December 2020

A Model of Scaffolded Encoding

Yanxiu Chen
Syracuse University

Abstract

Repetitions are ubiquitous and are the foundation of episodic memory. As we investigate episodic memory in a laboratory setting and with, usually, a study-test design, we have inaccurately assumed encoding happens only at study and retrieval happens only at test. However, it's entirely possible that encoding and retrieval happen simultaneously at both study, with explicit repetition, and test, with implicit repetition. The fallacy that encoding and retrieval are separate processes coincides with the unsatisfactory status of memory process model: there is no unified memory model. Most models of memory are models of retrieval tasks. Facing these two challenges, we proposed a model of scaffolded encoding. Building on previous modeling studies that include an updating mechanism (accumulating information in an old memory trace) for an old event and an adding mechanism (storing a new trace) for a new event, we presented a scaffolded encoding mechanism for a semi-old event. Scaffolding a new trace means a new trace is not encoded and stored from nothing. Instead, it is added to memory with already updated information from old memories. As a result, an exact repetition (an old event) leads to one single strong memory trace and a partial repetition (a semi-old event) is stored as a separate episode (new) with strengthened information (old). In this current project, we will present a thorough investigation implementing the scaffolded encoding mechanism within the retrieving effectively from memory model. Within this model, any study or test event can be identified as an old, new, or semi-old event, and then corresponding encoding process (updating, adding, or scaffolded encoding) takes place. Our model, unprecedentedly, successfully accounts for the hallmark findings that show effects of repetition, including list strength effect, output interference, proactive interference and proactive facilitation, from multiple retrieval tasks.

A MODEL OF SCAFFOLDED ENCODING

by

Yanxiu Chen

B.S., Guangzhou University of Chinese Medicine, 2015
M.S., Syracuse University, 2017

Dissertation
Submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Cognitive Psychology.

Syracuse University
June 2020

Acknowledgement

Part of the journey is the end. I have been looking forward to becoming an independent scientist. Now that I have to fly out this comfortable nest, I'm ready for the challenges but I'm not excited about putting an end to this journey. I'm used to asking Amy for advice casually. I'm used to being brutally challenged in almost every presentation. I'm used to celebrating every step of growth. I'm sadly aware these will be scarcity outside our lab and this department.

I want to express my deepest gratitude to my advisor, Amy Criss. As a young female scientist, it's the most fortunate coincidence having an intelligent and powerful woman to be my advisor. You are my role model. I also would like to thank the entire CBB program. I would not be in this place without the faculty members asking the hardest questions and sharing their most honest opinions. I would not have grown as much without the camaraderie from my fellow graduate students.

And for my parents. I thank my father for not understanding my passion but supporting me regardless. I thank my mother for always being my biggest fan and still believing I'm the smartest person in the galaxy. I'm proud of making you proud of me. I thank you both for raising me to be, what they call, sassy.

At last, I want to thank my friends. The weekly Murder Mystery Dinner party strangely keeps me sane. Dining and traveling with you guys are already an important part of my life. For my dearest friends on the other side of the earth, Xiao Meng and Zhi Lin, my life is enlightened with our daily random chats, giggling over a drama or novel, and the countless remote, free therapy sections. I love you 3000.

Table of Contents

Introduction

Having spent many long winters in Syracuse, I remember the episodic associations between this city and snow vividly. I remember the first time I experienced snow at Syracuse. I remember school was closed because snow was too heavy… Episodic memory is our life experiences. In the field of cognitive psychology, there are different ways to study episodic memory, one of which is developing and testing memory models. They describe and help us understand memory's underlying mechanisms. However, the current status of computational memory models is unsatisfactory, which we will unfold from two perspectives, the fallacy that encoding and retrieval are separate (Hintzman, 2011) and the lack of memory model for multiple paradigms and multiple retrieval tasks (Criss & Howard, 2015). They are the two primary motivations for our proposed and extensively examined model.

Concurrent Encoding and Retrieval

In a well-controlled laboratory situation, memory researchers typically ask participants to study a list of stimuli, and later test their memories, which we label study and test stages. Likewise, we typically implement computational models describing separate encoding and retrieval processes. For instance, in the simplest version of retrieving effectively from memory model (REM, Shiffrin & Steyvers, 1997), encoding occurs at the study stage and retrieval occurs at the test stage. At study, as a stimulus is presented, its features are encoded and stored imperfectly in the form of a memory trace. In a simple recognition memory test, as a cue is presented, a recognition decision (i.e., old or new) is made depending on whether the match between the cue and stored episodic memory traces exceeds a familiarity criterion. In a recall task, a trace is sampled in proportion to its match to the cue and then possibly recovered.

However, encoding and retrieval happen simultaneously rather than separately (Bjork, 1975; Hintzman, 2011). For an example, during each year's first snow (i.e., repeated presentation and encoding of the Syracuse-snow stimulus), I remember the heaviest snowy day at Syracuse in my memory (i.e., retrieval) and wonder when school is cancelled this year. Every time I fly back to Syracuse from my winter break, I am instantly reminded that snow is waiting for me (i.e., cued recall in real life), and my episodic association between Syracuse and snow is even stronger (i.e., encoding). Encoding and retrieval processes are not clear-cut but intertwined.

Concurrent encoding and retrieval imply that both processes happen at both study and test stages and there are examples of theories and models. Incidental study-phase retrieval (i.e., recursive reminding, Fisher & Nelson, 2006; Hintzman, 2004, 2010; Hintzman & Block, 1973; Jacoby & Wahlheim, 2013; Wahlheim & Jacoby, 2013) refers to the process that, at study, repeated presentation of a stimulus (e.g., Syracuse-snow) triggers memory retrieval and participants are reminded that they have encountered it before. Retrieval-phase encoding refers to the process that, at test, outputting a response comes with encoding and storing its information in memory. Within the search of associative memory (SAM; Raaijmakers & Shiffrin, 1981a, 1981b) model where memories are represented as associations among experimental context and items presented together, output of a target item under a presented item cue leads to the strengthening of their associations (i.e., incrementing). Within REM (Shiffrin & Steyvers, 1997) where memories are represented as memory traces, concatenated vectors of experimental context and item(s), an example of concurrent encoding and retrieval is ***Update+Add*** (Criss, Malmberg, and Shiffrin, 2011). In a recognition task, if a presented test item is recognized as old, the best matching memory trace is updated with the test item's information. If a presented test item is judged as unstudied, a new trace corresponding to the test item is added. That is, an updating mechanism is for an old event and an adding mechanism is for a new event.

Repetition is the main theme behind concurrent encoding and retrieval and undoubtfully is the most powerful factor affecting memory (Hintzman, 1970; Hintzman, 1976; Hintzman & Block, 1971). On one hand, encoding during retrieval happens at test and is an implicit repetition from generating an output. It takes the form of presenting a cue (recognition and recall) and recovering the remaining content of a memory (recall). Recognizing the presented cue as unstudied is a new event whereas recognizing the presented cue as studied and successfully recovering a memory trace is an old event. Correspondingly, as mentioned, a new trace is added and an old trace is updated (i.e., ***Update+Add*** at test). It has been used to account for changing memory performance across test trials in recognition (Criss et al., 2011; Kilic, Criss, Malmberg, & Shiffrin, 2017) and in cued recall (Wilson, Kellen, & Criss, 2020). On the other hand, retrieval during encoding happens at study and is an explicit repetition provided by the external environment. It could be an exact repetition, an old event where the same stimulus is presented more than once (e.g., restudying A-B pair) or a partial repetition, a semi-old event where an old stimulus is presented with a new partner (e.g., studying A-B and A-D pairs, Wahlheim & Jacoby, 2013; Jacoby, Wahlheim, & Yonelinas, 2013). However, adding for a new event and updating for an old event (i.e., ***Update+Add*** at study) doesn't consider an encoding mechanism specified for a partial repetition, or a semi-old event. Then, ***Update+Add*** can't account for data showing partial repetition benefits and harms memory simultaneously (proactive facilitation and proactive interference, PF and PI) in cued recall (Aue, Criss, & Fischetti, 2012; Aue, Criss, & Novak, 2017).

Therefore, the Criss lab proposed an encoding mechanism, scaffolded encoding (Criss, Cox, Chen, Wilson, & Aue, 2018). As a semi-old event has new component(s) yet is similar to a prior memory, the prior memory trace is updated with features of this current event and another trace representing this semi-old event is scaffolded, borrowing the updated features from the prior trace. Here, 'scaffold' means a new trace is not encoded and stored from nothing, but from already strengthened memories with updated information. Combining with the mentioned updating mechanism for an old event and adding mechanism for a new event, we have a model of scaffolded encoding. This model enforces simultaneous operation of encoding and retrieval at both study and test, by having every event evaluated first (i.e., new, old, or semi-old) then activating corresponding encoding mechanism (i.e., adding, updating, or scaffolded encoding).

One may think of our model as a combination of two classes of memory models, cumulative strength models (McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997) and multiple-trace models (Hintzman, 1986, 1988). Within the former models, information cumulates in one memory trace at the time of repetition whereas, within the latter models, another piece of memory is added when repetition happens. The model of scaffolded encoding, by having a single strong memory trace for exact repetition and multiple strong traces for partial repetition, has the potential to account for data from multiple paradigms across multiple retrieval tasks. This brings us to the second obstacle of current development of memory models, there is no unified model of multiple tasks.

A Unified Memory Model

Different memory models have been proposed to account for data of different paradigm from different retrieval tasks (Criss & Howard, 2015). For example, REM was developed to account for the absence of list strength effect (LSE) in recognition (Shiffrin & Steyvers, 1997).

The class of retrieved context models (Howard & Kahana, 1999, 2002) explains effects of response transitions in free recall well. However, different retrieval tasks, or testing memory in different ways, should not change how memories are represented or how encoding and retrieval processes are described. It's unsatisfactory that we don't have a unified memory model.

The model of scaffolded encoding, although was proposed to account for PF and PI in cued recall, has the potential to account for other effects of repetition across different retrieval tasks. With the scaffolded encoding mechanism embedded within REM, we seek to account for a series of hallmark findings from recognition, cued recall, and free recall. This is a significant theoretical advancement.

The hallmark findings are 1) LSE in recognition, cued recall and free recall (Wilson & Criss, 2017), 2) output interference (OI) in recognition (Criss et al., 2011) and cued recall (Wilson et al., 2020), and, most importantly, 3) PF and PI in single item recognition (Aue et al., 2012) and cued recall (Aue et al., 2017). The main reason we chose these paradigms is that they share the property that repetitions at study and/ or at test harm and/ or help memory. In other words, these paradigms are good testaments for the model of scaffolded encoding. The LSE is about if and how repeated presentation of some stimuli affects memories of others in the same list. OI is about how repetition during test affects memory performance. PF and PI show how repeating stimuli in different arrangements, or partial repetition, at study improves and harms memory at the same time. In addition, as mentioned before, LSE was the main reason behind REM's birth. Also, REM with concurrent encoding and retrieval implemented at test already predicts OI in recognition and cued recall. Therefore, it's important for our model to be applicable to LSEs and OIs. Detailed descriptions of these paradigms and data follow.

List Strength Effect (LSE)

The list strength paradigm investigates how memory of a stimulus is affected by the strength of other stimuli studied in the same list. The classic paradigm, shown in Figure 1, includes a manipulation of stimulus strength and list type. Stimulus strength can be weak or strong, which is manipulated through repetition, study time, or depth of processing (Craik & Lockhart, 1972; Hintzman, 1974; Malmberg & Shiffrin, 2005; Wilson & Criss, 2017). List type can be pure or mixed in strength. A pure list consists of stimuli with the same strength, whereas a mixed list consists of stimuli of different strengths. This results in 3 study lists containing four classes of target stimuli (mixed strong, mixed weak, pure strong, and pure weak). The critical comparison is between stimuli of the same strength but in different types of lists, i.e., memory for mixed strong stimuli verses pure strong stimuli, and mixed weak stimuli verses pure weak stimuli.

Consistently, we have observed LSEs in different directions from different retrieval tasks (Ratcliff, Clark, & Shiffrin, 1990, Shiffrin, Ratcliff, & Clark, 1990; Wilson & Criss, 2017). In free recall, we have observed a positive LSE (i.e., memories of the mixed strong are better than those of the pure strong and memories of the pure weak are better than those of the mixed weak, as in Figure 2c), meaning memory is harmed when other items are well-learned. In cued recall, we have observed a null LSE (i.e., memories of the mixed strong are roughly the same as those of the pure strong and memories of the pure weak are roughly the same as those of the mixed weak, as in Figure 2b), meaning memory of a stimulus is independent of the strength of other stimuli in the same list. In recognition, we have observed a null towards negative LSE (i.e., memories of the pure strong are better than those of the mixed strong and memories of the mixed weak are better than those of the pure weak, as in Figure 2a), meaning other strongly encoded items help memory. One reason we choose this paradigm is that there has not been a formal model explaining these different LSE patterns observed in different retrieval tasks. REM was designed to account for a null towards negative LSE in recognition (Shiffrin & Steyvers, 1997) and was also used to explain a positive LSE in free recall (e.g., Malmberg & Shiffrin, 2005) but was never officially implemented to explain a null LSE in cued recall (except a recent attempt made by Chen & Criss (under revision)). Also, as mentioned earlier, LSE reveals how repetitions of some stimuli in a list affects memories of others. Combining these reasons, it is important for the model of scaffolded encoding to account for the LSE.

Output Interference (OI)

OI is the finding that memory performance decreases across consecutive test trials. In single item recognition, OI shows as a declining hit rate and no change in false alarm rate across test trials (see Figure 3a). Criss et al. (2011) has captured these data patterns with a retrieval-phase learning mechanism embedded within REM. In their model, the best matching trace is updated with a test item's information that is recognized as old and a new trace is added for a test item that is recognized as new. OI in recognition was the first data pattern REM succeeds in explaining via the idea of encoding happening during retrieval (Criss et al., 2011; Kilic et al., 2017). It also demonstrates repetition of the already tested items harms memory for the to-be-tested items, and hence is important for the scaffolded encoding mechanism to account for.

In cued recall, participants study word pairs and recall a word when presented a cue. If no response is produced, this is a response failure. If a response is produced, it can be a correct response or an intrusion. OI in cued recall shows as robust increase in response failures, slight decrease in correct responses, and a null or slight decline in intrusions (see Figure 3b). Implementing a retrieval filter within SAM (Raaijmakers & Shiffrin, 1981a, 1981b), combined with retrieval-phase learning, Wilson et al. (2020) account for OI in cued recall. In this fresh work, Wilson et al. (2020) argue, filtering contributes to increasing proportion of response failures and retrieval-phase learning contributes to decreasing proportion of correct responses. Filtering does not belong to the realm of concurrent encoding and retrieval and examining the role of filtering in memory is beside our aim, so we treat it as an inherent mechanism within REM in our simulations.

In free recall, OI data patterns include decreasing performance across successive study-test blocks (Wickens, Borne, & Allen, 1963), declining proportion correct across consecutive free recall trials (Bregman & Wiener, 1970; Tulving, 1967), and longer recall latency across study-test blocks with longer retention intervals (Wixted & Rohrer, 1993). Mostly due to the complexity of these tasks, we won't consider the model's applicability to OI in free recall.

Proactive Interference (PI) and Proactive Facilitation (PF)

PI is the finding that an earlier event harms memory for a later presentation, evidence for that repetition creates confusion and in turn increases false memory. PF is the finding that an earlier event benefits the memory of its later presentation, evidence for that repetition leads to enhanced encoding and in turn increases correct recalls. In a typical design (see Figure 4), participants study a list of pairs (e.g., treat-chair and signal-point), and then a second list, which consists of rearranged pairs (e.g., treat-point) and novel pairs (e.g., hotel-water). In single item recognition, presented with an item from a studied pair (e.g., treat, hotel, chair) or a new item (e.g., apple), participants are asked to determine if it has been studied in the most recent list. Aue et al., (2012) observed both PF and PI in a recognition task. Specifically, there were more target responses of items from rearranged pairs (Target-Rearranged, e.g., “treat” is old) than from novel pairs (Target-Novel, e.g., “hotel” is old) and more foil responses of items from List 1 (Foil-List 1, e.g., “chair” is old) than novel items (Foil-New, e.g., “apple” is old) at test (see Figure 5a). In cued recall, participants are presented with a cue (e.g., treat, clock, hotel, south) and asked to remember its study partner in the most recent list. Aue et al. (2017) observed both PF and PI, in that there were more correct responses (e.g., "point" when cued by signal) and more incorrect responses (e.g., "hello" when cued by clock) from rearranged pairs than from novel pairs (see Figure 5b).

Aue and colleagues (2012, 2017) proposed and tested a variety of explanations for PF, including longer search time for a more familiar (rearranged) cue (Diller, Nobel, and Shiffrin, 2001; Mensink & Raaijmakers, 1988), lower threshold for accepting and recovering content of a memory active by a more familiar (rearranged) cue (e.g., Benjamin, 2005), and better encoding of the rearranged pairs during study (recursive reminding, e.g., Wahlheim & Jacoby, 2013). They found evidence for the last theory, that repeated items benefit the encoding of subsequently studied pairs. That is, in the second list (i.e., the most recent list), memories of the rearranged pairs are stronger than memories of the novel pairs. This verbal account captures the results and essence of the scaffolded encoding mechanism, and we expect REM with scaffolded encoding should account for PF as well as PI.

In short, most memory models have excluded concurrent encoding and retrieval and have been divided by retrieval tasks. This makes our proposed model unprecedented in two significant ways, implementing concurrent encoding and retrieval mechanism and accounting for data from multiple paradigms across multiple retrieval tasks. When the model was first presented (Criss et al., 2018), it was merely conceptually tested. Therefore, in this project, we intend to computationally implement this model and test it comprehensively.

Research Plan

The model of scaffolded encoding has encoding and retrieval processes at every step of what would be labelled as study and test stage in traditional models. It is important to identify the reasons, or specific data patterns, scaffolded encoding is needed for. In this project, we implemented five models (see Table 1) with different encoding mechanisms, ***Update+Add, Update+Scaffold***, active at study, or at test, or both and applied them to the described hallmark paradigms. The concept of concurrent encoding and retrieval implies that encoding occurs as a result of retrieval. Considering the candidate processes, encoding during retrieval requires ***Update+Add*** or ***Update+Scaffold*** implemented at test and retrieval during encoding requires ***Update+Add*** or ***Update+Scaffold*** implemented at study.

We structure this manuscript in three parts for the three hallmark paradigms, where we display and discuss simulation results of LSE, OI, and PF. In each part, we identify and discuss the relationship between the encoding mechanisms, parameter values and REMs' predictions. Detailed descriptions of the REM models are to follow.

Model Description

We implemented 5 models derived from REM in R studio (R Core Team, 2017). We first describe the identical mechanics, composition of stimuli and memory traces representations, shared by these models and then describe how each of them differ.

Shared Aspects of the Models: Representation

Across models, every stimulus, every memory trace is represented as a vector of item and context features. Item features are semantic information of the stimulus whereas context features are peripheral information. We note that there isn't an agreed definition of context. Nevertheless, in a typical experiment, as opposed to item information that is at the center of attention (e.g., Aue, Fontaine, & Criss, 2018; Malmberg & Shiffrin, 2005), context information is not instructed to attend to yet is present within the external and internal environment and naturally encoded with the item (Malmberg & Shiffrin, 2005; Murnane & Phelps, 1993, 1994, 1995; Murnane, Phelps, & Malmberg, 1999). Context could be the environmental contextual information shared by every item (e.g., Godden & Baddeley, 1975) or, if experimentally manipulated, experimental materials uniquely paired with items (e.g., Cox & Criss, 2017).

Item or context, feature values are generated randomly based on a geometric distribution with a parameter $\boldsymbol{g}$.

$$\boldsymbol{P(f = v) = (1 - g)^{v-1} g, where\ v = 1, \dots \dots \infty} \qquad (\boldsymbol{E1})$$

Feature values are generated independent of one another based on Equation 1, which says the probability of a feature $\boldsymbol{f}$ taking the value of $\boldsymbol{v}$. One can interpret parameter $\boldsymbol{g}$ as representation and manipulation of frequency and/ or similarity. Because a geometric distribution starts as describing one success with probability of $\boldsymbol{g}$ in the $\boldsymbol{v}^{\text{th}}$ time of attempt, a feature with low value occurs frequently and is common whereas a feature with high value occurs infrequently and is uncommon. There is also $\boldsymbol{g_{sys}}$, referring to the base rate of features. The parameter $\boldsymbol{g_{sys}}$ is used to generate a feature value that is failed to encoded correctly.

In general, the value of $\boldsymbol{g}$ is determined by properties of the stimulus. Throughout our simulations, for simplicity, we assume $\boldsymbol{g_c}$ ($\boldsymbol{g}$ of context) = $\boldsymbol{g_i}$ ($\boldsymbol{g}$ of item) = $\boldsymbol{g_{sys}}$ and use the default value of 0.4. In addition, instead of varying both the number of item and context features, we fix the number of item features ($\boldsymbol{Nf_i}$) at 20 per word and vary the number of context features ($\boldsymbol{Nf_c}$) from 0 to 20 by 1. When context has 20 features, context vs. item ratio is 1. For now, we ignore the possibility that context drifts from trial to trail (Howard &Kahana, 2002; Shiffrin & Steyvers, 1997).

Next, we describe the models to be tested differing at study stage and test stage via (in)active encoding mechanism(s). These mechanisms become added parameters in the model and can be turned on or off. Explanation of each parameter, or each mechanism can also be found in the REMs_README.R file.

REM0

REM0 partially comes from the basic model in Shiffrin and Steyvers (1997) and ARC-REM (assessment of retrieval completion) in Diller et al (2001). REM0 is the basic model in this current paper where encoding only happens at study and retrieval only happens at test. It also has a retrieval filtering (Wilson et al., 2020) when the retrieval task is recall. REM0 is the foundation of other models.

Study

Storage and encoding are imperfect. Every feature in a memory trace has a probability of $\boldsymbol{u}$ to be stored. Provided stored, every feature has a probability of $\boldsymbol{c}$ to be correctly encoded. In other words, every feature is not stored, is stored but encoded incorrectly, or is stored and encoded correctly. When it is not stored, feature value stays as zero, referring to empty information. When it is stored but encoded incorrectly, a value is randomly chosen from the geometric distribution of $\boldsymbol{g_{sys}}$. Note that a randomly drawn feature value could match the correct feature value by chance. When it is stored and encoded correctly, feature value is copied from the stimulus's corresponding feature value. Under no circumstance is a feature value overwritten.

In REM0, encoding process occur at study. As a stimulus, a word (i.e., a context-item vector) or a word pair (i.e., a context-item-item vector), is presented for the first time in a specific context (e.g., a study list in a lab), an episodic memory trace is stored and encoded, as in Figure 6a. When it is presented again in the following study trials, the trace representing the stimulus is updated, meaning the trace's zero features, those who are still lack of information, are stored and encoded from the presented stimulus' features following the same principles, as in Figure 6b. We refer this operation as updating.

In general, the value of $\boldsymbol{c}$ is treated as a fixed property of the person and the value of $\boldsymbol{u}$ is determined by the experimental or environmental conditions. In our simulations, for simplicity, we make $\boldsymbol{u}$ of context and $\boldsymbol{u}$ of item the same. We vary the value of $\boldsymbol{u}$ from 0.1 to 0.5 by 0.02 to observe its effect on the memory phenomenon. From past research experience with REM, 0.5 is the maximum possibly encoding rate of an event. We set $\boldsymbol{c}$ as 0.7, per REM simulation conventions.

Test

Retrieval is initiated by activating memory with a cue. For the hallmark findings we describe in this document, both recognition and cued recall use a retrieval cue is a vector of experimental context (i.e., this/ a study list) and item and is used to probe memory jointly. In free recall, a retrieval cue is a vector of context. When a cue, $\boldsymbol{q}$, is presented, it is compared to the corresponding information in every stored memory indexed by $\boldsymbol{j}$. For example, if $\boldsymbol{q}$ is a context cue, and every context vector of every memory trace tries to match this cue. Based on Equation 2, every memory trace is assigned a likelihood ratio $\boldsymbol{\lambda}$, indicating the match between $\boldsymbol{q}$ and $\boldsymbol{j}$. The likelihood ratio is computed from $\boldsymbol{P(Match|Y_{qj})}$, the probability of this trace, $\boldsymbol{j}$, matching the cue, $\boldsymbol{q}$, given this trace is produced by the cue, divided by, $\boldsymbol{P(Match|N_{qj})}$, the probability of the trace matching the cue given this trace is produced by other stimuli. It is then transformed as the probability of incorrectly and correctly encoding a feature. In the Equation 2, $\boldsymbol{n_{qj}}$ is the number of nonzero feature values in the trace that mismatch the cue. Every mismatching feature value is from incorrectly encoding that feature and takes the probability of (1 - $\boldsymbol{c}$), or 0.3. The likelihood evidence from a mismatching feature is the same regardless of its value. Then, $\boldsymbol{n_{ijm}}$ is the number of nonzero feature values in the trace that match and have the value $\boldsymbol{v}$. If the presented cue is a studied one (i.e., numerator of the second half of Equation 2), every matching feature value may be correctly encoded (i.e., $\boldsymbol{c}$) or incorrectly encoded but generated from the geometric distribution and match by accident (i.e., $(1-\boldsymbol{c})\, \boldsymbol{g_{sys}}\, (1-\boldsymbol{g_{sys}})^{v-1}$). If the presented cue is unstudied (i.e., denominator of the second half of Equation 2), a matching feature can only come from matching by accident. Given a feature value is from a geometric distribution, the higher the value, the less common it is, and the more likelihood evidence it provides. Features with the value of 0 (i.e., features that are not stored) refers to the lack of information and hence is not considered.

$$\boldsymbol{\lambda_j = \frac{P(Match|Y_{qj})}{P(Match|N_{qj})} = (1-c)^{n_{ij}} \prod_{\forall i}^{\infty} \left[\frac{c + (1-c)g_{sys}(1-g_{sys})^{v-1}}{g_{sys}(1-g_{sys})^{v-1}}\right]^{n_{ijm}}} \qquad \boldsymbol{(E2)}$$

Decision Rule: Recognition

In a (single item) recognition task, where participants are presented with a studied stimulus or an unstudied stimulus (i.e., a context-item vector) and need to provide an “old” or “new” response, such response requires a subjective memory decision made from a combined activation of memories, $\boldsymbol{\Phi}$, computed based on Equation 3,

$$\boldsymbol{\Phi} = \sum \lambda_j \,/\, \boldsymbol{j} \qquad \textbf{(E3)}$$

Where $\sum \lambda_j$ is the sum of likelihood ratios across all $\boldsymbol{j}$ traces. Then, $\boldsymbol{\Phi}$, familiarity, is compared to a criterion, $\boldsymbol{E}$. If familiarity exceeds this criterion, the presented item is considered old. Otherwise, it is considered new. In our simulations, we give $\boldsymbol{E}$ a standard value of 1 (Shiffrin & Steyvers, 1997).

