Developing a Neurocognitive Model of Temporal and Semantic Organization of Memory Search

By

Neal W Morton

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

December, 2014

Nashville, Tennessee

Approved:

Sean M. Polyn, Ph.D.

Gordon Logan, Ph.D.

Geoffrey F. Woodman, Ph.D.

Brandon A. Ally, Ph.D.

ACKNOWLEDGMENTS

This work was supported by the NSF (1157432) and a Vanderbilt University Discovery Grant.

Thanks to Cage Spoden, Jonathan Miller, John Burke, Matthew Mollison, Emily Rosenberg, and Richard Lawrence for help with data collection. Thanks to the members of my dissertation committee, Gordon Logan, Geoff Woodman, and Brandon Ally, for their valuable feedback on these projects. Thanks to the computational modeling community at Vanderbilt, which has provided helpful instruction and feedback on my modeling projects, especially Tom Palmeri and Mike Pratte. Thanks also to Andy Tomarken for advice on statistics. Thanks to Jim Kragel and Josh McCluey for many helpful discussions, and all the members of the Vanderbilt Computational Memory Lab. Thanks also to Michael Kahana and the Penn Computational Memory Lab for their continued support.

Thanks to my father, Jerald, for teaching me about physics and encouraging my interest in science, and my mother, Kris, for encouraging my initial foray into psychology research. Thanks to J'Nese Williams for her love and support through the home stretch.

A special thanks goes to Sean Polyn, for your example as a scientist, family man, and human being. Thank you for believing in me when I was a new psychology student, working an unpaid Summer internship in the Kahana lab, and helping me past many hurdles in the eight years since.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

Introduction

Episodic and semantic memory

Tulving (1972) proposed a distinction between two forms of memory, episodic and semantic memory. He described episodic memory as being inherently autobiographical, and tied to a specific time and place. In contrast, semantic memory was proposed to include general knowledge that is not tied to a particular context. Tulving (1972) noted that these different forms of memory interact, with existing semantic knowledge impacting encoding of episodic memories. Research in the free recall paradigm, where participants study a list, then are asked to recall items from the list in any order, has provided a rich set of findings relating to how semantic knowledge influences formation and expression of episodic memories. In the work presented here, we examine the cognitive mechanisms and neural substrates by which semantic knowledge influences memory search.

Organization of memory search

Models of episodic memory search have been strongly influenced by the free-recall paradigm. The relatively unconstrained nature of free recall allows researchers to examine the detailed dynamics of memory search. Participants tend to successively recall items that were studied near to each other in the list (Kahana, 1996; Howard and Kahana, 1999); this tendency is referred to as temporal organization. Researchers have also long noted that semantic associations between list items can have a strong influence on the order in which items are recalled (Bousfield, 1953; Cohen, 1963). When a studied list contains items drawn from distinct taxonomic categories, participants show a strong tendency to group together items from the same category during recall, a tendency known as *category clustering* (Puff, 1974). Category clustering is observed even when same-category items are never presented adjacent to each other in the list (Bousfield, 1953). Semantic organization of free recall is also observed when there is no systematic semantic structure to the studied list

(Romney et al., 1993; Howard and Kahana, 2002b). Although many behavioral investigations have focused on semantic organization, the findings are not easily systematized (Howard et al., 2007), and there is little consensus about the cognitive mechanisms involved in mediating the relationship between semantic associations and recall organization (e.g. Anderson 1974; Sirotin et al. 2005; Kimball et al. 2007; Polyn et al. 2009). One of the difficulties in understanding semantic organization lies in determining whether organization occurs during encoding or retrieval (e.g. Stricker et al. 2002). Recent results from neurorecording studies suggest that neural data may help to disambiguate semantic information active during encoding and retrieval (Kuhl et al., 2012; Morton et al., 2013).

Neural correlates of stimulus category

Research in neuroscience has identified neural signals that may be used to track the contents of memory. Early fMRI research found evidence of brain areas that respond selectively to faces (fusiform face area: Kanwisher et al. 1997) and places (parahippocampal place area: Epstein and Kanwisher 1998). Subsequent research with multi-variate pattern analysis (MVPA; Norman et al. 2006) revealed that category-sensitive signals are not limited to these areas, but rather are widely distributed in ventral temporal cortex (Haxby et al., 2001). Polyn et al. (2005) used MVPA of fMRI data to identify category-specific brain activity during free recall of categorized materials. They found that brain activity in the seconds leading up to a successful recall were predictive of the category of the item about to be recalled, demonstrating that brain activity could provide a window into the process of memory search. Using MVPA of electrocorticographic (ECoG) recordings, Morton et al. (2013) demonstrated that widespread brain regions exhibit category-specific oscillatory activity. They found that items associated with strong category-specific activity at temporal cortex electrodes were more likely to subsequently be recalled during free recall. In a scalp EEG experiment, they also showed that items associated with high-fidelity category-specific activity were more likely to be recalled as part of a cluster of same category items. Furthermore, they found evidence that oscillatory activity is sensitive not only to the item currently being viewed, but also is impacted by recently presented items: As multiple items from a category are studied in succession, category-specific activity gradually becomes stronger. The rate of this increase in category-specific activity correlates with individual differences in category clustering during recall, suggesting that this integrative activity during encoding may influence the organization of stimuli in memory. Finally, category-specific activity during recall increased during periods of increased category organization, possibly reflecting the deployment of a category-specific retrieval cue during search. We discuss these findings in more detail in Chapter II.

An important goal of cognitive neuroscience research is to understand the cognitive mechanisms being implemented by the brain. Behavioral research has yielded sophisticated computational models of memory, which describe sets of interacting cognitive mechanisms and representations that can account for many regularities in behavior (e.g. Raaijmakers and Shiffrin 1980; Howard and Kahana 2002a; Davelaar et al. 2005; Sederberg et al. 2008; Polyn et al. 2009; Farrell 2012). However, many questions remain about whether and how the cognitive mechanisms described by these models are implemented in the brain. A goal of this dissertation is to improve our understanding of how distributed neural signals relate to the cognitive mechanisms involved in memory search. We focused on the class of retrieved-context models of memory search, which have been shown to account for a wide range of influences on free recall behavior (Sederberg et al., 2008; Polyn et al., 2009). We used this framework to begin development of a model capable of bridging between the cognitive and neural domains of research on memory search.

Retrieved-context models

Retrieved-context models were originally developed to explain recency and contiguity effects observed in a range of free-recall paradigms (Howard and Kahana, 2002a; Sederberg et al., 2008). The retrieved-context framework has also proven useful for understanding spatial navigation (Howard et al., 2005), transitive inference (Howard et al., 2005; Rao and Howard, 2008), and memory changes with aging (Howard et al., 2006b). Here I give a brief overview of retrieved-context models; see Chapter II for a more detailed discussion. The retrieved-context modeling framework proposes that memory search is driven by a representation of temporal context, which serves as a cue to target specific temporal intervals. There are two layers in the model: an item feature layer

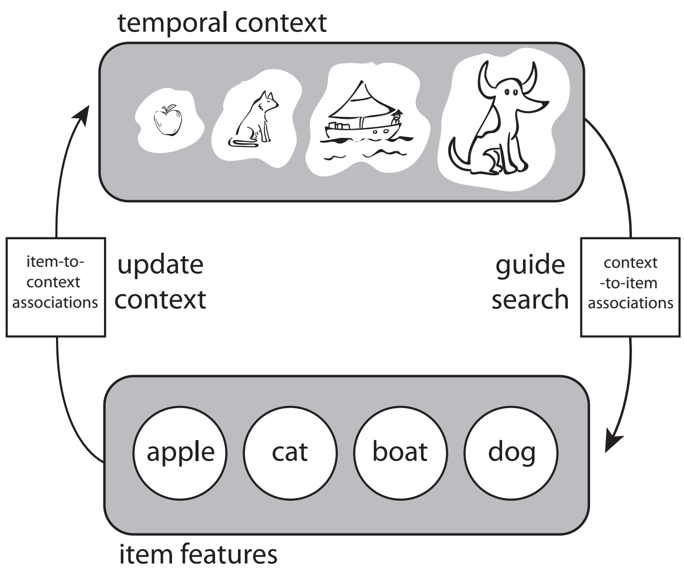


Figure 1. Schematic of retrieved-context models. Presentation of a studied item activates a unit on the item feature layer. This causes pre-experimental context associated with that item to be integrated into the current state of the temporal context layer. This continues for each item on a list, causing temporal context to change slowly over time. Each item becomes associated with the context active when it was presented. During memory search, the current state of context serves as a cue for items associated with similar contexts. Retrieving an item also causes retrieval of associated context, which updates the current context and alters the retrieval cue, giving rise to temporal organization. Illustration by Sean Polyn.

and a temporal context layer (Fig. 1). When a studied item is presented, it becomes activated on the item layer. This triggers retrieval of pre-experimental context associated with that item; for example, presentation of BOAT might cause associated things to come to mind, such as one's previous experiences traveling by boat and general information about boats and how they work. The context retrieved by the item then updates a representation of the current temporal context. As each item in a to-be-remembered list is presented, temporal context changes slowly. At any moment, context most strongly reflects information related to the most recent item, but also contains some information about other recently presented items. Each item becomes associated with the state of context that was active when it was presented; items can serve as a cue to retrieve associated context, and context can serve as a cue to retrieve associated items. During free recall, the current state of context is used to probe memory. Items are well-cued by a given state of context if they were associated with a similar state of context. For the first recall, context will provide a good cue for items presented near the end of the list, giving rise to the recency effect. When an item is recalled, the context associated with that item during encoding is reinstated. This reinstated context provides a good cue for items presented near to the recalled item, giving rise to temporal organization (Howard and Kahana, 2002a).

The context maintenance and retrieval model (CMR) is a retrieved-context model that extended the framework to explain effects of source contexts, such as encoding task, as well as effects of semantic associations (Polyn et al., 2009). CMR assumes that context-to-item associations reflect long-standing semantic associations between items; for example, activation of context associated with CAT would provide a strong cue for retrieval of DOG. These associations allow the model to account for the finding of semantic clustering during free recall. While CMR has been shown to account for the general finding of semantic organization (Polyn et al., 2009), this was not the main emphasis of the model, and it is unclear whether it can account for the wealth of findings from classic studies of categorized free recall, as well as recent neural findings.

Overview

In the present work we extend the CMR framework to account for recent findings about brain activity observed during study that predict subsequent semantic organization (Chapter II), test a critical prediction of this extended framework (Chapter III), and use the framework to test competing models of semantic associations and mechanisms that may be involved in memory cuing (Chapter IV).

In Chapter II, we focus on the scalp EEG study reported by Morton et al. (2013). This study, which included measurements of distributed category-specific activity, as well as detailed measures of recall behavior, provides useful constraints for developing a neurocognitive model of memory search. We developed a simple extension of CMR that proposes that the observation of persistent category-specific activity during study is related to the operation of a gradually evolving context representation, which is later used to guide memory search. We demonstrate that this model can simultaneously account for neural and behavioral data in the study. We also find that the model can account for the classic finding in the categorized free recall literature that category clustering is stronger when items in a category are presented in blocked, rather than interleaved, order (Puff, 1966).

The findings of Morton et al. (2013) established a link between persistent category-specific activity during encoding and subsequent organization by category; in Chapter III, we further explore this link using a manipulation of inter-item distraction. Based on prior work using retrieved-context models (Sederberg et al., 2008), we predicted that inter-item distraction would disrupt context, causing decreased accumulation of category information. Based on the findings of Morton et al. (2013), we further predicted that this disruption would cause decreased category clustering. Using MVPA of oscillatory activity measured using scalp EEG, we compared persistent category-specific activity in lists with no distraction to lists with long or short periods of inter-item distraction. Consistent with our prediction, we found evidence of persistent category-specific activity in right posterior electrodes in the beta band, which was attenuated in the long distraction condition. Furthermore, category clustering was decreased in the distraction conditions. In contrast, temporal organization was not affected by inter-item distraction, establishing a dissociation between semantic and temporal organization. I discuss the implications of these findings for theories of episodic memory.

While Chapters II and III focus on the impact of taxonomic category structure on free recall, Chapter IV tests the ability of finer-grained models of semantic associations to account for behavior in free recall. We created a modeling framework based on CMR, where different models of semantic associations could be paired with CMR, and evaluated based on their ability to predict participant responses on a recall-by-recall basis. We contrasted two prominent models of semantic associations, based on latent semantic analysis (LSA; Landauer and Dumais 1997) and word association spaces (WAS; Steyvers et al. 2004). We also tested two possible mechanisms through which semantic associations might influence memory search. The item-based cuing mechanism assumes that the last item recalled will tend to support associated items, leading to semantic clustering. Polyn et al. (2009) proposed a distinct mechanism where retrieved context is used to query semantic associations. We found that the item-based semantic cuing mechanism provided a better account of the data, and that WAS served as a better model of semantic associations. This work represents progress in evaluating different models of semantics and semantic cuing, all within a common quantitative framework that can control for other influences on behavior such as temporal organization.

CHAPTER II

A neurocognitive theory of episodic and semantic interactions during memory search

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 and Kahana, 2002a; Kahana, 2012), the Context Maintenance and Retrieval model (CMR; Polyn et al. 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 et al. (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; 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 and Kahana, 2002a; Howard et al., 2009), directed forgetting (Sahakyan and Kelley, 2002), interval timing judgments (Shankar and Howard, 2010), reconsolidation (Sederberg et al., 2011), and retrieval-induced forgetting (Jonker et al., 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 and 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 and Kahana, 2002a; Sederberg et al., 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 and Kahana, 2008; Polyn et al., 2012; Manning et al., 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 and 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 et al. 2006) and representational similarity analysis (RSA; Kriegeskorte et al. 2008a) 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 et al., 2005, 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 et al., 2008a; Mitchell et al., 2008; Manning et al., 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 and Rees, 2005; Polyn et al., 2005; O'Toole et al., 2005; Kuhl et al., 2007; Awipi and Davachi, 2008; Lewis-Peacock and Postle, 2008; Kriegeskorte et al., 2008b; Danker and Anderson, 2010; Kuhl et al., 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 et al., 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 et al. (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 encoding 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 and 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 et al., 2007; Jenkins and Ranganath, 2010; Ezzyat and 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 et al. 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 et al., 1957; Bower, 1967; Underwood, 1969; Murdock, 1982; Rumelhart et al., 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

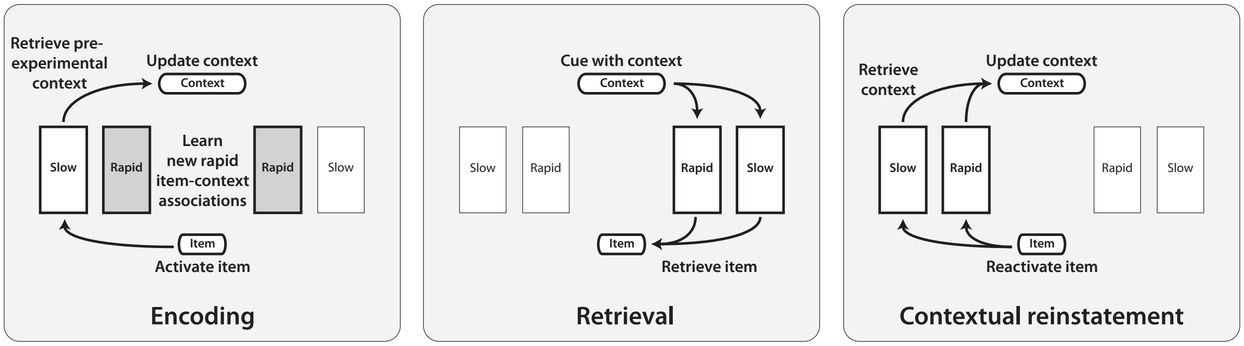


Figure 2. Schematic of model structure and mechanisms for encoding and memory search. An *item* representation and a *context* representation are connected by slowly changing and rapidly changing associations. *Slow* associations were formed before the experiment and hold semantic information. *Rapid* 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.

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 2 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.

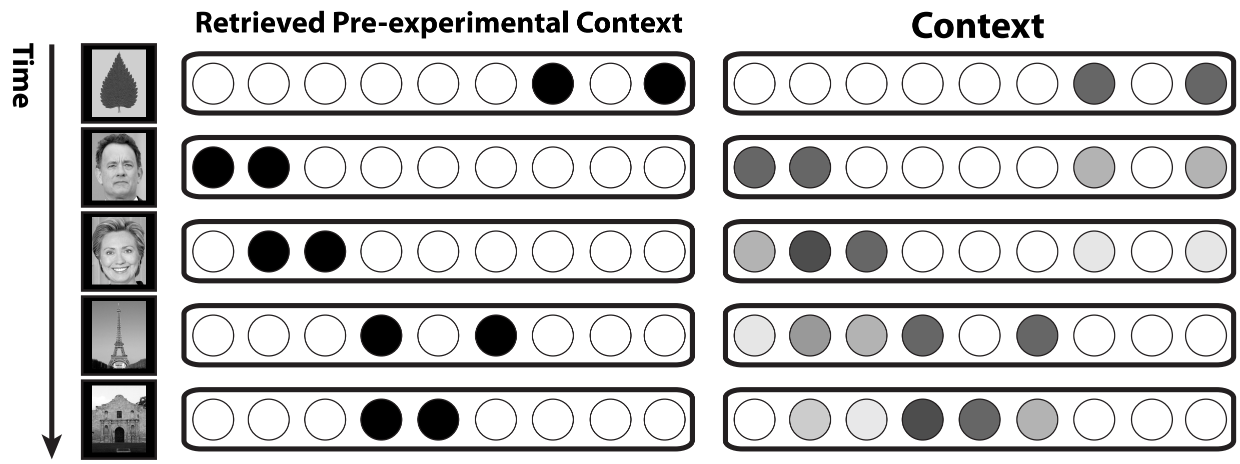


Figure 3. Schematic representation of context evolution during presentation of a list of 5 items from different taxonomic categories. The input representation shows the pre-experimental context associated with each item, which is retrieved when that item is presented. Items from the same category are 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.

When an item is presented, this causes a representation on the item layer to become activated (Fig. 2, left panel). This activated representation then projects through pre-experimental 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 3 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. 2, 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. 2, 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. 2, 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 and Kahana, 2002a; Sederberg et al., 2008; Polyn et al., 2009). The state of the contextual cue determines 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. 2, 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 and Kahana, 2002a).

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 et al., 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 et al., 2000; Quiroga et al., 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 and Kahana, 2002a; Howard et al., 2006a; 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 and 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 and Howard, 2008; Howard et al., 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 and 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 1*, 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 2 through 4 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 2*, 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 3*, 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 4* 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 1: 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 et al., 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 et al., 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).

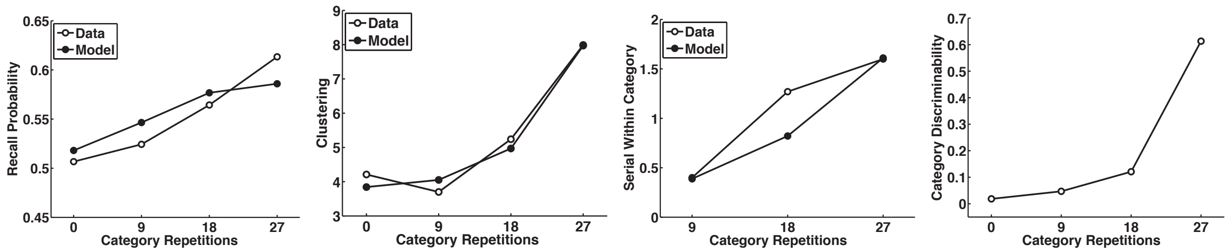
In the present theory, the order in which items are recalled is determined by the dynamics of a contextual cue that is a recency-weighted composite of the semantically laden contextual representations retrieved by representations of the studied items. The fact that the same associative structures are 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 et al., 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. 3 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 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 et al. (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. 4a–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



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

Figure 4. (**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.

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. 4d). 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. 4b) and recall (Fig. 4a). The model also simultaneously accounts for the serial organization observed in the data (Fig. 4c).

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 4 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.

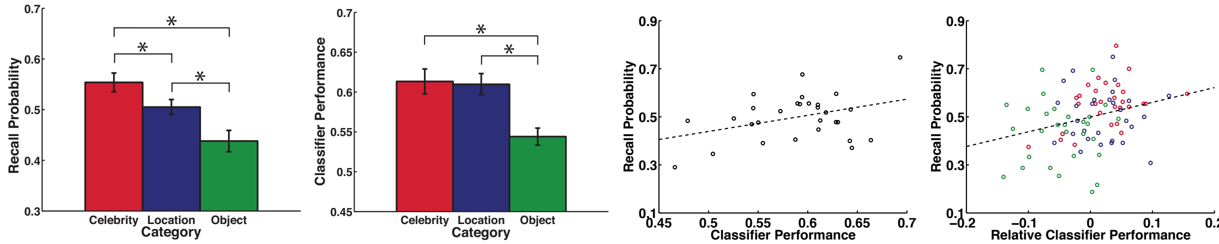
Prior work suggests that this problem could be effectively resolved through the addition of an executive process that detects when a particular category is becoming depleted of memories, prompting a strategic shift from a local search of memory to a more global search (Hills et al., 2012; Raaijmakers and 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 and 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 2: 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



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

Figure 5. (**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 ($r = 0.291$, $p = 0.0062$).

