

Running head: CONTEXT MAINTENANCE AND RETRIEVAL

A context maintenance and retrieval model of organizational processes in free recall

Sean M. Polyn

University of Pennsylvania

Kenneth A. Norman

Princeton University

Michael J. Kahana

University of Pennsylvania

Sean M. Polyn

University of Pennsylvania, Department of Psychology

3401 Walnut St., Room 303c, Philadelphia, PA 19104

e-mail: polyn@psych.upenn.edu

phone: (215) 573-3365

fax: (215) 746-6848

Version of October 7, 2008

Abstract

We present the Context Maintenance and Retrieval (CMR) model of memory search, a generalized version of the temporal context model (TCM) of Howard and Kahana (2002a), which proposes that memory search is driven by an internally maintained context representation composed of stimulus-related and source-related features. In the CMR model, organizational effects (the tendency for related items to cluster during the recall sequence) arise as a consequence of associations between active context elements and features of the studied material. Semantic clustering is due to longstanding context-to-item associations, whereas temporal clustering and source clustering are both due to associations formed during the study episode. A behavioral investigation of the three forms of organization provides data to constrain the CMR model, revealing interactions between the organizational factors. Finally, we discuss the implications of CMR for our understanding of a broad class of episodic memory phenomena, and suggest ways in which this theory may guide our exploration of the neural correlates of memory search.

Keywords: context; organization theory; source; encoding task; computational model.

**A context maintenance and retrieval model of organizational
processes in free recall**

Introduction

The free-recall paradigm has had an important role in the development of theories of memory search, and research in this domain has tended to follow two threads. The first deals with organization; the free-recall paradigm reveals structure and organization in memory through the ways items tend to cluster in recall sequences (for a review see Puff, 1979). When items are recalled successively, it indicates that they are somehow related or structurally connected in the memory system. The second thread deals more directly with the memorability or availability of individual items in memory, often in terms of the sort of processing the item received. In these studies theoretical attention focuses on whether particular items were recalled, and not on the order of those recalls. These studies have dominated the recent literature, and tend to employ a variant of the free-recall paradigm designed to minimize organizational influences, through the use of a single trial and randomly chosen items (Battig & Bellezza, 1979). However, recent work has shown that even in the single-trial free-recall paradigm, one can see substantial and reliable organizational influences, both in terms of the temporal contiguity of successively recalled items (Kahana, 1996) and in terms of the semantic relatedness of the studied items (Howard & Kahana, 2002b).

Models of the free-recall paradigm have tended to ignore order information as a simplifying assumption (Brown, Neath, & Chater, 2007; Wixted & Rohrer, 1993). However, a recent trend in the modeling literature has brought these organizational principles to center stage. The temporal context model (TCM) of Howard and Kahana (2002a) is a formal computational model of the human memory system designed to explain the phenomena of temporal organization. By this model, a slowly changing internal context representation is associated with each of the studied items, and is then used to guide memory search. TCM is a model of the interactions between context and content, but it lacks the machinery to explain the important role of non-temporal factors

in memory retrieval, such as semantic and source information (e.g., Howard & Kahana, 2002b; Hintzman, Block, & Inskip, 1972). We generalize TCM to model the semantic similarity relations between words, as well as the influence of source context on the recall process. We refer to this generalized model as the Context Maintenance and Retrieval (CMR) model. According to the CMR model, the most obvious behavioral manifestation of organization is the clustering of recalled items along a dimension of similarity. This similarity may arise due to the longstanding associative relations between studied items (giving rise to semantic organization), or due to the similarity structure of an internal context representation that is associated with the studied items during the learning episode (giving rise to episodic clustering). A fine-grained analysis of clustering behavior reveals the structure of the representations in the memory system, and provides insights into the dynamics of memory search. Before describing the machinery of CMR, we first review a range of organizational phenomena in the human memory literature.

Clustering and the organization of memory

Early studies of organization focused on clustering by the semantic category of the words (Bousfield & Sedgewick, 1944; Bousfield, 1953; Cofer, Bruce, & Reicher, 1966). While these early studies focused on the clustering of words drawn from taxonomic categories, even the weaker associations between randomly chosen words influence the output order of recalled items (Howard & Kahana, 2002b). These studies characterize *semantic clustering* related to the longstanding associations between words.

Kahana (1996), in reanalyzing a number of classic free-recall studies, showed that *temporal clustering* seems to be a ubiquitous property of the recall sequences (see also Kahana, Howard, & Polyn, 2008). This form of episodic clustering is perhaps best exemplified by the contiguity effect, the observation that items studied in neighboring list positions tend to be reported successively during the recall period, regardless of their

degree of semantic association (Kahana, 1996). Kahana (1996) introduced a conditional response probability analysis as a function of lag (or lag-CRP) to show that the probability of successively recalling two items falls off smoothly as the temporal distance (lag) between them increases (see also Howard & Kahana, 1999; Ward, Woodward, Stevens, & Stinson, 2003; Hulme, Stuart, Brown, & Morin, 2003; Unsworth, in press; Lewandowsky, Brown, & Thomas, submitted). Below, we will look closer at this phenomenon including at a recently characterized tendency to recall temporally distant items early in the recall sequence (Farrell & Lewandowsky, in press).

Another form of episodic clustering is observed based upon associations between the studied items and their source characteristics. The earliest observations of *source clustering* arose from within-list manipulations of the modality of the studied word (auditory or visual). Murdock and Walker (1969) showed that along with the superior recall of the auditory items, words associated with each modality clustered together during the recall sequence (see also Hintzman et al., 1972). Following the demonstration of organization by modality, researchers found that a number of associated source characteristics could induce clustering in the free-recall paradigm. These include similarities in shape or orientation of a picture of an item (Frost, 1971), the gender of the presenter's voice, the typeface of the word (Hintzman et al., 1972; Nilsson, 1974), and the spatial location of a word on a screen (Curiel & Radvansky, 1998).

Just as participants may organize material according to external source features, a similar kind of organization might be expected based on similarity in internal representations activated during processing (i.e., internal source features). For example, Cohen, Dunbar, and McClelland (1990) proposed that different processing tasks have distinct task representations that guide the cognitive system to flexibly process incoming stimuli in accordance with task demands. In the Stroop task, a task representation for "color naming" would allow one to name the color of ink a word was written in, instead of

reading the text of the word. The idea that task is part of context in memory has been advanced by a number of researchers to explain a diverse array of cognitive phenomena (e.g., Kolers & Ostry, 1974; Kolers & Roediger, 1984; Braver et al., 2001; Botvinick & Plaut, 2002). If internal task representations are associated with the features of items that are studied in their context, one would expect to observe clustering by orienting task during memory search for the studied items. Polyn (2005) reported direct evidence for organization by internal source features in a continuous-distraction free-recall paradigm (see also Polyn, Norman, & Kahana, Submitted), in which each studied item was encoded with one of two orienting tasks (a size judgment and a pleasantness judgment). In lists where half of the items were studied with each task, *task clustering* was observed during recall.

Although these distinct forms of clustering have been studied in separate experiments, it is easy to show that any free-recall paradigm will give rise to multiple forms of clustering. However, little is known about how different forms of clustering interact during recall. The CMR model is designed to explain simultaneous organization by multiple factors (semantic, temporal, and source), and suggests that two principles can explain clustering behavior in the free-recall paradigm: First, the principle of clustering by similarity states that clustering along a dimension of similarity arises when items are associated with similar contextual states, because the recall process is driven by the current state of the context representation. Second, the principle of clustering by isolation states that a sudden shift in context (caused by a disruptive cognitive event) can isolate a set of items from the items studied prior to the disruptive event. This causes the isolated items to cluster together in the recall sequence (relative to a condition without such a disruptive event). Here, we explore the possibility that such a disruptive context shift can be triggered by the detection of a sufficiently novel representation (e.g., a change in orienting task causing a new task representation to become active).

The idea that novelty is treated specially by the system has been used in a number of models of human memory. One prominent example arises from the work of Donchin and colleagues (Karis, Fabiani, & Donchin, 1984; Donchin & Coles, 1988; Fabianai & Donchin, 1995), who proposed that a context updating process is engaged whenever one encounters an item that mismatches previous items on some salient dimension, and used this theory to explain distinctiveness effects in free recall. Similar mechanisms have been proposed to explain distinctiveness effects (as well as the primacy effect) in serial recall (Lewandowsky & Farrell, in press; Farrell & Lewandowsky, 2002; Brown, Preece, & Hulme, 2000). Recently, Sahakyan and Kelley (2002) showed that if a period of elaborative mental activity (e.g., imagine what you would do if you were invisible) is inserted between two studied lists, one observes both a cost to memorability of the material studied prior to the elaborative activity as well as a benefit to the memorability of material studied after the activity. It is quite possible that performing this novel task disrupted temporal context, making the items studied in the first list less accessible, and also reducing the degree of interference between these items and the items in the second list. This hypothesis is quite consistent with their contention that the elaborative activity causes a shift in inner mental context, and that a mismatch between inner mental context during retrieval and inner mental context during study reduces accessibility of the studied material.¹

The spotlights of memory

The CMR model proposes that the process of memory search is driven by a set of internally maintained context representations that are used to probe associative weights in order to reactivate the features of studied items. Each maintained context representation plays the role of a spotlight, sweeping across a stage on which a set of items have been placed (Figure 1). For this example, imagine that some of the items have been studied with source A, and others with source B. The strength of the context representation

determines the overall intensity of the spotlight, and the set of associative links to the item features determines the intensity with which it illuminates any one item. The intensity with which each item is illuminated directly influences its likelihood of being recalled in a competition where all of the items compete in parallel to have their features reinstated in the system. The specific item recalled brings with it retrieved context which alters the context representation (shifting the spotlights), and another recall competition ensues. In other words, while the recall process is fundamentally parallel, the full recall period comprises a series of these recall competitions. Recall in the CMR model is perhaps best described as an iterative parallel process, where the result of each recall competition affects the course of the subsequent competition.

Figure 1 contains two spotlights, one for temporal context (illuminating items studied nearby in time, and weakening with temporal distance), and one for source context (illuminating items studied with the same source). The set of items illuminated by each spotlight depends on its current state (i.e., the currently active context representation). If the source context representation A is active, then items studied with source A are illuminated. The state of the context representation is determined in part by the context retrieved by the most recently recalled item. If an item studied with source B is recalled, then the source context spotlight will shift to illuminate source B items. Clustering occurs when a spotlight is trained on a particular set of items for more than one recall attempt. Recalled items trigger the retrieval of associated temporal and source information, which keeps the spotlights trained on items studied in a similar temporal context and source context. This raises the likelihood that the next recalled item will be from a nearby list position (giving rise to temporal clustering) or from the same source (giving rise to source clustering). As recall proceeds, context reinstatement from retrieved items causes the spotlights to sweep over the list, until time runs out or no more items can be retrieved.

The item information activated in temporal context concurrently illuminates the

semantic associates of the studied items, due to the longstanding associations connecting the context features to semantically related items (this aspect of the spotlights is not graphically represented in Fig. 1, but receives further attention below). The spotlight metaphor is useful for gaining an intuition regarding the associative basis of clustering in the CMR model. Below we discuss how a second mechanism (novelty-related context disruption) may produce clustering as well.

Insert Figure 1 about here

Precis

Memory search in free recall is best understood as a multiply constrained process; the probability of recalling an item, and the order with which items are recalled, are simultaneously influenced by semantic, temporal, and source information. In the following sections, we present simulations of a set of experiments using the immediate free-recall paradigm. First, we examine the results of a new experiment in which we manipulate the source context associated with studied items within list (details of this paradigm are provided in Appendix A). Then we examine the results of two classic studies of free recall, reported by Murdock (1962) and Murdock and Okada (1970). The model accounts for the benchmark results described by these studies, regarding how the shape of the serial position curve changes with list length, and the exponential growth of inter-response times during recall. It also provides a natural framework for understanding the multiple organizational influences giving rise to clustering during memory search.

- We begin with a description of the components and processes of the CMR model, detailing how semantic, temporal, and source information simultaneously influence recall dynamics. This is followed by four Simulation analyses, which examine the dynamics of the model across three free-recall paradigms.

- Simulation analysis I examines the simultaneous organization of memory search along semantic, temporal, and source dimensions, by examining clustering effects for each of these three factors. Both the data and model demonstrated reliable organizational effects by each of the three factors.

- Simulation analysis II examines the interaction between temporal organization and source organization. The CMR model captures the finding of source organization between both nearby and remote list items, but with much stronger source clustering for items studied in a nearby temporal context. A comparison of different variants of the model suggests that associations between source context and item features are critical to explain source clustering between remote items.

- Simulation analysis III examines memory performance as a function of serial position in the studied list. The CMR model is used to explain serial position effects across a list length manipulation described by Murdock (1962). The model also explains the serial position effects of a within-list source context manipulation. The model suggests that a task-shift related disruption of temporal context is important for understanding the perturbations in the serial position curve due to the within-list source context manipulation.

- Simulation analysis IV examines inter-response times (IRTs) between successively recalled items. The CMR model is used to explain IRT effects in a study described by Murdock and Okada (1970). The model also explains the effects of all three organizational factors (semantic, temporal, and source) on IRTs. The CMR model predicted that participants should show an increased latency to make recall transitions between items associated with different sources (tasks, in this case), which was confirmed upon examination of the empirical data.

The CMR network model of human memory search

Context-based models of free recall

The notion of context considered here is inspired by the stimulus sampling theory of Estes (1959) and Bower (1972), as well as the temporal context model (TCM) of Howard and Kahana (2002a). We conceive of context as a pattern of activity in the cognitive system, separate from the pattern immediately evoked by the perception of a studied item, that changes over time and is associated with other coactive patterns. Along with the theory of Howard and Kahana comes the notion that the elements of context are activated by some stimulus or event, tend to stay active past the time this stimulus leaves the environment, and are associated with the features of studied material. The consequences of this are explored below. While it is clear that static external (e.g., environmental) features can also play the role of context (Smith, 1988; Bjork & Richardson-Klavehn, 1989; Murnane, Phelps, & Malmberg, 1999), for the purposes of the present treatment, we restrict our consideration to context as an internally maintained stimulus.

The CMR model builds upon the TCM framework (Howard & Kahana, 2002a), which describes a mechanism for representing temporal context, and the dynamics of how this representation updates and is associated with the representations of the studied items. Finally, it describes how, given a particular context state, one can calculate the degree of support for each of the items in a lexicon having been in the most recent list. Howard and Kahana (2002a) used the choice probability framework of Luce (1959) as a simple decision mechanism. Recently, Sederberg, Howard, and Kahana (in press) described a variant model, TCM-A, which pairs TCM with a dynamical system capable of modeling a many-dimensional choice problem (Usher & McClelland, 2001), used to model the decision process leading to a recall. The CMR model described below utilizes the decision rule of TCM-A. While TCM-A provides an elegant explanation of temporal

clustering and the effects of distraction on memory search, it does not predict the existence of semantic clustering or other forms of episodic clustering, nor does it address the interactions between these factors.

As mentioned above, CMR is a generalized version of TCM, designed to capture the broader set of organizational effects observed in free recall. Semantic organization arises because when an item is recalled, it retrieves an associated temporal context representation. This retrieved representation contains a blend of all the temporal context representations this particular item has ever been associated with. Semantic associates of a particular item have historically tended to appear in similar contexts, so a recalled item's retrieved temporal context is associated with that item's semantic associates, and will tend to favor their subsequent retrieval. This addition of semantic information to the model is consistent with the principles of TCM, and allows the model to explain a number of aspects of the behavioral data.

The true generalizing principle of the CMR model is that the context-related mechanisms developed for temporal context apply to any sort of context, such as the physical source characteristics of an item, or the internal source context of an item (e.g., task). Associations between the features of studied items and this expanded set of context sub-regions allows the model to explain the simultaneous organization of studied material along multiple dimensions.

The structure and dynamics of the model

Representational structure.

In CMR any given environmental stimulus is composed of some number of features, and the presence of that feature in the environment corresponds to the activation of the corresponding element in the feature layer (Underwood, 1969; Bower, 1967). Figure 2 depicts the basic structure of the model. There are two representational sub-areas: the

feature layer F , and the context layer C . Studied items activate a representation \mathbf{f}_i in F , where the subscript i indexes the list position of the item. In other words, \mathbf{f}_i is a vector representation of the features of the studied item, which consists of both item features and source features concatenated into a single vector representation ($\mathbf{f} = \mathbf{f}^{item} \oplus \mathbf{f}^{source}$). For simplicity, we follow the tradition of TCM in treating the items as orthonormal on F^{item} (each item has a localist representation on F). Similarly, different sources are orthonormal on F^{source} . The state of the context layer at a given list position is \mathbf{c}_i ; the context layer is subdivided into elements corresponding to temporal context and source context ($\mathbf{c} = \mathbf{c}^{temp} \oplus \mathbf{c}^{source}$). In the TCM framework, temporal context was represented as a vector \mathbf{t} , which corresponds to \mathbf{c}^{temp} in the CMR framework. The F layer and the C layer interact through two associative matrices: M^{FC} , which describes the strengths of the feature-to-context associations; and M^{CF} , which describes the strength of the context-to-feature associations. A given element in an associative matrix describes the connection strength between a particular feature element, and a particular context element.

Whereas we only consider two classes of features in the current investigation (item and source, where the source considered is related to the task performed during study), the CMR framework can be easily extended to represent any aspect of a studied stimulus (simply by adding extra elements corresponding to the features in question, e.g., acoustic or orthographic properties of a word, or features of the local environment).

Insert Figure 2 about here

Updating temporal and source context.

When a feature representation is activated in F (following an item presentation during study, or the reactivation of an item representation during recall), information

about the event is integrated into the context representation; the inserted information weakens whatever information was already resident in context, such that the global state of the context representation changes slowly over time. Prior to the start of a trial, each context sub-region is initialized as a vector of unit length; this represents whatever information was present in context prior to the list.

