Frequency effects on memory: A resource-limited theory

Vencislav Popov (vpopov@cmu.edu)[1,2] & Lynne Reder (reder@cmu.edu)[1,2]

[1] Department of Psychology, Carnegie Mellon University, Pittsburgh, PA
[2] Center for the Neural Basis of Cognition, Pittsburgh, PA

Abstract

We present a review of frequency effects in memory, accompanied by a theory of memory, according to which the storage of new information in long-term memory (LTM) depletes a limited pool of working memory (WM) resources as an inverse function of item strength. We support the theory by showing that items with stronger representations in LTM (e.g. high frequency items) are easier to store, bind to context, and bind to one another; that WM resources are involved in storage and retrieval from LTM; that WM capacity is greater for stronger, more familiar stimuli. We present a novel analysis of preceding item strength, in which we show from eight existing studies that memory for an item is higher if during study it was preceded by a stronger item (e.g. a high frequency word; HF). This effect is cumulative (the more prior items are HF, the better), continuous (memory proportional to word frequency of preceding item), interacts with current item strength (larger for weaker items) and interacts with lag (decreases as the lag between the current and prior study item increases). A computational model that implements the theory is presented, which accounts for these effects. We discuss related phenomena that the model/theory can explain.

Contents

I. Introduction

When we learn new information, we have to combine or chunk multiple features of our experiences into a single whole. The building blocks of these experiences can be, for example, faces, names, concepts, individual sensations, or an experiential context. We can think of these units and the associations among them as traces in a memory system, and some of these traces are arguably stronger than others – people find it much easier to recall and use some words, concepts, or events, compared to others. When it comes to memory trace strength, there are a number of questions which any theory of memory has to answer: How should we conceptualize memory strength? What are the factors that make some memories stronger than others and what are the mechanisms through which changes in memory strength occur? How and why does the existing strength of a memory affect future learning?

Much of memory research revolves around the commonsense concept of memory trace strength[1], yet there is little agreement about how to operationalize it. Some models in the field consider strength to be an intrinsic property stored within memory traces themselves (Anderson, Matessa, & Lebiere, 1997; Reder et al., 2000), while others calculate strength at retrieval, a measure of evidence for the match between a cue and the contents of memory (e.g. Dennis & Humphreys, 2001; Hintzman, 1984; Murdock, 1982; Osth & Dennis, 2015; Shiffrin & Steyvers, 1997). While both classes of models can account for many key findings in the field, they significantly disagree whether these effects occur at encoding or at retrieval. Relatedly, memory theorists have long debated whether memory traces are best described by a single strength value or whether multiple signals and processes contribute to memory retrieval (Anderson & Bower, 1972; Diana, Reder, Arndt, & Park, 2006; Wixted & Mickes, 2010; Yonelinas, 2002). Some researchers even doubt the utility of the strength concept altogether (Hintzman, 2011).

In operationalizing memory strength, theorists have to deal with a number of intriguing behavioral patterns. As it turns out, memory strength is not a simple function of frequency of exposure. Rather, memory strength is influenced by many aspects of the learning procedure, such as the type of practice (restudy vs retrieval; Carrier & Pashler, 1992; Liu, Liang, Li, & Reder, 2014; Roediger & Karpicke, 2006; Rowland, 2014), the spacing of repetitions (massed vs distributed; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006; Glenberg, 1976; Pavlik & Anderson, 2005), and the delay between study and testing (immediate vs delayed testing; Bahrick, 1979). These factors interact in puzzling ways. For example, closely spaced repetitions lead to better memory performance with immediate testing, while more widely distributed repetitions lead to stronger memory after a delay (Bahrick, 1979; Cepeda et al., 2006; Glenberg, 1976; Pavlik & Anderson, 2005). Similarly, restudying the material usually leads to better immediate memory, while retrieving the studied information leads to better recall over time (Broek, Segers, Takashima, & Verhoeven, 2014). These crossover effects suggest that without any intervening information,

[1] For example, a Google Scholar search for the query ["trace strength" AND memory] returns 2520 results

the relative strength of two items can reverse merely with the passage of time, which presents a key challenge for models of human memory.