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 3 we examine the predictions of this model in terms of contextual dynamics, and in Simulation Analysis 4, 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.

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 ($\mathrm{LBC}_{\mathrm{sem}} = 3.66$ SEM 0.25; Stricker et al. 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 ($\mathrm{LBC_{sem}} = 0.808$, 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. 5a). 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. 7a). 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. 5b). 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. 5c; $r = 0.371$, $p = 0.048$). 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. 5d) 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 4.

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. 6a), probability of first recall by serial position (Fig. 6e), category clustering (Fig. 7a), 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 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. 5c,d). Here, we focus on the pro-

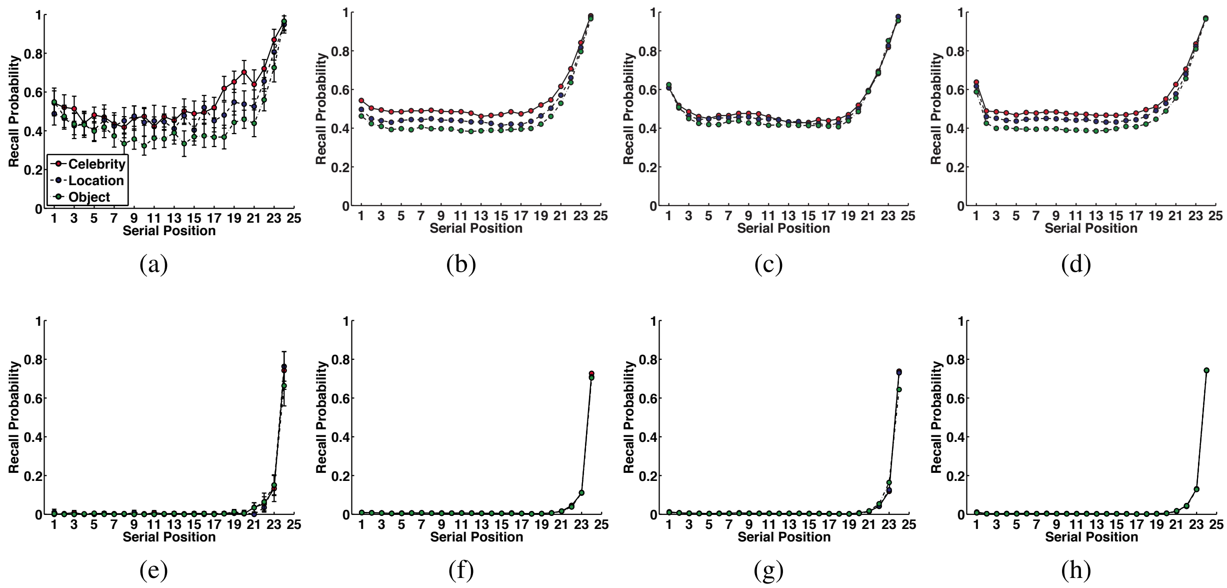


Figure 6. (**a**) Recall probability as a function of serial position and stimulus category. (**b**) Distributed CMR simulation where L^{CF} is free to vary between categories. (**c**) Distributed CMR simulation where β_{enc} is free to vary between categories. (**d**) Distributed CMR simulation where 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.

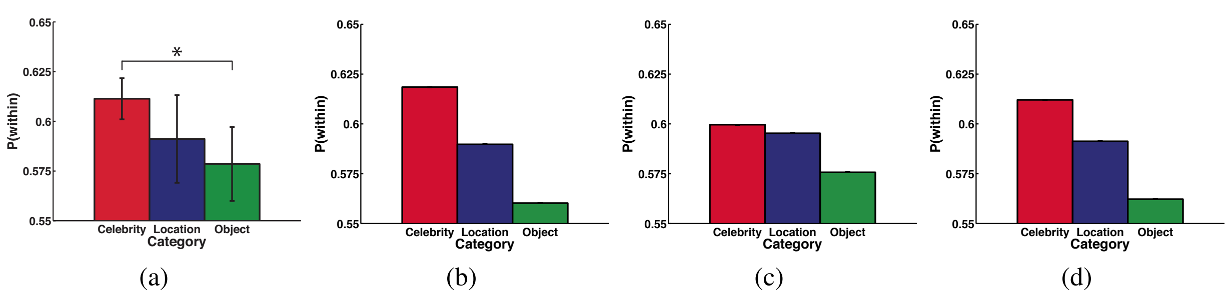


Figure 7. (**a**) Probability of making a within-category transition, conditional on the just-recalled category. (**b**) Distributed CMR simulation where L^{CF} is free to vary between categories. (**c**) Distributed CMR simulation with category-specific β_{enc}. (**d**) Distributed CMR simulation with category-specific within-category similarity.

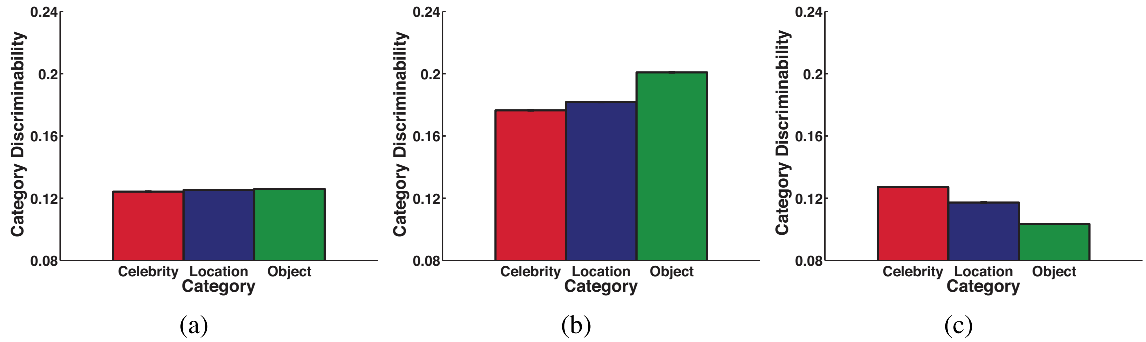


Figure 8. Prototypicality (cosine similarity to the category prototype representation) of states of context, for different variants of CMR. (**a**) Model variant where learning rate on M^{CF} is free to vary between categories. The learning rate manipulation has no effect on item prototypicality. (**b**) 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.

posal 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 1) 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. 6b) and clustering (Fig. 7b). 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 and Pearlstone, 1966; Puff, 1974).

However, this model variant fails to predict differences in category discriminability during encoding (Fig. 8a). 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 encoding context. During recall, the study-period context is reactivated and used as a retrieval cue; therefore,

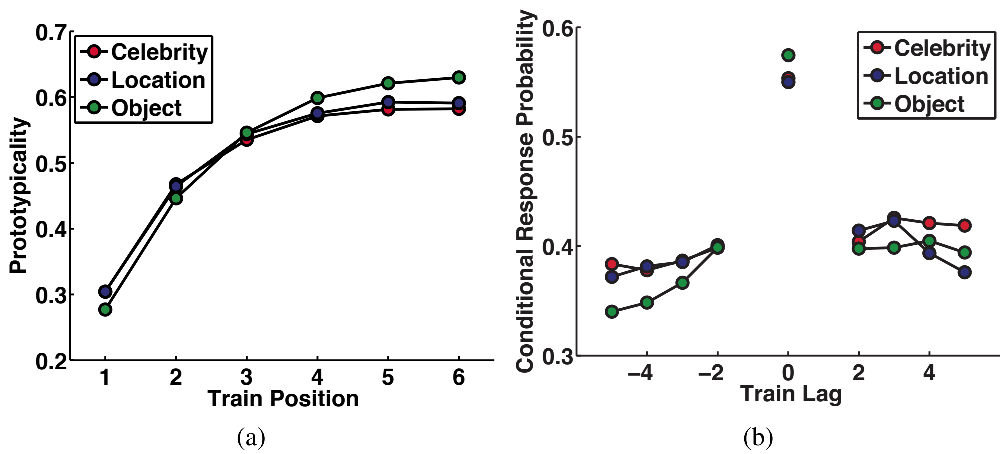


Figure 9. Category activation in context and behavior in the Integration model variant. (**a**) Similarity between the state of context and the category prototype for the current presented item, as a function of train position. Because context updating rate is low for objects, prototypicality at train 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.

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. 6g), which constrained the degree to which the different categories could be assigned different values of β_{enc}. This led to the model under-predicting the magnitude of category-related behavioral effects: celebrities were both better recalled (Fig. 6c) and more strongly clustered (Fig. 7c) 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. 8b). 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. 9a). 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. 9a). 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. 9b). 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 under-predicted. 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

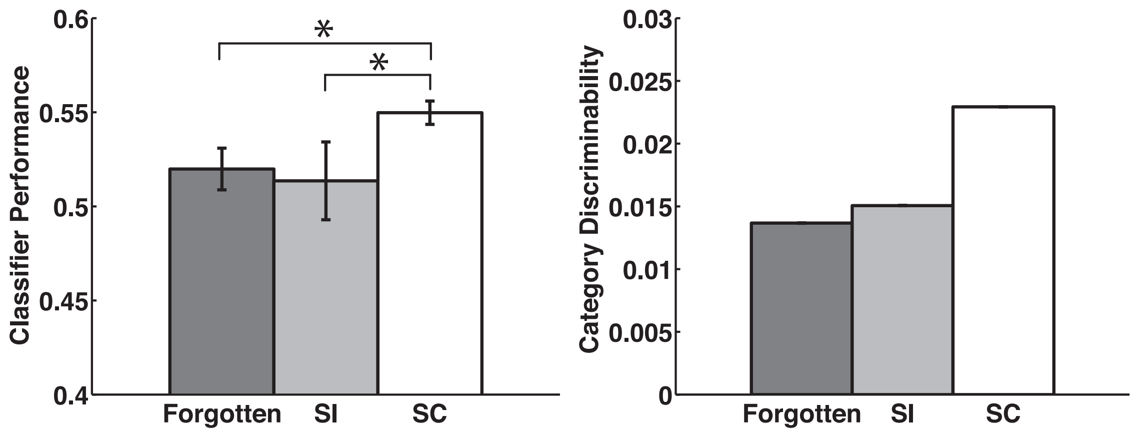
Our third model variant proposes that category-specific differences in behavioral and neural effects arise because of differences in semantic representational structure between the three categories. These structural differences exhibit themselves when an item is studied and retrieves a contextual representation. 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. 7d) and recall (Fig. 6d), while leaving recency unaffected (Fig. 6h).

As described by Howard and Kahana (2002a), 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 representational 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 3: 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



(a) (b)

Figure 10. (**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.

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. 2), 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 2, 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 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. 10a; $t(28) = 3.26$, $p < 0.005$). 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 2, 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 2, 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. 10b)[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 and 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 2.

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. 11a; 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 1 and 2). Consistent with the data, we found that the category discriminability of context increases as multiple items from a category are presented in succession (Fig. 11c). As described in Simulation Analysis 1, 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. 11d). 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. 11b). 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

[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.

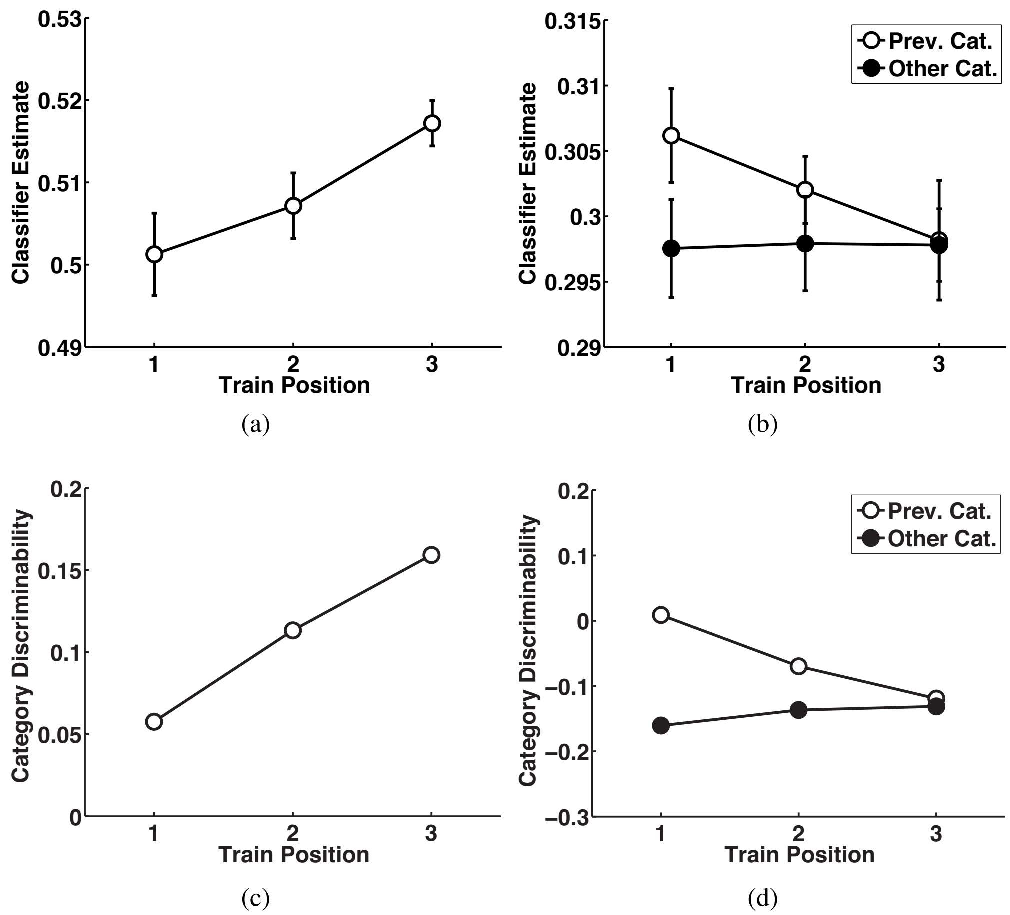


Figure 11. (**a**) Classifier evidence for the current stimulus category increases when multiple items from the same category are presented successively in a train, suggesting that neural activity during encoding reflects the recent history of stimulus presentations. (**b**) Activity related to the category 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.

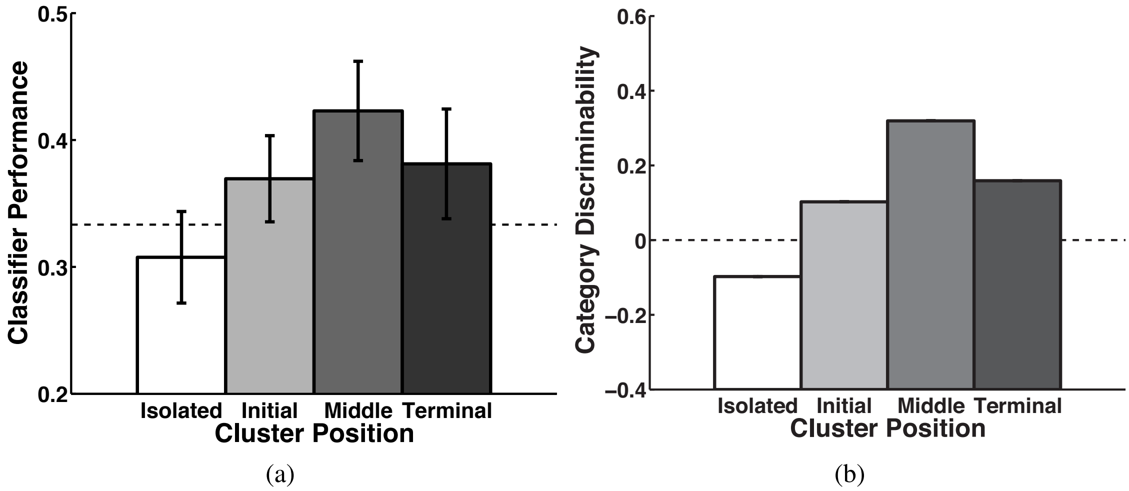


Figure 12. (**a**) Performance of a pattern classifier applied to oscillatory power averaged over 3–0.5 s before onset of a vocalized recall, to identify the category of the upcoming recall. Performance is shown for different cluster position bins (see text for details). (**b**) Strength of the current 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.

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; $t(28) = 0.107$, $p = 0.92$), 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 4, 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. 12a).

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. 12b). 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 another category.

Simulation Analysis 4: 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 2) 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, 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 13 gives an overview of our approach to testing the model's ability to account for individual differences in neural activity.

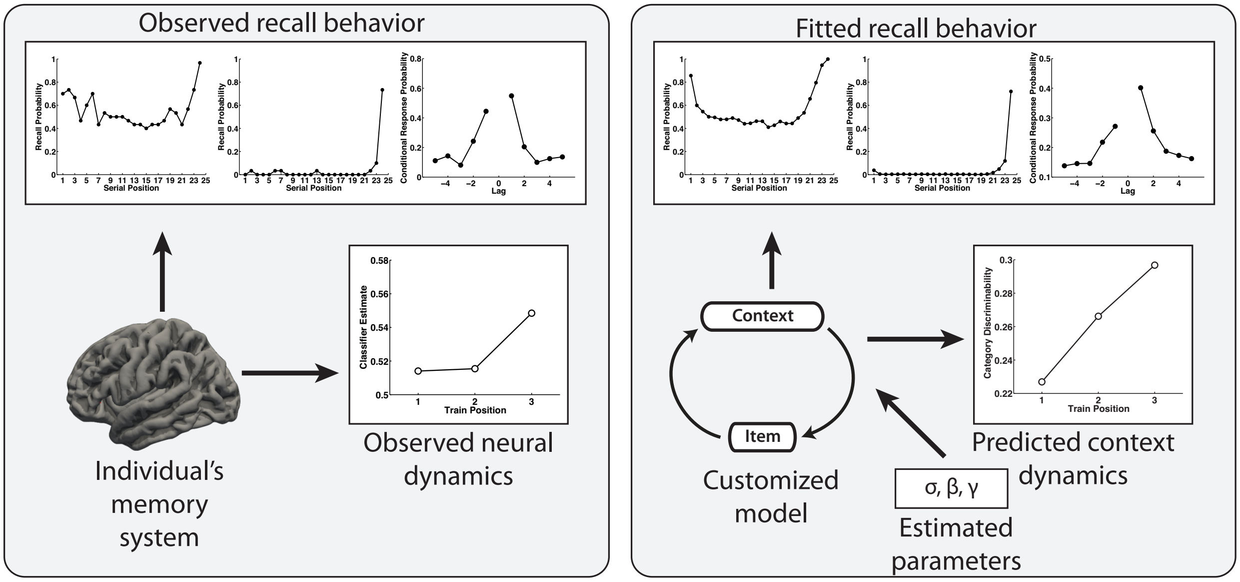
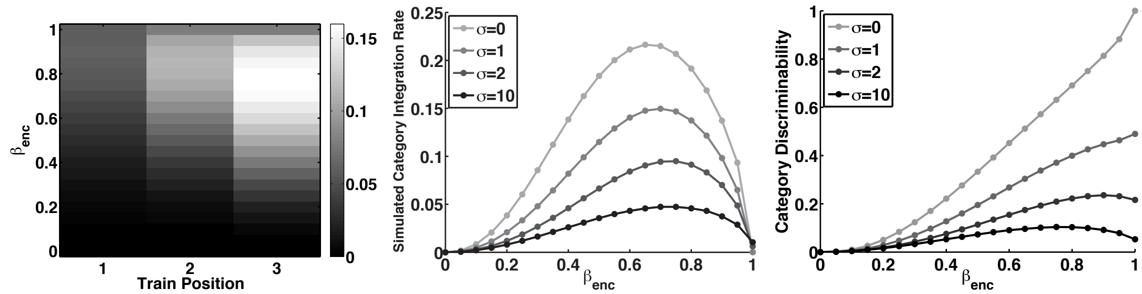


Figure 13. Left panel: Each individual participant has observed recall behavior and neural activity recordings during encoding. Right panel: for each participant, a customized model is constructed by optimizing model parameters to fit that participant's recall behavior. Based on those fitted parameters, 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.

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



(a) (b) (c)

Figure 14. 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 2. (**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}.

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 1. 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. 14). Simulated category integration rate is greatest for intermediate values of β_{enc} (Fig. 14a). 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. 14). We test these predictions by comparing neural activity patterns to states of simulated context (Fig. 13). In the next section, we use measures of recall behavior to estimate parameters for each participant in the scalp

[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 $\sigma > 0$.

