A neurocognitive theory of episodic and semantic interactions during memory search

Neal W Morton & Sean M. Polyn
Department of Psychology
Vanderbilt University

Abstract

The context maintenance and retrieval (CMR) model proposes that a contextual representation is constructed as an experience unfolds, continually being updated by semantic information associated with the details of the experience. Thus, context is a blend of recently retrieved semantic information, but also becomes a unique representation of a given temporal interval. The composite nature of context supports the flexible targeting of memories on the basis of both temporal and semantic constraints. The model accounts for behavioral phenomena reflecting the interaction between the temporal and category structure of studied material. The interacting cognitive mechanisms that give rise to this behavior also make specific predictions about the relationship between the representational structure of context and the behavioral phenomena exhibited by a participant. We found evidence supporting these predictions in category-sensitive distributed patterns of neural activity recorded during both study and recall. The model provides a mechanistic interpretation of several neural-behavioral phenomena, including the tendency for items with strong category-specific signals at encoding to later be recalled as part of a category cluster, and the tendency for neural signal to become progressively more category-specific as multiple same-category items are studied in succession. We also custom-fit the model to each participant's recall behavior, and found that differences in the parameterization of these models across participants explained individual differences in neural signal. These results establish the viability of using a neurocomputational framework to investigate the relationship between neural signal and the construction and retrieval of episodic memories. *Keywords*: free recall, computational modeling, scalp electroencephalography, oscillatory power, multivariate pattern analysis

This work was supported by the NSF (1157432) and a Vanderbilt University Discovery Grant. We thank Michael Kahana, Gordon Logan, Tom Palmeri, Jim Kragel, and Joshua McCluey for helpful discussions.

Introduction

With his proposal regarding the distinction between episodic and semantic memory, Tulving (1972) transformed the way psychologists and neuroscientists think about the human memory system. Under this framework (Tulving, 1983), episodic memories correspond to particular experiences, and contain information about the particular spatiotemporal context of an event. Semantic memories, in contrast, are not associated with a particular context; they correspond to stable, fact-based memories. Tulving proposed that episodic and semantic information are handled by independent but interacting memory systems. We revisit this issue, using a retrieved-context model of human memory (Howard & Kahana, 2002; Kahana, 2012), the Context Maintenance and Retrieval model (CMR; Polyn, Norman, & Kahana, 2009), to understand the behavioral and neural phenomena observed when participants study and then recall materials with strong category structure. This mechanistically explicit framework allows us to specify the nature of the interactions between episodic and semantic memory, and the sense in which they are independent. Our approach builds upon the Complementary Learning Systems (CLS) framework of McClelland, McNaughton, and O'Reilly (1995), in that distinct associative structures support the two forms of memory. However, CMR goes beyond the CLS framework, in describing how episodic and semantic information interact to construct a temporal-semantic context representation that is ever-changing, is associated to representations of studied material, and is used to guide memory search. We use CMR to define a set of cognitive processes that bridge between neural signal and behavioral phenomena, and examine the ability of different model variants to explain behavioral and neural variability.

The Complementary Learning Systems (CLS) framework of McClelland et al. (1995) provides a mechanistically explicit formulation of the episodic and semantic memory systems, that we will build upon here. Under this framework, episodic memory for particular events (which, by definition, occur just once) is supported by a system that rapidly forms associative structures which bind the details of a particular experience to one another. A second system creates associative structures over much longer timescales, corresponding to stable, reliable properties of the world that might repeat themselves many times. If a fact is presented in many different contexts, such as an image of a particular celebrity along with their name, this slow learning system will create a semantic memory structure that supports this association. While creating this structure, the slow learning process

will average out many inconsistent details, including the diverse spatial and temporal contexts in which the individual events occurred. As an ecological example of this, we can consider that the average undergraduate at a university has elaborate, longstanding memory structures containing conceptual knowledge about hundreds of celebrities. These structures contain associations not only between names and faces, but also to a network of trivia regarding a celebrities' general oeuvre, and often scattered gossip regarding their life events and relationship status.

If our hypothetical student goes out to the movies, the CLS framework describes the cognitive mechanisms necessary to create an episodic memory of the experience. Her longstanding semantic knowledge about the people, places, and things being experienced determines the form of the neural representations projected into the episodic memory system, which is proposed to reside in neuroanatomical structures in the medial temporal lobe. In the hippocampus, information about the features of the experience intermingles with contextual information unique to the event, and associative structures are rapidly created to link these elements to one another. These structures support retrieval of an episodic memory of the experience via a process known as pattern completion. If our undergraduate walked past the movie theater the next day, features of the spatial context could prompt reactivation of the events of the previous night.

An episodic memory, by definition, is associated with a given spatiotemporal context (Tulving, 1983; Schacter, 1987). If this contextual representation is activated, memories for experiences that took place within that context (or similar contexts) become more accessible (Bower, 1972; S. M. Smith, 1988). Research in episodic memory suggests that temporal context has an important influence on behavior in a range of tasks, including free recall (Howard & Kahana, 2002; Howard, Jing, Rao, Provyn, & Datey, 2009), directed forgetting (Sahakyan & Kelley, 2002), interval timing judgments (Shankar & Howard, 2010), reconsolidation (Sederberg, Gershman, Polyn, & Norman, 2011), and retrieval-induced forgetting (Jonker, Seli, & MacLeod, 2013). Theories involving temporal context require the creation of a temporal code that corresponds uniquely to the current moment, but retains some influence of prior events, such that states of the code corresponding to nearby time intervals are representationally similar (Estes, 1950; Yntema & Trask, 1963; Bower, 1972). While CLS describes how to create and retrieve an episodic memory, it does not describe the processes necessary to create a contextual representation, or the cognitive machinery that maintains and manipulates this representation to guide memory search. To understand these processes,

we turn to retrieved-context models, which describe how slowly and rapidly formed associative structures interact to create a contextual representation whose dynamics determine the course of memory search (Howard & Kahana, 2002; Sederberg, Howard, & Kahana, 2008; Polyn et al., 2009). In these models, there are two critical mechanisms that allow context to guide memory search: An integration mechanism which causes the state of the contextual representation to change slowly as an experience unfolds, and an associative mechanism that binds the contextual representation to feature-based representations of the details of the experience (i.e., the people, places, and things comprising the experience). Retrieved-context models have been very successful in describing the behavioral dynamics observed in memory tasks like free recall, and recent work suggests that these models may also provide insight into the neural dynamics observed in laboratory-based memory tests (Polyn & Kahana, 2008; Polyn, Kragel, Morton, McCluey, & Cohen, 2012; Manning, Polyn, Baltuch, Litt, & Kahana, 2011).

We use the retrieved-context framework as a starting point to develop an integrated neural-behavioral theory of human memory in which episodic and semantic structures, while supported by distinct model components, are part of a highly integrated memory system. The slowly learned semantic associations of CLS most closely respond to the pre-experimental associations of retrieved-context models. These are the set of associative structures formed prior to the experimental session in consideration, which are assumed to reside in cortex and change slowly over time (Rao & Howard, 2008). While retrieved-context models assume that pre-experimental associations reflect longstanding knowledge about the studied material, the effects of the structure of prior experience on formation of new episodic memories have not been explored. We modified CMR to have semantic structure similar to that described by Rao and Howard (2008).

In the modified model, the semantic structure of the studied materials is built into the pre-experimental associations linking the feature-based representation of a studied item to the contextual representation. Thus, when an item is studied, these associative structures retrieve a distributed, category-specific representation which is integrated into the contextual representation. As such, the contextual representation becomes a composite of semantic and temporal information. The theory is broadly consistent with the principles of CLS regarding the development of semantic structure, and it inherits the substantial successes of retrieved-context models in accounting for behavioral dynamics in memory tasks. Furthermore, the model can account for both behavioral

and neural dynamics in free-recall experiments where the temporal and semantic structure of study lists is experimentally manipulated.

Neural investigations of semantic and temporal representations

Over the past two decades, there have been great advances in our ability to characterize the representational structure of neural codes. Techniques such as multivariate pattern analysis (MVPA; Norman, Newman, Detre, & Polyn, 2006) and representational similarity analysis (RSA; Kriegeskorte, Mur, & Bandettini, 2008) reveal neural codes that reflect the semantic structure of studied materials, both at the coarse level in which items are assigned to distinct taxonomic categories (Haxby et al., 2001; Polyn, Natu, Cohen, & Norman, 2005; Polyn et al., 2012; Morton et al., 2013), and at a finer level in which items are assigned attribute-based representations that can be used to define the semantic relatedness of any pair of items, regardless of category (Kriegeskorte, Mur, & Bandettini, 2008; Mitchell et al., 2008; Manning, Sperling, Sharan, Rosenberg, & Kahana, 2012). These multivariate analysis techniques have become a major tool of cognitive neuroscientific investigations, allowing researchers to identify and track category-specific neural signals in a variety of psychological tasks (Haynes & Rees, 2005; Polyn et al., 2005; O'Toole, Jiang, Abdi, & Haxby, 2005; Kuhl, Dudukovic, Kahn, & Wagner, 2007; Awipi & Davachi, 2008; Lewis-Peacock & Postle, 2008; Kriegeskorte, Mur, Ruff, et al., 2008; Danker & Anderson, 2010; Kuhl, Rissman, Chun, & Wagner, 2011). Distributed neural signals are thought to reflect an underlying attribute-based cognitive representation that is sensitive to the semantic structure of presented items (Huth, Nishimoto, Vu, & Gallant, 2012). We seek to understand how the structure of distributed semantic representations affects memory search, by building semantic structure into a computational model of memory. The model provides a bridge between measures of distributed neural activity and cognitive theory, providing insight into how distributed neural activity relates to cognitive mechanisms involved in memory encoding and retrieval.

A number of reports make clear the utility of these neural analysis techniques for linking the semantic structure of neural activity to behavioral performance on memory tests. For example, Kuhl, Rissman, and Wagner (2012) showed that the strength or fidelity of category-specific neural activity at the time of encoding predicts whether a given item will be subsequently recalled. Furthermore, Morton et al. (2013) showed that category-specific neural activity at the time of en-

coding also reveals whether a given item will be recalled in sequence with other items from the same category, or in isolation from same-category items. Category-specific neural patterns exhibit behaviorally sensitive dynamics during recall as well, reactivating prior to the vocalization of an item from the corresponding category (Polyn et al., 2005), and rising in strength when multiple items from the same category are recalled in sequence (Morton et al., 2013). Furthermore, Manning et al. (2012) established that neural activity patterns observed during free recall of words reflected the semantic relations between those words as characterized by a corpus-based model of semantic meaning (Landauer & Dumais, 1997).

Other work has established that the degree to which neural patterns change over time also has predictive power regarding behavioral performance on memory tests. Temporally sensitive neural codes are hypothesized to support judgments regarding the memorability (Xue et al., 2010) or temporal organization of past experience (Manns, Howard, & Eichenbaum, 2007; Jenkins & Ranganath, 2010; Ezzyat & Davachi, 2014). Recently, Manning et al. (2011) used electrocorticography (ECoG) to measure oscillatory neural activity during encoding and retrieval in a free-recall task. They compared the pattern of neural activity recorded just before a given item was recalled to the patterns of neural activity recorded as each item was presented. The recall pattern showed the greatest degree of similarity to the original presentation of the about-to-be-recalled item, and showed graded similarity to neighboring items in the list (see also Howard, Viskontas, Shankar, & Fried, 2012). Taken together, these studies suggest that there is a time-sensitive code in the neural system that is reactivated when needed to support memory search through past experience.

Morton et al. (2013) found evidence of a time-sensitive neural code that is also sensitive to stimulus category, suggesting an interaction between temporal and semantic representations. Using scalp electroencephalography (EEG), they measured distributed patterns of category-specific oscillatory activity during a free-recall task. They observed category-specific activity that increased in strength as multiple items from a given category were presented, suggesting that there is a neural representation that integrates information over multiple items. Critically, this integrative neural activity was related to subsequent recall performance: Participants exhibiting faster neural integration also showed more category clustering (grouping together of items from the same category) in their recall sequences. Finally, they found evidence that category-specific activity during recall is stronger during periods of greater category clustering, suggesting that category-specific cues are

used to guide memory search, resulting in category clustering.

Our theory provides a mechanistically explicit description of the cognitive processes that support these behavioral and neural effects. By this theory, when an item is studied, longstanding associative structures allow one to reactivate knowledge about that item in the form of a semantic representation, whose attributes reflect the perceptual and conceptual characteristics of the item. As an experience unfolds, a succession of these semantic representations are elicited by the succession of items that make up the experience. Each time a new semantic representation is activated, it alters a temporal representation. The cognitive system maintaining the temporal representation contains integrative machinery that causes it to become a blend of whatever information it contained previously and the incoming semantic information. Thus, while the temporal code changes slowly and contains a unique representation for each temporal interval, it simultaneously contains a blend of semantic information related to the studied items. This temporal-semantic composite representation becomes more category-specific as multiple items are studied successively from a single category. We explore the consequences of this composite code in a series of simulation analyses using the CMR model. We find that an extended version of CMR can simultaneously account for both recall behavior and the dynamics of semantic neural representations during memory encoding and retrieval.

A computational model of episodic and semantic interactions during memory search

Overview

We introduce a mechanistically explicit cognitive theory designed to bridge between neural and behavioral dynamics during free recall. The theory builds upon the context maintenance and retrieval (CMR) model (Polyn et al., 2009), and examines the consequences of including associative structures that allow a stimulus representation to trigger the retrieval of a distributed representation that reflects the category structure of the stimulus space. The model is representative of a broader class of attribute-based theories, which characterize cognitive processes in terms of multicomponent, distributed representations (Osgood, Suci, & Tannenbaum, 1957; Bower, 1967; Underwood, 1969; Murdock, 1982; Rumelhart, McClelland, & the PDP Research Group, 1986; Nosofsky, 1986;

Shepard, 1987), where a given representation may correspond to one of a number of cognitive constructs, including an item, an event, a plan, or a context. These theories often use the language of linear algebra, in which a given representation is described as a vector of elements, with each element corresponding to an attribute or characteristic of the construct in question (see the Appendix for a description of our model and its dynamics in these terms). The attribute-based framework facilitates using the model to understand neural phenomena, which are naturally described in terms of vectors of numbers corresponding to a collection of neural readings from different topographic locations.

The model captures the major behavioral phenomena observed in the free-recall paradigm, and explains how temporal and semantic organization relate to the dynamics of neural representations recorded during study and memory search. We first describe our implementation of CMR by describing the major modifications to the theory. Generally speaking, the present model is consistent with the broader class of retrieved-context models, and structurally similar to the model variants described by Polyn et al. (2009), but assumes that semantically similar items reactivate similar pre-experimental contextual representations. This allows us to develop neural predictions regarding how the representational structure of context will change as a function of the construction of the study list, and behavioral predictions regarding how how this representational structure will influence the course of memory search. Critical aspects of the model's computational dynamics are examined in a series of Simulation Analyses.

Contextual dynamics and recall organization

Our model builds upon the version of CMR described by Polyn et al. (2009) by assuming that the similarity structure of the pre-experimental contextual states associated with items is influenced by prior experience. Here, we give an informal description of the model; see the Appendix for a mathematical description of the model structures and dynamics, and a description of the different model parameters (Table 1). Figure 1 depicts, in a schematic fashion, the important dynamics of the model as it encodes a stimulus during study, as it retrieves an item during memory search, and as that recalled item triggers contextual reinstatement.

When an item is presented, this causes a representation on the item layer to become activated (Fig. 1, left panel). This activated representation then projects through pre-experimental

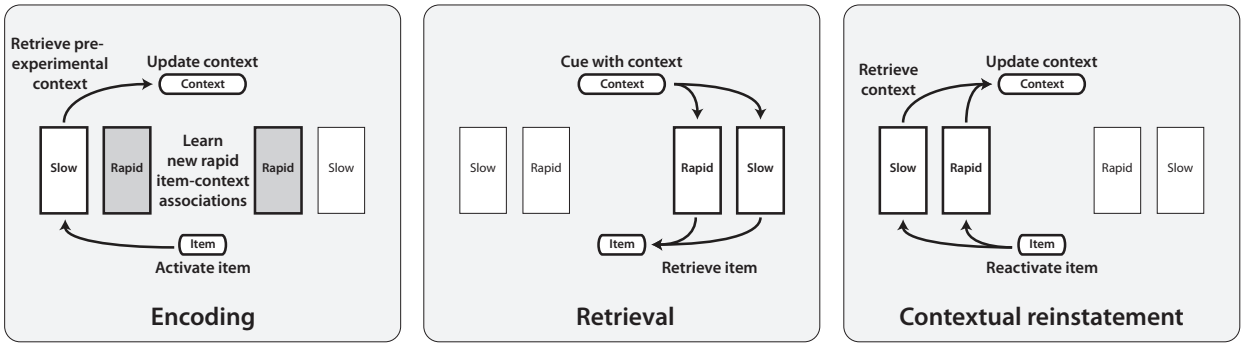


Figure 1. Schematic of model structure and mechanisms for encoding and memory search. An *item* representation and a *context* representation are connected by slowly changing and rapidly changing associations. *Slow* associations were formed before the experiment and hold semantic information. *Rapid* associations are formed during the experiment and are episodic in nature. **Encoding:** A representation of the studied item is activated on the item layer, causing retrieval of pre-experimental context associated with that item; this retrieved context is used to update context. New rapid (episodic) associations are formed between the item and context representations. **Retrieval:** During memory search, context is used as a cue to retrieve an item. Both experimental and pre-experimental associations influence what is retrieved. **Contextual reinstatement:** Recalling an item causes it to retrieve associated pre-experimental and experimental context, which is folded into the context cue. This updated context can be used for another retrieval attempt.

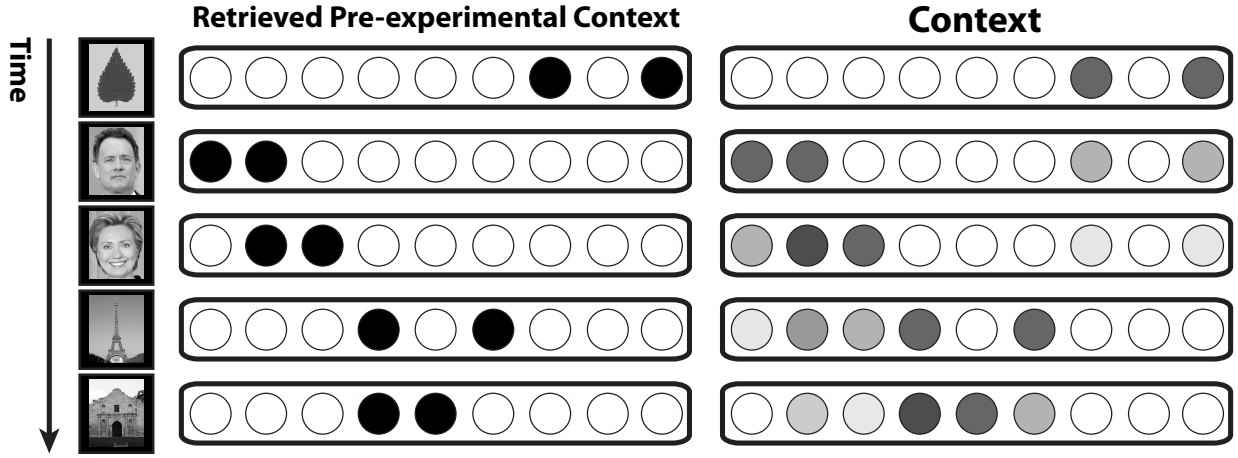


Figure 2. Schematic representation of context evolution during presentation of a list of 5 items from different taxonomic categories. The input representation shows the pre-experimental context associated with each item, which is retrieved when that item is presented. Items from the same category are associated with similar pre-experimental contexts. The context retrieved by each item provides an input to the context representation, which integrates inputs over time. Context reflects a blend of information related to the current item and information from recently presented items.

item-to-context associations to retrieve context previously associated with the item. This retrieved pre-experimental context contains information about the semantic relationships between this item and other studied items. The contextual representation is updated by this incoming information, but also maintains some information corresponding to its previous state. In this way, as a sequence of items is presented, the contextual representation evolves; it becomes a recency-weighted average of the pre-experimental contextual information associated with the studied items. Figure 2 illustrates how temporal context evolves as a series of items associated with distinct pre-experimental contexts is presented. After each item is presented, rapid associative processes (controlled by a Hebbian learning rule) modify the strength of the connections between the item and context layers, associating the active item representation with the active contextual representation. These rapidly formed associations (also referred to here as *experimental associations* to distinguish them from associations formed before the experiment) are critical for episodic memory, as they bind a stimulus to a particular context. Since the state of context changes gradually, neighboring items are associated with similar states of context.