Complicating matters further, the prior strength of items in memory has multifaceted and often opposing effects on memory encoding, binding and retrieval (for a review, see Reder, Paynter, Diana, Ngiam, & Dickison, 2007). Depending on the nature of the task, the procedure, the experimental design, the stimuli and the processes under investigation, prior item strength can either facilitate or impair memory performance. One major variable with which researchers have studied the effects of prior item strength is normative word frequency. A myriad of research indicates that high frequency hurts episodic recognition (Clark, 1992; Glanzer & Adams, 1990, 1990; Hockley, 1994; MacLeod & Kampe, 1996; Malmberg & Murnane, 2002; Reder et al., 2000; Schulman, 1967); however, on the basis of multiple indirect and direct recent findings, we strongly believe that prior exposure frequency can help encoding. Specifically, the evidence suggests that highly frequent and familiar items are easier to encode, to associate to one another, to bind to an episodic context and to hold in working memory (WM; e.g. DeWitt, Knight, Hicks, & Ball, 2012; Diana & Reder, 2006; Reder et al., 2013; Reder, Liu, Keinath, & Popov, 2016; Reder et al., 2007).

The interpretation of these findings is not widely accepted, however, and while many researchers recognize that people can retrieve stronger memory traces more quickly and more accurately, few acknowledge that the prior strength of a trace affects the ease with which it can be encoded, associated and manipulated. As a result, most memory models pose no role of prior item strength in episodic learning, knowledge formation, or WM capacity (for a review of WM models, see Oberauer, Farrell, Jarrold, & Lewandowsky, 2016).

Our main goal in this article is to describe a theory of human memory that can account for many of the paradoxes raised above. In the process, we will:

- introduce a theory of declarative memory, and its mathematical formulation, in which a key novel mechanism is that the encoding, binding and manipulation of stronger items require less WM resources (Section II)
- review evidence consistent with the claim that stronger items have an encoding advantage in both long-term memory and WM (Section III)
- support a key prediction of the theory, according to which memory for one item is influenced by how much resources are spent in processing preceding items (Section IV)

II. Overview of the theory

The theory we present is an evolution of the Source of Activation Confusion model (Reder et al., 2000; Reder & Schunn, 1996; Schunn, Reder, Nhouyvanisvong, Richards, & Stroffolino, 1997), which itself has roots in the ACT-R cognitive architecture (Anderson et al., 2004). SAC implements a dual-process spreading activation theory in which semantic and episodic memory traces are represented as localist nodes in a network. In SAC, memory strength is a continuous value stored within nodes and the links between them, and this strength increases through practice and decays with time. The model has been previously successful in fitting a variety of findings in recognition and cued-recall memory, including the key mirror frequency (Reder et al., 2000), list length (Cary & Reder, 2003) and list strength effects (Diana & Reder, 2005). For that reason, we have imported many of its assumptions in the current theory.

Despite its successes, the original version of SAC, just like most models of memory, did not assume that the current memory trace strength affects the probability of encoding. The lack of such a mechanism prevents it from explaining why stronger items would be easier to encode, bind and manipulate. This core insight was initially articulated by Reder et al. (2007) and Diana & Reder (2007) on the basis of quasi-experiments that tested memory as a function of normative word frequency, and since then we have accumulated in support of the theory a number of critical experimental findings that directly manipulated exposure frequency of the stimuli (Reder et al., 2016; Shen, Popov, Delahay & Reder, 2017).

The backbone of the current theory is that the storage, binding, and retrieval of stimuli deplete WM resources inversely proportional to the strength of the stimulus representation in long-term memory. This principle can be illustrated by a simple analogy to muscle fatigue. During any physical activity, substrates within muscle fibers fuel contractions of the muscle. These substrates get depleted with every contraction and they recover over time if the muscle is not used. When lifting a heavier weight, the muscle fibers use a greater amount of their available substrates. In this analogy, weaker items in memory are like heavier weights – they are more difficult to manipulate, and their retrieval, binding and encoding require more resources.

To be more specific, our theoretical claims are that:

1. the encoding, updating, and binding of stimuli to context, to other stimuli, or to relational structures, depend on a limited pool of WM resources
2. these operations deplete the WM resource pool in a continuous manner which in turn replenish over time
3. these operations deplete more WM resources for weaker, less familiar stimuli
4. as a result of maintaining or manipulating less familiar chunks of information, there are less WM resources available for performing additional operations or for processing additional stimuli.