In the original REM paper, $\boldsymbol{E}$ as 1 was used because Shiffrin and Steyvers suggested a two-step retrieval approach. Memories that are activated by the context cue, those with likelihood ratios over a threshold value (i.e., $\boldsymbol{\tau}$) of 2980.9, are then probed by the item cue. That is, context is excluded from making a recognition decision. However, this threshold value was meant for 40 context features (Shiffrin & Steyvers, 1997) and is not applicable to this current investigation as we are changing the number of context features from 0 to 20. We could change the value of $\boldsymbol{\tau}$ accordingly in the first stage of retrieval. However, this approach is unfounded, and it would add variance to our model evaluation. Therefore, for simplicity and for our primary interest of choosing the most powerful model, we stuck to using a joint cue to probe memory at test and choose a standard criterion value of 1. In the discussion section, we implemented several other methods to improve this process.

Decision Rule: Recall

In a cued recall task, participants are presented with a cue (i.e., a context-item vector) and asked to remember the other half studied with this trace. In a free recall task, participants are asked to remember as many items as possible given an experimental context (i.e., a context vector). In either case, recall consists of sampling, choosing a memory out of many, and recovery, generating and reporting the content of the chosen memory. Both sampling and recovery are probabilistic processes.

A single trace is sampled in proportion to the match between the cue and this trace.

$$\boldsymbol{P(S_j|q) = \frac{\lambda j}{\sum \lambda j}} \qquad (\boldsymbol{E4})$$

From Equation 4, a memory trace $\boldsymbol{j}$ has the probability of $\boldsymbol{P(S_j|q)}$ getting sampled under the presented cue $\boldsymbol{q}$. $\boldsymbol{P(S_j|q)}$ is calculated from λ_j, the likelihood ratio of memory trace $\boldsymbol{j}$, divided by and $\sum\lambda_j$, the sum of likelihood ratio of all $\boldsymbol{j}$ memory traces. In addition, λ_j needs to exceed a criterion, $\boldsymbol{E}$, to be activated. A criterion $\boldsymbol{E}$ is implemented to ensure the sampled memory must match the cue to a certain extent. Otherwise, a trace is sampled every time regardless whether it is similar to the cue. When a memory trace, $\boldsymbol{j}$, is sampled and activated, the probability of the remaining parts of the trace is recovered, $\boldsymbol{P(R_j|q)}$, is calculated based on Equation 5,

$$\boldsymbol{P(R_j|q) = \rho^{\tau}} \quad (\boldsymbol{E5})$$

where $\boldsymbol{\rho}$ is the proportion of features in this trace that are stored correctly, and $\boldsymbol{\tau}$ is a scaling parameter. In our simulations, we set $\boldsymbol{\tau}$ as 0.5 following (consistent with Malmberg and Shiffrin (2005)). The role of $\boldsymbol{\tau}$ and its psychological underpinnings has not yet been investigated.

In recall, when the system recovers a memory trace successfully and produces an output, it is a correct response or an intrusion. When the system fails to report a response, it is a response failure. Response failure could happen under several possible scenarios, which are 1) failure to recover a sampled memory trace, 2) sampling of a trace of an already outputted response (i.e., a retrieval filter), and 3) sampling of a trace whose activation doesn't exceed the criterion $\boldsymbol{E}$. We implement a retrieval filter such that sampling of a memory of an already generated response is prohibited and this test trial is counted as a response failure (SAM, Raaijmakers & Shiffrin, 1981a, 1981b; Wilson et al., 2020). In recall where multiple attempts at retrieval are made with the same context cue, a parameter, ***Kmax***, is needed to decide when to stop memory search. Search continues until the number of response failures (***K***) hits this limit (i.e., ***Kmax***). ***Kmax*** is subject to both person (e.g., effort) and experimental constraints (e.g., instructions). In our simulations, we set ***Kmax*** as the number of stimuli.

In the following models, we will start incorporating ***Update+Add*** or ***Update+Scaffold*** at study and/ or test within REM. They are variations that build up to the full model, the model of scaffolded encoding. As mentioned before, incorporating ***Update+Add*** or ***Update+Scaffold*** at study is actualizing retrieval during encoding and incorporating ***Update+Add*** or ***Update+Scaffold*** at test is actualizing encoding during retrieval.

The retrieval and encoding processes at study and at test are identical, except that, at test, a joint cue of context and item probes memory whereas, at study, context and item are evaluated separately. This is a post-hoc assumption considering the common context features shared by stimuli and memories. If a joint cue of context and item were used at study, identical context features would make new stimuli falsely recognized as old. If only item were used to probe memories at test, performance would not change across different number of context features. Both approaches would prevent us from evaluating the models extensively. Therefore, we made a compromise and adopted processes only different in the usage of context and item cue. Thorough investigation and discussion of the role of context will be presented in the discussion section.

REM1

REM1 is REM0 with ***Update+Add*** at test. Criss et al., (2011) were the first who implemented this model to account for OI in recognition. In this model, we modify and apply this mechanism to both recognition and recall.

Study

The study stage of REM1 is identical to that of REM0. Encoding and storage of every stimulus happens as already described, a new trace is added for a new word or new pair. Repetitions of a word or pair result in an updating of the trace, by assumption.

Test (Update+Add)

The test stage of REM1 includes an encoding process, with encoding rate of $\boldsymbol{u_{test}}$ and following the same principles in the study stage. The general rules of the ***Update+Add*** mechanism is that 1) a new memory trace is added when the current episode is new (as in Figure 6a), 2) an old memory trace is updated when the current episode is old (as in Figure 6b), and 3) an old memory trace is updated and a new trace is added when the current episode is semi-old (as in Figure 6c and 6e). As a joint cue of context and item is used at test to evaluate the test item in recognition or to sample a memory in recall, a semi-old event is when the test context is evaluated separately as new. This second retrieval step may not matter to our current investigation, given we don't vary context between study and test, but it leaves a window for future endeavor of applying this model to other memory effects with emphasis on drifted context.

In single item recognition, like in Criss et al., (2011), if familiarity (calculated from Equation 3) of a test probe (i.e., a context-item vector) exceeds a criterion (i.e., ***E***) (i.e., an old response), the best matching trace is updated with the (context and item) features of the current test event. If its familiarity doesn't exceed the criterion (i.e., a new response), a new trace is added with encoded features of the current test event. Unlike Criss et al., (2011), our REM1 is not a pure-item model. We consider the old/ new nature of context. As an old response is produced by the model, the model evaluates if familiarity of the context exceeds the criterion. If it doesn't, the context is recognized as new and this is considered as a semi-old event. In this scenario, in addition to updating, a new trace is added with newly encoded item features and current context features.

In recall, under a retrieval cue (a context in free recall, or context-item vector in cued recall), if a memory trace is sampled and successfully recovered, the recovered trace is updated with the cue features and features of the outputted response (as in Figure 6b if it is a correct response). Then, familiarity of the context is evaluated. If the context is recognized as new, a new trace is added representing this semi-old event (as in Figure 6c). If no memory trace is sampled, no encoding process takes place. Recognition of context follows the same rule described before (Equation 3 and comparison with the criterion $\boldsymbol{E}$).

In our simulations, we have an encoding probability at test ($\boldsymbol{u_{test}}$), which we set the same as the encoding probability at study ($\boldsymbol{u}$) for consistency between study and test.

REM2

REM2 is REM1 with the ***Update+Add*** (see Figure 6a, 6b, 6c, and 6e) mechanism implemented at study. In other words, REM2 has concurrent encoding and retrieval at study and test stage.

Study (Update+Add)

In REM2, we implement the process for updating at study - whether an additional trace is added to memory and/ or an old trace is updated depends on the recognition of the studied stimulus. In studying a word (i.e., a context-item vector), the model evaluates familiarity of the context and item and reaches a recognition decision according to the same rule described earlier. If the item is recognized as new, this is a new stimulus and a new trace is added. If both context and item are old, this is an old stimulus and the best matching trace is updated. If the item is old while the context is new, in addition to updating, a new trace is added representing this semi-old event.

The encoding and retrieval process in studying a word pair (i.e., context-item-item vector) (e.g., A-B, C-D) depends on the old/ semi-old/ new decision made based on associative recognition of the pair (Shiffrin & Steyvers, 1998). During retrieval, one of the items from the pair is recognized as old or new. If it is old, its best matching trace’s counterpart is then subject to recognition. If the counterpart is old, then this pair is intact (i.e., A-B). If the counterpart of the memory trace is new, the other item from the tested pair is then determined old (i.e., a rearrange pair, A-D) or new (i.e., A-Z). It is also likely that both items are new (i.e., a new pair, E-F). Correspondingly, an update and/or add/ scaffold encoding process is implemented following an old/ new/ semi-old recognition decision. A new event is when both items are new (i.e., E-F), in which case a new trace is added (as in Figure 6a). An old event is an exact repetition, when an item is old, its best matching trace's counterpart is old (i.e., A-B), and the context is old. In this case, the best matching trace is updated (as in Figure 6b). A semi-old event can happen with multiple possibilities, which include 1) both items are old but not from the same trace (i.e., a rearranged pair, A-D, as in Figure 6f), 2) one of the items is old while the other one is new (i.e., A-Z), and 3) both items are old and from the same trace (i.e., A-B) while the experimental context is new (as in Figure 6c). When one of these conditions is met, in addition to updating the best matching trace(s) with the current context features and the corresponding old item(s) features, a new trace is added representing this semi-old event.

Test (Update+Add)

The test stage of REM2 is identical to that of REM1.

<u>REM3</u>

REM3 is REM2 with ***Update+Scaffold*** at study. Theoretically, scaffolded encoding is incorporating information of repetition into formation of a memory. The idea being that when a memory is retrieved it is used to support the formation of new memory in a new context. Mechanically, scaffolding a trace means adding a trace with features borrowed from a previously updated memory, instead of with newly encoded features from feature values as zero. The advantage of scaffolded encoding is, from a conceptual perspective, combining and separating two events at the same time. Updating old memory trace(s) with new information and adding a new trace with both retrieved and new information represents combination of two events. Yet such having multiple traces means there are two episodes that took place separately. From a practical perspective, scaffolded encoding ensures the added trace is at least as strong as the old trace, such that sampling the added trace is at least as likely as sampling the old one.

As in Figure 6, the general rules of the ***Update+Scaffold*** mechanism is that 1) a new memory trace is added when the current episode is new (as in Figure 6a), 2) an old memory trace is updated when the current episode is old (as in Figure 6b), and 3) an old memory trace is updated and a new trace is scaffolded with updated features when the current episode is semi-old (as in Figure 6d and 6f). The only difference between the ***Update+Add*** the ***Update+Scaffold*** is the encoding process when the episodic event is semi-old (i.e., Figure 6c vs Figure 6d, and Figure 6e vs Figure 6f).

Study (Update+Scaffold)

At study, compared to REM2, REM3 is only different in its encoding mechanism when the study event is a semi-old. In studying a word (i.e., a context-item vector), a semi-old event is when item's familiarity exceeds the criterion (i.e., old), yet context's doesn't (i.e., new), in which case, the best matching trace is updated, and a new trace is scaffolded borrowing features from the updated trace. In studying a word pair, (i.e., a context-item-item vector), in deciding the old/new nature of the current stimulus (following the associative recognition process described earlier), a semi-old event can be one of the followings, 1) an intact pair presentation yet the study context is recognized as new (as in Figure 6d), 2) both items of the pair are old but from different traces (as in Figure 6f), or 3) one item of the pair is old while the other one is new. When a semi-old event occurs, the best matching trace(s) is updated with the current context features and the corresponding old item(s) features and a new trace is scaffolded borrowing features from the updated trace(s).

Test (Update+Add)

The test stage of REM3 is identical to that of REM2. That is, REM3 doesn't have a consistent encoding/ retrieval process between study and test.

REM4

REM4 is the full model of scaffolded encoding we propose in this research project. Compared to REM3 which implements scaffolded encoding only at study, REM4 implements scaffolded encoding at both study and test.

Study (Update+Scaffold)

The study stage of REM4 is identical to that of REM3.

Test (Update+Scaffold)

At test, compared to REM2 and 3, REM4 is only different when the test event is a semi-old. In recognition, it could be when the test stimulus's familiarity exceeds the criterion and an old decision is made yet test context has changed from study and a new decision is made. In recall, it could be when a trace is sampled and recovered under a retrieval cue, or when a trace is sampled but not recovered, but the context is recognized as new. In this case, the current test event is considered semi-old. Then, the recognized/ recovered trace is updated, or the sampled trace's cue component is updated with cue features, and a new trace is scaffolded borrowing features from the updated trace (as in Figure 6d).

Methods

To reiterate, the goal of this project is to examine the model of scaffolded encoding and evidence of 1) every REM's power in accounting for every hallmark finding, 2) every encoding mechanism's impact on every hallmark finding, and 3) potential interactions between the mechanisms and parameters.

This writing is a post-registration (Lee et al., 2019) of the model of scaffolded encoding, detailing the necessity of mechanisms and development of the model. For the sake of the incremental and exploratory nature of model development, we hope this paper is to inform and improve future modeling endeavor related to REM.

We implemented every hallmark paradigm (i.e., LSE in recognition, cued recall, and free recall, OI in recognition and cued recall, and PF and PI in recognition and cued recall) within every model described, varying the number of context features ($\boldsymbol{Nf_c}$, from 0 to 20 by 1) and encoding rate ($\boldsymbol{u}$, from 0.1 to 0.5 by 0.02). We vary these two parameters because different combinations of $\boldsymbol{Nf_c}$ and $\boldsymbol{u}$ lead to different relative strengths of item versus context features, which, from past research, tend to affect model predictions qualitatively or quantitatively in a significant scale (e.g., Shiffrin & Steyvers, 1997). We keep other parameters (e.g., $\boldsymbol{g}$, $\boldsymbol{c}$, $\boldsymbol{E}$) constant either because their variations correspond to specific manipulations (e.g., parameter $\boldsymbol{g}$ for word frequency) which we don't consider here or because they represent structural processes, as opposed to control processes (Atkinson & Shiffrin, 1968), and were carefully fitted and calibrated (e.g., $\boldsymbol{c}$) during initial model development (Shiffrin & Nobel, 1997; Shiffrin & Steyvers, 1997) and hence are anything but liberal. This is also standard practice following predecessors (e.g., Malmberg & Shiffrin, 20005; Kilic et al., 2017).

For each memory paradigm, we applied every model and generated "data" for 1000 subjects for each combination of $\boldsymbol{Nf_c}$ and $\boldsymbol{u}$ value. One exception is when $\boldsymbol{Nf_c}$ is less than 2 (i.e., $\boldsymbol{Nf_c} = 0$ or $\boldsymbol{Nf_c} = 1$) in free recall. Given a retrieval cue in free recall consists of context only, a context vector with 0 or 1 feature can't initiate memory retrieval. For each model, firstly, we computed memory performance (i.e., hit rate and false alarm rate in recognition, and proportion of correct responses, intrusions and response failures in cued recall, proportion of correct in free recall). Then, from overall performance, we calculated a measurement for each memory effect as a quantitative estimation. We also conducted hypothesis testing on these measurements, to evaluate models' predictions qualitatively. This allows us to measure if an effect is present and the magnitude of the effect as well as the relationship among model predictions, encoding mechanisms, and parameter values.

The main goal is to compare different models and see which one(s) successfully predict all the data patterns in all memory paradigms. Therefore, we didn't evaluate and discuss the impact of encoding mechanisms and parameter values on memory performance till such a successful model is chosen. For now, the focus is each model's capability (i.e., under which model with what encoding mechanism activated and with what parameter values is a memory phenomenon successfully predicted).

Next, we review each paradigm and discuss simulation results.

Table 1

Models with different mechanisms implemented at study and test.

Models	Active Mechanism (s)	
	Study	Test
REM0	\	\
REM1	\	***Update+Add***
REM2	***Update+Add***	***Update+Add***
REM3	***Update+Scaffold***	***Update+Add***
REM4	***Update+Scaffold***	***Update+Scaffold***

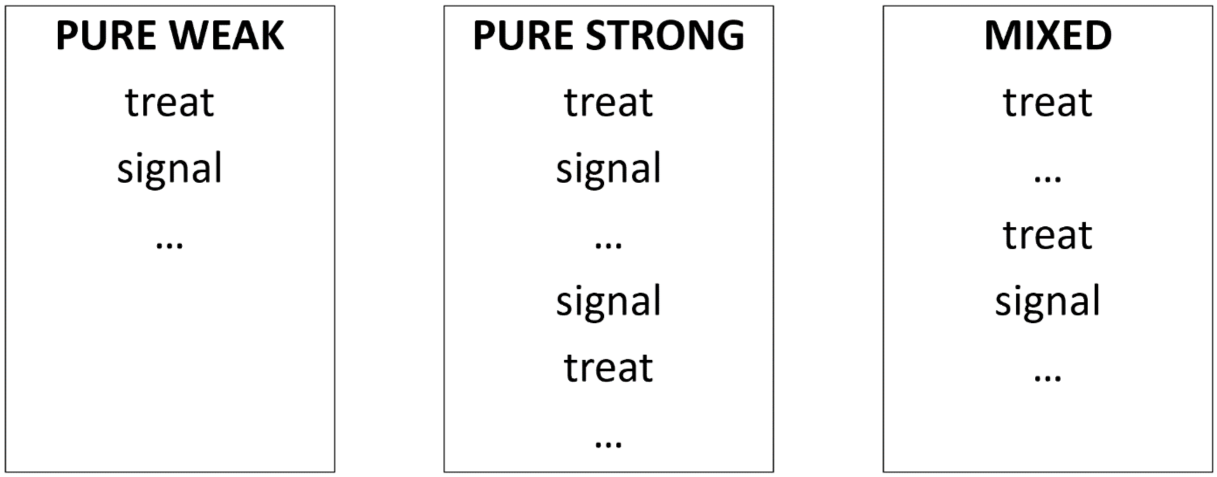


Figure 1

A classic list strength paradigm. It consists of a pure weak list (with weak stimuli), a pure strong list (with strong stimuli presented multiple times), and a mixed list (with half weak and the other half strong stimuli). Although only two stimuli are shown in each list, study lists are longer (and represented by … in the figure) in an actual experiment.

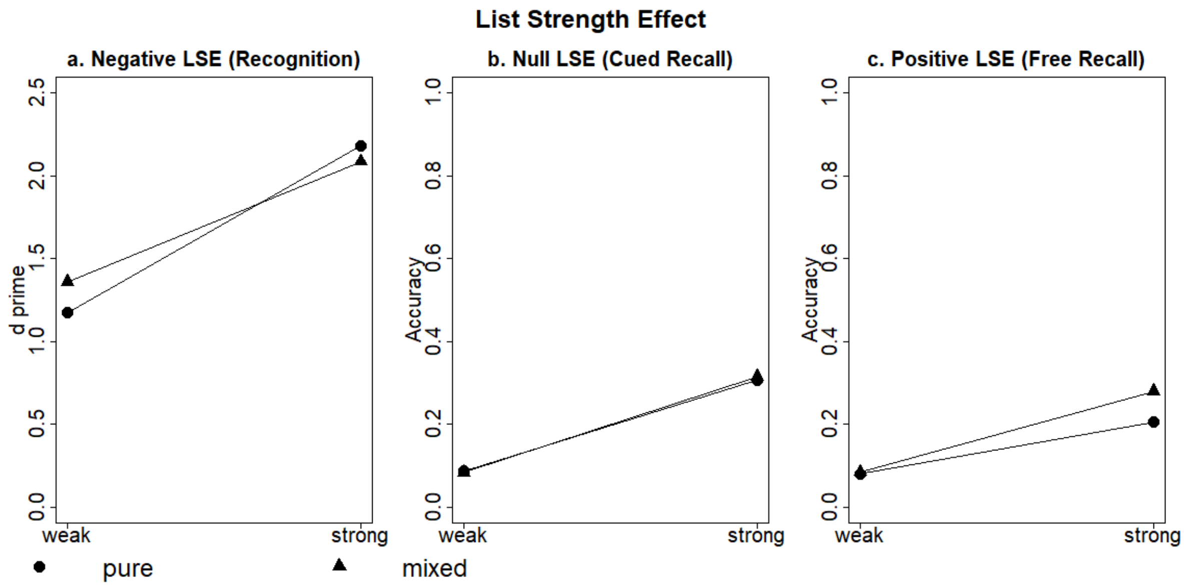


Figure 2

List strength effect data patterns from Wilson and Criss (2017)'s Experiment 4. There is a negative list strength effect in recognition (a), a null LSE in cued recall (b), and a positive LSE in free recall (c).

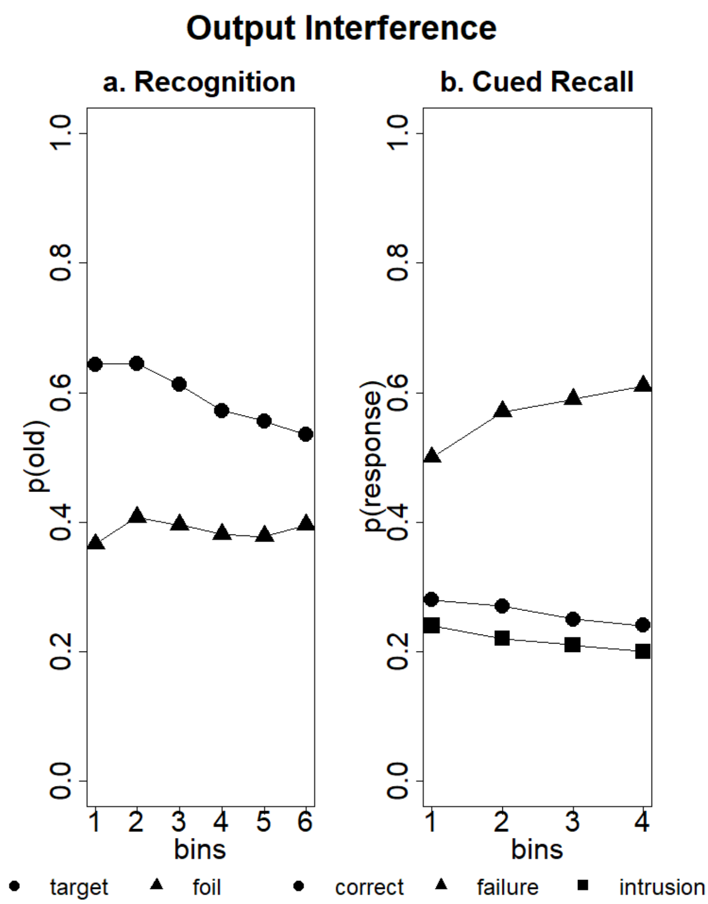


Figure 3

Output interference data patterns. In single item recognition (Experiment 1, Criss et al., 2011) (a), it demonstrates as decreasing hit rate and unchanging false alarm rate across 6 bins of 25 test trials from 6 lists. In cued recall (Wilson & Criss, 2017) (b), it demonstrates as robust increasing proportion of response failures, decreasing proportion of correct responses, and slightly decreasing proportion of intrusions across 4 bins of 4 test trials.

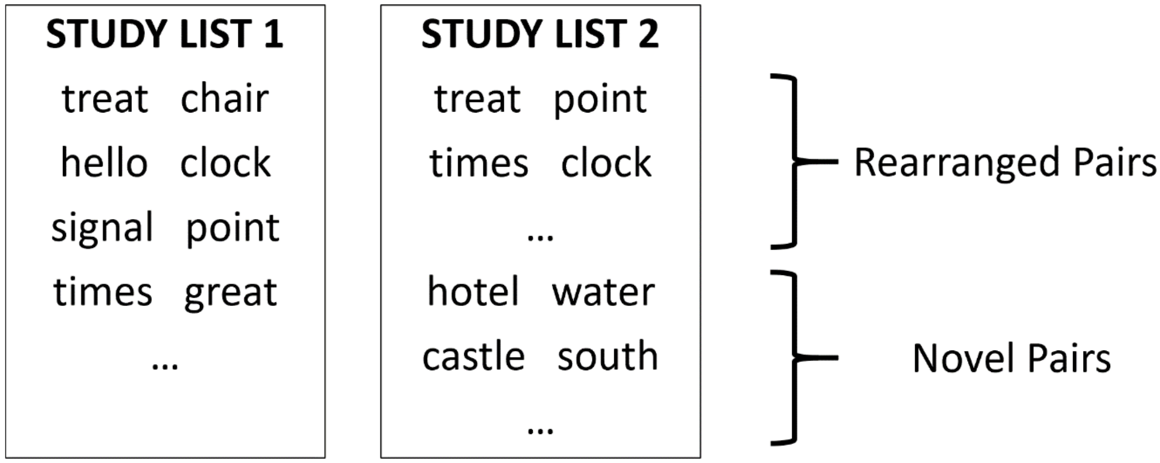


Figure 4

A typical proactive interference and proactive facilitation paradigm. It consists of two study lists. Study list 1 is made of randomly associated word pairs. Study list 2 is made of rearranged pairs (i.e., already presented words from study list 1 but paired differently) and novel pairs (i.e., newly composed pairs). For simplicity of presentation, only four pairs are displayed, but study lists are longer (and represented by … in the figure) in an actual experiment.