After a list of items has been presented, recall begins (Fig. 1, middle panel). The active state of context is used as a cue to guide memory search. This representation projects through context-to-item associations to activate units on the item layer to varying degrees. The activation state of each item determines how well it fares in a competitive decision process which determines which item will be recalled. The representation of the item that wins the decision process is reactivated on the item layer. This reactivated item representation is projected through item-to-context associations to retrieve a combination of pre-experimental and experimental context, which is integrated into the contextual representation (Fig. 1, right panel). This contextual reinstatement process causes the contextual representation to become more similar to the state of context that was active when the item was studied. Since the contextual representation is a composite of temporal and semantic information, both temporal and semantic neighbors of the just-recalled item will be supported in the next retrieval competition. This process (Fig. 1, center and right panels) repeats until the recall period ends or all studied items have been recalled.

The major behavioral phenomena of free recall can be understood in terms of the interaction between contextual and item representations (Howard & Kahana, 2002; Sederberg et al., 2008; Polyn et al., 2009; Lohnas, Polyn, & Kahana, submitted). The state of the contextual cue de-

termines how well any given item is supported in the decision competition; items associated with contexts similar to the active contextual state are more likely to be remembered. Since context changes gradually, when the free-recall period starts, the contextual representation provides good support to items from the terminal serial positions. Behaviorally, these items are more memorable than those from other serial positions, and tend to be recalled before other items; this is known as the recency effect (Murdock, 1962; Kahana, 1996; Howard, 2004).

The contextual reinstatement process (Fig. 1, right panel) causes there to be sequential dependencies in the recall process. In other words, if the participant recalls a particular item, the identity of that item alters the course of memory search, since it modifies the retrieval cue in a unique way. These sequential dependencies in free recall are referred to as organizational effects; here, we are most concerned with organizational effects reflecting the temporal and category structure of the studied items.

The temporal organization of recall sequences is perhaps most clearly demonstrated by the contiguity effect, whereby participants tend to successively recall items that were presented in adjacent list positions (Kahana, 1996). As described above, the model assumes that items from nearby list positions are associated with similar contextual states. When an item is recalled, its associated context is retrieved and integrated into the contextual representation, updating the retrieval cue to focus memory search on the part of the list when the item was presented. This updated cue provides enhanced support for items that were studied in adjacent list positions to the just-recalled item, giving rise to the contiguity effect (Howard & Kahana, 2002).

When recalling items from categorized lists, participants have a strong tendency to successively recall items from the same category; this is known as category clustering (Bousfield, 1953). When an item from a given category is recalled, the retrieved context contains category-specific information. This category-specific information is integrated into the contextual cue, which makes the contextual cue itself more category specific. This increases the likelihood of the next recall being an item from the same category, causing the model to exhibit category clustering.

When retrieved context contains both category-specific and item-specific information, the context representation can simultaneously support both category organization and temporal organization. Researchers have found that category clustering is increased when category items are blocked together during presentation, compared to when they are spaced apart in the list (Puff,

1966; D'Agostino, 1969). To explain this interaction of temporal and categorical list structure, researchers have proposed that persistent activity (either in the form of short-term priming or a short-term buffer) causes semantically related items to become more strongly associated when they are presented nearby in time (Puff, 1974; Glanzer, 1969). As has been noted, the contextual representation in a retrieved-context model also allows item-specific activity to persist in the cognitive system (Howard, Kahana, & Sederberg, 2008), though prior versions of the model have assumed that different items elicit orthogonal (i.e., structurally unrelated) states of context. In the simulation studies below, we show that these distributed, semantically laden, contextual representations allows the model to account for interactions between temporal and categorical information, without the use of a buffer or priming mechanism.

Neural and cognitive category structure.

When a picture of a celebrity, a landmark, or an object is presented to a participant, this elicits a distributed pattern of neural activity with category-specific features in every lobe of the brain (Haxby et al., 2001; Polyn et al., 2005; Morton et al., 2013). There is evidence that both the perceptual (O'Toole et al., 2005) and conceptual (Mitchell et al., 2008) characteristics of the stimuli affect the similarity structure of the elicited neural patterns. Conceptual similarity in neural representations is supported by the finding that a purely orthographic cue, such as the name of a celebrity, can elicit a neural pattern similar to that elicited by a picture of a celebrity (Kreiman, Koch, & Fried, 2000; Quiroga, Reddy, Kreiman, Koch, & Fried, 2005). In this case, the perceptual characteristics of the stimuli are utterly different, supporting the hypothesis that a component of the neural pattern reflects a high-level representation, which contains information reflecting the conceptual similarity of the items from a given category. In other words, perceptual similarity is not a necessary condition for neural similarity. For many stimulus sets, both perceptual and conceptual similarity will contribute to the neural similarity structure.

In terms of our modeling framework, representational overlap in the feature layer could reflect either perceptual or conceptual similarity, depending upon one's working hypothesis regarding the neuroanatomical region corresponding to this component of the model. For example, similarity in certain regions of visual cortex might reflect perceptual similarity, while similarity in ventral temporal lobe might reflect higher-level conceptual similarity. In preliminary work, we explored a model

variant that contained distributed representations in both the feature and context layers. However, we found that a simpler version of the model, in which distributed semantic representations were restricted to the contextual representation, did just as well explaining the behavioral and neural phenomena characterized in this report. This led to the decision to focus the present work on the simpler form of the model, and explore the interaction of perceptual and conceptual similarity in other work.

In many implementations of retrieved-context models, a simplifying assumption is used, whereby distinct studied items are assigned orthogonal (i.e., non-overlapping) representations, and furthermore, when these items are presented to the model, the contextual information that an item retrieves through the pre-experimental associations is orthogonal to the contextual information retrieved by any other item (Howard & Kahana, 2002; Howard, Kahana, & Wingfield, 2006; Sederberg et al., 2008, 2011). These simplifying assumptions have the consequence that semantic similarity between items will not be reflected in representational similarity in either the item or context representations. Despite these simplifications, this framework has been used to simulate the effect of semantic similarity on memory search, by building latent semantic structure into the associative structures of the model.

For example, the version of CMR described by Polyn et al. (2009) used a corpus-based model of semantic similarity (LSA: Landauer & Dumais, 1997) to create semantic structure. Each item was assigned an orthogonal representation, which, when studied, retrieved a contextual representation orthogonal to that retrieved by any other item (as if each item caused a person to revive past knowledge that was unrelated to the past knowledge revived by any other studied item). Information about semantic similarity was built into the pre-existing associations connecting the orthogonal context representations back to the feature layer. In an experiment with strong category structure, if the participant studied, for example, the celebrity Tom Hanks, this item would prompt the retrieval of pre-experimental information specific only to Tom Hanks. This retrieved context would be integrated into the contextual representation. During memory search, if this idiosyncratic Hanks-related context is active, the latent semantic associations connecting the contextual layer to the feature layer would support recall of the semantic associates of Tom Hanks (e.g., Meg Ryan, or John Candy). This causes the model to exhibit category clustering, whereby semantically related items (i.e., from the same category) tend to be recalled successively. This version of CMR can

account for semantic and temporal organizational effects in behavior, but it would be unsuitable as a model of neural dynamics, as neither the stimulus or contextual representations would reflect the category structure of the studied material.

A promising alternative was proposed by Howard and colleagues, where semantic information can be embedded in the latent associative structures of the model, and also exhibit itself in the contextual representations elicited by presented items. In this approach, semantic structure is learned through repeated experiences with items (Rao & Howard, 2008; Howard, Shankar, & Jagadisan, 2011). Word representations are orthogonal at the feature layer of the model, and initially elicit orthogonal contextual representations, as above. The model is sequentially presented pairs of synonyms, where a given word can appear in more than one pair: e.g., bread and butter; butter and knife. The model is shown to have created useful semantic associative structures; after training, the contextual representations elicited by the words reflects the higher order semantic structure of the full set of words in that, i.e., the contextual representation elicited by bread will be similar to that elicited by knife, despite the fact that they were never presented in sequence. A similar mechanism (though with a quite different implementation) causes semantic structure to emerge in the word representations of the BEAGLE model (Jones & Mewhort, 2007). The Rao and Howard (2008) model provides a suitable starting point for a neurocognitive model of episodic-semantic interactions in memory search, as it makes predictions about how the similarity structure of neural representations during encoding should be influenced by the semantic similarity of presented items, and how neural similarity should relate to the dynamics of memory search.

Rather than train our model to derive the semantic structure of the studied items from experience, we create pre-experimental associative structures that allow orthogonal item representations to retrieve contextual representations whose structure reflects the semantic similarity of the studied items. This allows us to build upon the work of Howard and colleagues in creating a model that utilizes distributed representations with semantic structure, while focusing on the question of how these semantic structures interact with episodic structures to produce the behavioral and neural phenomena observed in free-recall tasks.

Précis

In *Simulation Analysis I*, we apply the model to a classic study in which the temporal and semantic structure of the study list was simultaneously manipulated using categorized lists (Puff, 1966). We find that the contextual integration mechanism in the model allows it to capture interactions between temporal and semantic influences on recall behavior. Semantic information in context accumulates when semantically related items are presented near to each other, resulting in greater semantic organization during memory search.

Simulation Analyses II through IV examine behavioral and neural data from a recent free recall study with categorized stimuli (Morton et al., 2013), where neural oscillatory activity was measured using scalp electroencephalography (EEG). A number of novel analyses are reported that demonstrate the predictive power of the hypothesis linking category-specific neural representational structure to cognitive representational structure.

In *Simulation Analysis II*, we examine the divergent predictions of three model variants which specify how category membership of studied items influences cognitive dynamics at encoding. We demonstrate that category-level behavioral differences in recall performance and category clustering, along with category-level differences in classifier performance during encoding, constrain the relative viability of the three model variants. The best-fitting model proposes that the stimulus categories vary in the inter-item similarity of the contexts they were associated with prior to the experiment, resulting in category-level differences in neural similarity, recall organization, and recall performance.

In *Simulation Analysis III*, we test the best-fitting model's predictions for how the strength of category-specific activity should fluctuate during encoding and retrieval by comparing the contextual representation in the model to distributed patterns of neural activity. As predicted by the model, the category-specificity of the neural representation elicited by an item's presentation predicts whether that item will be remembered as part of a category cluster. We also show that the growth and decay of category-specific neural activity during study is consistent with the predictions of the model. Finally, we find that category-specific neural activity during free recall increases in strength during clusters of same-category recalls, consistent with the dynamics of the contextual retrieval cue used by the model to guide memory search.

Finally, *Simulation Analysis IV* demonstrates the ability of this framework to provide insight into participant-level differences in cognitive structure and dynamics. By creating a family of models tuned to account for the behavior of individual participants, we show that individual variability in recall performance and organization can be explained in terms of individual differences in category structure and contextual dynamics. Furthermore, although only behavioral observations are used to determine model parameters for each individual, we find that the family of models successfully predicts individual differences at the neural level, both in terms of category discriminability during encoding, and in terms of category integration as a set of same-category items are studied in succession.

Simulation Analysis I: Category clustering and spacing effects

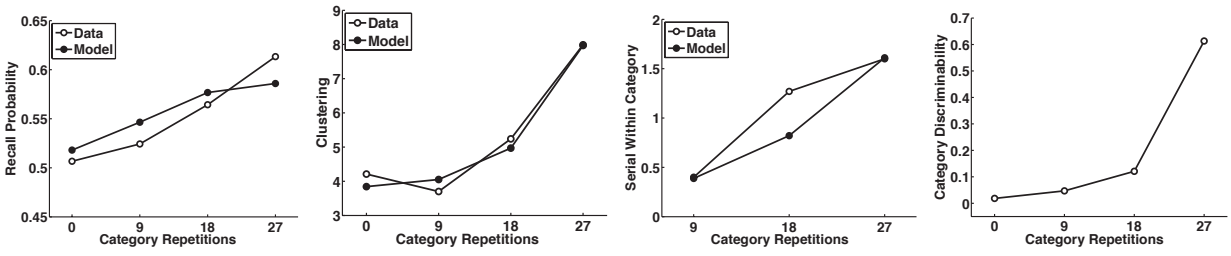
CMR suggests that the order in which one studies a set of materials of varied semantic similarity has important consequences for the subsequent memory of that material. When a study list is composed of groups of items from a number of taxonomic categories, memory performance is markedly better when items from a given category are presented in a block, as compared to when they are scattered about the list (Dallet, 1964; D'Agostino, 1969; Cofer, Bruce, & Reicher, 1966; Puff, 1966, 1974). In addition to affecting the number of recalled items, the stimulus-list organization (SLO) also influences the organization with which these items are pulled from memory, with blocked presentation of items from a particular category leading to increased category clustering. An important factor not taken into account in many of the classic studies exploring SLO effects is that strong temporal organization can greatly inflate estimates of category organization. If items are presented in adjacent list positions, then list position and category are confounded: Even if there were no trace of the category structure of the items in the cognitive system, a metric of category clustering would show above-chance organization. There has been much debate regarding the appropriate baseline for inferring a behavioral effect of category organization (Roenker, Thompson, & Brown, 1971), but even when temporal organization is taken into account, it is clear that category/semantic information has a strong effect on recall organization (Puff, 1966; Polyn et al., 2009).

In the present theory, the order in which items are recalled is determined by the dynamics of a contextual cue that is a recency-weighted composite of the semantically laden contextual

representations retrieved by representations of the studied items. The fact that the same associative structures are involved in both temporal and category organization places strong constraint on the patterns of behavior the model can exhibit. For example, there is no single model process that can alter the level of category organization without also affecting temporal organization.

In prior work, theorists have attributed the behavioral advantage on blocked categorized lists to enhanced discovery of semantic relations between items when they co-occur in a short-term store (Anderson, 1972; Glanzer, 1969), or to short-term priming, which can build when related items are presented near to each other (Puff, 1966; Kimball, Bjork, Bjork, & Smith, 2008; Puff, 1974). In contrast, in CMR, the advantage of studying same-category items in succession comes from the integrative mechanism that drives contextual evolution. We used a model of category structure in which each item representation is created by blending item-specific information with the representation of a category prototype (Hintzman, 1986). When multiple items from the same category are presented in succession, context becomes a weighted average of these items (see Fig. 2 for an example). As a result, the idiosyncratic characteristics of the items cancel each other out, causing the contextual representation to come to resemble the category prototype. This prototypical representation is a good cue for all of the items from the category. Thus, if this prototypical context is used as a cue during recall, it will cause an increase in the overall strength of category clustering. When categorized items are interspersed in the list, contextual integration causes blending of the retrieved context corresponding to different categories. In this case, context never becomes as representative of any one category, resulting in decreased category clustering during later recall.

We examined whether the context integration and cuing mechanisms proposed by CMR can account for SLO effects in free recall of categorized lists. We chose to focus on the study reported by Puff (1966), which parametrically manipulated the amount of SLO, and measured both temporal and category organization. In addition to overall recall and category clustering, Puff (1966) also reported the mean number of serial transitions between items in the same category (e.g., after recalling an item from category A in serial position N, the next recall is also from category A, and serial position N+1). This statistic was meant to estimate the effect of serial organization on category clustering. This is not a perfect measure of temporal organization, since the effects of temporal contiguity extend beyond just an item and its successor. However, the measure has some validity, as recall transitions of +1 lag (an item and its successor) are the most frequently



(a) (b) (c) (d)

Figure 3. (**a**) In both the data and the best-fitting model, overall recall increased as the number of category repetitions in the stimulus list increased. (**b**) Category clustering, measured as the number of output category repetitions corrected for chance, was greater when there were more category repetitions in the input order. (**c**) The number of serial transitions between items in the same category increased as stimulus-list organization increased, in both the data and the best-fitting model. (**d**) Due to integration of category information over time, the category discriminability of context increases as the number of category repetitions in the input list is increased. As a result, category clustering increases as a function of the number of input repetitions.

observed transition in free recall (Kahana, 1996), and account for much of the variability due to the contiguity effect.

Puff (1966) presented participants with lists of 30 words, with 10 words drawn from each of 3 taxonomic categories taken from the Cohen, Bousfield, and Whitmarsh (1957) norms (animals, vegetables, and professions). The SLO was manipulated between subjects; the number of category repetitions (C-Reps) in the stimulus presentation order was either 0, 9, 18, or 27. We simulated the pre-experimental semantic structure of the study materials by first generating a prototype pattern for each category, then adding an item-specific pattern (with features randomly drawn from a normal distribution) to create each exemplar. In order to make the model's behavior easier to interpret, the category prototypes were orthogonal to one another. At the beginning of the list, context was set to a random normally distributed vector normalized to have length 1. All four SLO conditions were fit using a single set of parameters; best-fitting parameters were found by minimizing RMSD (see Appendix for details of the fitting procedure). We allowed 7 model parameters to vary freely, fixing other parameters to the best-fitting values from a fit to the Murdock (1962) free recall dataset (details of this analysis and the values of fixed parameters are given in the Appendix: *Serial Position, List Length, and Contiguity Effects*).

We found that, using a single set of parameters, the model provides a good simultaneous fit to recall probability, category clustering, and temporal clustering, as a function of SLO (RMSD =

0.4013; Fig. 3a–c). Because parameters are the same for each condition, these changes in model recall behavior must result from the differences in list structure between conditions, which affect the evolution of temporal context during encoding. Since the context active during presentation of the list is later retrieved and used to guide subsequent search of memory, these changes in list structure alter retrieval dynamics. Category clustering is increased when the retrieved-context cue is strongly category-specific (when categorized items are blocked), and falls when the cue is a blend of categories (when categorized items are interspersed).

We defined a *category discriminability* metric, to allow us to describe how differences in list structure across conditions affected the category-specificity of the encoding context representation. Higher values of this metric indicate context more purely reflecting one category, and lower values indicate a blend of categories. The context associated with each item was defined as the state of context observed in the model after updating with retrieved pre-experimental context corresponding to that item. For each item, we calculated the mean cosine similarity between its context and the states of context of other items on the list from that category; we refer to this as the *same-category similarity*. We also calculated *different-category similarity* as the mean cosine similarity between each item's context and the contexts of items on the list from different categories. We defined the category discriminability for each item as the difference between same-category similarity and different-category similarity (see Polyn et al., 2012 for a similar approach). A category discriminability of 1 indicates that context is exactly the same for each item in a given category, and orthogonal to context for the other categories. Lower values indicate that context reflects a blend of categories. For each SLO condition, category discriminability was averaged over 100 replications of the Puff (1966) experiment.

We found that category discriminability of temporal context increases directly with the number of category repetitions in the input list (Fig. 3d). In the model, context represents a recency-weighted average of recently presented items. When items from the same category are presented in blocked order (i.e. 27 C-Reps), during each category block context comes to strongly represent the current category. In the spaced conditions (0, 9, and 18 C-Reps), there are more transitions between categories, and thus more items associated with a contextual representation reflecting a blend of category contexts. When SLO increases, the context retrieved by each recalled item provides a good cue for other items from that category, resulting in increases in category clustering

(Fig. 3b) and recall (Fig. 3a). The model also simultaneously accounts for the serial organization observed in the data (Fig. 3c).

We find that our extended model can account for effects of stimulus-list organization on category clustering observed by Puff (1966); in a separate simulation, we verified that the changes to the model do not affect its ability to fit benchmark results observed in a standard free-recall paradigm (Murdock, 1962). A previously reported version of CMR was able to account for list-length, serial position, and temporal clustering effects in the Murdock (1962) dataset (Polyn et al., 2009). We find that the extensions to the model described here do not affect the ability of the model to account for these effects; see the Appendix (*Serial Position, List Length, and Contiguity Effects*) for details.

The model fits presented in Figure 3 do a good job explaining the variability in several dependent measures across different list structures. However, a closer look at model dynamics reveals some underlying tensions that may help to drive future development of the model. Wide regions of model parameter space cause the model to produce category clustering behavior in line with the experimental results: As items from the same category are presented in closer temporal proximity, category clustering increases. In the Puff (1966) dataset, as well as in a number of similar prior studies, recall has also been found to increase when categories are blocked together during presentation (Puff, 1974). However, under many parameter sets, the model exhibits a *decrease* in overall recall performance under blocked presentation.

The model dynamics giving rise to this tension are clear. When the items from the three categories are associated with very distinct contextual representations, the contextual retrieval cue tends to become more and more focused on the most recently retrieved category. When the contextual representation becomes highly category-specific during search, items from the most recently retrieved category are well supported, at the expense of items from any other categories. This effect is strongest in the fully blocked (27 C-Reps) condition, where temporal clustering does not provide an effective means to bridge between categories.

