning and Verbal Behavior, 12,* 436–439.

Usher, M., & Cohen, J. D. (1999). Short-term memory and selection processes in a frontal-lobe model. In D. Heinke, G. W. Humphries, & A. Olsen (Eds.), *Connectionist models in cognitive neuroscience* (pp. 78–91). New York: Springer-Verlag.

Usher, M., Cohen, J. D., Servan-Schreiber, D., Rajkowski, J., & Aston-Jones, G. (1999, January 22). The role of locus coeruleus in the regulation of cognitive performance. *Science, 283,* 549–554.

Usher, M., & Davelaar, E. J. (2002). Neuromodulation of decision and response selection. *Neural Networks, 15,* 635–645.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review, 108,* 550–592.

Wagner, A. D., Pare-Blagoev, E. J., Clark, J., & Poldrack, R. A. (2001). Recovering meaning: Left prefrontal cortex guides controlled semantic retrieval. *Neuron, 31,* 329–338.

Ward, G. (2002). A recency-based account of the list length effect. *Memory & Cognition, 30,* 885–892.

Warrington, E. K., & Shallice, T. (1969). The selective impairment of auditory verbal short-term memory. *Brain, 92,* 885–896.

Watkins, M. J. (1974). Concept and measurement of primary memory. *Psychological Bulletin, 81,* 695–711.

Waugh, N. C. (1970). Retrieval time in short-term memory. *British Journal of Psychology, 61,* 1–12.

Waugh, N. C., & Norman, D. A. (1965). Primary memory. *Psychological Review, 72,* 89–104.

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

Wickelgren, W. A. (1973). The long and short of memory. *Psychological Bulletin, 80,* 425–438.

Wickens, D. D. (1970). Encoding categories of words: An empirical approach to meaning. *Psychological Review, 77,* 1–15.

Wickens, D. D., Born, D. G., & Allen, C. K. (1963). Proactive inhibition and item similarity in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 2,* 440–445.

Wixted, J. T., & Rohrer, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1024–1039.

(*Appendixes follow*)

Appendix A

Modeling Protocol and Parameter Values

In all simulations, the same set of parameters was used. The general procedure for all free recall simulations was as follows. The total number of item-representing units in the network was 40. Only those units receiving sensory input were part of the list. We modeled a sequential list presentation by changing the input to the unit for the next item every $t = 500$ time steps. At each time step, the activation function (Equation 1) is updated for all 40 units. The distractor interval is approximated by sequentially activating 12 distractor units (i.e., units that are not part of the list), each for 500 iterations with the same amount of sensory input as list items. These units enter the short-term buffer, displacing its current contents. The distractor units do not take part in the selection and recovery processes.

Retrieval starts with the unloading of active items, followed by the competitive recall of inactive items, which are encoded in episodic memory. During list presentation (every 500 time steps) and at every retrieval attempt, context changes probabilistically according to a random walk with a drift.

On every trial the following protocol is used:

```
0 Set all activations to zero (for $i = 1$ to $40 x[i] = 0$); start with context unit $c = 0$

Encoding
1 For every list item do
  1.1 Maintain context unit $c$ with probability $1 - P^+ - P^-$ or deactivate context unit $c$ and activate context unit $c + 1$ with probability $P^+$ or $c - 1$ with probability $P^-$
  1.2 Reset the input to unit $i$, $I[i] = I_0$, $I[j] = 0$ $(j \neq i)$
  1.3 For 500 time steps do
    1.3.1 Update activations of all units in parallel according to Equation 1
    1.3.2 Increase connection weight between active buffer units and the active context unit according to Equation 2

Retrieval
2 Check for active items above the buffer threshold and output those in order of their episodic strength
3 For $k$ attempts do
  3.1 Maintain context unit $c$ with probability $1 - P^+ - P^-$ or deactivate context unit $c$ and activate context unit $c + 1$ with probability $P^+$ or $c - 1$ with probability $P^-$
  3.2 Select one item from the lexicon (40 units) according to Equation 4
  3.3 If selected item was not already retrieved, attempt a recovery according to Equation 5; else do nothing (silent event)

4 Reset context to $c = 0$ (i.e., the start context)
5 Repeat Step 3 and then end
```