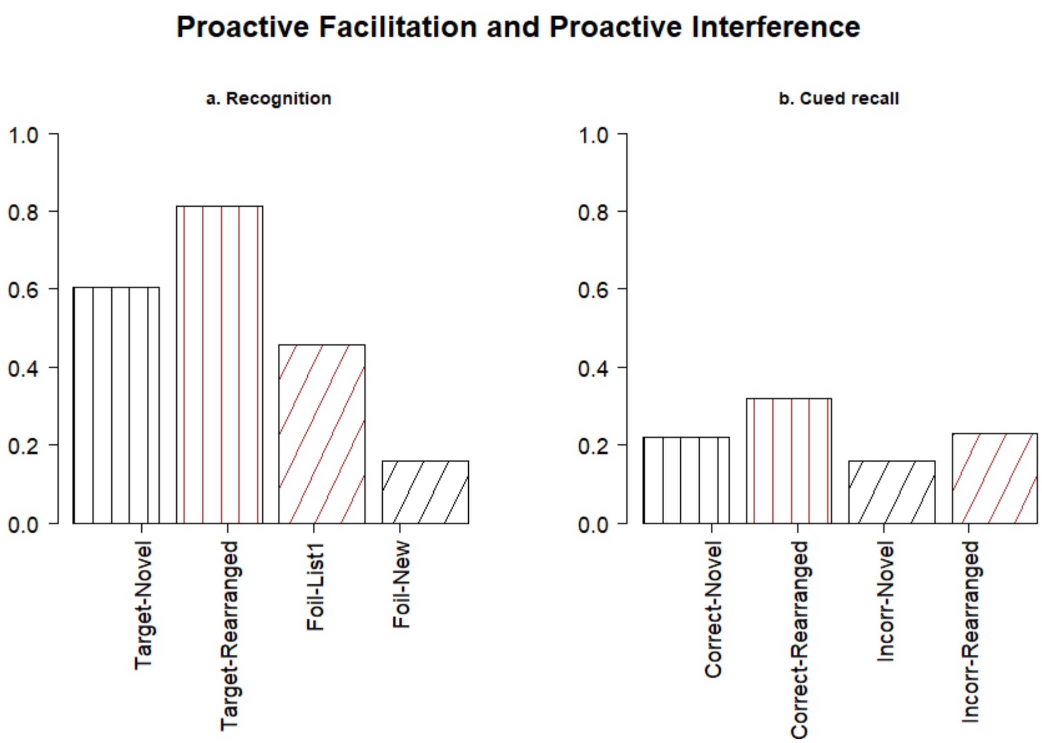
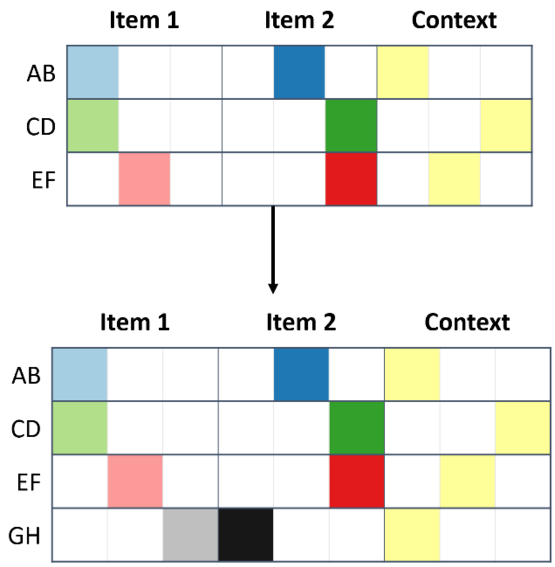
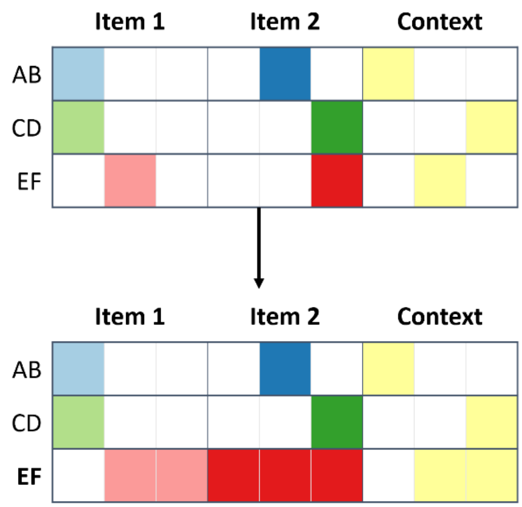


Figure 5

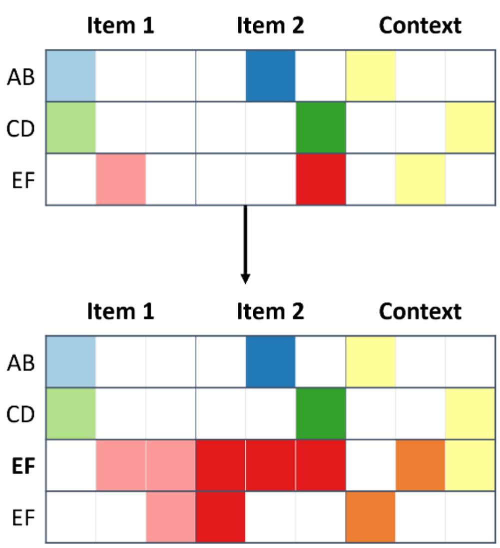
Proactive facilitation and proactive interference data patterns (Aue et al., 2012). In single item recognition (a), facilitation is more target responses (vertical) for items from rearranged pairs (brown) than from novel pairs (black) and interference is more foil responses (diagonal) for items from List 1 than new items. In cued recall (b), facilitation is more correct responses and interference is more incorrect responses from rearranged pairs than from novel pairs.



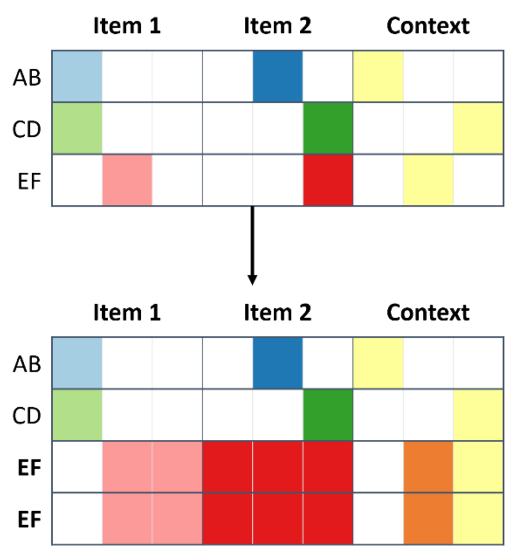
(a) adding a new trace
when the episode is new



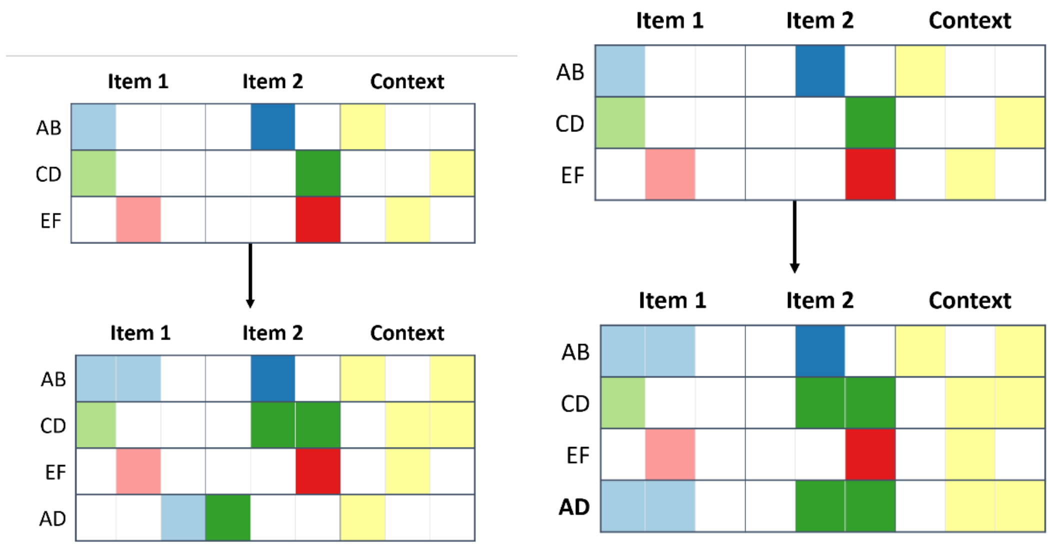
(b) updating an old trace
when the episode is old



(c) updating and adding
when the episode is semi-old with a new context



(d) updating and scaffolding



(e) updating and adding (f) updating and scaffolding

when the episode is semi-old with a rearranged pair.

<u>Figure 6</u>

A graphic illustration of the encoding mechanisms. ***Update+Add*** is made of a, b, c, and e.

Update+Scaffold is made of a, b, d, and f.

List Strength Effect (LSE)

LSE or the lack of it evidences the impact of better encoding of some stimuli on memory for other stimuli in the same list. In a typical list strength paradigm (see Figure 1), participants study and are tested (via recognition, cued recall, or free recall) on three lists with four types of stimuli (words or word pairs). They are, a pure weak list with weak stimuli, a pure strong list with strong stimuli, and a mixed list with weak and strong stimuli.

The research question is how the described REM models (with different retrieval/ encoding mechanisms at study and/ or test and with different parameter values) differently predict LSEs in different tasks (as in Figure 2). REM was developed to account for a null towards negative LSE in recognition (Shiffrin & Steyvers, 1997) and has been able to account for a positive LSE in free recall (Malmberg & Shiffrin, 2005). What we are interested in is if it can do that and to account for a null LSE in cued recall (Chen & Criss, under revision),

In our implementation, every study list has 16 words in recognition and free recall, and 16 word pairs in cued recall. With varying active encoding mechanisms and parameter values, we computed difference of difference (*DoD*) score (Wilson & Criss, 2017) based on Equation 6, where *MS*, *MW*, *PS*, and *PW* refer to, respectively, accuracy in recall or d prime in recognition of the mixed strong (*MS*), mixed weak (*MW*), pure strong (*PS*), and pure weak (*PW*) items.

$$DoD = (MS\text{-}MW) - (PS\text{-}PW) \quad (E6)$$

As an interaction term, *DoD* is a standard measure for the LSE (Ratcliff et al., 1990). The positive or negative value of an *DoD* indicates the direction of LSE and the absolute value of an *DoD* indicates its magnitude. We evaluated models via hypothesis testing and computation of *DoD*s. We conducted t-test comparing *DoD* values to 0, with significance level (i.e., α) as 0.05 to measure the direction of the DoD (see Figure 7). If the *DoD* value is significantly bigger than 0, it's a positive LSE and is displayed in blue. If it is not significantly different from 0, it is a null LSE and is displayed in white. If it is significantly smaller than 0, it is a negative LSE and is displayed in red. To illustrate the magnitude of the LSE, we also drew *DoD* values in color depths (see Figure 8). That is, the redder it gets, the more negative LSE produced, and the bluer it gets, the more positive LSE produced. From Figure 7 and 8, we observe the relationship between model predictions and parameter values within every REM model in a qualitative and quantitative way. We emphasize that, with 1000 subjects, changes in effects may be amplified in Figure 7 depicting significance whereas they stick to their actual values in Figure 8 depicting magnitudes.

Results and Discussions

From Figure 7 and 8, all model versions, including the model of scaffolded encoding (i.e., REM4), predict the same list strength patterns as from data (see Figure 2). Overall, negative LSEs are predicted in recognition, albeit with variance depending on active encoding mechanisms and parameter values. Null LSEs in cued recall and positive LSEs in free recall are predicted across all models. This informs us that REMs predict LSEs due to core assumptions inherent to REM, rather than the encoding mechanisms we added. In the following paragraphs, we discuss them in detail by encoding mechanisms and retrieval tasks.

Separate Encoding and Retrieval (REM0)

Recognition

When encoding happens (only) at study and retrieval happens (only) at test, negative LSEs (see Figure 7a and 8a) are predicted in recognition with encoding rate higher than 0.1 and the number of context features smaller than 10. In addition, as the number of context increases (e.g., 15), stronger encoding rate (e.g., 0.3) is required to generate a negative LSE.

To understand why a negative LSE requires these combinations of parameter values, we need to understand the source of LSE in recognition in the original REM. Models (global matching models, Gillund & Shiffrin, 1984; Hintzman, 1984; Humphreys, Bain, & Pike, 1989; Murdock, 1982) before REM predicted a positive LSE in recognition, contrary to data (Ratcliff et al., 1990; Shiffrin et al., 1990). According to these models, memories of strong stimuli provide more signals for recognizing themselves as studied items but also produce more noise interfering other stimuli from being recognized. In this sense, mixed-strong memories harm mixed-weak memories (i.e., positive LSE). Then, models built with differentiation (SLiM, McClelland & Chappell, 1998; REM, Shiffrin & Steyvers, 1997) were adopted.

A null towards negative LSE in recognition was one of the incentives of REM's birth (Shiffrin & Steyvers, 1997). In this seminal REM paper, differentiation is implemented and is used to account for a null towards negative LSE. Differentiation refers to the idea that, the more a stimulus is presented, the clearer its memory becomes and the less confusable its memory is from others. Within REM, when a stimulus is presented again, differentiation is actualized by updating features of its memory trace, such that its memory is has fewer empty features and is more completed. Then, in recognition, where an old/ new judgment is made, a more completed memory trace provides more evidence that a target matches its target trace and more evidence that a foil mismatches. Therefore, we observe that strong stimuli in a pure list are better remembered than those in a mixed list, and weak stimuli in a mixed list are better remembered than those in a pure list, yielding a negative LSE (i.e., having strong stimuli helps memories of others).

Under the prerequisite that a negative LSE in recognition is due to differentiated memories, the combinations of parameter values are required to generate the correct patterns because they produce memories that are differentiated enough. In this sense, it's obvious why encoding rate matters. When any feature value has little probability to be stored, differentiation can't happen. In turn, a negative LSE can't happen. In terms of context features, it is a bit complicated. The fact that not all $\boldsymbol{Nf_c}$ values predict negative LSE has to do with the way items and contexts are set up in typical experimental design and model implementation (Malmberg & Shiffrin, 2005; Wilson & Criss, 2017). Usually, items (i.e., words) are unique and experimental context is shared by all traces in a list, unless manipulated otherwise (Chen & Criss, under revision; Murnane & Phelps, 1993, 1994, 1995; Murnane et al., 1999). Correspondingly, in the model, context features are similar or even identical among memories whereas item features are different. That is, context features don't differentiate memories. On the contrary, more context (identical) features make memories more obscure, counteracting the effects of differentiation from item (unique) features. Therefore, memories with more context features in a list don't help memories of others, failing to predict a negative LSE. It explains why negative LSE disappears when there are too many context features (e.g., 15) with small encoding rate (e.g., 0.1) (as in Figure 7a and 8a). In addition, for the model to generate a negative LSE, increased number of context features (e.g., 15) requires a higher encoding rate (e.g., 0.3), where differentiated item features counteract undifferentiated context features (as in Figure 7a and 8a). This argument is in accord with the literature where item is assumed weighted more than context in recognition task and especially in list strength paradigm (Shiffrin et al., 1990; Wilson & Criss, 2017).

In short, the prediction of a negative LSE in recognition depends on the level of differentiation of memories, or differentiated item features outperforming undifferentiated context features, which is affected by encoding rate and the number of context features.

<u>Cued Recall</u>

Under separate encoding and retrieval mechanism, from qualitative evaluation (i.e., Figure 7a), null LSEs are predicted in cued recall but with limited parameter space, where encoding rate is relatively weak (e.g., 0.1) and the number of context features is relatively small (e.g., 6), where encoding rate is moderate (e.g., 0.3) and the number of context features is moderate (e.g., 12), and where encoding rate is relatively strong (e.g., 0.5) and the number of context features is relatively large (e.g., 18). Other combinations of parameter values (e.g., weak encoding rate and larger number of context features) generate either positive or negative LSEs. However, as mentioned before, because we simulated a large number of subjects, small changes in *DoD* values can be amplified to be significant. As we look at quantitative evaluation (i.e., Figure 8a), null LSEs, or LSEs with *DoD* values extremely close to 0, are predicted. This finding match results in literature where null LSEs have been consistently found in cued recall (Wilson & Criss, 2017), even with manipulations pushing the weighting of context or item (Chen, Wilson, & Criss, in writing).

Within REM, the level of competition at retrieval induced by the retrieval cue determines LSE in cued recall (Chen & Criss, under revision). Competition comes from memories getting sampled in proportion to their match to the retrieval cue. Because any memory is made of context and item (Criss & Shiffrin, 2004; Cox & Criss, 2017; Ratcliff et al., 1991; Shiffrin et al., 1991; Wilson & Criss, 2017) and, with standard manipulation, contexts are common whereas items are distinct, competition then is affected by the amount of common/ context information, relative to distinct/ item information, memories contain. When memories are similar to each other with many common/ context information, the retrieval cue has the potential to match any stored memory trace, making sampling a highly competitive process. In this case, memories are equally likely to be sampled in a pure list whereas mixed strong memories with more encoded features are more likely to be sampled than mixed weak memories in a mixed list. This leads to better memory for mixed strong than pure strong and better memory for pure weak and mixed weak, or a positive LSE. Now think about the opposite scenario where memories are highly different with few common/ context information. Sampling becomes less competitive in that a retrieval cue matches a specific memory trace and mismatches others. The stronger the memory is encoded, the more evidence of match and mismatch it provides. In this case, in any type of list, the more strongly encoded memories, the better list performance. More to our point, in a list mixed in strength, when a mixed weak cue is presented, it matches a specific mixed weak memory and mixed strong memories with more encoded features provide strong evidence that they themselves are mismatches. Mixed weak memories are helped by mixed strong memories and are better than pure weak memories. When a mixed strong cue is presented, it matches a specific mixed strong memory and mixed weak memories with less encoded features provide weak mismatch evidence. Mixed strong memories are harmed by mixed weak and are worse than pure strong memories, or, a negative LSE. There is also a third scenario in between the two extreme conditions just described, where memories have almost equal amount of distinct/ item and common/ context information and the level of competition during retrieval is moderate. In this case, the probability of a memory trace getting sampled is only affected by its strength but isn't affected by other memories, or a null LSE. Therefore, in cued recall, using a joint cue of context and item generates null LSEs (see Figure 8a)

Free Recall

When encoding and retrieval are separate, positive LSEs are predicted across all combinations of encoding rates and number of context features (as in Figure 7a and 8a), except when the number of context features is less than 2 and free recall can't operate. In addition, we also observed more positive LSEs with encoding rate higher than 0.2 and the number of context features more than 5 (as in Figure 8a).

As said before, the source of LSE in recall is the level of competition at test, which depends on how much memories match the cue and different from each other. In free recall, as the retrieval cue consists of only context and context is shared among memories, stored contexts, undifferentiated parts of the memories, engage in highly competitive sampling during retrieval, resulting in a positive LSE (as in Figure 7a and 8a). When the encoding rate increases and/ or when the number of context features increases, there are more context (shared) features stored in memories, increasing the level of competition and hence increasing the magnitude of LSE (as in Figure 8a).

One may wonder if it's possible to generate or observe LSE other than positive in free recall, and the answer is yes when the storage of context features is impaired. As the strengthening method was changed to extending study time or encoding task with deeper processing (Malmberg & Shiffrin, 2005), encoding and storage of context features in strong memories are not superior than those in weak memories. During sampling, contexts of the mixed strong memories can't outcompete those of the mixed weak memories, producing a null LSE.

Update+Add at Test (REM1)

When the ***Update+Add*** mechanism is active, a new trace is added when the current event is judged as new (see Figure 6a), an old trace is updated when the current event is judged as old (see Figure 6b), and an old trace is updated as well as a new trace is added when the current event is judged as semi-old (see Figure 6c and 6e).

Overall, compared to REM0, major changes in simulation results from REM1 are in recognition task. In cued recall, although effects change qualitatively (see Figure 7a and Figure

7b), possibly due to 1000 simulated subjects, LSEs remain as null quantitatively (see Figure 8a and Figure 8b). In free call, the predicted positive LSEs are not affected. This is because, as elaborated before, the source of LSE in recall is the level of competition at test, or how much a retrieval cue matches the stored memories. Under the conventional setting that context is common whereas items are distinct information, a context cue can match every memory whereas an item cue specifically match a single memory trace. Therefore, regardless of the added encoding mechanisms, the usage of a joint cue of context and item produces a moderate level of competition, which, in turn, leads to null LSE in cued recall and the usage of a pure context cue produces a high level of competition, which, in turn, leads to positive LSE in free recall.

Now we spent more words elaborating changes in predicted LSEs in recognition.

Recognition

From Figure 8b and 9b, negative LSEs are predicted with encoding rate higher than 0.12 and the number of context features smaller than 10. Positive LSEs are predicted with encoding rate smaller than 0.35 and the number of context features bigger than 12.

As elaborated before, the main factor affecting LSE in recognition is differentiation. Differentiation, affected by encoding rate and the number of context features, also explains why, compared to REM0, REM1 (i.e., REM0 with ***Update+Add*** at test) has a more restricted parameter space for negative LSEs. When memories are not differentiated enough (e.g., encoding rate is 0.2 and the number of context features is 15), a studied stimulus presented at test may be incorrectly recognized as new (i.e., miss). Accordingly, information of this old stimulus, instead of accumulated in an old memory trace, is encoded and stored in a new copy added to memory. The memory list grows longer, and denominator gets bigger in Equation 3. It counteracts the idea of differentiation, creates more interference for the following trials, and produces a null towards positive LSE (as in Figure 7b and 8b).

Update+Add at Study and at Test (REM2)

Similarly, major changes in simulation results are from recognition task.

Recognition

When the ***Update+Add*** mechanism is active both at study and at test, unless encoding rate is higher than 0.2 and the number of context features is less than 5, a null or a positive LSE is predicted (as in Figure 8c and 9c).

Compared to REM1 where the ***Update+Add*** mechanism is only active at test, in REM2, there are less combinations of parameters that generate negative LSEs. Again, we attributed this pattern to differentiation. As every studied word is evaluated before encoding happens, restudying no longer results in updating an old memory trace automatically. Now, when memory is weakly encoded (i.e., $\boldsymbol{u} < 0.2$), it provides little familiarity evidence for whether the presented stimulus is old or new and an already studied word may be recognized as new. Then, a completely new trace is added, rather than updating an old trace. The study list ends up with multiple copies of one stimulus. 'Strongly' encoded memories are simply multiple weak memories. As a result, later at test, given a recognition decision is made from averaged activation from memories, in a mixed list, weakly encoded memories are harmed by the 'strongly' encoded memories in a similar fashion to list length effect and are worse than weak memories in a pure list. A positive LSE is generated. A null towards positive LSE is also predicted by the multi-trace memory models (e.g., MINERVA2, Hintzman, 1984; 1986), which assumes repetition leads to storing additional exemplars or memory traces.

Update+Scaffold at Study and/ or at Test (REM3 and REM4)

When the Update+Scaffold mechanism is active, a new trace is added when the current event is new (see Figure 6a), an old trace is updated when the current event is old (see Figure 6b), and an old trace is updated as well as a new trace is scaffolded borrowing updated features from the old trace when the current event is semi-old (see Figure 6d and 6f).

It appears implementing Update+Scaffold at study and/ or at test don't affect the LSE predictions in recognition, cued recall, or free recall (see Figure 7d, 7e, 8d and 8e), compared to implementing ***Update+Add*** (see Figure 7c and 8c). This is because the only difference between the Update+Scaffold and the ***Update+Add*** is the encoding mechanism when the current event is semi-old and there is a small chance for a stimulus to be detected as semi-old (i.e., an old item with a new context), due to our simplified implementation that context stays the same among study trials as well as between study and test.

<u>Summary</u>

All the implemented models successfully predict a negative LSE in recognition, a null LSE in cued recall, and a positive LSE in free recall, because of core mechanisms inherent to REM. Nevertheless, we found model dependence of predicting negative LSE in recognition on encoding rate. This conclusion can be tested in the future in an experiment where we vary the number of presentations for both weak and strong stimuli, matching the encoding rate in the model, and observe if and how LSE changes.