The first claim above is shared by a number of other researchers (Blumenfeld, Parks, Yonelinas, & Ranganath, 2010; Geldorp, Parra, & Kessels, 2015; Peterson & Naveh-Benjamin, 2017); however, to our knowledge it has not been implemented before into a mechanistic model of memory. In contrast, the remaining claims about how stimulus familiarity affects WM resources are unique to our proposal (also see Reder et al, 2007). We believe that these key new extensions of the theory hold true regardless of how they are implemented. Below we present a complete description of the theory and its computational implementation. Some of the core assumptions are inherited from the previous version of SAC (Reder et al, 2000), and we will note explicitly where the new model deviates.

A. Full description of the theory

1. Representation

Information in memory is represented as a set of interconnected nodes in an associative network. These nodes reflect semantic or perceptual concepts (e.g., the concept of a "bush" or the representation of a face), events ("I planted a bush yesterday" or "I studied the word "bush" in this experiment"), or contextual information (the internal and external context associated with an experience). For simplicity, we consider these nodes to be the basic representational units, but in principle they also include detailed information about their semantic content. In other words, the localist nodes are high-level abstractions, and the theory is potentially consistent with a number of different lower level representations. A basic schematic illustration of the model representations in a typical paired-associate list learning memory experiment is presented in Figure 1. Semantic nodes represent existing concepts in memory, and they can be thought of as conceptual structures in the neocortex. Context nodes represent the encoding environment, which is specific to the time and place in which the experiment occurs. Episodic nodes represent that a concept or a combination of concepts/features was experienced in a specific context, and they link together the individual aspects of experiences. Episodic nodes can be thought of as hippocampal traces linking incoming neocortical information about concepts and context. Despite the interpretation of the nodes, there is no difference between the properties of each node type.

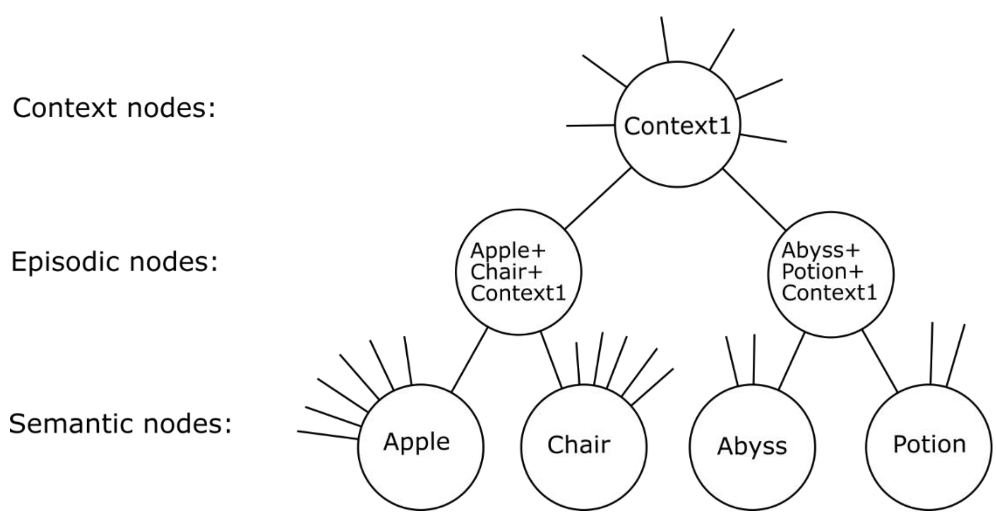

Figure 1. *Illustration of the SAC model structure for a paired-associate experiment. Participants have studied the word pairs Apple-Chair and Abyss-Potion (among others). Each concept has a pre-existing semantic node, which has connections to multiple episodes in which it has been experienced over time. High frequency words (e.g. Apple) are experienced more often, thus have a greater fan of preexisting connections, compared to low frequency words (e.g. Abyss). The current list context also has a separate node, which is connected to all the episodes (i.e. different trials)*

experienced in the current list. There is a unique episode node that connects all features of an experience, i.e. the two concepts and the context in which they are experienced.