Whenever a representation is activated in F , the input to C is determined as follows:

$$\mathbf{c}^{\text{IN}} = M^{FC} \mathbf{f}_i. \quad (1)$$

The vector \mathbf{c}^{IN} represents the net input to the context layer. As with \mathbf{c} , \mathbf{c}^{IN} consists of two sub-regions, corresponding to temporal and source elements. Each of the two sub-components of \mathbf{c}^{IN} are normalized to be of unit length prior to updating context.

Given \mathbf{c}^{IN} , context integration proceeds as follows:

$$\mathbf{c}_i = \rho_i \mathbf{c}_{i-1} + \beta \mathbf{c}^{\text{IN}}, \quad (2)$$

where

$$\rho_i = \sqrt{1 + \beta^2 [(\mathbf{c}_{i-1} \cdot \mathbf{c}^{\text{IN}})^2 - 1]} - \beta (\mathbf{c}_{i-1} \cdot \mathbf{c}^{\text{IN}}). \quad (3)$$

Here, β is a scaling parameter that determines how much new information (\mathbf{c}^{IN}) is placed in context, and ρ_i weakens the current state of context (\mathbf{c}_{i-1}) such that the overall level of contextual activation remains constant (for details concerning the form of this expression see Howard & Kahana, 2002a). The above updating process is applied separately to C^{temp} and C^{source} . Each of these subdivisions of context has a distinct drift rate parameter (β^{temp} and β^{source}), such that the two pools of units can update at different rates. The β^{temp} parameter was allowed to vary between the study ($\beta_{\text{enc}}^{\text{temp}}$) and recall periods ($\beta_{\text{rec}}^{\text{temp}}$), reflecting the hypothesis that the rate of context integration may be different depending upon whether a stimulus was externally presented or recalled. To simplify matters, in the current simulations the β^{source} parameter was fixed between study and recall.

The β_{enc}^{temp} and β_{rec}^{temp} parameters are important in determining the nature of temporal clustering: higher values for these parameters (with other parameters held constant) will increase the degree of temporal clustering in the model. The β^{source} parameter is important in determining the degree of source clustering; here, a large value will cause source context to update quickly given a shift in task, increasing the magnitude of the source clustering effect (Table 3).

Novelty-related context disruption.

In the introduction, we suggested that the sudden appearance of a novel representation in one context sub-region (such as a shift from one task to another) triggers a system-wide event that causes other context representations to update, by increasing the rate at which new information is integrated. This disrupts the accessibility of all items studied prior to the novel event, since the temporal context associated with all of those prior items has been pushed out in favor of novel information.

The rate at which context integrates new information is related to the overall novelty or salience of that information. When the model experiences a large shift in source context, as when the participant must switch from one task to another during the study period, all context regions (here, temporal is the only other region) increase their rate of integration, which disrupts or weakens the currently active context representation for the sake of new incoming information. One parameter (d) controls the degree to which a shift in source context disrupts the temporal context representation. This disruption was simulated by presenting a new, orthogonal item in F , and allowing this item to update context by Equation 2, where d serves as the temporary value of β . This “disruption item” is not learned by the network, and does not enter into the recall competition. If d is 0, temporal context and source context are independent; a task shift does not influence temporal context.

It is important to note that disruptive is a relative term here, referring to a small

but detectable behavioral effect. The types of context disruption that one experiences in everyday life are likely to be orders of magnitude more powerful (but much more difficult to manipulate experimentally). Just as the reaction time for a simple judgement is disrupted by a few hundred milliseconds (or sometimes much less) by a shift in task (e.g., Allport, Styles, & Hsieh, 1994), so may we observe small but reliable decrements in memorability and increments in recall latencies given task shifts during a study period. To anticipate the later results, we report significant disruptions to both memorability and latency to recall studied material in a condition where participants shift between two orienting tasks within a list.

In the current simulations we only investigate the effect of a sudden change of source context on the state of temporal context, however, it is reasonable to consider that whenever any context sub-region experiences a large update other representations are disrupted. Since all items are quite similar (they are all visually presented words drawn from a set of concrete nouns) there are never any similarly large shifts in temporal context. It is possible that the first items presented on each list trigger a novelty-related signal (for a similar idea see Laming, 2006). This possibility receives further attention below.

Associative connections: learning and semantic structure.

As described above, two associative matrices represent the feature-to-context connections (M^{FC}) and the context-to-feature connections (M^{CF}). Each of these matrices contains a pre-experimental component (M_{pre}^{FC} and M_{pre}^{CF}), representing the set of associative connections in memory prior to performing the free-recall task, and an experimental component (M_{exp}^{FC} and M_{exp}^{CF}), representing the set of associations learned during the study period of the free-recall task.

M_{pre}^{FC} represents the set of existing associations between the item features and the context elements. This component is initialized as an identity matrix (under the simplifying assumption that any feature of an item has a corresponding element in context

that it can activate). Since we use a localist code for the items in the current simulations, this means that the first time an item is encountered in the context of a given experiment, it activates a single feature element, which activates a corresponding context element. Below we describe how the *pre* and *exp* connections are weighted to create the full matrix M^{FC} .

M_{pre}^{CF} represents the set of existing associations between the context elements and the item features. On the assumption that semantically related items have been associated with one another's temporal contexts, we have implemented semantic associations in this matrix (Rao & Howard, 2008). Each element in M_{pre}^{CF} is determined by taking the $\cos \Theta$ similarity value between two items (with indices a and b) derived using latent semantic analysis (LSA; Landauer & Dumais, 1997), scaling that value by a parameter s , and placing that value in position (a,b) of M_{pre}^{CF} . In this way, the semantic memory of CMR simulates that of a participant whose semantic memory is identical to LSA. This is clearly a drastic simplification of the variability between the semantic memories of individual participants, and tends to inflate the degree of semantic organization produced by the model, since the same association values that are used to assess semantic organization are also used to create the semantic associations in the CMR model. Thus, we were motivated to estimate the degree to which an individual person's semantic memory might mismatch the LSA-derived values, to allow us to create a correction factor to apply to the semantic clustering scores produced by the model.

The correction factor was estimated with a simple simulation of an independent data set: the University of South Florida (USF) Free Association Norms (Nelson, McEvoy, & Schreiber, 2004). We used the CMR model of semantic memory (derived from the LSA database) to predict the distribution of responses produced by participants in the USF study. To simulate variability between participants, we added normally distributed random noise to each simulated participant's semantic memory, and searched for the level

of noise (by manipulating the distribution’s variance) that would minimize the difference between the CMR model’s predictions and the USF data. This procedure is described in more detail in Appendix B.

This estimate of variance between individual participant’s semantic memory structures was used to derive a corrected semantic score for each variant of the CMR model, using the following steps: The recall sequences produced by the CMR model were analyzed to determine the degree of semantic clustering produced by the model; this analysis uses the LSA-derived association values as a normative matrix to determine the degree of clustering. We added normally distributed random noise (with the variance estimated by our simulation of the USF study, described in Appendix B) to the normative LSA matrix, and calculated the degree of semantic clustering. This is equivalent to adding noise to each simulated participant, and simplifies the procedure greatly. The addition of noise to the normative LSA matrix produces a correction factor that reduces the degree of semantic clustering produced by the model. The semantic clustering analysis, and the corrected results, are described below (*Simulation analysis I: Semantic clustering*, and Tab. 2).

As mentioned above, the semantic connections are among the pre-existing associations in the CMR model. The second class of associations are episodic in nature, and are formed as the experiment proceeds. The set of experimental associations M_{exp}^{FC} are initialized to zero, and are updated each time an item is studied using a simple Hebbian outer-product learning rule:

$$\Delta M_{exp}^{FC} = \mathbf{c}_i \mathbf{f}_i^\top, \quad (4)$$

where \mathbf{f}_i^\top is the transpose of \mathbf{f}_i . The relative strengths of the pre-experimental and experimental associations are controlled by a parameter γ_{FC} ; as described by Howard and Kahana (2002a), this parameter influences the magnitude of the forward asymmetry (the

tendency to make forward transitions during recall):

$$M^{FC} = (1 - \gamma_{FC})M_{pre}^{FC} + \gamma_{FC}\Delta M_{exp}^{FC} \quad (5)$$

The associative processes on M^{FC} treat item and source features equivalently. However, the return connections on M^{CF} allow temporal and source context to scale independently to capture the different magnitudes of temporal and source clustering. The set of experimental associations M_{exp}^{CF} are initialized to zero, and are also updated using the Hebbian outer-product learning rule:

$$\Delta M_{exp}^{CF} = \phi_i L^{CF} \mathbf{f}_i \mathbf{c}_i^\top. \quad (6)$$

Both ϕ (described below) and L^{CF} scale the magnitude of particular connections during learning. The matrix L^{CF} allows CMR to separately scale the magnitude of source associations relative to temporal associations (as we will discuss below, the CMR model estimates that associations between source context and item features are about 13% as strong as associations between temporal context and item features). The matrix contains four sub-components:

$$L^{CF} = \begin{bmatrix} L_{tw}^{CF} & L_{ts}^{CF} \\ L_{sw}^{CF} & L_{ss}^{CF} \end{bmatrix}. \quad (7)$$

Where the subscript t refers to temporal context, the subscript s refers to source context if it comes first, and source features if it comes second, and the subscript w refers to item features (this was chosen instead of i , for ‘‘item’’, to avoid confusion with the i subscript indicating list position). For example, the sub-matrix L_{sw}^{CF} scales the associative connections between source context and item features. The magnitude of L_{sw}^{CF} is a manipulable parameter of the model; L_{ts}^{CF} and L_{ss}^{CF} are set to zero (since source features in F do not currently play a role during the recall process), and L_{tw}^{CF} is fixed at 1. In summary, there are three types of organizational information whose strength is varied on

M^{CF} : semantic information (built into M_{pre}^{CF} and scaled by s), temporal information (added into M_{exp}^{CF} as the list proceeds), and source information (added into M_{exp}^{CF} and scaled by L_{sw}^{CF}). Since the recall process is competitive, only two of these strengths need to vary (source and semantic) relative to a fixed amount of temporal information.

The scalar ϕ_i factor is introduced to describe the recall advantage for items in early serial positions (its value is determined by two manipulable parameters of the model). This factor starts at a value above 1, and as the list progresses it decays to 1, at which point it has no effect on the dynamics of the model:

$$\phi_i = \phi_s e^{-\phi_d(i-1)} + 1. \quad (8)$$

Here, ϕ_s is a scaling parameter controlling the magnitude of the primacy effect, and ϕ_d is a decay parameter, which controls the rate at which this advantage decays with serial position (i). This primacy mechanism was added to CMR to explain the behavioral dynamics associated with the primacy effect. While the appearance of the first list item might trigger a novelty-related context disruption similar to that elicited by a change in source, we have opted to use this relatively simple model of the primacy effect. A full treatment of the primacy effect may require a more complete model of encoding dynamics (see discussion).

The recall process.

The context updating and associative learning processes determine the state of the context cue for each recall attempt, which activates each of the item features to a different degree. The degree of activation for a particular item feature determines how well it fares in an ensuing competition, in which all of the items on the most recently studied list compete to be recalled. The recall period is modeled as a series of competitions, each of which takes a certain amount of time and either produces a winning item which is recalled, or the system runs out of time, and the next trial begins. The competition is

mediated by a set of accumulators using the framework described by Usher and McClelland (2001) and applied in the domain of free recall by TCM-A (Sederberg et al., in press). The use of this framework allows us to make predictions regarding both the order in which the items will be recalled, as well as the inter-response times.

The input to the accumulators is determined as follows:

$$\mathbf{f}^{\mathbf{IN}} = M^{CF} \mathbf{c}_i, \quad (9)$$

where $\mathbf{f}^{\mathbf{IN}}$ is then used to guide a leaky, competitive accumulation process:

$$\begin{aligned} \mathbf{x}_s &= (1 - \tau\kappa - \tau\lambda\mathbf{N}) \mathbf{x}_{s-1} + \tau\mathbf{f}^{\mathbf{IN}} + \epsilon, \\ \mathbf{x}_s &\rightarrow \max(\mathbf{x}_s, 0). \end{aligned} \quad (10)$$

This process runs until one of the accumulating elements crosses a threshold (which is set at 1) or the recall period is over. Each element of the vector \mathbf{x}_s (where s indexes the number of steps in the accumulation process) corresponds to an element in $\mathbf{f}^{\mathbf{IN}}$ (in other words, one accumulator for each studied item). As in TCM-A, κ is a constant which determines the rate at which the activation of a given element decays, and λ is a constant which controls lateral inhibition, by scaling the strength of an inhibitory matrix \mathbf{N} which connects each accumulator to all of the others (except itself). ϵ is a noise vector drawn from a random normal distribution with mean zero and standard deviation η , and τ is a time constant determining the rate of the accumulation process. The second line of Eq. 10 means that the accumulating elements can not take on negative values. Items that have already been recalled still take part in the competition, but the threshold serves as an upper limit on their activity value, and they can not be recalled again. When an item wins the recall competition, its item features are reinstated in F . This allows the system to revive the contextual state associated with the item, by allowing it to update context according to Eq. 2. The input $\mathbf{f}^{\mathbf{IN}}$ is updated, and the competition begins again, with \mathbf{x}_s initialized to $\mathbf{0}$.

The decision parameters are critical for obtaining a good fit to the behavioral data, especially the recall latencies. Increasing the value of κ (decay) increases the minimum amount of contextual support an item must have to be able to cross the recall threshold. Increasing the value of λ (lateral inhibition) increases the degree to which the items with the most contextual support inhibit those with less support, enhancing organizational effects in the model. Increasing η (noise) works against organization, by increasing the likelihood that random items are recalled.

Simulation analyses of the CMR model

In the following sections we compare the behavior of the CMR model to human behavior observed in three studies of immediate free recall. The first study was designed and carried out for the purposes of this report, and includes a within-list manipulation of source context (we refer to this as the source-manipulation experiment, described in Appendix A). The second study is a subset of the conditions reported by Murdock (1962), where list length was manipulated. The third study was reported by Murdock and Okada (1970), and includes a detailed analysis of the inter-response times (IRTs) between successive recalls.

We separately fit the model to the behavioral data from each experiment. For each experiment, a single parameter set was found that simultaneously provided good fits to all of the relevant behavioral measures. All model parameters were fixed within-experiment, that is to say, for the experiments that contained multiple conditions (Murdock, 1962 and the source-manipulation experiment) the same set of parameters was used to fit the results from all conditions. This is a high bar for a model of recall; for example, Brown et al. (2007) recently showed that the SIMPLE model fit the data from Murdock (1962), but only if the parameters of the model were allowed to vary for different conditions (Brown et al., 2007).

For each experiment, a genetic algorithm was used to extensively search the parameter space of the model to find the best-fit parameter set (see Appendix C). For the Murdock (1962) and Murdock and Okada (1970) experiments, this search was across 10 parameters. For the source-manipulation experiment, three source-context relevant parameters were included, so the search was across 13 parameters.² Table 1 presents the best-fit parameter set for each experiment. Each parameter set was used to simulate a large number of trials of the particular paradigm, and the resulting recall sequences (and accompanying latencies, where appropriate), were analyzed to create behavioral measures analogous to those carried out on the original behavioral data. Each experiment yielded a different number of behavioral measures, which are detailed in Appendix C. For example, the fitting procedure for the source-manipulation experiment minimized the difference between the simulated data and behavioral data across 93 data-points (including serial position curves, inter-response latencies, and clustering measures). The goodness-of-fit for a particular simulated data set was quantified using a χ^2 statistic.

Insert Table 1 about here

Each of the best-fitting parameter sets reported in Table 1 yielded a χ^2 value. The number of degrees of freedom for each model was equal to the number of data points being fit minus the number of model parameters. We first report the fit to the classic studies of Murdock (1962) [10 parameters; $\chi^2(140) = 760.6$], and Murdock and Okada (1970) [10 parameters; $\chi^2(55) = 3046$].

As mentioned above, we think two mechanisms are important to understand the behavioral effects of a source manipulation: associations between source context and item features, and disruptions to temporal context due to task-shifts. To assess the importance of these mechanisms, we compared the ability of three CMR model variants to fit the data

from the source-manipulation experiment. These were the full model [13 parameters; $\chi^2(80) = 235.5$], and two nested variants of the full model, the Pure association model [12 parameters, d was set at zero; $\chi^2(81) = 327.6$], and the Pure disruption model [12 parameters, L_{sw}^{CF} was set at zero; $\chi^2(81) = 290.8$].

We carried out χ^2 comparisons between the Full model and each of the nested models (since one parameter was fixed in each of the nested models, these tests were carried out with one degree of freedom). This indicated that the Full model gave the best fit to the data. Full versus Pure association: $\chi^2(1) = 92.0, p < 0.0001$; Full versus Pure disruption: $\chi^2(1) = 55.3, p < 0.0001$.

We also calculated the Bayesian Information Criterion (BIC; Schwarz, 1978; Kahana, Zhou, Geller, & Sekuler, 2007), a quantity used to compare goodness-of-fit for models with different numbers of parameters (lower values of BIC indicate better fit, accounting for the number of parameters of the model) using the following equation:

$$BIC = k \ln(n) + n \ln\left(\frac{RSS}{n}\right), \quad (11)$$

where k is the number of model parameters, n is the number of data points being simultaneously fit, and RSS is the residual sum of squares (for more details see Appendix C). This calculation gave BIC values of -602, -574 and -595 for the Full model, the Pure association model, and the Pure disruption model, respectively. These tests agree with the χ^2 measure, indicating that the Full model provides a significantly better fit than each of the nested variants.

The χ^2 statistics reported above indicate significant deviation between the model fit and the behavioral data. A common complaint about χ^2 -based analyses of model fit is that given enough data, any model will be invalidated for failing to fit the fine-grained nuances of the data. In the end, the χ^2 measure is more useful for allowing us to compare model variants to one another (as described above). Regardless of the statistical deviation of the

best-fit models from the behavioral data, the CMR model provides a good qualitative fit of a diverse range of behavioral phenomena across a range of experimental manipulations.