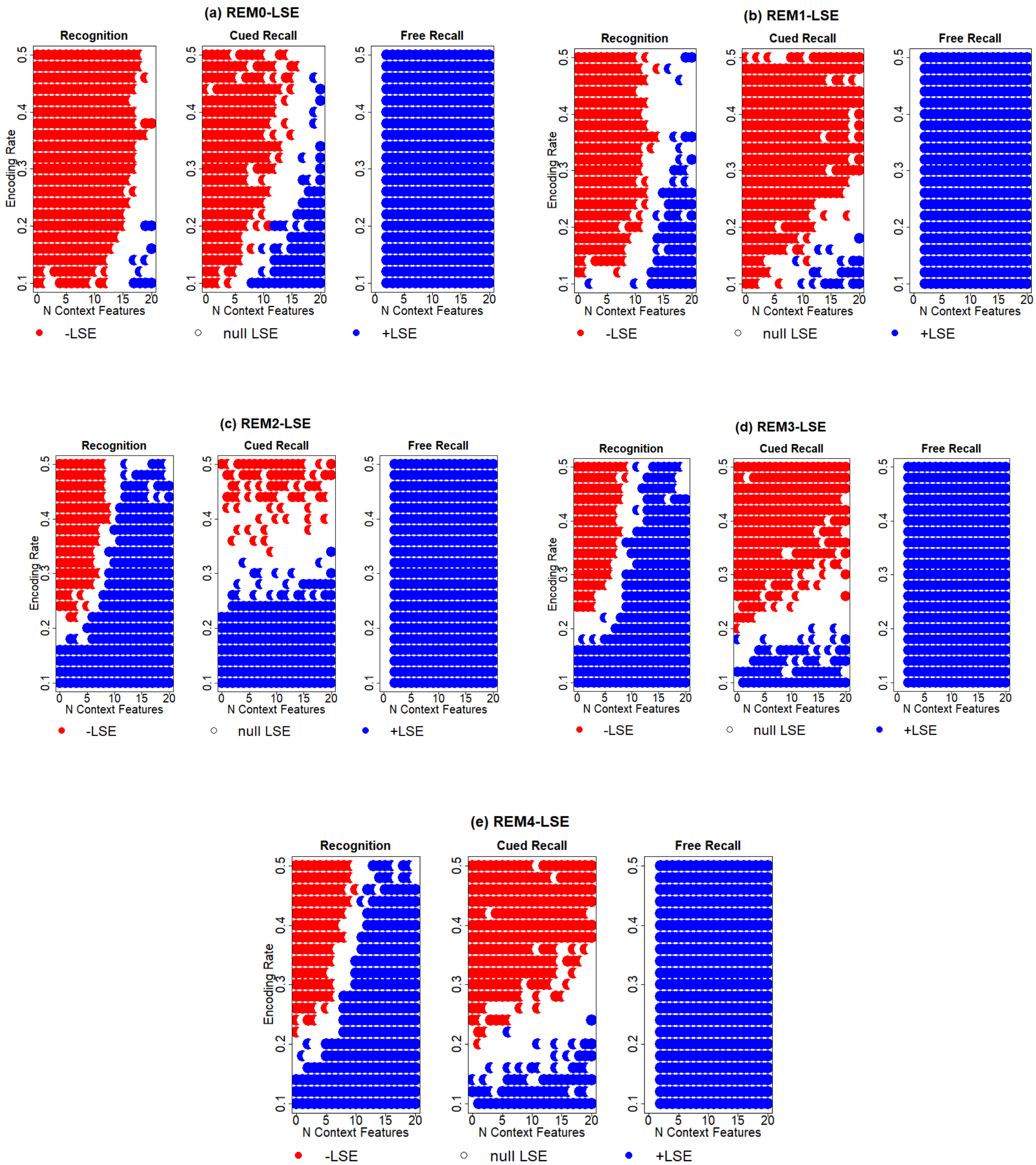
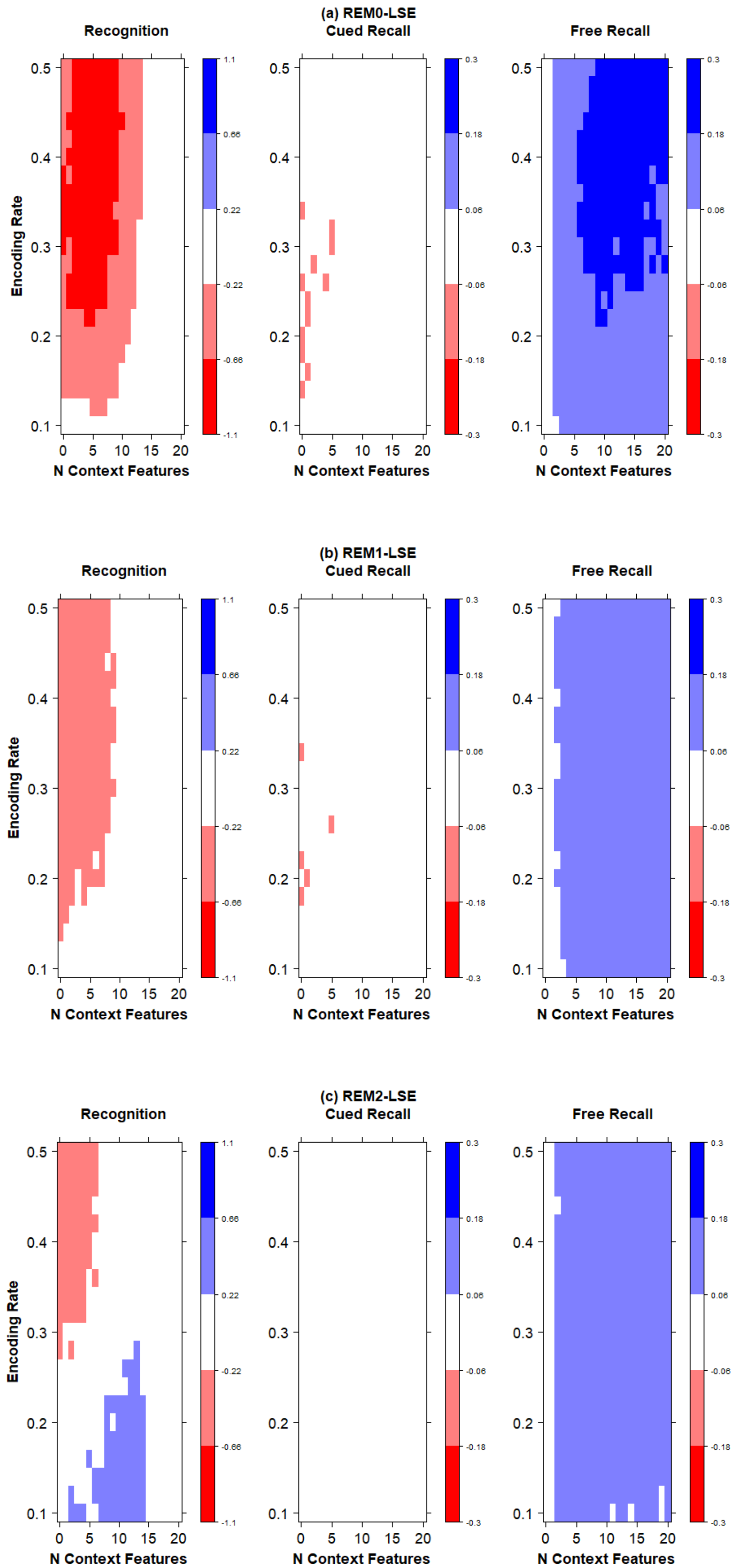


Figure 7

Qualitatively list strength effects (LSEs) from REM0 (a), REM1 (b), REM2 (c), REM3 (d), and REM4 (e) simulations. It demonstrates possible directions of LSEs under different models, different encoding rates ($\boldsymbol{u}$) (y axis), and different numbers of context features ($\boldsymbol{Nf_c}$) (x axis). A red, white, and blue circle represents, respectively, a negative, null, and positive LSE.


(a) REM0-LSE
Recognition
Cued Recall
Free Recall
Encoding Rate
N Context Features
(b) REM1-LSE
Recognition
Cued Recall
Free Recall
Encoding Rate
N Context Features
(c) REM2-LSE
Recognition
Cued Recall
Free Recall
Encoding Rate
N Context Features

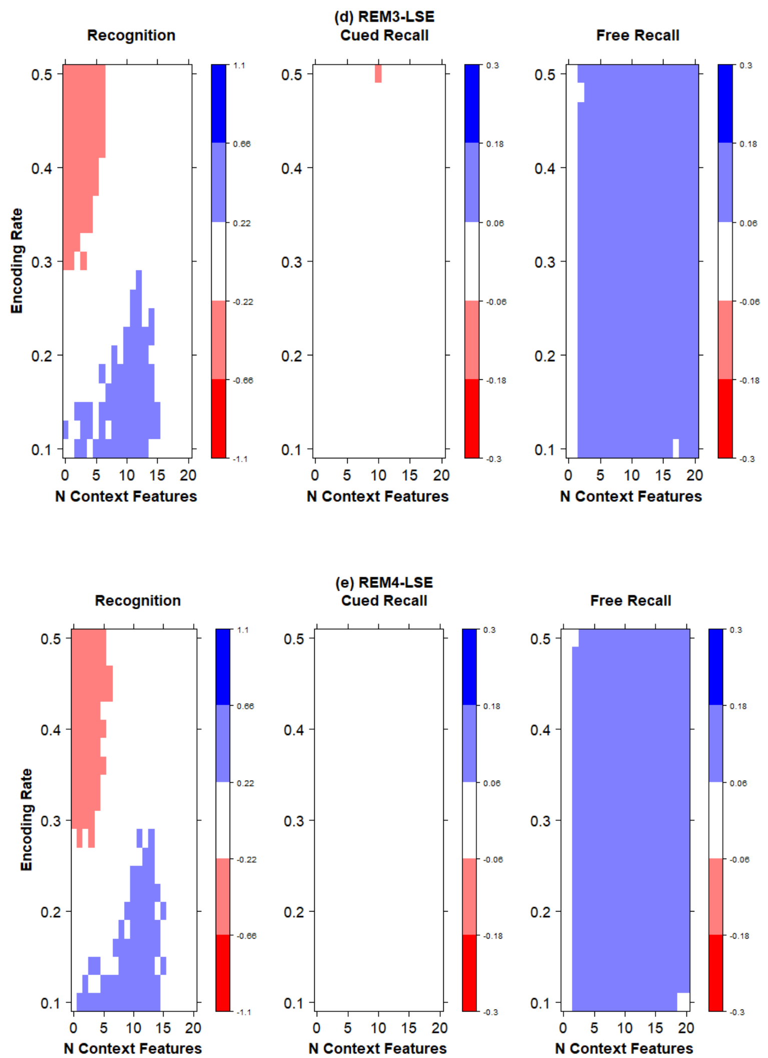


Figure 8

Quantitatively predicted list strength effects (LSEs) from REM0 (a), REM1 (b), REM2 (c), REM3 (d), and REM4 (e) simulations. It demonstrates the magnitude of the predicted LSEs under different models, different encoding rates ($\boldsymbol{u}$) (y axis), and different numbers of context features ($\boldsymbol{Nf_c}$) (x axis). The redder, the more negative list strength effect. The bluer, the more positive list strength effect.

Output Interference (OI)

OI is the finding that the already-tested items affect memories of the to-be-tested items. The OI paradigm is a simple study-test design, where participants study and are tested (via recognition or cued recall) on a list of stimuli (words or word pairs) and performance is plotted as a function of test trial.

The research question is how different models, with different encoding mechanisms and combinations of parameter values, predict OI patterns in recognition and cued recall. The model of scaffolded encoding is built on Criss et al., (2011) which implemented concurrent retrieval and encoding at test and accounted for OI in recognition. In addition, Wilson et al., (2020), with concurrent retrieval and encoding at test and a filtering mechanism (included in all REMs here), accounted for OI in cued recall. Therefore, we are interested to see if our model predicts OI in recognition (i.e., decreasing hit rate and unchanging false alarm rate) and in cued recall (i.e., increasing proportion of response failures and decreasing proportion of correct responses) as in data (see Figure 3).

In our implementations, in recognition, a study list is made of with 30 words and a test list is made up with 30 targets (studied words) plus 30 foils. In cued recall, a study list has 30 word pairs. By dividing 60 test trials in recognition and 30 test trials in cued recall evenly into 6 test blocks and averaging performances, we computed a standard measure of OI (Criss et al., 2011; Malmberg et al., 2012; Kilic et al., 2017; Wilson et al., 2020), performance slope, or the correlation coefficient between memory performance and test blocks for each subject. According to the literature in recognition the hit rate has a slope less than 0, and false alarms has a slope close to 0. In cued recall, OI reflected in the proportion of correct responses has a slope less than 0, and OI reflected in the proportion of response failures has a slope greater than 0. Via hypothesis testing, we conducted t-test comparing slope values to 0, with significance level (i.e., α) as 0.05. A slope significantly less than 0 is displayed in red, one significantly more than 0 is displayed in blue, and one insignificantly different from 0 is displayed in white (see Figure 9). To evaluate the magnitude of the slope, we also drew slope values in color depths (see Figure 10). The depth of red shows the magnitude of decreasing performance indicators (i.e., hit rate, false alarm rate, correct response, or response failure) and the depth of blue shows the magnitude of increasing performance indicators. Therefore, if a model successfully predicts OI in recognition and in cued recall, we should see a red dot in Recognition-Hit, a white dot in Recognition-False alarm, a red dot in Cued Recall-Correct, and a blue dot in Cued Recall-Failure from Figure 9 and 10. We evaluate qualitative and quantitative evidences of model predictions and their relationships with parameter values.

Results and Discussions

The basic REM (i.e., REM0) is unable to predict the OI patterns in recognition (as in Figure 9a and 10a) in that, without encoding at test, no change happens to memories. REM0 produces OI in cued recall (as in Figure 9a) (i.e., decreasing proportion of correct responses and increasing proportion of response failures in cued recall), although the predicted magnitudes of the declines in correct responses are relatively small (as in Figure 10a). This is due to filtering, as described later. Other models, including the model of scaffolded encoding (i.e., REM4), are able to produce all the proper OI patterns in both recognition and cued recall (see Figure 9b, 9c, 9d, 9e, 10b, 10c, 10d, and 10e) with some combinations of parameter values. Given these successful models (i.e., REM1, 2, 3, and 4) are in common that there is an encoding mechanism, ***Update+Add*** or Update+Scaffold, during retrieval and there is a retrieval filter, we conclude that concurrent encoding and retrieval at test and filtering already outputted responses are essential to

OI. These findings are consistent with the literatures (Criss et al., 2011; Kilic et al., 2017; Wilson et al., 2020). Nevertheless, prediction and magnitude of OI depend on parameter values. In the following paragraphs, we discuss them in detail by encoding mechanisms and retrieval tasks.

Separate Encoding and Retrieval (REM0)

Recognition

When there is no concurrent encoding and retrieval at either study or test, the model fails to predict OI and generates slopes close to 0 in recognition (see Figure 9a and 10a). Any deviation from 0 reflects the stochastic nature of the model. This result is sensible given the spirit of OI is that the retrieved memories interfere with the to-be-retrieved ones. The basic REM doesn’t have a mechanism for such interference.

Cued Recall

Different from the case of recognition, OI patterns are present even without concurrent encoding and retrieval. From Figure 9a, with almost all combinations of encoding rate and the number context features, REM0 predicts decreasing proportion of correct responses and increasing proportion of response failures. From Figure 10a, the magnitude of decreasing proportion of correct responses is relatively small, or even 0 when the encoding rate is below 0.3 and/ or when the number of context features is below 5. The magnitude of increasing proportion of response failures enhances when encoding rate is more than 0.2 and when the number of context features is more than 10.

The reason the basic REM predicts OI in cued recall is because we implemented the filtering mechanism. OI in cued recall has two signature patterns, decreasing correct responses and increasing response failures. The former has been found in cued recall task and reported in the literature (Raaijmakers & Shiffrin, 1981b; Tulving & Arbuckle, 1963) whereas the latter has been noticed less often (Tulving & Arbuckle, 1966). Both patterns are accurately predicted by a SAM with learning during retrieval and a retrieval filter (Wilson et al., 2020). Filtering (Raaijmakers & Shiffrin, 1981; Raaijmakers & Shiffrin, 1981b) refers to the idea that, a once outputted response is filtered and not produced again under the same cue. If a memory trace of such response is sampled again, the response is filtered resulting in a response failure. The number of response failures increases over test trials because participants give an incorrect response earlier at test which is supposed to be a correct answer for a later trial yet is not allowed to be outputted again. Wilson et al. (2020) found the retrieval filter is essential to produce increasing response failures from simulations of the SAM model which is the predecessor to REM. Here we found similar results applying the filtering mechanism to REM.

The reason the magnitude of increasing proportion of response failures depends on parameter values is because there are combinations of encoding rate and the number of context features that make some of the memory traces sampled repeatedly. As mentioned earlier, context is identical among memories whereas items are unique under conventional setting. Therefore, as there are more encoded context (identical) features, a memory trace is likely to be sampled repeatedly for having these matching features, increasing the proportion of response failures across blocks.

In addition, REM0 predicts that the proportion of correct responses decreases as a function of test block (see Figure 9a), but the magnitude is small (see Figure 10a). We think the presence of this pattern is because, as an already outputted response is not allowed to be produced again, this prohibited response might be an incorrect response in a previous trial yet ought to be a correct response in a future trial. In other words, a response is denied a chance to be correctly outputted in the right trial. This leads to decreasing proportion of correct responses across test blocks. For the same reason, its magnitude, albeit small, becomes more robust as the encoding rate increases and/ or the number of context features increases (see Figure 10a). As argued before, there are combinations of encoding rate and the number of context features that result in repeated sampling of one memory trace. In these scenarios, the number of responses that are available to be outputted as a correct response decreases, decreasing the proportion of correct responses across blocks.

Lastly, the reason the magnitude of decreasing proportion of correct responses is small is because the filtering mechanism is not its main source. This point is evidenced by SAM simulations of Wilson et al. (2020). When they implemented an encoding mechanism during retrieval, decreasing proportion of correct responses is present whereas increasing proportion of response failures is absent. When they implemented an encoding mechanism during retrieval as well as a retrieval filter, both patterns are present.

Update+Add at Test (REM1)

Recognition

With the ***Update+Add*** mechanism added, during recognition, as the test stimulus is recognized as old, the best matching trace is updated with the current context features and the test item features (as in Figure 6b). If the test context is recognized as new, it is a semi-old event and a new trace is added in addition to updating (as in Figure 6c). As the test stimulus is recognized as new, a new trace is added (as in Figure 6a). Based on Figure 9b and 10b, OI patterns in recognition (i.e., decreasing hit rate and unchanging false alarm rate) are predicted when the number of context features is 0 or 1. In addition, based on Figure 10b, as encoding rate increases, the magnitude of decreasing hit rate reverses to the opposite direction.

Having the ***Update+Add*** mechanism at test is essential to OI in recognition, as suggested by previous thorough studies (Criss et al., 2011; Kilic et al., 2017; Kilic, Fontaine, Malmberg, & Criss, submitted). Think about the possible responses, hit, false alarm, correct rejection, and miss. As a test stimulus is correctly recognized as old (i.e., a hit response), and as the corresponding old trace is located, updating the old trace accumulates information and differentiates it from others. This updated trace becomes less likely to match any to-be-tested stimulus in the future trials, decreasing hit rates and false alarm rates. As a test stimulus is incorrectly recognized as old (i.e., a false alarm), updating a memory trace supposedly for another test item accumulates wrong information, decreasing hit rates even more. As a test stimulus is recognized as new, correctly (i.e., a correct rejection) or incorrectly (i.e., a miss), adding a new trace extends the length of the memory list and adds variance, which is the opposite of differentiation, decreasing hit rates and increasing false alarm rate. These possibilities collectively contribute to the OI patterns (i.e., decreasing hit rates and unchanging false alarm rate across test blocks) in recognition under the ***Update+Add*** mechanism at test.

However, correct OI patterns in recognition are restricted to extremely small number of context features. This is because the choice of criterion value (i.e., $\boldsymbol{E} = 1$). According to the original REM (Shiffrin & Steyvers, 1997), context is used merely to active memories and does not participate in making a recognition decision. Then, these activated memories are probed by item and an old/ new response is generated. Here, we used a joint cue to probe memory and a criterion value of 1 for simplicity and for our main goal of choosing a model that is capable of accounting for all effects. This is also partially because we made the over-simplified assumption that context stays the same from trial to trail. Later in discussion, we adopted and discussed different adjustments to make model predictions less restricted to the number of context features.

In addition, consistent with Kilic et al. (2017)'s simulation study, the magnitude of decreasing hit rate depends on encoding rate (see Figure 10b). As mentioned before, decrease in hit rate comes updating memories following correctly (i.e., a hit response) and incorrectly (i.e., a false alarm) endorsing a test stimulus as old. When memories are more strongly encoded and more differentiated to begin with, there are fewer empty features left that can be encoded and stored, diminishing the effects of the updating process and producing less robust decrease in hit rate across test trials (as in Figure 10b).

Cued Recall

A model with ***Update+Add*** active at test produces decreasing proportion of correct responses (see Figure 9b and 10b) in a greater magnitude compared to a model without one (comparing Figure 10a to 10b). In addition, the magnitude of the decrease in correct responses grows when encoding rate is higher than 0.2 and the number of context features is more than 10. The magnitude of the increase in response failures (caused by the filtering mechanism as discussed) grows when the number of context features is more than 10 and as the encoding rate increases, as mentioned earlier.

Results here are consistent with Wilson et al. (2020) simulations in SAM and our previous argument, that an encoding mechanism during retrieval is the main reason for the decline in correct responses, although the filtering mechanism also contributes to by prohibiting potentially correct responses for the future test trials from being outputted. Now that ***Update+Add*** is implemented, a sampled and recovered memory trace is updated with the retrieval cue features (current context and item cue) and the recovered item features (as in Figure 6b). Because items features are distinct and context features are common among memory traces, updating memories with item features differentiates them from each other whereas updating memories with context features make them more confusable. In this sense, updating item features doesn't affect the probability other memories getting sampled in future trials whereas context features do. Therefore, when there are enough context features (e.g., 10) and the encoding rate is strong enough (e.g., 0.3), the already recovered and updated memories have more stored context features, outcompete other memories during sampling and are likely to be sampled again. Repeated sampling the same memory traces leads to worsening accuracy in future trials and an increase in response failures (as in Figure 10b). This explains why implementing ***Update+Add*** predicts a decreasing proportion of correct responses and such prediction depends on parameter values. It also explains why the magnitude of response failure depends on encoding rate and the number of context features.

Update+Add at Study and at Test, Update+Scaffold at Study and/ or at Test (REM2, REM3 and REM4)

Similar, if not identical, OI patterns are generated by REMs with ***Update+Add*** at study (see Figure 9c and 10c), Update+Scaffold at study (see Figure 9d and 10d), and Update+Scaffold at test (see Figure 9e and 10e), so we discuss them together here. In addition, OI patterns predicted by these models are like those predicted by REM with ***Update+Add*** at test (comparing Figure 9c, 9d, 9e, 10c, 10d and 10e to Figure 9b and 10b). This implies two questions, 1) why doesn't a retrieval mechanism at study affect the OI patterns and 2) why doesn't a different encoding mechanism at test affect the OI patterns?

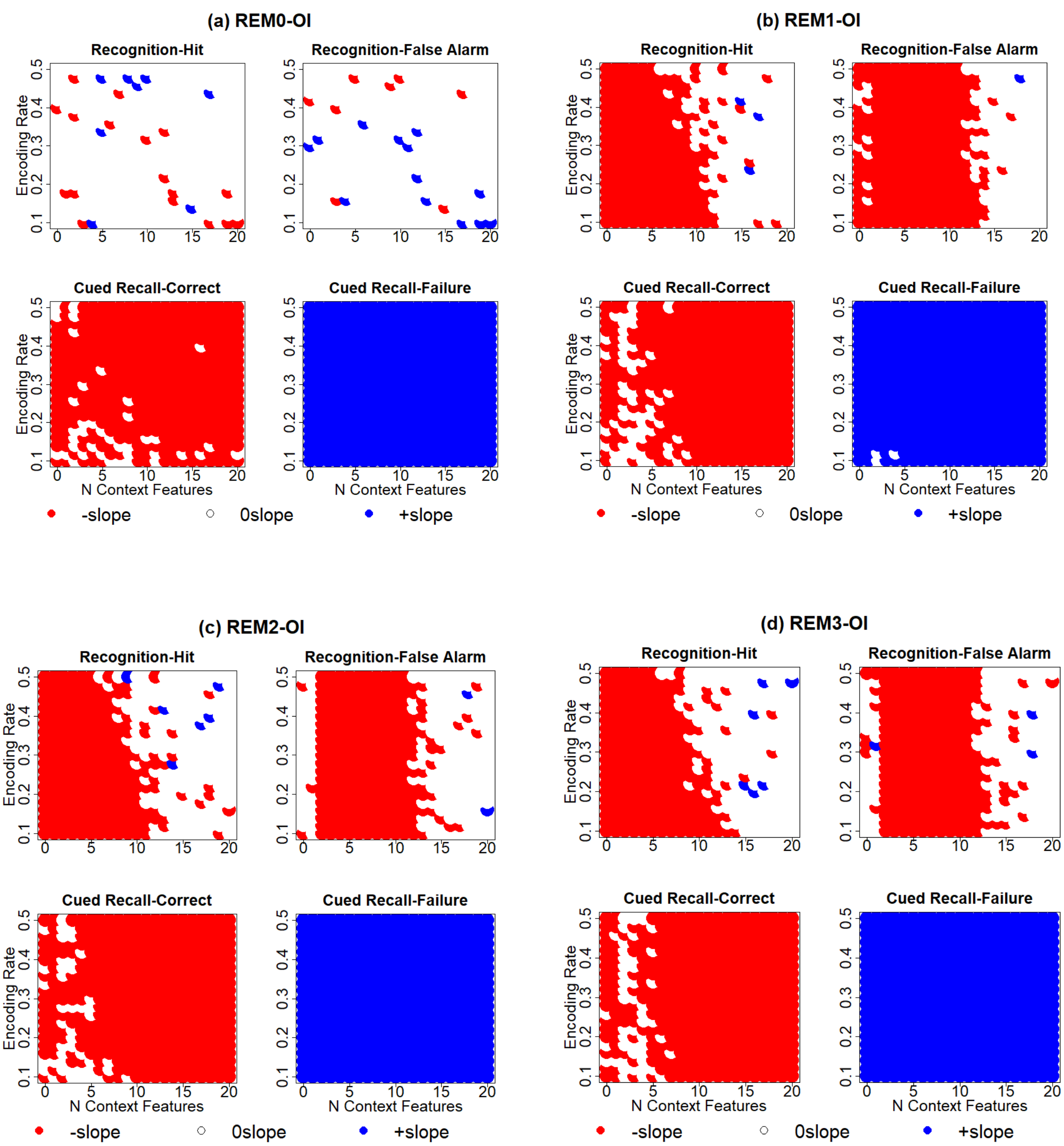
To answer the first question, it is simply because there was no repetition of stimuli implemented in the current OI design. As a retrieval mechanism is implemented at study, a presented stimulus is accessed via recognition and a decision is made whether it is old, new, or semi-old. Correspondingly, the best matching trace is updated, a new trace is added, a new trace is added (i.e., ***Update+Add***) or scaffolded (i.e., Update+Scaffold) in addition to updating the old trace. This mechanism affects memory the most when there are stimuli that are presented repeatedly and there are failures of recognizing them as old. In the previous LSE section, it leads to storing multiple memory traces for one stimulus and affects the data patterns related to strength. Therefore, in the OI design where stimulus strength isn't actively manipulated at study, implementing a retrieval mechanism at study doesn't affect the generated OI patterns.