We make a clear distinction between stimuli that have existing representations in memory, and stimuli that do not. For example, when people see a novel Chinese character for the first time, they have no choice but to encode it as a configuration of simpler visual features that they already know, such as a small triangle, the letter X, etc. With repeated experience, the feature nodes get associated to a single node, which eventually, with repetition, becomes able to represent the character as a whole. The feature nodes themselves remain attached to the character node and the character is represented by both its independent node, and the individual feature nodes that construct it. This process, which we refer to as *chunking*, is not unique to our model, and bears similarities to many influential proposals in theories of visual perception (Gilbert, Boucher, & Jemel, 2014; Palmer, 1977), statistical learning (Fiser & Aslin, 2005; Perruchet & Vinter, 1998), and has a long history in memory research (Chase & Simon, 1973; Gilchrist, 2015; Gobet et al., 2001; Simon, 1974).

We do, however, consider that the same chunking mechanism operates in most forms of knowledge formation, including episodic, perceptual and statistical learning. In essence, the same process that forms an episode node to represent experiencing a pair of words in a list learning experiment, is the process that forms perceptual nodes to represent complex visual stimuli such as novel Chinese characters, or nodes that represent co-occurring features in a statistical learning paradigm. For all intents and purposes, these nodes are equivalent, and we can consider semantic and perceptual nodes to evolve from episodic nodes through strengthening by repeated exposure.

2. Learning, forgetting, and base-level strength

Two important properties characterize the strength of nodes in memory. We distinguish between the current activation of a node and its base-level strength. Both values differ between nodes and are a function of experience. Every time a certain node is activated, for example by repeatedly studying it in an experimental setting, its base-level strength increases by a discrete amount, and these increments decay as a function of time. These principles are illustrated in Figure 2, where we simulate the base-level strength of a node as a function of the repeated study of a concept. Deviating from the previous implementation of SAC, in the current model the increase in base-level strength depends on how strong the node is at the time of study. Specifically, we assume that there is a maximum strength that nodes can reach, B_{max} (which we fix to 1), and that each repetition strengthens the node as a proportion δ (learning rate) of the maximum strength minus its current base-level strength, *B*:

$$s = \Delta B = \delta(1 - B) \tag{1}$$

This equation governs both the strengthening of existing traces, and the creation of novel traces. We initialize new nodes with a base-level strength of δ, because they have no prior strength. As a result of Equation 1, weaker memory traces are strengthened more than stronger traces and the increment size is controlled by the learning rate, δ. For example, a node whose base-level strength is $B = 0.8$ will be strengthened by $s = 0.2 * \delta$, while a node with a base-level strength of $B = 0.1$ will be strengthened by $s = 0.9 * \delta$. Similar equations have been successful in modeling classical conditioning in the influential Rescorla-Wagner theory (Rescorla & Wagner, 1972) and it is the first main deviation from our previous SAC model[2]. We have found that a learning rate of δ = 0.8 works well for most simulations, and this is the default value.

We chose this learning function, rather than the one used previously in SAC and ACT-R, because it has several desirable properties. First, the amount of resources depleted from WM by the processing of a stimulus is inversely related to its current strength, so we needed a way to quantify the resource cost of this operation. Within this equation the cost of an increment of size *s* can be set to be proportional to *s*. Since weaker items are strengthened more (i.e., $\delta(1 - B_{weak}) > \delta(1 - B_{strong})$), the resource cost of strengthening them is also larger. Second, a single learning parameter, δ, can be modulated by attention. For example, if the participant is less attentive during some stimuli, either due to instructions or divided attention conditions, the learning proportion δ can be decreased, which would also decrease the resource cost. Third, the equation allows us to specify a maximum level of activation that arises naturally from the form of the learning function, rather than arbitrarily cutting off strength values at a certain level.[3] Finally, it allows our model to exhibit crossover effects in strength as a function of spacing and study-test lag[4].