Prior work suggests that this problem could be effectively resolved through the addition of an executive process that detects when a particular category is becoming depleted of memories, prompting a strategic shift from a local search of memory to a more global search (Hills, Jones, & Todd, 2012; Raaijmakers & Shiffrin, 1980). In the model, parameters affecting context updating

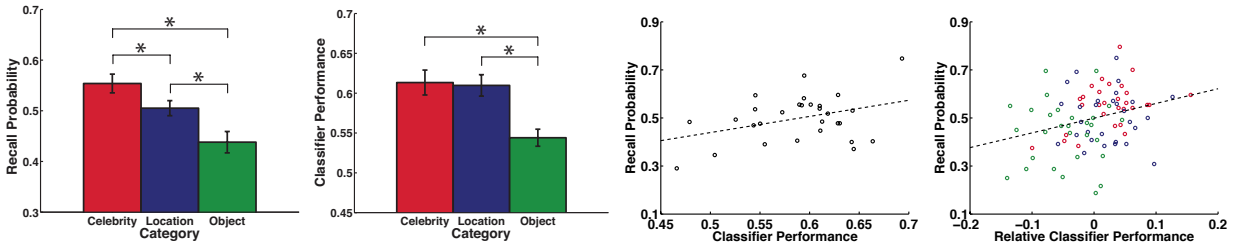
and the retrieval competition are assumed to be fixed throughout memory search; an expanded model could instead allow these parameters to be strategically modified during search (cf. Logan & Gordon, 2001), to allow switching between local and global search of memory. During retrieval, the model's λ parameter controls the amount of contextual support an item must have for it to compete effectively with other items in the retrieval competition. When λ is high, only items strongly cued by the current state of context will be retrieved, resulting in relatively local memory search. When λ is low, relatively weakly supported items still have a chance to be recalled, resulting in more global search. In order to improve the efficiency of recall from blocked categorized lists, λ could be strategically modified to be higher when starting recall from a given category (causing more local recall), and lower when the current category is exhausted (causing a shift to more global recall, and a better chance of discovering a new category). This approach to memory search could be useful not just in recall from blocked categorized lists, but more generally for facilitating recall from any targeted memory set with a "patchy" structure, i.e. where there are multiple groups of targeted items, and each group is associated with a relatively distinct temporal-semantic context. A related dynamic memory targeting approach was used in a model of memory search developed by Becker and Lim (2003) to explain deficits in recall of categorized lists exhibited by patients with frontal damage.

While Puff (1966) characterized recall in terms of category organization, serial organization, and recall performance, the above discussion suggests that the model is still potentially under-constrained by these data. Computational models of memory could not move forward without constraint through the characterization of phenomena that challenge the model. In the rest of the Simulation Analyses, we describe an experiment that manipulates temporal and category structure within list. We characterize recall performance in terms of more detailed organizational metrics, and characterize the neural activity patterns observed during both encoding and retrieval. The model is able to simultaneously account for these neural and behavioral phenomena, and allows us to relate neural measures with the cognitive constructs described by the model. This unified neurocognitive framework allows us to develop more precise hypotheses regarding the cognitive mechanisms involved in memory search, allowing us to explain how neural signal relates to behavioral phenomena.

Simulation Analysis II: Encoding processes involved in semantic organization

Morton et al. (2013) manipulated the temporal and semantic structure of study lists in a free-recall paradigm, as participants studied items from three distinct taxonomic categories (celebrities, landmarks, and common objects). They used multivariate pattern analysis (MVPA) to characterize category-specific patterns of oscillatory activity measured with scalp EEG. They observed category-specific patterns of oscillatory activity during both encoding and retrieval that were sensitive to variability in the strength of category clustering during free recall. In the next section, we present new analyses of these data, which reveal reliable category differences in overall recall and the strength of category clustering, with celebrities being best remembered and most reliably clustered, followed by landmarks, and then objects. Neural classification performance shows a similar ordering, with celebrity items better classified than landmarks, which in turn are better classified than objects. We use the CMR framework to specify three hypotheses regarding the cognitive mechanisms giving rise to these category differences.

The first model proposes that items from the different categories engage associative processes to different degrees, with celebrity items becoming most strongly associated to the contextual representation, followed by landmarks, followed by objects. We demonstrate that this model variant is able to explain category-level differences in recall and clustering, but is unable to account for the neural differences in classifiability across the three categories. The second model proposes that items from the different categories trigger different amounts of contextual integration. As with the first model, this model variant is able to explain category differences in recall and clustering, but not the neural differences. The third model proposes that items from the three categories are not treated differently during study, but that differences in the representational structure of each category give rise to the behavioral and neural differences. This model is able to account for both behavioral and neural differences between categories. In Simulation Analysis III we examine the predictions of this model in terms of contextual dynamics, and in Simulation Analysis IV, we show that by customizing model parameters to fit the behavior of individual participants, we can explain individual differences in both neural activity patterns and recall behavior.



(a) (b) (c) (d)

Figure 4. (**a**) Overall recall performance on mixed lists, as a function of category. (**b**) Performance of a pattern classifier trained to discriminate between the stimulus categories based on patterns of oscillatory power recorded during stimulus presentation. Performance is shown as the fraction of stimuli classified correctly in a cross-validation procedure. (**c**) Classifier performance during encoding predicts individual differences in recall performance ($r = 0.371$, $p = 0.048$). (**d**) Classifier performance during encoding predicts subsequent recall performance, at the level of categories. Recall probability is shown for each participant and stimulus category, as a function of classification performance for that category, relative to the participant's mean classifier performance ($r = 0.291$, $p = 0.0062$).

Categorized free recall experiment

Twenty-nine participants performed a free recall task while scalp EEG was recorded. Stimuli were photographs of famous landmarks, celebrity faces, and common objects, with the name of the stimulus presented in text above the picture. Participants studied 48 lists, each of which was immediately followed by free recall. Each list was composed of 24 stimuli. Lists were either all drawn from the same category (pure; 18 lists per participant) or contained 8 items from each of the 3 categories (mixed; 30 lists per participant). Here, we focus on the mixed lists. In the mixed lists, items were presented in trains of same-category items, with each train containing 2–6 items. The order of category trains was pseudorandom, with the constraints that all categories appeared in each set of 3 trains, and that adjacent trains did not contain the same category. Each item was presented for 3.5 s, during which participants made a 4-point semantic judgment that was specific to the category of the stimulus. An interstimulus interval of 0.8–1.2 s separated each item on the list. After presentation of the last stimulus, participants were given 90 s to recall items from the list in any order.

Morton et al. (2013) examined oscillatory power over the scalp at a range of frequencies from 2 to 100 Hz. Using pattern classification techniques, they found that oscillatory power during the encoding period could be used to decode the category of the stimulus currently being studied, with

accuracy of 0.589 (chance is 1/3). They also found that activity observed just prior to (3 s to 0.5 s before) vocalization of recalled items could be used to predict the category of the item about to be recalled, even when epochs that overlapped with previous recalls were excluded (Morton et al., 2013). Furthermore, they found that classification accuracy during both encoding and retrieval was related to item-level and participant-level variations in category clustering. We first focus on simulating behavior in this experiment, and then examine whether CMR can account for the neural effects observed.

As described by Morton et al. (2013), there was substantial category clustering on the mixed lists, as measured by the semantic list-based clustering metric ($LBC_{sem} = 3.66$ SEM 0.25; Stricker, Brown, Wixted, Baldo, & Delis, 2002). Given that the lists were organized in trains of items from the same category, it is important to take temporal contiguity into account, since a tendency for participants to make transitions between adjacent items will also tend to increase the number of transitions between same-category items. Morton et al. (2013) used a relabeling procedure to estimate the amount of category clustering predicted due to temporal organization, and found that the category clustering observed on the mixed lists was greater than expected based on temporal clustering and serial position effects, as estimated based on recall behavior on the pure lists ($LBC_{sem} = 0.808$, SD 0.061, $p < 0.0002$; cf. Polyn et al., 2009).

Further analysis not reported by Morton et al. (2013) showed that recall performance varied markedly by stimulus category (Fig. 4a). Recall was significantly greater for celebrities, compared to locations ($t(28) = 2.91$, $p = 0.007$) and objects ($t(28) = 5.88$, $p = 3x10^{-6}$). Recall was also significantly greater for locations, compared to objects ($t(28) = 3.69$, $p = 0.001$). We also examined whether the categories varied in the degree to which they were clustered during recall. We calculated the probability of making a within-category transition (vs. a transition between categories), conditional on the category of the item just recalled (Fig. 6a). The conditional probability of within-category transitions was greater for celebrities than objects ($t(28) = 2.17$, $p = 0.039$). While the numerical value for landmark clustering fell between celebrities and objects, the other pairwise comparisons were not significant ($p > 0.05$).

We found that the category differences in recall behavior were mirrored by differences in classification performance (Fig. 4b). Celebrities were classified significantly more accurately than objects ($t(28) = 4.20$, $p = 0.0002$), and locations were classified significantly more accurately than

objects ($t(28) = 4.30$, $p = 0.0002$). Furthermore, individual variability in classification performance was positively correlated with recall performance (Fig. 4c; $r = 0.371$, $p = 0.048$). In other words, participants with neural category representations that were more distinct from one another tended to recall more of the studied material overall.

We ran a second analysis to determine whether, within participant, items from a particular category were better remembered if the neural category representations associated with items from that category were more distinct. For a given participant, we calculated how well items from a given category were classified relative to mean classification performance across the three categories; this gave us three numbers describing the relative classifier performance for each category. The relative classification score for a given category (Fig. 4d) shows a reliable correlation with recall of items from that category ($r = 0.291$, $p = 0.0062$), suggesting that neural category discriminability is predictive of recall performance both at the participant and category levels. A closer examination of these individual differences, in terms of model dynamics, is presented in Simulation Analysis IV.

Models of category influence during encoding

The finding that neural discriminability is related to recall performance is consistent with previous neurorecording studies (Kuhl et al., 2012). It also provides neural validation for classic behavioral studies suggesting that representational similarity at encoding has important consequences for subsequent recall (Deese, 1959a; Cohen, 1963). As mentioned above, we used CMR to propose three hypotheses regarding the cognitive mechanisms underlying these empirical effects.

Each model variant was first fit to a number of measures of behavioral performance, including serial position curves (Fig. 5a), probability of first recall by serial position (Fig. 5e), category clustering (Fig. 6a), and temporal clustering, separately for each stimulus category, to minimize χ^2 error (see Appendix for details of the fitting procedure). We compared the ability of each model variant to account for the behavioral data while minimizing model complexity using the Bayesian information criterion (BIC; Schwarz, 1978).

We propose that a number of mechanisms give rise to category-specific differences in neural signal. As mentioned above (see *Neural and cognitive category structure*), we expect that similarity in the perceptual characteristics of stimuli from each category support neural discriminability. We return to this source of variability in later sections. Some of the category differences in neural

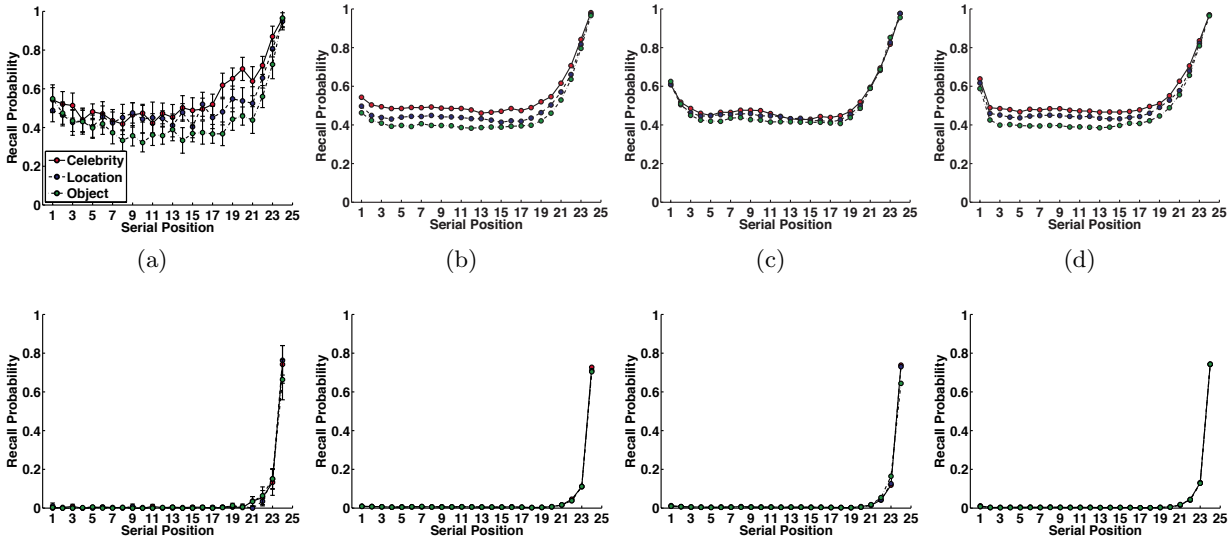


Figure 5. (**a**) Recall probability as a function of serial position and stimulus category. (**b**) Distributed CMR simulation where L^{CF} is free to vary between categories. (**c**) Distributed CMR simulation where β_{enc} is free to vary between categories. (**d**) Distributed CMR simulation where within-category similarity is free to vary between categories. (**e**) Probability of first recall. (**f**) Distributed CMR simulation where L^{CF} is free to vary between categories. (**g**) Distributed CMR simulation with category-specific β_{enc}. (**h**) Distributed CMR simulation with category-specific within-category similarity.

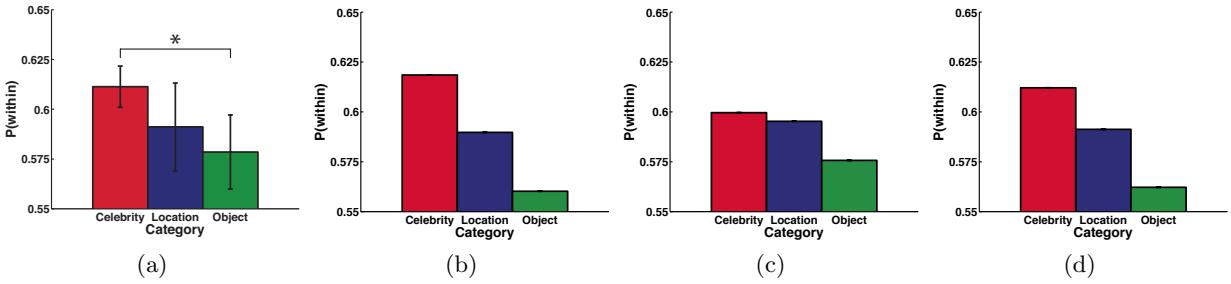
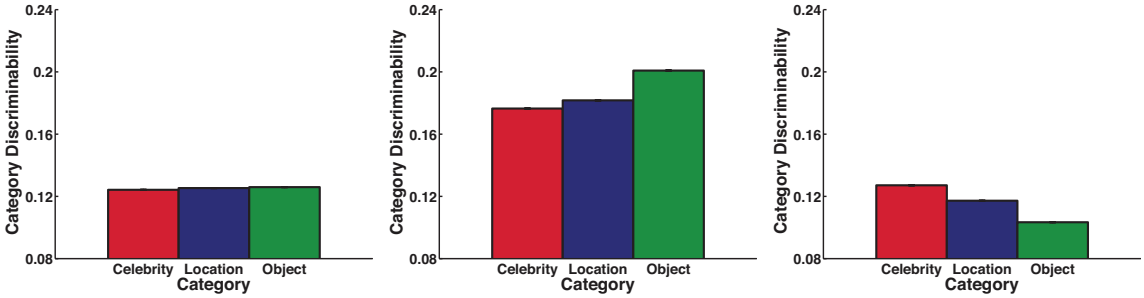


Figure 6. (**a**) Probability of making a within-category transition, conditional on the just-recalled category. (**b**) Distributed CMR simulation where L^{CF} is free to vary between categories. (**c**) Distributed CMR simulation with category-specific β_{enc}. (**d**) Distributed CMR simulation with category-specific within-category similarity.



(a) (b) (c)

Figure 7. Prototypicality (cosine similarity to the category prototype representation) of states of context, for different variants of CMR. (**a**) Model variant where learning rate on M^{CF} is free to vary between categories. The learning rate manipulation has no effect on item prototypicality. (**b**) Model variant where within-category similarity is the same for each category, but integration rate (β_{enc}) varies between categories. Integration rate is highest for celebrities, followed by locations, then objects. For objects, category-specific activity is integrated over a longer time window, resulting in higher average activation. (**c**) Model variant where within-category similarity is free to vary by category. Celebrity item representations have higher within-category similarity than locations, and objects have the lowest within-category similarity.

discriminability may also be due to non-cognitive sources of variance, such as differences in signal strength given the different locations of brain regions with activity selective for the different categories. However, this type of variability in neural category discriminability would be unable to explain the correlation between classifier performance and recall probability observed both at the level of participants and at the level of individual categories (Fig. 4c,d). Here, we focus on the proposal of Morton et al. (2013) that a significant proportion of the variance in neural discriminability is related to representational similarity in the context representation of the model. To facilitate comparison of model predictions with neural data, we compared the category discriminability of simulated states of context (see Simulation Analysis I) with the neural category discriminability taken from MVPA analysis of the scalp EEG data. For each model variant, the average category discriminability in context during encoding was calculated for each category, averaged over 40 simulations of the categorized free recall experiment. We then examined whether each model variant's predictions for category discriminability provided a qualitative fit to the neural measure of category discriminability from the scalp EEG data.

Learning rate.

Given the reliable differences in recall performance across the three categories, a simple

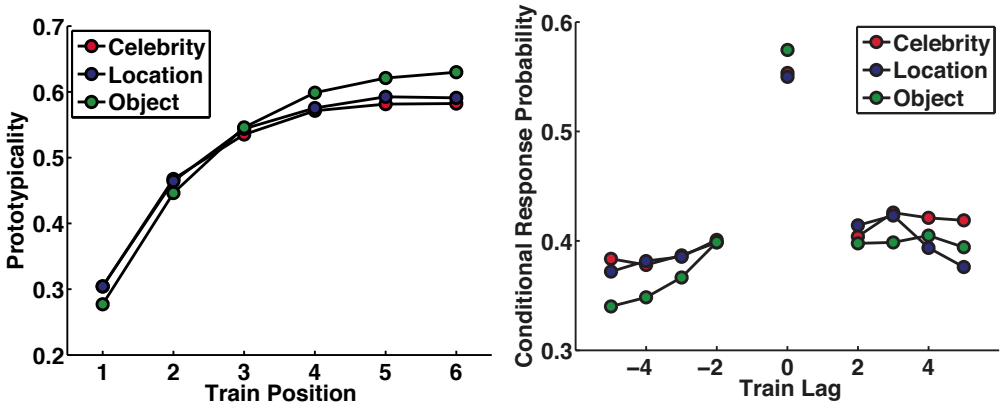
explanation of category-level differences is that items from each category vary in the strength with which they are encoded. In the model, episodic learning involves the rapid formation of item-to-context and context-to-item associations. We followed previous implementations of variability in associative strength during encoding (e.g., accounting for the primacy effect, Sederberg et al., 2008; Polyn et al., 2009), by allowing the strength of experimental context-to-item associations (L^{CF}) to vary between categories. We fixed the strength of pre-experimental associations, determined by γ^{CF}, to be the same for each category.

This model variant provided a good fit to category differences in both overall recall (Fig. 5b) and clustering (Fig. 6b). The ability of the model to capture variability in both clustering and recall by modulating a single parameter (context-to-item learning rate) is consistent with a wealth of experimental work suggesting a close relation between recall performance and organization (Cohen, 1963; Dallet, 1964; Cofer et al., 1966; Tulving & Pearlstone, 1966; Puff, 1974).

However, this model variant fails to predict differences in category discriminability during encoding (Fig. 7a). The states of context during encoding are determined by pre-experimental associations between presented items and context. Since each category is associated with a separate category prototype, from which exemplars are derived, each category is discriminable from each other, but the discriminability does not vary with category. This case provides a simple example of how multivariate pattern analysis can provide constraint on models of cognition. This analysis does not rule out the possibility that learning rate varies by category, but it suggests that learning rate variability alone is not enough to explain both the behavioral and neural findings.

Integration.

The second model variant proposes that neural and behavioral category differences arise due to differences in contextual dynamics associated with each set of stimuli. At any moment, context is a blend of information, a weighted average of the retrieved contextual states associated with the past several studied items. As each new item is studied, it retrieves pre-experimental context, which is integrated into the current state of context. Model parameter β_{enc} controls this integration process; higher values of β_{enc} increase the rate of integration, causing the most recent item to have a larger weight in the weighted average. If the retrieved context contains category-specific information (as in our model), changes in β_{enc} will affect the category discriminability of



(a) (b)

Figure 8. Category activation in context and behavior in the Integration model variant. (**a**) Similarity between the state of context and the category prototype for the current presented item, as a function of train position. Because context updating rate is low for objects, prototypicality at train position 1 is lowest for objects. However, at later train positions during object trains, context contains a recency-weighted average over a number of object exemplars, causing context to be more prototypical on average. (**b**) Probability of making transitions between different trains of same-category items, for each of the three categories. Results are shown for the model variant where context updating rate was allowed to vary by category. Conditional response probabilities are shown for within-category transitions, as a function of distance in the list in terms of train number, conditional on at least one item from that train being available, and conditional on the transition being between items of the same category. For celebrities and locations, response probabilities do not change much as a function of train lag, while for objects, probability decreases substantially with increasing train lag.

encoding context. During recall, the study-period context is reactivated and used as a retrieval cue; therefore, differences in β_{enc} during encoding will influence subsequent recall behavior. Here, we allowed β_{enc} to have a distinct value for each of the three categories, and examined whether this model could simultaneously account for category differences in neural discriminability, recall performance, and category clustering.