As for the second question, let's review the difference between ***Update+Add*** and Update+Scaffold. These two encoding mechanisms are identical when the current event is new (i.e., a new trace is added, as in Figure 6a) and when the current event is old (i.e., an old trace is updated, as in Figure 6b). However, they are different when the current event is semi-old. According to ***Update+Add***, in addition to updating an old trace, a new trace is added (as in Figure 6c and 6e) while, according to Update+Scaffold, a new trace is scaffolded with copied updated features from the old trace (as in Figure 6d and 6f). However, given we assume context doesn't change from trial to trial, when a test stimulus is recognized as old or when a memory trace is successfully sampled or recovered, it is unlikely that the test context is recognized as new and the current event is identified as semi-old. In other words, changes in the encoding processes corresponding to a semi-old event, adding or scaffolding a new trace, should not affect simulation results for this paradigm

More importantly, in recognition, although updating and adding collectively contribute to OI recognition, updating an old trace and adding a new trace affect OI patterns, or memory of the to-be-tested items differently. Updating an old trace leads to a substantial decrease in hit rate from correctly and incorrectly endorsing a test stimulus as old. Such impact on hit rate comes from accumulated signal, condensed evidence from the updated memory trace. Adding a new trace merely leads to a slight decrease in hit rate. Such impact on hit rate comes from added noise, scattered evidence from at least two memory traces. This makes the updating operation steepens hit rates more drastically than the adding operation (Kilic et al., 2017; Kilic et al., submitted). In other words, the updating operation plays a major role contributing to OI in recognition. Therefore, whether it is ***Update+Add*** or Update+Scaffold implemented at test doesn't affect OI patterns in recognition. Similarly, in cued recall, because the source mechanisms of decreasing proportion of correct responses and increasing proportion of response failures are, respectively, updating recovered memories and the retrieval filter, their patterns are not affected by whether a new trace is added (i.e., ***Update+Add***) or scaffolded (i.e., Update+Scaffold).

Summary

All the implemented models, except for the basic REM (i.e., REM0), successfully predict the OI patterns in recognition and in cued recall, for having an encoding mechanism at test, either ***Update+Add*** or Update+Scaffold, and for having a retrieval filter, built in all the models here. Nevertheless, our compromising choice of using a joint cue in recognition as well as using a criterion value of 1 renders model predictions extremely limited to parameter values. We will modify this process in the Discussion section. Furthermore, parameter values affect the amount of stored and updated undifferentiated information in memories, which in turn affects the probability of repeated sampling of memories in cued recall, making models' predictions depend on encoding rate and the number of context features.


(a) REM0-OI
(b) REM1-OI
(c) REM2-OI
(d) REM3-OI
Recognition-Hit
Recognition-False Alarm
Cued Recall-Correct
Cued Recall-Failure
Encoding Rate
N Context Features
-slope
0slope
+slope

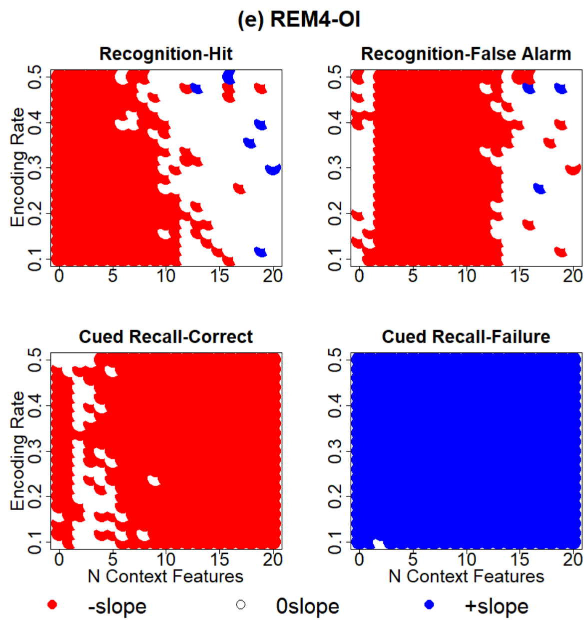
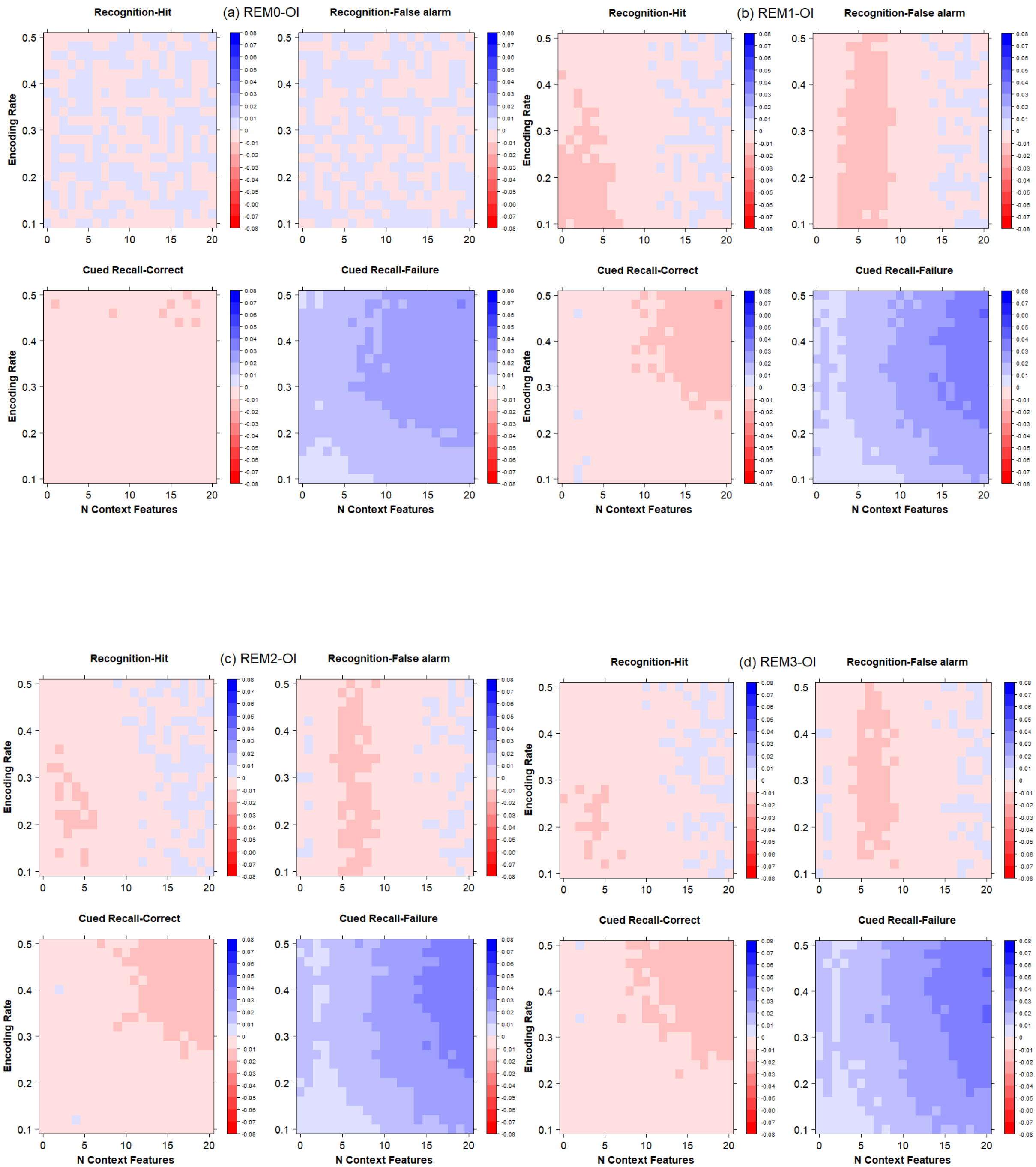


Figure 9

Qualitatively predicted output interference (OI) from REM0 (a), REM1 (b), REM2 (c), REM3 (d), and REM4 (e) simulations. It demonstrates whether OI is present under different models, different encoding rates (***u***) (y axis), and different numbers of context features ($\boldsymbol{Nf_c}$) (x axis). A red, white, and blue circle represents, respectively, a decreasing, unchanging, and increasing slope of that performance indicator.


(a) REM0-OI
Recognition-Hit
Recognition-False alarm
Cued Recall-Correct
Cued Recall-Failure
(b) REM1-OI
Recognition-Hit
Recognition-False alarm
Cued Recall-Correct
Cued Recall-Failure
(c) REM2-OI
Recognition-Hit
Recognition-False alarm
Cued Recall-Correct
Cued Recall-Failure
(d) REM3-OI
Recognition-Hit
Recognition-False alarm
Cued Recall-Correct
Cued Recall-Failure
Encoding Rate
N Context Features

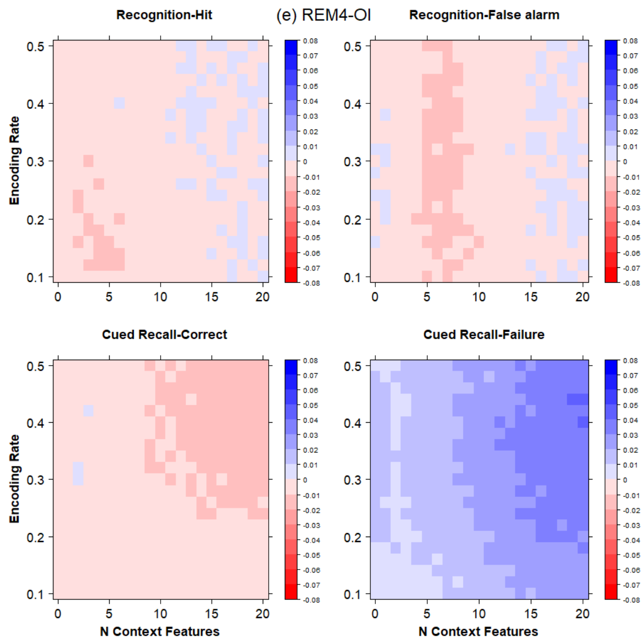


Figure 10

Quantitatively predicted output interference (OI) from REM0 (a), REM1 (b), REM2 (c), REM3 (d), and REM4 (e) simulations. It demonstrates the magnitude of the predicted OIs under different models, different encoding rates ($\boldsymbol{u}$) (y axis), and different numbers of context features ($\boldsymbol{Nf_c}$) (x axis). The redder, the more robust decreasing slope. The bluer, the more robust increasing slope.

Proactive Facilitation (PF) and Proactive Interference (PI)

When an item is repeated with two different associates, memory for one of the pairs is simultaneously improved (PF) and impaired (PI). For example, imagine a study with two study lists, list 1 and list 2 (as in Figure 4), each of which has 30 word pairs. Half of list 2 pairs (rearranged pairs, e.g., treat - point) are rearranged from list 1 (e.g., treat – chair, signal – point) whereas the other half pairs are new (novel pairs, e.g., castle – south). In cued recall, participants are asked to remember the word studied with the presented cue in the identified list (i.e., list 2). Every test trial may be categorized as a correct response, an incorrect response, or a response failure. In such an experiment, PF is observed as more correct responses for the rearranged pairs than the novel pairs and PI is observed as more incorrect responses for the rearranged pairs than the novel pairs (Aue et al., 2012; Aue et al., 2017; Burton, Lek, & Caplan, 2017; Wahlheim & Jacoby, 2013; Wahlheim et al., 2013). Figure 5b shows example data. In recognition, a similar study paradigm can be used testing memories of target words from a specified list (i.e., list 2) (Target-Rearranged and Target-Novel), foils from the other list (e.g., list 1) (Foil-List1) and foils from outside the experiment (Foil-New). The repeated targets (words from rearranged pairs) have higher hit rates than the novel targets. False alarms are higher for foils from list 1 than the new foils (as in Figure 5a).

Aue et al (2017) proposed and tested several mechanisms to account for PF in cued recall and concluded that the most plausible idea is an encoding advantage for the rearranged pairs. Various implementations suggested that the original REM model could not account for the data. This is one of the primary reasons that the Criss lab began considering a scaffolded encoding mechanism. Therefore, in this current investigation, the research question is whether implementing a scaffolded encoding mechanism within REM can account for PF and PI in recognition and cued recall.

In our simulations, as mentioned, there are two study lists of 30 word pairs each. List 2 pairs contain half rearranged pairs and half novel pairs, as in Figure 4. In recognition, a test list has 15 targets words that have been studied in both lists (Target-Rearranged), 15 targets words that have been studied only in list 2 (Target-Novel), 15 foils that have been studied in list 1 only (Foil-List1), and 15 foils that have not been studied in either list (Foil-New). Participants/ models judge if a test word is studied in the most recent list (i.e., list 2) and are provided with the options of “Yes” and “No”. In cued recall, we observe if participants/ models remember the targets studied with the cue from list 2.

From model simulations, we computed the memory differences between pair types as an indicator of whether and how much PF and PI are generated. In recognition, we subtracted proportion of responding old of Target-Novel from that of Target-Rearranged as a value for PF and subtracted the proportion of responding old of Foil-New from that of Foil-List1 as a value for PI. In cued recall, we subtracted proportion of correct responses of the novel pairs (i.e. Correct-Novel) from that of the rearranged pairs (i.e., Correct-Rearranged) as a value for PF and subtracted proportion of incorrect responses of the rearranged pairs (i.e., Incorrect-Rearranged) from that of the novel pairs (i.e., Incorrect-Novel) as a value for PI. Via hypothesis testing, we conducted t-test comparing these values against 0, with significant level (i.e., α) as 0.05. In Figure 11, a value indicating PF (better memory for items from repeated pairs) is displayed in blue and a value indicating PI (worse memory for items from repeated pairs) is displayed in red. A white dot indicates, under these parameter values, there isn’t a significant predicted PF or PI.

To evaluate the magnitude of the effect, we plotted the magnitude of the values with more blue indicating a bigger positive value and deeper red indicating more interference (see Figure 12).

Results and Discussions

From Figure 11 and 12, all REMs predict simultaneous PF and PI in recognition whereas only REMs with scaffolded encoding implemented at study (i.e., REM3 and 4) predict simultaneous PF and PI in cued recall, consistent with data (Aue et al., 2012; 2017) and the idea that item information and associative information are accessed differently (e.g., Criss, 2005; Hockley & Cristi, 1996a, 1996b). In addition, all predictions depend on parameter values. In the following paragraphs, we display and discuss simulation results of recognition and cued recall separately.

Recognition

Because whether and how much PF and PI in recognition are predicted vary little among different models (as in Figure 11 and 12), we discuss them together here.

PF and PI in recognition are predicted by all models (as in Figure 11). However, they are only predicted in noticeable magnitude (blue) when the number of context features is less than 10 in the first three models or 15 in the models with scaffolded encoding (as in Figure 12). In addition, the magnitude of PI increases as the encoding rate increases and as the number of context features decreases (as in Figure 12).

The reason why REMs, with or without additional encoding mechanism, account for simultaneous PF and PI in recognition is because a recognition decision is made based on global familiarity, or averaged activations of memory traces. Repeated items gain more activation regardless of how they are repeated and hence are more likely to be correctly recognized as old (i.e., higher hit rate) than the items that only appeared once. The same logic applies to higher false alarm rate for items that appeared on List 1 than items that are completely new (i.e., PI). A List 1 foil was studied in a somewhat similar context and are more likely to be incorrectly recognized as studied in List 2, compared to a new item that is dissimilar to any stored memory. This replicates the findings that both hit rate and false alarm rate are higher for repeated items in a paradigm with rearranged stimuli (e.g., Aue et al., 2012; Criss, 2005; Dyne et al., 1990; Kelly & Wixted, 2001; Overman & Becker, 2009; Rotello & Heit, 2000).

As for why predictions of PF and PI in recognition depend on the number context features (as in Figure 12), we argue it is due to context features being identical and the fact the we used a joint cue of context and item in recognition task. At test, as there are more context (identical) features stored in memories, memories are less differentiated with each other and any test word is more likely to be endorsed as old. Both hit rates and false alarm rates of both repeated items and unrepeated items raise to a point there is a decrease in magnitude for both PF and PI.

<u>Cued Recall</u>

Now we discuss the predicted PF and PI in cued recall.

Research related to interference among similar memories is well established (Anderson, 1974; Anderson & Neely, 1996; Postman & Underwood, 1973). When there are more than one response associated with the same cue, or multiple memories are encoded with similar information, memories compete to be accessed and retrieved (e.g., Barns & Underwood, 1959; Burns, 1989; Postman & Gray, 1977; Postman, Stark, & Burns, 1974). In the PF and PI experimental design, with rearranged pairs and two memories with overlapping item and possibly context, retrieval under a rearranged pair cue is destined to suffer more interference than retrieval under a novel pair cue. In addition, such interference comes from previously encoded memory with the shared item, evidenced by the analysis that an intrusion of rearranged pair cued recall is most likely to be its List 1 partner (Aue et al., 2012).

However, explanations for PF (e.g., extended retrieval search within SAM, Mensink & Raaijmakers, 1988; a flexible short-term decay within the Adaptive Control of Thought – Rational model, Thomson, Bennati, & Lebiere, 2014) are relatively poorly established, and the evidenced (e.g., Aue et al., 2017) verbal theory, recursive reminding (Hintzman, 2004; Wahlheim & Jacoby, 2013) has never been rigorously examined within computational memory models. Here, a scaffolded encoding mechanism, a computational implementation of recursive reminding, predicts PF and PI simultaneously in cued recall. Next, we discuss simulation results from models with and without scaffolded encoding.

REMs without scaffolded encoding (REM0, REM1, and REM2)

Similar, if not identical, PF and PI patterns in cued recall are generated by the REM without scaffolded encoding. PI is predicted for both correct and incorrect responses.

PI and the lack of PF can be explained by competing memories of list 1 and list 2 rearranged pairs. They are two memory traces having the same rearranged cue component. Memories of novel pairs, on the other hand, contain specific novel cue information. In a recall task, presentation of a rearranged cue induces more competition than a novel cue.

Within REM0 and REM1, information of pairs automatically encoded into memories per presentation of the pair. The products of studying two list are two equally weak memories for a rearranged pair that compete later at test and one weak memory for a novel pair. In REM2 where ***Update+Add*** is active at study, still no PF is predicted (see Figure 11c and 12c) because of the encoding operation on a semi-old event (as in Figure 6f). Presentation of a rearranged pair should be recognized as a semi-old event, because such pair has old items, yet their association is new.

According to the ***Update+Add*** mechanism, the corresponding old parts of memories are updated, and a new trace is added (as in Figure 6f). Note that this added trace is encoded and stored from nothing. That is, the stored memories for list 2 rearranged pairs within REM2 are just as week as those within REM0 and REM1. These models lacking a scaffolding operation can't predict PF.

REMs with scaffolded encoding (REM3 and REM4)

Implementing Update+Scaffold at study (i.e., REM 3) predicts PF in correct responses and PI in incorrect responses simultaneously (as in Figure 11d and 12d). Implementing Update+Scaffold at both study and test (i.e., REM4) predicts identical patterns (comparing Figure 11e and 12e to Figure 11d and 12d). Therefore, it is safe to conclude Update+Scaffold is essential and sufficient for simultaneous PF and PI in cued recall. The full model of scaffolded encoding (i.e., REM4) is proposed for having Update+Scaffold implemented at both study and test for consistency.

Update+Scaffold at study predicts PF, more correct responses for rearranged pairs than for novel pairs, because of its operation on a semi-old event. When a rearranged pair is presented, ideally, both items are recognized as old and from different memory traces, which is a semi-old event. Per scaffolded encoding, old memory traces are updated with features of this study event and a new trace is added with the already updated features (as in Figure 6g). When a novel pair is presented, it is recognized as a new pair with new items, which is a new event. Correspondingly, a new trace is added (as in Figure 6a). As a result, the scaffolded trace, or the memory trace of the rearranged pair from List 2, is more complete than the new trace, or the memory trace of the novel pair from List 2. Then, at test, when an item from the List 2 pair is presented as a retrieval cue, the memory trace of the rearranged pair is more likely to be sampled and recovered than the memory trace of the novel pair, producing PF (as in Figure 11d, 11e, 12d, and 12e).

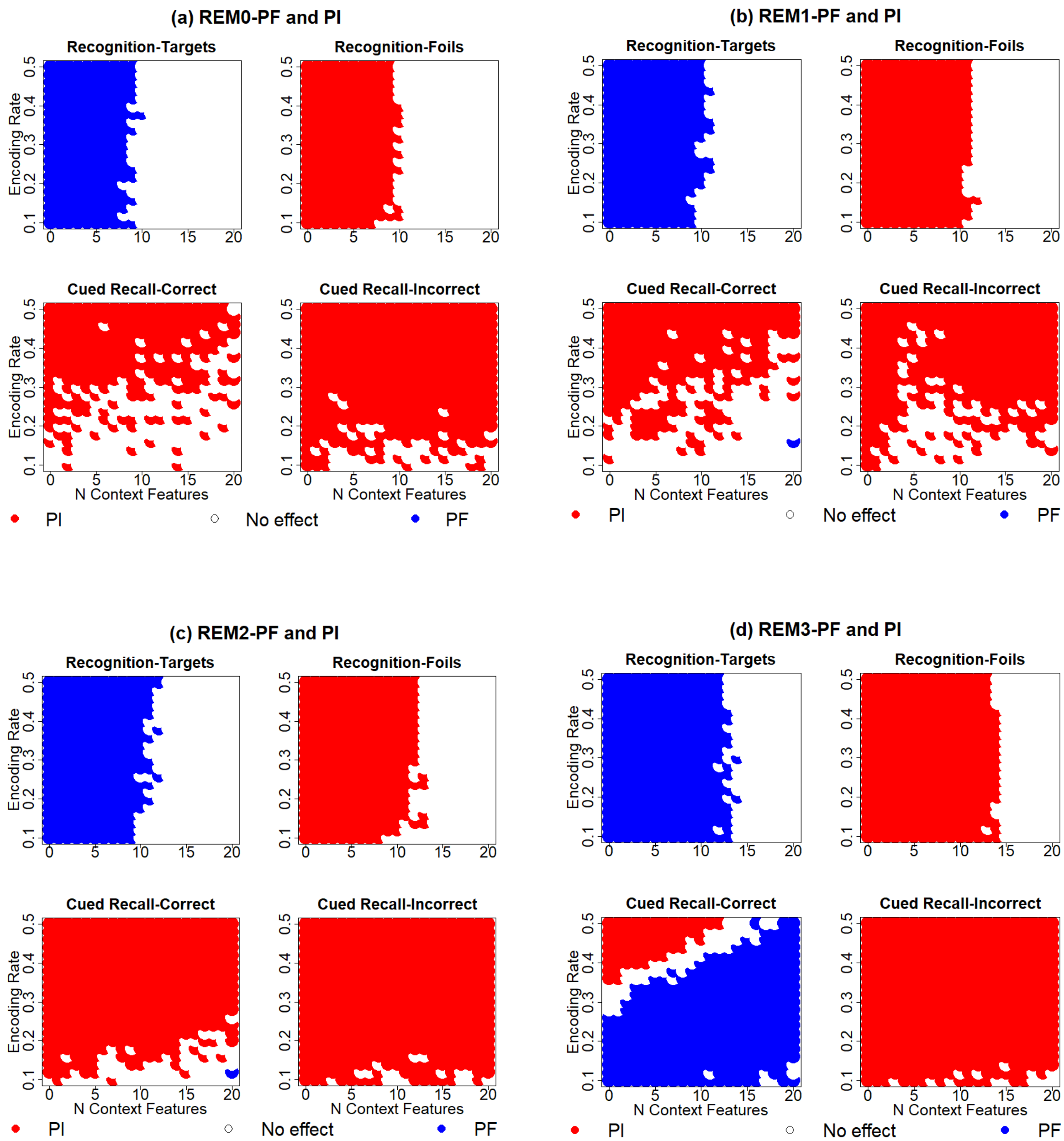
For the same reason, the predicted PF starts to show magnitude when encoding rate is high enough (i.e., 0.2) (as in Figure 12d and 12e). Because only when encoding quality is ensured from the beginning, rearranged pairs can be correctly detected as semi-old episodes and the Update+Scaffold mechanism functions at study to produce facilitation in correct responses later at test. We also see PF disappears when higher encoding rate is paired with less context features (as in Figure 11d, 11e, 12d, and 12e), and the predicted PF is more robust when higher encoding rate is paired with more context features (as in Figure 12d and 12e). Given PF equals more correct responses answering a cue from a rearranged pair than answering a cue from a novel pair, more robust PF equals sampling even more List 2 scaffolded memories than List 2 new memories. When there are more context features, memories have more features matching the retrieval cue, but, more to our point, the enhanced match is greater for List 2 scaffolded memories since they are more complete. Therefore, more encoded context features facilitate PF.

The idea of scaffolded encoding is consistent with Aue et al. (2017), who found evidence for that memories of rearranged pairs in List 2 are stronger, most likely from remembering old memories in List 1 during List 2 studying (i.e., study-phase retrieval). In their Experiment 3, when cued recall was conducted on List 1 (i.e., please remember the word presented with the cue in the first list), instead of List 2, facilitation was extended to the rearranged pairs in List 1 (i.e., retroactive facilitation). This data suggests the necessity of an encoding mechanism (i.e., scaffolded encoding) which ensures rearranged pairs' traces from List 2 are equally strong as the old ones from List 1 for both proactive and retroactive facilitation to happen. The idea of scaffolded encoding is also consistent with the recursive reminding concept (Hintzman, 2004, 2010). As an item is reencountered and it leads to retrieval of its previous encounter, this previous encounter's information is incorporated into this present encounter such that this new memory contains information of both events. One may think of scaffolded encoding at study embedded in REM is a model implementation of this concept.

Lastly, one should note the sources of PF reflected in correct responses and PI reflected in incorrect responses are different. The former is the scaffolded memories outcompeting the novel memories, coming from the scaffolded encoding mechanism implemented at study whereas the latter is the updated and old memories of List 1outcompeting the scaffolded memories of List 2, coming from the inherent competitive nature of memory retrieval (Anderson & Neely, 1996). This echoes with Aue et al. (2017) who proposed PI is not produced by the same mechanism as PF. In Aue et al. (2017)'s Experiment 5, when there are consecutive blocks of study and test, and novel pairs in the previous trial became rearranged pairs in the next trial, PF was observed whereas PI was absent.