Some component of the deviation of the model predictions from the behavioral data may arise because we force the model to find a single parameter set that accounts for the data across all participants. This may represent an impossible task for the model, if subsets of participants are better represented by models with distinct parameter settings. Fitting the model separately to each participant's data would give us further insight into these issues but requires more behavioral data per participant to obtain stable behavioral estimates. Current work in our laboratories is focused on gathering much more behavioral data per participant, making such an analysis more feasible.

Each of the following four sections examines the recall dynamics of the CMR model in a different way. The first section focuses exclusively on the source-manipulation experiment, examining the three basic forms of clustering (semantic, temporal, and source) observed simultaneously therein. The second section examines the interaction between temporal and source clustering in the source-manipulation experiment. The third section examines the classic serial position effects of immediate free recall reported by Murdock (1962), as well as the perturbation of these effects by a within-list manipulation of source context in the source-manipulation experiment. The fourth section examines the classic inter-response time effects reported by Murdock and Okada (1970), and the perturbation of these effects by the semantic, temporal, and source relations between the studied items in the source-manipulation experiment.

Simulation analysis I: Basic clustering effects

The influence of each organizational factor (semantic, temporal, and source) is demonstrated with a series of clustering analyses on the recall sequences of the source-manipulation experiment described in Appendix A. The basic form of the

clustering analysis is similar for each of the three factors. One can step through the set of recall sequences generated by each participant, and label each word by its semantic identity, list position, and study task. This information is used to calculate measures of semantic, temporal, and source clustering, as described below.

Semantic clustering.

The degree of semantic association between two words is represented by a single number, represented by the cosine distance between the vector representations of those words derived with latent semantic analysis (LSA; Landauer & Dumais, 1997). These semantic association values are used to generate a semantic clustering score for each recall transition (representing how related the two successively recalled words are relative to the other words the participant could have recalled next), and the average of these scores across participants provides us with a summary of the degree of semantic clustering for the experiment. Specifically, for each recall transition we calculate a distribution of semantic association values between the just-recalled word, and the set of words that have not yet been recalled. A percentile score is generated by comparing the association value corresponding to the next item in the recall sequence to the rest of the distribution. Thus, if the participant always chose (from the set of remaining items) the strongest semantic associate, then this semantic clustering measure would yield a value of 1, representing perfect semantic organization. Likewise, a value of 0 would indicate that the participant always chose the least semantically related of the remaining items for their next recall. A value of 0.5 indicates no effect of semantic clustering. As Table 2 shows, the observed value for the behavioral data (0.545) is significantly greater than 0.5, indicating a reliable effect of semantic clustering ($t(44) = 104.2; p < 0.0001$).

Table 2 also presents the estimates of semantic clustering produced by the CMR model, which exceed the behavioral data by about 0.02. As described elsewhere (*Associative connections: learning and semantic structure* and Appendix B), a correction

factor was applied to the model's estimates of semantic clustering to account for the fact that while the LSA-derived semantic association values mismatch each human participant's semantic memory, they are a perfect match for the CMR model's semantic memory, which inflates the model's estimate of the expected degree of semantic clustering. The correction factor reduced the degree of semantic clustering from 0.791 to 0.566 for the Full variant of the model; from 0.792 to 0.569 for the Pure association variant; and from 0.781 to 0.568 for the Pure disruption variant.

Insert Table 2 about here

Temporal clustering.

A similar technique is used to quantify the magnitude of the temporal clustering effect (Tab. 2). For each recall transition we created a distribution of temporal distances between the just-recalled word and the set of words that have not yet been recalled. These distances are simply the absolute value of the difference between the serial position of the just-recalled word and the set of not-yet-recalled words (here, these can range between 1 and 23). A percentile score is generated by comparing the temporal distance value corresponding to the next item in the recall sequence to the rest of the distribution. Specifically, we calculate the proportion of the possible distances that the observed value is less than, since strong temporal clustering will cause observed lags to be smaller than average. As is often the case, when there is a tie, we score this as the percentile falling halfway between the two items. If the participant always chose the closest temporal associate (which is only possible for pure serial recall in the forward or backward direction), then the temporal clustering measure would yield a value of 1 (as there would never be an opportunity for a tie). A value of 0.5 indicates no effect of temporal clustering. The observed value of the behavioral data (0.638) is significantly greater than 0.5,

indicating a reliable effect of temporal clustering ($t(44) = 60.3; p < 0.0001$). It is worth noting that while we describe this as a measure of temporal distance, this is not meant in the sense of “clock time” as in recent models of temporal distinctiveness (e.g., Brown et al., 2007), but rather in the sense of positional lag between items in the study list.

At the suggestion of a reviewer, we examined more closely the pattern of temporal clustering for early and later output positions, for both near and distant lags. Recently, Farrell and Lewandowsky (in press) carried out a similar analysis over a number of free recall data sets. They found a marked tendency for participants to make transitions to distant serial positions early in recall. They suggested that these nonmonotonicities in the probabilities of recall by lag challenge the generality of the contiguity effect, which, in a pure form, suggests that probability of recall by lag should drop smoothly as lag increases. Here, we first examine the behavioral data, before turning to the simulated data generated by the CMR model.

Figure 3 (A and C) shows the probability of making recalls of various temporal distances to the just-recalled item. The analysis is presented separately for output positions 1–3 and for output positions 4 onward. We restricted our analysis to mid-list output positions (serial positions 5–19), as recalls from these positions allow us to separately examine the influence of the contiguity effect and the recency effect on recall transitions. The analysis was carried out separately for each originating serial position and participant, and then aggregated across serial positions and participants, such that each originating serial position is given equal weight in the analysis. Due to the relatively small number of transitions being examined in certain cases (e.g., the more distant bins for output positions 1–3), we aggregated our conditional probabilities over multiple lags, such that the dot in a particular bin represents the probability of transitioning to any lag in the specified range.

Insert Figure 3 about here

The contiguity effect and forward asymmetry for early output positions can be clearly seen in Figure 3A, bins -1 and 1 . The nonmonotonicity can be seen as well, in bins 11.5 and 18.5 . When recall transitions fall into this bin, it means that the participant made an early recall of a mid-list item, followed by a recall of an item from the end of the list. In other words, the nonmonotonicity is due to the persistence of the recency effect, even after a mid-list item is recalled. Figure 3C shows that later in the recall sequence, the influence of the recency effect recedes, but the contiguity effect remains.

Figure 3B and D presents these same analyses, but on the CMR Full model (the two variants show similar results). The model captures the basic behavioral pattern whereby early recall transitions show evidence of both a contiguity effect and a recency effect (Fig. 3B). While the simulated recency effect matches the behavioral data, the model underpredicts the exact magnitude of the contiguity effect. The simulated data match the behavioral data well for the later output positions (Fig. 3D), both in terms of the size of the contiguity effect and the attenuation of the recency effect. We return to this issue in the Discussion (“Relationship to other models of free recall”).

Source clustering.

The degree of organization by source (in this case, encoding task identity), was estimated by tallying the number of recall transitions that were between items studied with the same task, and dividing by the total number of recall transitions, to give a proportion of same-source recall transitions for each participant. This number was averaged across all participants and is reported in Tab. 3. A relabeling technique, in which each control list was randomly assigned the task ordering from one of the task-shift

lists, was used to create a baseline for the source clustering measure. By relabeling the control list items with the same sequence of task alternations used in the experimental lists (and aggregating over many such assignments), we calculated the proportion of same-task transitions one would expect to observe solely on the basis of the temporal contiguity of items studied with the same task.³ The overall source clustering effect was both numerically large (a increase from 0.536 to 0.606 in within-task transitions between the relabeled control and task-shift lists; see Tab. 3, *All transitions*) and statistically reliable (paired sample t-test across participant means; $t(44) = 6.83, p < 0.001$).

Table 3 (*All transitions* column) also presents the source clustering results for each of the three model variants. The Full model and Pure association variants provide a very good fit to the behavioral data, but the Pure disruption variant underpredicts the degree of source clustering. This is because the disruption mechanism simply weakens item-related contextual support every time there is a task shift, which serves to isolate items in different trains from one another. Clustering arises for same-train items because given the recall of one item in a train, the retrieved context supports recall of other items in that train, but not items in other trains.

Insert Table 3 about here

Thus, the recall sequences exhibit an influence of (at least) these three organizational factors: semantic, temporal, and source. Characterizing the ways in which these factors interact is not a trivial matter. As we shall elaborate, the factors do not seem to combine in a linear manner. Rather, these combined factors set in motion a nonlinear recall competition that leads to the actual response. Clouding the picture further, each recall updates the state of internal context, which alters the course of all successive recalls. The highly interactive and dynamic nature of the recall process makes

the CMR model a valuable tool for interpreting the rich behavioral interactions between these organizational factors.

Simulation analysis II: An interaction between source and temporal information

Source clustering is observed at all transition distances, but the magnitude of source clustering is enhanced for items studied nearby in time. Table 3 (*Remote transitions*) describes the proportion of recall transitions to same-task items, conditional on the fact that this is a remote transition (outside of the local train of same-task items). A significantly greater proportion of these remote transitions are to same-task items in the task-shift lists compared to the relabeled control lists (0.391 and 0.334 respectively; paired-sample t -test on participant means: $t(44) = 5.11; p < 0.001$). Thus, longer-range transitions also exhibit clustering by task identity, although the magnitude of the effect is diminished relative to nearby transitions.

Table 3 (*Remote transitions*) also presents the remote source clustering results for each of the three model variants. The Full model and the Pure association variants provide a reasonable fit to the data. However, the Pure disruption variant underpredicts the magnitude of the remote source clustering.

Insert Figure 4 about here

The interaction of the source and temporal factors is apparent in Figure 4A, which plots the results of a conditional response probability analysis of the recall transitions. Here, instead of grouping items by serial lag to the just-recalled item, they are grouped by train position relative to the just-recalled item (creating a *train-lag CRP* analysis). A *train lag* of zero corresponds to the recall of an item from the same train as the just-recalled item, whereas a train lag of one corresponds to the recall of an item from the

next train in the study sequence. The source-clustering effect is revealed most clearly in the difference plots (lower panels B, D, F, and H) between the control and shift conditions. In the difference plots, *same-task* (black squares) and *between-task* (white squares) transitions are marked differently, revealing that the tendency to cluster by task is a decaying function of temporal distance from the just-recalled item. A strong source clustering effect is exhibited by an increase in the black squares above zero, and a decrease of the white squares below zero, meaning that same-task items are more likely to be clustered with one another than between-task items.

Inspection of the behavior of the Pure disruption and Pure association model variants (Fig. 4E–H) suggests that two factors play a large role in shaping the interaction between temporal clustering and source clustering: the task-shift disruption mechanism and the nonlinear recall competition. The disruption of temporal context given a shift in source context (novelty-related disruption) causes items in the same train to be tightly coupled to one another; when a particular item is recalled, the reinstated temporal context representation overlaps well with same-train items but not with items in neighboring trains. Since same-train items are also same-task items, this mechanism inflates the degree of source clustering for nearby items. This can be seen in the plot for the Pure disruption model variant (Fig. 4E), which even without associations between source context and item features shows source clustering for nearby (same-train) items. The task-shift disruption mechanism, in general, sharpens the train-lag CRP plot for the task-shift condition relative to relabeled control condition. This sharpening causes a decreased likelihood of transitions to neighboring trains (-1 and +1 train lag) relative to the likelihood of making these transitions in the control condition (where item context is not disrupted between trains).

The associations between source context and item features produce the source clustering observed in the Pure association variant of the model. Here, source context provides equivalent support to all items associated with a particular task; however, there

is still an interaction between the degree of source clustering and the temporal proximity of the items (the peak at train-lag 0 in the Pure association panel of Fig. 4H). This interaction arises as a consequence of the nonlinear recall competition; since the items compete with one another to cross the recall threshold, an increase in support can have a supra-linear effect on the likelihood of an item winning the competition. Even though task information provides equivalent support to all same-task items, this task support combines with the temporal support for nearby items to result in increased source clustering for temporally proximal items.

Simulation analysis III: Serial position effects

The experimental study of the free-recall paradigm has been long dominated by the analysis of probability of recall by serial position on the study list. Murdock (1962), in a classic paper, described behavior in an immediate free recall paradigm across several conditions, where both list length and presentation rate were manipulated. The CMR model has not yet been extended to deal with presentation rate, so we focus here on the effect of changes in list length on the probability of the recall of items by serial position (with a one-second presentation rate). In his investigation, Murdock (1962) described the effect of changing list length on three features of the serial position curve. He showed that increasing list length did not change the shape of the primacy portion of the curve, though the overall probability of recall for the primacy positions did drop with increases in list length. He showed that increasing list length caused the mid-list asymptote to drop. Finally, he showed that increasing list length had no effect on the slope of the recency portion of the curve.

The CMR model fits the effects of list length on the shape of the serial position curve with a single set of parameters. Figure 5A and B respectively present the original behavioral data reported by Murdock (1962) and the simulated data from the best-fit

parameter set of the CMR model (M62 in Tab. 1). This fit was arrived at using a parameter search that minimized the difference between the simulated data and the behavioral data for the three serial position curves as well as the lag-CRP curves for the three conditions (for ± 10 lag positions; results not shown).

The major behavioral effect, in which the probability of recall for early and mid-list items decreases as a function of list length is observed for wide regions of the parameter space of the model, and can be understood in terms of the dynamics of the recall competition. The λ term in Equation 10 causes the support for any one item to decrease as the number of items competing for recall increases. However, the recent items are somewhat insulated from this lateral inhibition as they are strongly supported by the end-of-list context cue. The nonlinear nature of the recall competition allows these well-supported items to cross threshold quickly, whereas items retrieved later in the recall sequence are relatively less well-supported (owing to a β_{rec}^{temp} parameter that is lower than β_{enc}^{temp}), and feel the effects of competition more sharply.

Insert Figure 5 about here

The lack of an effect of a manipulation of list length on the slope of the recency effect has traditionally been taken as evidence for a short-term buffer where any items whose representations were active in the buffer were shielded from proactive interference from the preceding items in the study list. Recently, Sederberg et al. (in press) showed that TCM-A could account for the insensitivity of the probability of recall for recently studied items to the overall list length, although they did not present a quantitative fit of the Murdock (1962) data. TCM-A would likely provide as good a fit to the Murdock (1962) data as the CMR model; however, TCM-A is unable to fit effects related to semantic and source organization. We argue that while TCM-A provides a good model of

temporal organization, understanding the dynamics of free recall requires consideration of the role of semantic and source information as well.

The CMR model provides a good fit to the probability of recall by output position (Figure 6) as well as the basic serial position curve (Figure 7) of the source-manipulation experiment. Figure 6 depicts the probability of recalling an item studied in a particular serial position, as a function of output position. The temporal component is obvious in the recency effect observed in the first output position; however, by the third recalled item, all serial positions are nearly equiprobable for recall.

Insert Figure 6 about here

Were temporal and source information the only two factors driving recall order, the CMR model would predict that recall would generally proceed backwards from the end of the list. While a strong source cue would produce some temporal spread in the initial recall transitions, source context provides equivalent support for all items associated with a particular task; once source and temporal cues combine, nearby same-task items are still more supported than remote same-task items. The semantic associations (s) and the decision noise (η) both have a stochastic effect on the recall sequence. Both of these parameters contribute to the flattening of the serial position curve across the first three output positions.

The CMR model also provides a good fit to the serial position curve of the source-manipulation experiment (across all output positions; Figure 7). The overall percentage of items recalled is much higher than in the comparable list length of the Murdock (1962) study. This is due to a number of differences between the experiments, including an increased presentation time in the current experiment (3 sec.) and the use of an orienting task. The best fitting parameter set for the CMR model is presented in Table

1; small adjustments in a number of the model parameters allow the model to explain the differences in the serial position curves between the experiments.

Insert Figure 7 about here

Figure 7 reveals that when participants are asked to shift between encoding tasks within-list, they are less likely to remember items from early and mid-list serial positions. It also shows that the items in recent positions are just as well recalled. In the task-shift condition, participants switch between variably lengthed trains of items associated with each task. Each task-shift list contains 6 or 7 trains of items. By examining probability of recall as a function of train position, we gain some insight into the influence of source information in this domain. Figure 8 presents the proportion of items recalled from each train in the study list. In order to present the 6- and 7-train lists on the same plot, we divided the 6-train lists into two parts (the first 3 and last 3 trains), and “end justified” the last three trains to correspond to trains 5, 6, and 7 on the 7-train lists. Separate analysis of the 6- and 7-train lists revealed similar results, providing justification for this aggregation method.

Insert Figure 8 about here

A relabeling procedure was used on the data from the control condition to construct the *Relabeled Control* line in Figure 8, allowing us to more directly compare performance between the task-shift and control conditions. By this procedure, each control list was assigned the task ordering from a randomly selected task-shift list. This task ordering was used to assign each serial position in the control list a train position (numbering from 1–6

if the randomly selected task-shift trial was a 6-train list, and 1–7 if it was a 7-train list). The recall sequences from these relabeled control trials were analyzed to count the proportion of items recalled from each relabeled train, giving rise to the *Relabeled control* line in Figure 8. Many random assignments of task-shift lists to control lists were aggregated to create a stable baseline measure, against which to observe the influence of source information on the probability of recall of studied items.