In all simulations, 1,000 trials were run and averaged to obtain serial-position functions. The parameters for all simulations were the following: $\lambda = .98$; $\alpha_1 = 2.0$; $\beta = .20$; $I = .33$; $\sigma = 1.0$; $\varepsilon = .02$; $\varphi_1 = .20$; $\varphi_2 = 4.0$; $\rho = 1.0$; $\tau = 2.0$ (1.0 for Simulation 4); $k = 20$; $P^+ = .10$; $P^- = .05$; number of nonlist distractor items = 12. Typical values of W in our simulations ranged from 2 to 6.

In the proactive interference simulation, two consecutive trials were simulated in which the last active context unit at the end of List 1 retrieval was the start context for List 2 encoding. During retrieval (of both List 1 and List 2), in the selection and recovery formulas (Equations 3 and 4, respectively), W_i was replaced by $W_i + g$, where $g = 1.2$ and represents a constant semantic input. In the amnesia simulation, 50% of the context units had their item–context connections set to 0.

Appendix B

Dynamics of the Retrieval Process

End-Context Retrieval Dynamics (Order of Retrieval From the Buffer)

In the Model section, we assume that the order of output from the buffer follows the episodic strengths between the context and the buffer layer. Naively, however, one may expect that the order of retrieval from the buffer has to follow the activation levels (i.e., that items with higher activations are retrieved first), but we show here that this is not necessarily the case, because of the interaction between the buffer and the episodic system during the retrieval process.

To model the retrieval, we made a simplifying assumption (consistent with much of the cognitive literature; Ratcliff, 1978; Usher & McClelland, 2001) that items are retrieved for output when they reach a retrieval threshold. This threshold needs to be higher than the typical activation of the items in the buffer (to prevent output during encoding) and surely higher than the activation threshold (.2), which determines encoding into the episodic system. In Figure B1, we chose the value of this threshold to be .65. We further assume that the retrieval is context driven, even when items are in the activation buffer. Accordingly, at retrieval, the active context sends activation to the item units that are connected with it, in proportion to the strength of the corresponding episodic links (Davelaar, 2003; Humphreys, Bain, & Pike, 1989). Furthermore, we assume that, as proposed in the *Extending to Response Latencies* section, at retrieval the parameters of the recurrent buffer network are modulated by increasing both the mutual inhibition (β to .50) and the self-excitation (α_1 to 3.5).[B1] We find that under this condition, the units in the activation buffer have a start advantage relative to nonbuffer items and therefore are retrieved first. Nevertheless, the order of retrieval depends jointly on both the buffer activation levels and the episodic strengths. To illustrate this dependency, we show in Figure B1 a noiseless simulation in which a list of six items is presented and is followed by the cue-driven retrieval. We assume that during retrieval, once an item reaches the retrieval threshold it is inhibited to 0, enabling the other items to compete for retrieval. That is,

---

[B1] This is different from the situation in category-cued recall (see the Utility of the Dynamic Buffer section) in which only the inhibition was modulated in order to select a single item. In free recall, a single item selection, however, is counterproductive. We assume, therefore, that retrieval neuromodulation is adaptive and that as a result, both α_1 and β increase. The result is a less restrictive selection process in which all the active items compete in an accelerated way for selection.