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 2), 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.

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

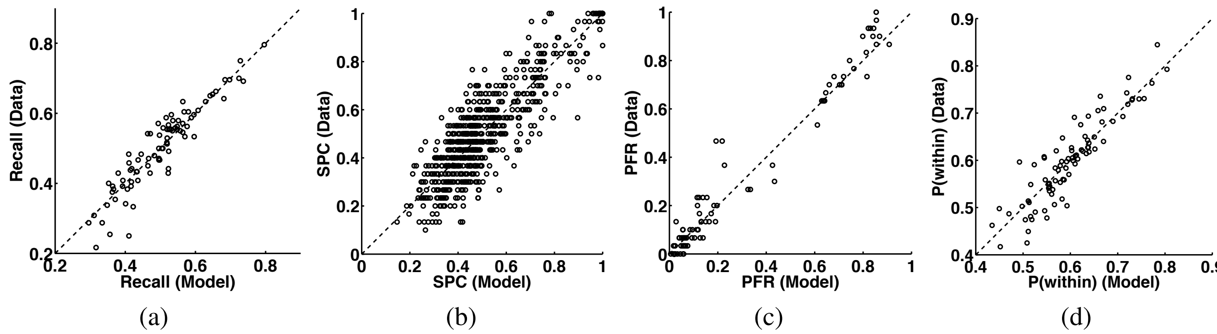


Figure 15. Individual participant fits for model M4 (see text for details). Dotted lines indicate the identity function. (**a**) Recall probability, for the best-fitting model and the observed data. Each point indicates overall recall probability for one participant, for one of the three categories. (**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.

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 et al. 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.

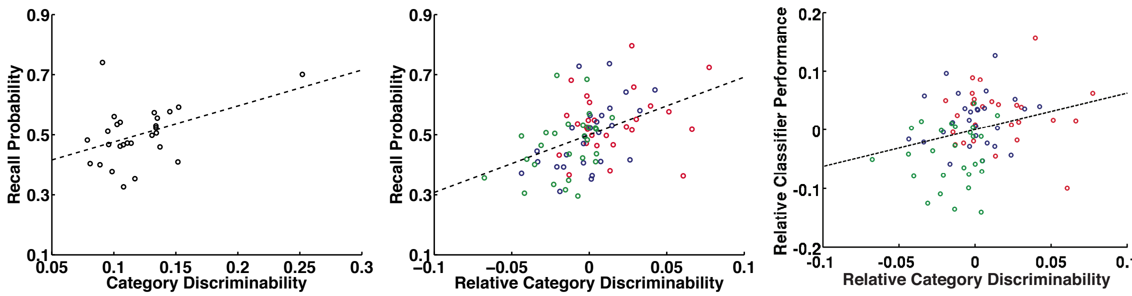
As shown in Figure 15, the M4 model variant demonstrates a good fit to individual variability in recall probability as a function of category, recall by serial position, probability of starting recall at each serial position, probability of making a within-category transition conditional on category, 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 2; Fig. 5c). Furthermore, for a given category, classifier performance relative to the other categories correlated with recall performance for that category (Fig. 5d). 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



(a) (b) (c)

Figure 16. (**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.

between category discriminability and recall probability (Fig. 16a; $r = 0.431$, $p = 0.025$), 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. 16c; $r = 0.273$, $p = 0.014$).

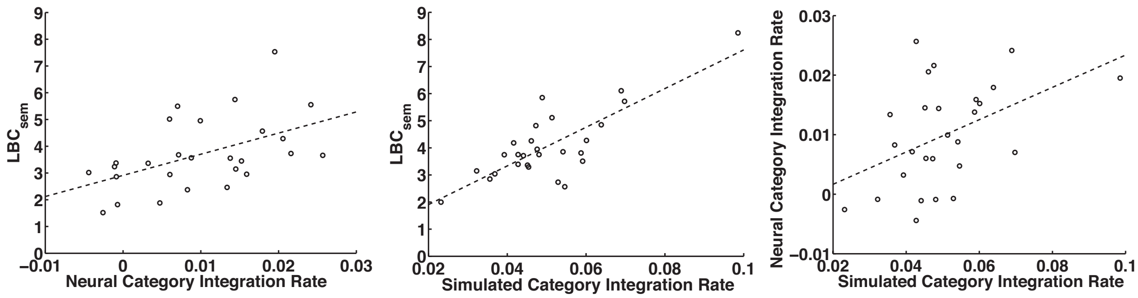
These results demonstrate that, using the model, it is possible to use measures of recall behavior to predict individual differences in encoding-related brain activity. In this case, the model provides accurate predictions about the relative neural discriminability of the different stimulus categories. 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 2 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 3). 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. 17a; $r = 0.503$, $p = 0.0075$; 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).

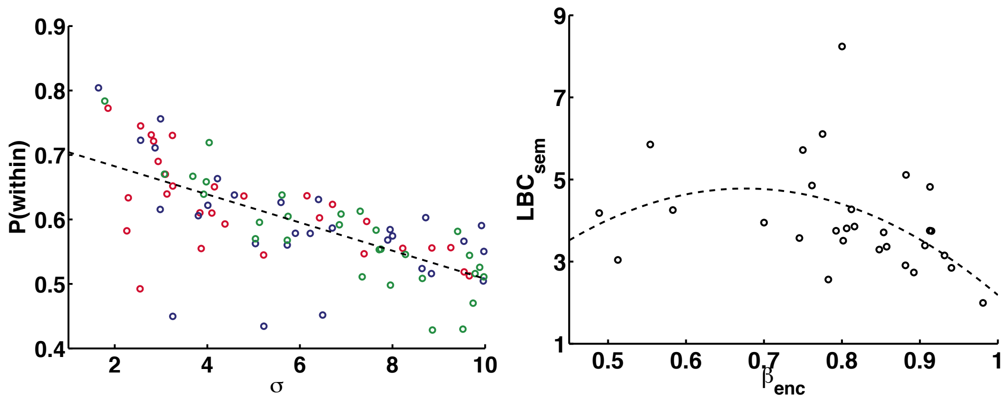


(a) (b) (c)

Figure 17. (**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. 17b; $r = 0.792$, $p = 9x10^{-7}$). 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. 17c; $r = 0.454$, $p = 0.017$). 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



(a)

(b)

Figure 18. (**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.

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. 18a; $r = -0.687$, $p = 1.9x10^{-13}$). This is because 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. 18b; linear trend, $p = 0.074$, quadratic trend, $p = 0.053$). There is a nonmonotonic relation between β_{enc} and LBC$_{\text{sem}}$, 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 et al., 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 and Mewhort, 2007; Rao and 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 1 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. 4a–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. 4).

These dynamics are examined in more detail in Simulation Analyses 2 through 4, 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. 5–7). 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. 8).

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. 5 & Fig. 16), and the tendency for items that elicit strong category-specific patterns to be subsequently recalled in a category cluster (Fig. 10). 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. 11a & c) while the representation of the previous category declines smoothly (Fig. 11b & 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. 12).

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. 16c), and in neural category integration rate (Fig. 17c).

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 and Pearlstone, 1966; Puff, 1966, 1974; Raaijmakers and 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 and 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 et al., 2001).

Kimball et al. (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 and 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. 12). 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 and Pearlstone, 1966; Puff, 1966, 1974; Raaijmakers and Shiffrin, 1980; Becker and 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 et al., 2011) and category fluency tasks (Taler et al., 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 and Pearlstone 1966; 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 and 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. 11, 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 et al., 2012; Polyn et al., 2012; Polyn and 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 2 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 $\mathbf{c}_i^{\text{IN}}$ is retrieved and is input to the context layer to update the current state of context. The input to context is

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

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

[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.

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.

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

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

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

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

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 and Kahana 2002a; Howard et al. 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

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

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:

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

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}:

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

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

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

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

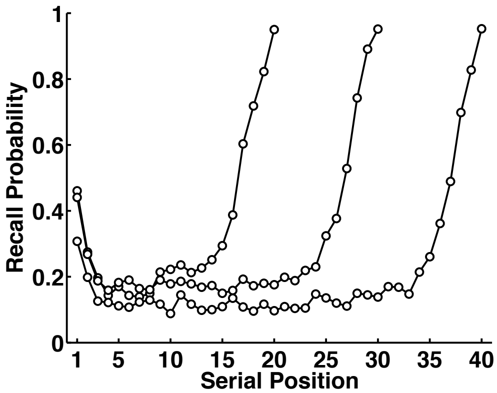
Once $\mathbf{f}^{\text{IN}}$ has been determined, there is a competition between items to determine which will be recalled. The support for each item $\mathbf{f}^{\text{IN}}$ enters into a competition of competing, leaky accumulators (Usher and McClelland, 2001; Sederberg et al., 2008), where each item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

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

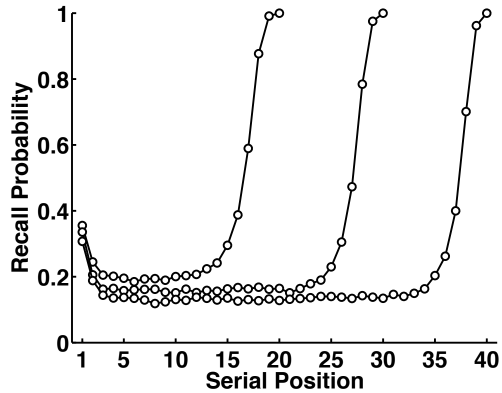
where each element of the vector $\mathbf{x}_s$ corresponds to a studied item. $\mathbf{x}$ is initialized to $\mathbf{0}$. It is updated on each step s of the accumulation process, until one of the accumulating elements crosses a threshold (which is set at one), or the recall period is over. Each accumulator 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

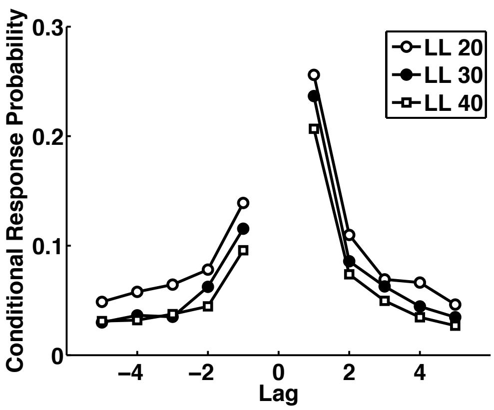
$$\mathbf{c}_i^{\text{IN}} = \mathbf{M}^{FC}\mathbf{f}_i. \tag{II.10}$$



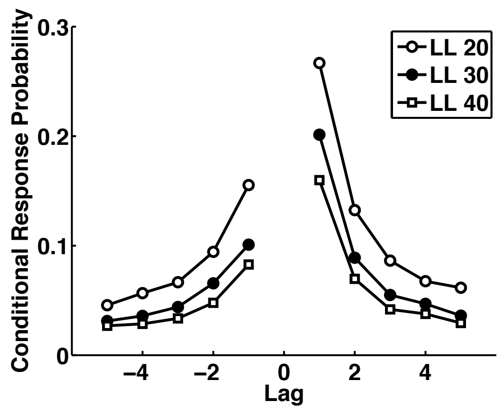
(a)



(b)



(c)



(d)

Figure 19. 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.

Context is then updated according to

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

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 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 and 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 et al. 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 et al., 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 19 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

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).

RMSD, based on the summary statistics of interest. We used a MATLAB-based implementation of differential evolution, based on code developed by Price et al. (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 re-

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

	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.

sponse 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 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 1

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, 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

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

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 2

For fitting group statistics in Simulation Analysis 2, 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

[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.

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 4

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 2.

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 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.

CHAPTER III

Inter-item distraction disrupts accumulation of semantic context

Introduction

Studies of free recall have consistently found that participants tend to successively recall items that were presented adjacent to one another (Kahana, 1996). This finding of temporal organization has provided important constraints for theories of episodic memory. Two influential classes of models, dual-store models and retrieved-context models, use distinct but related mechanisms to account for the finding of temporal organization. Dual-store models assume that studied items are placed into a multi-item working-memory buffer. When the buffer becomes full, incoming items displace older items in the buffer. Items that are maintained in the buffer at the same time are assumed to become associated in long-term memory (Raaijmakers and Shiffrin, 1980). Items that are presented near to each other will tend to be in the buffer at the same time, and therefore will tend to be more strongly associated than distant items. This difference in inter-item association strength allows dual-store models to account for temporal organization in free recall (Kahana, 1996). In contrast, retrieved-context models assume that studied items become associated to a slowly changing representation of temporal context. Items presented near to each other will become associated with similar states of context. When an item is retrieved, its associated context is reactivated and used to cue for other items. This state of context provides a good cue for items associated with similar states of context; therefore, nearby items will be relatively well-supported, leading to temporal organization (Howard and Kahana, 2002a).

Although the two classes of models differ in important ways, both dual-store and retrieved-context models rely on a cognitive representation that persists over presentation of multiple items, allowing items presented at different times to become associated. Dual-store models assume persistent activation in the form of a buffer that holds a discrete number of items (Raaijmakers and Shiffrin, 1980; Davelaar et al., 2005), while retrieved-context models assume persistent activation in the form of a gradually changing context representation (Howard and Kahana, 2002a; Sederberg et al., 2008; Polyn et al., 2009). Collectively, we will refer to the combination of persistent activation and association formation used in both models as *cue construction*, as in both models persistent activation during encoding allows formation of associations that will later guide memory search.

Actively maintained representations of studied items are have been proposed to be important for recognizing semantic relationships between items, which may enhance memory (Glanzer, 1969). Anderson (1974) argued that the dominant strategy in free recall is to identify connections between the studied items in semantic memory, and that this process is easier when related items are held in working memory at the same time. Consistent with the idea that working memory facilitates discovery and utilization of semantic relationships, the spacing between related items in a list affects later recall behavior. When related items are presented near to each other in the list, this leads to enhanced recall (Glanzer, 1969) and semantic organization (Howard and Kahana, 2002b; Puff, 1966). These results are consistent with the hypothesis that the relationship between a given pair of items is easier to discover when they are simultaneously held in working memory, and that recognizing this relationship somehow aids later recall. Recognition of a relationship between a set of items may enhance later recall by allowing them to be stored as a single "chunk" of information (Cohen, 1963).

Morton et al. (2013) found neural evidence consistent with an integrative cue-construction process during episodic encoding. They found that patterns of oscillatory activity related to simulus category increase in fidelity as multiple items from the same category are presented in succession. Furthermore, they found that the rate of this increase predicts individual differences in category clustering, the tendency to group together items from the same category during free recall. These results suggest that persistent category-specific neural activity may aid integration of related memories, resulting in increased organization during memory search. However, many questions remain about the properties of integrative oscillatory activity. Morton et al. (2013) used multivariate pattern analysis to decode category information from a wide range of neural signals at different frequency bands and electrodes. This approach provides a summary measure of how category-specific activity changes over time, but does not provide information about the topography and frequency distribution of integrative activity.

Both dual-store and retrieved-context theories assume that a neural correlate of cue construction should be susceptible to disruption by distracting activity. In order to account for attenuation of the recency effect when recall initiation is delayed by a distracting task (Glanzer and Cunitz, 1966), both types of models propose that distraction causes disruption of actively maintained information (Raaijmakers and Shiffrin, 1980; Sederberg et al., 2008). In order to test these predictions, we measured scalp EEG during study and free recall of lists of categorized stimuli. We varied the amount of distraction between list items, with the prediction that distraction would disrupt the accumulation of category-specific neural activity. Given prior evidence of a link between integrative category-specific neural activity and category clustering (Morton et al., 2013), we also predicted that increasing distraction would lead to a decrease in category clustering.

Consistent with dual-store and retrieved-context theories of memory, we found evidence of oscillatory activity that becomes more category-specific as multiple items from the same category are presented in succession. We also found evidence that activity related to a previously presented category decreases gradually as other items are presented. This integrative activity was attenuated, but not eliminated, by inter-item distraction. We found that persistent category-specific activity was specific to right posterior electrodes, in the beta frequency band. Furthermore, consistent with prior work (Howard and Kahana, 2002b), we found that inter-item distraction attenuated category organization, while leaving temporal organization unaffected. Our results demonstrate a dissociation between temporal and semantic organization, as well as a potential oscillatory signature of the disruption of semantic encoding processes during study in the presence of distraction. These findings provide important constraint for models of memory search.

Methods

Participants

Ten paid volunteers (5 female, age 18-30 years) participated in the study. The research protocol was approved by the Institutional Review Board of Vanderbilt University.

Stimuli

Stimuli consisted of photographs of famous landmarks, celebrity faces, and common objects. There were 256 stimuli for each category. The pool was based on a previous experiment (Morton et al., 2013), with some changes to increase the homogeneity of stimuli within each category. A norming study was used to determine exemplars of the categories that were familiar to our participant pool. The norming study included 19 undergraduate Vanderbilt University students (9 female, age 18–35 years) that received course credit for participation. Participants performed a category fluency task, where they wrote down exemplars of several categories, chosen to be sub-categories of the 3 main categories. Celebrity sub-categories were actors, athletes, entertainers, and politicians. Location sub-categories were natural landmarks, manmade landmarks, and cities. Object sub-categories were tools, foods, furniture, clothing, toys, and instruments. The order of presenting sub-categories was randomized across participants. High-frequency responses from the norming study were used to replace some items from the original pool that failed to meet the new criteria in terms of content (e.g. all celebrities were contemporary) or picture (if no picture matching the requirements could be found; sometimes the case for older celebrities with no available high-resolution photo).

Only contemporary famous people were included in the celebrity pool; non-contemporaries such as Abraham Lincoln were replaced. Recently deceased celebrities such as Whitney Houston were also replaced, in order to limit variability in emotional valence associated with different stimuli. Pictures of most stimuli were replaced with new higher-resolution photographs taken from Google image search. All pictures were cropped and scaled to 600 x 750 resolution. For each celebrity, a picture was selected to maximize image quality, emotionally neutral expression, facing straight toward the camera, and simplicity of the background. Pictures were cropped approximately from the chin to the top of the head (excluding hair). Pictures of locations were selected to minimize the number of people in them (it is difficult to eliminate all people, since many of the pictures are tourist attractions). Some of the included pictures contain people, but only at a distance. Stimuli were chosen to limit overlap with the other categories (e.g. the Liberty Bell was not included, since it could be considered either a location or an object). With the exception of two performance venues (known most for their interiors), all locations were shown from the exterior and taken during the day. Finally, all object pictures were taken on a white background. Using the SHINE toolbox (Willenbockel et al., 2010), images were converted to grayscale and normalized so that mean image intensity and contrast were the same for each image.

Unlike in the experiment presented by Morton et al. (2013), names of presented stimuli were presented auditorially instead of visually, in order to limit eye movements during encoding. A male speaker (NWM) recorded sound clips of all stimulus names with neutral affect. Sound clips ranged in duration from 690 to 1498 ms (mean 1056, S.D. 159). It is worth noting that the mean length of stimulus names (and therefore the length of the sound clips) varied by category (celebrities: mean 1105 ms, S.D. 104 ms; locations: mean 1166 ms, S.D. 124 ms; objects: mean 898 ms, S.D. 104 ms). This is an issue that is difficult to avoid, given the stimulus categories used here.

Experimental paradigm

In a preliminary session, participants rated their familiarity with each stimulus used in the experiment while scalp EEG was recorded. This was done to assess participants' pre-experimental familiarity with each stimulus, to provide participants at least a minimal familiarity with each stimulus, and to provide us with category-specific oscillatory responses in the absence of the cognitive demands of an episodic encoding task. Items were presented in blocks; the order of stimulus presentation within each block was designed to match the structure of stimulus presentation in the later free recall sessions (described below). Each block contained 24 items, with 8 items from each category. Stimuli were presented in trains of items from the same category, with 2–4 items in each train. The order of the trains was randomized, with the constraint that all categories appeared in each set of 3 trains, and that adjacent items did not contain the same category. Stimuli were chosen so that items of a similar type (e.g. stadiums, presidents) did not appear in the same list. Each stimulus image was presented for 2000 ms, during which participants rated their familiarity with the stimulus' referent on a four-point scale. The name of the stimulus was presented auditorially, beginning at the same time the image was presented. Participants were given a chance to rest after each block. The familiarization session included 32 blocks; all 768 stimuli were presented.

The subsequent 3 sessions each included 18 free recall trials, followed by a final free recall period; scalp EEG was recorded throughout each session. In each free recall trial, 24 items were presented for later recall. The structure of each list was the same as each of the familiarization blocks (described above), with 8 items from each of the 3 categories. Each stimulus was presented for 2000 ms, during which participants rated how much they liked or disliked the item on a 4-point scale.