Figure 8B presents the difference between these curves for the relabeled control and task-shift conditions. Overall, participants recalled a lower proportion of items in the task-shift condition relative to the relabeled control condition (as measured with a paired-sample t -test; $t(44) = -5.19$; $p < 0.001$). It may seem a bit counterintuitive that memory was worse for lists with multiple retrieval cues (appealing to the generally beneficial effect of encoding variability), however the literature is unclear on what one ought to expect in this situation. For example, Tulving and Colotla (1970) found a decrease in mean recall for lists that were composed of items drawn from multiple languages, relative to unilingual lists (the participants were trilingual), while Murdock and Walker (1969) found a small increase in mean recall for mixed-modality lists (auditory and verbal presentation) relative to lists that were purely of one modality or the other. It is interesting to note that one can observe significant organization by source even when this source variation is harmful to overall recall of the list; we suggest that this is because temporal information is a powerful cue, and the disruption to temporal context due to repeated task switches harms recall to a greater extent than the variability in source helps recall.

Closer examination of recall by train position revealed that while participants indeed have worse memory for items in early and mid-list trains ($t(44) = -5.83$; $p < 0.001$), memory is actually improved for the final train of items in the task-shift condition ($t(44) = 2.30$; $p < 0.05$). Figure 8C and D show that the CMR model

is able to capture these effects. According to the model these two effects mostly arise because the act of switching tasks is disruptive to temporal context. Thus, early and mid-list items are difficult to retrieve, because temporal context has changed more over the course of the task-shift list than over the control list. This same mechanism explains the improved recall for the final set of items. Since every train of items except the last one is less well supported in the retrieval competition, these final items benefit from the reduction in proactive interference and are more likely to be recalled than the equivalent items in the relabeled control lists. It is worth noting that this increase in proportion recalled is obscured in the standard serial position curve (Fig. 7), since the length of the final train varies randomly between 2 and 6.

Insert Figure 9 about here

As mentioned above, the Full CMR model provides a significantly better fit to the global pattern of behavioral data than the two variants, but by examining the best-fit versions of these variants (one without associations to source context [Pure disruption] and one without a disruptive task shift [Pure association]), we can gain insight into the roles of these mechanisms on recall dynamics. Figure 9A shows that the Pure disruption model variant easily fits the decreased memorability of items in the task-shift condition (although it can not account for the full pattern of source clustering). This is due to the small disruption to temporal context that occurs with each shift between tasks. However, the Pure association model is unable to provide a good fit to the train serial position curves (Figure 9). In particular, it has trouble fitting the decreased memorability of items in the task-shift condition. Specifically, Figure 9D shows that the model can not fit the impaired recall of items from trains 5 and 6 in the task-shift condition. This is because source and temporal context naturally work together in the model, as follows: Given the strong

associations between source context and item representations, the model often jumps back to the fifth train (which was always studied with the same task as the final train). Once the context associated with the fifth train is reinstated, there is a strong bias to step forward to the sixth train, due to both the general forward bias in recall transitions, as well as the still lingering end-of-list context present in the context representation. Thus, the Pure association variant overpredicts the the percent recall for these later trains. This is true not only of the best-fit parameter set, but of all of the Pure association parameter sets that we inspected that provided reasonable fits to the other aspects of the data. Taken together, these simulations suggest that disruption of organizational processes due to task switching is behind the reduced memorability of items in the task-shift condition.

Both variants explain the increased memorability of the final train in a similar way: The Pure disruption variant fits this because the most recent task shift disrupts the temporal context associated with earlier items (leading to a reduction in proactive interference from earlier items during the recall competition). The Pure association fits this because the source representation associated with the final train of items is still active, giving these items a boost in the recall competition. Thus, the model reveals the influence of multiple organizational factors on the memorability of studied material. In the next section we examine how these organizational factors affect the speed with which these items are retrieved.

Simulation analysis IV: Inter-response latencies

The ability of the CMR model to capture basic serial position effects places it alongside several models of the recall process, such as SAM, ACT-R, TCM-A, and SIMPLE (Raaijmakers & Shiffrin, 1980; Anderson, Bothell, Lebiere, & Matessa, 1998; Sederberg et al., in press; Brown et al., 2007). Many fewer models have been developed to account for the fine-grained temporal dynamics of retrieval, including both output order

effects and the inter-response times between successive recalls. Raaijmakers and Shiffrin (1980) applied SAM to the problem of serial position effects and inter-response times, but not simultaneously, as the temporal dynamics of retrieval from short-term store had not been worked out.

Insert Figure 10 about here

A classic study by Murdock and Okada (1970) examined inter-response times between successive recalls in the free-recall paradigm, and showed that they increase exponentially with each response produced, and that the rate of decay of the exponential curve varies as a function of the total number of items recalled on the trial in question. The exponential character of the growth in inter-response times has been replicated and extended in an elegant series of papers by Wixted and colleagues (Wixted & Rohrer, 1993; Rohrer & Wixted, 1994; Wixted & Rohrer, 1994).

Figure 10A shows an analysis of the original Murdock and Okada (1970) data, plotting inter-response times for trials on which 4, 5, 6, and 7 total items were recalled. Figure 10B shows the same analysis on the best-fitting version of the CMR model. The model was fit to the serial position curve, lag-CRP curve (± 10 lags), and the set of inter-response curves presented in Fig. 10A. This pattern of inter-response time increases seems to be a robust property of the model, and was exhibited for a wide region of the parameter space explored.

Two parameters are critical for observing this pattern in the inter-response times. Perhaps the most important is the lateral inhibition parameter in the recall competition (λ). Increasing λ increases the competitiveness of the recall competition, such that well-supported items race past threshold quickly, and items with less support take a much longer time to reach threshold. Exploration of parameter space revealed the importance of

the semantic association parameter (s), for the model to account for inter-response timing. A search of parameter space with s fixed at zero (making the CMR model nearly isomorphic with TCM-A) was unable to discover any parameter sets that allowed the model to simultaneously fit the pattern of inter-response times, serial position curve and lag-CRP curve. This is because the semantic associations play an important role in the recall dynamics in the CMR model. Since the words are randomly selected from a large word pool (and are randomly placed in the study list), some items tend to be well connected to one another, and others tend to have fewer strong semantic associations with the other list members. The well-connected items will tend to be retrieved quickly and early in the recall sequence, and later recalls will tend to be to less well supported items. Temporal context, of course, also plays an important role here. The end-of-list temporal context is a good cue for the recent items, thus these items are recalled quickly and early in the recall sequence. The model suggests that the temporal context retrieved during the recall period (β_{rec}^{temp}) is weaker than that retrieved during the study period (β_{enc}^{temp}), which means that later recalls are not as well supported in the decision competition.

The rise of IRTs with output position has been traditionally explained as arising from a sampling-with-replacement rule (Raaijmakers & Shiffrin, 1980; Wixted & Rohrer, 1994): As recall progresses, more of the items sampled for recall are ones that have already been recalled. Given that each sampling attempt takes a fixed amount of time, the IRTs grow in a roughly exponential manner. The CMR model suggests that the increase in IRTs with output position is not due to ever-increasing intrusions of already-recalled items, but rather due to the relative lack of support for these later items in the recall competition, paired with interference from the set of already-recalled items. However, the CMR mechanism is similar in spirit to the sampling-with-replacement rule, in that at least one source of interference originates with the already-recalled items.

The CMR model can account for the major source of variance in IRTs, the

exponential rise with output position and total number of recalled items. This allows us to examine the model's predictions regarding the influence of semantic, temporal, and source information on IRTs. A number of groups have reported decreased IRTs for items that are similar on various dimensions. For example, a number of studies of recall of categorized stimuli observed decreased inter-response times when the participant shifted to a new category (Patterson, Meltzer, & Mandler, 1971; Pollio, Richards, & Lucas, 1969; Wingfield, Lindfield, & Kahana, 1998). Similarly, successively recalled items that are semantically similar have smaller IRTs than items that are dissimilar (Howard & Kahana, 2002b). Finally, items that were studied in a similar temporal context (nearby list positions) have smaller IRTs than items studied in distant temporal contexts (distant list positions; Kahana, 1996). The CMR model suggests that these effects arise due to the context reinstatement that occurs as each item is recalled. Furthermore, the model predicts an effect of source context on IRTs, whereby items associated with similar sources are recalled with a shorter IRT than items associated with distinct sources. The source-manipulation experiment was designed to allow us to simultaneously examine effects of semantic, temporal, and source information on retrieval. While the effects of semantic and temporal information on inter-response times has been documented (Howard & Kahana, 2002b; Kahana, 1996), we are unaware of any report of a source-related IRT effect. As we will detail below, all three effects were found in the behavioral data.

For our three inter-response time analyses, we calculated the mean inter-response time for each output position within each participant, allowing us to remove these sources of variance from the inter-response time measure. For each organizational factor, we calculated a threshold to classify each recall transition as being between similar or dissimilar items. For the semantic factor, an LSA-derived similarity score of 0.2 was used as the threshold; items with a similarity greater than or equal to 0.2 were classified as semantically similar, and below 0.2, dissimilar. For the temporal factor, items with a lag

difference of 3 or less were temporally similar, greater than 3, temporally dissimilar. For the source factor, items studied with the same task were similar, and items studied with distinct tasks were dissimilar.

The behavioral data from the control condition were examined for semantic and temporal effects. On average, the IRT between semantically similar items was 1,771 msec faster than between dissimilar items, an effect that was statistically significant ($S.E.M. = 443, t(44) = 4.04, p < 0.0001$). The Full CMR model showed a similar effect ($Mean = 2387, S.E.M. = 88, t(539) = 26.96, p < 0.0001$). The mean inter-response time between temporally similar items was 700 msec faster than temporally distant items, which was also significant ($S.E.M. = 334, t(44) = 2.12, p = 0.02$). Again, the Full CMR model showed a similar effect ($Mean = 367, S.E.M. = 102, t(539) = 3.60, p < 0.0001$).

Insert Table 4 about here

The behavioral data from the task-shift condition were examined for the source effect. As reported in Table 4, the inter-response time between same-task items was 1,302 msec faster than between-task items, which was significant ($t(44) = 4.78; p < 0.002$). All three models variants showed this effect: Full ($t(539) = 5.26; p < 0.002$), Pure association ($t(539) = 3.37; p < 0.002$), and Pure disruption ($t(539) = 6.97; p < 0.002$). These results suggest that both source-to-item associations and task-shift disruption are sufficient to explain the source-related cost on inter-response times. The association mechanism does this in a positive way, causing same-task items to get a boost from the source context reinstated by the just-recalled item. The disruption mechanism does this in a negative way, decreasing support for all list items in a different train than the just-recalled item. This causes the same-train items (which by definition are same-task) to be favored relative to the other items in the competition. While these model fits show some quantitative

deviations from the behavioral data, it is worth noting that these measurements were not included in the fitting procedure, so it is possible that the model could produce a better fit to these data with some parameter adjustment.

Insert Table 5 about here

The CMR model is a model of memory search; as such, it has little to say about the processes driving performance of the orienting task judgments. However, the CMR model attempts to explain the role of task representations in memory, so it is worth documenting that in the behavioral data we observe a shift cost in the reaction time to make the orienting task judgments in the task-shift condition (Table 5), as has been observed many times in the task performance literature (e.g., Allport et al., 1994). We explore the connection between these two domains in the discussion.

General Discussion

The Context Maintenance and Retrieval (CMR) model of memory search proposes that features of the study episode activate an internally maintained context representation which is used to search through one's recent memory. The model is designed to explain three forms of organization: semantic clustering, temporal clustering, and source clustering. We introduced the *spotlights of memory* analogy to describe the process by which the maintained context representation sweeps across the associative structures of the memory system, searching for the representations of recently studied items. Furthermore, we described two ways in which these context representations can interact: First, by combining to drive a non-linear recall competition they can have super-additive effects on the likelihood of recalling particular items. Second, a large influx of novel information to one context representation can disrupt other context representations.

We examined three studies of free recall: a new experiment where we manipulated source context within list (the source-manipulation experiment), a classic study of serial position effects (Murdock, 1962), and a classic study of inter-response times (Murdock & Okada, 1970). The source-manipulation experiment yielded reliable effects of three forms of clustering: semantic (assessed using latent semantic analysis), temporal (serial lag), and source (encoding task). This observation of clustering by encoding task is, to our knowledge, the first such observation reported in the literature, suggesting that some representation related to the operations carried out during study is associated to the representation of the to-be-remembered item. This finding allows us to add task features to the set of attributes that can be used by the memory system to target particular memories (Underwood, 1969). Furthermore, closer inspection of the recall sequences revealed that, whereas task clustering was observed for both nearby and remote transitions, the effect is greatly enhanced for words studied nearby in time.

The CMR model shows the classic insensitivity of the recency effect to the length of

the studied list, while the overall proportion of recalled items drops as list length increases (Murdock, 1962). When source context was varied within-list, there was a decrease in memorability for all studied items, except for the most recent, which showed enhanced memorability. Exploration of variants of the model suggest that a disruption of temporal context with each task shift was necessary to explain this decreased memorability, but that all variants of the model predicted the enhanced memorability of the final items.

The CMR model shows the classic exponential rise in inter-response times (IRTs) with output position, modulated by the total number of items to be recalled on that trial (Murdock & Okada, 1970). Finally, the model accounts for the cost (to IRTs) for shifting between items that were dissimilar on any of the three organizational dimensions (semantic similarity, temporal distance, and task identity). The shift cost on IRTs for between-task recall transitions is another observation that we believe is novel in the literature.

Clustering: Association vs. disruption

The CMR model provides a framework in which to explore the mechanistic basis of recall clustering. A (perhaps) counterintuitive finding of the current simulations was that associations between context representations and item features was not enough for the model to explain the full pattern of data observed in the source-manipulation experiment. We found that the addition of a mechanism whereby shifts in source context (such as the one elicited by a task switch) during the study period also disrupt temporal context was also necessary to explain two facets of the behavioral data. First, as detailed in an analysis of the interaction between temporal and source information during recall (Simulation analysis II, Fig. 3), the disruption mechanism allows the model to show the behavioral pattern whereby source clustering is greater for items studied nearby in time, and tapers off with temporal distance. Second, as detailed in an analysis of serial position effects (Simulation analysis III, Fig. 7), the disruption mechanism allows the model to

account for the decreased memorability of early and mid-list items in the task-shift condition. In effect, shifts in task context cause items studied nearby in time to become more distant from one another, relative to two items studied with the same task.

We also observed that the model, without associations between source context and item features, showed some degree of source clustering, albeit only for items studied nearby in time. How can a purely disruptive mechanism support increased clustering? Above (in Simulation analysis II), we showed that since recall is a competitive process, the task-shift disruptions to temporal context create isolated islands of same-train items whose associations to items studied in other trains has been weakened relative to a control condition. While the overall probability of recalling any one item is lower, conditional on the recall of a particular item, the likelihood of recalling another item from the same island is enhanced.

An interesting question arises as to whether the disruption of context due to exposure to novel information is specific to shifts between tasks, or whether it is a general principle of the memory system. A more definitive answer may await further research; however, a number of findings in the literature provide converging evidence for such a hypothesis. The study by Sahakyan and Kelley (2002), described in the introduction, introduces a disruption to inner mental context perhaps related to the context disruption due to task switching. Parallel findings can also be observed with manipulations of external context.

For example, Strand (1970) carried out a classic study of environmental context change using a retroactive interference paradigm (in which participants studied two lists in sequence and were then tested on their memory for the first). Participants were run in one of three conditions: A neutral condition, where participants studied both lists in the same room; a context-change condition, where participants walked to another room between lists; and a context-disruption condition, where participants walked into the hall

between lists but returned to the same room. Interestingly, Strand found that the context-change and the context-disruption conditions elicited an equivalent degree of reduction of retroactive interference, suggesting that the primary factor at work was the disruption due to traversing the halls, and not the removal of the contextual associations of the surrounding environment.⁴ Here, the novel interpolated activity involves simply walking into the hall; presumably this drives an updating of temporal context, which causes the two lists to be encoded more distinctly in memory.

The CMR model provides a framework in which to examine the interactions between context disruption and context associations. These two factors may be important in understanding a number of classic findings in the free-recall literature which explore the disruptive effects of interpolated mental activity on the memorability of studied items. Of particular relevance are a set of experiments that compared performance in immediate free-recall (IFR), delayed free-recall (DFR), and continual distraction free-recall (CDFR) paradigms (Glanzer & Cunitz, 1966; Postman & Phillips, 1965; Bjork & Whitten, 1974). In these paradigms, the participant engages in a short period of a distraction task (e.g., mental arithmetic) either just prior to beginning the recall period (DFR), or before and after every item (CDFR).

Recently, Sederberg et al. (in press) showed that TCM-A can account for a number of dissociations between IFR, DFR, and CDFR, by assuming that performing the distraction task disrupts the temporal context representation (using a similar mechanism to our task-shift disruption). While this mechanism proved sufficient to explain performance across these paradigms, the addition of a task context representation may be required to explain performance in a set of paradigms that manipulated the identity of the distraction task within-list.

Specifically, Koppenaal and Glanzer (1990) introduced a variant of the CDFR paradigm in which participants perform one distraction task in the intervals between each

list item, but a second task in the interval just prior to the start of the recall period. This shift in distraction task just before recall causes an attenuation of long-term recency usually observed in a standard CDFR paradigm. The novelty-related disruption mechanism of CMR would be triggered by a shift in distraction task, which would disrupt temporal context to a greater extent than in the standard CDFR paradigm, where distraction task is consistent throughout. Even without the novelty-related disruption, the CMR model would predict worse performance with a shift in distraction task: Since distraction task identity is integrated into context and associated with the studied items, shifting to another distraction task just before recall removes a cue that could be used to support those items in the recall competition.

Finally, Thapar and Greene (1993) demonstrated that when one performs a different distraction task after each list item (including between the study period and the recall period), the recency effect reemerges (see also Neath, 1993). The CMR model should handle this finding as well; since there is an equivalent amount of disruption to task and temporal context after every list item, the most recent list items will again be favored relative to the more distant items. By allowing task to be represented within source context, the CMR theory provides a straightforward explanation of the effect of the manipulation of distractor tasks on the memorability of studied material. Furthermore, this approach allows us to treat manipulations of encoding and distractor tasks in a common framework, and predicts that the same organizational effects that were observed by manipulating encoding task should also obtain with within-list manipulations of distractor task.