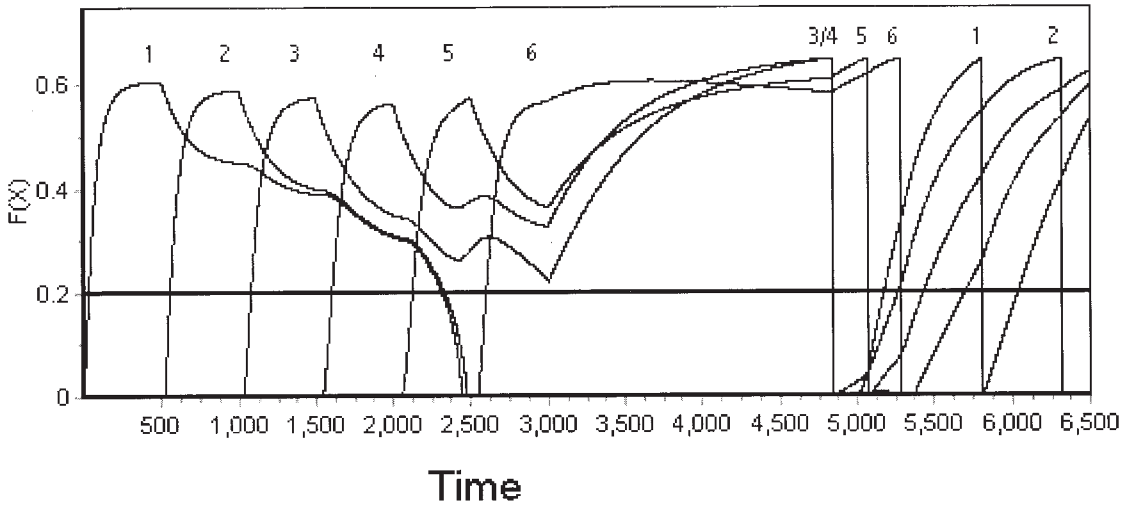

Figure B1. Activation trajectories of a six-item list during list presentation and dynamical retrieval. At the end of the list presentation (at t = 3,000), the last four items are active above threshold (φ_1 = .20). From t = 3,000, all items receive context-driven input that is proportional to the episodic strengths. Immediately after an item reaches an upper threshold (at .65), the item is inhibited to 0. F(x) = output activation value.

we assume that context does not change, and therefore the same context unit is connected to all the items, with encoding strengths given by integrating Equation 2 (β = .15; ε = .02) during list presentation.

We can see that at the end of the list presentation, four out of the six items are still active, and their activation magnitudes increase with serial position (the most recent items are more active; see Figures B1 and B2). We can first see that the activated items in the buffer (3, 4, 5, and 6) are retrieved first (before nonactive items 1 and 2). Second, we can see that the output order among these items is mainly in the forward direction. This is because the order of retrieval depends jointly on the activation levels and on the episodic strengths, with the latter being the dominant factor. As discussed in Simulation 5, unlike the activation levels in the buffer, the episodic strengths decrease with serial position (see Figure B2). As a result, Items 3 and 4 reach the retrieval threshold first, at t = 4,840 iterations. Those items have higher episodic strengths than the other active items, and their relative difference in activation levels (at the end of list presentation, t = 3,000) compensates for their relative difference in episodic encoding. The next items retrieved are 5 (at t = 5,066) and 6 (at t = 5,285). For these items the small difference in activation levels is overridden by the larger difference in episodic strengths. The nonactive units, 1 and 2, become active only later (even though all units receive context-driven input simultaneously) and reach the retrieval threshold after a relatively larger time gap (t = 5,812 and t = 6,317, respectively). Thus, the overall tendency is that buffer items are reported first, but their order is mainly in the forward direction. This justifies the model assumption (order of retrieval from the buffer according to episodic strengths), which was critical for accounting for the dissociation between output order and lag recency in Figure 11 (left). This assumption is definitely a simplification, and it may overestimate the forward lag-recency asymmetry, but it provides an upper boundary for this effect. A smaller asymmetry can be easily obtained by increasing the retrieval noise or by assuming a lower weight for the episodic component during retrieval.