There were three types of lists: immediate free recall (IFR), short continual distraction free recall (CDS; 2.5 s of distraction), and long continual distraction free recall (CDL; 7.5 s of distraction). In the IFR condition, each stimulus was followed by a fixation cross for 1000$\pm$200 ms. After the last stimulus in the list, a fixation cross was presented for 1300$\pm$100 ms, followed by a row of asterisks and a 300-ms tone signaling the start of a 70 s free recall period. Participants were instructed to say the names of items from the list, in any order, and to keep trying to remember items throughout the recall period. Digital recordings of vocal recalls were scored using Penn TotalRecall. The CD conditions were the same as the IFR condition, except each item was preceded and followed by a math distraction task. After each item, a fixation cross was presented for 500$\pm$200 ms. Next, a series of three digits was presented, each for 400 ms (with no gap in between). After the digits, a proposed answer was immediately shown along with a question mark until the participant responded or the distraction period ended. Participants were instructed to indicate by button press whether the last number equaled the sum of the first three numbers. The proposed answer was incorrect for 50% of problems; incorrect answers were generated by adding 1, -1, 2, or -2 to the correct answer. Immediate feedback was given in the form of a beep to indicate incorrect responses. After the response, there was a fixation cross for 300$\pm$200 ms. Another problem was presented if there was at least 2400 ms left in the distraction period; otherwise, a fixation cross was presented for the remainder of the distraction period. Each distraction period was followed by a fixation cross for 500±200 ms. Distraction period timing was designed so that participants were presented with one problem in the 2.5 s condition, and up to three problems in the 7.5 s condition.

Each of the three free-recall sessions included 18 lists, 6 from each condition. Conditions were randomly assigned to lists, with the constraints that each group of three lists included all conditions, and no adjacent lists were from the same condition. At the end of each session, there was a final free recall (FFR) period where participants were given 360 s to recall names of stimuli from any of the lists presented during the session.

Behavioral analysis

In addition to standard measures of free-recall performance such as recall as a function of serial position, we also calculated measures of recall organization. To obtain a measure of temporal clustering independently of category clustering, we first focused on transitions between recalls of items in the same category. For these within-category transitions, we calculated the temporal organization score (Polyn et al., 2009). For each included transition, we ranked all possible recalls by their distance from the just-recalled item in the list, then calculated the percentile of the item that was recalled. This percentile was averaged over all transitions to obtain a summary measure of temporal clustering, with higher values indicating greater clustering. Category clustering was measured using the semantic list-based clustering measure (LBC_{sem}; Stricker et al. 2002).

Scalp electroencephalography recordings and data processing

EEG measurements were recorded using 128-channel HydroCel Geodesic Sensor Nets and a Net Amps 300 Amplifier (Electrical Geodesics, Inc.). Recordings were initially referenced to Cz. Voltage was digitized at 500 Hz, and a digital bandpass filter of 0.1–200 Hz was applied.

A digital Butterworth notch filter with zero phase distortion at 60 Hz was used to remove electrical noise. Electrodes with poor contact were identified through manual inspection of event-related potential (ERP) images (Jung et al., 2001); these electrodes were omitted from the independent components analysis (ICA) procedure described below. Continuous EEG was segmented into epochs during the study period of free-recall trials, from 2000 ms before stimulus onset to 1500 ms after stimulus offset. Epochs were manually inspected using EEGLAB (Delorme and Makeig, 2004) to reject epochs containing non-stereotyped artifacts such as skin potentials. ICA was then used to decompose the signal (Onton et al., 2006). Components reflecting artifacts, such as eye movements, blinks, and muscle artifacts, were manually identified; non-artifactual components were projected back into the original sensor space to obtain a cleaned signal (Junghöfer et al., 2000; McMenamin et al., 2010). Bad electrodes were then replaced using spherical spline interpolation (Nolan et al., 2010).

Remaining artifacts on individual electrodes on specific epochs were removed using spherical spline interpolation. Our procedure was based on a combination of the SCADS (Junghöfer et al., 2000) and FASTER (Nolan et al., 2010) artifact-rejection methods. Bad electrode-epochs were detected using a number of statistics; this procedure is modified from the FASTER algorithm proposed by Nolan et al. (2010). For each epoch, an electrode was considered bad if there was a voltage change of greater than 150 μV, or if the electrode-epoch was an outlier (defined as more than six times the interquartile range above or below the first or third quartile, relative to all electrode-epochs) on any of three statistical measures. Because the variance of electrodes vary depending on their distance from the reference, quadratic regression was used to estimate the effect of distance from reference on each of the statistics; rejection was performed on the residuals of this regression (Junghöfer et al., 2000). The first statistic is designed to detect fast changes related to muscle artifacts. First SS_{fast} is calculated as the sum of squared differences between successive samples; SS_{fast} is then divided by the total summed squared deviations, SS_{total}, to obtain a measure proportional to the fraction of variance in voltage explained by fast changes. Informal investigation confirmed that this measure has high sensitivity and selectivity for detecting EMG artifacts. The other statistics are the variance over time and the difference between the maximum and minimum voltages. Any electrodes identified as bad for a given epoch were replaced using spherical spline interpolation (Nolan et al., 2010). If more than 10% of electrodes were bad for a given epoch, that epoch was excluded. After interpolation of bad electrode-epochs, the data were converted to an average reference. Finally, the electrode-epoch interpolation procedure was run again on the re-referenced data (Junghöfer et al., 2000).

Oscillatory analysis

We measured oscillatory power using a Morlet wavelet transform with a wavenumber of 6. Oscillatory power was measured at 37 logarithmically spaced frequencies from 2 to 128 Hz. Power was then log-transformed and down-sampled to 20 Hz. Log power was normalized relative to a baseline period 500–400 ms before stimulus onset.

Multivariate pattern analysis

We used multivariate pattern analysis (Norman et al., 2006) to decode stimulus category based on patterns of oscillatory power. Classification was carried out using penalized logistic regression (penalty parameter = 10), using L_2 regularization (Duda et al., 2001). Classification analyses were carried out using APERTURE (available at: http://code.google.com/p/eeg-analysis-toolbox) and the Princeton MVPA Toolbox (available at: http://www.pni.princeton. edu/mvpa). A cross-validation procedure was used to determine the strength of category-specific activity during each item presentation, based on the distributed pattern of oscillatory power. The classifier was trained on all but one list, then applied to the epochs on the left-out list to evaluate its performance, measured as the fraction of items classified correctly.

We first examined which frequencies and times relative to stimulus onset contained information about stimulus category. A separate classification analysis was performed at each time and frequency bin, to decode category based on the pattern of oscillatory power over the scalp. We next examined the category discriminability of oscillatory activity at each individual electrode. At each electrode, we carried out a classification analysis where the classifier simultaneously used features taken from two time bins (power was averaged within these bins: 0–500 ms and 500–2000 ms) and all frequencies. Based on the results of this analysis, we defined two regions of interest (ROIs) where classification accuracy was highest (Fig. 23; see Results for details). We focused subsequent analysis on six frequency bands of interest: delta (2–4 Hz), theta (4–8 Hz), alpha (10–14 Hz), beta (16–25 Hz), low gamma (25–55 Hz), and high gamma (65–128 Hz). Power was averaged within each of these ranges to obtain power values for each band. A final series of classification analyses examined power in these frequency bands. For each electrode and frequency band, we performed a cross-validation analysis, with power at each time bin during stimulus presentation (0–2000 ms, sampled at 20 Hz) as features. Classifier performance was then averaged over all electrodes within each ROI.

At each electrode and frequency band, we calculated the classifier's estimate of the category of each stimulus, based on its neural activity over time. We averaged this estimate over all electrodes in each ROI. For each ROI/frequency band, we calculated the average classifier evidence as a function of the position of the item in a train of items from the same category, separately for each subject. For each subject, we calculated the slope of classifier evidence over the first three train positions. We focused on the first three train positions because Morton et al. (2013) found that classifier estimates did not increase beyond the third train position. We used a t-test to assess whether the slope over train position was significantly positive across subjects. Significance values were Bonferroni corrected across ROIs and frequency bands. We also examined the classifier's estimate of activity related to the category presented in the previous train of stimuli. For comparison, we also examined activity related to a baseline category, which was neither the currently presented category or the previously presented category. We used t-tests to determine whether the previous or baseline categories changed as a function of train position.

Results

Inter-item distraction dissociates temporal and semantic organization

Recall probability by serial position and condition (Fig. 20) showed a striking interaction between condition and serial position, where distraction only affected recall for earlier positions on the list. To characterize differences in serial position effects between conditions, we first examined the serial position curve averaged over conditions, to identify serial positions corresponding to primacy, asymptote, and recency positions. Using a series of t-tests, we found that the earliest

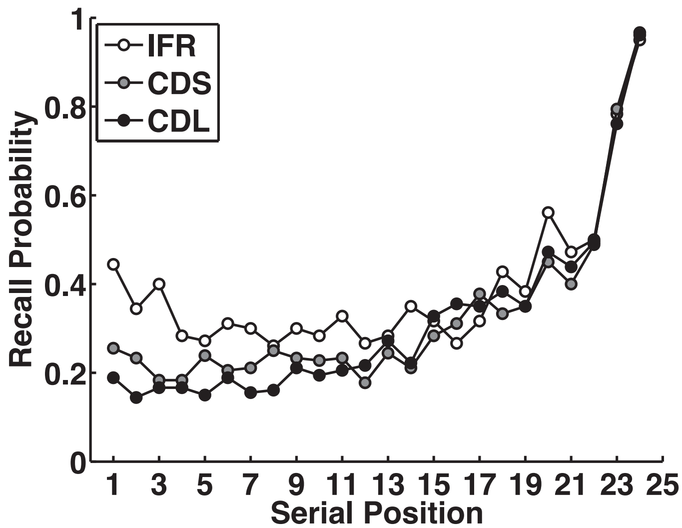


Figure 20. Recall probability as a function of serial position. IFR: immediate free recall (no distraction). CDS: continual distraction for 2.5 s before and after each item. CDL: continual distraction for 7.5 s before and after each item.

serial position pair showing a decrease in recall was the comparison between serial positions 12 and 13 ($t(9) = 1.96$, $p = 0.041$, one-sided test). We found evidence of only a 1-position primacy effect (serial position 1 vs. serial position 2: $t(9) = 2.50$, $p = 0.017$, one-sided test). We therefore defined three serial position bins: primacy (1), asymptote (2–12), and recency (13–24). A two-way repeated measures ANOVA with Greenhouse-Geisser correction for nonsphericity revealed a significant main effect of serial position ($F(2,18) = 40.81$, $p = 7.0x10^{-7}$), a main effect of distraction ($F(2,18) = 34.44$, $p = 3.3x10^{-5}$), and an interaction ($F(4,36) = 8.25$, $p = 0.0015$). Followup one-way t-tests tested our prediction that recall would decrease with increasing distraction. For the primacy position, we observed significant differences between IFR and CDS ($t(9) = 4.54$, $p = 0.0007$) and IFR and CDL ($t(9) = 4.54$, $p = 0.0007$), but no difference between CDS and CDL ($t(9) = 1.72$, $p = 0.06$). At asymptote positions, we observed significant differences between IFR and CDS ($t(9) = 6.75$, $p = 0.00005$), IFR and CDL ($t(9) = 7.55$, $p = 1.7x10^{-5}$, and CDS and CDL ($t(9) = 2.15$, $p = 0.03$). At recency positions, we observed a significant difference between IFR and CDS ($t(9) = 2.18$, $p = 0.029$); all other pairwise comparisons were not significant ($p > 0.05$).

At asymptote positions, we observed a significant decrease in recall performance from the

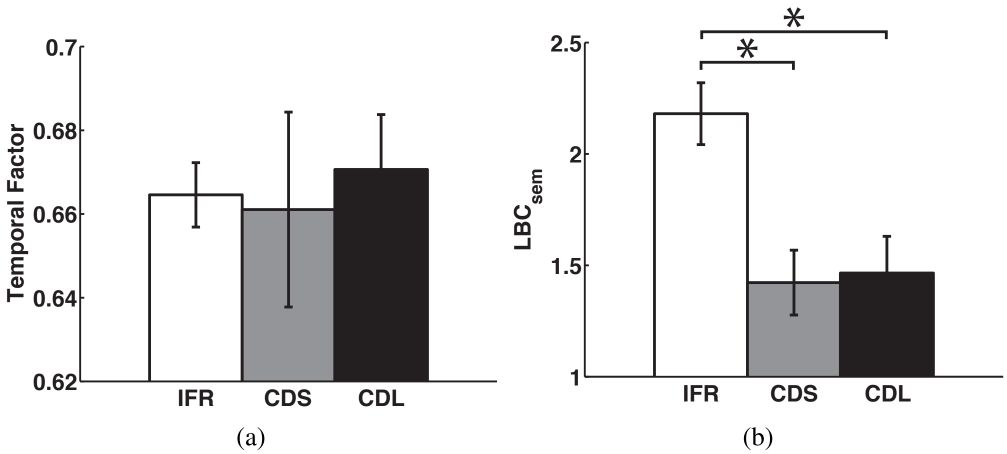
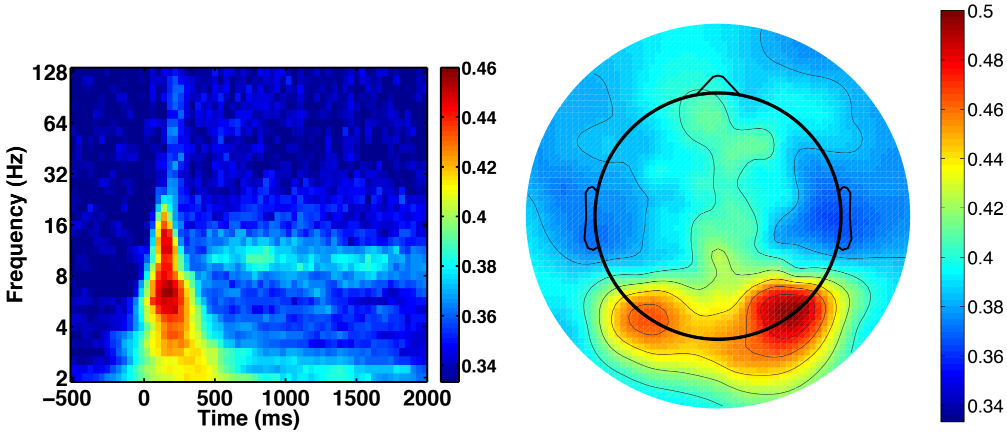


Figure 21. (**a**) Distraction had no effect on temporal organization scores. To control for the effect of category, only within-category transitions were included. (**b**) Distraction significantly attenuated category clustering, as measured by LBC$_{sem}$. IFR: immediate free recall; CDS: short continual distraction; CDL: long continual distraction.

CDS condition to the CDL condition, suggesting that our parametric manipulation of distraction had a similarly continuous effect on recall. In contrast with prior results, where the addition of continual distraction attenuated recall at all serial positions (Bjork and Whitten, 1974), distraction only attenuated recall for the first half of the list.

We measured temporal clustering using the temporal organization score introduced by Polyn et al. (2009). To control for effects of category, we separated transitions by whether they were within category or between category (Polyn et al., 2011). Significant within-category temporal clustering (temporal organization score > 0.5) was observed in each condition, even the distraction conditions (Fig. 21a; IFR: $t(9) = 21.37$, $p = 5.1x10^{-9}$; CDS: $t(9) = 6.92$, $p = 6.9x10^{-5}$; CDL: $t(9) = 13.02$, $p = 3.8x10^{-7}$). There were no significant differences in within-category temporal organization between conditions (all $p > 0.05$). Similar results were observed for between-category transitions. There was significant temporal organization in between-category transitions for all distraction conditions (IFR: mean 0.554, SEM 0.011, $t(9) = 49.47$, $p = 2.8x10^{-12}$; CDS: mean 0.595, SEM 0.020, $t(9) = 30.48$, $p = 2.2x10^{-10}$; CDL: mean 0.577, SEM 0.011, $t(9) = 52.14$, $p = 1.8x10^{-12}$). There were no significant differences in temporal organization for between-category transitions between distraction conditions (all $p > 0.05$).



(a) (b)

Figure 22. (**a**) Classifier accuracy for separate cross-validation analyses at each time and frequency bin. Stimuli were classified based on oscillatory power over all electrodes. Deep blue corresponds to chance performance ($0.3\bar{3}$). (**b**) Classifier accuracy for separate cross-validation analyses at each electrode. Stimuli were classified based on oscillatory power at all frequencies and two time bins. Performance was significantly above chance ($0.3\bar{3}$) at all electrodes ($p < 0.05$, corrected).

Category clustering was quantified using the list-based semantic clustering (LBC_{sem}) measure introduced by Stricker et al. (2002). Category clustering was decreased in the distraction conditions (Fig. 21b). Category clustering in IFR was greater than in CDS ($t(9) = 4.18$, $p = 0.0024$) and CDL ($t(9) = 5.22$, $p = 5.5x10^{-4}$). There was no difference in clustering between CDS and CDL ($t(9) = 0.29$, $p = 0.78$).

Neural category discriminability during encoding

We first characterized which frequencies, times, and electrodes exhibited category-specific activity during encoding, over all distraction conditions. We carried out a separate cross-validation classification for each time and frequency bin, providing the classifier with topographic patterns of oscillatory activity to distinguish between categories. Classifier performance early after stimulus onset (0–500 ms) is relatively high for a wide range of frequencies from about 2–25 Hz (Fig. 22a), with peak performance in the theta band. In the later period (500–2000 ms), classifier performance was above chance, with peaks in the delta and alpha bands.

We next examined which electrodes exhibited category-specific oscillatory activity, collapsing

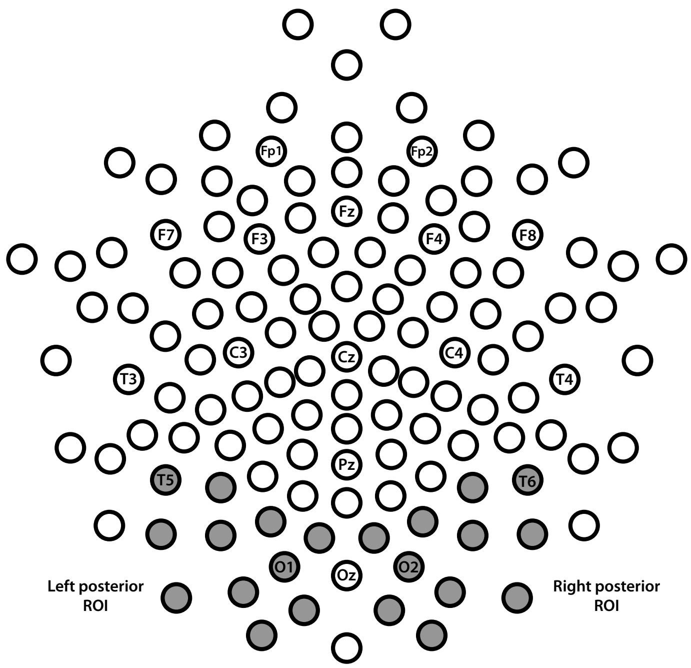
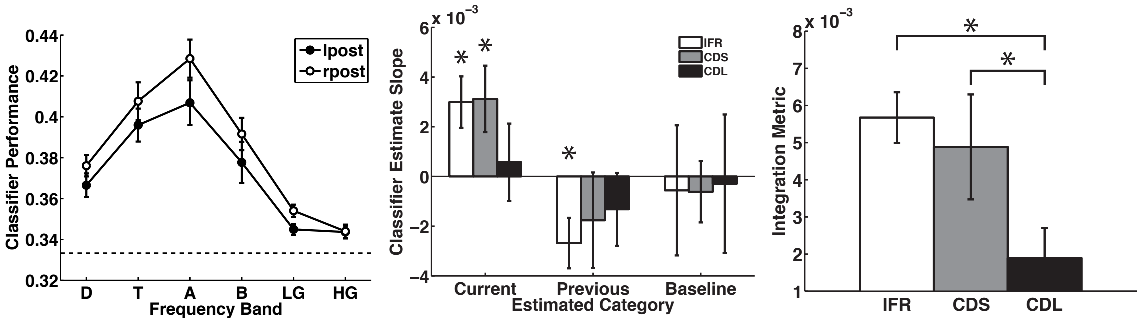


Figure 23. Approximate sensor locations and selected locations from the 10–20 system. Regions of interest are highlighted; ROIs are based on thresholding classifier accuracy arbitrarily to obtain electrodes with accuracy > 0.43. Selected electrodes were then divided into right and left hemispheres; the left ROI was expanded to be symmetric with the right ROI, so it would contain the same number of electrodes.