Summary

The source of both PF and PI in recognition is more averaged activation from repeated items. PF in cued recall comes from scaffolded encoding at study and PI in cued recall is a product of competition inherent to retrieval. Our proposed model of scaffolded encoding (i.e., REM4) successfully accounts for PF in cued recall while no other model does.



(a) REM0-PF and PI
Recognition-Targets
Recognition-Foils
Cued Recall-Correct
Cued Recall-Incorrect
Encoding Rate
N Context Features
PI
No effect
PF
(b) REM1-PF and PI
Recognition-Targets
Recognition-Foils
Cued Recall-Correct
Cued Recall-Incorrect
Encoding Rate
N Context Features
PI
No effect
PF
(c) REM2-PF and PI
Recognition-Targets
Recognition-Foils
Cued Recall-Correct
Cued Recall-Incorrect
Encoding Rate
N Context Features
PI
No effect
PF
(d) REM3-PF and PI
Recognition-Targets
Recognition-Foils
Cued Recall-Correct
Cued Recall-Incorrect
Encoding Rate
N Context Features
PI
No effect
PF

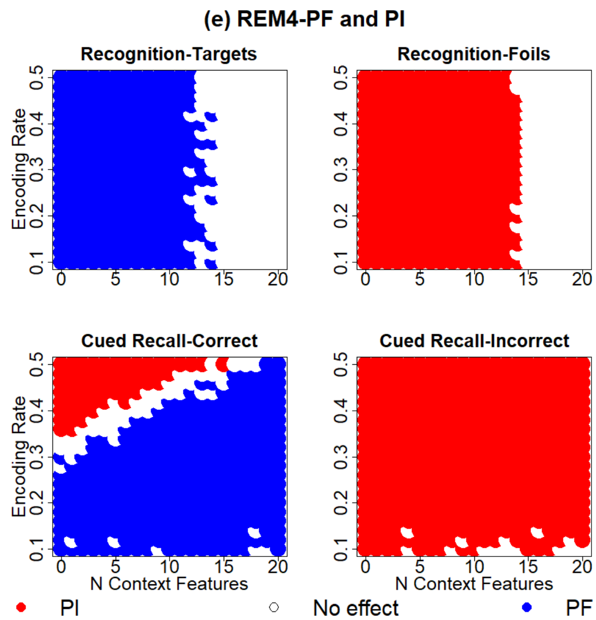
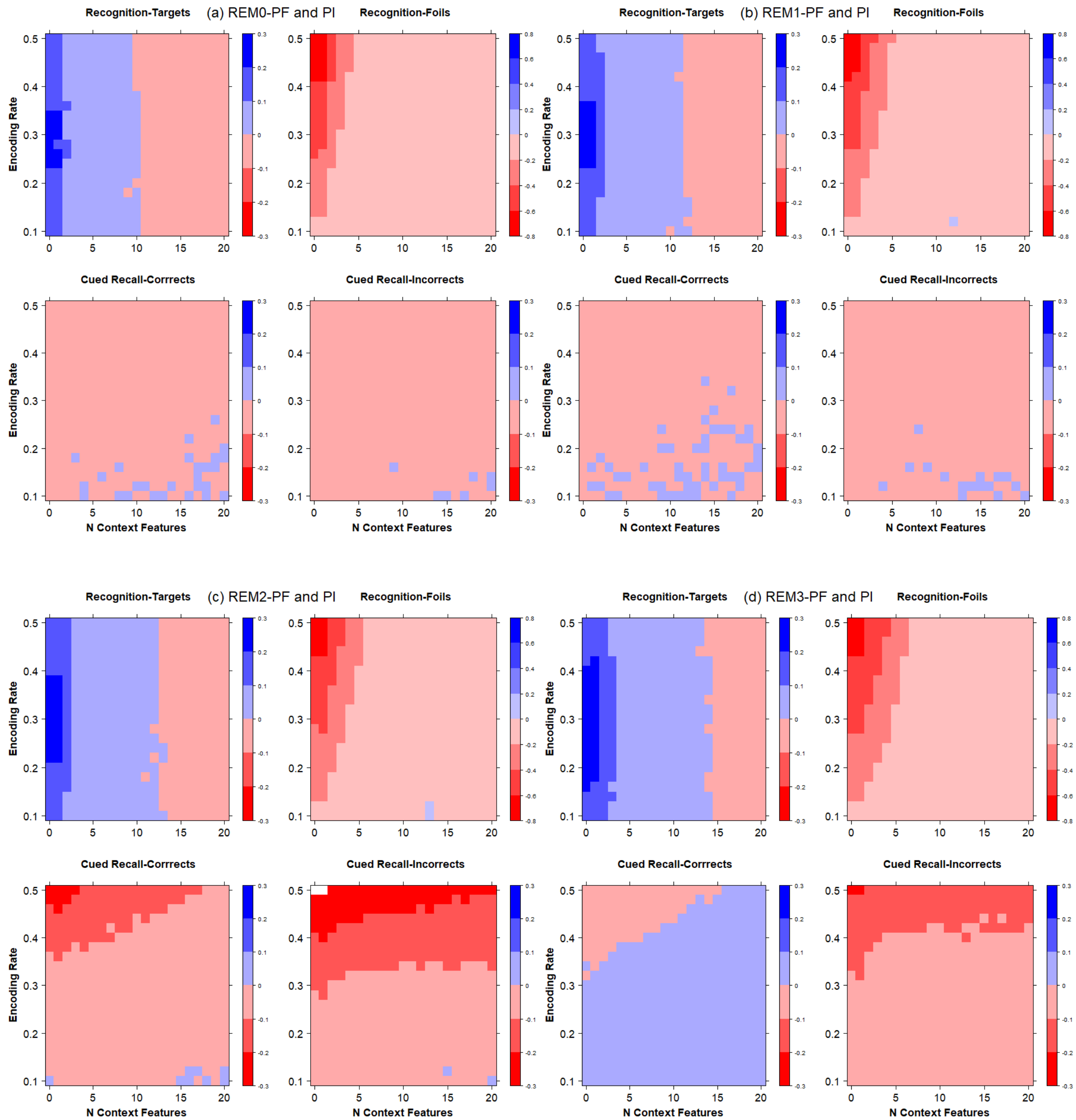


Figure 11

Qualitatively predicted proactive facilitation (PF) and proactive interference (PI) from REM0 (a), REM1 (b), REM2 (c), REM3 (d), and REM4 (e) simulations. It demonstrates whether PF and/ or PI is present under different models, different encoding rates (***u***) (y axis), and different numbers of context features (Nf_c) (x axis). A red and blue circle represents, respectively, a PI and PF. A white circle means either PF or PI is present.


(a) REM0-PF and PI
(b) REM1-PF and PI
(c) REM2-PF and PI
(d) REM3-PF and PI
Recognition-Targets
Recognition-Foils
Cued Recall-Corrrects
Cued Recall-Incorrects
Encoding Rate
N Context Features

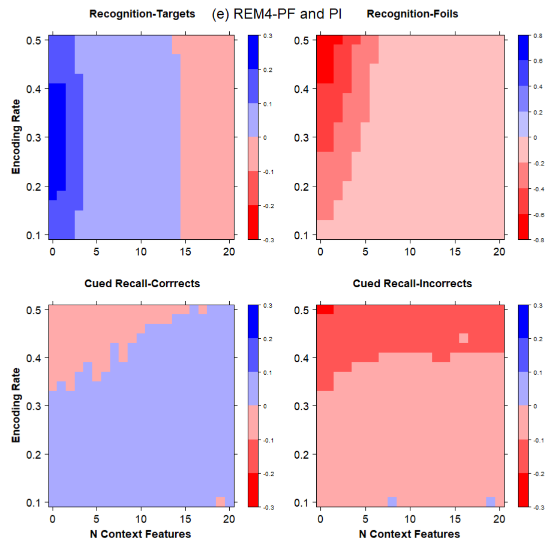


Figure 12

Quantitatively predicted proactive facilitation (PF) and proactive interference (PI) from REM0 (a), REM1 (b), REM2 (c), REM3 (d), and REM4 (e) simulations. It demonstrates the magnitude of the predicted PF and/ or PI under different models, different encoding rates ($\boldsymbol{u}$) (y axis), and different numbers of context features ($\boldsymbol{Nf_c}$) (x axis). The redder, the more robust PI. The bluer, the more robust PF.

General Discussion

In this current paper, we proposed and comprehensively tested a unified model of memory, the model of scaffolded encoding. Our simulation results show that our model, although was initially designed to account for cued recall data and mainly for PF, accounts for multiple memory phenomenon of repetition in both recall and recognition.

Like previous models (e.g., Criss et al., 2011), our model assumes that an old memory trace is updated (i.e., differentiation, Criss & Koop, 2015; Kilic et al., 2017) corresponding to an old episode and a new trace is added corresponding to a new episode. Unlike previous models, our model allows an episode to be semi-old and includes the scaffolded encoding mechanism. A semi-old event may be studying old word or old pairs under a new context, studying a rearranged pair, or retrieval under a new context. That is, a semi-old event has both old and new episodic components. Then, the old memory trace is updated, corresponding to the old components, and a new trace is added with the updated features from the old memory trace, corresponding to the new components.

Reiterating our findings, our model, REM with the scaffolded encoding mechanism implemented at both study and test (i.e., REM4), accounts for LSEs in recognition, cued recall, and free recall (see Figure 13a), OIs in recognition and cued recall (see Figure 13b), as well as PF and PI in recognition and cued recall (see Figure 13c). A negative LSE in recognition is predicted because repeated information is accumulated in one memory (i.e., differentiation) and provides more evidence when a recognition decision is made. A null LSE and a positive LSE in recall are predicted for, respectively, moderate and high level of competition at sampling. OIs in recognition are produced by encoding happening at test, ***Update+Add*** or Update+Scaffold. OIs in cued recall are produced by an encoding process as well as a retrieval filter at test. PF and PI in recognition are given by more activations from memories of repeated items. The source of PF in cued recall is storing memories of the rearranged pairs with already updated features at study (i.e., scaffolded encoding) and the source of PI in cued recall is retrieval competition inherent to memory search. In short, mechanisms prior to our model are sufficient to explain LSEs, OIs, PF and PI in recognition, and PI in cued recall, however, the scaffolded encoding mechanism is essential to PF in cued recall.

Nevertheless, our model isn't perfect. There are several limitations worth discussing. We also discuss its broader implication and future research.

Scaffolded Encoding and Context

The first and foremost limitation regards the role of context in episodic memory and within our model. As we examine what combinations of parameter values produce all the mentioned effects under the model of scaffolded encoding and to compare the difference between recognition and recall task, we found that there are limited sets of parameter values predicting all the memory effects in both recognition and recall task (see Figure 14).

It is obvious that predictions in recognition are subject to the number of context features or require context features that are less than in recall. As we averaged list performance from various indicators from different paradigms in different retrieval tasks implemented within the model of scaffolded encoding (see Figure 15), we see the common theme that, in recognition task, as the number of context features increases, false alarm rate raises quickly towards ceiling, making model predictions under these parameter values uninformative or inapplicable to data. We have argued repeatedly in previous sections that this is because context is identical among memories and our choice of including context to make a recognition decision, unlike the original study (Shiffrin & Steyvers, 1997). As the number of context features increases, the match between stored memories and a context cue, likelihood ratio calculated based on Equation 2, increases exponentially (as in Figure 16a) given an encoding rate of 0.1. It makes using a joint cue to probe memory produce too much evidence for match (see Figure 16b). In this sense, increasing the number of context features is increasing memory evidence for match and increasing "old" responses but without merit.

However, we don't consider our current model is concerning due to its limited parameter space. Compared to a working model with ample combinations of parameter values that may be unfalsifiable (Popper, 1934), our model is robust for predicting data and generalizing to multiple paradigms and tasks (Lee et al., 2019).

Regardless, we intend to discuss several amendments to make model predictions less sensitive to context features and to enhance the model's applicability to recognition and recall. They are 1) implementing a context drift at study, 2) partitioning the retrieval process into two phases (Shiffrin & Steyvers, 1997), context-dependent activation and item-dependent recognition or sampling and recovery, 3) diluting memory evidence of a matching feature (Shiffrin & Steyvers, 1997), and 4) weighting (Criss & Shiffrin, 2004b) calculated likelihood ratio from context. Among these potential solutions, the first one aims to make context less similar whereas the rest aim to downplay the effects from identical context via various quantitative methods. We emphasize that these adjustments to the model are merely ancillary modeling assumptions, whereas the scaffolded encoding mechanism is the core modeling assumption and is the key accounting for data.

Context Drift

Let's first discuss the possibility that context drifts, as we have assumed that context doesn't change from trial to trail, or between study and test. May we clarify that keeping context constant is, as a matter of fact, a standard practice in REM modeling (e.g., Malmberg & Shiffrin, 2005). More to our primary interest of having a viable model that qualitatively explains a series of observations, drifted context and its role in memory were beyond our concerns. Nevertheless, we admit fixing context contradicts the idea that context varies slowly and randomly within a study list (Estes, 1955; Howard &Kahana, 2002; Lohnas & Kahana, 2014; Mensink & Raaijmakers, 1988; Polyn, Norman, & Kahana, 2009; Shiffrin & Steyvers, 1997) and between study and test (Karpicke, Lehman, & Aue, 2014; Murnane & Phelps, 1994; Polyn et al., 2009; Sederberg, Howard, & Kahana, 2008).

Now that we have the model of scaffolded encoding as the winner model and are perplexed by constant context features, we further test our model of scaffolded encoding with added auxiliary assumptions regarding context change. We assumed, at study, that there is a random context drift (parameter $\boldsymbol{d_1}$) from trial to trial, referring to the probability a context feature value changes. Such context drift may arise from temporal factors (Howard & Kahana, 1999, 2002), encoding tasks or semantic organization of the list (Polyn et al., 2009a, 2009b), or other random variations (Estes, 1955; Mensink & Raaijmakers, 1988). When a context feature value changes, it is filled by a randomly generated value from the geometric (i.e., $\boldsymbol{g}$) distribution (Shiffrin & Steyvers, 1997). At test, a context cue is composed from drifted study contexts. Specifically, first, a prototype context is composed with each feature value being the most frequent feature value from all the study contexts. That is, the constructed prototype test context has feature values overlapping with most of the studied memories within a specific study list. This process represents context reinstatement, that an internally maintained (Howard & Kahana, 1999, 2002) and a broad sense of gestalt-like context is created or retrieved by the participants (Dennis & Humphreys, 2001; Lehman & Malmberg, 2013). We use similar logic and consider a potential context drift (parameter $\boldsymbol{d_2}$) from study to test, referring to the probability a context feature value from the prototype context changes and is replaced by a random value generated by the geometric (i.e., $\boldsymbol{g}$) distribution. Such context drift may be time-dependent, given there is a time lag between study and test and a delay may be implemented, and task-dependent, given instructions are different between study and test. These probabilistic drifted context features make up a test context which is used to probe memory within a test block. We note that implementing a context drift at study (i.e., $\boldsymbol{d_1}$) has the potential to explain the spacing effect (Lohnas & Kahana, 2014; Malmberg & Shiffrin 2005) and implementing a context drift between study and test (i.e., $\boldsymbol{d_2}$) has the potential to explain the testing effect (e.g., Karpicke et al., 2014), which we will elaborate later.

Because context changes slowly during study, we assign $\boldsymbol{d_1}$ the value of 0.008, following Shiffrin and Steyvers (1997). To directly observe and only observe the impact of drifted contexts at study on memory, we assign $\boldsymbol{d_2}$ the value of 0 since context isn't changed drastically from study to test (e.g., Smith & Verla, 2001). We implemented context drift within the model of scaffolded encoding (i.e., REM4) and simulated 1000 subjects for each combination of parameter values, as the former formal modeling method. We expected a drifted context makes memories less confusable and makes model predictions in recognition less constrained by the number of context features. More importantly, context drift makes our model more applicable to both recognition and recall.

For our interest, we drew the combinations of parameters that quantitatively produce all the memory effects in recognition, recall, and both (see Figure 17a). Comparing Figure 14 and Figure 17a, although model predictions in recognition are indeed less constrained by the number of context features, as expected, we replicated that less context features (i.e., less than 15) are required in recognition than in recall. In addition, there are merely slightly more possible combinations of encoding rate and number of context features to predict the critical data patterns simultaneously in both recognition and recall (comparing Figure 17a to Figure 14). This is a sensible result given context drifting at the rate of 0.008 is not that distinct from identical contexts.

The added assumptions of context drift are not critical to our model of scaffolded encoding. Without explicit instructions or manipulations, it is unlikely for contexts to be different enough such that memory evidence provided by context features contributes to retrieval in the same fashion as item. Next, one try to attenuate context's contribution to retrieval.

Two-Phase Retrieval

Potential problems caused by identical contexts were foreseen by Shiffrin and Steyvers (1997). In "REM.4" of the original REM paper, wanting to separate within-list memories from extralist memories, Shiffrin and Steyvers included 40 context features and raised the concern that matching context might pollute memory evidence and falsely contribute to an "old" response in recognition. Therefore, they divided the recognition process into two components, context-based activation of memories and item-based recognition. Specifically, the first step of retrieval is using context to activate memories. The match between the context cue and context part of the memories are calculated based on Equation 2 and the memories whose calculated likelihood higher than an activation threshold (i.e., parameter $\boldsymbol{\tau}$) are activated. Then, the second step of retrieval is using item to make a recognition decision. The match between the item cue and stored item features of the activated memories was calculated based on Equation 2, and an averaged activation is computed, according to Equation 3, and compared to a criterion (i.e., $\boldsymbol{E}$) of 1. If it exceeds this criterion, an "old" response is produced. Otherwise, a "new" response is produced. In this way, only item features that are distinct among memories contribute to making a recognition decision.

However, in our case where the number of context features is varied and, in turn, the likelihood ratio of context is varied drastically (as in Figure 16a), a two-phase retrieval is inapplicable to our modeling exploration. Ideally, in the activation phase, most but not all memories within the study list the context cue is designated for are activated and are certified to contribute to the second phase of retrieval. The value of the activation threshold, $\boldsymbol{\tau}$, determines how many memories can be activated. If it is too small, all memories, even those from another study list, are activated. If it is too large, few memories within the list can be activated and memory performance is worsened. It means the number of context features has to be taken into account when it comes to the choice of $\boldsymbol{\tau}$ value. In the original paper where 40 context features were used, a threshold value (i.e., $\boldsymbol{\tau}$) of 2980.9 was adopted. In this current paper where we vary the number of context features from 0 to 20, the threshold value must also be varied such that there is a rightful amount of activated memory traces. However, what this variation may be is unfound in literatures and previous modeling documentations. For those who are interested in knowing what happens if we simply adopt the same $\boldsymbol{\tau}$ value (i.e., 2980.9), memory performances were all at floor because few memories exceed this threshold supposedly meant for 40 context features. Therefore, we decide to forego this amendment.

<u>Diluted Likelihood Ratio</u>

In addition to two-phase retrieval, Shiffrin and Steyvers (1997) proposed a conceptually similar approach in their Appendix B. Via this approach, in recognition and cued recall, the retrieval process consists of one phase of a joint cue (context + item) probing memory, but the way likelihood ratio is calculated is different such that the evidence for a match between the retrieval cue and stored memories is diluted. Superior to the two-phase retrieval approach and more suitable to our modeling exploration, the level of dilution can be adjusted accordingly to the number of context features ($\boldsymbol{Nf_c}$). Now, given a cue $\boldsymbol{q}$, every memory trace is assigned a likelihood ratio $\boldsymbol{\lambda}$, calculated according to Equation 7. Compared to Equation 2, Equation 7 has a new parameter $\boldsymbol{b}$, calculated according to Equation 8, where $\boldsymbol{Nf_c}$ is varied from 0 to 20 by 1 and $\boldsymbol{Nf_i}$ is fixed as 20.

$$\boldsymbol{\lambda_j = \left[\frac{1-c}{1-bc}\right]^{n_{ij}} \prod_{\forall i}^{\infty} \left[\frac{c+(1-c)g_{sys}\left(1-g_{sys}\right)^{v-1}}{bc+(1-bc)g_{sys}\left(1-g_{sys}\right)^{v-1}}\right]^{n_{ijm}}} \qquad (\boldsymbol{E7})$$

$$\boldsymbol{b = \frac{Nf_c}{(Nf_c + Nf_i)}} \qquad (\boldsymbol{E8})$$

In this form of calculating likelihood ratio and evaluating evidence for a match between a retrieval cue and memories, the more the number of context features there are, the less credits we give to a matching feature.

We applied this calculation of likelihood ratio to the model of scaffolded encoding (i.e., REM4) and simulated 1000 subjects for each combination of parameter values. With diluted likelihood ratio (see Figure 16c), discrediting memory evidence for a matching feature in proportion to the number of context features, we expected to see less restricted predictions of memory effects in recognition by identical contexts and increased number of combinations of parameters that quantitatively predict the critical memory phenomenon in recognition, recall, and both.

Modeling findings match our expectation (see Figure 17b). With diluted likelihood ratio, predicting the memory effects in both recognition and recall is possible even when the number of context feature is large (e.g., 20) (comparing Figure 14 and 17b). Compared to the context drift approach that supposedly improve model's applicable to task by making context less similar, the current approach is more effective (comparing Figure 17a and 17b).

However, this approach only applies to recognition and cued recall. In free recall, likelihood ratio is computed the old way (i.e., Equation 2), because when the cue is context only, $\boldsymbol{b}$ takes the value of 1 and the likelihood ratio calculated from Equation 7 is not diagnostic. It was neglected given the original REM model (Shiffrin & Steyvers, 1997) wasn't applied to free recall. Moreover, this approach underestimated evidence of a matching feature of both context and item, whereas the main issue resides in matching context features. Memory evidence provided by distinct item features is informative and shouldn't be discredited. Doing so may affect model predictions in effects determined by item similarity. This brings us to our final potential adjustments of the model that targets at likelihood ratio computation of contexts (Criss & Shiffrin, 2004a).

Weighted Context

Criss and Shiffrin (2004a) weighted context and item differently. After the likelihood ratio between context part of a retrieval cue and a memory trace $\boldsymbol{j}$ (i.e., $\boldsymbol{\lambda_{jC}}$) and that between context part of a retrieval cue and memory $\boldsymbol{j}$ (i.e., $\boldsymbol{\lambda_{jI}}$) are computed, weights, respectively, 1-$\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}$, are assigned to them. Adding weighted context and weighted item, according to Equation 9, outputs the likelihood ratio of the match between the retrieval cue and this memory (i.e., $\boldsymbol{\lambda_j}$).

$$\boldsymbol{\lambda_j = \left[\alpha {\lambda_{jI}}^{-1} + (1-\alpha){\lambda_{jC}}^{-1}\right]^{-1}} \quad (\boldsymbol{E9})$$

In this way, in recognition and cued recall, memory evidence from identical context features contributes less than distinct item features to the likelihood ratio used to make a recognition decision or sample a memory trace. Averaged likelihood ratios computed this way are no longer skewed (see Figure 16d). In free recall, given a retrieval cue is made of pure context features, a full weight of 1 is assigned to context. That is, the likelihood ratio in free recall is computed the old way as in Equation 2.

Compared to the diluted likelihood ratio approach that underweights memory evidence for match from both context and item, the current approach that lasers at context features is more logical. In addition, unlike the diluted likelihood ratio approach that only applies to recognition and cued recall where a joint cue of context and item is utilized, differently weighting information within a retrieval cue applies to all retrieval tasks. The idea that there are different weightings of context and item was also implemented within SAM with a differentiation mechanism and advocated as the source of LSE in memory (Shiffrin et al., 1990). They argued that as the weighting of context vs. item within a retrieval cue shifts from context to item in recognition, cued recall and free recall, mean and variance of memory activation increases, predicting LSEs ranging from negative, null to positive in respective retrieval tasks.

However, we are concerned about, like with the two-phase retrieval approach, fixating the weighting on item vs. context (i.e., $\boldsymbol{\alpha}$) contradicts our varying number of context features, as the averaged likelihood ratio of a joint cue with weighted context changes noticeably (as in Figure 16d). Nevertheless, we assigned $\boldsymbol{\alpha}$ the value of 0.6 as Criss and Shiffrin (2004b) who had 20 context features in their implementations and expected the model of scaffolded encoding with weighted context generate proper predictions in both recognition and recall at least when the number of context features is close to 20.

As before, we varied encoding rate and the number of context features, within the model of scaffolded encoding (i.e., REM4) with 1000 simulated subjects. To evaluate model applicability to recognition and recall directly, we displayed where in the parameter space the model quantitatively generates the predictions in recognition, recall, and both (see Figure 17c). We found, contrary to the standard model (see Figure 14), model predictions of effects are absent in recognition when there are less than 10 context features (see Figure 17c). Matching our expectations, however, the model with weighted context performs well when there are more than 10 context features, predicting the effects simultaneously in recognition and recall (see Figure 17c). This result validates the effectiveness of the weighting approach but also backs up our prior concern about fixated $\boldsymbol{\alpha}$ value. With a constant value of 0.6, adopted from Criss and Shiffrin (2004b) with 20 context features, a joint cue with context features less than 10 can't produce an averaged likelihood ratio more than 1 (see Figure 16d), affecting recognition performance and preventing the model to predict the effects. One could vary the weighting parameter accordingly to the number of context features such that the model predicts the effects at the same level across different number of context features, but its technicalities are unknown and won't be discussed further here.

In summary, the fact that the context component of a cue matches all memories and therefore provides compromised evidence during retrieval has made our model predictions confined to a small number of context features. Proposing and testing several approaches, either making context a bit more distinguishable or downplaying its role during retrieval, we found they adjusted our model to be more applicable to recognition and recall to various extent and with various limitations.