Task context and human memory

The role of the processing task on later memorability of studied material has a long history in the literature. A major thread, levels of processing, was concerned with the

finding that deeply processed items (i.e., where semantic features were emphasized) are remembered better than items receiving shallow processing (where phonological or orthographic features were emphasized; Craik & Lockhart, 1972; Hyde & Jenkins, 1969; Postman & Adams, 1956). One reaction to the levels of processing approach, transfer-appropriate processing, pointed out that one performs best on a memory test when the processes engaged at study emphasize the attributes of the studied material that are most relevant for the upcoming memory test (Morris, Bransford, & Franks, 1977; Blaxton, 1989). This is also closely related to the concept of encoding specificity (Tulving & Thompson, 1973), which states that retrieval of a piece of information will be facilitated if the retrieval cue used to recall that information matches the specific attributes that were emphasized during encoding. At a mechanistic level, the CMR model is consistent with the principles of transfer-appropriate processing and encoding specificity, in that one will be better able to remember a particular piece of information the more one's context representation (which is used as a retrieval cue) matches the context representation that was present when that item was originally studied. Thus, activating the source context associated with a subset of the studied material makes that material more accessible, relative to material studied in another source context.

While we focused specifically on task context in this article, the CMR model was developed to account for manipulations to many varieties of source characteristics. As such, some broad parallels might be seen between this effort, and the source monitoring framework of Johnson and colleagues (Hashtroudi, Johnson, & Chrosniak, 1989; Johnson, Hashtroudi, & Lindsay, 1993). In the source monitoring framework, source is meant to refer broadly to the set of characteristics that specify the conditions under which a given memory was acquired, much like the notion of context being explored here. In other words, source in the source monitoring framework actually subsumes all of what we refer to as context here. Perhaps most directly related to this endeavor is the work by Jacoby and

colleagues characterizing the role of source-constrained retrieval in recognition memory paradigms (Jacoby, Shimizu, Daniels, & Rhodes, 2005). Below (“Future directions”), we explore the possibility of applying the CMR model fruitfully in this domain.

Another domain in which the notion of task context receives much attention is in the behavioral study of task performance. Researchers have shown that there is a cost to performance associated with switching from one task to another task (e.g., Allport et al., 1994; Wylie & Allport, 2000). This phenomenon most reliably exhibits itself as an increased latency to respond to the stimulus following the switch, also known as a switch cost. As presented in Table 5, a reliable switch cost was observed in the latencies to make judgments to post-switch items in the behavioral data from the current paradigm (this decision process was not part of the CMR model). The switch cost is taken as evidence that task representations are being updated when one switches between tasks. These task representations are thought to be a component of an executive control system that guides processing of incoming stimuli, in accordance with the particular demands of a given task (Cohen et al., 1990).

Theories of task performance suggest that associative interference arising from previously active task representations is an important factor underlying reaction time shift costs (Monsell, 2003). These interference effects are thought to arise from rapidly formed associations between the features of the stimuli and the task representation guiding the processing of those stimuli. Thus, when a participant switches to performing a new task, associations between the stimuli and the now-inappropriate other task representation slow processing.

The CMR model predicts that two dependent measures in the current experiment reflect the influence of task representations: the task clustering effect (in particular the remote task clustering effect) and the cost to IRTs after a task shift. According to the CMR model, both of these behavioral effects arise from a combination of

context-association and context-disruption mechanisms. The context-association mechanism is similar to the above described associative interference mechanism from the task switching literature, in that it describes the process by which features of studied items are associated with a concurrently active task representation.

At first glance, these mechanisms seem quite different: in the free-recall paradigm, the associations formed between items and task context exhibit themselves minutes later during the recall period, but in the task-switching paradigm, these effects are observed only seconds later, and seem to dissipate once a few stimuli have been judged with the new task. However, some researchers have hypothesized that in the task switching paradigm, these interfering associations are still present, but control processes detect the conflict between the two competing tasks and support the current task, allowing it to overcome any interference arising from the now-inappropriate task. Thus, the associations between stimuli and tasks only exhibit themselves behaviorally in the first trials following a shift, before control processes have had a chance to activate (Botvinick, Cohen, & Carter, 2004). If true, this predicts that the magnitude of a participant's switch cost will be positively correlated with the degree of task clustering observed for that participant during a later recall period (as well as that participant's IRT shift cost). A future study blending the techniques of task-switching paradigms with free-recall paradigms will be able to test this prediction.

Developing the CMR model of encoding dynamics

In the present investigation, we have chosen to focus on retrieval-period dynamics in order to simplify the model under consideration. However, a number of researchers have established the importance of study-period rehearsal dynamics in understanding performance in the overt rehearsal free-recall paradigm (Rundus & Atkinson, 1970; Rundus, 1971; Brodie & Murdock, 1977; Tan & Ward, 2000; Laming, 2006). It is likely

that covert rehearsal processes also play a role in the immediate free recall paradigm.

Laming (2008) suggests that the mechanisms underlying rehearsal are the same as those underlying recall. If so, it would be a straightforward extension of the model to include a short burst of retrieval in the interval between the presentation of each studied item. Such an addition to the model would allow us to simulate the pattern of rehearsals during free recall. The work of Murdock and Metcalfe (1978) and Tan and Ward (2000) suggests that this might provide a rehearsal-based explanation for the primacy effect, which would allow us to remove the two model parameters controlling the primacy gradient, although several parameters would likely need to be added in order to characterize the rehearsal process. While it is not clear whether this addition would shed additional light on the dynamics of the organizational processes under investigation here, the CMR model predicts that these organizational factors (semantic, temporal, and source) will influence the order of rehearsals in the overt-rehearsal free-recall paradigm. The influence of temporal context during rehearsal is clear from a number of studies (Friendly, 1979; Ward et al., 2003; Bhatarah, Ward, & Tan, 2006; Laming, 2008), but the influence of semantic and source information during rehearsal has not been characterized.

One aspect of the model fits that could potentially be related to these covert rehearsal processes is the inability of the CMR model to match the exact shape of the probability of first recall (PFR) curve (the probability of recalling a particular serial position in the first output position) for the Murdock (1962) dataset (depicted in Howard and Kahana (1999), Figure 1). The CMR model predicts that the participant is most likely to start recall with the final item, and that the probability of earlier items initiating recall falls off rapidly and monotonically (possibly with a small bump for the primacy positions). However, in the Murdock (1962) study, participants more often began recall two or three positions from the end of the list (producing a nonmonotonic, or “humped” PFR curve).

The nonmonotonicity in the PFR curve has been taken as support for the notion of a short-term buffer, where items in the buffer are reported in the order in which they were inserted into the buffer (in other words “oldest first”). This has been cast as a challenge for context-based theories of recency (Farrell, 2008). However, it is possible that if participants covertly start to recall list items prior to the onset of the recall start signal, the first “overt” recall could come from a slightly earlier serial position (Brodie & Murdock, 1977; Rundus, 1971; Tan & Ward, 2000). Finally, many free-recall studies do not exhibit this nonmonotonicity in the PFR curve (producing instead a curve that peaks with the final item; Howard & Kahana, 1999; Kahana, Howard, Zaromb, & Wingfield, 2002; Howard, Youker, & Venkatadass, 2008; as well as the present study). It is unclear which experimental variables are responsible for this nonmonotonicity. It is important that we understand what causes this effect to come and go before we modify the model to account for it.

Relationship to other models of free recall

Models of memory search have often emphasized that memory search involves a strategic, systematic, and serial inspection of a number of locations in the memory system (James, 1890; Shiffrin, 1970; Kintsch, 1970; Raaijmakers & Shiffrin, 1980; Burgess & Shallice, 1996). These models often detail a process by which the memory system retrieves as many items as possible with one set of cues before assembling a new set of cues in an attempt to recall more items. These models of strategic retrieval would presumably make similar predictions to the CMR model regarding clustering by similarity and increased IRTs with a shift in similarity across any of the organizational dimensions. However, the CMR model describes an automatic process that produces these phenomena, obviating the need for an executive system that determines the optimal state of the next set of cues. While there is certainly room in the cognitive system for such executive processes, it

seems that for these data, the simpler CMR model is sufficient.

The CMR model is part of a longstanding tradition of context-based theories of human memory (McGeoch, 1942; Bower, 1972; Murdock, 1997; Dennis & Humphreys, 2001; Howard & Kahana, 2002a; Sederberg et al., in press), whereby an internal context representation is principally responsible for the guidance of memory search. These context-based models may be contrasted with dual store models of memory search (Raaijmakers & Shiffrin, 1980; Gillund & Shiffrin, 1984; Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005; Sirotin, Kimball, & Kahana, 2005; Kimball, Smith, & Kahana, 2007), which posit that in addition to a context-guided search of long-term memory, a major proportion of the behavioral variance is explained by a short-term buffer that can concurrently maintain the representations of several recently studied items.⁵

The modern debate between context-based and dual store models of human memory centers on the notion of a short-term store, which has been conceived as a buffer-like structure that can concurrently maintain a few item representations (e.g., Davelaar et al., 2005). When recall begins, the items still active in the buffer are read out, and then a context-based recall process is responsible for the rest of memory search. There has been longstanding friction among researchers regarding whether the short-term store is a necessary component of the memory system, or whether the dynamics of recall can be explained by a system with a single context-based search mechanism (as implemented, for example, in TCM-A; Sederberg et al., in press). Recently, Usher, Davelaar, Haarmann, and Goshen-Gottstein (in press) argued that TCM-A is itself a dual component model, the two components being the context representation and the episodic association matrix (referred to as M^{FC} and M^{CF} in both TCM-A and CMR). Of course, by this scheme, the Davelaar et al. (2005) model has three components, a context representation, an associative matrix, and a short-term store. This contrast highlights what some may see as a shortcoming of a context-based memory system: it has no explicit provision for a

separate working memory component (Baddeley, 1986), thought to be critical for explaining, for example, the ability of amnesic patients to recall several items in a free-recall experiment, without the ability to remember, a moment or so later, that they even participated in such an experiment (Carlesimo, Marfia, Loasses, & Caltagirone, 1996). However, as Sederberg et al. (in press) showed, the context representation can be used to momentarily keep a set of recently presented items in an enhanced state of accessibility (using the pre-experimental connections between item and context representations), allowing the model to fit the performance of amnesic patients. By this view, the amnesic syndrome is best characterized by an inability to form new associations between item features and context representations.

A second behavioral phenomenon raised as a critical marker of the presence of a short-term store is the sigmoidal shape of the serial position curve for the recency items often seen in immediate free recall (e.g., Murdock, 1962; Fig. 5). Usher et al. (in press) suggested that TCM (and by extension TCM-A and CMR) could not account for this effect because of the exponential decay of the elements comprising the context representation. However, recall is a dynamic process in these context-based models, and the probability of recall of the studied items does not map linearly onto the activation of a particular context unit at the time of recall initiation (owing both to context reinstatement and the nonlinear recall competition). As can be seen in Fig. 5, the CMR model is able to fit the sigmoidal shape of the serial position curve (for another example see Kahana, Sederberg, & Howard, in press).

While both the short-term store and the context representation are activation-based components, there are important differences between them. The short-term store account suggests that the first few recalls are items read out of the buffer, whereas context-based accounts suggest that these items are recalled because of their strong associations with the current state of the context representation. A recent study by Howard, Vankatadass,

Norman, and Kahana (2007) throws these two accounts into sharp contrast, by showing that when participants first recall an item that was presented twice in the list, the temporal neighbors of the original presentation of the item show an enhanced probability of being recalled. This is a natural consequence of a context-based model, where recall of the repeated item will retrieve context related to both presentations. It is difficult to see how a short-term store would account for this phenomenon.

A related issue was explored in the current manuscript, in our analysis of the effect characterized by Farrell and Lewandowsky (in press) (also explored thoroughly by Howard, Sederberg, and Kahana (2008)). This issue is described in *Simulation analysis I: temporal clustering*. In brief, if one of the first items recalled comes from a mid-list position, the next recall is sometimes another mid-list item (giving rise to a contiguity effect), and is sometimes an end-of-list item (giving rise to a recency effect). This finding was presented as a challenge for context-based models, because a version of TCM was unable to fit the data (Farrell & Lewandowsky, in press). As seen in Fig. 3, the CMR model is able to fit the effect qualitatively, showing both a contiguity effect and a recency effect for these first recalls from mid-list positions. This is because, after the first recall of the mid-list item, context is updated, creating a blend of mid-list context and end-of-list context. However, the CMR model underpredicts the magnitude of the contiguity effect in this situation, and it is worth reviewing how this underprediction arises.

Exploration of the parameter space of the model suggests that the CMR model's underprediction of the magnitude of the contiguity effect for early output positions is not something that can be simply remedied by adjusting the model parameters. Raising the recall period context-retrieval parameter (β_{rec}^{temp}) can bring the simulated results for the early output positions (Fig. 3B) into line with the behavioral data (Fig. 3A). Raising β_{rec}^{temp} causes the just-recalled item to more strongly reinstate its associated temporal context, which causes contiguous items to compete more effectively against the end-of-list

items in the decision competition. However, this adjustment causes the model to overpredict the magnitude of the contiguity effect for later output positions, reducing the overall goodness-of-fit of the model.

Two parameters control the likelihood that an early recall will come from a mid-list serial position: the semantic association parameter (s), and the noise parameter on the recall competition (η). Increases in each of these parameters tend to wrest recalls away from the end-of-list serial positions; s because semantic associations cause transitions that are random with respect to list structure, and η because all items are supported by noise equivalently. However, while increasing either of these parameters will increase the proportion of mid-list recalls, they will not increase the contiguity effect for those recalls. In fact, each will tend to work against the contiguity effect by increasing the likelihood that the next item recalled comes from a random serial position.

One mechanism that could both increase the likelihood of mid-list recalls early in the recall sequence, and increase the size of the contiguity effect for those recalls, would be the addition of variability in encoding strength. If some items are better encoded than others (by boosting the strength of the associative connections between the feature representation and the context representation; L_{tw}^{FC} and L_{tw}^{CF}), then these items would tend to be recalled earlier in the recall sequence, and once recalled, would more strongly reinstate context than a less well encoded item, causing a boost in the contiguity effect. The remaining items would be less well encoded, leading to a gradual reduction in the magnitude of the contiguity effect over the course of the recall period.

While the current version of CMR underpredicts the magnitude of the recency effect in this situation, it is unclear how a short-term store model would predict a contiguity effect at all in this situation. Davelaar et al. (2005), in a discussion of response latencies, suggest that when the first recall of an item comes from an early or mid-list position, that recall was likely because the item managed to remain in the buffer throughout the list

presentation (pg. 32). While it is clear why one would see a recency effect in this situation (the mid-list item shares the buffer with some of the final items from the list), it is unclear why reading out a mid-list item from the short-term store would then lead to recall of another mid-list item from a neighboring position in the study sequence (for a more thorough examination of this issue see Howard, Sederberg, & Kahana, 2008).

While the current study makes some points relevant to the debate between dual store and context-based models, the focus of the work is on developing our understanding of the nature of the context representation, a central component of both classes of models. Specifically, we show how the mechanisms developed in recent work on the nature of temporal context can be generalized to a much broader range of context-related phenomena, in particular, those corresponding to the source characteristics of the studied items. The CMR model describes the dynamics by which these source representations influence memory search, as well as how they interact with other components of the memory system. The goal of this endeavor is to create the simplest possible model that is consistent with the widest range of behaviors exhibited by the human memory system, in the domain of free recall. It is our hope that the CMR model will prove useful in determining the role of context representations in other related paradigms, beginning with the wide range of free-recall variants that have arisen over the many decades of research in this domain.

The control of memory search: Future directions

The best-fit version of the CMR model provides a set of parameter values that characterize the operation of the human memory system in terms of several relatively simple mechanisms (e.g., context updating, association formation, and decision making). The critical next step in this endeavor is to evaluate the utility of the model as an interpretive tool. It is our hope that the model can be used to disentangle the factors at

work during memory search, both in the behavioral and neural domains. Here, we outline some future work that may benefit from the CMR model: First, a description of how the model may extend our understanding of the memory deficit observed in healthy aging (Hasher & Zacks, 1988; Kahana & Wingfield, 2000; Naveh-Benjamin, 2000; Kahana et al., 2002). Second, a description of how the model may shed light on neural reinstatement effects during memory retrieval.

Howard, Kahana, and Wingfield (2006) showed that TCM could be used to investigate the age-related associative impairment by examining the temporal clustering behavior of young and older participants in a free-recall task. They concluded that the reduced temporal clustering observed in older participants (paired with intact recall initiation) was consistent with a variant of TCM in which older adults have an impaired ability to retrieve the temporal context associated with each studied item. The idea that the ability to retrieve associations is impaired in older participants is consistent with the finding of impaired performance in the associative recognition paradigm (e.g., Naveh-Benjamin, 2000), and in a number of source recognition paradigms (Johnson et al., 1993; Chalfonte & Johnson, 1996; McIntyre & Craik, 1987; Hashtroudi et al., 1989), in which one must retrieve associations between item and source information to respond correctly.

The generalization of TCM to create the CMR model primarily involved the addition of machinery to handle source information; this substantially broadens the class of memory phenomena that the model can be applied to. For example, the CMR model could be used to investigate whether older adults' source association deficits extend into the domain of free recall, and whether these deficits are best explained as an inability to retrieve associations between item features and source context, or an inability to inhibit the activation of competing representations (such as a competing source representation Hasher & Zacks, 1988). A study of older adults in a continuous performance task suggests

that the memory deficit in older adults may be related to an inability to properly maintain contextual representations (Braver et al., 2001), which will cause older adults to be more sensitive to context disruption than younger participants. The source-manipulation paradigm of the current article is well designed to assess this hypothesis: By examining the best-fit parameters for a set of older adults, we can assess the level of source association as well as the level of task-shift related disruption. In line with the arguments presented above, we expect that the best-fit model will show decreased associative strength between source context and item features, as well as increased task-shift related disruption of temporal context, leading to decreased remote same-task clustering, but increased local clustering (much like that shown by the Pure disruption model variant in Fig. 4E). Furthermore, older participants will likely show a larger decrement to memorability of items due to task shifts (and a larger relative increase in memorability of the final train), owing to the increase in the disruption parameter. Obtaining these results would extend our understanding of the age-related associative deficit, and provide further evidence for the utility of CMR as a general model of free recall.