The best-fitting model assigned the highest integration rate to celebrities, with a slightly lower rate for locations, and the lowest rate for objects. Integration rate is directly related to the magnitude of the recency effect; increased integration will cause the final item to be more prominently represented in the initial contextual retrieval cue, which will increase the likelihood of the final item being the first item recalled (Polyn et al., 2012). In the experimental data, the size of the recency effect did not vary based on the category of the terminal item (Fig. 5g), which constrained the degree to which the different categories could be assigned different values of

β_{enc}. This led to the model underpredicting the magnitude of category-related behavioral effects: celebrities were both better recalled (Fig. 5c) and more strongly clustered (Fig. 6c) than other categories, but these differences were smaller than was observed in the actual data.

This model variant also made an incorrect prediction about overall category discriminability in context: It predicts that objects are the most discriminable, followed by locations, then celebrities (Fig. 7b). This result is counterintuitive; if each celebrity pushes more information into context, why would objects be more discriminable? The answer lies in how context changes over presentation of multiple items. At any given point in a list of stimuli, context contains a recency-weighted average of the stimuli seen so far. Lower β_{enc} corresponds to a larger window of items contributing substantially to this average. When the representations of multiple exemplars from a category are averaged together, the resulting representation will be more similar to the category prototype, and more discriminable from other categories.

We carried out a follow-up simulation to examine the effect of the value of β_{enc} on category discriminability as a function of train position (Fig. 8a). For simplicity, context was initialized to a random state that was orthogonal to all categories, followed by presentation of 6 items drawn from the same category. This simulation was repeated 1000 times for each of the three categories. We then calculated the mean cosine similarity between the category prototype and each of the 6 states of context associated with a given train; we refer to this measure as *prototypicality*. At train position 1, context was least prototypical for the objects; however, context was the most prototypical for objects at later train positions (Fig. 8a). This follow-up simulation makes it clear that the predicted classification advantage for objects arises from their high discriminability at later train positions.

Why does the high celebrity β_{enc} for the best-fitting model result in greater clustering and recall? Given that context during encoding is later retrieved and used as a memory cue, increasing the prototypicality of context will tend to increase category clustering, all else being equal. However, there is another factor at play: temporal organization. The cue strength of a given item, for a given context cue, depends on the overlap between the cue and the states of context associated with the item. Context may overlap due to similarity in either pre-experimental or experimental context. Through a series of simulations, Polyn et al. (2009) demonstrated that varying β_{enc} during encoding can alter the similarity structure of the experimental contexts associated with studied items. They

proposed that switching between encoding tasks during encoding results in a transient disruption to context. In their version of CMR, this disruption occurred at each transition between trains of items studied with the same encoding task. To examine potential effects of disruption, they introduced a measure analogous to the conditional response probability by lag (lag-CRP) measure of Kahana (1996), but examining transitions at the level of trains of same-task items. This measure determines the probability of making a transition of a given train lag (the number of trains between the just-recalled item and the next recalled item), conditional on at least one item from that train being available. Using this train-CRP analysis, they found that task-shift-related context disruption had the effect of isolating the trains from one another during recall, making transitions between different trains less likely.

In order to examine disruption in the present model variant, we calculated the train-CRP, examining only within-category transitions. We calculated these transition probabilities separately for each of the three categories (Fig. 8b). Because β_{enc} is allowed to vary between categories, a similar disruption effect is observed to that examined by Polyn et al. (2009) in their simulations. For objects, the probability of making a within-category transition is actually higher than the other categories for train lag 0 (that is, a transition within the same train). However, the probability of making a within-category transition falls off quickly with increasing train lag for objects, but not for celebrities or locations, resulting in lower overall clustering for objects. Celebrity and location trains cause relatively large changes in temporal context, effectively isolating object trains from one another and making long-distance transitions between objects less likely, which lowers the overall probability of making a transition between two objects.

Allowing β_{enc} to vary by category results in a qualitative fit to the behavioral data, though the magnitude of category differences is underpredicted. However, the model's predictions for the discriminability of the different categories in context are incorrect, suggesting that category differences in the similarity structure of context during encoding cannot be accounted for by variation in the rate of context updating.

Representational similarity.

Our third model variant proposes that category-specific differences in behavioral and neural effects arise because of differences in semantic representational structure between the three cat-

egories. These structural differences exhibit themselves when an item is studied and retrieves a contextual representation. Previous published versions of CMR assumed that these retrieved contextual representations were unrelated to one another, but here, we propose that items from the same category are associated with similar contextual representations. Model parameter σ controls within-category similarity by scaling the amount of item-specific noise added to a category prototype to create the contextual representation associated with a particular item. We allow this parameter to take a distinct value for each category. This gives the model the freedom to control inter-item similarity on a category-by-category basis. The consequences of this freedom can be somewhat complex: When an item is presented, pre-experimental context is retrieved and allowed to update context. As a result, differences in pre-experimental similarity will also affect the similarity structure of experimental context during encoding.

We found the best-fitting parameters for this model variant, and found that σ was lowest for celebrities, greater for locations, and greatest for objects. That is, the best-fitting model assigned highly similar pre-experimental contexts to celebrities, while objects were associated with highly variable pre-experimental contexts. Because retrieved pre-experimental context drives context evolution during encoding, category discriminability in context follows a similar pattern: Celebrities are most discriminable, followed by locations, then objects. Whereas the other model variants failed to predict the similarity structure observed in the neural oscillatory data, this variant produces the correct qualitative predictions. These differences in within-category similarity cause differences in cue strength during recall. Categories with high within-category similarity tend to provide good cues for one another, causing a relative increase in both category clustering (Fig. 6d) and recall (Fig. 5d), while leaving recency unaffected (Fig. 5h).

As described by Howard and Kahana (2002), items associated with similar states of context tend to provide good cues for one another. Here, we demonstrate that similarity can be simultaneously influenced by context dynamics during encoding (resulting in recency and temporal contiguity effects) and the similarity structure of pre-experimental context (resulting in category-specific activation in context during encoding, and category clustering during recall). By incorporating a more detailed representation of pre-experimental associations into CMR, we also gain the ability to make predictions about item-level variability in encoding and recall behavior; we turn to this topic next. Of the three models considered, only the third, which had the ability to alter semantic representa-

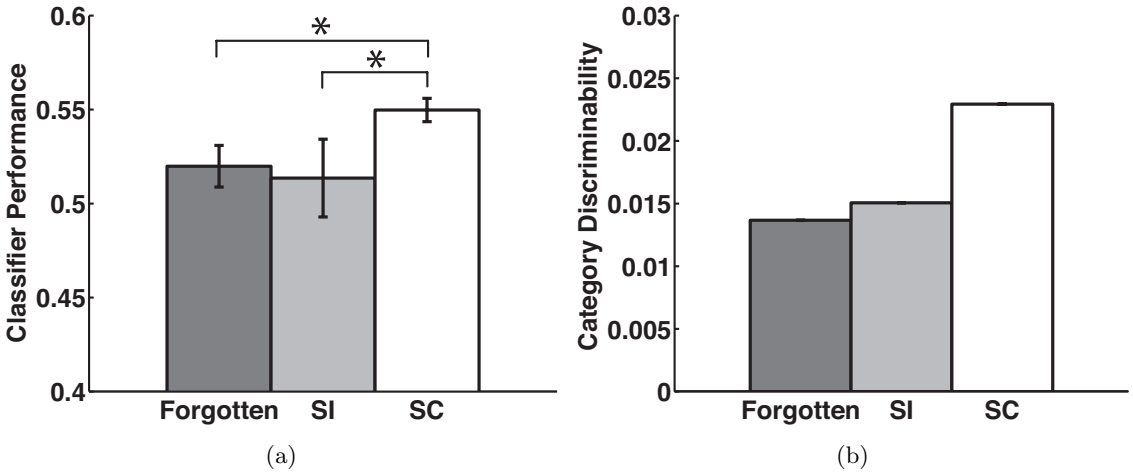
tional structure between categories, could simultaneously account for behavioral performance and category differences in the neural data. Thus, we focus on this model variant for the remainder of this report, as we examine item-level and subject-level variability in behavior and neural measures.

Simulation Analysis III: Contextual dynamics

Our theory (and retrieved-context theory, more generally) describes how ongoing experience (described in terms of the activation of a succession of featural item representations), is used to construct a temporal context representation whose dynamics control search through memory. This contextual representation is constructed during the study period through a series of retrievals; each item representation is projected through the associative structures connecting the item and context layers (Fig. 1), retrieving a contextual state laden with semantic information. This semantic information is integrated into context over time, resulting in a temporal-semantic cue that can guide both temporal and semantic organization during memory search. Here, we describe in more detail how these cognitive mechanisms determine the representational characteristics of neural signal during both study and memory search, and establish how these representational characteristics relate to the organization of memory search.

Item-level fluctuations in category discriminability

In Simulation Analysis II, we showed that category discriminability (in both the model and the neural data) during encoding correlates with variability in subsequent category clustering at the subject and category levels. Morton et al. (2013) showed that fluctuations in neural category discriminability also predict subsequent clustering at the level of individual items. They split studied items by whether they were subsequently recalled or not recalled, and split the recalled items by whether they were recalled as part of a category cluster (two or more items from the same category recalled successively) or were isolated from same-category items during recall (recalled adjacent to items from other categories). To evaluate whether neural category discriminability predicted subsequent recall or subsequent clustering, they first trained a pattern classifier to discriminate between the stimulus categories. They trained the pattern classifier on a separate session that participants completed prior to the free-recall task, where they were presented with each stimulus and rated their prior familiarity with each of the 256 items in each category. They then applied



(a) (b)

Figure 9. (**a**) Performance of a classifier applied to oscillatory activity during the study period. The classifier was trained on a separate session where participants performed a non-episodic-memory task where they rated their familiarity with each item. Classifier performance was greater for subsequently clustered (SC) items, compared to subsequently isolated (SI) and subsequently forgotten items. (**b**) Category discriminability of context during study, for the best-fitting CMR with variable within-category similarity.

this pattern classifier to the study period of the free-recall task, and found that subsequently clustered items were associated with higher neural category discriminability than both subsequently isolated items ($t(28) = 2.39$, $p < 0.05$) and subsequently forgotten items (Fig. 9a; $t(28) = 3.26$, $p < 0.005$). They also found a similar pattern of results when the pattern classifier was trained on the study period of the free-recall task, using a cross-validation scheme; however, in that case, the difference in neural category discriminability between subsequently clustered and subsequently isolated items was not significant. For that reason, we focus here on the model's predictions for the familiarization-to-study classification analysis.

We examined whether the best-fitting model from Simulation Analysis II, which was only fit to behavior, correctly predicts the relation Morton et al. (2013) observed between item-level fluctuations in neural category discriminability during encoding and subsequent clustering. Because the pre-experimental context associated with each item contains both category-specific information (the prototype representation) and item-specific information (the "item noise" added to the prototype representation), the model assumes there will naturally be fluctuations in the category discriminability of presented items. We examined whether these fluctuations in encoding context result in an increased tendency toward clustering during recall.

We first developed an analogue to the familiarization-to-study classification analysis used by Morton et al. (2013). This is similar to a pattern classification analysis, but instead of examining neural data, we examine the representational states of context pulled from the model as the simulated lists are presented. Each item is given a label corresponding to its category, and the state of context that is active when that item is presented is assigned to the set of context representations associated with that category. Morton and colleagues suggested that the familiarization period might not exhibit integrative activity, in that the familiarization judgment encouraged participants to focus on one item at a time, and didn't require any sort of association formation for a later memory test. Therefore, we assumed that context during the familiarization period only reflects the pre-experimental context retrieved by the current item (that is, $\beta_{\text{enc}} = 1$, and therefore $\mathbf{c}_i = \mathbf{c}_i^{IN}$). We then simulated a pattern classification analysis in which the classifier was trained on the neural data from the familiarization period and then used to estimate the category discriminability of each item. We simulated the familiarization session, using the best-fitting parameters obtained in Simulation Analysis II, but setting $\beta_{\text{enc}} = 1$. We then simulated the free-recall lists, and compared the representational states of context during encoding of each item to the representational states of context from the familiarization period. We calculated, for the context during each studied item, the mean similarity to same-category items in the simulated familiarization session, and subtracted the mean similarity to different-category items. Similar to the neural results observed by Morton et al. (2013), we found that subsequently clustered items were associated with higher category discriminability in context than both subsequently isolated items and subsequently forgotten items (Fig. 9b)[1].

These results are consistent with the hypothesis that the neural data reflect differences in category discriminability between different states of context during study, which predict subsequent recall performance. These results may also be compatible with the hypothesis that item representations vary in prototypicality, and that, during recall, retrieved items serve as cues to retrieve other items (cf. Raaijmakers & Shiffrin, 1980). We will return to the distinction between item and context representations in the Discussion. Next, we examined model predictions that follow from the integrative nature of the context representation.

[1]Similar results were observed when we calculated category discriminability by comparing context during each item to the context of other items on the same list, as in the category discriminability analyses presented in Simulation Analysis II.

Integrative activity

Morton et al. (2013) found evidence that neural category discriminability changes over time during encoding. They used pattern classification with a cross-validation scheme to examine how neural category discriminability changed as a train of items from the same category is studied. They found that the pattern classifier's estimate of neural evidence for the stimulus category increases over the first three train positions, leveling off beyond that (Fig. 10a; Morton et al., 2013). We examined whether the context representation of the model demonstrates similar changes in category discriminability (calculated as described in Simulation Analysis I and II). Consistent with the data, we found that the category discriminability of context increases as multiple items from a category are presented in succession (Fig. 10c). As described in Simulation Analysis I, the model predicts an increase in category discriminability because context contains a recency-weighted average of the pre-experimental contexts of presented items. At later train positions, context contains an average composed mainly of items from the same category, resulting in greater category discriminability.

The model also makes the prediction that context specific to a given category should decay slowly after switching to a different category (Fig. 10d). The category of the previous train decays as items from the current train are presented. As a baseline, we examine the "other" category, which is neither the category of the current train nor the category of the previous train. In contrast to the immediately previous category, the discriminability of the other category shows a negligible change with train position[2]. We re-analyzed the data reported by Morton et al. (2013) to examine whether the data support this prediction. Consistent with the model, we found that the classifier's estimate of the strength of the previous train's category decreases with train position (Fig. 10b). For each participant, we performed a linear regression of classifier evidence on train position, weighted by the number of samples available for each train position. Across participants, this slope was significantly negative (mean -0.0041, SEM 0.0020; $t(28) = 2.02$, $p = 0.027$, one-sided test). Importantly, we did not find a similar decrease in the baseline category (mean 0.00018, SEM 0.0017; $t(28) = 0.107$,

[2]Numerically, there is a slight increase in the discriminability of the baseline/other category with train position. This is because the baseline category is sometimes presented in the following train, but never in the previous train (by definition). Because context changes gradually over time, adjacent states of context are similar to one another. As a result, similarity to the baseline category tends to increase at later train positions, because the next train (which may include the baseline category) is getting closer. We carried out a follow-up analysis in which the contextual states were not compared to one another within the list, but rather were compared to randomly generated sets of items from the three categories. This follow-up analysis produced similar results for the current and previous categories, and showed a slight decrease in the discriminability of the other category with increasing train position.

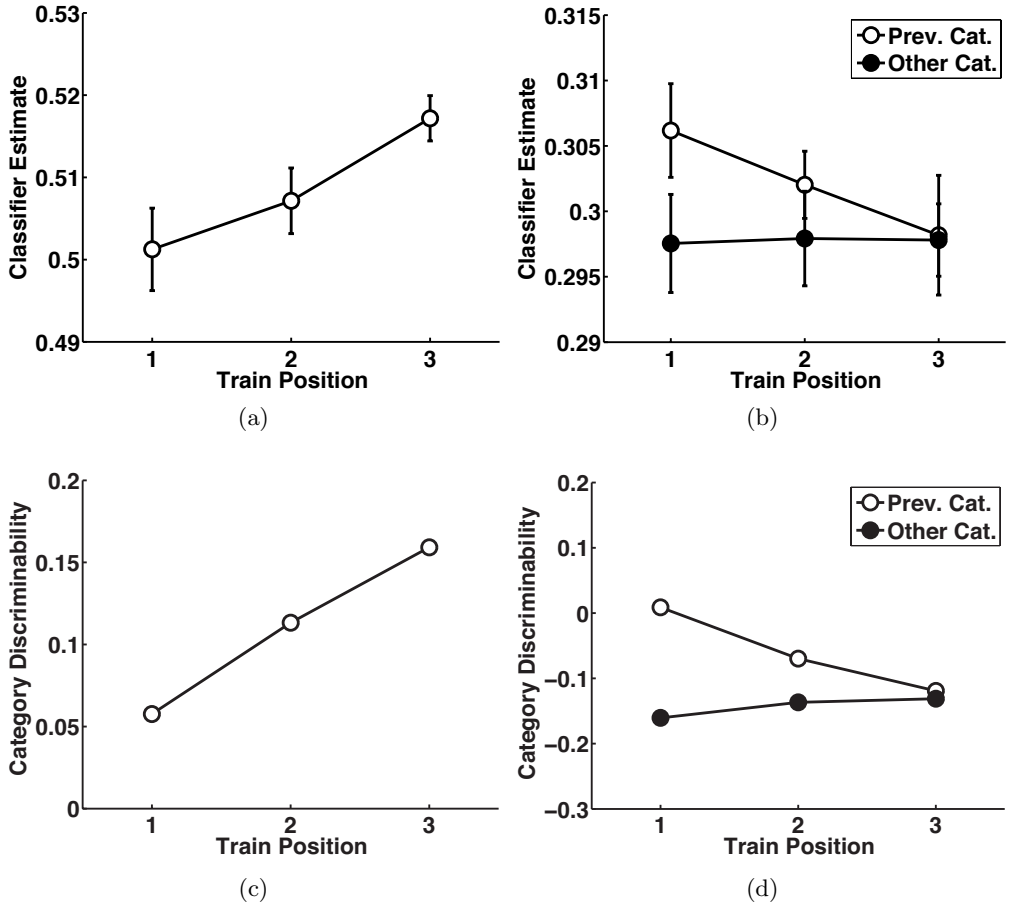
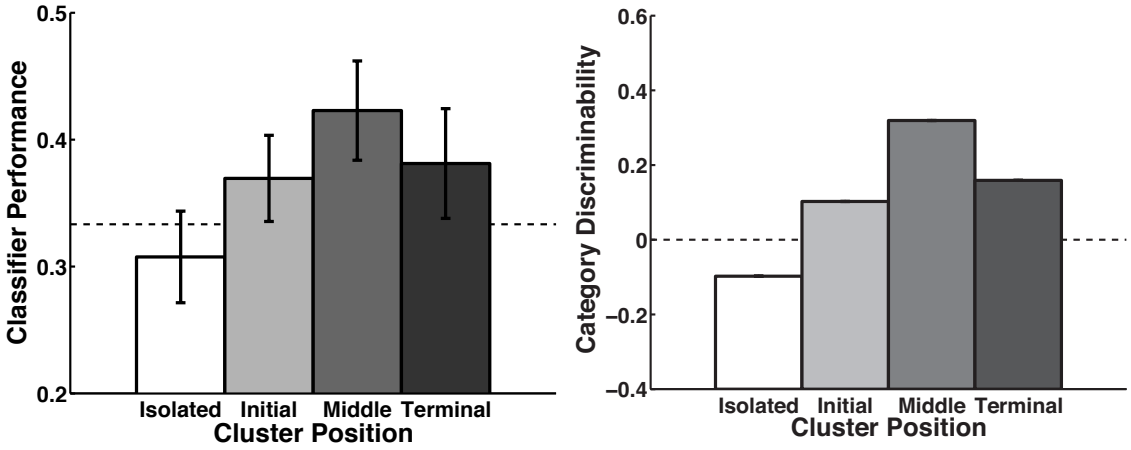


Figure 10. (**a**) Classifier evidence for the current stimulus category increases when multiple items from the same category are presented successively in a train, suggesting that neural activity during encoding reflects the recent history of stimulus presentations. (**b**) Activity related to the category of the previous train of items decreases with train position, while the "other" less-recently presented category does not change with train position. (**c**) Category discriminability of context in the best-fitting model. As in the neural data, similarity to other items from the current category increases as a function of train position. (**d**) Same as (c), but showing similarity between the current state of context and the previous category, as well as similarity to the item that is neither the current or the previous category.



(a) (b)

Figure 11. (**a**) Performance of a pattern classifier applied to oscillatory power averaged over 3–0.5 s before onset of a vocalized recall, to identify the category of the upcoming recall. Performance is shown for different cluster position bins (see text for details). (**b**) Strength of the current category in context during the recall period at different cluster position bins, for the best-fitting CMR with variable within-category similarity. Category discriminability for a given recalled item is calculated based on the context cue used to retrieve that item.

$p = 0.92$), suggesting that the decrease in the classifier's estimate of the previous train's category is not merely due to the classifier's constraint that category estimates sum to 1.