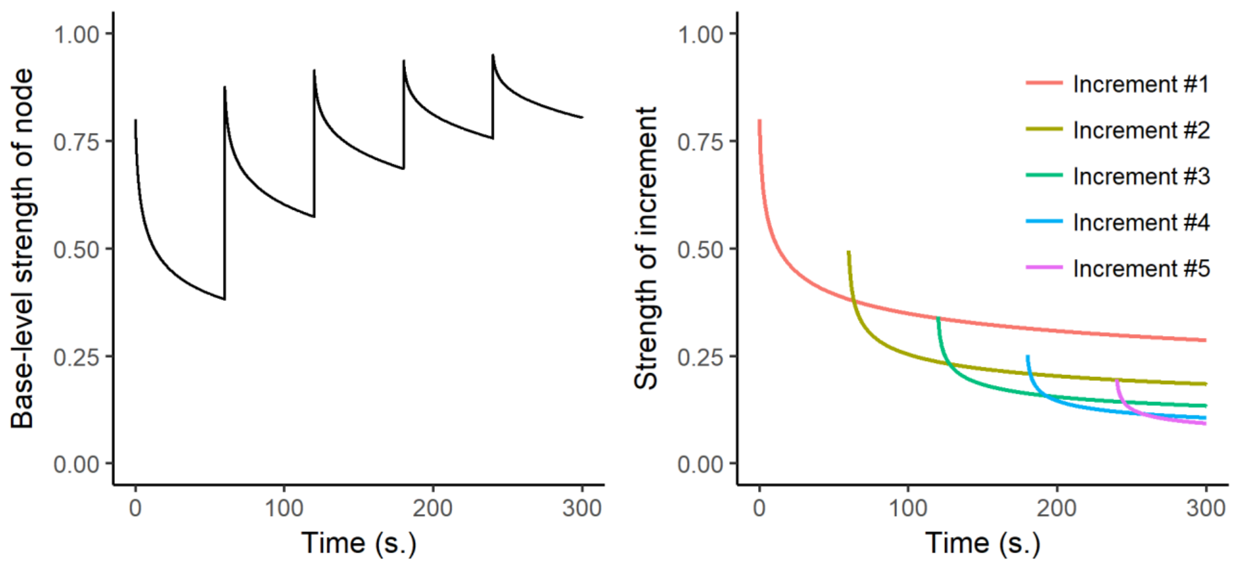

Figure 2. *Illustration of repeated strengthening and decay of a single episode node in the model. The same stimulus is experienced 5 times with 60 seconds in between repetitions. Left panel shows the total base-level strength of the episode node as a function of time. Right panel shows the strength of each individual increment to the node strength. Later increments are smaller, because the size of the increment is a function of the current base-level strength of the*

[2] In previous models, each repetition strengthened memory by the same fixed amount, and the summed strength was then log transformed. Log transformation squashes larger values more, which accounts for the diminishing returns of practice. The same principle is behind ACT-R declarative memory module.

[3] The significance of this point is not the focus of this paper

[4] Modeling the spacing effect is also beyond the scope of this paper

node (see the text for details). The simulation for this figure had a learning rate of δ = 0.8 and a forgetting rate d = -0.18 (default values for all model simulations).

In addition to being strengthened with each repetition, base-level strength decays with time, which is the main cause of forgetting in the model. The strength increment from each repetition decays independently[5] as a function of how much time has elapsed since its occurrence (as illustrated in Figure 2). Thus, at any time, *t*, the base-level strength of a node is:

$$B = B_0 + \sum_{i=1}^{n-1} s_i \times (1 + t - t_i)^{-d}, \tag{2}$$

where s_i is the strength increment produced by the *i*-th repetition, $t - t_i$ is the time since the *i*-th repetition, *d* is the decay rate, and B_0 is the preexisting base-level strength. Researchers have long debated whether forgetting is better described by an exponential or a power function (Rubin & Wenzel, 1996). Consistent with the prior SAC model and ACT-R, we selected a power decay function. The initial time value was offset by 1, so that immediately after encoding the increment size is not infinite.

The current learning and forgetting reformulations of SAC allow for an intuitive interpretation of increment size, base-level strength and forgetting rate for episode nodes. The base-level strength varies between 0 and 1, so does each increment, *s*, as well as the forgetting component, $(1 + t - t_i)^{-d}$. We can interpret each increment as the probability of having stored the node in memory, and the forgetting component as the probability that a node is still in memory. The base-level strength, then, is the probability of having stored the node multiplied by the probability of it still being in memory, which is simply the probability that the node exists.