(a) (b) (c)

Figure 24. (**a**) Classifier performance for the two ROIs (right posterior and left posterior) and six frequency bands (D: delta, T: theta, A: alpha, B: beta, LG: low gamma, HG: high gamma). Performance was significantly above chance ($0.3\bar{3}$; indicated by the dotted line) for all frequency bands for both ROIs. (**b**) Slope of classifier evidence over train positions 1–3, for the current category, the previously presented category, and the baseline category, for beta power measured at the right posterior ROI. A significantly positive slope is observed for the current category in the IFR and CDS conditions. In the IFR condition, the activation of the previous category decreases over time, while the baseline category does not change in activation. (**c**) For each distraction condition, the difference between the slope of the current category and the slope of the previous category. IFR: immediate free recall; CDS: short continual distraction; CDL: long continual distraction.

over time and frequency bin. We carried out a separate cross-validation classification analysis at each electrode, with power at two time bins (0–500 ms and 500–2000) and all frequencies included as features. Classifier performance was significantly above chance at all electrodes ($p < 0.05$, Bonferroni corrected). The highest classifier accuracy was found at posterior electrodes (Fig. 22b). We defined a posterior region of interest by thresholding mean classifier performance to include electrodes where fraction correct was greater than 0.43. We then divided the posterior region into left and right posterior ROIs, excluding two electrodes on the midline. We added 2 electrodes to the left ROI to make it symmetric with the right ROI so that it included the same number of electrodes (Fig. 23).

Neural evidence of persistant category-specific activity

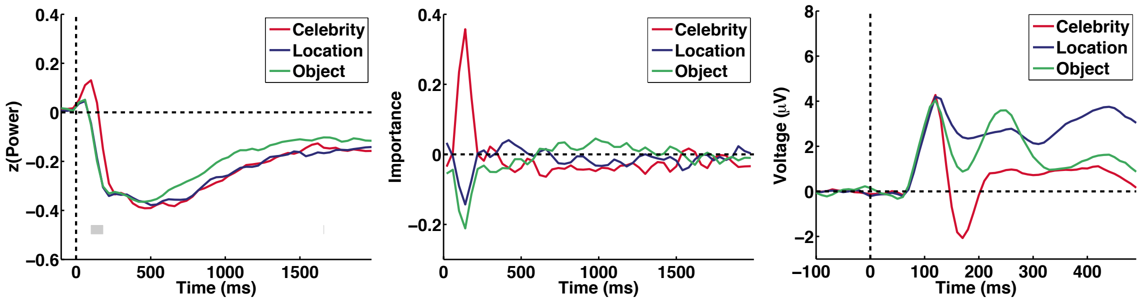
Morton et al. (2013) found that the classifier's estimate of the activation of the category of presented stimuli increased as multiple items from a category were presented in succession. Using a classifier that aggregated information over electrodes, time bins, and frequencies, they observed that classifier estimates increased for the first three successive presentations of items from a given category, then leveled off beyond that. Here, we focused on our two ROIs and six frequency bands of interest in an attempt to gain more information about what oscillatory activity exhibits category-specific activity that persists over multiple stimuli.

Collapsing over distraction conditions, we found that classifier performance was significantly above chance in every frequency band, in both ROIs (Fig. 24a; all $p < 0.05$, Bonferroni corrected). For each distraction condition, ROI, and frequency band, we examined whether classifier estimates of the activation of the currently presented category changed with train position. In the IFR and CDS conditions, we observed a significantly positive slope in the right posterior ROI, in the beta band (IFR: $t(9) = 5.62$, $p = 0.0039$; CDS: $t(9) = 5.23$, $p = 0.0065$). All other slopes were not significantly different from zero. We further examined the slope over train position of classifier evidence in the beta band in the right posterior ROI using a two-way repeated measures ANOVA, with stimulus category and distraction condition as factors. We found no main effect of category ($F(2,18) = 3.00$, $p = 0.075$), a main effect of distraction ($F(2,18) = 4.32$, $p = 0.029$), and no interaction ($F < 1$). This suggests that the increase in classifier evidence with train position was not restricted to one category. This was further confirmed using t-tests of each category and distraction condition. In IFR, there was a significantly positive slope for every category (celebrities: $t(9) = 2.71$, $p = 0.024$, locations: $t(9) = 2.42$, $p = 0.039$, objects: $t(9) = 3.15$, $p = 0.012$). In the CDS condition, locations and objects had a significantly positive slope (locations: $t(9) = 1.85$, $p = 0.049$, one-sided test; objects: $t(9) = 2.44$, $p = 0.037$); the slope for celebrities was not different from zero ($p > 0.05$). In the CDL condition, celebrities had a significantly positive slope ($t(9) = 1.92$, $p = 0.043$, one-sided test); the slopes for the other categories did no differ significantly from zero ($p > 0.05$).

We also investigated whether activity related to a presented category persists into the next train of stimuli from a different category. We used a similar procedure as above, but instead of examining the classifier's estimate of the neural strength of the currently presented category, we examined the estimate of the category presented in the previous train. We calculated the slope of this estimate over train position, to determine whether activity related to the previously presented category slowly decays as a new category is presented. We only observed a significantly negative slope in the right posterior ROI, in the beta band, during IFR ($t(9) = 4.80$, $p = 0.012$). Slope in this ROI and frequency band did not vary by category ($F < 1$) or distraction condition ($F(2,18) = 1.34$, $p = 0.29$), and there was no interaction ($F < 1$). In IFR, celebrities ($t(9) = 2.34$, $p = 0.044$) and objects ($t(9) = 2.08$, $p = 0.034$, one-sided test) had significantly negative slope, while slope for locations was not different from zero ($p > 0.05$). For comparison, we also examined classifier estimates of a third, baseline category, which was neither the currently presented category or the previously presented category. We observed a significantly positive slope in the right posterior ROI in the delta band ($t(9) = 4.28$, $p = 0.025$) and the low gamma band of the left posterior ROI ($t(9) = 4.54$, $p = 0.017$), but no change in the right posterior ROI in the beta band. Subsequent analyses focused on the right posterior ROI in the beta band.

To investigate whether there was any evidence of integrative activity in the CDL condition, we combined the current and previous slope measures by taking the difference between them (Fig. 24c). With this statistic, we found a significant effect in CDL ($t(9) = 2.35$, $p = 0.043$), suggesting that weak but reliable integrative activity is observed even in the long distraction condition. We used this metric to compare the distraction conditions. We found no difference between IFR and CDS ($t(9) = 0.56$, $p = 0.59$), a significant difference between IFR and CDL ($t(9) = 3.08$, $p = 0.013$), and a significant difference between CDS and CDL ($t(9) = 1.85$, $p = 0.049$, one-tailed test).

We next focused on understanding the signal in the beta band at the right posterior ROI, that demonstrated evidence of integrative activity. Overall classifier performance (averaged over electrodes in the ROI) was 0.3913, where chance performance is $0.3\bar{3}$; this was significantly above chance ($t(9) = 7.18$, $p = 5.2x10^{-5}$). Classifier performance varied substantially by category, with the greatest performance for celebrities (mean 0.4677, SEM 0.0270, $t(9) = 4.98$, $p = 0.0008$), and objects (mean 0.3992, SEM 0.0287, $t(9) = 2.30$, $p = 0.047$), and chance performance for locations (mean 0.3070, SEM 0.0142, $t(9) = 1.86$, $p = 0.096$). Classifier performance was signifi-



(a) (b) (c)

Figure 25. (**a**) Average z-scored Beta band power for each category at the right posterior ROI. Shaded areas indicate significant category differences ($p < 0.05$, Bonferroni corrected). (**b**) Importance map for Beta power, averaged over electrodes in the right posterior ROI. (**c**) Average event-related potentials for the three stimulus categories, at an electrode near P010. Robust category differences in amplitude were observed in the N170 component.

cantly greater for celebrities compared to locations ($t(9) = 5.23$, $p = 5.4x10^{-4}$); all other pairwise comparisons were not significant ($p > 0.05$). Figure 25a shows average z-scored beta-band power for each category. A one-way repeated measures ANOVA at each time bin revealed significant category differences in beta power early after stimulus presentation ($p < 0.05$, Bonferroni corrected). There were also trends toward later category differences, around 500–2000 ms after stimulus onset. We calculated an importance map to better understand which features the classifier relied on the most. Classifier weights for each category were multiplied by the average normalized power value at each time bin. Importance values were averaged over all folds of the cross-validation and all electrodes in the right posterior ROI. This analysis revealed that the classifier was most influenced by power early after stimulus presentation, but also had some weighting during the later period (Fig. 25b).

Given our finding of category-specific beta power at right posterior electrodes, with the highest weighted activity being around 100–200 ms, we next asked whether our results were related to variability in the N170 face-selective ERP component. To measure the N170, voltage was low-pass filtered below 40 Hz, down-sampled to 100 Hz, and baseline corrected relative to the average voltage in the 100 ms before stimulus onset. We defined N170 amplitude as the minimum voltage (for each subject and condition) in the range of 135–205 ms post-stimulus onset, at an electrode near PO10 (this electrode was included in the right posterior ROI). There was a significant effect of stimulus category on N170 amplitude (Fig. 25c; $F(2,18) = 34.83$, $p = 6.5x10^{-7}$). For each subject and distraction condition, we calculated the slope of N170 amplitude over train position. We found that the amplitude of the N170 did not change with train position for any of the distraction conditions ($p > 0.05$), suggesting that the beta-band power results were not related to fluctuations in the N170.

Discussion

Computational models of episodic memory propose that persistent activity during encoding carries information about recently presented stimuli, allowing stimuli in a list to become associated. Morton et al. (2013) found evidence of category-specific oscillatory activity that persisted over multiple presented stimuli, and found that this activity predicted individual differences in category clustering. Here, we tested whether category-specific integrative oscillatory activity persists when there is distraction between stimuli, or whether distraction prevents information about multiple stimuli from being represented. We found evidence for persistent category-specific brain activity: Patterns of oscillatory activity in the beta band became more category-specific as multiple items from a category were presented in succession. Furthermore, there was evidence that activity related to recently presented categories decayed slowly as items from a different category were presented. This integrative activity was attenuated by inter-item distraction. Consistent with the hypothesis that integrative category-specific neural activity during encoding supports subsequent organization by category, we found that category clustering was decreased in the inter-item distraction conditions. In contrast, temporal clustering was not affected by inter-item distraction, consistent with prior research (Howard and Kahana, 1999).

It should be noted that the effect of distraction on category clustering and neural integration were somewhat different. Category organization was reduced in both the short and long distraction conditions (Fig. 21b), while neural integration was only significantly reduced in the long distraction condition (Fig. 24b). One possible reason for this discrepancy is that partial co-activation of related items is not sufficient for semantic relationships between them to be discovered; as a result, partial integration in the short distraction condition may not lead to enhanced semantic organization among adjacent items. Another possibility is that there are multiple mechanisms by which inter-item distraction can disrupt semantic organization; for example, distraction may increase the influence of proactive interference by making it more difficult to target the just-presented list. As a result, participants may rely less on category cues during recall in order to reduce proactive interference. Such an effect might be partially independent of changes to integrative activity during encoding caused by distraction.

The dissociation between temporal and category clustering in their response to inter-item distraction provides a puzzle for existing models of episodic memory search to address. Pure dual-store models like SAM (Raaijmakers and Shiffrin, 1980) or FRAN (Anderson, 1974) may be able to account for the decrease in semantic organization with increasing inter-item distraction, by assuming that distraction empties out short-term memory, preventing semantic relationships from being discovered during encoding. However, buffer-based models have difficulty accounting for the finding, reported here and elsewhere (Howard and Kahana, 1999), that temporal organization is not affected by inter-item distraction (Davelaar et al., 2005). A hybrid model with both a short-term store (responsible for augmenting semantic organization) and a temporal context representation (responsible for temporal organization at multiple timescales) such as the context-activation model (Davelaar et al., 2005) may be able to account for the effects we observed. Alternatively, a purely context-based account may be possible. According to retrieved-context models, inter-item distraction should disrupt evolution of context (Sederberg et al., 2008), preventing the accumulation of semantic information over multiple items. While temporal organization can, in principle, be driven from any slowly changing signal that becomes associated with studied items (e.g. contrast Howard and Kahana 1999, 2002a), the enhancement of semantic organization in adjacent items relies on a buildup of information related to the specific studied items. This information may be disrupted by inter-item distraction, leading to a disproportionate decrease of semantic organization.

We found that classification accuracy at right posterior electrodes was above chance in each of the six frequency bands we examined (Fig. 24a); however, classifier evidence was only sensitive to the recent history of stimuli in the beta band (16–25 Hz; Fig. 24b). These results are consistent with the findings of Morton et al. (2013), and extend them by localizing integrative activity to a specific topography and frequency band. While activity related to stimulus identity has previous been found in the gamma band (Jacobs and Kahana, 2009), the present study and Morton et al. (2013) both found evidence that category-specific activity is present in a wide range of frequencies. As discussed above, one possible origin of the integrative activity in the beta band is maintenance of an increasing number of items from the same category in a working-memory buffer. Although some prior research suggests items in a working-memory buffer are represented in gamma-band activity (Fuentemilla et al., 2010), other researchers have found evidence that the beta-band is also involved in working-memory maintenance (Axmacher et al., 2010). One possible reason we did not find evidence of integrative activity in the gamma band is because attenuation of high-frequency activity at the scalp may have decreased the signal-to-noise ratio in the gamma band (Nunez and Srinivasan, 2006). Although classifier performance in the gamma band is above chance, it is substantially lower than beta-band performance (Fig. 24). Morton et al. (2013) compared classifier accuracy for scalp EEG and ECoG recordings of a similar experiment with similar stimuli, and found that the difference in accuracy of a pattern classifier between scalp EEG and ECoG was greater for high gamma (65–100 Hz) activity than for low gamma (25–55 Hz) activity, suggesting that category-specific activity above about 65 Hz is particularly attenuated at the scalp. However, this explanation for our lack of evidence for gamma-band integrative activity is preliminary, especially since category-specific activity related to working memory maintenance has been observed in the gamma band using MEG (Fuentemilla et al., 2010). Future work should examine whether beta-band category-specific activity is coupled to the phase of theta oscillations; this type of cross-frequency interaction has been proposed to underly representation of items in a multi-item working memory buffer (Jafarpour et al., 2013).

CHAPTER IV

Using a joint model of episodic and semantic associations to predict recall sequences

Introduction

Theories of episodic memory have often focused on findings that are observed regardless of the specific material being learned, such as the primacy, recency, and contiguity effects observed in free recall (Raaijmakers and Shiffrin, 1980; Howard and Kahana, 2002a; Brown et al., 2007; Sederberg et al., 2008). However, research has demonstrated that semantic associations can strongly influence performance in free recall, a nominally episodic task (Bousfield, 1953; Glanzer, 1969; Romney et al., 1993; Howard and Kahana, 2002b). Participants tend to recall semantically related items successively, even when related items are never studied adjacent to one another (Bousfield, 1953). Semantic clustering is observed both when the study list contains obvious taxonomic category structure (Bousfield, 1953; Puff, 1966) and when there is no systematic semantic structure (Schwartz and Humphreys, 1973; Romney et al., 1993; Howard and Kahana, 2002b).

Despite the importance of semantic knowledge for shaping new episodic memories, there is little consensus about the specific mechanisms involved (for various proposals, see Sirotin et al. 2005; Polyn et al. 2009; Cohen 1963; Kimball et al. 2007). The complicated relationship between semantic associations and other influences on recall (Howard et al., 2007) further complicates this effort. Given the complexity of memory search, development of computational models that take semantic associations into account is important to improve understanding in this domain.

A number of researchers have proposed models of memory search that include hypotheses about how associations in semantic memory contribute to search (Anderson, 1972; Romney et al., 1993; Sirotin et al., 2005; Kimball et al., 2007; Socher et al., 2009; Polyn et al., 2009). A difficulty in evaluating these models is that they must include some model of semantic knowledge. Furthermore, measures of semantic organization must also assume some model of semantics, and these measures may be used to evaluate the performance of models of memory search (e.g. Raaijmakers and Shiffrin 1980; Polyn et al. 2009). This can lead to a situation where the same semantic model is used both as part of the internal structure of a model and as a means to evaluate it, causing a circularity that can complicate interpretation (Manning and Kahana, 2012; Polyn et al., 2009).

A related issue complicates the measurement of semantic organization in the first place. Measurement of semantic organization requires specification of a modeling framework that includes a baseline model that estimates the level of organization expected due to chance, a model of semantic associations, and a model of how these semantic associations influence search (Stricker et al., 2002). The baseline model has often been assumed to be random sampling from the list (Roenker et al., 1971; Stricker et al., 2002). However, the assumption of random sampling fails to take into account the influence of regularities in recall on the incidence of clustering; this is a particularly critical issue when considering the effects of a manipulation of presentation order (e.g. blocked vs. random presentation of categorized items) on category clustering (Puff, 1966). However, little work has been done to control for non-semantic influences on recall performance, such as primacy, recency, and temporal contiguity effects (Murdock, 1962; Kahana, 1996). Romney et al. (1993) demonstrated a method for controlling for overall recall of individual items, but this method fails to take into account sequential dependencies such as temporal organization, the tendency of participants to successively recall items presented near to one another on the list (Kahana, 1996). Temporal organization is an important and near-ubiquitous phenomenon observed in memory search tasks (Kahana, 1996; Sederberg et al., 2010) that can affect measures of semantic organization (Puff, 1966; Morton et al., 2013). Because of temporal organization, traditional measures of semantic organization which do not take the ordering of the input list into account, such as ratio of repetition (Bousfield, 1953), adjusted ratio of clustering (Roenker et al., 1971), and list-based clustering (Stricker et al., 2002), will be inflated whenever semantically related items are presented in proximity. This is an important problem to address, since much of the classic literature on semantic organization has focused on manipulations of presentation order (e.g. Glanzer 1969; Borges and Mandler 1972; for a review, see Puff 1974), and these results have motivated theory development (Anderson, 1972).

To address these issues with measurement of semantic organization in memory search, and to aid testing of models of semantic organization, we propose a new modeling framework that involves evaluating models of memory search that take both semantic and non-semantic influences into account. In this framework, models are evaluated based on their ability to predict behavior on the basis of individual recall sequences, rather than summary statistics; this allows us to evaluate different models of semantic influences on recall while avoiding the circularity of assuming a "true" model of semantic associations. Furthermore, the framework allows measurement of semantic organization while controlling for other influences on free recall behavior.

We base this framework on the context maintenance and retrieval model (CMR), a well-developed model of memory search that successfully accounts for a range of behavior in free recall (Polyn et al., 2009). In order to account for the influence of semantic associations, we also incorporated a model of semantic association. We tested two competing models of semantic similarity, based on latent semantic analysis (LSA; Landauer and Dumais 1997) and word association spaces (WAS; Steyvers et al. 2004). A full model of search also requires specification of how semantic associations influence search. While each model we examined assumes that temporal organization is driven by cuing with temporal context, we tested two distinct mechanisms that might contribute to semantic organization. One semantic cuing mechanism was based on context-based semantic cuing as in the previously reported version of CMR (Polyn et al., 2009); we contrasted this with item-based cuing, which has been assumed by a number of researchers (e.g. Sirotin et al. 2005).

Rather than evaluating the ability of each model to capture summary statistics reflecting effects of recency, primacy, contiguity, semantic organization, etc., we instead evaluate models based on their ability to predict the sequence in which individual items are recalled. For a given model variant and set of model parameters, we determine the likelihood of that model producing the entire set of recall sequences observed in the experiment. The set of parameters that maximizes the likelihood of the model is calculated for each model variant, and model variants are compared based on their maximum likelihood fits. While maximum likelihood estimation provides important benefits such as high consistency and efficiency in parameter estimation (Myung, 2003), little work has used this technique with models of free recall (see Farrell and Lewandowsky 2008 for a likelihood-based fit of partial sequences). The dearth of likelihood-based fitting in models of free recall may stem from the historical emphasis on fitting certain summary statistics such as the serial position curve (e.g. Sederberg et al. 2008), as well as the common use of simulation models for which exact likelihoods cannot easily be calculated (e.g. Raaijmakers and Shiffrin 1980; Sederberg et al. 2008; Polyn et al. 2009; Davelaar et al. 2005; Farrell 2012). Furthermore, free recall sequences exhibit a number of sequential dependencies (Polyn et al., 2009), so accurate prediction of recalls requires taking the prior history of recalls into account. CMR makes predictions based the entire history of recalls (Socher et al., 2009), and has been shown to account for a number of important sequential dependencies (Polyn et al., 2009; Healey and Kahana, 2014), including higher-order effects of compound temporal cuing (Lohnas and Kahana, 2014). Here, we use a variant of CMR that allows direct calculation of the probability of entire recall sequences, allowing exact calculation of likelihood for free-recall data. Maximum likelihood provides an unbiased measure for evaluating competing models of memory search, without having to define an arbitrary subset of measures to emphasize, and without assuming a "true" model of semantic associations.

We find that both models of semantics allow an improvement over a base model with no semantics, and that WAS-based models provide the best prediction accuracy. We also found that an item-based semantic cuing mechanism is most consistent with the data, in contrast to the context-sensitive semantic cuing assumed by the original version of CMR. Finally, we discuss the potential of this modeling framework to allow measurement of variation in semantic organization (between condition, participant group, etc.) while controlling for other factors such as temporal organization.