Nevertheless, we are not stating the formal model, the model of scaffolded encoding, is an inferior model. Firstly, comparing our formal model (i.e., REM4) to the models with added parameter and additional assumptions, even though the latter are more generous at predicting the desired memory effects with different number of context features, the former, the formal model, is more parsimonious and robust. Secondly, having a small number of context features is conceptually different but practically similar for the model to having a large number of context features but underweighted. In an experiment, it is difficult to test if participants encode as many contexts as items but don't use them to their full potential during retrieval or encode fewer contexts and use them fully during retrieval. It may depend on task demands and retrieval strategies differently and flexibly used by participants (e.g., Malmberg & Xu, 2007). The role of context in episodic memory is far more complicated than we discuss and implement here.

Scaffolded Encoding and Models for Context

As demonstrated, even without assumptions regarding context drift or downplaying the role of context during retrieval, the model of scaffolded encoding performs well at predicting the benchmark memory findings we pointed out at the beginning of this paper. Nevertheless, context is more substantially critical for other models and theories.

In recognition, if we think about two extremes where item or context solely contributes to making a recognition decision, there are item-noise models and context-noise models (Criss & Shiffrin, 2004b). One example of the former is the template of our model, the original REM model (Shiffrin & Steyvers, 1997), where context, if present, is used to active memories whereas items majorly contribute to and interfere the retrieval process. One context-noise example is the bind cue decide model of episodic memory (BCDMEM, Dennis & Humphrey, 2001) with the strong claim that the only interference to memory is context information. Every time an item is encountered, prior to and during experiment, it is associated with context. At test, as an item is presented, it triggers retrieval of its previously associated contexts. Then, a recognition decision is made based on the match between the current test context and the retrieved contexts. Since other items in the same list don't contribute to recognition at all, strength of other items doesn't matter and hence BCDMEM naturally accounts for a null LSE in recognition. With the added assumption that context drifts and the match between the test context and the retrieved contexts decreases across test trials (Criss et al., 2011), it accounts for OI in recognition. To account for PI and PF, though never implemented, it could be argued that repeated items have more and different associated contexts, causing more interference in recognition (Fox, Dennis, & Osth, 2020) as well as more likely matching the current context by chance. However, one must admit eliminating the role of item in any episodic memory tasks is exceedingly radical (Criss et al., 2011; Criss & Howards, 2015) and the idea that context and item jointly interfere retrieval is more reasonable and supported by behavioral evidences (Criss & Shiffrin, 2004b; Fox et al., 2020).

In free recall, with the absence of an explicit cue, context is impossible to omit. Critical computational problems surrounding free recall (Criss & Howard, 2015) are how recall is initiated and how recall continues, which are tackled by the class of retrieved context models (Howard & Kahana, 1999, 2002). Unlike the model of scaffolded encoding, they have detailed and mechanic assumptions regarding context and are specified for context-sensitive effects from free recall. According to this class of models, every item is associated with a history of slowly drifting temporal context. At test, the current context is used as cue to initiate recall, therefore predicting increased probability of recalling items close to the end of the study list (i.e., the recency effect). When this item is repeated or recalled, its pre-existing contexts are retrieved and used as cue to continue recall, therefore predicting the tendency to recall neighboring items (i.e., the contiguity effect). The models' predicting free recall dynamics is the model of scaffolded encoding can't outperform.

Nevertheless, with the context drift adjustment mentioned earlier, REM with scaffolded encoding may account for the spacing effect, as the retrieved context models (Lohnas & Kahana, 2014). The spacing effect refers to the memory phenomenon that memory is improved for items that are repeated in a spaced fashion, compared to items that are repeated in a massed fashion (Melton, 1970). Within the framework of retrieved context models, given context is retrieved when an item is repeated, an item of massed repetition is associated with similar contexts whereas an item of spaced repetition is associated with different contexts (Polyn et al., 2009). Then, at test, as the current test context is used for recall, when the associated contexts are more distinct, there are additional retrieval paths to the item (Lohnas, Polyn, & Kahana, 2011), increasing the probability of recalling the item repeated in either presentation position (i.e., the spacing effect) (Lohnas & Kahana, 2014).

Consider how one anticipate the model of scaffolded encoding to account for the spacing effect with implementation of a context drift at study (i.e., $\boldsymbol{d_1}$). On one hand, when an item that is repeated in a presentation position close to the previous one (i.e., massed repetition), context hasn't drifted much and the current study episode is decided by the model to be old, therefore the old memory trace is updated. On the other hand, when an item that is repeated in a further presentation position (i.e., spaced repetition), context has drifted enough such that the model categorizes the current event as semi-old. Correspondingly, the old memory trace is updated, and a new trace is scaffolded just as completed as the old one. Then, at test, memory of the item of spaced repetition, having two strong traces, is more accessible than that of the item of massed repetition, having one strong trace. This logic echoes with the additional retrieval paths proposed by Lohnas and colleagues (2011, 2014).

Similar rationales can also be applied to the testing effect (Tulving, 1967). Visiting literatures of testing effect, it's easy to notice the reappearance of the role of context (e.g., an episodic context account, Karpicke et al., 2014). However, there has been few computational memory models (Hopper & Huber with SAM framework, 2016, 2018; Lohnas and colleagues with the retrieved context framework, in progress). With a context drift implemented between study and test (i.e., $\boldsymbol{d_2}$), our model has the potential to account for the testing effect. At test, the test context, used as retrieval cue and having shifted from study, is recognized as new and the current event is categorized as a semi-old event by the model. Correspondingly, the recovered memory trace is updated with the current test context features and the recovered content, and a new trace is scaffolded to memory, providing additional retrieval routes to memory in the future.

In general, the mentioned models emphasize the irreplaceable role of context in episodic memory and excel at other aspects our model isn't designed for. The model of scaffolded encoding could learn from them and take context-dependent effects into consideration to be more profuse.

Scaffolded Encoding and Reconsolidation

Considering broader implication of this current paper, we are interested in conceptually compatible ideas that have been proposed and discussed in a line of research in the field of neuroscience, consolidation and reconsolidation.

Consolidation refers to the process where a newly formed and labile memory becomes stabilized. This process involves hippocampus and cerebral cortex (McGaugh, 2000), from the earliest study linking damage in hippocampal region and retrograde amnesia (Scoville & Milner, 1957) to more recent studies (e.g., Takashima et al., 2009) measuring functional connectivity migrating from hippocampus-cortex to cortex-cortex as retrieval happens longer after learning.

Reconsolidation (Misanin, Miller, & Lewis, 1968) refers to the process that, as a presented reminder or a new related experience causes memory retrieval, an initially consolidated memory is reactivated and modified with new learning (McKenzie & Eichenbaum, 2011). It is considered that reconsolidation is simply an extended process of consolidation (Dudai & Eisenberg, 2004; McKenzie & Eichenbaum, 2011).

Reconsolidation and scaffolded encoding are in common and relating them has the potential to bridge memory research in mathematically psychology and neuroscience. As Marr (1982) pointed out, there are three levels to understand any information-processing machine (e.g., human cognition), abstract, algorithmic, and physical. It is entirely possible that reconsolidation and scaffolded encoding are conceptually identical in the abstract level, but the former is actualized in the synaptic or physical level whereas the latter is described in the computational or algorithmic level. Therefore, unifying them is theoretically meaningful and has the potential to advance our understanding of human memory.

We are not the first aiming for such unified framework. It was long assumed that reconsolidation strengthens reactivated memories (Gordon, 1981). Reconsolidation following retrieval has also been used to explain behavioral findings related to the malleability of memory (Hardt, Einarsson, & Nader, 2010). In their opinions, new information initiates memory retrieval, which then initiates memory modifications, causing memory interference. Hupbach, Gomez, and Nadel (2009), with a list-learning paradigm, manipulated that participants were reminded, or not, the studying of List 1 prior to studying List 2, with the same global context (i.e., experimenter and room). Then, a final memory test took place either immediately after List 2 or one day after. They found that intrusions from List 2 during List 1 recall were substantially more than intrusions from List 1 during List 2 recall (i.e., asymmetric intrusion effect), when the test was delayed. It was argued that the reminder procedure made the stored List 1 memories malleable and susceptible to updating with information of the List 2 items. It was also argued that such interference from reconsolidation happens only after a delay because its corresponding mechanism at the molecular level happens slowly, matching results from rodent studies (Nader, Schafe, & Doux, 2000).

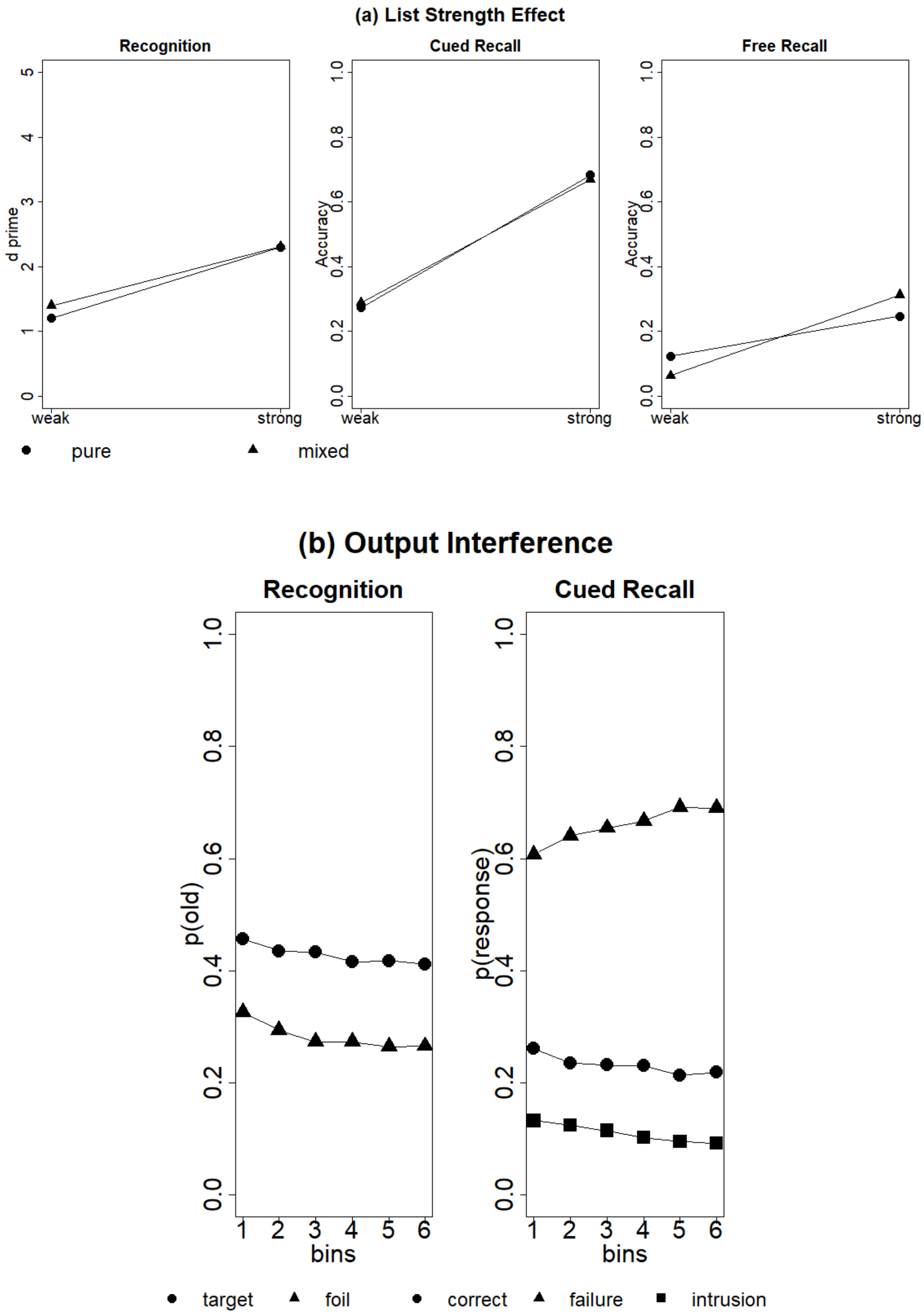
More interestingly to us, a computational memory model, temporal context model of memory (TCM, Howard & Kahana, 2002) with context reinstatement, context-item binding, and a recall-to-reject strategy (Hintzman, Caulton, & Levitin, 1998), accounts for the data of Hupbach et al. (2009) (Sederberg, Gershman, Polyn, & Norman, 2011). According to TCM, during List 1 studying, List 1 items are associated with List 1 context. In the reminder conditions, List 1 context is reinstated, and List 2 studying makes List 2 items are associated with both List 1 and List 2 contexts. Then, at test during List 1 recall, a retrieval cue made of List 1 context activates memories of both List 1 and List 2 items, whereas, during List 2 recall, a retrieval cue made of List 2 context can only activates memories of List 2 items. This accounts for the asymmetric intrusion effect. In addition, it was assumed that participants retrieve context, compare it to the current test context, and reject it if they are too similar, accounting for the absence of the asymmetric intrusion effect when the test is immediate.

The mentioned work are admirable efforts unifying memory research from two different realms. Given our model is capable of explaining various of effects of repetition in different retrieval tasks, the scaffolded encoding mechanism has great potential to contribute to this endeavor. Although reconsolidation has been thought to disrupt consolidated memories, it is also proposed that reconsolidation and consolidation are similar processes or one process that naturally unfolds at different time points (Dudai & Eisenberg, 2004; McKenzie & Eichenbaum, 2011). Similarly, within the model of scaffolded encoding, different mechanic operations, forming a new episodic memory (consolidation), combining information within a memory trace, and scaffolding another memory for repetition (reconsolidation), follow the same principles and occur spontaneously as a stimulus is encountered and reencountered.

Future Research

Finally, as extensive as the current project is, it's only a tree in a forest. Looking into the future, we consider potential research topics related to scaffolded encoding. The first research question is if scaffolded encoding can be manipulated experimentally. We suppose we could learn from Wahlheim and Jacoby (2013), who found successful detection of change and recollection of a previously associated target were critical for PF and PI in a similar paradigm (A-B, A-D). One explanation is that the effects depend on scaffolded encoding, which in turn depends on accurate recognition of study events. Therefore, similarly, in our described PF and PI paradigm, we could introduce item (is the item old?) and associative recognition task (is the pair old?) during List 2 study. Doing so would help us understand if scaffolded encoding is a fundamental mechanism that underlies memory or a control process.

In addition, the current modeling work presents group-level and qualitative fit for a range of effects. May we stress that a coarse fit for all the hallmark findings is already an important theoretical accomplishment. Nevertheless, we ought to calibrate and polish our model to a finer state. This implies another decade of work including 1) model generation to other effects, 2) data prediction in a quantitative and individual level, and 3) covariation of effects within the model and among participants. The last one calls for a large-scale experiment similar to Cox et al. (2018) for a comprehensive understanding of the relationship among tasks.


(a) List Strength Effect
Recognition
Cued Recall
Free Recall
d prime
Accuracy
Accuracy
weak
strong
pure
mixed
(b) Output Interference
Recognition
Cued Recall
p(old)
p(response)
bins
bins
target
foil
correct
failure
intrusion

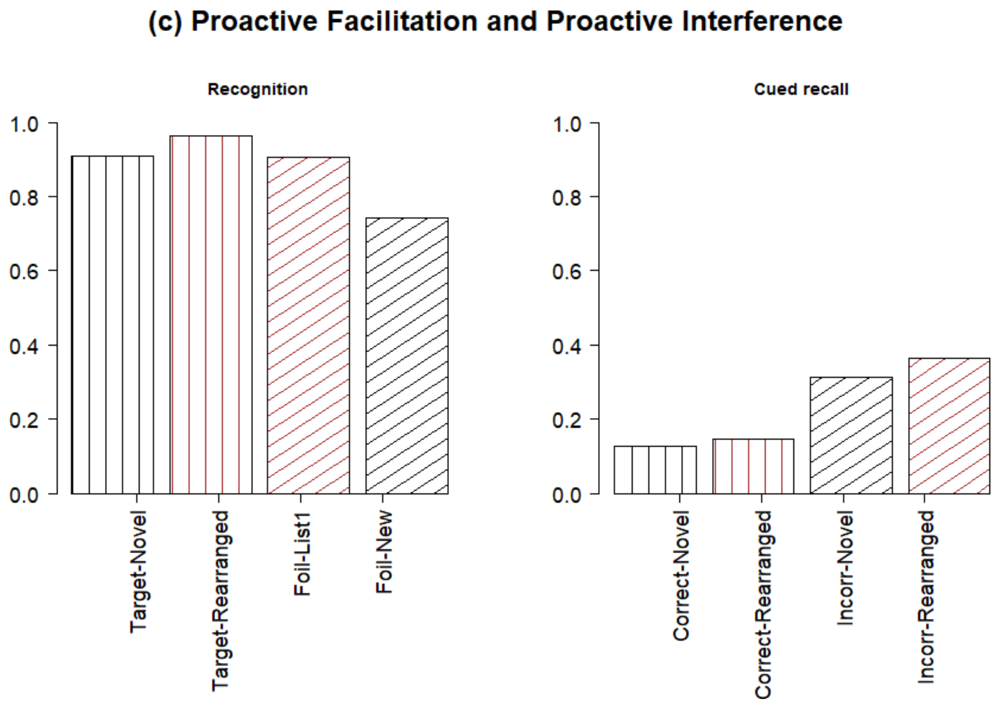


Figure 13

From the model of scaffolded encoding (i.e., REM4), with encoding rate (***u***) as 0.3, the number of context feature ($\boldsymbol{Nf_c}$) as 5, the predicted list strength effect (a), output interference (b), and proactive interference and proactive facilitation (c).

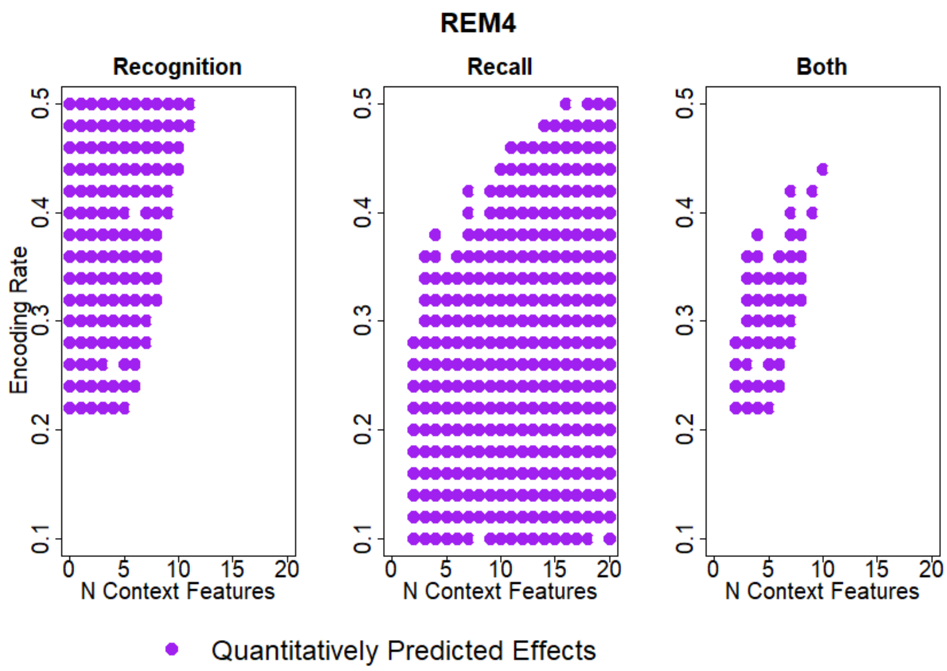


Figure 14

The quantitatively predicted memory effects (list strength effect, output interference, proactive facilitation and interference in recognition and recall) from the model of scaffolded encoding (i.e., REM4). Every purple dot indicates this combination of encoding rate and number of context features predicts all the effects.

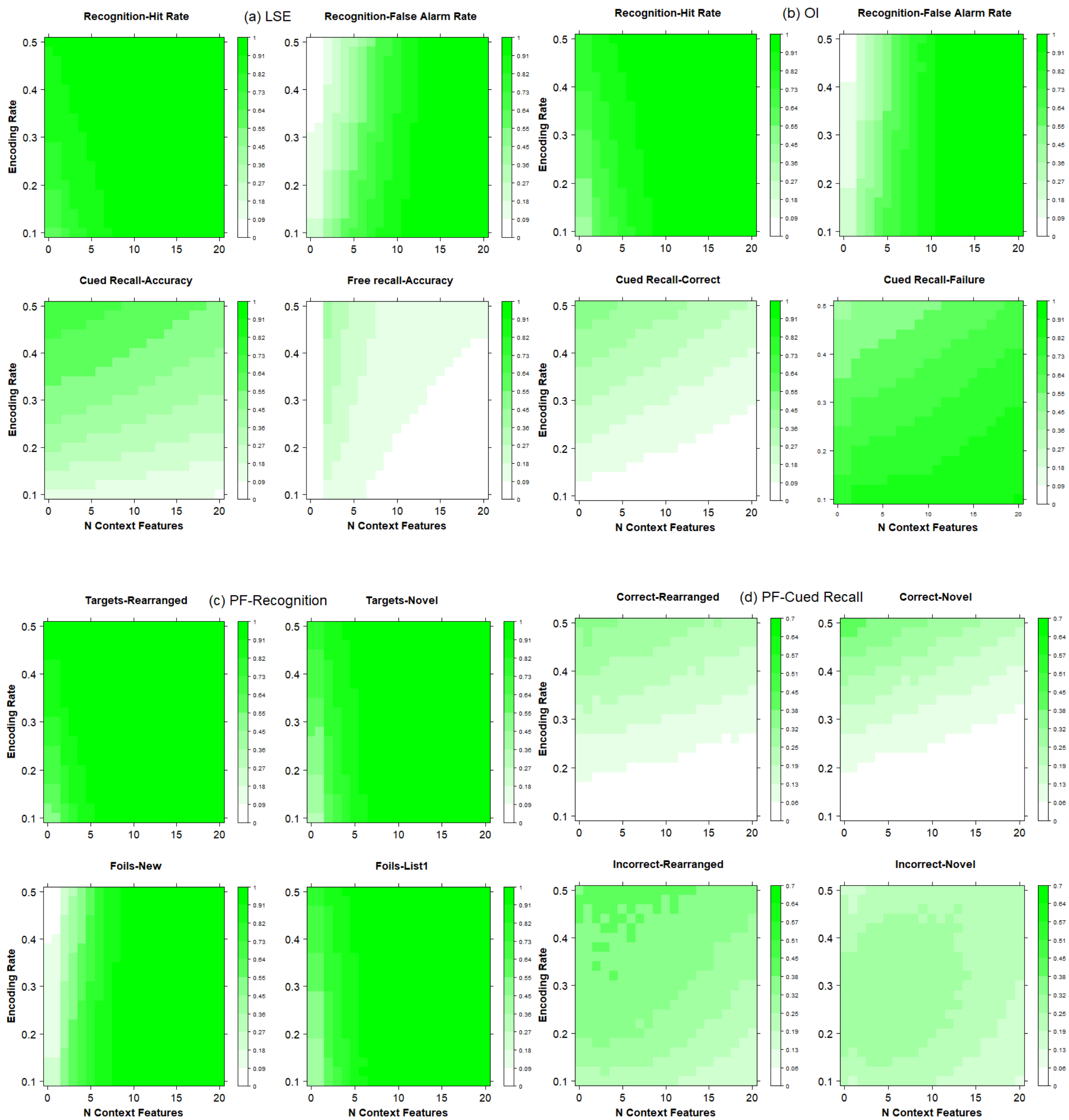


Figure 15

The predicted and averaged accuracy from the model of scaffolded encoding (i.e., REM4) of list strength paradigm (a), output interference (b), proactive facilitation in recognition (c), and proactive facilitation in cued recall (d).