A second future direction involves bridging relatively abstract cognitive theories of memory with the burgeoning literature on the neural substrate of memory. Models of memory search such as TCM and CMR provide a framework with which to interpret the functional significance of patterns of brain activity observed during free-recall performance. Recent studies have begun to visualize the process of memory reinstatement, whereby the pattern of neural activity observed when one studies a particular item is revived when that item is later recalled (Wheeler, Petersen, & Buckner, 2000; Polyn, Natu, Cohen, & Norman, 2005; Prince, Daselaar, & Cabeza, 2005; Sederberg et al., 2007).

These studies raise the possibility that we can identify the neural correlates of the item and context representations characterized by the CMR model, and track the coming and going of these representations over the course of memory search. Polyn et al. (2005)

took a step in this direction, by having participants study items drawn from three distinct categories (celebrities, landmarks, and objects). Using machine-learning techniques (Norman, Polyn, Detre, & Haxby, 2006), Polyn et al. (2005) characterized the pattern of brain activity associated with each study category, on a participant-by-participant basis. These machine-learning techniques were then used to assess the relative strengths of each category-specific pattern on a second-by-second basis over the course of the recall period. They found that the reinstatement of a particular category pattern predicted the upcoming recall of an item from that category. The CMR model suggests that the patterns identified in the Polyn et al. (2005) study were likely a blend of item and context information. According to the model, when an item is recalled, the system revives not only that item's representation, but also the pattern of context activity associated with that item. This context pattern then shapes the course of the following search, determining the probability of recalling any given studied item.

The next frontier in understanding the neural basis of memory search involves identifying the neural substrate of context. The CMR model provides a precise specification of the functional properties of the context representation: it must reflect the features and statistical properties of studied items, integrate information over long time-scales, and return to a prior state, given the recall of an item (Polyn & Kahana, 2008). Using these specifications, researchers may be able to identify candidate anatomical regions for the neural seat of context in the human brain. An emerging view, summarized by Polyn and Kahana (2008), is that prefrontal cortex is centrally involved in contextual processing. Patterns of activity in prefrontal cortex play a double role, both guiding how item representations in more posterior brain regions are processed (Miller & Cohen, 2001) and also serving to contextualize these patterns through associations formed between the two sets of patterns by the hippocampal formation (see also Norman, Detre, & Polyn, 2008). This hypothesis is consistent with evidence drawn from neuroimaging studies of

memory retrieval (Blumenfeld & Ranganath, 2007), neuropsychological studies of patients with prefrontal damage (Schacter, 1987), and computational models of the role of prefrontal cortex in free recall (Becker & Lim, 2003). A mechanistic specification of the role of prefrontal cortex in memory search will prove quite valuable in integrating the current framework with computational models of the medial temporal memory system (McClelland, McNaughton, & O'Reilly, 1995; Norman & O'Reilly, 2003).

The CMR model, coupled with a set of neural linking hypotheses, may serve as a valuable tool in interpreting the neural patterns observed in prefrontal and other brain regions during study as well as during the recall period. The CMR model is a predictive framework; given a particular set of studied items (which vary in semantic relatedness, list position, and source characteristics) the model can provide the most likely recall sequences. As the preceding analyses show, these predicted recall sequences match the characteristics of the observed sequences quite well. Patterns of neural activity observed during the study and recall periods, in various brain regions, can be used to gain predictive power regarding the order of recalls, allowing us to gain insight into the functional contribution of these regions in the domain of memory search.

References

- Allport, A., Styles, E., & Hsieh, S. (1994). Shifting intentional set: exploring the dynamic control of tasks. *Attention and Performance, 15*, 421-452.
- Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language, 38*, 341-380.
- Baddeley, A. D. (1986). *Working memory*. Oxford, England: Clarendon Press.
- Battig, W. F., & Bellezza, F. S. (1979). Organization and levels of processing. In C. R. Puff (Ed.), *Memory organization and structure* (pp. 321-346). New York: Academic Press.
- 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.
- Bhatarah, P., Ward, G., & Tan, L. (2006). Examining the Relationship Between Free Recall and Immediate Serial Recall: The Effect of Concurrent Task Performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(2), 215-229.
- Bjork, R. A., & Richardson-Klavehn, A. (1989). On the puzzling relationship between environmental context and human memory. In C. Izawa (Ed.), *Current issues in cognitive processes: the Tulane Flowerree symposium on cognition* (chap. 9). New Jersey: Lawrence Erlbaum Associates.
- Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology, 6*, 173-189.
- Blaxton, T. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*(4), 657-668.
- Blumenfeld, R., & Ranganath, C. (2007). Prefrontal cortex and long-term memory

- encoding: An integrative review of findings from neuropsychology and neuroimaging. *The Neuroscientist*, 13(3), 280.
- Botvinick, M., Cohen, J., & Carter, C. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, 8, 539–546.
- Botvinick, M., & Plaut, D. C. (2002, Nov). Representing task context: proposals based on a connectionist model of action. *Psychol Res*, 66(4), 298–311.
- Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 49, 229-240.
- Bousfield, W. A., & Sedgewick, C. H. W. (1944). An analysis of sequences of restricted associative responses. *Journal of General Psychology*, 30, 149-165.
- Bower, G. H. (1967). A multicomponent theory of the memory trace. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation : Advances in research and theory* (Vol. 1, p. 229-325). New York: Academic Press.
- Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (p. 85-121). New York: John Wiley and Sons.
- 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.
- Brodie, D. A., & Murdock, B. B. (1977). Effects of presentation time on nominal and functional serial position curves in free recall. *Journal of Verbal Learning and Verbal Behavior*, 16, 185-200.
- Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, 114(3), 539–576.
- Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial

- order. *Psychological Review*, 107(1), 127–181.
- Burgess, P., & Shallice, T. (1996). Confabulation and the Control of Recollection. *Memory*, 4(4), 359–412.
- 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.
- Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition*, 24, 403-416.
- Cofer, C. N., Bruce, D. R., & Reicher, G. M. (1966). Clustering in free recall as a function of certain methodological variations. *Journal of Experimental Psychology*, 71, 858-866.
- Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the control of automatic processes: A parallel distributed processing model of the Stroop effect. *Psychological Review*, 97(3), 332-361.
- Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing : A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11, 671-684.
- Curiel, J. M., & Radvansky, G. A. (1998). Mental organization of maps. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(1), 202-214.
- Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review*, 112, 3–42.
- Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, 108, 452-478.
- Donchin, E., & Coles, M. G. (1988, September). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, 11(3), 357-427.

- Estes, W. K. (1959). *Component and pattern models with markovian interpretations. studies in mathematical learning theory*. Stanford, CA: Stanford University Press.
- Fabianai, M., & Donchin, E. (1995). Encoding processes and memory organization: A model of the von restorff effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 224-240.
- Farrell, S. (2008). Dissociating conditional recency in immediate and delayed free recall: A challenge for unitary models of recency. *Submitted*.
- Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review*, *9*, 59-85.
- Farrell, S., & Lewandowsky, S. (in press). Empirical and theoretical limits on lag recency. *Psychonomic Bulletin and Review*.
- Friendly, M. (1979). Methods for finding graphic representations of associative memory structures. In *Memory organization and structure* (pp. 85–129). New York: Academic Press.
- Frost, N. (1971). Clustering by visual shape in the free recall of pictorial stimuli. *Journal of Experimental Psychology*, *88*, 409–413.
- Geiselman, R. E., Bjork, R. A., & Fishman, D. (1983). Disrupted retrieval in directed forgetting: a link with posthypnotic amnesia. *Journal of Experimental Psychology: General*, *112*, 58–72.
- Geller, A. S., Schleifer, I. K., Sederberg, P. B., Jacobs, J., & Kahana, M. J. (2007). PyEPL: A cross-platform experiment-programming library. *Behavior Research Methods*, *39*(4).
- Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review*, *91*, 1-67.
- Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, *5*, 351-360.

- Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (p. 193-225). San Diego: Academic Press.
- Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1989). Aging and source monitoring. *Psychology and Aging, 4*(1), 106–112.
- Hintzman, D. L., Block, R. A., & Inskip, N. R. (1972). Memory for mode of input. *Journal of Verbal Learning and Verbal Behavior, 11*, 741–749.
- Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 923–941.
- Howard, M. W., & Kahana, M. J. (2002a). A distributed representation of temporal context. *Journal of Mathematical Psychology, 46*, 269-299.
- Howard, M. W., & Kahana, M. J. (2002b). When does semantic similarity help episodic retrieval? *Journal of Memory and Language, 46*, 85–98.
- Howard, M. W., Kahana, M. J., & Wingfield, A. (2006). Aging and contextual binding: Modeling recency and lag-recency effects with the temporal context model. *Psychonomic Bulletin & Review, 13*(3), 439–445.
- Howard, M. W., Sederberg, P. B., & Kahana, M. J. (2008). Reply to Farrell & Lewandowsky: Changes in the shape of the lag-CRP predicted by TCM due to recency. *Submitted*.
- Howard, M. W., Vankatadass, V., Norman, K. A., & Kahana, M. J. (2007). Associative processes in immediate recency. *Memory & Cognition, 35*(7), 1700–1711.
- Howard, M. W., Youker, T. E., & Venkatadass, V. (2008). The persistence of memory: Contiguity effects across several minutes. *Psychonomic Bulletin & Review, 15*, 58–63.
- Hulme, C., Stuart, G., Brown, G., & Morin, C. (2003). High-and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial

- recall. *Journal of Memory and Language*, 49(4), 500–518.
- Hyde, T. S., & Jenkins, J. J. (1969). Differential effects of incidental tasks on the organization of recall of a list of highly associated words. *Journal of Experimental Psychology*, 82, 472–481.
- Jacoby, L., Shimizu, Y., Daniels, K., & Rhodes, M. (2005). Modes of cognitive control in recognition and source memory: Depth of retrieval. *Psychonomic Bulletin & Review*, 12(5), 852–857.
- James, W. (1890). *The principles of psychology*. New York, NY, US: Henry Holt and Co, Inc.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114(1), 3–28.
- Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, 24, 103–109.
- Kahana, M. J., Howard, M. W., & Polyn, S. M. (2008). Associative retrieval processes in episodic memory. In H. L. Roediger III (Ed.), *Cognitive psychology of memory. Vol. 2 of Learning and memory: A comprehensive reference, 4 vols. (J. Byrne, Editor)*. Oxford: Elsevier.
- 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.
- Kahana, M. J., Sederberg, P. B., & Howard, M. W. (in press). Putting short-term memory into context: Reply to Usher and colleagues. *Psychological Review*.
- Kahana, M. J., & Wingfield, A. (2000). A functional relation between learning and organization in free recall. *Psychonomic Bulletin & Review*, 7, 516–521.
- Kahana, M. J., Zhou, F., Geller, A. S., & Sekuler, R. (2007). Lure-similarity affects visual episodic recognition: Detailed tests of a noisy exemplar model. *Memory &*

Cognition, 35, 1222–1232.

Karis, D., Fabiani, M., & Donchin, E. (1984). P300 and memory: Individual differences in the Von Restorff effect. *Cognitive Psychology*, 16, 177-216.

Kimball, D. R., Smith, T. A., & Kahana, M. J. (2007). The fSAM model of false recall. *Psychological Review*, 114(4), 954–93.

Kintsch. (1970). *Learning, memory, and conceptual processes*. New York, N.Y.: John Wiley & Sons.

Kolers, P. A., & Ostry, D. J. (1974). Time course of loss of information regarding pattern analyzing operations. *Journal of Verbal Learning and Verbal Behavior*, 13, 599-612.

Kolers, P. A., & Roediger, H. L. (1984). Procedures of mind. *Journal of Verbal Learning and Verbal Behavior*, 23, 425-449.

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

Laming, D. (2006). Predicting free recalls. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(5), 1146-1163.

Laming, D. (2008). An improved algorithm for predicting free recalls. *Cognitive Psychology*, 57, 179–219.

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

Lewandowsky, S., Brown, G. D. A., & Thomas, J. (submitted). Output order in short-term memory: Evidence for strategic retrieval.

Lewandowsky, S., & Farrell, S. (in press). Short-term memory: New data and a model. *The Psychology of Learning and Motivation*, 49.

Luce, R. D. (1959). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (p. 103-189). New York: Wiley.

- McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, *102*(3), 419–57.
- McGeoch, J. A. (1942). *The psychology of human learning: An introduction*. New York: Longmans.
- McIntyre, J., & Craik, F. (1987). Age differences in memory for item and source. *Psychology and Aging*, *4*, 106–112.
- Mensink, G.-J. M., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, *95*, 434–455.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.
- Monsell, S. (2003, Mar). Task switching. *Trends in Cognitive Sciences*, *7*(3), 134–140.
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, *16*, 519–533.
- Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*, 482–488.
- Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review*, *104*, 839–862.
- Murdock, B. B., & Metcalfe, J. (1978). Controlled rehearsal in single-trial free recall. *Journal of Verbal Learning and Verbal Behavior*, *17*, 309–324.
- Murdock, B. B., & Okada, R. (1970). Interresponse times in single-trial free recall. *Journal of Verbal Learning and Verbal Behavior*, *86*, 263–267.
- Murdock, B. B., & Walker, K. D. (1969). Modality effects in free recall. *Journal of Verbal Learning and Verbal Behavior*, *8*, 665–676.

- Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: the ICE theory. *Journal of Experimental Psychology: General*, *128*, 403.
- Naveh-Benjamin, M. (2000). Adult-age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1170-1187.
- Neath, I. (1993). Contextual and distinctive processes and the serial position function. *Journal of Memory and Language*, *32*, 820-840.
- Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments and Computers*, *36*(3), 402-407.
- Nilsson, L.-G. (1974). Further evidence for organization by modality in free recall. *Journal of Experimental Psychology*, *103*(5), 948-957.
- Norman, K. A., Detre, G. J., & Polyn, S. M. (2008). Computational models of episodic memory. In R. Sun (Ed.), *The cambridge handbook of computational cognitive modeling*. in press.
- 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.
- Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424-430.
- Patterson, K. E., Meltzer, R. H., & Mandler, G. (1971). Inter-response times in categorized free recall. *Journal of Verbal Learning and Verbal Behavior*, *10*, 417-426.
- Pollio, H. R., Richards, S., & Lucas, R. (1969). Temporal properties of category recall. *Journal of Verbal Learning and Verbal Behavior*, *8*, 529-536.
- Polyn, S. M. (2005). *Neuroimaging, behavioral, and computational investigations of*

- memory targeting*. Unpublished doctoral dissertation, Princeton University.
- Polyn, S. M., & Kahana, M. J. (2008). Memory search and the neural representation of context. *Trends in Cognitive Sciences, 12*, 24–30.
- Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science, 310*, 1963–1966.
- Polyn, S. M., Norman, K. A., & Kahana, M. J. (Submitted). Task context and organization in free recall.
- Postman, L., & Adams, P. A. (1956). Studies in incidental learning: Iv. the interaction of orienting tasks and stimulus materials. *Journal of Experimental Psychology, 51*(5), 329–333.
- Postman, L., & Phillips, L. W. (1965). Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology, 17*, 132–138.
- Prince, S., Daselaar, S., & Cabeza, R. (2005). Neural Correlates of Relational Memory: Successful Encoding and Retrieval of Semantic and Perceptual Associations. *Journal of Neuroscience, 25*(5), 1203–1210.
- Puff, C. R. (1979). Memory organization research and theory: The state of the art. In C. R. Puff (Ed.), *Memory organization and structure* (pp. 3–17). New York: Academic Press.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 14, p. 207–262). New York: Academic Press.
- Rao, V. A., & Howard, M. W. (2008). Advances in neural information processing systems 20. In Y. S. J.C. Platt D. Koller & S. Roweis (Eds.), (chap. Retrieved context and the discovery of semantic structure). MIT Press: Cambridge, MA.
- Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free

- recall. *Memory & Cognition*, *22*, 511- 524.
- Rundus, D. (1971). An analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, *89*, 63-77.
- 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.
- Rutherford, A. (2000). The ability of familiarity, disruption, and the relative strength of nonenvironmental context cues to explain unreliable environmental-context-dependent memory effects in free recall. *Mem Cognit*, *28*(8), 1419–28.
- Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(6), 1064-1072.
- Schacter, D. L. (1987). Memory, amnesia, and frontal lobe dysfunction. *Psychobiology*, *15*, 21-36.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6* (2), 461-464.
- Sederberg, P. B., Howard, M. W., & Kahana, M. J. (in press). A context-based theory of recency and contiguity in free recall. *Psychological Review*.
- Sederberg, P. B., Schulze-Bonhage, A., Madsen, J. R., Bromfield, E. B., Litt, B., Brandt, A., et al. (2007). Gamma oscillations distinguish true from false memories. *Psychological Science*, *18*(11), 927–932.
- Shiffrin, R. M. (1970). Forgetting: Trace erosion or retrieval failure? *Science*, *168*, 1601-1603.
- Sirotin, Y. B., Kimball, D. R., & Kahana, M. J. (2005). Going beyond a single list: Modeling the effects of prior experience on episodic free recall. *Psychonomic Bulletin & Review*, *12*(5), 787-805.