Although the model was only fit to measures of recall behavior, it generates correct predictions for the dynamics of representational structure during encoding. Both the general trends (increase of activity related to the current category, decrease of the previous category) and the temporal scale of these changes are predicted by the model. It is also important to establish the relation between these dynamics and subsequent recall behavior; we address this in Simulation Analysis IV, which examines the relation between individual differences in integrative activity and category clustering.

Category-specific cues during retrieval

Morton et al. (2013) also examined the dynamics of category-specific neural activity during the recall period. They examined oscillatory activity in the 3–0.5 s before onset of vocalization of correct recalls, excluding epochs that included vocalizations related to previous items. Using pattern classification with a cross-validation procedure, they demonstrated that the oscillatory activity leading up to a vocalization can be used to predict the category of the item being recalled (classifier performance was 0.366 [SEM 0.013], which was significantly above chance [$0.\bar{3}$]; $t(28) = 2.47$,

$p < 0.01$). They found that classifier performance was greatest during recall of clusters of items from the same category, compared to periods where the participant was switching between categories. They divided recall epochs into bins corresponding to the sequence of recalls: *isolated* items were preceded and followed by recalled items from different categories; *initial* items were preceded by a different category, and followed by the same category; *middle* items were preceded and followed by recalls from the same category; and *terminal* items were preceded by a recall from the same category, and followed by a recall from a different category. Items in the middle cluster position bin were classified most accurately; items in the initial and terminal bins were classified less accurately; and isolated items were classified poorly, not exceeding chance levels of performance (Fig. 11a).

We hypothesized that the category-specific activity observed by Morton et al. (2013) in the 3 s prior to each recall corresponds to the contextual cue used to retrieve the recalled item. To examine whether the model that best fit the behavioral data predicts similar changes in category activation during recall, we calculated the category discriminability of the context cue used to retrieve each item. Similarly to their pattern classification procedure, we compared different states of context during recall. The category discriminability of each recall cue was defined as the mean cosine similarity to other items from the same category recalled during that list, relative to the mean cosine to recalled items from different categories. If only one category was recalled on a given simulated list, that list was excluded from the analysis.

We examined whether the model correctly predicts the neural dynamics observed by Morton et al. (2013). The model's predictions were strikingly similar to the observed changes in category-specific activation in the neural data (Fig. 11b). The category discriminability for isolated items was negative, indicating that activation specific to the category of the item about to be recalled was less than activation related to the other two categories. This may be related to the fact that, in the neural data, the pattern classifier was actually numerically below chance when applied to isolated recalls, suggesting that other categories were more active than the category of the upcoming recall. The model also predicts the observed increase related to clustering; middle items were associated with more discriminable context cues than initial and terminal items, suggesting that the strength of category-specific activity in context is related both to the category of the previous recall and the category of the next recall. In both the model and the neural data, category-specific activity builds during a series of recalls from the same category, and decreases when recall is about to switch to

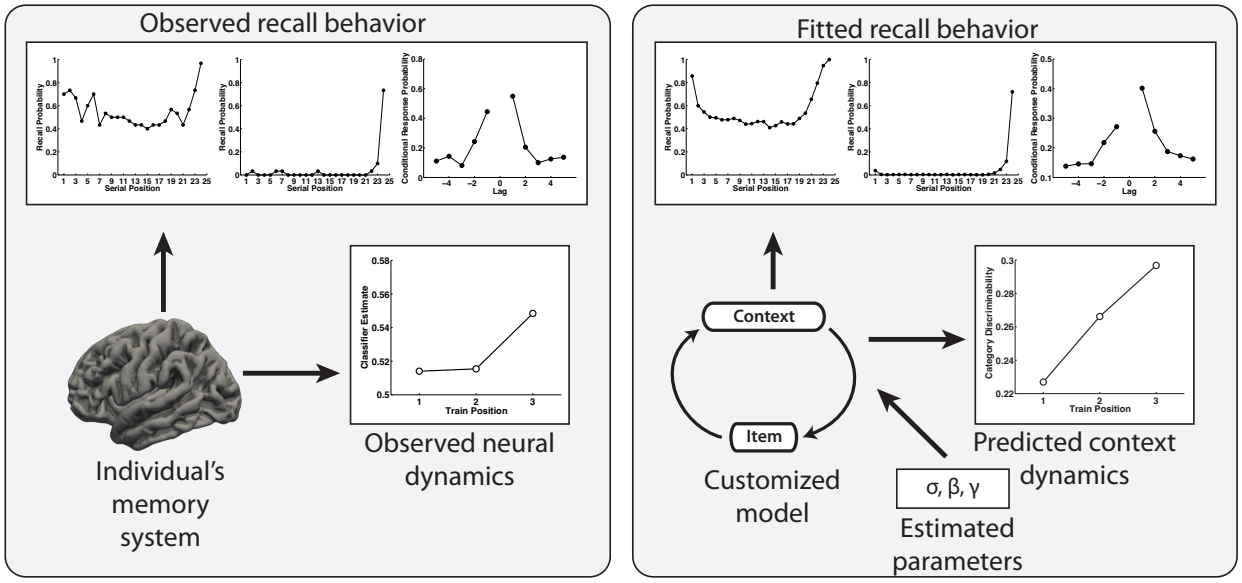


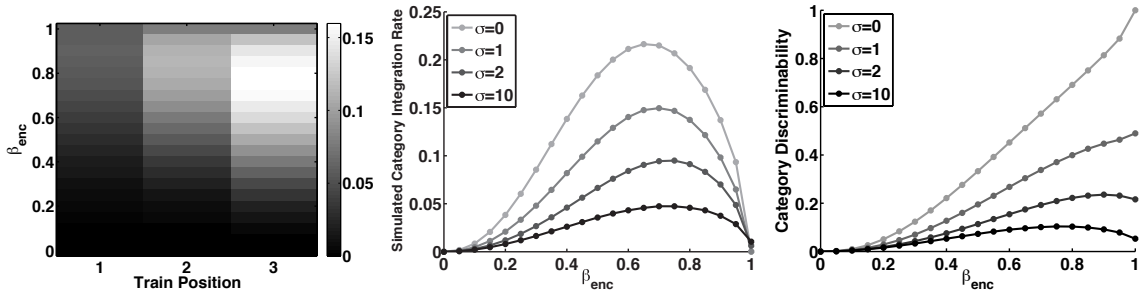
Figure 12. Left panel: Each individual participant has observed recall behavior and neural activity recordings during encoding. Right panel: for each participant, a customized model is constructed by optimizing model parameters to fit that participant's recall behavior. Based on those fitted parameters, the model produces predictions about how distributed neural patterns should evolve during encoding. We then compare the observed neural dynamics to the context dynamics predicted by the model.

another category.

Simulation Analysis IV: Individual differences

Morton et al. (2013) found that the category discriminability of neural activity during encoding is related to subsequent organization by category. Further analysis (Simulation Analysis II) showed that individual differences in category discriminability also correlate with overall recall performance. Furthermore, Morton et al. (2013) found that neural category discriminability increases as a series of items from the same category are presented in succession, and that the rate of this increase (which we refer to here as *neural category integration rate*) correlates with individual differences in category clustering. We examine whether the model can account for these individual differences in distributed neural activity during encoding and subsequent recall behavior.

To examine whether the model can account for individual differences in recall behavior and neural activity, we fit the model to a range of measures of recall behavior. This fitting process allowed us to estimate model parameters for each participant, giving a customized model of each participant's memory system. Each customized model makes predictions, for a given participant,



(a) (b) (c)

Figure 13. Category discriminability in context for a range of possible parameter values in the model. (**a**) The shade of a given point indicates category discriminability of simulated context as a function of train position, for a range of values of β_{enc}. All σ parameters are set to their best-fitting values from Simulation Analysis II. (**b**) The rate of increase in category discriminability by train position (simulated category integration rate) is dependent on β_{enc} and σ (here, set the be the same for all categories). Simulated category integration rate is highest when σ is low and β_{enc} is at an intermediate value. (**c**) Average category discriminability of context as a function of σ and β_{enc}. When $\sigma = 0$, category discriminability increases with increasing β_{enc}, as context becomes more focused on the most recently presented item. When $\sigma > 1$, category discriminability is highest at intermediate values of β_{enc}.

about how context should change over time during encoding. We then compared these predicted context dynamics to observed neural dynamics, to test the hypothesis that distributed neural oscillatory activity reflects temporal context. Figure 12 gives an overview of our approach to testing the model's ability to account for individual differences in neural activity.

First, we examine how the model parameters shape the simulated category discriminability of context during encoding. We find that both prior experience (which shapes the structure of pre-experimental context) and the rate of contextual drift (which determines how quickly pre-experimental context is integrated during encoding) affect how context unfolds during the study period. Next, we fit the model to the recall behavior of individual subjects, and find that the customized models can account for a range of variability in recall behavior. Finally, we test the model's predictions for individual variability in category-specific neural activity during encoding. We find that the model successfully predicts individual differences both in overall category discriminability and in the rate at which category discriminability changes over time.

Variability in context dynamics

There are four model parameters that determine the similarity structure of context during encoding: β_{enc}, σ_c, σ_l, and σ_o. The σ parameters determine the similarity of the pre-experimental

contexts associated with different items in a given category. Because contextual drift is assumed to be driven by pre-experimental contexts associated with presented items, the σ parameters will influence contextual evolution during encoding. β_{enc} determines how much of the state of context during encoding of an item is composed of its retrieved pre-experimental context, and how much of the context is carried over from the previous state of context. If $\beta_{\text{enc}} = 1$, then context will only contain the pre-experimental context associated with the item; when $\beta_{\text{enc}} < 1$, context will also reflect the history of recently presented items. Together, β_{enc} and the σ parameters determine the category discriminability of context during learning of each item in a simulated experiment.

In order to illustrate how the composition of context depends jointly on context integration and the composition of pre-experimental context, we examined how category discriminability evolves during encoding for a wide range of the σ and β_{enc} parameters. For simplicity, we set all σ parameters to be equal; here, we refer to the value for all categories as σ. At each point in this two-dimensional parameter space, we simulated 20 replications of the experiment. We then calculated two measures: mean category discriminability and *simulated category integration rate*. For each item, we calculated the category discriminability of context, as described in Simulation Analysis I. We calculated the mean category discriminability, averaged over all simulated item presentations. We also calculated simulated category integration rate, the slope of the change in category discriminability over the first three train positions, with the regression weighted by the frequency with which each train position appears in the experiment.

Simulated category integration rate is jointly determined by σ and β_{enc} (Fig. 13). Simulated category integration rate is greatest for intermediate values of β_{enc} (Fig. 13a). When $\beta_{\text{enc}} = 0$, context is static, so the integration rate is 0. When $\beta_{\text{enc}} = 1$, context is changing quickly, so that it contains only the pre-experimental context of the most recent item; however, the *category discriminability* does not change with train position. This is because, on average, the pre-experimental context of the most recent item will not vary with train position. Simulated category integration rate is inversely proportional to σ. When σ is low, the pre-experimental contexts being integrated into the current context are highly prototypical, causing increased category discriminability.

While σ and β_{enc} do not interact much in determining simulated category integration rate, average category discriminability is determined by a more complex combination of the two parameters. When $\sigma = 0$, category discriminability is greatest when $\beta_{\text{enc}} = 1$, maximizing context

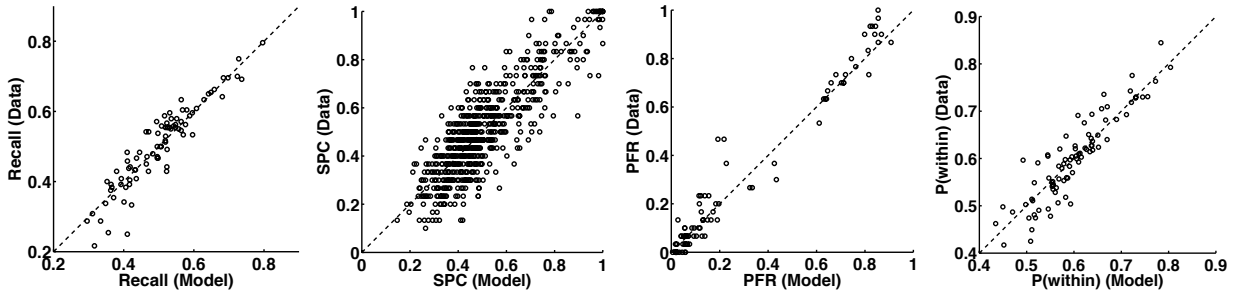
integration rate[3]. When $\sigma >= 2$, intermediate values of β_{enc} allow context to accumulate category information, averaging over the pre-experimental contexts associated with multiple items to produce context that is more prototypical than the pre-experimental context associated with any given item. In this way, context integration allows the model to extract a summary, or gist, representation of recently presented items; we discuss this property of the model further in the Discussion.

As described above, the model makes predictions for how cognitive representations should change during encoding, as a function of the σ and β_{enc} parameters (Fig. 13). We test these predictions by comparing neural activity patterns to states of simulated context (Fig. 12). In the next section, we use measures of recall behavior to estimate parameters for each participant in the scalp EEG categorized free recall experiment. Based on the estimated parameters for each participant, the model provides predictions for the category discriminability of context during encoding. We test these predictions by comparing each participant's simulated category discriminability to their measured neural category discriminability during encoding.

Model selection

First, we aimed to determine which model parameters are critical to account for individual variability in recall performance. Starting from the best-fitting group parameters for the best-fitting model variant (which allowed representational similarity to vary between categories; see Simulation Analysis II), we examined nested model variants that allowed different parameters to vary between subjects. We then examined whether the best-fitting model is able to predict individual differences in neural data during encoding, despite having been fit only to recall behavior. Morton et al. (2013) noted that two participants (of 29) were outliers in terms of their neural data; in contrast to the other participants, they showed a strong trend toward category-specific activity decreasing over successive presentations of items from that category (neural integration rate -0.0183 and -0.0225; both are more than one interquartile range below the first quartile). This trend may be a result of noise in the neural signal, or could reflect that these two participants approached the task in an unusual manner. These participants are excluded from the following analyses of individual differences, leaving 27 participants.

[3]The simulations of encoding-task context reported by Polyn et al. (2012) are similar to this condition, since they assume that each item is associated with an identical encoding task context. Therefore, they found a monotonic relation between β_{enc} and task discriminability, unlike the generally nonmonontic relation that we observe in the present simulations, since in our fits $\sigma > 0$.



(a) (b) (c) (d)

Figure 14. Individual participant fits for model M4 (see text for details). Dotted lines indicate the identity function. (**a**) Recall probability, for the best-fitting model and the observed data. Each point indicates overall recall probability for one participant, for one of the three categories. (**b**) Fit to individual serial position curves (SPCs). Each point gives recall probability for one participant at one serial position. (**c**) Probability of first recall (PFR), for each participant and each of the 3 last serial positions on the list. (**d**) Probability of making a within-category transition, conditional on the category of the just-recalled item. Each point indicates P(within) for one participant, conditional on a given category.

We examined four model variants, which progressively allowed more parameters to reflect individual differences. The first two models allowed the parameters controlling the representational structure of context at encoding to vary by individual. The simplest model we examined, designated M1, only allowed study-period integration rate (β^{enc}) to vary between participants. M2 also allowed representational similarity for each category (σ_c, σ_l, and σ_o) to vary between participants. The next two models also allowed the parameters directly affecting recall behavior to vary by individual. M3 allowed contextual retrieval (β^{rec}) to vary by individual (along with the free parameters of M1 and M2). M4 was the most flexible of the set, allowing the parameters controlling the balance of influence between pre-experimental and experimental associations (γ^{FC} and γ^{CF}) to vary by participant (along with all of the above mentioned parameters). Other parameters were fixed at the best-fitting group values from the previous search. These included the parameters controlling the primacy effect (ϕ_s and ϕ_d), the number of dimensions of the context representation (N), and the decision parameters (κ, λ, η, and τ). A separate model fit was carried out for each participant (see Appendix for details of the fitting procedure).

We used the Bayesian Information Criterion (BIC; Kahana, Zhou, Geller, & Sekuler, 2007; Polyn et al., 2009; Schweickert, 1978) to quantify whether the improved fit of the more complex models justifies the large number of free parameters required to capture individual differences. This statistic suggests that the increased complexity is not justified: model M1 has the lowest (i.e.

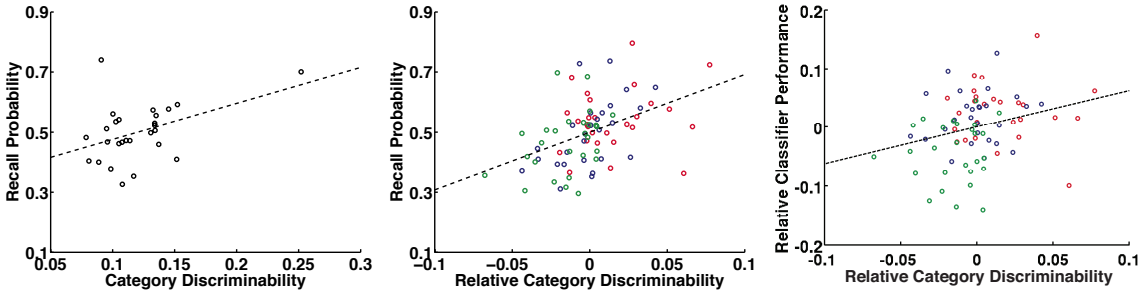
best) BIC score, whereas M4 has the second highest score, despite it having the best overall fit to the data (RMSD= 0.0746). This is unsurprising, given that it had the highest number of free parameters (203 parameters to fit 1537 data points). However, the specific pattern of successes and failures of the four model variants in accounting for both behavioral and neural phenomena justifies a closer look at M4. All four model variants provide a good fit of individual variability in serial position effects and overall recall, and at least a fair fit of individual variability in temporal contiguity. However, M1 provides a poor fit to clustering (RMSD= 0.1026), while M4 provides a good fit (RMSD= 0.0394). See the Appendix for best-fitting parameters for each model.

As shown in Figure 14, the M4 model variant demonstrates a good fit to individual variability in recall probability as a function of category, recall by serial position, probability of starting recall at each serial position, probability of making a within-category transition conditional on category, and a fair fit to variability in temporal clustering. Most importantly, M4 provides a starting point for understanding how inter-participant variability in cognitive processes (estimated based on behavioral differences between subjects) can provide insight into individual differences in neural dynamics, in terms of the category discriminability of neural representations.

In the following sections, we examine the whether the M4 variant of the model can account for individual differences in category-specific neural activity during encoding. For each simulated individual, we simulated 80 replications of that participant's 30 mixed-category lists, using the best-fitting parameters based on that participant's recall behavior. Based on the simulated states of context for each simulated participant, we then calculated mean category discriminability and simulated category integration rate (as described in *Variability in context dynamics*). We compared these measures to their analogous neural measures of mean classifier accuracy and neural category integration rate. As we discuss below, we find that the model successfully predicts a number of attributes of individual differences in neural activity during encoding.

Category discriminability and recall performance.

We first examine mean category discriminability in the neural data and simulations. In the actual experiment, individual differences in neural classifier performance correlated with overall recall performance (Simulation Analysis II; Fig. 4c). Furthermore, for a given category, classifier performance relative to the other categories correlated with recall performance for that category



(a) (b) (c)

Figure 15. (**a**) With the exception of one outlier, the model does not predict a strong relation between a given participant's overall category discriminability and recall probability ($r = 0.431$, $p = 0.025$; with outlier excluded: $r = 0.173$, $p = 0.40$). (**b**) However, the model predicts a strong relation between recall performance and the discriminability of each category relative to the mean discriminability over all the categories ($r = 0.438$, $p = 4.4x10^{-5}$). (**c**) The model successfully predicts individual variability in classifier performance at the level of categories ($r = 0.273$, $p = 0.014$). Relative category discriminability is defined as the model's predicted category discriminability for each category, relative to the mean for that simulated participant. Similarly, relative classifier performance is the classifier accuracy for each participant and category relative to the overall accuracy for that participant.

(Fig. 4d). We examined the model's predictions for individual differences in category discriminability and recall performance, based on the M4 model simulations described in *Model selection*.

For each participant, we calculated mean category discriminability in context during encoding, averaged over all train positions and categories. In the model, we found a significant relation between category discriminability and recall probability (Fig. 15a; $r = 0.431$, $p = 0.025$), but this correlation was dependent on one participant with unusually high estimated category discriminability (with outlier excluded, $r = 0.173$, $p = 0.40$). Furthermore, individual estimates of model category discriminability did not correlate with neural classifier accuracy ($r = -0.135$, $p = 0.49$), suggesting that the model failed to accurately capture the relation between recall behavior and individual differences in overall category discriminability during encoding. These results suggest that some variable not present in the model affects the relation between category discriminability and recall at the level of participants. In the actual experiment, category discriminability may be an important factor in mitigating proactive interference and improving recall performance; for example, if the previous list contained celebrities, then focusing recall on locations and objects from the current list will be relatively easy since they are released from proactive interference. However, the model does not simulate interference from prior lists, so it would not be able to account for

such an effect of proactive interference.