Equations 1 and 2 are recursive and to calculate the current base-level strength of a node, we need to know how much each repetition incremented it. In turn, in order to know how much each repetition incremented the base-level strength, we need to know what the strength was prior to that repetition. Thus, to calculate the strength increment on a certain repetition *n*, we need to combine Equations 1 and 2:

$$s_n = \delta \left(1 - B_0 - \sum_{i=1}^{n-1} s_i \times (1 + t - t_i)^{-d} \right) \tag{3}$$

The links that connect individual nodes also vary in strength, depending on how often the two nodes have been co-active. The increment and decay of link strength also follow Equations 2 and 3 and the only difference is in the values of the decay parameter.

[5] see Pavlik & Anderson, 2005, for a neural justification behind this assumption

Learning in the model consists of two main operations – creating novel nodes and links and strengthening them with repetition. For example, when a participant studies the word "dog" in an episodic memory experiment, they:

- Strengthen the semantic node for "dog" according to Equation 3
- If this is the first occurrence of "dog" in the experiment, then:
    - Create a new episode node to reflect that the concept "dog" was seen in the experimental context
    - Create links between the "dog" semantic node, the episode node, and the current context node
- If this is a repeated occurrence of the word "dog" in the experiment, then:
    - Strengthen the episode node and its links to the semantic and context nodes

3. Strengthening and binding deplete working memory resources

The key novel aspect of our theory is that a shared pool of resources fuels these learning processes, and that they cost more in resources when applied to items with weaker base-level strength. We assume that people have different amount of WM resources (Daily, Lovett, & Reder, 2001; Lovett, Daily, & Reder, 2000; Lovett, Reder & Lebiere, 1998), which is denoted by a W_{max} parameter. Every time they create a node/link and they strengthen it by an amount s, s^2 amount of resources is depleted. Under most circumstances, this defaults to Equation 1, squared:

$$W_{default\_cost} = s^2 = \left(\delta(1 - B)\right)^2 \tag{4}$$

Since items with stronger base-level strength, such as high frequency words, are incremented less, their processing depletes fewer resources. We chose the cost of strengthening to be s^2 because the square exponent slightly increases the cost difference between small and big increments relative to the overall cost of the operations, which leads to better fits of most models.

We also assume that the resource pool replenishes over time (also see, Reder et al, 2007; Buchler, Faunce, Light, Gottfredson & Reder, 2011). There is limited data to estimate the exact recovery function. The simplest option is that the resource pool recovers at a fixed rate function f of time since the last operation, $t - t_i$, and the remaining resources at time t_i, such that:

$$W_t = \min\left(W_{max}, f(W_{t_i}, t - t_i)\right) \tag{5a}$$

where W_{t_i} is the amount of resource remaining after operation i. Thus, after the last operation, the resource pool recovers at a fixed rate until it reaches W_{max}. If we assume that the recovery function is linear at rate w_r, then:

$$W_t = \min\left(W_{max}, W_{t_i} + w_r(t - t_i)\right) \tag{5b}$$

WM depletion and recovery are illustrated in Figure 3. The left panel shows the available resources at the beginning and end of each study trial in a single list that contains both weak and strong items (i.e. items whose existing representations differ in strength). These simulations used a learning rate of $\delta = 0.8$, resource capacity of $W_{max} = 2.8$, and recovery rate $w_r = 0.45$. Weaker items deplete more resources and the available resources are reduced when more items in a row are weak. The right panel shows the available resources on average after studying strong and weak items in either pure lists of only weak or strong items or mixed lists of both weak and strong items.