- Smith, S. M. (1988). Environmental context-dependent memory. In G. M. Davies & D. M. Thomson (Eds.), *Memory in context: Context in memory*. (p. 13-34). Oxford, England: John Wiley & Sons.
- Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, 6(4), 342–353.
- Strand, B. Z. (1970). Change of context and retroactive inhibition. *Journal of Verbal Learning & Verbal Behavior*, 9(2), 202–206.
- Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1589-1626.
- Thapar, A., & Greene, R. L. (1993). Evidence against a short-term- store account of long-term recency effects. *Memory & Cognition*, 21, 329-337.
- Tulving, E., & Colotla, V. A. (1970). Free recall of trilingual lists. *Cognitive Psychology*, 1, 86–98.
- Tulving, E., & Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80, 352-373.
- Underwood, B. J. (1969). Some correlates of item repetition in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 8, 83-94.
- Unsworth, N. (in press). Exploring the retrieval dynamics of delayed and final free recall: Further evidence for temporal-contextual search. *Journal of Memory and Language*.
- Usher, M., Davelaar, E. J., Haarmann, H. J., & Goshen-Gottstein, Y. (in press). Short-term memory after all: Comment on Sederberg, Howard and Kahana (2008). *Psychological Review*.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108(3), 550-592.
- Ward, G., Woodward, G., Stevens, A., & Stinson, C. (2003). Using overt rehearsals to

explain word frequency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 29, 186-210.

Wheeler, M. E., Petersen, S. E., & Buckner, R. L. (2000). Memory's echo: Vivid remembering reactivates sensory-specific cortex. *Proceedings of the National Academy of Sciences*, 97, 11125.

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

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.

Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin and Review*, 1, 89-106.

Wylie, G., & Allport, A. (2000). Task switching and the measurement of "switch costs". *Psychological Research*, 63, 212-233.

Appendix A

Experimental methods for the source-manipulation experiment

45 participants (28 female) from the University of Pennsylvania community received payment in accordance with the University's IRB guidelines. Stimuli were presented with a computer running PyEPL (Python Experiment Programming Library:

<http://pyepl.sourceforge.net>, Geller, Schleifer, Sederberg, Jacobs, & Kahana, 2007).

Verbal responses were recorded with a microphone and parsed with the pyParse package.

On each trial, a list of 24 words was presented; each item was concurrently presented with a task cue, indicating the judgment that the participant should make for that word. Each word was presented for 3 seconds. The two tasks were a size judgment ("Will this item fit into a shoebox?") and an animacy judgment ("Does this word refer to something living or not living?"). Immediately following the list, a row of asterisks appeared, along with a beep, indicating the start of the recall period. Participants were given 90 seconds to recall as many words as they could remember from the most recent list, in any order.

There were two conditions, control and task shift. On control lists, every word was judged with the same encoding task. On the task-shift lists, participants shifted back and forth between the two tasks, first judging a short train of items from one task, and then switching to the other task for another short train. Each train was constrained to be between 2 and 6 items long (inclusive), and the ordering of these trains was randomized. We counterbalanced (across lists) the task used to start the list and the number of trains in the list (6 or 7).

The words on a given list were chosen such that in total there would be a roughly equivalent number of items associated with each response ("big", "small", "living", and "nonliving"). Many words are ambiguous with regard to the "correct" judgment (e.g., given the word "dog", an image of a chihuahua might elicit a "small" judgment, while an

image of a Great Dane might elicit a “big” judgment). We ran a small norming study in which 12 participants judged each of 1297 words using these tasks. This allowed us, in the current experiment, to avoid words that were ambiguous; to choose words that tended to be quickly judged; and to include roughly equal numbers of words associated with each response. In general, the responses that participants made in the free-recall experiment were highly correlated with the responses of the participants in the norming study.

Appendix B

Details of the USF free-association simulations

As described in *Associative connections: learning and semantic structure*, the semantic clustering estimates of the CMR model are inflated relative to the behavioral data. Here, we describe a simulation study designed to estimate the degree to which an average human participant’s semantic memory mismatches the LSA semantic association values, in order to create a correction factor to apply to the semantic clustering scores produced by the model.

In the USF free-association study, a large number of participants were asked to respond to a given cue word with the first word that came to mind. We randomly chose five words from the USF database which were also words in the word pool of our source-manipulation experiment (*agent, bracelet, elephant, glove, plane*). Between these five cue words, 74 distinct target words were produced by the USF participants. We obtained the LSA association values (as were used to create the semantic memory for CMR) for each cue word to the full set of 74 targets, excluding words not in the LSA corpus. We then created a very simple choice model, using the LSA association values (CMR’s semantic memory): Given a particular cue word, the model selects the target with the largest LSA association value as its response. The model had one parameter, the variance of randomly distributed noise that was added to each LSA association value to simulate the variability between participants. We searched for the value of this variance that minimized the difference (measured with RMSD) between the responses produced by the CMR semantic memory, and the USF behavioral data. If the variance is too low, the model does not produce enough variability in its responses to match the USF data. If the variance is too high, the model may produce an entirely unrelated target (e.g. “secret” given “elephant”). The technique produced a smooth and stable curve with a minimum

when the variance of the noise distribution was set to 0.41. This variance estimate was then used, as described in the main text, to correct the estimates of semantic clustering produced by the CMR model.

Appendix C

Details of the genetic algorithm fitting technique

A genetic algorithm was used to find the parameter set for each variant of the CMR model that allowed the model to best fit the behavioral data. In order to determine the best-fitting parameter set, we attempted to simultaneously minimize the deviation between the model predictions and the behavioral data for a large number of behavioral measures. The following aspects of the behavioral data were used to assess the goodness of fit of a given parameter set of the model (93 data points in total; each point contributed equally to the overall χ^2 goodness of fit):

- The overall probability of making a same-task transition, as well as the probability of making a remote same-task transition, for both the task-shift and relabeled control conditions [4 data points].
- The binned lag-CRP values for recall transitions originating from serial positions 5 through 19, both for early output positions (1 to 3) and later output positions (4 onwards), from the control condition (lag bins: -19 to -18, -17 to -6, -5 to -2, -1, 1, 2 to 5, 6 to 17, and 18 to 19) [16 data points].
- The final three serial positions of the probability of first, second, and third recall curves from the control condition [9 data points].
- All points from the train serial position analysis for the task-shift and relabeled control conditions, as well as the differences between the conditions at each train position [21 data points].
- The points from -5 to +5 from the train-lag CRP analysis for the task-shift and relabeled control conditions, as well as the differences between the conditions at each train lag [33 data points].
- The mean IRT for the first 10 output positions from the control condition [10

data points].

For each of the model variants, the following procedure was used to find the best-fit parameter set: The first generation of the genetic algorithm consisted of eight thousand points uniformly randomly selected from pre-determined ranges along each of the parameters. Then the algorithm was run for 15 generations, where each successive generation took the most fit 20% of the previous generation, and used these “parent” parameter sets to form 1024 new parameter sets to simulate, by randomly repairing the parameters and adding random “mutation” to all values (using a random normal distribution with mean zero and standard deviation 0.1). Then another 10 generations were run, with the mutation standard deviation dropped to 0.05, and 512 parameter sets per generation. For these 10 generations each simulated experiment generated 3 times as much data as the original experiment. Finally, the top 256 best-fitting parameter sets were each re-run (generating 12 times as much data as the original experiment) to find the final parameter set.

In calculating BIC goodness of fit scores reported in the main text, we scaled down the contribution of the IRTs (by dividing these observations by 10^5). This served to bring the variability in IRTs into the same range as the performance scores and conditional probabilities reported in the other analyses, and was done in order to ensure that the IRTs had a similar influence on the goodness of fit of the model as the other behavioral measures. Such scaling was not necessary for the χ^2 goodness of fit, which is already normalized by the standard error of the observations (causing the influence of the IRTs, which have large error terms, to be roughly comparable with the other data points).

Author Note

Special thanks to Jonathan Miller and Neal Morton for assistance with the programming of behavioral analyses, running participants, and analyzing data. Thanks to Simon Dennis for providing LSA similarity values. Thanks to Per Sederberg, Bennet Murdock, Christoph Weidemann, Stephan Lewandowsky, John Wixted, Colleen Kelley, and Marc Howard for helpful comments on the manuscript. This work was supported by NIH grants: MH078513 to SMP, MH069456 to KAN, MH055687 to MJK. The simulation code and supporting documentation is available from <http://memory.psych.upenn.edu/CMR>.

Notes

¹They further proposed that a similar mechanism underlies the reduced accessibility of materials cued to be forgotten in the directed-forgetting paradigm (Geiselman, Bjork, & Fishman, 1983).

²This number of parameters is comparable to other dynamical models of free recall. For example, eSAM contains 11 parameters (Sirotin et al., 2005), and TCM-A contains 12 parameters (Sederberg et al., in press).

³A similar method was used by Nilsson (1974) to quantify the degree of clustering by modality.

⁴Subsequent studies of environmental context change revealed that with careful control of the amount of disruption each group of participants received (e.g., by sending all groups to a waiting room between lists), an effect of context change could also be observed (Smith, Glenberg, & Bjork, 1978; Rutherford, 2000)

⁵Mensink and Raaijmakers (1988) straddles these two traditions in that it uses the dual store SAM framework but posits a critical role for context in explaining interference effects in paired associates learning.

Best-fit parameters of the CMR model.

	M62	MO70	Full	P.A.	P.D.
β_{enc}^{temp}	0.745	0.621	0.776	0.767	0.772
β_{rec}^{temp}	0.36	0.179	0.510	0.468	0.510
β^{source}	–	–	0.588	0.681	0.743
L_{sw}^{CF}	–	–	0.129	0.171	0
d	–	–	0.767	0	0.880
γ^{FC}	0.581	0.559	0.898	0.799	0.889
s	1.80	3.06	2.78	2.71	2.80
κ	0.091	0.166	0.111	0.053	0.092
λ	0.375	0.284	0.338	0.272	0.349
η	0.182	0.072	0.159	0.126	0.183
τ	0.242	0.323	0.174	0.145	0.201
ϕ_s	5.39	6.0	1.07	0.881	1.83
ϕ_d	1.41	0.916	0.981	0.641	0.942
$d.pt.s.$	150	65	93	93	93

Table 1

These parameters were discovered using a genetic algorithm fitting technique described in Appendix C. Codes for the different simulations are as follows: M62, Murdock (1962) data set; MO70, Murdock and Okada (1970) data set; Full, Full variant applied to the source-manipulation experiment; P.A., Pure association variant applied to the source-manipulation experiment; P.D., Pure disruption variant applied to the source-manipulation experiment. $d.pt.s.$ corresponds to the number of behavioral datapoints fit by each simulation.

Clustering scores during recall.

	Semantic	Temporal
<i>Behavioral Data</i>	0.545 (0.005)	0.638 (0.011)
<i>CMR: Full model</i>	0.566 (0.001)	0.636 (0.001)
<i>CMR: Pure association</i>	0.569 (0.001)	0.630 (0.001)
<i>CMR: Pure disruption</i>	0.568 (0.001)	0.647 (0.001)

Table 2

Note. Clustering scores are followed by standard error of subject means in parentheses. All results are significant at the $p < 0.001$ level. See text for details of the statistical analysis. A correction factor has been applied to the semantic clustering scores produced by the CMR model, as described in the text (*The structure and dynamics of the model. Associative connections: learning and semantic structure*).

Source clustering during recall.

<i>Behavioral Data</i>	All transitions	Remote transitions
Relabeled control	0.536 (0.009)	0.334 (0.011)
Shift	0.606 (0.008)	0.391 (0.012)
<i>CMR: Full model</i>		
Relabeled control	0.528 (0.001)	0.336 (0.001)
Shift	0.611 (0.002)	0.410 (0.003)
<i>CMR: Pure association</i>		
Relabeled control	0.523 (0.001)	0.336 (0.001)
Shift	0.604 (0.002)	0.413 (0.002)
<i>CMR: Pure disruption</i>		
Relabeled control	0.530 (0.001)	0.333 (0.001)
Shift	0.578 (0.002)	0.368 (0.002)

Table 3

Note. Transition probabilities are followed by standard error of subject means in parentheses.

Source shift cost in recall latencies (output transitions 1–8).

	Shift cost
<i>Behavioral Data</i>	1302 (276)
<i>CMR: Full model</i>	432 (82)
<i>CMR: Pure association</i>	217 (65)
<i>CMR: Pure disruption</i>	586 (84)

Table 4

Note. Inter-response times in milliseconds are followed by standard error on subject means.

Shift-cost is the difference between the same-task and between-task transitions. See text for details of the statistical analysis.

Mean reaction times for judgments during the study period				
Task	Control	Shift, repeat	Shift, new	Switch cost
Size	1336 (33)	1405 (29)	1626 (33)	221 ^a (18)
Animacy	1319 (34)	1414 (32)	1578 (35)	164 ^a (18)
Combined	1328 (33)	1409 (30)	1601 (33)	192 ^a (15)

Table 5

Note. These data are from the judgments in the behavioral data from the source-manipulation experiment; the CMR model did not simulate these judgments. Reaction times are reported in milliseconds and are followed by standard error of subject means. Shift, repeat refers to items in the task-shift condition where the encoding task is the same as that for the preceding item. Shift, new refers to items in the task-shift condition where the encoding task is different from the preceding item. ^a Significant at $p < 0.0001$ (t -test).

Figure Captions

Figure 1. Context as the spotlight of memory. Here we envision context as a set of spotlights, each shining into memory. Each lamp can illuminate a different subset of memories. The temporal lamp always illuminates a set of traces that were stored nearby in time (the light becoming more diffuse for more distant items). The source lamp illuminates memories that were associated with similar source characteristics. The context retrieved upon successful recall of an item may swing each lamp to illuminate a different set of items.

Figure 2. The basic structure of the Context Maintenance and Retrieval (CMR) model. Item features can update the state of context via an associative matrix M^{FC} . The context representation itself has maintenance capabilities which allows context information to be integrated over long temporal intervals. During recall, the state of context can bias the retrieval competition towards items associated with the active context elements, through the associative matrix M^{CF} . When a particular item is recalled, its features are reactivated; it can then update context to bias further retrieval attempts.

Figure 3. Conditional probability of recall transitions as a function of lag, for the behavioral data from the source-manipulation experiment. Transitions originating from mid-list items (serial positions 5 through 19) are considered. Each dot represents the aggregate probability for a set of lags, marked according to the mean lag of that bin; from left to right these are: -19 to -18, -17 to -6, -5 to -2, -1, 1, 2 to 5, 6 to 17, and 18 to 19. A) The behavioral data, probabilities for output positions 1 to 3. B) The CMR Full model, probabilities for output positions 1 to 3. C) The behavioral data, probabilities for output positions 4 onward. D) The CMR Full model, probabilities for output positions 4 onward.

Figure 4. Conditional response probability analysis of recall transitions by train lag, for

the relabeled control and task-shift conditions. A and B correspond to the analysis on the behavioral data. A) Each point corresponds to the probability of making a transition of a particular train lag from the just-recalled item. B) The difference in probabilities between the relabeled control and task-shift conditions. Black squares correspond to transitions between items studied with the same task, and white squares to transitions between items studied with different tasks. C and D) The same analysis of the simulated data from the CMR Full model. E and F) The same analysis of the simulated data from the CMR Pure disruption model. G and H) The same analysis of the simulated data from the CMR Pure association model.

Figure 5. The probability of recall by serial position, across list lengths (LL) of 20, 30, and 40 words. A) Behavioral data from a subset of the conditions reported in Murdock, 1962. B) Simulated data from the best-fitting parameter set of the CMR model (parameter set M62 in Table 1).

Figure 6. Recall probability by serial position for each of the first three output positions. The simulation fits use the CMR Full model, but all three variants show similar fits to the serial position curves for the first three output positions.

Figure 7. The probability of recall by serial position. A) Behavioral data from the source-manipulation experiment. B) Simulated data from the best-fitting parameter set of the CMR model (parameter set Full in Table 1).