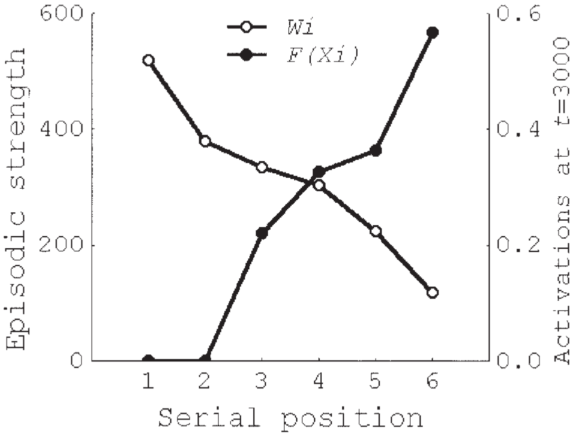

Figure B2. Episodic strength values and activation levels at t = 3,000 for the six items in Figure B1, showing the primacy gradient in the episodic strengths (*Wi*) and the recency gradient in the activation levels (*F*(*Xi*)).

In addition to accounting for the asymmetry in lag recency, this simulation can also explain why in some studies (e.g., Murdock & Okada, 1970) the item presented three positions from the end of the list has the highest first-recall probability. The interaction between the activation buffer and the episodic system is such that depending on the balance between the activation levels and the episodic strengths, the item that has the highest probability of being retrieved first is a nonterminal list item (that is still in the activation buffer). As current single-store models assume that all items have equal episodic strengths and that context favors later items, there is no mechanism that can account for situations in which first-recall probabilities are larger for nonterminal than for terminal items without introducing additional processes.

Start-Context Retrieval Dynamics

When participants are instructed to report items that were presented at the beginning of the list, the start context is used to drive retrieval (see Simulation 3). In our model, we assumed that in this situation, the buffer units are deactivated so that the items linked to the

(*Appendixes continue*)

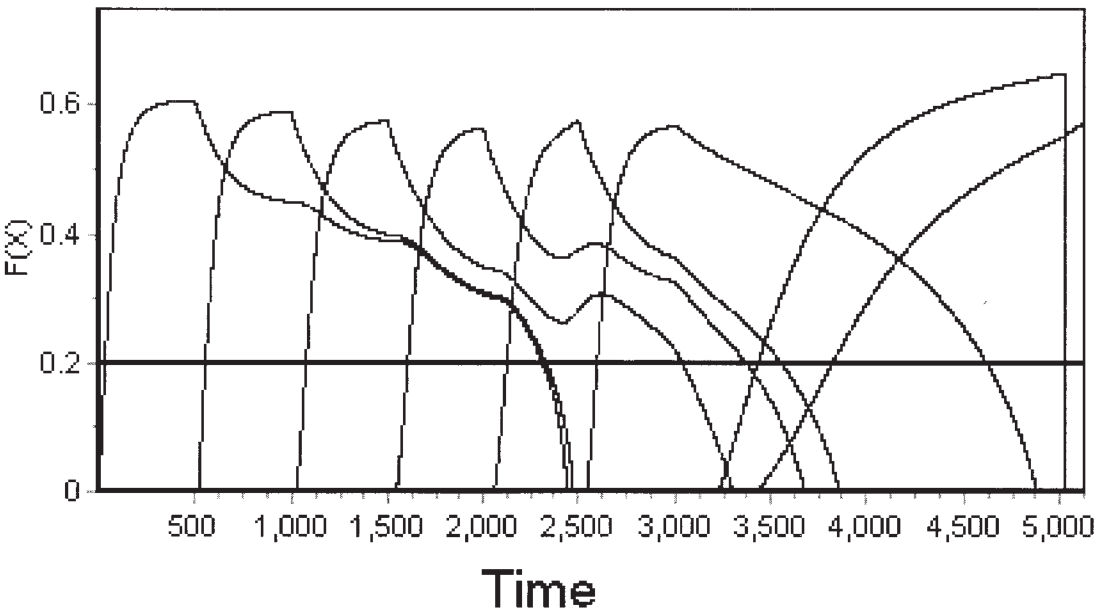

Figure B3. Activation trajectories of a six-item list during list presentation and dynamical retrieval. At the end of the list presentation (at $t = 3{,}000$), the last four items are active above threshold ($\varphi_1 = .20$). From $t = 3{,}000$, only Items 1 and 2 receive context-driven input, because of the use of the start context. When an item reaches an upper threshold (at .65) the item is inhibited to 0. Note that buffer items do not reach this threshold and are instead deactivated. F(x) = output activation value.