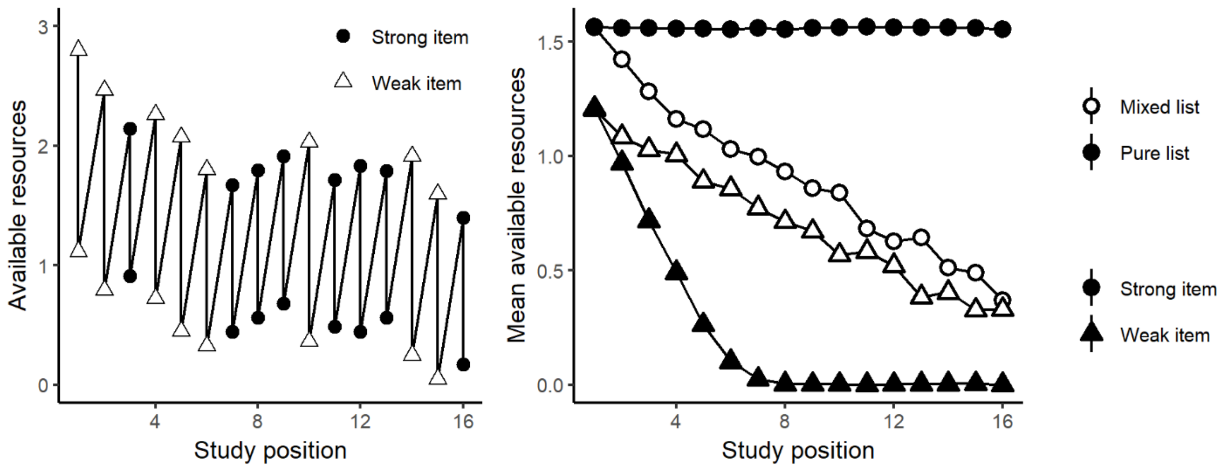

Figure 3. *Illustration of resource depletion and recovery in the model. Left – amount of available resources at the beginning and end of each trial during a single study list, as a function of item position on the list and its current strength. Right – the mean available resources at the end of each trial in pure and mixed lists of weak and strong items.*

Finally, we have to address what happens if the default cost of a process is more than the remaining resources[6]. One option is that encoding fails because of insufficient resources. However, a simpler modeling alternative is that the system uses whatever resource remains, and the strength increment in Equation 1 and 3 is adjusted by the proportion $\frac{W_t}{W_{defaultcost}}$:

$$\begin{aligned} s = \min\left(\frac{W_t}{W_{defaultcost}}, 1\right) \times s = \\ \min\left(\frac{W_t}{s_n}, 1\right) \times s = \\ \min(W_t, s) = \\ \min\big(W_t, \delta(1 - B)\big) \end{aligned} \qquad (6a)$$

Thus, if the default learning requires more resources than there are available, the strength of the memory trace is incremented by just the remaining resources, W_t.

[6] See the General Discussion section on Partial Matching for a more in-depth examination of this topic

A similar situation arises when multiple stimuli have to be encoded at the same time. For example, if you have to remember a display of several different items, then, depending on their number, there might not be sufficient resources to encode them all. One possibility is that the system allocates the default proportion of resources to as many items as possible and fails to encode the remaining items. Another possibility is that all k stimuli share the resources proportionally to their default cost, such that the strength of item i is increased by[7]:

$$s_i' = \min\left(\frac{W_t}{k} \times \frac{(1-B_i)}{\sum_{i=1}^{k}(1-B_i)}, \delta(1-B_i)\right), \qquad (6b)$$

where B_i is the base-level activation of item i. If the number of items encoded at the same time is small enough, then the increment of each node is the default. However, if there are too many items, then they share the resource proportionally to their needs. Equation 6b is a generalized version of Equation 6a, which is a generalized version of Equation 1. When k=1, Equation 6b reduces to Equation 6a, and when sufficient resources remain, Equation 6a reduces to Equation 1.

These equations introduce only two new free parameters to the model – W_{max}, WM capacity, and w_r, WM recovery rate[8]. Yet, these equations allow us to account not only for why stronger items are easier to encode, bind and manipulate, but also how processing some items on a list affects other items on the same list. We can make a novel straightforward prediction – encoding weaker items, or attending more to some items on a list, should impair memory for subsequent items, because fewer resources will be available for their encoding. We present evidence for this prediction in Section IV.

4. Current activation and spreading activation

Nodes are also characterized by their current level of activation. Nodes become active when a concept is perceived, or when activation spreads from other nodes. In contrast to the base-level strength, which decays according to a power-law, the current activation decays exponentially, and its size is dependent on the node base level strength:

$$A = B \times A_{boost} \times e^{-\gamma t} \qquad (7)$$

where B is the current base-level strength of the node, t is the time since activation, γ is the exponential decay parameter, and A_{boost} is the amount of activation received by the node.

The multiplication of the activation boost by the node base-level strength is another deviation from the previous version of SAC, where the activation was added to the base-level strength. This formulation led to more gradual changes in current activation. For example, very

[7] Whether the resource is spread amongst all items or whether a fixed amount is allocated for a limited number of items is currently under debate in the literature. See the General Discussion for more on this topic.