Methods

Participants

Participants included 41 people (14 female) between the ages of 18 and 30. Participants were recruited as part of a series of studies designed to examined electrophysiological correlates of encoding and retrieval in free recall. We focus on the first study of the series, which included 4 sessions for each participant.

Stimuli

The stimulus pool was designed to facilitate analysis using word association spaces (WAS) analysis. Words were chosen from the pool of 5018 words used by Steyvers et al. (2004) to construct their word association spaces. Using the CELEX2 English database (Baayen et al., 1995), we identified a subset of this pool that contained 3925 nouns. Three raters judged whether each word in this pool was appropriate for the judgment tasks used in the task (size and animacy judgments; see *Procedure* for details); words were excluded if they were abstract or were highly ambiguous for either of the judgment tasks. The exclusion process yielded a pool of 1655 words[1]. Three additional raters performed both judgment tasks for each word, providing a rough estimate of the average response for each word.

Procedure

On each trial, a list of 24 words was presented; each word was presented with a task cue above it, indicating the judgement that the participant should make for that word ("Size" or "Living/Non-living"). Each stimulus was presented for 3 s. The two tasks were a size judgment ("Does this word refer to an object that could fit into a regular shoebox?") and an animacy judgment ("Does this word refer to something that is alive?"). Participants made their response to each word using the index and middle fingers of their right hand to press one of four keys labeled "Big", "Small", "Living", and "Nonliving". If participants failed to respond during the time the item was onscreen, a beep was presented along with a message asking them to respond more quickly. If participants pressed one of the two keys that was inappropriate for the current task (e.g. "Living" if asked to make a size judgment), a beep was presented with a message asking them to press one of the appropriate keys.

Using the pilot data described in *Stimuli*, where three participants were asked to rate each word

[1] During the course of the study, an additional 17 words were excluded due to homophony with other words in the pool.

using both encoding tasks, we calculated the average response for each word and encoding task. Lists were chosen so that the average rating for each possible response was no more than 70%, averaged over all words judged with the relevant task.

The screen was blank for a 0.8–1.2 s inter-stimulus interval between each word. Immediately following the list, a row of asterisks appeared, along with a beep, indicating the start of the recall period. Participants were given 90 seconds to recall as many words as they could remember from the most recent list, in any order.

There were two trial conditions: control and task shift. On control lists, each word was studied with the same encoding task. On task-shift lists, half of the items were studied with each encoding task. The number of control and shift lists was balanced within each session. Lists were ordered in groups of four, where each group contained two control (one of each task) and two shift lists, and the order of lists was randomized within each group. There were four sessions held on separate days; each session included 12 lists. Here, we focus on the control lists (24 for each subject); all analyses are collapsed over encoding task.

Models of semantic associations

Word association spaces (WAS) similarity was derived from University of South Florida free-association norms (Nelson et al., 2004; Steyvers et al., 2004). We used the 400-dimension singular value decomposition of the $S_{ij}^{(2)}$ measure described by Steyvers et al. (2004), which is freely available online[2]. We defined the WAS similarity between two words as one minus the cosine of the angle between their corresponding vectors. We also computed similar vectors by applying the latent semantic analysis (LSA) algorithm (Landauer and Dumais, 1997) to the Touchstone Applied Science Associates, Inc. (TASA) corpus, resulting in a 400 dimensional vector for each word. We then calculated the LSA similarity between each pair of words as one minus the cosine of the angle between their LSA vectors. LSA similarity values were not available for two words. Therefore, we excluded from all analyses 27 lists that included either of those words, leaving 957 lists considered here.

[2]http://psiexp.ss.uci.edu/research/software.htm

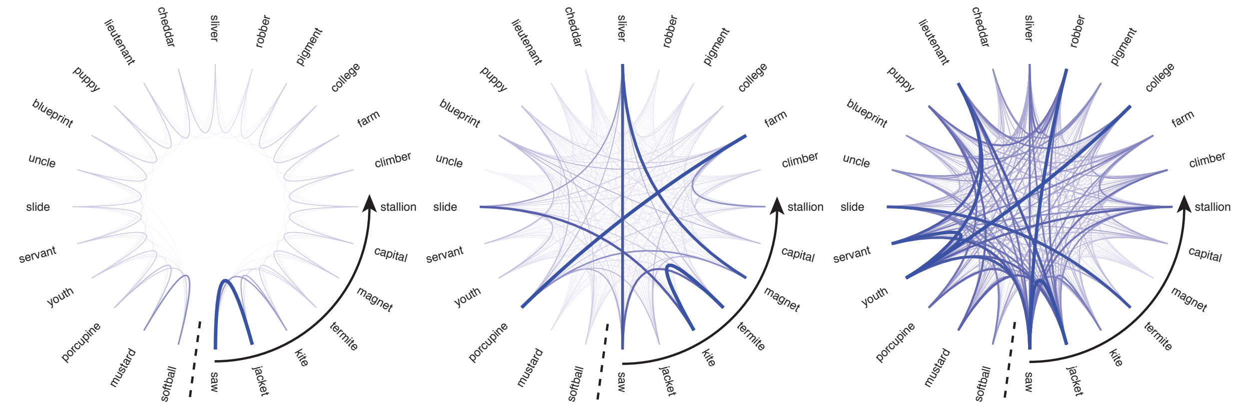


Figure 26. Visualization of similarity between pairs of items, for a list from the experiment. (**a**) Similarity between words, based on the best-fitting base CMR model with no semantics. Here, similarity between a pair of words is defined as the probability of transitioning between them (averaged over both directions), for a version of the model with no semantic associations (see text for details). The list starts at the bottom and procedes counterclockwise. Greater line saturation and thickness indicates greater similarity between the words being connected. Versions of the model that add semantic similarity are valuable inasmuch as they can predict deviations from this pattern of behavior. (**b**) Similarity between words, based on cosine similarity in word association spaces (WAS). (**c**) Cosine similarity based on latent semantic analysis (LSA). Values have been scaled to cover the same range as in (b).

Visualizations of WAS and LSA similarity for a sample list in the experiment are shown in Figure 26b,c[3]. Note that WAS has relatively sparse connectivity, while LSA similarity shows many connections of moderate strength. For comparison, Figure 26a shows a measure of similarity derived from the predictions of a version of CMR with no semantic associations (see *Model of memory search* for details, and Table 6 for parameter values). We simulated the retrieval of each item, and calculated the probability of next recalling each other item in the list. For this simplified example, we assume that all list items are available (have not already been recalled), and only consider the influence of the last item recalled. For the purposes of this visualization, similarity between two items was defined as the probability of transitioning between them (in either direction). As shown in Figure 26a, learning of item-context associations causes items presented in adjacent positions become more strongly associated than items from distant positions. In our simulations, we examined whether including semantic associations based on WAS or LSA in the

[3]Visualization created using code modified from the Schemaball package: http://www.mathworks.co.uk/matlabcentral/fileexchange/42279-schemaball

model would allow it to improve its ability to predict recall sequences. The next section discusses the mechanisms the model uses to form these associations and guide memory search.

Model of memory search

We used a modified version of the context maintenance and retrieval model (CMR) as a framework to evaluate the impact of different models of semantic associations and different semantic cuing mechanisms on behavior in free recall. In CMR, there are two interacting representations, a context layer and a feature layer. When an item is studied, a representation of it becomes active on the feature layer. This also causes information about the item to be retrieved and pushed into the context representation. Context changes slowly over time and reflects a recency-weighted average of information related to recently presented stimuli (Fig. 27). Studied items become associated to the context that was active when they were presented, so that context can serve as a cue to retrieve items, and recalled items can retrieve the context that is associated with them. When an item is recalled, its associated context is reinstated, and used to cue for another item on the list. Items that are associated with similar states of context (such as adjacent items in a list) tend to be good cues for one another (Fig. 26a). See *Formal description of the model* for further details about model mechanisms. The mechanisms of item-context association, contextual cuing, and context reinstatement allow the model to account for a number of effects in recall, including recency and temporal contiguity effects (Howard and Kahana, 2002a; Howard, 2004; Howard et al., 2005; Sederberg et al., 2008; Polyn et al., 2009).

Polyn et al. (2009) introduced CMR, which is based on the temporal context model (TCM; Howard and Kahana 2002a). CMR added, among other things, a mechanism to explain how pre-experimental semantic associations influence recall. For CMR, Polyn et al. (2009) assumed that semantic associations influence contextual cuing. They assumed that items will be well-supported by context to the extent to which semantically related items are reactivated in context (Fig. 27). Notably, this context-based semantic cuing contrasts with an alternative mechanism that has often been assumed, where the just-recalled item serves as a cue for semantically related items (e.g. Sirotin et al. 2005). For this type of cuing, which we refer to as item-based semantic cuing, only

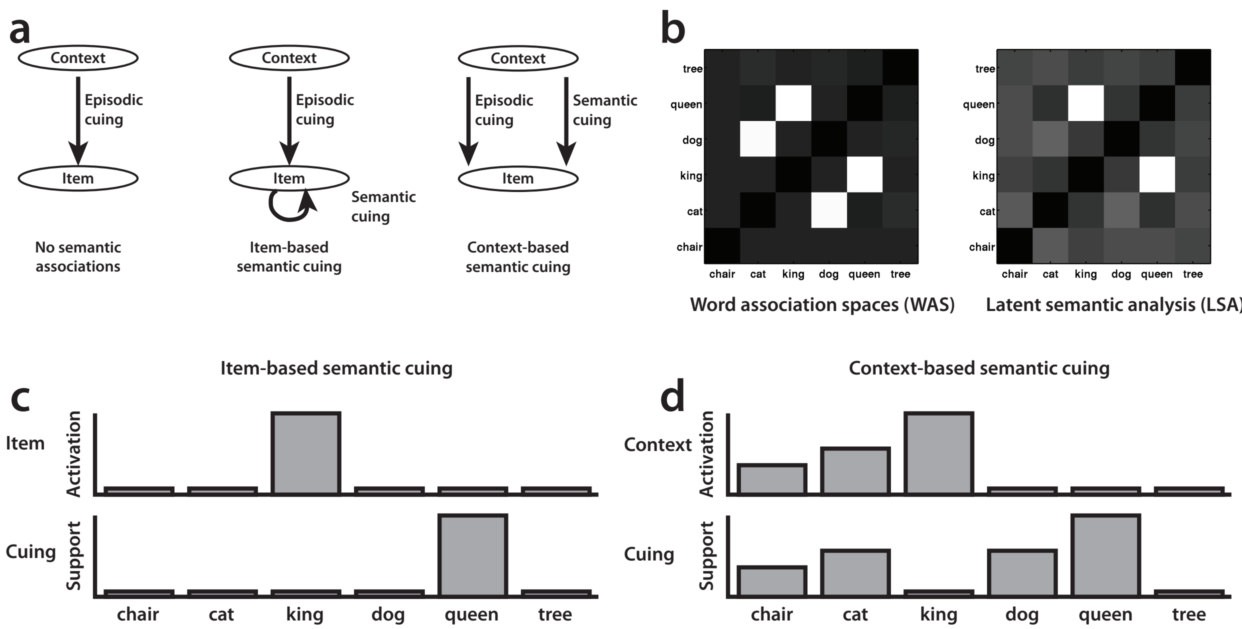


Figure 27. (**a**) Schematic of cuing mechanisms. Left: Model with no semantic associations. Recall is driven solely by episodic associations between items and context. Center: In addition to episodic associations, retrieved items will cue for other semantically related items. Right: Only context is used as a cue to retrieve items; context projects through both episodic and semantic associations. (**b**) Semantic similarity according to cosine similarity in the word association spaces (WAS) and latent semantic analysis (LSA) models, for a sample list of words. Lighter colors indicate greater similarity. Self-similarities are set to 0. (**c**) Example of item-based semantic cuing using WAS. In this example, king has just been recalled, and has been reactivated on the item layer. It is then used as a memory cue to retrieve related items. During the subsequent recall attempt, we assume that the king unit is inhibited, so that it does not become activated again. The queen unit becomes most strongly activated, since it is strongly semantically related to king. (**d**) Context-based semantic cuing, using the same example as in (c). In the context layer, the context associated with king in the study list has become reactivated. This reactivated context contains information about king, as well as chair and cat, since they were active in context when king was studied. When context projects through semantic associations, it primarily activates queen, but dog also becomes activated since it is semantically associated to cat, which has become partially reactivated in context. Cat and chair also become activated, because context units are connected to their corresponding item units.

the identity of the just-recalled item matters. In contrast, context-based semantic cuing assumes that the items presented just prior to the studied item will also have an impact on semantic cuing. Furthermore, context integrates retrieved context from all retrieved items (with the most recently retrieved having highest weight). Therefore, CMR predicts that semantic organization after recall of a given item should be sensitive to (1) the items preceding that item in the list and (2) the items that were recalled prior to that item. Although Polyn et al. (2009) showed that CMR can produce a reasonable overall amount of semantic organization, they focused on organization conditional only on the just-recalled item. The more nuanced predictions of the model have not been evaluated.

Using the framework of CMR, we contrasted the candidate mechanisms of item- and context-based semantic cuing. For item-based cuing, we assumed that the item representation of a recalled item is reactivated and used to retrieve items that are semantically related. For context-based cuing, we assumed that all items activated in the context representation reactivated their semantic associates (Fig. 27). Both of these cuing mechanisms were paired with both of the models of semantic association we examined (described in *Models of semantic associations*). We evaluated five variant models: (1) a base model with no semantics, (2) LSA with context-based semantic cuing, (3) LSA with item-based semantic cuing, (4) WAS with context-based semantic cuing, and (5) WAS with item-based semantic cuing. We compared these models based on their ability to predict the sequences of individual recalls that we observed in the experiment.

Likelihood calculation

At the beginning of each recall period, we use the model to calculate the probability of recalling no items and the probability of recalling each individual item on the list. From these probabilities, we take the probability of whatever recall event actually took place (for example, recalling item 24 in the list), and take the log of this probability (to avoid precision issues caused by very low probabilities). If the participant recalled item 24, we then simulate recall of item 24 in the model, and update the context cue accordingly. This new state of the model is then used to calculate stopping and recall probabilities, and the probability of the recall event that occurred next in the experiment is added to the previous log probability. This process is repeated until we reach the end of the recall sequence being examined. We sum the log likelihood over all lists and subjects, to obtain the log likelihood of the entire dataset, given the specified model and parameters.

For a given model variant with specified parameters, we can calculate the probability of generating the recall sequences observed in the experiment. There is one caveat, however; since the current implementation of the model does not account for repeats and intrusions, it cannot estimate the probability of these recall events. To avoid this issue, we removed repeats and intrusions from the recall sequences. We excluded 9.42% of recall attempts; of all recall attempts, 4.52% were repeats, 1.74% were prior-list intrusions, and 3.32% were extra-list intrusions.

Model comparison

For each model variant, we estimated the maximum likelihood parameter set using a version of the differential evolution algorithm (Storn, 2008). We used a MATLAB-based implementation of differential evolution, based on code developed by Price et al. (2005)[4]. For each search, we began with 50 individuals at randomly chosen points in the parameter space. Mutation was done using the local-to-best strategy, with a step weight of 0.85 and crossover probability of 1. Iterations of the algorithm were run until the maximum log likelihood over all individuals had not changed more than 0.01 over the last 50 generations.

Model performance was quantified using the Aikake Information Criterion (AIC; Wagenmakers and Farrell 2004). For each model, we calculated AIC with a correction for finite samples:

$$AIC_c = -2\log L + 2V + \frac{2V(V+1)}{(n-V-1)}, \tag{IV.1}$$

where L is the maximum likelihood value for the candidate model, V is the number of free parameters, and n is the number of estimated data points.

We compared model performance using AIC weights, defined as

$$w_iAIC = \frac{exp(-\frac{1}{2}\Delta_i AIC)}{\sum_{k=1}^{K} exp(-\frac{1}{2}\Delta_k AIC)}, \tag{IV.2}$$

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

where $\Delta_i AIC$ is the difference in AIC_c between a given candidate model and the best-fitting model in the set.

Analysis of recall behavior

In addition to the likelihood-based model selection procedure described above, we examined standard summary statistics to gain further information about the performance of participants and the performance of each model. We used the model to generate simulated recall sequences. For each recall attempt, we calculated the probabilities of each recall event (recalling an item or stopping recall), using the same procedure described in *Likelihood calculation*. We then randomly chose an event based on this probability distribution, and updated the state of the model accordingly. Each recall period was simulated in this manner until a stop event was chosen. To calculate summary statistics for each model, we simulated each list in the experiment 100 times, and calculated each statistic averaged over the 100 simulated replications of the experiment.

Behavior in free recall can be described in terms of three stages: initiation, transitions, and termination (Kahana, 2012). We measured recall initiation by calculating the probability of first recalling an item as a function of the serial position in which it was presented in the list. After the first recall, transitions between recalled items exhibit two major forms of organization: temporal clustering and semantic clustering.

Temporal clustering is the tendency of participants to successively recall items that were presented adjacent to one another in the list (Kahana, 1996). We measure temporal clustering using the conditional response probability as a function of lag (lag-CRP; Kahana 1996). The lag-CRP divides transitions between recalled items in terms of distance, or lag, between the positions in the list in which they were presented. For example, if a participant just recalled item 10 in the list, and then recalled item 12, that transition would have a lag of +2. We calculated the probability of making a transition of each lag, conditional on that lag being available for recall (an item was considered unavailable if there was no item presented at that serial position, or if that item had already been recalled previously).

We also measured semantic clustering using a related measure, the semantic-CRP (Howard and Kahana, 2002b; Sederberg et al., 2010). First, we tallied the number of times each participant made a transition from item i to item j, for each item in the stimulus pool. We also tallied a separate count of the number of times that each participant *could have* made each possible transition between words, given the words that were still available at each point in recall. A given transition between items i and j was not counted as possible if item i was never recalled. We then determined a set of semantic similarity bins that we used to group together inter-item transitions. Prior implementations of the semantic-CRP analysis have generally used bins that contain deciles (Healey and Kahana, 2014) or percentiles (Howard and Kahana, 2002b; Howard et al., 2007). However, because semantic similarity values based on WAS and LSA are highly positively skewed (Manning and Kahana, 2012), this results in many bins at low similarity values, and very few bins at higher similarity values. This skew is particularly problematic when estimating the slope of the semantic-CRP, since the regression will be disproportionately affected by the small number of high-leverage points from high-similarity bins. To better estimate CRPs for the full range of similarity values, we took a different strategy of determining bin sizes so that we obtain a minimal sample size at each bin (see Sederberg et al. 2010 for another example of unequal bin sizes used for this analysis). First, we obtained the semantic similarities for each inter-item transition that was possible at least once over all recall sequences in the study, based on the semantic similarity measure of interest (LSA or WAS). Starting from the highest similarity value, we decreased the lower limit of the bin by increments of 0.01 until there were at least 10 possible transitions per subject on average. After defining a bin, the lower limit of that bin became the upper limit of the next bin, and the process was repeated. The center of each bin was defined as the mean similarity value over all possible transitions within that bin. We determined the bins from the actual data, then applied these bins to the simulated data from our model variants.

We also measured the properties of recall termination by calculating stop probability as a function of output position. We excluded repeats and intrusions when calculating output position, so that the probability of stopping at output position list length $+1$ is unity. Finally, we calculated the standard summary measure of the serial position curve, which shows the probability of recalling

Parameter Type	Parameter	Description
Context Updating	β_{enc}	Rate of context drift during encoding
	β_{start}	Amount of start-list context retrieved at start of recall
	β_{rec}	Rate of context drift during recall
Associative Structure	α	Initial strength of context-to-item connections
	δ	Initial strength of the diagonal of M^{CF}
	s	Scaling of semantic association strengths
	γ	Amount of experimental context retrieved by a recalled item
	ϕ_s	Scaling of primacy gradient in learning rate on M^{CF}
	ϕ_d	Rate of decay of primacy gradient
Recall Dynamics	τ	Sensitivity parameter of the Luce choice rule
	θ_s	Scaling of the stop probability over output position
	θ_r	Rate of increase in stop probability over output position

Table 5. List of model parameters, with a brief description of each.

each item as a function of its serial position in the list.

For each measure of recall behavior, we calculated confidence intervals using a bootstrap procedure. For each of 5000 samples, we sampled subject means with replacement, and calculated a simulated group mean. We set the confidence interval to include the middle 95% of the simulated group means.

Formal description of the model

Here, we give a formal description of the equations that define CMR's structure and behavior. Table 5 provides an overview of the parameters that control the behavior of the model.

There are two interacting representations, a context layer F and a context layer C. The feature layer is connected to the context layer through associative connections represented by $\mathbf{M}^{FC}$, 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 $\mathbf{M}^{CF}_{\text{exp}}$.