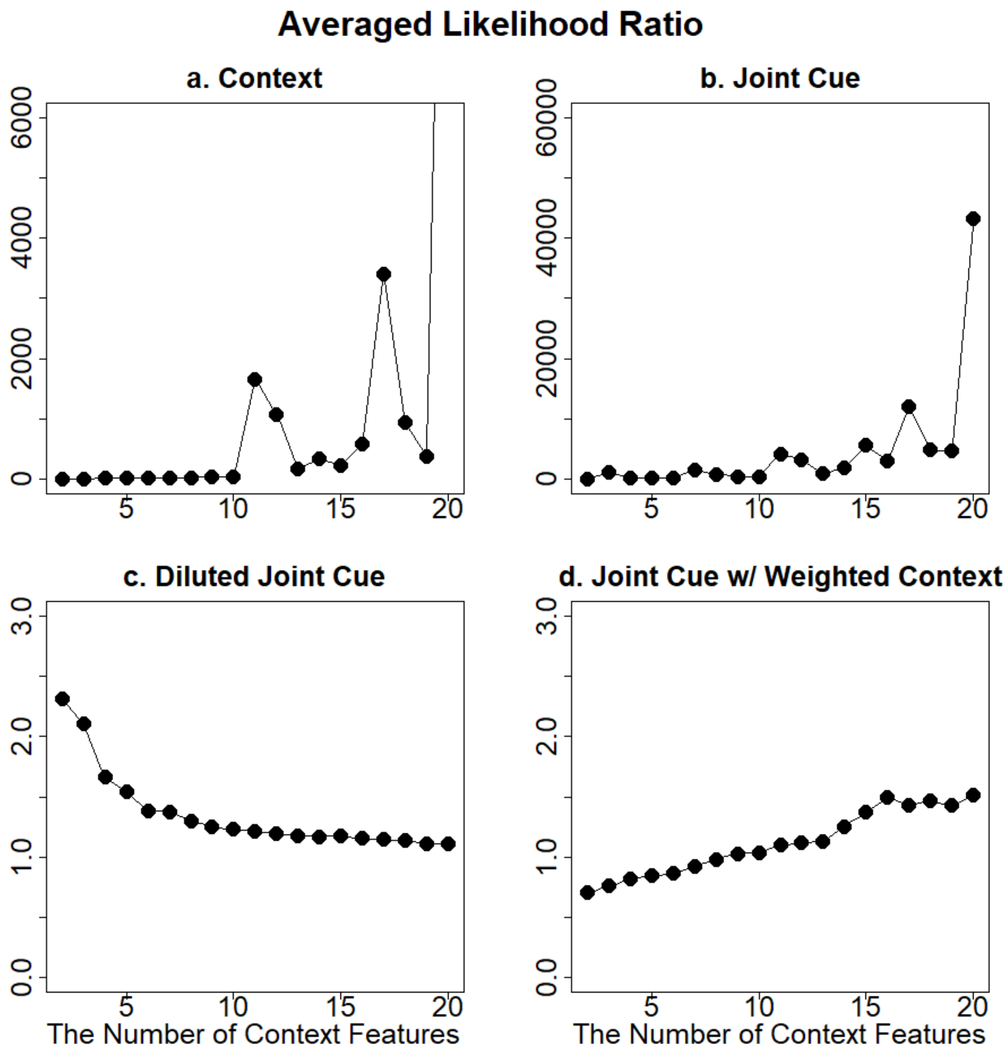
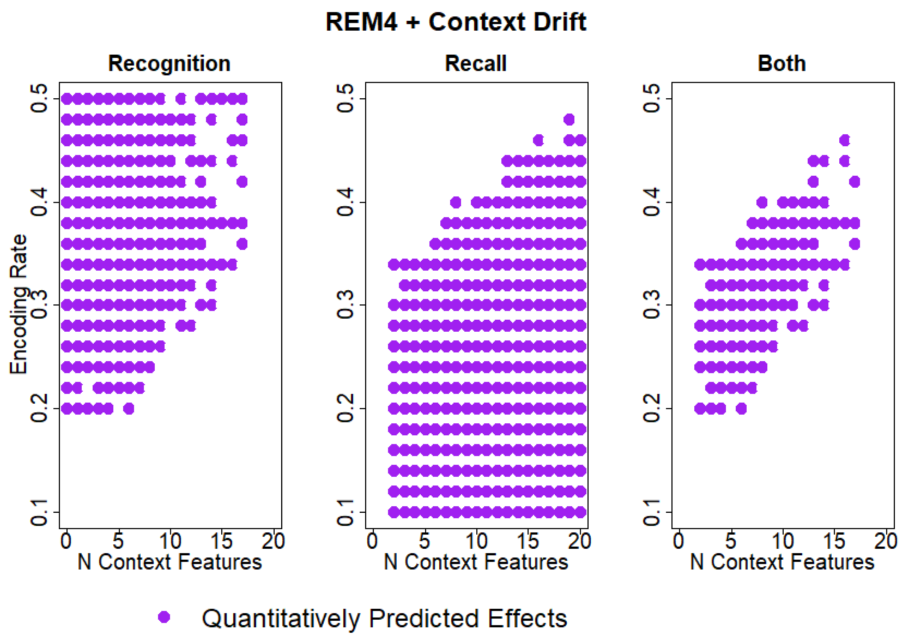
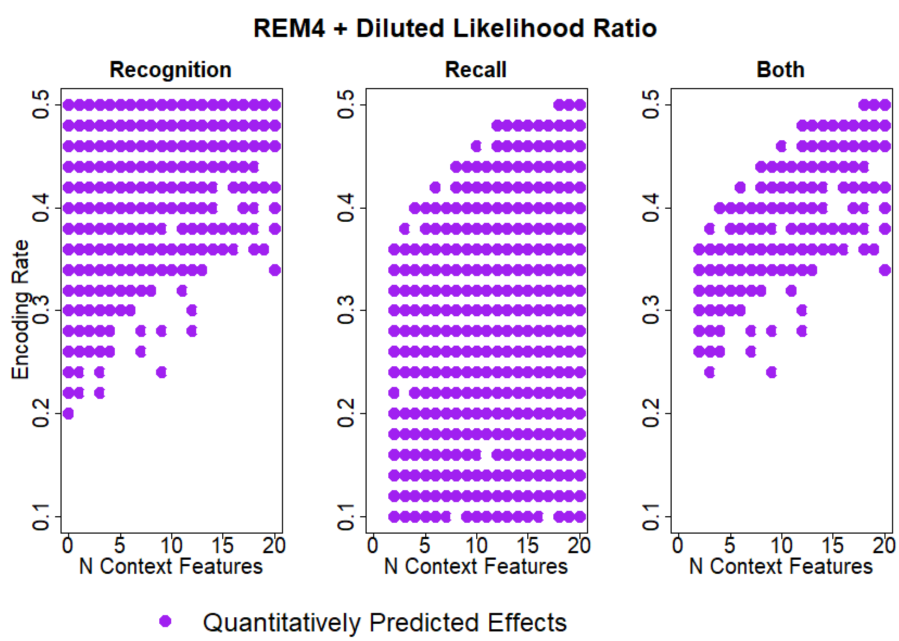


Figure 16

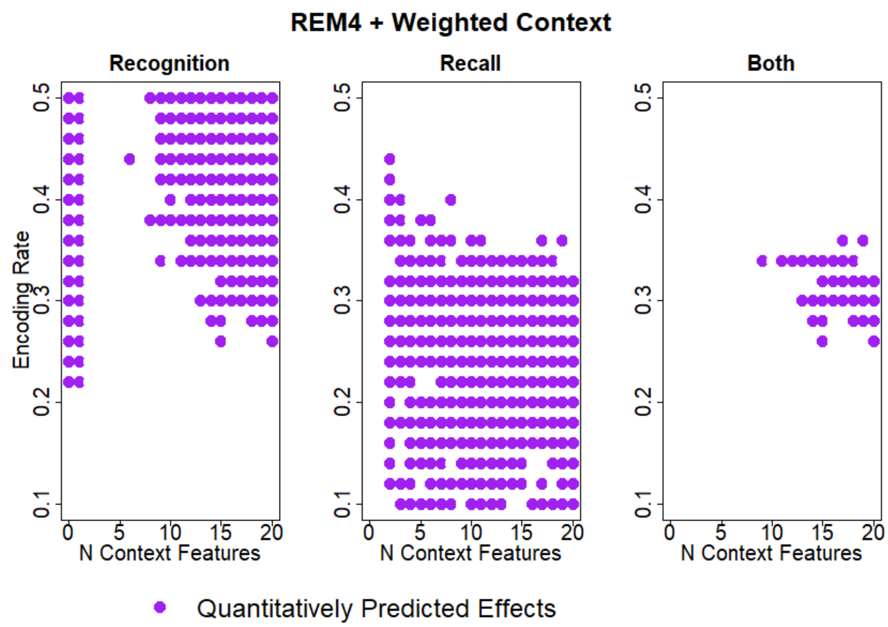
Averaged likelihood ratios from the match between a list of 16 memory traces and a context cue (a) calculated from Equation 2, a joint cue of context and item calculated from Equation 2 (b), a joint cue calculated from Equation 7 and 8 (c), and a joint cue calculated rom Equation 9 (d).


REM4 + Context Drift
Recognition
Recall
Both
Encoding Rate
N Context Features
Quantitatively Predicted Effects

(a)


REM4 + Diluted Likelihood Ratio
Recognition
Recall
Both
Encoding Rate
N Context Features
Quantitatively Predicted Effects

(b)



(c)

<u>Figure 17</u>

The quantitatively predicted memory effects (list strength effect, output interference, proactive facilitation and interference in recognition and recall) from the model of scaffolded encoding (i.e., REM4) with context drift (a), diluted likelihood ratio (b), and weighted context (c). Every purple dot indicates this combination of encoding rate and number of context features predicts all the effects.

Summary

We proposed and tested extensively a unified model of memory, the model of scaffolded encoding. We demonstrate the mechanisms of simultaneous encoding and retrieval within this model are justified. They altogether account for beneficial and detrimental effects of repetition from multiple paradigms in multiple tasks.

References

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

Anderson, M. C., & Neely, J. H. (1996). Interference and inhibition in memory retrieval. In E. Bjork & R. A. Bjork (Eds.), *Memory* (pp. 237–313). San Diego, CA, USA: Academic Press.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes1. In *Psychology of learning and motivation* (Vol. 2, pp. 89-195). Academic Press.

Aue, W. R., Criss, A. H., & Fischetti, N. W. (2012). Associative information in memory: Evidence from cued recall. *Journal of Memory and Language*, *66*(1), 109-122.

Aue, W. R., Criss, A. H., & Novak, M. D. (2017). Evaluating mechanisms of proactive facilitation in cued recall. *Journal of Memory and Language, 94*, 103-118.

Aue, W. R., Fontaine, J. M., & Criss, A. H. (2018). Examining the role of context variability in memory for items and associations. *Memory & cognition, 46*(6), 940-954.

Barnes, J. M., & Underwood, B. J. (1959). ’Fate’ of first-list associations in transfer theory. *Journal of Experimental Psychology*, *58*(2), 97–105.

Benjamin, A. S. (2005). Response speeding mediates the contributions of cue familiarity and target retrievability to metamnemonic judgments. *Psychonomic Bulletin & Review, 12*, 874–879.

Bjork, R. (1975). Retrieval as a memory modifier: An interpretation of negative recency and related phenomena. In R. L. Solso (Ed.), *Information processing and cognition: The lyola symposium* (pp. 123–144). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Bregman, A., & Wiener, J. (1970). Effects of test trials in paired-associate and free recall learning. *Journal of Verbal Learning and Verbal Behavior, 9,* 689-698.

Burns, D. J. (1989). Proactive interference: An individual-item versus relational processing account. *Journal of Memory and Language, 28*(3), 345–359.

Chen, S., & Criss, A.H. (Under Revision). The Source of List Strength Effect in Recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Cox, G. E., & Criss, A. H. (2017). Parallel interactive retrieval of item and associative information from event memory. *Cognitive psychology*, *97*, 31-61.

Cox, G. E., Hemmer, P., Aue, W. R., & Criss, A. H. (2018). Information and processes underlying semantic and episodic memory across tasks, items, and individuals. *Journal of Experimental Psychology: General, 147*(4), 545.

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior, 11*(6), 671-684.

Criss, A. H. (2005). The representation of single items and associations in episodic memory. (Doctoral dissertation, Indiana University, 2004). *Dissertation Abstracts International B, 65,* 6882.

Criss, A.H., Cox, G., Chen, S., Wilson, J.H, & Aue, W. R. (2018). *A model of cued recall*. Talk presented at Psychonomic Society's 59th Annual Meeting, New Orleans, Louisiana.

Criss, A.H. & Howard, M. (2015). Models of episodic memory. In J. Busemeyer, J. T. Townsend, Z. Wang, & A. Eidels (Eds.) *Oxford handbook of computational and mathematical psychology*. New York, NY: Oxford University Press.

Criss, A. H., & Koop, G. J. (2015). Differentiation in episodic memory. In J. Raaijmakers, A. H. Criss, R. Goldstone, R. Nosofsky, & M. Steyvers (Eds.), *Cognitive modeling in* *perception and memory: A festschrift for Richard M. Shiffrin* (pp. 112–125). Psychology Press.

Criss, A.H., Malmberg, K.J., & Shiffrin, R.M. (2011). Output interference in recognition memory. *Journal of Memory and Language*, *64*(4), 316-326.

Criss, A. H., & Shiffrin, R. M. (2004a). Context noise and item noise jointly determine recognition memory: a comment on Dennis and Humphreys (2001).

Criss, A. H., & Shiffrin, R. M. (2004b). Pairs do not suffer interference from other types of pairs or single items in associative recognition. *Memory & Cognition*, *32*(8), 1284-1297.

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

Diller, D. E., Nobel, P. A., & Shiffrin, R. M. (2001). An ARC–REM model for accuracy and response time in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(2), 414.

Dudai, Y., & Eisenberg, M. (2004). Rites of passage of the engram: reconsolidation and the lingering consolidation hypothesis. *Neuron*, *44*(1), 93-100.

Dyne, A. M., Humphreys, M. S., Bain, J. D., & Pike, R. (1990). Associative interference effects in recognition and recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(5), 813.

Estes, W. K. (1955). Statistical theory of spontaneous recovery and regression. *Psychological review, 62*(3), 145.

Fisher, S. L., & Nelson, D. L. (2006). Recursive reminding: Effects of repetition, printed frequency, connectivity, and set size on recognition and judgments of frequency. *Memory & cognition*, *34*(2), 295-306.

Fox, J., Dennis, S., & Osth, A. F. (2020). Accounting for the build-up of proactive interference across lists in a list length paradigm reveals a dominance of item-noise in recognition memory. *Journal of Memory and Language*, *110*, 104065.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological review*, *91*(1), 1.

Godden, D. R., & Baddeley, A. D. (1975). Context‐dependent memory in two natural environments: On land and underwater. *British Journal of psychology*, *66*(3), 325-331.

Gordon WC. 1981. Mechanisms of cue-induced retention enhancement. In: Spear NE, Miller RR, editors. *Information processing in animals: memory mechanisms*. Hillsdale, NJ: Erlbaum. p. 319–339.

Hardt, O., Einarsson, E. Ö., & Nader, K. (2010). A bridge over troubled water: Reconsolidation as a link between cognitive and neuroscientific memory research traditions. *Annual review of psychology*, *61*, 141-167.

Hintzman, D. L. (1970). Effects of repetition and exposure duration on memory. *Journal of Experimental Psychology*, *83*(3p1), 435.

Hintzman, D. L. (1974). Theoretical implications of the spacing effect. In R. L. Solso (Ed.), *Theories in cognitive psychology: The Loyola Symposium* (pp. 77–99). Hillsdale, NJ: Erlbaum.

Hintzman, D. L. (1976). Repetition and memory. In *Psychology of learning and motivation* (Vol. 10, pp. 47-91). Academic Press.

Hintzman, D. L. (1984). MINERVA 2: A simulation model of human memory. *Behavior Research Methods, Instruments, & Computers*, *16*(2), 96-101.

Hintzman, D. L. (1986). " Schema abstraction" in a multiple-trace memory model. *Psychological review*, *93*(4), 411.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological review*, *95*(4), 528.

Hintzman, D. L. (2004). Judgment of frequency versus recognition confidence: Repetition and recursive reminding. *Memory & Cognition*, *32*(2), 336-350.

Hintzman, D. L. (2010). How does repetition affect memory? Evidence from judgments of recency. *Memory & Cognition*, *38*(1), 102-115.

Hintzman, D. L. (2011). Research strategy in the study of memory: Fads, fallacies, and the search for the "coordinates of truth". *Perspectives on Psychological Science*, *6*(3), 253-271.

Hintzman, D. L., & Block, R. A. (1971). Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology*, *88*(3), 297.

Hintzman, D. L., & Block, R. A. (1973). Memory for the spacing of repetitions. *Journal of Experimental Psychology*, *99*(1), 70.

Hintzman, D. L., Caulton, D. A., & Levitin, D. J. (1998). Retrieval dynamics in recognition and list discrimination: Further evidence of separate processes of familiarity and recall. *Memory & Cognition*, *26*(3), 449-462

Hockley, W. E., & Cristi, C. (1996a). Tests of encoding tradeoffs between item and associative information. *Memory & Cognition*, *24*(2), 202-216.

Hockley, W. E., & Cristi, C. (1996b). Tests of the separate retrieval of item and associative information using a frequency-judgment task. *Memory & Cognition*, *24*(6), 796-811.

Hopper, W. J., & Huber, D. E. (2016). The Primary and Convergent Retrieval Model of Recall. In *CogSci*.

Hopper, W. J., & Huber, D. E. (2018). Learning to recall: Examining recall latencies to test an intra-item learning theory of testing effects. *Journal of Memory and Language*, *102*, 1-15.

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning*, *Memory, and Cognition*, *25*(4), 923.

Howard, M., & Kahana, M. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299.

Humphreys, M., Bain, J., & Pike, R. (1989). Different ways to cue a coherent memory system: A theory for episodic, semantic, and procedural tasks. *Psychological Review*, *96*(2), 208–233.

Hupbach, A., Gomez, R., & Nadel, L. (2009). Episodic memory reconsolidation: updating or source confusion?. *Memory*, *17*(5), 502-510.

Jacoby, L. L., & Wahlheim, C. N. (2013). On the importance of looking back: The role of recursive remindings in recency judgments and cued recall. *Memory & cognition*, *41*(5), 625-637.

Jacoby, L. L., Wahlheim, C. N., & Yonelinas, A. P. (2013). The role of detection and recollection of change in list discrimination. *Memory & cognition*, *41*(5), 638-649.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: An episodic context account. In *Psychology of learning and motivation* (Vol. 61, pp. 237-284). Academic Press.

Kelley, R., & Wixted, J. T. (2001). On the nature of associative information in recognition memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27*, 701-722.

Kilic, A., Criss, A.H., Malmberg, K.J., & Shiffrin, R.M. (2017). Models that allow us to perceive the world more accurately also allow us to remember past events more accurately via differentiation. *Cognitive Psychology, 92*, 65-86.

Kilic, A., Fontaine, J., Malmberg, K. J., & Criss, A. H. (submitted). The moderating role of feedback on forgetting in item recognition. *Computational Brain & Behavior.*

Lee, M. D., Criss, A. H., Devezer, B., Donkin, C., Etz, A., Leite, F. P., ... & Vandekerckhove, J. (2019). Robust modeling in cognitive science. *Computational Brain & Behavior*, *2*(3-4), 141-153.

Lehman, M. & Malmberg, K. J. (2013). A buffer model of encoding and temporal correlations in retrieval, *Psychological Review*, *120*(1), 155-189.

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

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2011). Contextual variability in free recall. *Journal of memory and language*, *64*(3), 249-255.

Malmberg, K. J., Criss, A. H., Gangwani, T. H., & Shiffrin, R. M. (2012). Overcoming the negative consequences of interference from recognition memory testing. *Psychological Science*, *23*(2), 115-119.

Malmberg, K. J. & Shiffrin, R. M. (2005) The "one-shot" hypothesis for context storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(2), 322-336.

Malmberg, K. J., & Xu, J. (2007). On the flexibility and the fallibility of associative memory. *Memory & Cognition*, *35*(3), 545-556.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information.

McClelland, J. L., & Chappell, M. (1998). Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. *Psychological review*, *105*(4), 724.

McGaugh, J. L. (2000). Memory--a century of consolidation. *Science*, *287*(5451), 248-251.

McKenzie, S., & Eichenbaum, H. (2011). Consolidation and reconsolidation: two lives of memories?. *Neuron*, *71*(2), 224-233.

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory.

Mensink, G. J., & Raaijmakers, J. G. (1988). A model for interference and forgetting. *Psychological Review*, *95*(4), 434.

Misanin, J. R., Miller, R. R., & Lewis, D. J. (1968). Retrograde amnesia produced by electroconvulsive shock after reactivation of a consolidated memory trace. *Science*, *160*(3827), 554-555.

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*(6), 609.

Murnane, K., & Phelps, M. P. (1993). A global activation approach to the effect of changes in environmental context on recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(4), 882.

Murnane, K., & Phelps, M. P. (1994). When does a different environmental context make a difference in recognition? A global activation model. *Memory & Cognition*, *22*(5), 584-590.

Murnane, K., & Phelps, M. P. (1995). Effects of changes in relative cue strength on context-dependent recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 158.

Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General*, *128*(4), 403.

Nader, K., Schafe, G. E., & Le Doux, J. E. (2000). Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. *Nature*, *406*(6797), 722-726.

Overman, A. A., & Becker, J. T. (2009). The associative deficit in older adult memory: Recognition of pairs is not improved by repetition. *Psychology and Aging*, *24*(2), 501.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009a). A context maintenance and retrieval model of organizational processes in free recall. *Psychological review*, *116*(1), 129.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009b). Task context and organization in free recall. *Neuropsychologia*, *47*(11), 2158-2163.

Postman, L., & Gray, W. (1977). Maintenance of prior associations and proactive inhibition. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(3), 255.

Postman, L., & Underwood, B. J. (1973). Critical issues in interference theory. *Memory & Cognition*, *1*(1), 19-40.

Postman, L., Stark, K., & Burns, S. (1974). Sources of proactive inhibition on unpaced tests of retention. *The American Journal of Psychology*, 33-56.

Popper, K. R. (1934). 1959. *The logic of scientific discovery*.

Team, R. C. (2017). R Core Team (2017). R: A language and environment for statistical computing. *R Found. Stat. Comput. Vienna, Austria. URL http://www. R-project. org/., page R Foundation for Statistical Computing*.

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of verbal learning and verbal behavior*, *6*(2), 175-184.

Raaijmakers, J., & Shiffrin, R. (1981a). Search of associative memory. *Psychological Review*, *88(*2), 93–134.

Raaijmakers, J., & Shiffrin, R. (1981b). Order effects in recall. *Attention and performance*. 403-415.

Ratcliff, R., Clark, S. E., & Shiffrin, R. M. (1990). List-strength effect: I. Data and discussion. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 163.

Rotello, C. M., & Heit, E. (2000). Associative recognition: A case of recall-to-reject processing. *Memory & Cognition*, *28*(6), 907-922.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry*, *20*(1), 11.

Sederberg, P. B., Gershman, S. J., Polyn, S. M., & Norman, K. A. (2011). Human memory reconsolidation can be explained using the temporal context model. *Psychonomic bulletin & review*, *18*(3), 455-468.

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

Shiffrin, R. M., & Nobel, P. A. (1997). The art of model development and testing. *Behavior Research Methods, Instruments, & Computers*, *29*(1), 6-14.

Shiffrin, R. M., Ratcliff, R., & Clark, S. E. (1990). List-strength effect: II. Theoretical mechanisms. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(2), 179.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic bulletin & review*, *4*(2), 145-166.

Shiffrin, R. M., & Steyvers, M. (1998). The effectiveness of retrieval from memory. *Rational models of cognition*, 73-95.

Takashima, A., Nieuwenhuis, I. L., Jensen, O., Talamini, L. M., Rijpkema, M., & Fernández, G. (2009). Shift from hippocampal to neocortical centered retrieval network with consolidation. *Journal of Neuroscience*, *29*(32), 10087-10093.

Thomson, R., Bennati, S., & Lebiere, C. (2014). Extending the influence of contextual information in ACT-R using buffer decay. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 36, No. 36).

Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *6*(2), 175–184.

Tulving, E., & Arbuckle, T. (1963). Sources of intratrial interference in immediate recall of paired associates. *Journal of Verbal Learning & Verbal Behavior*, *1,* 321-334.

Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive remindings in proactive effects of memory. *Memory & Cognition*, *41*(1), 1-15.

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

Wilson, J. H., & Criss, A. H. (2017). The list strength effect in cued recall. *Journal of Memory and Language*, *95*, 78-88.

Wilson, J.H., Kellen, D., Criss, A.H. (2020). Mechanisms of output interference in cued recall. *Memory & Cognition*, *48*, 51–68.

Wixted, J. T., & Rohrer, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(5), 1024.

Curriculum Vitae

EDUCATION

- 2011-2015. B.A., Guangzhou University of Chinese Medicine, Applied Psychology, Department of Psychology
- 2015 – 2017. M.S., Syracuse University, Experimental Psychology, Department of Psychology
- 2015 – present. PhD., Syracuse University, Cognitive Psychology, Department of Psychology

PUBLICATIONS

- Chen, S., & Criss, A.H. (Under Revision). The Source of List Strength Effect in Recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*
- Chen, S., Malmberg, K., Prince, M. & Criss, A.H. (2018). The Effect of Perceptual Information on Output Interference. *Psychonomic Bulletin & Review*. doi: https://doi.org/10.3758/s13423-018-1521-y (see stimuli and data: https://osf.io/znk4f/)
- Kellen, D., Singmann, H., Chen, S., & Winiger, S. (2018). Assumption Violations in Forced-Choice Recognition Judgments: Implications from the Area Theorem. In C. Kalish, M. Rau, J. Zhu, & T. T. Rogers (Eds.), *Proceedings of the 40th Annual Conference of the Cognitive Science Society*. Madison, WI: Cognitive Science Society.

PRESENTATION

- Chen, S., & Criss, A.H (2016, July). *Weighting of Item versus Context in Memory Retrieval: List Strength Effect(s) in Cued Recall.* Computational and Mathematical Modeling of Cognition 2016, Dobbiaco, Italy.

- Chen, S., Wilson, J., & Criss, A.H (2016, November*). Investigation of the Source of List Strength Effect within The Retrieving Effectively from Memory (REM) Framework: Types of Cue versus Levels of Competition*. Poster Presented at Psychonomic Society's 57th Annual Meeting, Boston.
- Chen, S., & Criss, A.H (2018, November). *Testing Effect and Output Interference in Recall.* Poster Presented at Psychonomic Society's 59th Annual Meeting, New Orleans, Louisiana.
- Chen, S., & Criss, A.H (2019, November). *A Unified Account of Episodic Memory: A Model of Scaffolded Encoding That Accounts for Multiple Tasks.* Poster Presented at Psychonomic Society's 60th Annual Meeting, Montreal, Canada.
- Chen, S., & Criss, A.H (2019, November). *The Source of List Strength Effect in Recall.* Poster Presented at Psychonomic Society's 60th Annual Meeting, Montreal, Canada.

INVITED TALKS

- Chen, S., & Criss, A.H. (2017, July). *Investigation of the Source of List Strength Effect within The Retrieving Effectively from Memory Framework*. Sixteenth Annual Summer Interdisciplinary Conference, Interlaken, Switzerland.
- Chen, S. (2017, September). *Rookie's Guide to Open Science in Psychology*. CBB seminar, Syracuse, NY.
- Chen, S., Malmberg, K. J., Prince, M., Peckoo, S., & Criss, A.H. (2018, July). *The Effect of Perceptual Information on Output Interference*. The 51st Annual Meeting of the Society for Mathematical Psychology, Madison, Wisconsin.
- Chen, S., & Criss, A.H. (2019, July). *The Source of List Strength Effect in Recall*. The 52nd Annual Meeting of the Society for Mathematical Psychology, Montreal, Canada.

AWARDS AND HONORS

- Travel and Networking Award for Women of Mathematical Psychology (2019)

TEACHING

- PSY205 Foundations of Human Behavior
- PSY332 Cognitive Lab
- PSY313 Introduction to Research Method in Psychology