Although the model only showed a weak relation between overall category discriminability and recall performance, we found that the model category discriminability of individual categories, relative to the overall discriminability for that participant, was strongly related to the recall of each category ($r = 0.438$, $p = 4.4x10^{-5}$). This may reflect the competitive nature of free recall; if one category has relatively similar pre-experimental contexts compared to the other categories, items from that category will provide good cues for one another, causing them to be clustered and recalled more frequently at the expense of items from other categories. Furthermore, the model succeeded in predicting individual differences in relative classifier performance: Relative category discriminability and relative classifier performance were significantly correlated (Fig. 15c; $r = 0.273$, $p = 0.014$).

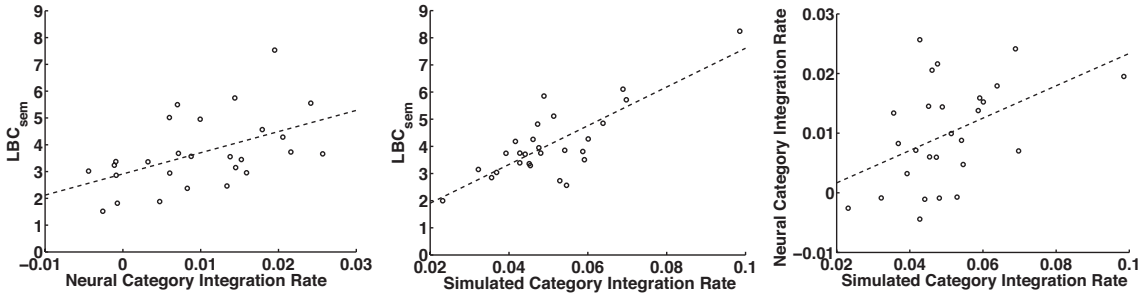
These results demonstrate that, using the model, it is possible to use measures of recall behavior to predict individual differences in encoding-related brain activity. In this case, the model provides accurate predictions about the relative neural discriminability of the different stimulus categories. This successful prediction provides support for the link between encoding representations and recall behavior hypothesized by the model. It is important to note that this property is not true of every model that can fit recall behavior reasonably well; see Simulation Analysis II for examples of model variants that capture recall behavior but fail to correctly predict the relation between recall behavior and neural category discriminability.

In addition to the predictions about overall category discriminability presented above, the model also makes predictions about how category-specific activity in context should change throughout the study list (see Simulation Analysis III). In the next section, we test the model's predictions for individual differences in context dynamics.

Integrative activity.

In the results reported by Morton et al. (2013), neural category integration rate was significantly correlated with individual differences in category clustering (Fig. 16a; $r = 0.503$, $p = 0.0075$; with all subjects included, $r = 0.422$, $p = 0.023$)[4]. We found a similar relation in the model: across

[4]Morton et al. (2013) calculated LBC_{sem} over all lists, including control lists; here, we only include mixed lists. We observe a very similar correlation between LBC_{sem} and the slope of classifier estimates as the one reported by Morton et al. (2013).



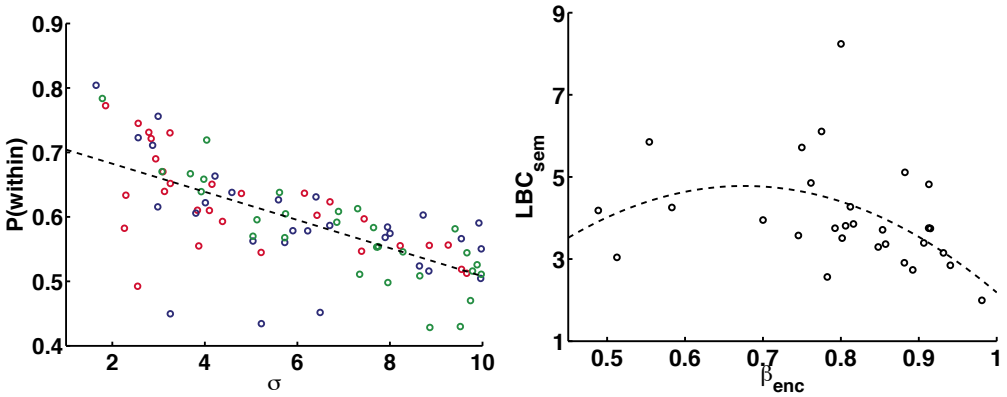
(a) (b) (c)

Figure 16. (**a**) The rate with which classifier estimates increase in strength correlates with individual differences in clustering (as measured by LBC_{sem}; $r = 0.503$, $p = 0.0075$). (**b**) A version of CMR fit to individual differences (model M4; see text for details) shows a similar correlation between representational integration rate and category clustering ($r = 0.792$, $p = 9x10^{-7}$) (**c**) There is a significant correlation between neural integration rate and the rate of change of category activation in the model's context representation ($r = 0.454$, $p = 0.017$).

participants, simulated category integration rate was strongly correlated with category clustering, as measured by LBC_{sem} (Fig. 16b; $r = 0.792$, $p = 9x10^{-7}$). The model successfully accounts for the link between neural category integration rate and category clustering observed by Morton et al. (2013). Furthermore, we found that the individually customized model, which was fit only to individual recall behavior, was able to predict individual differences in neural dynamics during encoding: Simulated category integration rate was significantly correlated with neural category integration rate (Fig. 16c; $r = 0.454$, $p = 0.017$). These results demonstrate that the model can make successful predictions not only about individual differences in average category-specific activity at encoding, but also correctly predict the rate at which category-specific activity changes in discriminability during encoding of a series of items.

In order to determine why the model predicts a link between category integration rate and category clustering, we examined how model parameters influence these two measures. In the model, simulated category integration rate is determined by β_{enc} and the σ parameters for each category (see *Variability in context dynamics*). Category organization in recall is also influenced by these parameters: They determine the state of context associated with each item, which is subsequently retrieved and used to guide recall. Because recall depends on pre-experimental associations as well as experimental associations, the σ parameters also influence recall directly.

We found that, in the simulated participants, σ in each category was correlated with the amount of clustering in that category (Fig. 17a; $r = -0.687$, $p = 1.9x10^{-13}$). This is because



(a) (b)

Figure 17. (**a**) Across simulated participants, σ is correlated with category clustering ($r = -0.687$, $p = 1.9x10^{-13}$). (**b**) In the same simulated participants, β_{enc} shows marginal linear ($p = 0.074$) and quadratic ($p = 0.053$) relations to category clustering.

decreasing σ increases the category discriminability of both pre-experimental context and experimental context, leading to greater category clustering. As discussed in *Variability in context dynamics*, decreasing σ also leads to increased simulated category integration rate, contributing to the relation between simulated category integration rate and category clustering. We also found marginal effects of β_{enc} on category clustering (Fig. 17b; linear trend, $p = 0.074$, quadratic trend, $p = 0.053$). There is a nonmonotonic relation between β_{enc} and LBC_{sem}, where clustering is greatest for intermediate values of β_{enc}. As discussed in *Variability in context dynamics*, for most values of σ, category discriminability of encoding context is greatest when β_{enc} is at an intermediate value. This increase in encoding context category discriminability will cause increased category clustering, since encoding context is retrieved during recall and used to guide memory search. Simulated context integration rate also peaks at intermediate values of β_{enc} (see *Variability in context dynamics*), suggesting that variability in β_{enc} contributes to the link between simulated category integration rate and category clustering.

These results suggest that much of the relation between simulated category integration rate and category clustering is due to variability between participants in the semantic structure of their categories. For example, some participants may have celebrity representations that are quite similar to one another, whereas others have more diffuse celebrity representations. The model suggests that, if items in a category are associated with similar pre-experimental states, this will cause faster category integration and increased category clustering. Both simulated category integration rate

and category clustering are determined by a number of factors, including learning rate (in both cases) and contextual dynamics during recall (in the case of category clustering). The model allows us to account for these many interacting cognitive processes, and reveals this relationship between semantic structure, simulated category integration rate, and category clustering. The model also reveals a relationship between individual differences in β_{enc}, the rate of context change during encoding, and our two measures of simulated category integration rate and category clustering. However, this relationship is much weaker than the one involving semantic structure.

We also examined the relations between the σ and β_{enc} parameters and recall performance, and found similar relations as we observed for category clustering (i.e. an inverse relation for σ and a nonmonotonic relation for β_{enc}). It is unsurprising that recall and clustering depend on model parameters in similar ways, since recall and clustering are highly correlated in the simulated subjects (P(within): $r = 0.687$, $p = 2.1x10^{-13}$; LBC_{sem}: $r = 0.828$, $p = 3.0x10^{-8}$). A similar pattern is observed in the actual data (P(within): $r = 0.630$, $p = 6.3x10^{-11}$; LBC_{sem}: $r = 0.847$, $p = 6.7x10^{-9}$), consistent with literature suggesting a close link between category organization and recall performance (Puff, Murphy, & Ferrara, 1977).

Discussion

The present work with the Context Maintenance and Retrieval (CMR) model specifies the interactions between semantic memory and episodic memory at a mechanistic level. The theory is implemented as a computational model that describes a set of interacting cognitive mechanisms that bridge behavioral and neural phenomena observed in free recall. Studied material activates feature-based representations, and these representations trigger the retrieval of semantically laden pre-experimental knowledge. This retrieved contextual information is used to construct a temporal context representation which is associated with the studied material. During memory search, this contextual representation is used to probe the associative structures of the memory system to reactivate the feature-based representations of the sought-after past experience.

This idea that a temporal code is built from retrieved semantic information (Socher et al., 2009) suggests that episodic and semantic memory are inextricably intertwined. Long-standing knowledge allows us to interpret unfolding experience, and integration of successively reactivated details of long-standing knowledge yields a temporal code that is unique to a given moment. Other

theorists have proposed that an integrative mechanism could be used to create representations reflecting semantic structure (Elman, 1990; Jones & Mewhort, 2007; Rao & Howard, 2008; Howard et al., 2011). Here, we examined the influence of this semantic structure in an episodic free-recall task, in terms of both behavioral and neural dynamics.

The intertwining of semantic and temporal information in the model causes it to be considerably more constrained in terms of its predictions regarding the interaction of temporal and semantic information, as compared to previous versions of CMR. These previous versions had a parameter that controlled the strength of semantic associations between item representations. While the prior implementation gave the model considerable flexibility in accounting for the precise amount of semantic organization, it was unable to account for the representational structure of neural signal and its relation to behavioral performance.

In our Simulation Analysis sections, we established the viability of this framework for understanding behavioral and neural phenomena in free recall. Simulation Analysis I examines the interaction of temporal and semantic information in terms of the often-demonstrated performance advantage when semantically related items are studied in close temporal proximity, as compared to when they are spaced throughout the study list. To investigate this phenomenon, we simulated a classic experiment reported by Puff (1966), in which the temporal structure of categorized materials on the study list was parametrically manipulated. With a single set of parameters, the model is able to capture the increase in recall probability, category clustering, and within-category temporal organization as list structure changes from completely interspersed (no within-category items in adjacent positions) to completely blocked (Fig. 3a–c). The model also makes the prediction that if neural recordings were collected with this paradigm, category-specific neural signal would be much stronger in the blocked condition than the interspersed condition, suggesting a link between neural category discriminability and behavioral performance. Specifically, when same-category items are presented successively, the integrative dynamics of context cause the contextual representation to become progressively more similar to a category prototype, which both increases the neural discriminability of categorized items, and supports efficient recall of the items from a particular category (Fig. 3).

These dynamics are examined in more detail in Simulation Analyses II through IV, all of which are concerned with an experiment that manipulated temporal and category structure of the

study list while recording neural data with scalp EEG (Morton et al., 2013). We went beyond the analyses reported by Morton and colleagues, demonstrating substantial category-level differences in overall recall performance, category clustering, and neural classifier performance, which showed a general advantage for celebrities on all these measures, followed by landmarks, followed by objects (Figs. 4–6). In order to demonstrate the constraint provided by the behavioral and neural phenomena, we examined three variant models, each of which embodied a distinct hypothesis regarding how the cognitive system might treat items from the three categories differently. Of these, the only viable model involved allowing each of the three categories to have different internal structure, with celebrities having the highest inter-item similarity, followed by landmarks, which were less similar to one another, followed by objects, whose representations were the most diffuse (Fig. 7).

The interaction of integrative and associative processes provides a mechanistic explanation for a number of neural-behavioral relationships, including the positive correlation between classifier performance and overall recall performance (Fig. 4 & Fig. 15), and the tendency for items that elicit strong category-specific patterns to be subsequently recalled in a category cluster (Fig. 9). The integration process can make the context representation highly category-specific. When a train of same-category items is studied, the integration mechanism causes the representation of the current category to increase in fidelity (Fig. 10a & c) while the representation of the previous category declines smoothly (Fig. 10b & d), matching the pattern seen in the neural data. If an item is strongly associated to the contextual cue, then that item will be strongly supported when that cue is used during memory search. When an item is recalled from a given category, the item representation is reactivated, and this reactivated item representation retrieves category-specific context, which makes the contextual representation even more category-specific, leading to multiple successive recalls from the same category (Fig. 11).

A core assertion of the model is that the category structure of the neural signal is determined by four parameters: the three sigma parameters controlling within-category similarity, and the parameter controlling rate of context integration during study. To test this assertion, we created a family of models, each of whose parameters were custom-fit to an individual participant (using only the behavioral data). We found that the customized parameter sets for each participant allowed the family of models to account for individual differences on a number of behavioral measures. Even though each customized model was only ever given access to behavioral data during the fitting

process, we found that the family of models predicted individual differences in neural category structure (Fig. 15c), and in neural category integration rate (Fig. 16c).

Consequences of a temporal-semantic contextual cue

Classic work in categorized free recall led theorists to propose that a superordinate category representation could be used as a retrieval cue to target items from a particular category (Bousfield, 1953; Cofer et al., 1966; Tulving & Pearlstone, 1966; Puff, 1966, 1974; Raaijmakers & Shiffrin, 1980). The contextual evolution process during study gives insight into how a participant could activate the superordinate category representation common to a set of items. Context represents a recency-weighted average of the retrieved context from the succession of study items. If this integrative mechanism is paired with a prototype/exemplar model of category structure, integration of these retrieved contextual states will cause the contextual representation to become progressively more similar to the prototype representation. This prototype representation is similar to the classic notion of a superordinate category representation, in that it will support recall of all items from a given category.

As the contextual representation approaches a category prototype, there are a number of potential consequences, both positive and negative. As explored above, when a highly prototypical representation is associated with studied items, this increases the likelihood that those items will be recalled as part of a category cluster. While the computational cost of simulations led us to simulate a single list at a time in the current report, we have explored basic model predictions when prior lists are included in memory. A highly prototypical representation targets all items from a given category, including those from prior lists. Thus, the model predicts that a highly category-specific neural signal will be associated with an increased likelihood of prior-list intrusions. Furthermore, if a large number of items from the targeted category were studied on the prior list, a highly category-specific neural signal will be related to increased proactive interference from that prior list, which will tend to decrease the number of items recalled from that category.

Given that the contextual representation is a composite of temporal and semantic information, the more this representation comes to resemble the prototypical representation of a particular category, the less influence temporal information will have on memory search. Thus, a highly prototypical category context will support not only prior-list items, but also items from the targeted

category that were not studied (i.e., extra-list intrusions). As such, CMR provides a framework to link category-specific neural activity to behavior in false memory paradigms (Deese, 1959b; Roediger & McDermott, 1995). In these paradigms, the list structure is manipulated such that a list contains several items all highly related to a particular critical item that is not studied. A substantial literature examines the experimental factors that influence the likelihood of the participant intruding this critical item (Roediger, Watson, McDermott, & Gallo, 2001).

Kimball, Smith, and Kahana (2007) used a variant of the Search of Associative Memory model (fSAM) to examine the interaction of encoding and retrieval processes in false memory paradigms. Their work highlights a number of empirical effects consistent with the principles of CMR. The Search of Associative Memory model (SAM; Raaijmakers & Shiffrin, 1980) has three interacting components that allow it to capture much of the behavioral dynamics observed in free recall: a short-term buffer, a representation of list context, and associative structures connecting the item representations to one another and to the context representation. When an item is studied, a representation of the item is activated in a short-term buffer, which can simultaneously maintain the representations of a fixed number of studied items. While items reside in the buffer, the system creates associative structures linking the item representations to one another, as well as to the representation of list context. During memory search, first the items still active in the buffer are reported, and then the list-context representation is used to probe memory. When a particular item is recalled, the representation of that item is reactivated, and a compound cue (the item representation and list context) is used to guide the next retrieval attempt.

Kimball et al. (2007) propose that in order to adequately explain the rates of false recall in a number of experiments, it is necessary to include two mechanisms in fSAM. The first mechanism is consistent with spreading activation theories of semantic memory, in which activating a particular item's representation causes activation to spread to that item's semantic associates. Specifically, they proposed that when multiple same-category items are co-activated in the buffer, the semantic associates that are common to those items get associated to the list context, and are more likely to be falsely remembered. The second mechanism is a compound cuing process that takes place at retrieval. In classic implementations of SAM, just the most recently recalled item is used as part of the compound retrieval cue. Kimball and colleagues proposed that multiple recalled items can become part of the compound cue (along with list context). When several of these items are from

the same category, they will support the false recall of a common semantic associate (the critical item).

Broadly speaking, these two mechanisms have effects consistent with the integration mechanism of CMR. During study, the contextual representation will become more prototypical as multiple same-category items are studied. The activation of the prototypical representation will cause it to be associated with the studied items, which will make it more likely that it is retrieved during memory search. This will increase the likelihood of making a critical intrusion. During retrieval, the same integrative dynamics will cause the contextual retrieval cue to become more prototypical when multiple same-category items are recalled successively (Fig. 11). We hypothesize that neural category discriminability during recall (as measured using pattern classification) is related to the prototypicality of the contextual retrieval cue. If so, then providing CMR with information about both list structure and neural signal will allow it to estimate the likelihood of a critical intrusion during recall more accurately than a baseline model without access to these details.

The executive control of memory search

To explain a rich set of results in categorized free recall, theorists have suggested that category representations may be used in a strategic way to guide memory search (Bousfield, 1953; Cofer et al., 1966; Tulving & Pearlstone, 1966; Puff, 1966, 1974; Raaijmakers & Shiffrin, 1980; Becker & Lim, 2003). Tulving and Pearlstone (1966) proposed that categorized free recall is best described by a two-stage retrieval process, in which first the participant searches amongst the superordinate category representations, and then uses the retrieved superordinate to probe memory for the studied items from that category. Simulation of an explicit shift between memory search on different classes of materials (superordinate representations vs. item representations) would require the addition of executive processes to CMR.

Elaboration of the executive processes engaged to control memory search would allow the model to be applied to a broader range of experimental paradigms, including recall-by-category (Polyn, Erlikhman, & Kahana, 2011) and category fluency tasks (Taler, Johns, Young, Sheppard, & Jones, 2013), where the participant must constrain recalled items to come from a particular category (as opposed to a particular temporal interval in free recall). Such an extension requires consideration of two mechanisms explored in a computational theory developed by Becker and Lim

(2003): A mechanism to target items from a particular category, and a mechanism for post-retrieval monitoring. They used these mechanisms, along with a reinforcement-based learning rule, to create a model specifying the role of prefrontal neural circuitry in the executive control of memory search.

A memory targeting mechanism would facilitate the application of the model to paradigms in which participants are given a category label and asked to selectively target and retrieve items from that category (e.g. Tulving & Pearlstone, 1966; A. D. Smith, 1971; Polyn et al., 2011). The category clustering in the current version of the model, in a sense, arises spontaneously. The model tends to recall related items in sequence, but a tendency would not be enough to simulate paradigms where there is a rigid requirement that responses come exclusively from a given category. Such a targeting mechanism could involve the deliberate reactivation of a prototypical category representation, whose retrieval could be prompted by the presentation of the cue indicating the target category for that recall period. The second mechanism would involve a post-retrieval decision regarding whether or not to report a given retrieved item. Work by Lohnas et al. (submitted) explores incorporating such a mechanism into CMR, to make decisions regarding the temporal source of a recalled item. This mechanism is perhaps most critical in their simulations of the list-before-last paradigm (Shiffrin, 1970; Jang & Huber, 2008), in which a participant must target memory search not on the most recent list of studied items, but rather on the list before last. Together, these two mechanisms would allow the model to focus search on a particular category, and determine when an item inconsistent with task demands was retrieved (potentially prompting a refreshing of the prototypical category context).