start context can be retrieved. This assumption can be justified if one assumes that participants have control over the level of inhibition and excitation so that when the start context is accessed, only the inhibition (and not excitation) increases. (This is an adaptive strategy, because maintaining the content of the buffer is counterproductive for this situation.) As can be observed in Figure B3, this leads to the beginning-of-the-list items being reported first and displacing items that were still in the buffer (see Simulation 3). Notice also that unlike in the end context retrieval dynamics, here the buffer units receive no contextual support.

Appendix C

Comparing the Traditional Buffer With the Activation Buffer

Here, we compare displacement in the activation buffer with displacement in traditional mathematical buffers that are based on the computer metaphor. The traditional buffer is a system with a fixed number of r slots that can be filled by items (chunks of information). Whenever the system is filled to capacity, subsequent items displace items that are maintained in the system. The displacement process can be either random (Atkinson & Shiffrin, 1968; Raaijmakers & Shiffrin, 1980) or biased toward items that have been in the buffer longest (knock-out buffer; Philips et al., 1967), as governed by Equation C1. Here, d_i is the probability that the ith item (with $i = 1$ being the oldest buffer item and $i = \mathrm{r}$ being the most recent item) will be displaced from the buffer (with capacity r) and replaced with the incoming item, and δ is a parameter that corresponds to the slope of the displacement function:

$$d_i = \delta(1 - \delta)^{i-1}/[1 - (1 - \delta)^r]. \tag{C1}$$

The serial-position functions generated by these two variants with capacity $r = 3$, $\delta_{\text{knock-out}} = .5$ and $\delta_{\text{random}} = .001$, are presented in Figure C1. The important point to note is that the random buffer has a recency function that is J-shaped (exponentially decaying), whereas the knock-out buffer shows an S-shaped recency function. Both shapes have been reported (for a

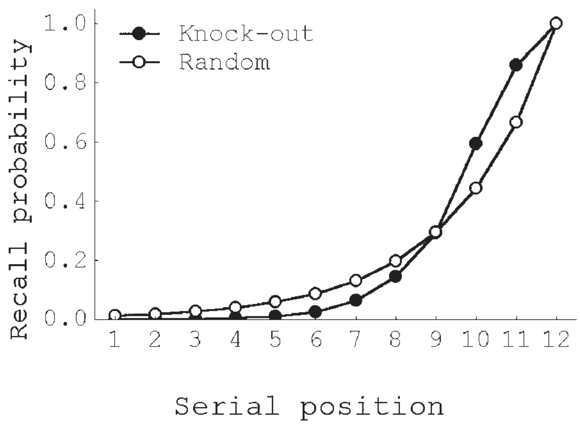

Figure C1. Serial-position functions for a traditional buffer model with a random or knock-out displacement process.