In the present simulations, we are particularly interested in structure of of the pre-experimental weights. For all model variants, we set the pre-experimental item-to-context associations according to

$$\mathbf{M}^{FC}_{\text{pre}(i,j)} = \begin{cases} 1-\gamma, & \text{if } i = j \\ 0, & \text{if } i \neq j. \end{cases} \tag{IV.3}$$

This simply connects each unit on F to each corresponding unit on C. The γ parameter scales the relative strength of newly formed experimental associations, relative to pre-experimental associations.

For the base model, which included no model of semantic similarity, we set the pre-experimental context-to-item associations according to

$$\mathbf{M}^{CF}_{\text{pre}(i,j)} = \begin{cases} \delta, & \text{if } i = j \\ \alpha, & \text{if } i \neq j. \end{cases} \tag{IV.4}$$

We also tested a form of the model where $\mathbf{M}^{CF}_{\text{pre}}$ was set to $\mathbf{0}$. Through a series of model comparison analyses (not reported here), we found that freeing both the δ and α parameters substantially improved the fit according to AIC. The δ parameter is similar to the γ^{CF} parameter used by Sederberg et al. (2008). Our implementation is different from theirs in that α is free to be non-zero, and some model variants also include the addition of semantic similarity strengths. For model variants including context-based semantic cuing, we set context-to-item associations according to

$$\mathbf{M}^{CF}_{\text{pre}(i,j)} = \begin{cases} \delta, & \text{if } i = j \\ \alpha + s\mathbf{M}^{\text{sem}}_{i,j}, & \text{if } i \neq j, \end{cases} \tag{IV.5}$$

where $\mathbf{M}^{\text{sem}}_{i,j}$ gives the semantic similarity between items i and j according to either WAS or LSA, and s is a scaling parameter (cf. Polyn et al. 2009). In other words, we used a linear transform to map WAS- or LSA-based semantic cosine similarity values to semantic strengths in the model, where α is an intercept parameter, and s is a slope parameter. The diagonal of $\mathbf{M}^{\text{sem}}$ is set to 0, so that self-strengths are solely determined by the δ parameter.

At the start of the list, context is initialized with a state that is orthogonal to the pre-experimental context associated with each item. Similarly, items are assumed to be orthonormal to each other; 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 $\mathbf{c}_i^{\mathrm{IN}}$ is retrieved and is input to the context layer to update the current state of context. The input to context is

$$\mathbf{c}_i^{\mathrm{IN}} = \mathbf{M}^{FC}\mathbf{f}_i = \mathbf{M}^{FC}_{\mathrm{pre}}\mathbf{f}_i, \tag{IV.6}$$

since $\mathbf{M}^{FC}_{\mathrm{exp}}$ is assumed to be zero at the start of the list. The retrieved pre-experimental context $\mathbf{c}_i^{\mathrm{IN}}$ is then normalized to have length 1.

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

$$\mathbf{c}_i = \rho_i \mathbf{c}_{i-1} + \beta_{\mathrm{enc}}\mathbf{c}_i^{\mathrm{IN}}, \tag{IV.7}$$

where ρ_i is set so that the length of $\mathbf{c}_i$ is 1, according to

$$\rho_i = \sqrt{1 + \beta_{enc}^2[(\mathbf{c}_{i-1} \cdot \mathbf{c}_i^{\mathrm{IN}})^2 - 1]} - \beta_{enc}(\mathbf{c}_{i-1} \cdot \mathbf{c}_i^{\mathrm{IN}}). \tag{IV.8}$$

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

$$\Delta\mathbf{M}^{FC}_{\mathrm{exp}} = \gamma \mathbf{c}_i \mathbf{f}_i'. \tag{IV.9}$$

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

$$\Delta\mathbf{M}^{CF}_{\mathrm{exp}} = \phi_i \mathbf{f}_i \mathbf{c}_i'. \tag{IV.10}$$

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 (Sederberg et al., 2008). ϕ_i depends on the serial position i of the studied item:

$$\phi_i = \phi_s e^{-\phi_d (i-1)} + 1. \tag{IV.11}$$

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

Before initiating recall, we assume that some amount of the pre-list context is reinstated. We assume that context is updated according to

$$\mathbf{c}_{\text{start}} = \rho_{N+1}\mathbf{c}_N + \beta_{\text{start}}\mathbf{c}_0, \tag{IV.12}$$

where $\mathbf{c}_{\text{start}}$ is the state of context at the start of free recall, N is the number of items in the list, $\mathbf{c}_0$ is the state of context at the start of the list before any items have been presented, and ρ_{N+1} is calculated according to Equation IV.8. This mechanism is consistent with evidence that participants sometimes recall the start of the list and use that event as a cue (Laming, 1999). We found that addition of this start-list context reinstatement allowed a better fit of the primacy effect than the learning-rate gradient alone.

At each recall attempt, the current state of context is used to attempt retrieval of an associated item. First, the activation of each item $\mathbf{a}$ is determined according to

$$\mathbf{a} = \mathbf{M}^{CF}\mathbf{c}'. \tag{IV.13}$$

In order to avoid the possibility of the model assigning a probability of 0 to any possible recall, we set a minimal activation for each item of 10^{-6}.

At each recall attempt, we calculated the probability of stopping recall and outputting no item. Probability of stopping recall varies as a function of output position j (where $j = 0$ for the first attempt), according to

$$P(\text{stop}, j) = \theta_s e^{j\theta_r}, \tag{IV.14}$$

where θ_s and θ_r are free parameters that determine the scaling and rate of increase, respectively, of the exponential function.

The probability $P(i)$ of recalling a given item i is defined conditional on recall not stopping at that position, and varies with activation strength, according to

$$P(i) = (1 - P(\text{stop})) \frac{\mathbf{a}_i^\tau}{\sum_k^N \mathbf{a}_k^\tau}, \tag{IV.15}$$

where τ is a sensitivity parameter that determines the contrast between well-supported and poorly supported items. High values of τ will cause a greater influence of differences in support, while low values will cause relatively uniform probabilities of recalling each item.

If an item is recalled, then that item is reactivated on F. The reactivated item is then used to retrieve both experimental and pre-experimental context, according to

$$\mathbf{c}_i^{\text{IN}} = \mathbf{M}^{FC}\mathbf{f}_i. \tag{IV.16}$$

Context is then updated using Equation IV.7, and used to cue for another recall attempt. The process continues until the model reaches the end of the recall sequence.

Item-based semantic cuing

In addition to context-based cuing, we examined a model that used item-based semantic cuing. In this model, contextual cuing worked as before, but semantic associations were not included in $\mathbf{M}^{CF}$. For each recall attempt, the feature-layer vector corresponding to the last recalled item (or, for recall initiation, the last item on the list), $\mathbf{f}_i$, was projected through the scaled semantic similarity matrix (the diagonal, representing item self-strengths, was set to 0). The item activations corresponding to contextual cuing and item cuing were added to obtain the total item activation:

$$\mathbf{a} = s\mathbf{M}^{sem}\mathbf{f}_i' + \mathbf{M}^{CF}\mathbf{c}' \qquad \text{(IV.17)}$$

The activation values **a** were then used with Equation IV.15 to determine recall probabilities.

We also examined a model that combined context- and item-based semantic cuing. This was the same as the item-based semantic cuing model, but rather than cuing semantics using just the item vector, we used a weighted combination of context and item:

$$\mathbf{a} = s\mathbf{M}^{sem}(w\mathbf{f}_i + (1-w)\mathbf{c})' + \mathbf{M}^{CF}\mathbf{c}', \qquad \text{(IV.18)}$$

where w is a parameter controlling the relative weighting of the item cue compared to the context cue. Note that this model is equivalent to the item-based cuing model when $w = 1$, and the context-based cuing model when $w = 0$.

Results

Recency and contiguity

Table 6 shows best-fitting parameters and AIC values for each model variant. Although the base model had the lowest AIC of the five model variants, it provided a qualitative fit to a number of standard summary statistics, including the recency, primacy, and contiguity effects. The generative Base model provided a good fit of recall as a function of serial position (Fig. 28), though primacy was under-predicted. This was an issue with each model variant. Given that retrieved-context models have successfully accounted for the magnitude of primacy in prior work (e.g. (Polyn et al., 2009)), it appears that this under-prediction of primacy is caused by our different emphasis on fitting entire recall sequences, rather than traditional summary statistics such as the serial position curve, as in prior work with retrieved-context models. The model also provides a qualitative account of the probability of initiating recall at each serial position (Fig. 28). The model accounts for the temporal contiguity effect, including the tendency to make forward transitions more often than backward transitions (Fig. 28c). The model slightly over-predicts nearby transitions, and under-

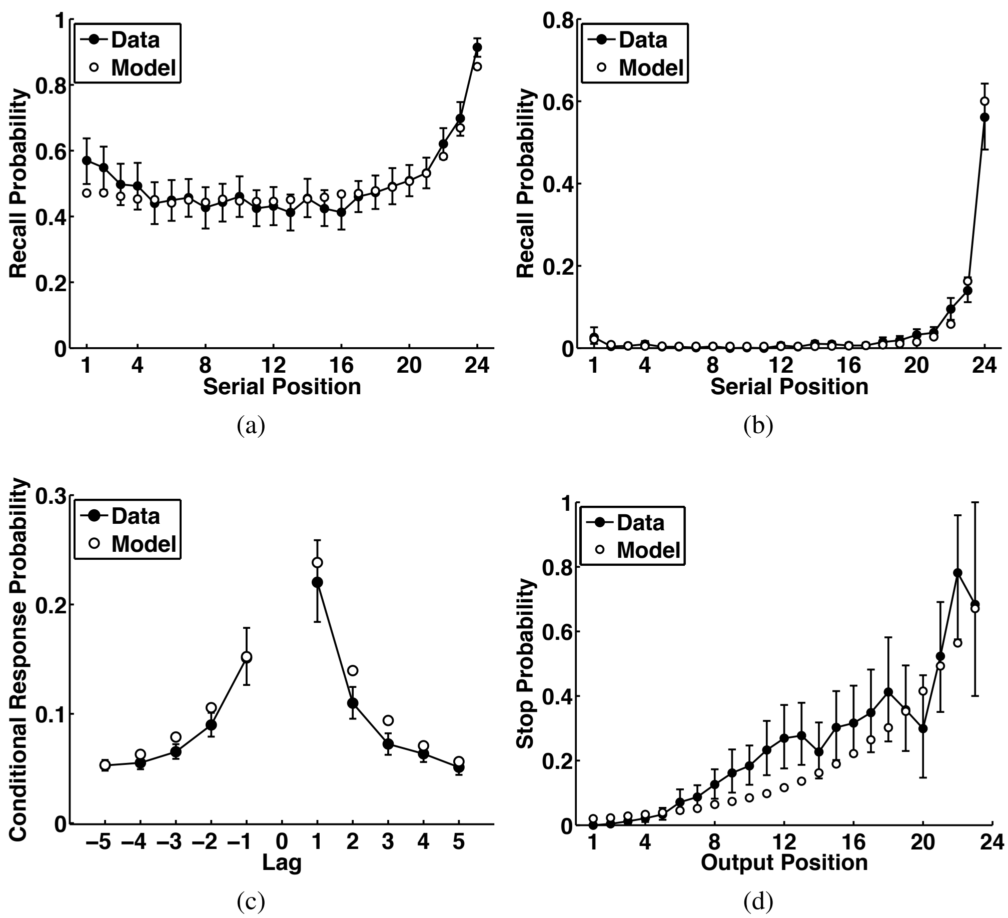


Figure 28. (**a**) Recall probability as a function of serial position, for the data and the best-fitting model with no semantic associations. (**b**) Probability of starting recall with each serial position. (**c**) Conditional response probability as a function of lag. (**d**) Stop probability by output position.

	Base	LSA Context	LSA Item	WAS Context	WAS Item
β_{enc}	0.73	0.73	0.72	0.73	0.73
β_{rec}	0.82	0.83	0.81	0.84	0.81
β_{start}	0.22	0.23	0.25	0.24	0.26
α	0.89	0.77	8.57	0.64	5.60
δ	1.66	1.73	9.19	1.83	6.28
γ	0.36	0.34	0.41	0.31	0.39
ϕ_s	1.74	1.86	1.37	1.97	1.43
ϕ_d	1.09	1.11	1.13	1.13	1.14
s	–	0.19	0.45	0.77	0.90
τ	8.81	7.96	76.48	7.31	50.30
θ_s	0.02	0.02	0.02	0.02	0.02
θ_r	0.16	0.16	0.16	0.16	0.16
AIC	60343	60306	60196	60152	60051
wAIC	3.02e-64	3.776e-56	3.615e-32	1.236e-22	1

Table 6. Best-fitting parameters for the model variants.

predicts distant transitions (Fig. 28c). Finally, the model accounts for the finding of a positively accelerated increase in stop probability with output position (Fig. 28d), though the imperfect fit suggests that the actual pattern in the data may not be perfectly accounted for by an exponential function.

Semantic similarity

Given that our Base model with no semantic associations was able to account for benchmark findings in free recall, we examined whether the fit could be improved by the addition of semantics. The addition of either LSA- or WAS-based associative structure to context-to-item associations led to a substantially better fit, as measured by AIC (Table 6). However, for both semantic models, AIC was lower (improved) when an item-based, rather than context-based, cuing mechanism was used. We also examined models that allowed semantic cuing to involve a weighted combination of context and item information. For both LSA and WAS, the best-fitting value of the w parameter was 1, indicating that the addition of context to the semantic cue provided no benefit to the fit.

Regardless of the cuing mechanism used, WAS provided a better fit to behavior than LSA.

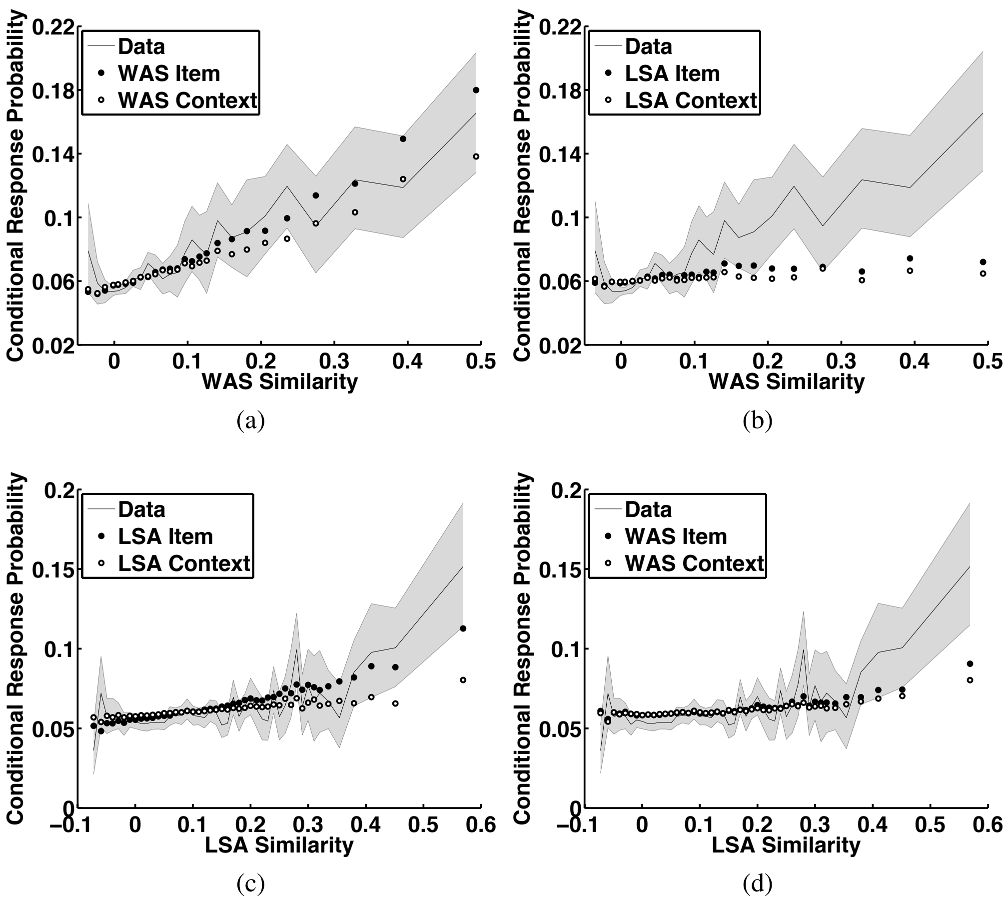


Figure 29. (**a**) Conditional response probability as a function of WAS semantic similarity bin. The line indicates the mean value in the data, and the shaded region represents the 95% confidence interval. Also shown is the performance of the WAS-based models. Black dots indicate the performance of the item-based cuing model, while white dots indicate the performance of the context-based cuing model. (**b**) Same as (a), but with performance of the LSA-based models. (**c**) Conditional response probability as a function of LSA semantic similarity bin, with performance of the LSA-based models. (**d**) Same as (c), but with performance of the WAS-based models.

According to wAIC, the item-based-cuing model with WAS was by far the most probable model of the models considered, with a weighted AIC that was virtually 1. Critically, this analysis uses a likelihood-based measure of fitness that makes no assumptions about the actual structure of our participants' semantic knowledge, and therefore avoids complications that arise when a semantic model is used to both generate and evaluate model predictions (Polyn et al., 2009; Manning et al., 2012).

In order to better understand the influence of cuing mechanism and the relation between the different semantic models, we calculated the LSA-CRP and WAS-CRP for every model (Fig. 29). As expected, both the LSA-CRP and the WAS-CRP were flat for the Base model[5]. For each semantic model-analysis combination, we found that context-based cuing lead to under-prediction of the slope of the semantic-CRP, suggesting that item-based cuing better describes the data regardless of the semantic model used. Both WAS- and LSA-based models were limited in their ability to explain the semantic-CRP based on the other semantic model (Fig. 29b,d). This suggests that both semantic models explain unique variance in recall sequences, although the wAIC scores suggest that WAS is overall more successful in predicting recall sequences.

Although all parameters were allowed to vary freely for each of the five model variants, many of the best-fitting parameters were quite similar across all models (Table 6). Parameters controlling the rate of context evolution (β_{enc}, β_{rec}), parameters involved in the primacy effect (ϕ_s, ϕ_d, and β_{start}), and stopping parameters (θ_s and θ_r) were all comparable across the five model variants. The semantic scaling parameter, s, was greater for better-fitting models, suggesting that the influence of semantics is scaled up as the model of semantic cuing is improved. Other parameters also varied substantially depending on the semantic cuing mechanism used. α, δ, and τ all increased for item-based semantic cuing models. Increasing α causes recall to be more stochastic (less dependent on the particular context cue used), and increasing τ causes recall to become more deterministic; therefore, there is a degree of parameter tradeoff, which may explain some of the variability in

[5] We found that the semantic-CRPs for the Base model showed an increased probability of making very low- or high-similarity transitions when the semantic-CRPs were calculated as described by (Howard and Kahana, 2002b). This lead us to implement a version of the analysis more similar to that proposed by Sederberg et al. (2010), which did not demonstrate this distortion.

best-fitting parameters between different variants. Similarly, inspection of the best-fitting values across models suggests that the optimal δ parameter may also depend on the value of τ. Another influence on the value of α may be the tendency of context-based semantic cuing models to predict more diffuse cuing of multiple items in the list (see Fig. 27d for an illustration), somewhat similar to the effect of increasing α. To replicate this effect, item-based semantic cuing models may rely on a larger value of α.

Discussion

We developed a likelihood-based modeling framework that allowed us to test competing models of semantic organization in free recall while controlling for many of the complexities of memory search. Using this framework, we found evidence that the influence of semantic associations on free recall is more consistent with item-based, rather than context-based, cuing. Furthermore, we found that a word-association spaces (WAS)-based measure of semantic similarity was better able to predict the order of recalls than a standard latent semantic analysis (LSA) measure of similarity. Below we discuss how our findings bear on evaluating models of semantic memory and semantic cuing, and discuss how our modeling framework may aid measurement of semantic organization.

Models of semantic association strength

Both WAS and LSA have been used to characterize behavior in free recall (Howard and Kahana, 2002b; Howard et al., 2007; Manning and Kahana, 2012) and have been used as components of models of memory search (Sirotin et al., 2005; Polyn et al., 2009). The present results suggest that WAS is better able than LSA to predict behavior in recall of lists of words with no systematic semantic structure. Our results complement those of Sirotin et al. (2005), who compared the ability of WAS and LSA to explain behavior in free recall of categorized materials. They developed a version of the search of associative memory (SAM) model (Raaijmakers and Shiffrin, 1980) that included semantic associations between items. Sirotin et al. (2005) assumed that search of long-term memory is driven by both episodic and semantic inter-item associations. They contrasted WAS and LSA as different models of semantic association strength to incorporate into their model.