Disentangling item and context representations

Perhaps the most obvious limitation of the current model is its assumption that there is no category structure to the featural representations of the studied items. In other words, each item is given a featural representation that is orthogonal to all the others. Upon presentation of an item, the semantically laden pre-experimental associations allow the model to retrieve a contextual representation containing category structure. On one hand, there is something theoretically advantageous to a model that can identify category relationships between two stimuli that are presented in such a way that they have nothing in common in terms of their perceptual characteristics. For example, a person would be able to determine that a visual stimulus depicting Jack Nicholson, and

the auditory stimulus "Robert DeNiro" both correspond to items from the same category, despite the fact that the stimuli are presented in different perceptual modalities. On the other hand, it is clear that neural representations elicited by stimuli reflect category structure at many levels of the cortical hierarchy (Haxby et al., 2001; O'Toole et al., 2005; Morton et al., 2013).

We have examined the performance of a more elaborate version of the model, where the similarity structure of both featural and contextual representations reflected category structure. However, since the simplified model (without category structure in the feature layer) is able to adequately fit the observed empirical phenomena, there is little unexplained behavioral or neural variability to support the more elaborate model. One piece of suggestive neural evidence was examined in Fig. 10, where the model's estimate regarding the category of the current item is low relative to the other categories. This is in comparison to the neural estimates which show a strong advantage for the correct category identity relative to the other categories. If the neural signal contains low-level visual category-specific features, or other information that is sensitive to taxonomic category but is not integrated into context, then this will lead to better classifier performance on the neural signal as compared to the model-generated representations.

Another consequence of the assumption that items are featurally independent of one another is that the model does not suffer from interference between items. Imagine a scenario in which two items have similar representations in the featural layer, but are presented widely spaced apart in the study list, such that each is associated with a distinct contextual state. Upon recall of either of these items, the model would reactivate a blend of the two contexts, despite the fact that the items were not studied together. Future work is necessary to determine whether this sort of interference would help or harm the model's ability to account for behavioral and neural empirical phenomena. If the model was unable to cope with such interference, the Complementary Learning Systems theory provides a potentially important mechanism in the form of pattern separation (McClelland et al., 1995). Such a mechanism would allow the model to take items that are representationally similar at the feature layer and recode them to be less similar, associating these pattern separated representations to the contextual representation.

A critical line of development for the current theory involves specifying more precisely the relationship between retrieved context and semantic knowledge. The current version of CMR proposes that a single set of associative connections allows the memory system to reactivate all of

one's prior knowledge regarding a studied item. However, this retrieved information is immediately integrated into the contextual representation. It is reasonable to think that there exist cortical regions (perhaps in ventral temporal lobe) that maintain this semantic information while the item is being considered. These high-level semantic regions would presumably project to the brain regions supporting the contextual representation itself. This would allow the participant to more effectively probe their semantic knowledge to answer some question about the item (e.g., name a movie this celebrity appeared in) before that information is blended with the other information in the contextual representation.

The future of memory modeling

In the current work, we examine the behavioral and neural consequences of allowing studied items to retrieve distributed patterns reflecting their semantic characteristics. We used behavioral data to determine the optimal parameter settings for the model, and then compared the neural predictions of these models to the observed neural data. By simultaneously considering the constraints that neural and behavioral phenomena place on cognitive theories, researchers have made important advances in a number of cognitive domains (Purcell et al., 2010; Manning et al., 2011; Davis, Love, & Preston, 2012; Polyn et al., 2012; Polyn & Sederberg, 2014; Turner et al., 2013).

The traditional approach in the domain of cognitive neuroscience has been to characterize neural phenomena using statistical techniques like general linear modeling, and then to verbally relate these statistical models to cognitive theory. One drawback of this approach is that these statistical models make strict assumptions regarding the nature of interactions between different factors. In contrast, mechanistically explicit cognitive models are much more flexible in terms of the interactions that can be examined. As an example, consider our analysis of the relationship between the model parameters controlling the structure category representations, the model parameter controlling integration, and the representational structure of context. We propose that this approach has the potential to create highly integrated theories allowing us to understand neural phenomena in terms of cognitive processes.

Parameter Type	Parameter	Description
Context Updating	β_{enc}	Rate of context drift during encoding
	β_{rec}	Rate of context drift during recall
Retrieved Context	γ^{FC}	Amount of context retrieved through experimental vs. pre-experimental associations
	γ^{CF}	Weighting of experimental vs. pre-experimental associations in cuing with context to retrieve items
Primacy	ϕ_s	Size of the learning rate boost for early items in a list
	ϕ_d	Rate of decay of learning rate gradient
Accumulator	κ	Rate of leakage of activity in each accumulator
	λ	Amount of lateral inhibition between accumulators
	η	Amount of noise input to accumulators
	τ	Time constant mapping decision competition steps into time in the experiment
Context Similarity	N	Number of context units
	σ	Noise added to category prototype to create each exemplar

Table 1. Description of the parameters of the model. σ applies only to simulations with synthetic category similarity structure.

Appendix

Formal description of the model

Here, we give a formal description of the equations that define CMR's structure and behavior. See Figure 1 for an overview of model structures and processes. Table 1 provides an overview of the parameters that control the behavior of the model.

In CMR, there are two representations: a feature layer F, and a context layer C. The feature layer is connected to the context layer through $\mathbf{M}^{FC}$, and the context layer is connected to the feature layer through $\mathbf{M}^{CF}$. Each of these weight matrices contains both pre-experimental associations and new associations learned during the experiment. Pre-experimental weights are designated $\mathbf{M}^{FC}_{\text{pre}}$ and $\mathbf{M}^{CF}_{\text{pre}}$; the experimental weights are $\mathbf{M}^{FC}_{\text{exp}}$ and $\mathbf{M}^{CF}_{\text{exp}}$.

Items are assumed to be orthonormal; each unit of F corresponds to one item. When an item i is presented during the study period, its representation on F, $\mathbf{f}_i$, is activated. Pre-experimental context $\mathbf{c}^{\text{IN}}_i$ is retrieved and is input to the context layer to update the current state of context. The input to context is

$$\mathbf{c}_i^{\text{IN}} = \mathbf{M}^{FC}\mathbf{f}_i = \mathbf{M}_{\text{pre}}^{FC}\mathbf{f}_i, \tag{1}$$

since $\mathbf{M}_{\text{exp}}^{FC}$ is assumed to be zero at the start of the list.

Previous versions of CMR have assumed that the pre-experimental context representations retrieved by items are orthonormal. Here, we assume that items that are similar to one another (e.g., in the same category) retrieve similar pre-experimental contexts. We assume for simplicity that items in different categories are associated with orthogonal pre-experimental contexts. Each category is assigned a separate group of context units. For an item in a given category, only the units corresponding to that category have nonzero activation levels; all other units are set to 0. For each category, we first generate a random prototype by drawing from the $N(0, 1)$ distribution. We then generate exemplars of that category by adding noise distributed as $N(0, \sigma_j)$, where σ_j determines the exemplar variability of category j. Each exemplar representation was normalized to have a length of 1. The distributed pattern of pre-experimental context associated with each item is stored in the corresponding column of $\mathbf{M}_{\text{pre}}^{FC}$ so that presentation of $\mathbf{f}_i$ causes activation of its corresponding pattern of pre-experimental context $\mathbf{c}_i^{\text{IN}}$.

It is unclear *a priori* what information should be in context before the start of each list. One possibility is that, for the majority of lists (all except for the first list in each session), there is still residual category information in context, left over from the previous list. To implement this idea in the model, we set the pre-list context to a combination of all three categories. Each set of category units was set to the corresponding category prototype; the vector was then normalized to have length 1. We examined an alternate initial state of context where the activation of each unit was drawn from a random normal distribution; this version of the model demonstrated poor recall for the primacy items on the list[5]. Because the context at the beginning of the list (which had no category-specific information) and the context later in the list (which always had category-specific activity) were quite different, items at the beginning of the list were not well-cued by context at the time of test. We also examined alternate mechanisms such as increasing integration rate at the start of the list, but chose to use the prototype-combination method since it requires no additional parameters and allows a satisfactory fit to the primacy effect observed in the data.

[5]This disadvantage interacted with category, and therefore could not be completely canceled out by the non-category-specific primacy gradient in learning rate discussed below.

After retrieval of pre-experimental context $\mathbf{c}_i^{\text{IN}}$, the current state of context is updated according to

$$\mathbf{c}_i = \rho_i \mathbf{c}_{i-1} + \beta_{\text{enc}} \mathbf{c}_i^{\text{IN}}, \tag{2}$$

where ρ_i is a scaling factor chosen to satisfy $||\mathbf{c}_i|| = 1$. After context is updated, the current item $\mathbf{f}_i$ and the current state of context $\mathbf{c}_i$ become associated, through simple Hebbian learning. After each item presentation, the experimental associations are updated according to

$$\Delta \mathbf{M}_{\text{exp}}^{FC} = \mathbf{c}_i \mathbf{f}_i'. \tag{3}$$

Each item is assumed to become associated to the new state of context $\mathbf{c}_i$ rather than the previous state of context $\mathbf{c}_{i-1}$. Implementations of CMR have varied in whether items are associated with the new state of context (Polyn et al., 2009) or the previous state of context (e.g., Howard & Kahana, 2002; Howard, Fotedar, Datey, & Hasselmo, 2005). In the case of orthogonal pre-experimental context representations, this choice is not important (both choices lead to identical behavior, though the parameters may be different). However, in the context of categorized free recall, the content of the context with which items become associated is an important factor. If an item is associated with context related to its category, it will provide an excellent cue for other items from the same category, causing category clustering; if an item is associated with a different category's context, it will provide a good cue for items in the other category, decreasing category clustering. If items are associated with the previous state of context, then an item presented immediately after an item from a different category will become associated with context related to the previous category, rather than the current one. In contrast, if items are associated with context after updating, then they will be associated with experimental context related to the category of that item. We chose the latter option, assuming that this version of the model would be better equipped to account for the strong category clustering apparent in the data.

When an item is presented, the network also learns associations from the current state of context to the current item, according to

$$\Delta \mathbf{M}_{\text{exp}}^{CF} = \phi_i \mathbf{f}_i \mathbf{c}_i'. \tag{4}$$

ϕ_i scales the amount of learning, simulating the increased attention to initial items in a list that has been proposed to explain the primacy effect, the recall advantage for early list items typically observed in free recall. ϕ_i depends on the serial position i of the studied item:

$$\phi_i = \phi_s e^{-\phi_d(i-1)} + 1. \tag{5}$$

The free parameters ϕ_s and ϕ_d control the magnitude and decay of the attentional boost, respectively.

There are also assumed to be pre-experimental associations between the context and feature layers. For simplicity, we assume that $\mathbf{M}^{CF}_{\text{pre}}$ is the same as the transpose of $\mathbf{M}^{FC}_{\text{pre}}$. These associations allow a "read-out" of context, making it possible to retrieve recently presented items even in the absence of new learning; this allows the model to capture the ability of amnesic patients to recall recently presented items (Sederberg et al., 2008). The pre-experimental associations could also be used to perform a free association task; the cue item would be activated on F, and allowed to retrieve pre-experimental context to update the state of context. Context would then be projected through $\mathbf{M}^{CF}_{\text{pre}}$, activating items associated with similar states of pre-experimental context.

The relative strength of experimental and pre-experimental associations is determined by the free parameters γ^{FC} and γ^{CF}:

$$\mathbf{M}^{FC} = \gamma^{FC}\mathbf{M}^{FC}_{\text{exp}} + (1-\gamma^{FC})\mathbf{M}^{FC}_{\text{pre}} \tag{6}$$

$$\mathbf{M}^{CF} = \gamma^{CF}\mathbf{M}^{CF}_{\text{exp}} + (1-\gamma^{CF})\mathbf{M}^{CF}_{\text{pre}} \tag{7}$$

At the end of a list, the current state of context $\mathbf{c}_{\text{test}}$ is used to cue for items on the list. The activation of each item $\mathbf{f}^{\text{IN}}$ is

$$\mathbf{f}^{\text{IN}} = \mathbf{M}^{CF}\mathbf{c}_{\text{test}}. \tag{8}$$

Once $\mathbf{f}^{\text{IN}}$ has been determined, there is a competition between items to determine which will be recalled. The support for each item $\mathbf{f}^{\text{IN}}$ enters into a competition of competing, leaky accumulators (Usher & McClelland, 2001; Sederberg et al., 2008), where each item corresponds to

one accumulator. The accumulators evolve according to

$$\begin{aligned} \mathbf{x}_s &= (1 - \tau\kappa - \tau\lambda\mathbf{L})\mathbf{x}_{s-1} + \tau\mathbf{f}^{\text{IN}} + \epsilon \\ \mathbf{x}_s &\rightarrow max(\mathbf{x}_s, \mathbf{0}), \end{aligned} \tag{9}$$

where each element of the vector $\mathbf{x}_s$ corresponds to a studied item. $\mathbf{x}$ is initialized to $\mathbf{0}$. It is updated on each step s of the accumulation process, until one of the accumulating elements crosses a threshold (which is set at one), or the recall period is over. Each accumulator is constrained to always have a positive activation. κ determines the "leakage" of each unit, that is, the rate at which each accumulator decays. λ controls the strength of lateral inhibition by scaling an inhibitory matrix $\mathbf{L}$, which connects each accumulator to every other accumulator. ϵ is a normally distributed random vector, where each element has mean zero and standard deviation η. Finally, τ is a time constant determining the rate of the accumulation process.

Items that have already been recalled still take part in the competition. However, if a previously recalled item reaches the threshold, it is not recalled, and the activity of the accumulator is set to 95% of the threshold. When the accumulator corresponding to an item that has not previously been recalled reaches the threshold, it is reactivated on F. The reactivated item is then used to retrieve both experimental and pre-experimental context, according to

$$\mathbf{c}_i^{\text{IN}} = \mathbf{M}^{FC}\mathbf{f}_i. \tag{10}$$

Context is then updated according to

$$\mathbf{c}_i = \rho_i\mathbf{c}_{i-1} + \beta_{\text{rec}}\mathbf{c}_i^{\text{IN}}, \tag{11}$$

and used as a cue for another recall attempt.

Serial position, list length, and contiguity effects

In the retrieved-context framework, pre-experimental associations link each item representation with a particular contextual state. Prior work examining the dynamics of retrieved-context models in free recall has assumed that these contextual states were unrelated to one another. That

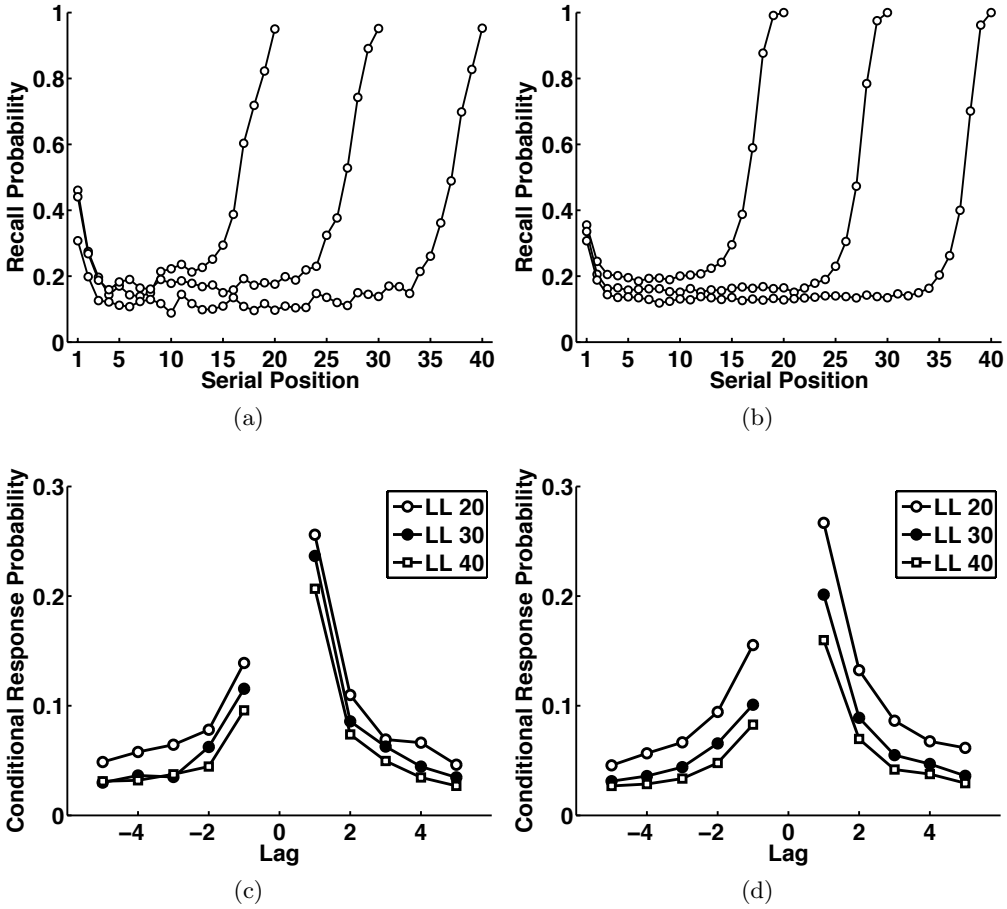


Figure 18. Data and simulation results for Murdock (1962). (**a**) Recall probability as a function of serial position, for conditions with list length 20, 30, and 40 and presentation time of 1 s. (**b**) Recall probability by serial position, for the best-fitting model. (**c**) Conditional response probability as a function of lag. (**d**) Conditional response probability as a function of lag, for the best-fitting model.

is, each item retrieved a contextual state that was orthogonal to the contextual states retrieved by any of the other items. Our extended version of CMR assumes that the contextual state retrieved by an item has representational structure that reflects the semantic relations between the items (Rao & Howard, 2008; Howard et al., 2011). Altering the representational structure of pre-experimental contextual states has the potential to drastically alter the dynamics of the model. The relative likelihood of recalling a particular study item depends upon the similarity between the state of context being used as a cue, and the state of context that was associated with that item during the study period. Thus, altering the similarity structure of contextual states will alter the dynamics of memory search. Here, we demonstrate that the altered model is able to account for effects of serial position on recall (including primacy and recency), changes in recall performance due to list

length, and the effect of temporal contiguity on recall transition probabilities, as observed in several conditions of an experiment reported by Murdock (1962).

A previous version of CMR showed that the model could account for the benchmark phenomena from Murdock (1962), as well as the simultaneous presence of semantic and temporal organization in recall sequences (Polyn et al., 2009). This prior version of the model allowed items to retrieve distinct pre-experimental contextual states; semantic structure arose from associative connections linking these distinct contextual states back to the feature layer of the model, fueling a decision competition. In the current version of the model, we build semantic structure into the contextual representations themselves, which causes the model to lose a degree of freedom. The prior version of the model had a free parameter that could scale the strength of the semantic associative structure without influencing temporal organization. However, in the current version of the model, the pre-experimental associations associating item representations to contextual states are doing double-duty. They contain semantic structure, and they drive the contextual evolution responsible for temporal organizational effects. A more detailed explanation of this intertwining of semantic and temporal dynamics in the model is presented in *Appendix: Formal Description of the Model*.

We demonstrate that these alterations do not affect the ability of the model to account for the classic data of Murdock (1962), including temporal organization and effects of serial position and list length. We focused on the conditions involving presentation of words for 1 s each, and examined recall performance on lists with 20, 30, and 40 words. While we were unable to obtain the original word pool used by Murdock (1962), we used the word association spaces (WAS; Steyvers, Shiffrin, & Nelson, 2004) model of semantic structure to create semantic representations for a comparable set of high-frequency concrete nouns (following the methodological details reported by Murdock). These 400-dimensional representations are derived from a singular value decomposition analysis of free-association norms (Nelson, McEvoy, & Schreiber, 2004). We assumed that, as in previous implementations of CMR, that item representations on the feature layer are orthogonal. In contrast, the context layer contains 400 units; changing states of context in this model may then be thought of as moving through the vector space defined by the WAS solution (see Socher et al., 2009 for a similar interpretation of contextual evolution).

We used a differential evolution search to determine the parameters that best fit the data,

including serial position curves and conditional response probability as a function of lag (Kahana, 1996; see *Appendix: Parameter Searches* for details of the parameter search). Figure 18 shows the serial position and conditional response probability as a function of lag curves for the data and the best-fitting model. The addition of WAS-based similarity structure to the pre-experimental context assumed by the model does not change the model's ability to account for the major effects observed by Murdock (1962), including primacy, recency, list-length effects on pre-recency recall, and temporal contiguity (cf. Polyn et al., 2009). This suggests that many of the basic effects predicted by the model are observed regardless of the specific form of contextual inputs used to drive context evolution, rather than depending on the common assumption that items are associated with orthogonal pre-experimental contexts.

Parameter searches

To determine the best-fitting parameters for each simulated experiment, we used a version of the differential evolution algorithm (Storn, 2008) to find the parameter set that minimized χ^2 error or RMSD, based on the summary statistics of interest. We used a MATLAB-based implementation of differential evolution, based on code developed by Price, Storn, and Lampinen (2005)[6]. Mutation was done using the local-to-best strategy, with a step weight of 0.85 and crossover probability of 1. We found that the best fitness often did not change for several iterations, before eventually decreasing. Therefore, we determined convergence by examining the median fitness value over all parameter sets.

Murdock 1962 simulation.