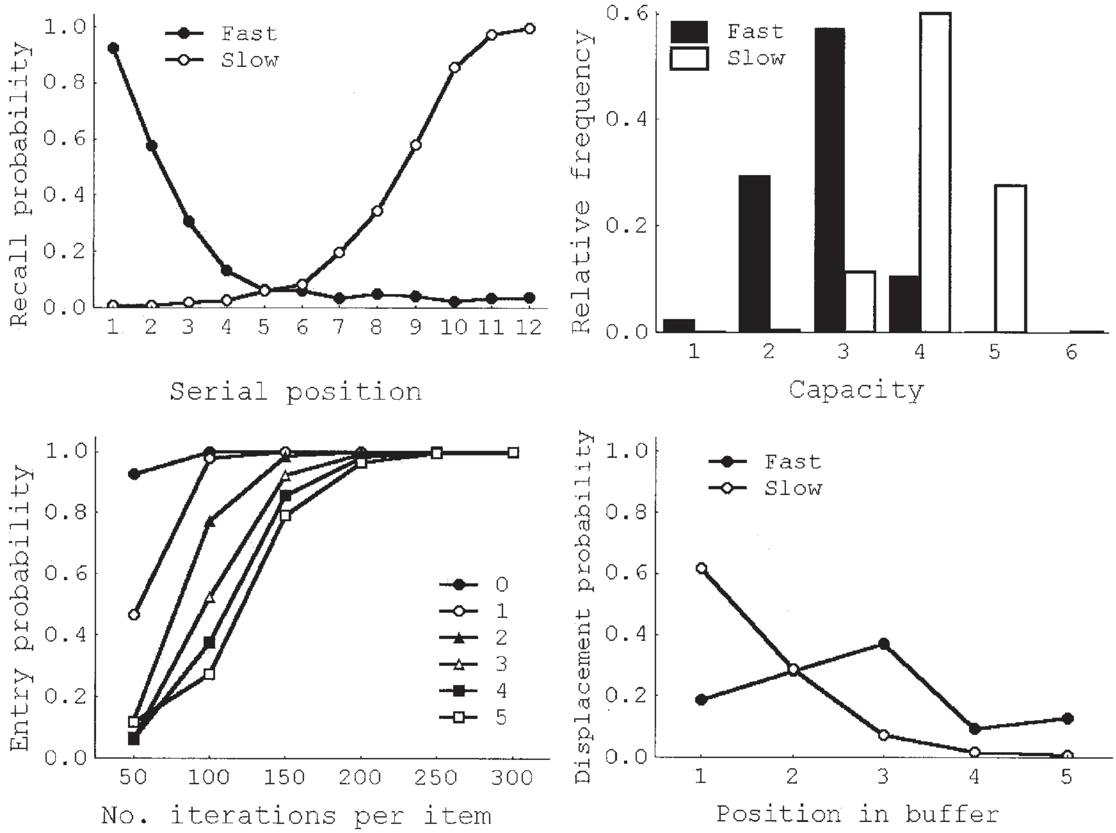

Figure C2. Details of the effect of presentation rate on the dynamics of the activation buffer. Top left: Serial-position functions for slow (250 iterations per item) and fast (50 iterations per item) presentation rates, showing a switch from recency to primacy, respectively. Top right: Distribution of capacity (i.e., the number of items active at the end of the sequence). Bottom left: Probability that a presented item will enter the buffer, as function of presentation rate and the number of already-active items. Bottom right: Distribution of displacement probabilities, as function of the buffer position and presentation rate.

discussion of an S-shaped recency function, see Murdock, 1962), suggesting that the knock-out buffer captures an aspect of the human short-term memory system. However, neither buffer variant (as used in the literature) is sensitive to presentation rate.

As shown in Figure 14 in the main text, the activation-based recency function of the activation buffer is sensitive to presentation rate. Figure C2 (top left) presents recency functions of the activation buffer for slow and fast presentation rates. As discussed in the main text and verified in