They found that the WAS-based model was better able to account for category clustering in a multi-trial free recall study (Kahana and Wingfield, 2000), because WAS similarity values better reflected the category membership of the stimuli. That is, the discriminability (d') between the similarity distributions of pairs of items in the same category compared to pairs of items in different categories was greater for WAS than for LSA. Note that their analysis of recall behavior focused on only one aspect of semantic organization, namely clustering by taxonomic category. In contrast, our likelihood-based framework for evaluating models does not rely on summary statistics such as measures of category clustering. This means that our framework can be applied to studies where the studied items have no systematic semantic structure, without having to assume a "true" model of semantic associations.

While we focused on two vector-space models, WAS and LSA, our framework can be used to evaluate any model of semantic memory that provides estimates of association strengths between items. These associations do not need to be symmetric; the association strength from item i to item j does not have to be the same as the association from item j to item i.

Mechanisms of semantic cuing

Polyn et al. (2009) extended the temporal context model (TCM) to account for multiple influences on recall organization, including source context and semantic similarity. Their model, the context maintenance and retrieval (CMR) model, proposed that recall of an item causes retrieval of the temporal context that was active when that item was originally studied. This retrieved context is assumed to contain a weighted average of information related to the items presented before the recalled item. The context then projects through semantic associations, providing a good cue for items that are semantically related to any of the items represented in the context cue.

We contrasted this context-based cuing mechanism with a simpler item-based cuing mechanism where the last item recalled provides support for items that semantically related to it. We found evidence that an item-based semantic cuing mechanism provides a better account of our free recall data than context-based semantic cuing. We also tested a model variant that used a combination of context and item information as a semantic cue; however, we found that the best-fitting model used only item information. This suggests that adding context-based semantic cuing did not improve the model's ability to fit behavior.

Although our results are inconsistent with the semantic cuing mechanism proposed by Polyn et al. (2009), this does not necessarily suggest that context is uninvolved in semantic organization. Using scalp EEG during encoding of categorized materials, Morton et al. (2013) found evidence of persistent category-specific activity, which accumulated over multiple stimulus presentations. The rate of this accumulation predicted individual differences in organization by stimulus category during recall. They noted that this accumulative activity is consistent with the operation of temporal context, if context is sensitive to the properties of studied materials. The simulation studies presented in Chapter II confirmed that the results of Morton et al. (2013) are consistent with a version of CMR where semantic information is integrated into context during study. Future work should adapt the distributed-CMR model presented in Chapter II to the likelihood-based framework presented here. This will allow us to contrast the mechanisms presented here, which operate only during retrieval, with the distributed-CMR model, where semantic information is integrated into context during encoding. It will also be important to investigate whether different semantic cuing mechanisms are involved when there is strong semantic structure to the studied materials (as in the Morton et al. 2013 study) compared to when there is no systematic semantic structure (as in the study examined in this chapter).

False memory paradigms will provide important constraint on joint models of semantic associations and cuing, since fitting the data requires a combination of making the critical false recalls observed in the data and not making false recalls to many other words (Kimball et al., 2007). Through simulations of false-memory paradigms using a variant of SAM, Kimball et al. (2007) found evidence that both persistent semantically related activity during encoding and compound cuing during recall may be necessary to fully account for data in false memory experiments. This suggests that a model with a semantic-sensitive context representation, such as that proposed in Chapter II, may be well-suited for accounting for findings in the false memory paradigm.

Measurement of semantic organization

The model-based framework used here may be useful for measuring semantic organization while controlling for other influences on recall behavior. Controlling for temporal organization is critical when considering a manipulation like blocked vs. random presentation of categorized stimuli (Puff, 1966), but most research on blocked-random effects has not accounted for this influence (e.g. Cofer et al. 1966; D'Agostino 1969; Borges and Mandler 1972). Through bootstrapping techniques, it is possible to estimate the amount of semantic organization due to temporal clustering (Polyn et al., 2009; Morton et al., 2013). However, this technique requires collecting data from baseline lists with no category structure, and involves the assumption that most aspects of recall behavior will be the same when comparing baseline and categorized lists. To avoid these issues, CMR could be fit separately to blocked and random lists; the semantic scaling parameter would then provide an estimate of the strength of semantic organization, separate from temporal organization and other influences that might vary between conditions. The ability to measure the strength of semantic influence during retrieval independent of temporal organization will provide an important tool to better understand how presentation order affects semantic organization during recall.

Conclusions

While research has found that semantic knowledge exerts an important influence on search of episodic memory, many questions remain about the cognitive mechanisms that mediate this influence. We developed a modeling framework that allows selection between different models of semantic associations, as well as comparison of different mechanisms through which these associations may influence memory search. We hope that the computational modeling framework presented here will help shed light on how prior semantic knowledge shapes new memory formation and expression.

CHAPTER V

Conclusions

Through a combination of behavioral measures, scalp EEG recordings, and computational modeling, we investigated the interplay of temporal and semantic information in driving search of episodic memory. We focused on the framework of retrieved-context models of episodic memory, which have successfully accounted for many findings in free recall. Retrieved-context models were originally developed to explain findings from studies using random word lists, which often ignored the impact of semantic associations. We made progress in understanding the influence of semantic associations on episodic memory by modifying and extending the retrieved-context framework to quantitatively test competing mechanisms by which semantics might influence encoding and retrieval. Furthermore, through scalp EEG recordings, we found a potential neural correlate of the representation of semantic knowledge of studied materials; the properties of this neural signal will provide constraint to neurocognitive models of memory search going forward.

Prior versions of retrieved-context models such as the context maintenance and retrieval model (CMR) were challenged by a recent finding that semantic associations influence the neural representations active during encoding, resulting in changes in later organization by taxonomic category (Morton et al., 2013). The original version of CMR assumed that semantic knowledge does not affect representations active during encoding, and instead only influences retrieval. However, we demonstrated that the larger theoretical framework of retrieved-context models actually predicts that semantic information will be integrated into context during encoding, resulting in activity similar to that observed by Morton et al. (2013). According to the retrieved-context framework, gradual learning will cause related items to become associated with similar states of context (Rao and Howard, 2008). We developed a new version of CMR that takes prior experience into account, and showed that it can account for the finding of Morton et al. (2013) that neural oscillatory activity during encoding integrates information about stimulus category over time. Furthermore, by fitting customized models to the recall behavior of individual participants, we were able to predict individual differences in integrative neural activity, thus validating the relationship between encoding representations and recall behavior predicted by the model. We also demonstrated that the model's context integration mechanism can explain the finding from the classic categorized free recall literature that recall performance and category organization are increased when items from the same category are presented adjacent to one another.

Using scalp EEG, we sought to better understand the properties of the integrative oscillatory activity observed by Morton et al. (2013). We localized integrative category-specific activity to fluctuations in power in the beta band, at right posterior electrodes, and found that this integrative activity was disrupted by inter-item distraction. These results provide a potential oscillatory substrate of the type of cue-construction mechanism assumed by major models of episodic memory (Raaijmakers and Shiffrin, 1980; Howard and Kahana, 2002b). At present, it is unclear whether this integrative activity is more consistent with the operation of a gradually changing context representation (as in CMR; see Chapter II) or a multi-item working memory buffer (Axmacher et al., 2010). Further modeling work may be helpful to understand whether a buffer-based model like SAM (Raaijmakers and Shiffrin, 1980) can account for the findings of Morton et al. (2013) and Puff (1966). Research on working memory suggests that multi-item buffers may be implemented in the brain through phase-amplitude coupling between theta and beta/gamma activity (Axmacher et al., 2010); therefore, it will also be important to determine whether category-specific beta-band activity is coupled to the phase of theta activity (cf. Fuentemilla et al. 2010). Another important goal for future research will be to test whether category-specific beta-band activity during encoding is predictive of subsequent recall behavior.

Interestingly, we found that inter-item distraction has a different effect on temporal and semantic organization. Consistent with prior results, distraction decreases semantic organization (Howard and Kahana, 2002b) but does not affect temporal organization (Howard and Kahana, 1999). Our results contribute to knowledge of this important dissociation by demonstrating a potential neural correlate of the disruption of semantic information in the presence of distraction.

This disruption of integrative neural activity may be related to the subsequent decrease in semantic organization. Future work should examine whether a combination of the retrieved-context model presented by Sederberg et al. (2008), which addressed the impact of distraction on non-semantic aspects of free recall, and the model presented in Chapter II, is able to account for this dissociation between temporal and semantic organization.

While the model presented in Chapter II provided insight into the encoding processes that contribute to recall organization by category, questions remain about the structure of the semantic associations that influence memory search, as well as the mechanisms that allow semantic associations to constrain search. We developed a novel modeling framework for testing both different models of semantic similarity, and different cuing mechanisms through which semantic similarity influences search. This framework allows evaluation of models without requiring specification of an arbitrary set of summary statistics (which themselves may require assumption of a model of semantic associations). Instead, models are evaluated based on their ability to predict the sequence of individual items that are recalled by each participant. Using this framework, we tested two models of semantic associations, WAS (Steyvers et al., 2004) and LSA (Landauer and Dumais, 1997), which have been used in a number of studies to measure semantic organization in free recall (Howard and Kahana, 2002b; Sirotin et al., 2005; Polyn et al., 2009). We found that WAS was better able to predict individual recall sequences; this suggests that WAS is better suited for informing experimental design of free-recall studies and for measuring recall behavior. Furthermore, we propose that our likelihood-based framework (paired with a model of semantic associations such as WAS) can allow measurement of semantic organization while controlling for other influences on recall behavior such as temporal organization.

We also contrasted two distinct mechanisms by which semantic associations might influence recall. Semantic organization has been proposed to be caused by direct inter-item associations (Raaijmakers and Shiffrin, 1980) or associations between distributed states of temporal context and items (Polyn et al., 2009). These different mechanisms make different predictions about the impact of semantic associations on memory search; item-based cuing predicts that recall will be more focused on associates of individual recalled items, while context-based cuing predicts that cuing will depend on the context surrounding the recalled item during encoding. We found that data in a free-recall study with lists of random words was more consistent with item-based cuing. This is somewhat surprising, since researchers have suggested that recall is driven by a broader cue related to the context in which an item was studied (Kimball et al., 2007). However, while our results suggest that the cuing mechanisms employed by the model of Polyn et al. (2009) may be flawed, it is possible that other context-based cuing mechanisms will provide a better account of the data. In particular, future work should examine an encoding-based model of semantic influence, as in the model presented in Chapter II, by incorporating that model into the likelihood-based framework presented in Chapter IV. It is possible that participants rely on different cuing strategies depending on the content of studied material. When learning lists of materials with no systematic semantic organization, participants may rely more on item-based cuing; in contrast, when learning lists of categorized materials, participants may rely more on context-based cuing. Applying the likelihood-based modeling framework developed here to categorized and uncategorized free recall may help clarify this issue.

While a number of questions remain about how semantic knowledge impacts episodic encoding and retrieval, we have made progress not only in understanding the behavioral consequences and neural correlates of prior knowledge, but also in understanding the links between neural and behavioral measures and the cognitive mechanisms that allow us to efficiently recall knowledge of past events.

REFERENCES

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

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive Psychology*, 6(4):451–474.

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

Axmacher, N., Henseler, M., Jensen, O., Weinreich, I., Elger, C., and Fell, J. (2010). Cross-frequency coupling supports multi-item working memory in the human hippocampus. *Proceedings of the National Academy of Sciences, USA*.

Baayen, R., Piepenbrock, R., and Gulikers, L. (1995). CELEX2 LDC96L14. Web download, Linguistic Data Consortium, Philadelpha, PA.

Becker, S. and 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.

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

Borges, M. A. and Mandler, G. (1972). Effect of within-category spacing on free recall. *Journal of Experimental Psychology*, 92:207–214.

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 Spence, K. W. and Spence, J. T., editors, *The Psychology of Learning and Motivation : Advances in Research and Theory*, volume 1, pages 229–325. Academic Press, New York.

Bower, G. H. (1972). Stimulus-sampling theory of encoding variability. In Melton, A. W. and Martin, E., editors, *Coding Processes in Human Memory*, chapter 5, pages 85–121. John Wiley and Sons, New York.

Brown, G. D. A., Neath, I., and Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, 114(3):539–576.

Cofer, C. N., Bruce, D. R., and 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., and 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. and Anderson, J. R. (2010). The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. *Psychological Bulletin*, 136(1):87–102.

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

Davis, T., Love, B. C., and 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.

Delorme, A. and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods*, 134:9–21.

Duda, R. O., Hart, P. E., and Stork, D. G. (2001). *Pattern classification, second edition*. Wiley, New York.

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

Epstein, R. A. and Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392:598–601.

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

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

Farrell, S. (2012). Temporal clustering and sequencing in working memory and episodic memory. *Psychological Review*, 119(2):223–271.

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

Fuentemilla, L., Penny, W. D., Cashdollar, N., Bunzeck, N., and Düzel, E. (2010). Theta-coupled periodic replay in working memory. *Current Biology*, 20(7):606–612.

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

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

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

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

Healey, M. K. and Kahana, M. J. (2014). Is memory search governed by universal principles or idiosyncratic strategies? *Journal of Experimental Psychology—General*, 143(2):575–596.

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

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

Howard, M. W. and Kahana, M. J. (2002b). When does semantic similarity help episodic retrieval? *Journal of Memory and Language*, 46:85–98.

Howard, M. W., Kahana, M. J., and 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., and Wingfield, A. (2006a). 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., and 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., Venkatadass, V., Norman, K. A., and Kahana, M. J. (2007). Associative processes in immediate recency. *Memory & Cognition*, 35(7):1700–1711.

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

Howard, M. W., Wingfield, A., and Kahana, M. J. (2006b). Modeling the age-related associative deficit in self-initiated recall. *Psychonomic Bulletin & Review*, 13(3):439–445.

Huth, A. G., Nishimoto, S., Vu, A. T., and 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.

Jacobs, J. and Kahana, M. J. (2009). Neural representations of individual stimuli in humans revealed by gamma-band ECoG activity. *Journal of Neuroscience*, 29(33):10203–10214.

Jafarpour, A., Horner, A. J., Fuentemilla, L., Penny, W. D., and Düzel, E. (2013). Decoding oscillatory representations and mechanisms in memory. *Neuropsychologia*, 51(4):772–780.

Jang, Y. and 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. and 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. and Mewhort, D. J. K. (2007). Representing word meaning and order information in a composite holographic lexicon. *Psychological Review*, 114(1):1–37.

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

Jung, T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., and Sejnowski, T. J. (2001). Analysis and visualization of single-trial event-related potentials. *Human Brain Mapping*, 14(3):166–185.

Junghöfer, M., Elbert, T., Tucker, D. M., and Rockstroh, B. (2000). Statistical control of artifacts in dense array EEG/MEG studies. *Psychophysiology*, 37(4):523–532.

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

Kahana, M. J. (2012). *Foundations of Human Memory*. Oxford University Press, New York, NY, 1st edition.

Kahana, M. J. and Wingfield, A. (2000). A functional relation between learning and organization in free recall. *Psychonomic Bulletin & Review*, 7:516–521.

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

Kanwisher, N., McDermott, J., and Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *J. Neurosci.*, 17:4302.

Kimball, D. R., Bjork, E. L., Bjork, R. A., and 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., and Kahana, M. J. (2007). The fSAM model of false recall. *Psychological Review*, 114(4):954–93.

Kreiman, G., Koch, C., and 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., and Bandettini, P. A. (2008a). 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., Tanaka, K., and Bandettini, P. A. (2008b). Matching categorical object representations in inferior temporal cortex of man and monkey. *Neuron*, 60:1126–1141.

Kuhl, B. A., Dudukovic, N. M., Kahn, I., and Wagner, A. D. (2007). Decreased demands on cognitive control reveal the neural processing benefits of forgetting. *Nat Neurosci*, 10(7):908–914.

Kuhl, B. A., Rissman, J., Chun, M. M., and 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., and Wagner, A. D. (2012). Multi-voxel patterns of visual category representation during episodic encoding are predictive of subsequent memory. *Neuropsychologia*, 50:458–469.

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

Landauer, T. K. and 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. and Postle, B. R. (2008). Temporary activation of long-term memory supports working memory. *Journal of Neuroscience*, 28(35):8765–8771.

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

Lohnas, L. J. and Kahana, M. J. (2014). Compound cuing in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(1):12–24.

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

Manning, J. R. and Kahana, M. J. (2012). Interpreting semantic clustering effects in free recall. *Memory*, 20(5):511–517.

Manning, J. R., Polyn, S. M., Baltuch, G., Litt, B., and 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., and 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., and Eichenbaum, H. (2007). Gradual changes in hippocampal activity support remembering the order of events. *Neuron*, 56(3):530–540.

McClelland, J. L., McNaughton, B. L., and 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.

McMenamin, B. W., Shackman, A. J., Maxwell, J. S., Bachhuber, D. R. W., Koppenhaver, A. M., Greischar, L. L., and Davidson, R. J. (2010). Validation of ICA-based myogenic artifact correction for scalp and source-localized EEG. *NeuroImage*, 49(3):2416–2432.

Mitchell, T. M., Shinkareva, S. V., Carlson, A., Chang, K. M., Malave, V. L., Mason, R. A., and 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., Sperling, M. R., and 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.

Myung, I. J. (2003). Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47(1):90–100.

Nelson, D. L., McEvoy, C. L., and 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.

Nolan, H., Whelan, R., and Reilly, R. B. (2010). FASTER: Fully Automated Statistical Thresholding for EEG artifact Rejection. *Journal of Neuroscience Methods*, 192(1):152–162.

Norman, K. A., Newman, E., Detre, G., and 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.

Nunez, P. and Srinivasan, R. (2006). *Electric fields of the brain: The neurophysics of EEG*. Oxford University Press, New York.

Onton, J., Westerfield, M., Townsend, J., and Makeig, S. (2006). Imaging human EEG dynamics using independent component analysis. *Neuroscience and Biobehavioral Reviews*, 30(6):808–822.

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

O'Toole, A. J., Jiang, F., Abdi, H., and 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., and 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. and 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., and 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., and Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science*, 310:1963–1966.

Polyn, S. M., Norman, K. A., and 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. and Sederberg, P. B. (2014). Brain rhythms in mental time travel. *NeuroImage*, 85(2):678–684.

Price, K., Storn, R. M., and Lampinen, J. A. (2005). *Differential Evolution: A Practical Approach to Global Optimization*. Natural Computing Series. 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., and 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., and 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., and Fried, I. (2005). Invariant visual representation by single neurons in the human brain. *Nature*, 435(23):1102–1107.

Raaijmakers, J. G. W. and Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In Bower, G. H., editor, *The psychology of learning and motivation: Advances in research and theory*, volume 14, pages 207–262. Academic Press, New York.

Rao, V. A. and Howard, M. W. (2008). Retrieved context and the discovery of semantic structure. In Platt, J. C., Koller, D., Singer, Y., and Roweis, S., editors, *Advances in Neural Information Processing Systems*, pages 1193–1200. MIT Press, Cambridge, MA.

Roediger, H. L. and 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., and 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., and Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, 76(1):45–48.

Romney, A. K., Brewer, D. D., and Batchelder, W. H. (1993). Predicting clustering from semantic structure. *Psychological Science*, 4:28–34.

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

Sahakyan, L. and 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.

Schwartz, R. M. and Humphreys, M. S. (1973). Similarity judgements and free recall of unrelated words. *Journal of Experimental Psychology*, 101:10–13.

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

Sederberg, P. B., Miller, J. F., Howard, W. H., and Kahana, M. J. (2010). The temporal contiguity effect predicts episodic memory performance. *Memory & Cognition*, 38(6):689–699.

Shankar, K. H. and Howard, M. W. (2010). 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.

Sirotin, Y. B., Kimball, D. R., and Kahana, M. J. (2005). Going beyond a single list: Modeling the effects of prior experience on episodic free recall. *Psychonomic Bulletin & Review*, 12(5):787–805.

Smith, 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 Davies, G. M. and Thomson, D. M., editors, *Memory in context: Context in memory.*, pages 13–34. John Wiley & Sons, Oxford, England.

Socher, R., Gershman, S. J., Perotte, A. J., Sederberg, P. B., Blei, D. M., and Norman, K. A. (2009). A bayesian analysis of dynamics in free recall. In Bengio, Y., Schuurmans, D., Lafferty, J., Williams, C. K. I., and Culotta, A., editors, *Advances in Neural Information Processing Systems*. MIT Press.

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

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

Stricker, J. L., Brown, G. G., Wixted, J. T., Baldo, J. V., and 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., and 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 Tulving, E. and Donaldson, W., editors, *Organization of Memory.*, pages 381–403. Adademic Press, New York.

Tulving, E. (1983). *Elements of Episodic Memory*. Oxford, New York.

Tulving, E. and 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., and 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. and McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108(3):550–592.

Wagenmakers, E.-J. and Farrell, S. (2004). AIC model selection using akaike weights. *Psychonomic Bulletin & Review*, 11(1):192–196.

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., and Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods*, 42(3):671–684.

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

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