In order to establish whether the distributed-context variant of CMR can still account for benchmark free recall data, we simulated the classic experiment by Murdock (1962). We simulated the 1 s presentation time conditions from that study, where participants performed free recall of lists of length 20, 30, or 40. We minimized χ^2 error for the serial position curve and conditional response probability as a function of lag ($-5 <= \text{lag} <= 5$), for each of the three conditions. When calculating χ^2 error, each curve from each condition was weighted equally. We first evaluated 5000 randomly chosen points in parameter space, using a single simulation of the experiment. We took the best-fitting 50 individuals from this, and then used differential evolution to find the best-fitting

[6]Thanks to Joshua McCluey for adapting this code for parallel execution.

Parameter Type	Parameter	M62	P66	Similarity	Integration	Learning Rate
Context Updating	β_{enc}	0.82	0.31	0.82	—	0.87
	β_{rec}	0.35	0.98	0.55	0.52	0.53
Retrieved Context	γ^{FC}	0.28	0.99	0.43	0.59	0.62
	γ^{CF}	0.60	0.16	0.26	0.30	0.08
Primacy	ϕ_s	3.58	(3.58)	1.18	0.51	0.19
	ϕ_d	2.29	(2.29)	3.53	1.19	1.09
Accumulator	κ	0.54	(0.54)	0.57	0.41	0.31
	λ	0.01	0.00	0.04	0.87	0.17
	η	0.26	(0.26)	0.33	0.29	0.46
	τ	0.17	0.03	0.45	0.85	0.16
Context Similarity	N	(400)	(60)	48	69	33
	σ	—	0.12	—	2.28	4.67
	Data points	120	11	124	124	124
	No. free par.	10	7	14	14	15
	χ^2	641.69	—	326.84	422.18	335.16
	χ^2 *(weighted)*	533.00	—	232.18	262.11	258.98
	RMSD	0.0430	0.4013	0.0495	0.0587	0.0481
	BIC (weighted)	-746	-14.4	-754	-738	-751

Table 2. Best-fitting values of free non-category-specific parameters for the CMR simulation analyses. Parameters shown in parentheses were fixed to that value; other parameters were free to vary. M62: Murdock (1962); P66: Puff (1966). Similarity, Integration, and Learning Rate are model variants used to simulate Morton et al. (2013).

parameter values (as described above). For each individual, we evaluated fitness based on two replications of the experiment. Once the best-fitting parameters were determined, we carried out a final simulation with four replications of the experiment; χ^2 error and analysis of model predictions were based on this final simulation. Best-fitting parameters are shown in Table 2.

Simulation Analysis I.

We examined the effects of manipulating stimulus list organization by simulating the study reported by Puff (1966). For each of the four conditions (input category repetitions, or C-Reps, of 0,

	Similarity (σ)	Integration (β_{enc})	Learning Rate (L^{CF})
Celebrity	4.65	0.77	0.74
Location	5.58	0.76	0.66
Object	7.71	0.70	0.60

Table 3. Best-fitting values of category-specific parameters for model variants used to simulate Morton et al. in press.

9, 18, or 27), we generated random lists with the specified number of C-Reps. For each condition, we determined all possible groupings of the items in each category. For example, consider the condition with 9 C-Reps. In this case, each category must have 3 C-Reps. To meet this condition, a given category could have: a single group of four category items in succession, with the other six items presented adjacent to other category items; or one group of three items, and one group of two items; or three groups with two items each. For each condition, each possible grouping was sampled randomly to create 1500 random lists (i.e., 100 replications of their experiment).

We analyzed the simulated recall sequences in the same manner described by Puff (1966). Category repetitions during recall were calculated, as well as the number of repetitions expected due to chance, based on the number of recalls. Expected repetitions were calculated as

$$\mathrm{E(C-Reps)} = \frac{m_1^2 + m_2^2 + m_3^2}{n} - 1, \tag{12}$$

where m_1, m_2, and m_3 are the number of words recalled from categories 1, 2, and 3, and n is the total number of words recalled. The expected repetitions were subtracted from the actual repetitions to obtain repetitions beyond chance. We calculated serial repetitions as the mean number of forward adjacent transitions between items in the same category; serial repetitions were impossible in the 0 C-Reps condition. We optimized model parameters to minimize error in fitting overall recall percentage, C-Reps beyond chance, and serial repetitions, for each of the spacing conditions (except serial repetitions in the 0 C-Reps condition). All data points were weighted equally. No measure of variability in the data was available for the measures reported by Puff (1966), so we minimized RMSD instead of χ^2 error. We used differential evolution with 50 individuals to find the best-fitting parameter values. We selected the best parameters based on the search and ran a final simulation; this was used to determine model predictions and RMSD. Best-fitting parameters are shown in Table 2.

Simulation Analysis II.

For fitting group statistics in Simulation Analysis II, we used a number of measures of recall behavior. We included the serial position curve, probability of first recall curve (last 3 points only), and conditional response probability as a function of lag (conditional on within-category

transitions[7]; $-5 <= \text{lag} <= 5$), each calculated separately for each category. We also included conditional response probability as a function of lag, conditional on the transition being between items of different categories. Finally, we also fit category clustering using the probability of making a within-category transition, conditional the category of the previous recall.

In calculating the fitness value associated with a given parameter set, we weighted the contribution of each data point in order to emphasize the importance of certain measures of interest, which in certain cases were represented by only a few numbers (e.g. category clustering, a measure of great importance to us, comprised only 3 data points out of 124). We weighted each curve equally, except for category clustering, which was weighted to be 4 times more important than the other curves. This was done in order to force each model variant to fit category clustering at the expense of other measures. For each model variant, we first evaluated 5000 randomly chosen points in parameter space, simulating 15 replications of the study. We then selected the 50 best-fitting individuals, which were used as the initial values for a differential evolution search. Each individual was evaluated based on 20 simulated replications of the study. After the search converged, we ran 40 replications of the best-fitting parameter set from the search; this final simulation was used to determine χ^2 and summary statistics such as the serial position curve. Table 2 presents both weighted and unweighted χ^2 values for each parameter search, as well as the best-fitting values for parameters that were common to all the model variants. Table 3 shows best-fitting values for variant-specific parameters.

Simulation Analysis IV.

To account for individual differences in recall behavior, we evaluated a series of model variants, with progressively more parameters allowed to vary between participants. For all model variants, parameters that were not varied by individual were set to the best-fitting group parameters from the Similarity variant of Simulation Analysis II.

Since the fit to each participant is based on fewer data points than the group fit, we summarized certain statistics to reduce variability due to noise. We fit the serial position curve, the probability of initiating recall with each of the last 3 serial positions, and conditional response

[7]This analysis is similar to that reported by Polyn et al. (2011). However, we examined conditional response probability as a function of lag, based on the raw serial position of each item. This contrasts with their analysis, which ignored items not included in the analysis when calculating lag.

Parameter	M1	M2	M3	M4
β_{enc}	0.82 (0.02)	0.83 (0.01)	0.82 (0.01)	0.80 (0.02)
σ_c	—	5.75 (0.49)	4.98 (0.37)	4.92 (0.46)
σ_l	—	6.71 (0.47)	5.70 (0.40)	6.06 (0.47)
σ_o	—	6.74 (0.40)	7.28 (0.39)	6.93 (0.44)
β_{rec}	—	—	0.54 (0.03)	0.55 (0.03)
γ^{FC}	—	—	—	0.93 (0.12)
γ^{CF}	—	—	—	0.48 (0.09)
Data points	1537	1537	1537	1537
No. free par.	29	116	145	203
RMSD	0.0946	0.0947	0.0973	0.0899
RMSD (weighted)	0.0933	0.0897	0.0873	0.0746
RMSD (cat. clustering)	0.1026	0.0731	0.0601	0.0394
BIC (weighted)	-7109	-6682	-6585	-6706

Table 4. Individual participant fits for the Morton et al. (2013) study. Each of the four model variants allows a different number of parameters free to vary between participants. Dashes indicate that this parameter was set to the value defined by the group fit reported in Table 2. For the parameters allowed to vary across individuals, the mean value of the parameter is shown, with the standard error of the mean across participants shown in parentheses. RMSD (cat. clustering) shows the RMSD just for the values of $P(within)$ for each category.

probability as a function of lag (separately for within- and between-category transitions; only lags from -5 to +5 were included). Each of these measures was collapsed over category in order to obtain adequate sample sizes. In addition, we examined recall probability as a function of category, and probability of making a within-category transition, separately for each category. Each of the curves described above were weighted equally (regardless of the number of data points in the curve).

As for the previous models, the best-fitting parameters were determined using a differential evolution search. Parameters were estimated for each participant separately. For each participant, we first evaluated 50 randomly chosen individuals in the parameter space being searched (other parameters were fixed to the best-fitting group parameters throughout), with 10 replications of the trials they ran in the actual experiment. Once the fit appeared to be converged for all participants, the number of replications was increased to 20. We then ran more generations until the search was again converged for all participants. Finally, 80 replications of each participant's trials were simulated using their best-fitting parameters determined from the search; results presented are based on this final simulation.

References

Anderson, J. A. (1972). A simple neural network generating an interactive memory. *Mathematical Biosciences*, *14*, 197–220.

Awipi, T., & Davachi, L. (2008). Content-specific source encoding in human medial temporal lobe. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(4), 769–779.

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.

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

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* (pp. 85–121). New York: John Wiley and Sons.

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, B. H. (1963). An investigation of recoding in free recall. *Journal of Experimental Psychology*, *65*(4), 368-376.

Cohen, B. H., Bousfield, W. A., & Whitmarsh, G. A. (1957). Cultural norms for verbal items in 43 categories. (Technical Report No. 22).

D'Agostino, P. R. (1969). The blocked-random effect in recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, *8*, 815–820.

Dallet, K. M. (1964). Number of categories and category information in free recall. *Journal of Experimental Psychology*, *68*(1), 1–12.

Danker, J. F., & Anderson, J. R. (2010). The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. *Psychological Bulletin*, *136*(1), 87–102.

Davis, T., Love, B. C., & Preston, A. R. (2012). Learning the exception to the rule: Model-based fMRI reveals specialized representations for surprising category members. *Cerebral Cortex*, *22*(2), 260–273.

Deese, J. (1959a). Influence of inter-item associative strength upon immediate free recall. *Psychological Reports*, *5*, 305–312.

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

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Estes, W. K. (1950). Toward a statistical theory of learning. *Psychological Review*, *57*, 94–107.

Ezzyat, Y., & Davachi, L. (2014, March). Similarity breeds proximity: Pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. *Neuron*, *81*(5), 1179–1189.

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

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425–2429.

Haynes, J. D., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience*, *8*(5), 686–691.

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review*, *119*(2), 431–440.

Hintzman, D. L. (1986). 'Schema abstraction' in a multiple-trace memory model. *Psychological Review*, *93*, 411–428.

Howard, M. W. (2004). Scaling behavior in the temporal context model. *Journal of Mathematical Psychology*, *48*, 230–238.

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

Howard, M. W., Jing, B., Rao, V. A., Provyn, J. P., & Datey, A. V. (2009). Bridging the gap: Transitive associations between items presented in similar temporal contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 391–407.

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

Howard, M. W., Kahana, M. J., & Sederberg, P. B. (2008). Postscript: Distinguishing between temporal context and short-term store. *Psychological Review*, *115*(4), 1125–1126.

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., Shankar, K. H., & Jagadisan, U. K. K. (2011). Constructing semantic representations from a gradually changing representation of temporal context. *Topics in Cognitive Science*, *3*(1), 48–73.

Howard, M. W., Viskontas, I. V., Shankar, K. H., & Fried, I. (2012). Ensembles of human MTL neurons "jump back in time" in response to a repeated stimulus. *Hippocampus*, *22*(9), 1833–1847.

Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A continuous semantic space describes the

representation of thousands of object and action categories across the human brain. *Neuron*, *76*(6), 1210–1224.

Jang, Y., & Huber, D. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 112–127.

Jenkins, L. J., & Ranganath, C. (2010). Prefrontal and medial temporal lobe activity at encoding predicts temporal context memory. *Journal of Neuroscience*, *30*(46), 15558–15565.

Jones, M. N., & Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, *114*(1), 1–37. doi: 10.1037/0033-295X.114.1.1

Jonker, T. R., Seli, P., & MacLeod, C. M. (2013). Putting retrieval-induced forgetting in context: An inhibition-free, context-based account. *Psychological Review*, *120*(4), 852–872.

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

Kahana, M. J. (2012). *Foundations of human memory* (1st ed.). New York, NY: Oxford University Press.

Kahana, M. J., Zhou, F., Geller, A. S., & Sekuler, R. (2007). Lure-similarity affects visual episodic recognition: Detailed tests of a noisy examplar model. *Memory & Cognition*, *35*, 1222–1232.

Kimball, D. R., Bjork, E. L., Bjork, R. A., & Smith, T. A. (2008). Part-list cuing and the dynamics of false recall. *Psychonomic Bulletin & Review*(15), 296–301.

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

Kreiman, G., Koch, C., & Fried, I. (2000). Category-specific visual responses of single neurons in the human medial temporal lobe. *Nature Neuroscience*, *3*, 946–953.

Kriegeskorte, N., Mur, M., & Bandettini, P. A. (2008). Representational similarity analysis—connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 1-28.

Kriegeskorte, N., Mur, M., Ruff, D. A., Kiani, R., Bodurka, J., Esteky, H., ... Bandettini, P. A. (2008). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, *60*, 1126–1141.

Kuhl, B. A., Dudukovic, N. M., Kahn, I., & Wagner, A. D. (2007). Decreased demands on cognitive control reveal the neural processing benefits of forgetting. *Nat Neurosci*, *10*(7), 908–914. Retrieved from `http://dx.doi.org/10.1038/nn1918` doi: 10.1038/nn1918

Kuhl, B. A., Rissman, J., Chun, M. M., & Wagner, A. D. (2011). Fidelity of neural reactivation reveals competition between memories. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(14), 5903–5908.

Kuhl, B. A., Rissman, J., & Wagner, A. D. (2012). Multi-voxel patterns of visual category representation

during episodic encoding are predictive of subsequent memory. *Neuropsychologia*, *50*, 458-469.

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.

Lewis-Peacock, J. A., & Postle, B. R. (2008). Temporary activation of long-term memory supports working memory. *Journal of Neuroscience*, *28*(35), 8765–8771.

Logan, G., & Gordon, R. (2001). Executive control of visual attention in dual-task situations. *Psychological Review*, *108*(2), 393–434.

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (submitted). Expanding the scope of memory search: Modeling intralist and interlist effects in free recall.

Manning, J. R., Polyn, S. M., Baltuch, G., Litt, B., & Kahana, M. J. (2011). Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(31), 12893–12897.

Manning, J. R., Sperling, M. R., Sharan, A., Rosenberg, E. A., & Kahana, M. J. (2012). Spontaneously reactivated patterns in frontal and temporal lobe predict semantic clustering during memory search. *Journal of Neuroscience*, *32*(26), 8871–8878.

Manns, J. R., Howard, M. W., & Eichenbaum, H. (2007). Gradual changes in hippocampal activity support remembering the order of events. *Neuron*, *56*(3), 530–540. doi: 10.1016/j.neuron.2007.08.017

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.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K. M., Malave, V. L., Mason, R. A., & Just, M. A. (2008). Predicting human brain activity associated with the meanings of nouns. *Science*, *320*(5880), 1191.

Morton, N. W., Kahana, M. J., Rosenberg, E. A., Baltuch, G. H., Litt, B., Sharan, A. D., ... Polyn, S. M. (2013). Category-specific neural oscillations predict recall organization during memory search. *Cerebral Cortex*, *23*(10), 2407–2422.

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

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*, 609–626.

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.

Norman, K. A., Newman, E., Detre, G., & Polyn, S. M. (2006). How inhibitory oscillations can train neural networks and punish competitors. *Neural Computation*, *18*, 1577–1610.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning*. University of Illinois Press.

O'Toole, A. J., Jiang, F., Abdi, H., & Haxby, J. V. (2005). Partially distributed representations of objects and faces in ventral temporal cortex. *Journal of Cognitive Neuroscience*, *17*(4), 580–590.

Polyn, S. M., Erlikhman, G., & Kahana, M. J. (2011). Semantic cuing and the scale-insensitivity of recency and contiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 766–775.

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., Kragel, J. E., Morton, N. W., McCluey, J. D., & Cohen, Z. D. (2012). The neural dynamics of task context in free recall. *Neuropsychologia*, *50*(4), 447–457.

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. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156.

Polyn, S. M., & Sederberg, P. B. (2014). Brain rhythms in mental time travel. *NeuroImage*, *85*(2), 678-684.

Price, K., Storn, R. M., & Lampinen, J. A. (2005). *Differential evolution: A practical approach to global optimization*. Springer.

Puff, C. R. (1966). Clustering as a function of the sequential organization of stimulus word lists. *Journal of Verbal Learning and Verbal Behavior*, *5*, 503-506.

Puff, C. R. (1974). A consolidated theoretical view of stimulus-list organization effects in free recall. *Psychological Reports*, *34*(1), 275-288.

Puff, C. R., Murphy, M. D., & Ferrara, R. A. (1977). Further evidence about the role of clustering in free recall. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(6), 742–753.

Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2010). Neurally constrained modeling of perceptual decision making. *Psychological Review*, *117*(4), 1113–1143.

Quiroga, R. Q., Reddy, L., Kreiman, G., Koch, C., & Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, *435*(23), 1102–1107.

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). Retrieved context and the discovery of semantic structure. In J. C. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems* (p. 1193-1200). Cambridge, MA: MIT Press.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*, 803-814.

Roediger, H. L., Watson, J., McDermott, K. B., & Gallo, D. A. (2001). Factors that determine false recall: A multiple regression analysis. *Psychonomic Bulletin & Review*, *8*(3), 385-407.

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, *76*(1), 45–48.

Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel distributed processing.* MIT Press.

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.

Schweickert, R. (1978). Separable effects of factors on speed and accuracy: Memory scanning, lexical decision, and choice tasks. *Psychological Bulletin*, *97 (3)*, 530-546.

Sederberg, P. B., Gershman, S. J., Polyn, S. M., & Norman, K. A. (2011). Human memory consolidation can be explained using the temporal context model. *Psychonomic Bulletin & Review*, *18*, 455–468.

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, *115*(4), 893–912.

Shankar, K. H., & Howard, M. W. (2010, dec). Timing using temporal context. *Brain Research*, *1365*, 3–17.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, *237*, 1317-1323.

Shiffrin, R. M. (1970). Forgetting: Trace erosion or retrieval failure? *Science*, *168*, 1601-1603.

Smith, A. D. (1971). Output interference and organized recall from long-term memory. *Journal of Verbal Learning and Verbal Behavior*, *10*(4), 400-408.

Smith, S. M. (1988). Environmental context-dependent memory. In G. M. Davies & D. M. Thomson (Eds.), *Memory in context: Context in memory.* (pp. 13–34). Oxford, England: John Wiley & Sons.

Socher, R., Gershman, S. J., Perotte, A. J., Sederberg, P. B., Blei, D. M., & Norman, K. A. (2009). A

bayesian analysis of dynamics in free recall. In Y. Bengio, D. Schuurmans, J. Lafferty, C. K. I. Williams, & A. Culotta (Eds.), *Advances in neural information processing systems.* MIT Press.

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2004). Word association spaces for predicting semantic similarity effects in episodic memory. In A. F. Healy (Ed.), *Cognitive psychology and its applications: Festschrift in honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer.* (pp. 237–249). Washington, DC: American Psychological Association.

Storn, R. (2008). Differential evolution research—Trends and open questions. In U. K. Chakraborty (Ed.), *Advances in differential evolution* (pp. 1–31). Heidelberg, Germany: Springer Berlin.

Stricker, J. L., Brown, G. G., Wixted, J. T., Baldo, J. V., & Delis, D. C. (2002). New semantic and serial clustering indices for the California Verbal Learning Test–Second Edition: Background, rationale, and formulae. *Journal of the International Neuropsychological Society*, *8*, 425–435.

Taler, V., Johns, B. T., Young, K., Sheppard, C., & Jones, M. N. (2013). A computational analysis of semantic structure in bilingual verbal fluency performance. *Journal of Memory and Language*, *69*(4), 607–618.

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory.* (p. 381-403). New York: Adademic Press.

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

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, *5*, 381-391.

Turner, B. M., Forstmann, B. U., Wagenmakers, E.-J., Brown, S. D., Sederberg, P. B., & Steyvers, M. (2013). A Bayesian framework for simultaneously modeling neural and behavioral data. *NeuroImage*, *72*, 193–206.

Underwood, B. J. (1969). Attributes of memory. *Psychological Review*, *76*(6), 559–573.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*(3), 550-592.

Xue, G., Dong, Q., Chen, C., Lu, Z., Mumford, J. A., & Poldrack, R. A. (2010). Greater neural pattern similarity across repetitions is associated with better memory. *Science*, *330*, 97–101.

Yntema, D. B., & Trask, F. P. (1963). Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, *2*, 65-74.