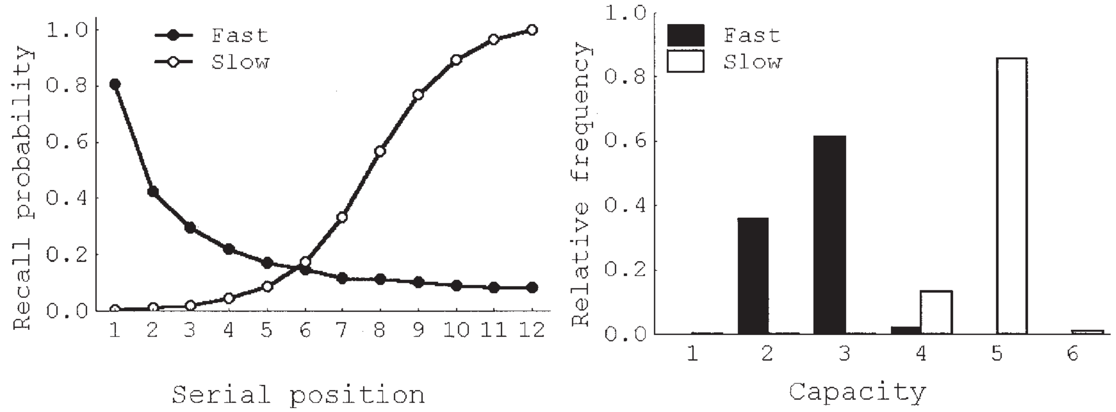

Figure C3. Results of a knock-out buffer in which the probability of entering the buffer is made conditional on the presentation rate and the number of items in the buffer. Note the shift from recency to primacy (left) and the decrease in the distribution of capacities (right) with the increase in presentation rate.

(*Appendixes continue*)

Experiment 2, the activation buffer predicts an S-shaped recency function at slow presentation rates and a primacy function at fast presentation rates. The activation buffer differs from the traditional buffer in the distribution of capacities. By implementation, the traditional buffer model has a fixed capacity, whereas the activation buffer shows a distribution of capacities (see Figure C2, top right), which shifts toward lower capacities with an increase in presentation rate.[C1]

Figure C2 (bottom left) also presents the probabilities that a presented item will enter the buffer (become active above threshold) as a function of a range of presentation rates and as a function of the number of active items at the time the item is presented. This is dramatically different from the traditional buffer, in which the entry probability is always unity. The dynamics of the activation buffer are such that at slow presentation rates, presented items have a high probability of entering the buffer, independent of the number of items that are already in the buffer. However, the faster the presentation rate the less likely a presented item's representation will be activated sufficiently to overcome the inhibition in the system. By the same token, the probability of entering the buffer decreases with the number of active items (which increases the amount of inhibition in the system). Finally, Figure C2 (bottom right) presents the distribution of displacement probabilities as a function of buffer position for the slow and fast presentation rates (averaged for situations with four or five items in the buffer). Of importance, this panel shows that with an increase in presentation rate, the displacement process becomes more random, as illustrated by a flat probability distribution for a fast presentation rate.

The analysis presented above suggests that one can see the activation buffer as a dynamic model that interpolates between and extends different types of traditional buffer models. Because such models are simple and easy to follow, one way to further explicate the activation buffer's function is to approximate it in terms of traditional buffers. To do this, we chose the knock-out buffer and added a parameter for entering the buffer based on the number of items that are already active and the presentation rate. The values for the probability of entering the buffer were taken from Figure C2 (bottom left) with $\delta_{\text{slow}} = .5$ and $\delta_{\text{fast}} = .01$. Figure C3 shows that the addition of this parameter captures the results that with an increase in presentation rate, there is a shift from recency to primacy (see Figure C3, left) and that the distribution of capacities becomes centered around a lower average (see Figure C3, right).[C2]

Received July 21, 2003
Revision received March 18, 2004
Accepted March 18, 2004 ■

[C1] The traditional buffer could be extended by having the capacity of the buffer being drawn from a distribution on every trial (e.g., Kahana, 1996), but that would not change the next difference between the two buffers.

[C2] It should be noted that in the original model of Atkinson and Shiffrin (1968), there was a parameter that governed the probability of entering the buffer (varying between .39 and .65 in model fits), but this parameter was fixed at unity in later work (e.g., Raaijmakers & Shiffrin, 1980). This was probably because these researchers modeled data from experiments that used slow presentation rates (larger or equal to 1 s per item).