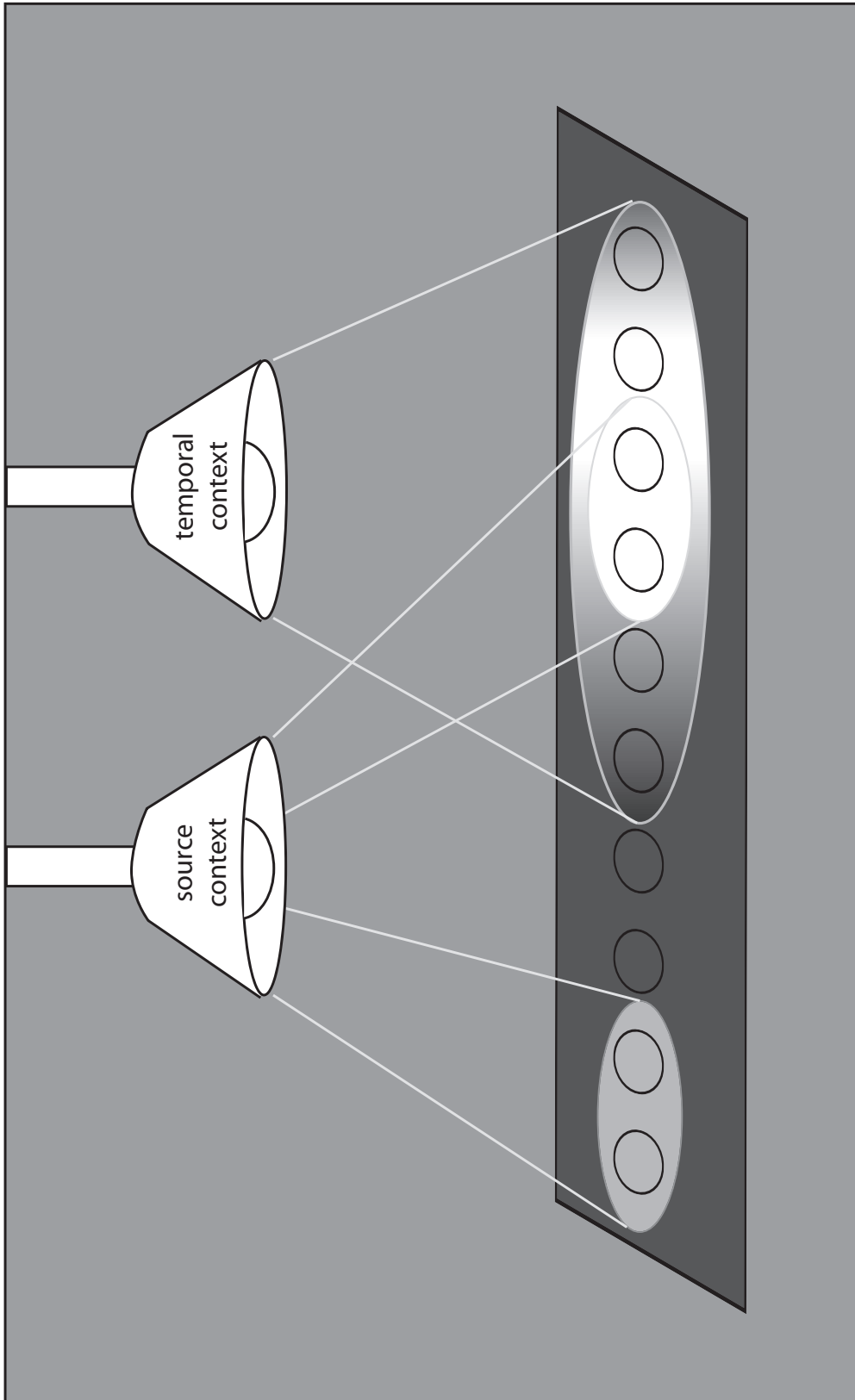
Figure 8. The proportion of items recalled from each train of same-task items in the source-manipulation experiment. A) Behavioral data for the relabeled control condition and the task-shift condition. B) The difference between these two conditions for each train, a negative value means worse performance in the task-shift condition. Error bars depict the standard error on the participant means. C and D) The same results calculated

with the simulated data from the Full CMR model.

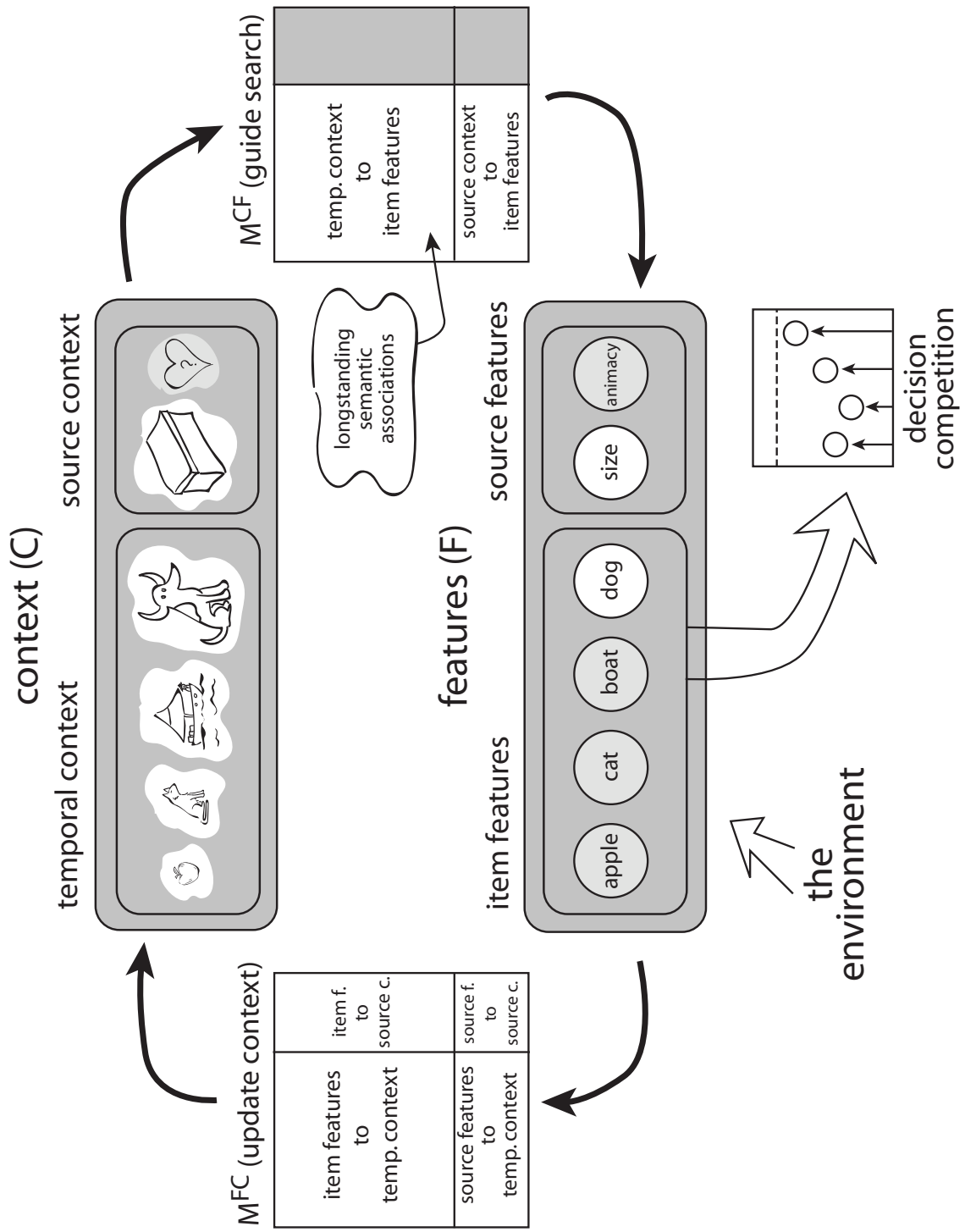
Figure 9. The proportion of items recalled from each train of same-task items in the source-manipulation experiment. A) Simulation results for the Pure disruption CMR variant, for the relabeled control condition and the task-shift condition. B) The difference between these two conditions for each train. Error bars represent the standard error on the participant means. C and D) The same results calculated for the Pure association model variant.

Figure 10. Inter-response time as a function of output position and total number of items recalled on that trial (the four lines correspond to trials on which 4, 5, 6, and 7 items were recalled). A) Behavioral data from Murdock and Okada, 1970. B) Simulated data from the best-fitting parameter set of the CMR model (parameter set MO70 in Table 1).

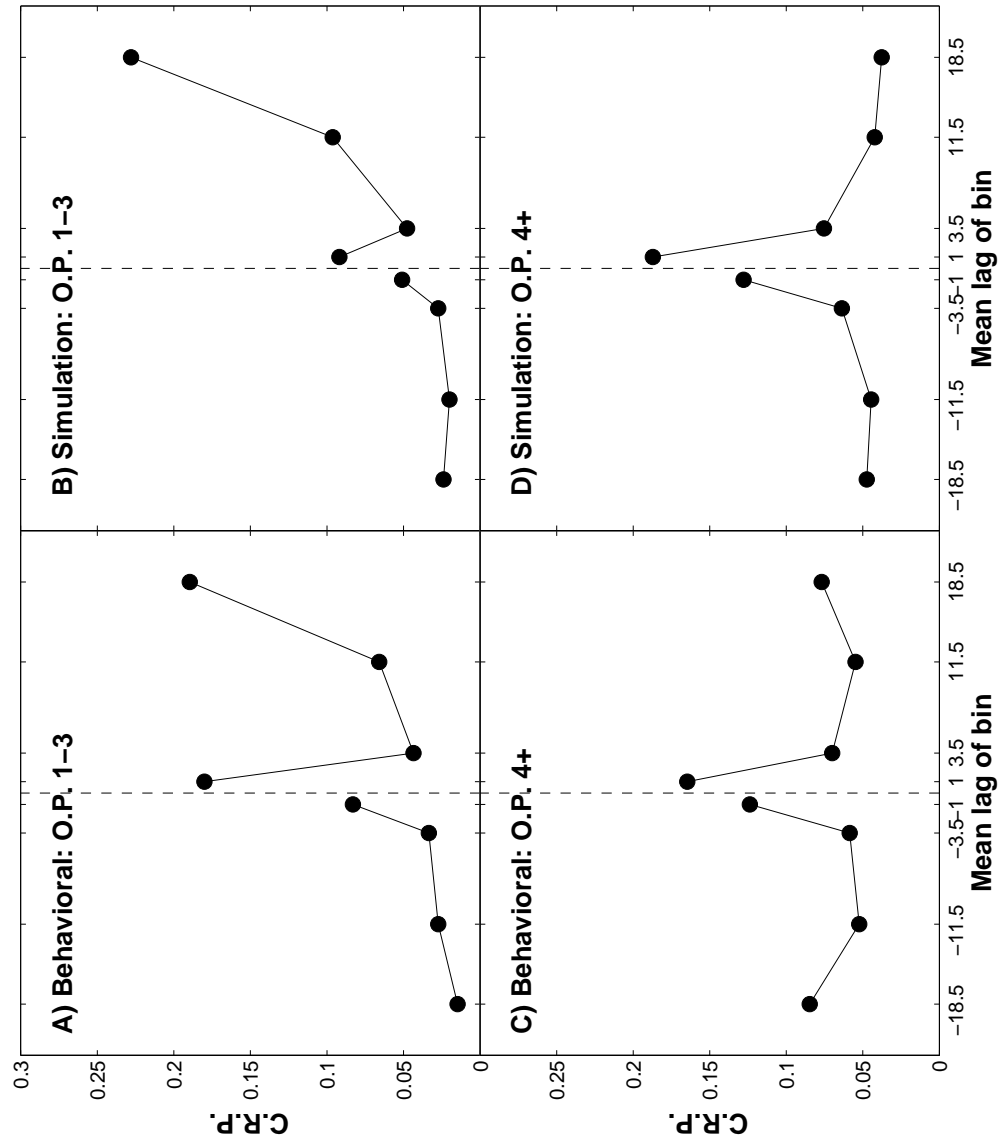
Context maintenance and retrieval, Figure 1



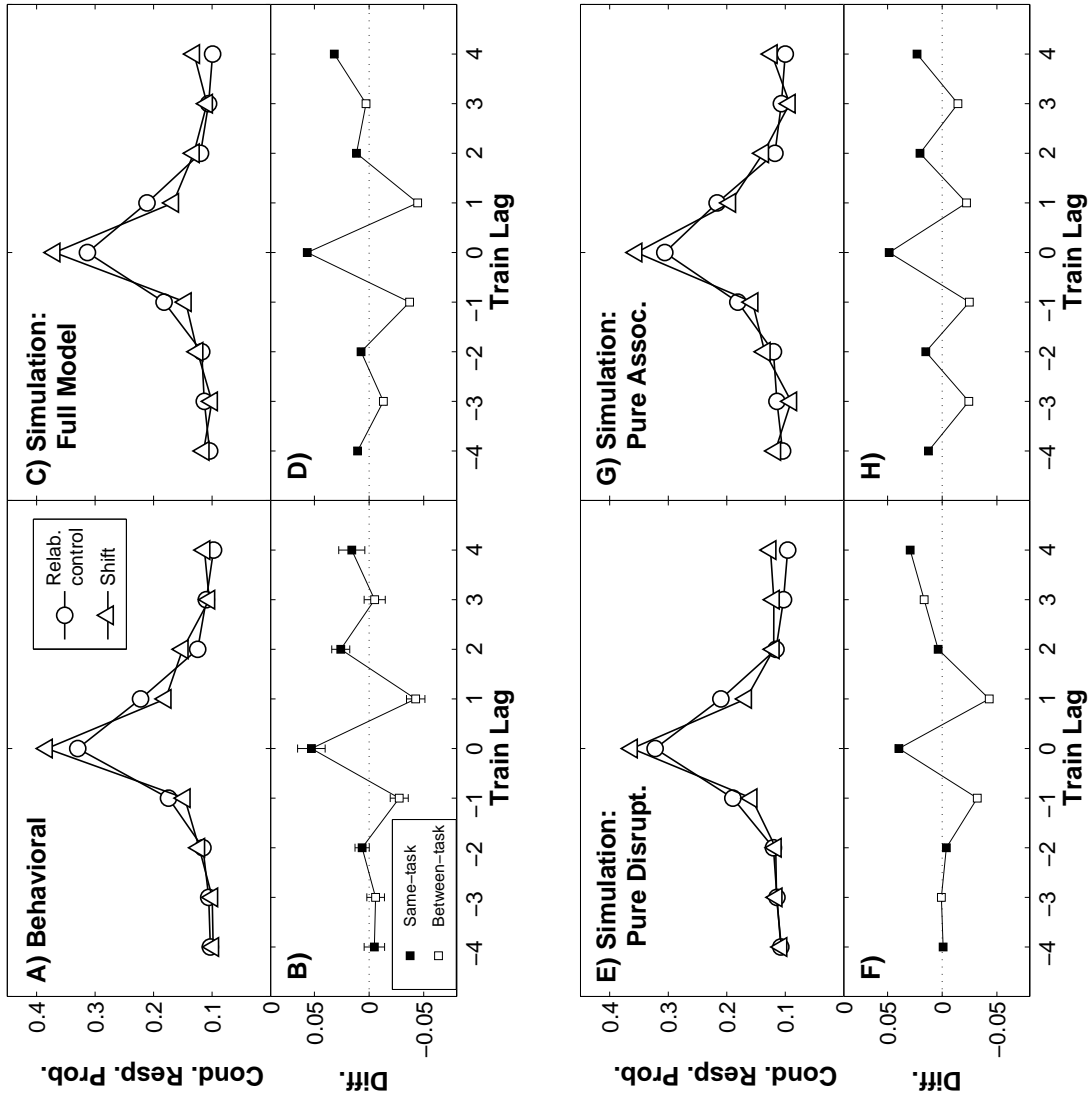
Context maintenance and retrieval, Figure 2



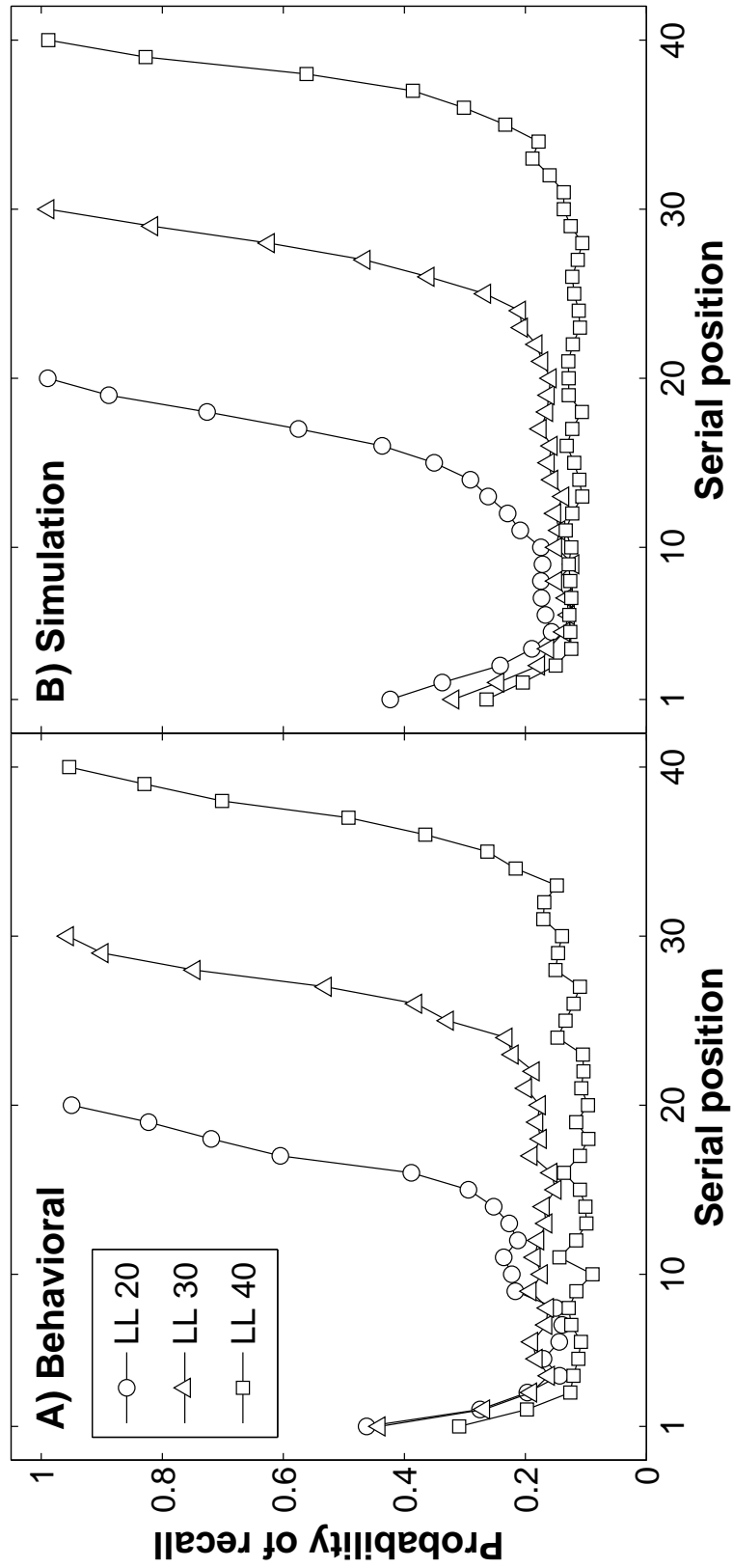
Context maintenance and retrieval, Figure 3



Context maintenance and retrieval, Figure 4



Context maintenance and retrieval, Figure 5



Context maintenance and retrieval, Figure 6