[8] k is a fixed property of the stimulus set and W_t is uniquely determined by the other parameters in the model

weak nodes with strength close to 0 used to get a big boost compared to non-existent nodes, which presented a problem for some of our simulations. With the multiplication equation, the current activation gradually increases from 0 without discontinuities. Furthermore, the multiplication leads to a nice interpretation. As we detailed in section A.2, the base-level strength of episode nodes can be interpreted as the probability of a node being in memory. The result of multiplying the activation boost by the base-level strength thus can be interpreted as the expected activation, given the probability that the node is still in memory.

Under normal conditions, directly perceived nodes become maximally active (i.e. A=1). Nodes can also be activated by their neighbors through spreading activation. When a person perceives the word "dog", this activates not only the semantic node for "dog", but also all events in which it was experienced, as well as other concepts related to it. Following Reder et al. (2000) and Anderson, Bothell, Lebiere, & Matessa (1998), we assume that all nodes connected to a source of activation compete with each other – the activation each node receives depends not only on the strength of its connections with the source nodes, but also on the strength and number of links emanating from each source node:

$$A_{boost,r} = \sum_{s=1}^{n} \left(A_s \times \frac{S_{s,r}}{\sum_{i=1}^{k} S_{s,i}} \right) \tag{9}$$

where $A_{boost,r}$ is the boost in activation in the receiving node, A_s is the activation of the source node, $S_{s,r}$is the strength of the link between the source and the receiving node, and $\sum_{i=1}^{k} S_{s,i}$ is the summed strength of all links emanating from the source node. In summary, the more links that a source node has, and the stronger they are, the less activation is received by nodes connected to the source. This assumption is based on decades of research on fan effects in memory, which show that it is more difficult to retrieve information when the cue has a greater fan of associates (Anderson & Reder, 1999; Anderson, 1974; Schneider & Anderson, 2012). It has also succeeded in modeling many key results in the literature, such as effects of list length (Cary & Reder, 2003), list strength (Diana & Reder, 2005), and word frequency (Reder et al., 2000).

5. Memory retrieval

The outcome of memory retrieval depends on whether the current activation level of the episode node and/or the semantic node, depending on the task, passes a retrieval threshold. SAC is a dual-process model, in which the activation of the episode node represents the recollection signal, while the semantic node's activation level represents the familiarity signal (Reder et al., 2000; Yonelinas, 2002). In recognition tests, a "remember" response occurs if the episode node is above a threshold; however, if the episode activation is below the retrieval threshold, the semantic node is evaluated. Then, a "know" response occurs if the semantic activation is above a threshold; otherwise, the item is judged as "new". In free, cued and serial recall, we assume that a response can be retrieved if the episode node's activation is above its retrieval threshold. Formally, we follow the signal detection theory tradition and assume that there is noise in the signal and that the probability of a response is the area to the right of a threshold under the normal distribution curve with a mean equal to the node's activation. Thus, if the node activation is equal to the retrieval threshold, the probability of a response is 50%.

III. Existing challenges for a theory of frequency effects

Rather than reviewing the evidence for all of the theory assumptions, which has been done elsewhere (Reder et al., 2007) we have a more specific focus – we will review evidence consistent with the claim that weaker items deplete more WM resources during memory formation. The argument goes as follows – first, WM capacity is greater for HF items, a fact demonstrated by both quasi-experimental and experimental studies. We argue that while potential confounds with other variables might explain effects in quasi-experiments, the results from Reder et al (2016) and the additional analyses of those data finesse any interpretation problems from quasi-experiments **(Section III.A).** Second, items with stronger memory representations are easier to bind to other items and to an experiential context in order to form novel episodic traces in LTM **(Section III.B)**. Finally, other items are also easier to store when presented close in time to stronger items, which suggests that LTM formation and binding might draw on a limited resource that is depleted with each operation and recovers over time **(Section III.C)**. In addition to reviewing well established findings, we present a reanalysis of multiple datasets in which we demonstrate that memory performance for one item depends on the frequency of the items that immediately preceded it during study **(Section IV)**. Throughout the text we will show and refer to simulation results and fits to existing data. In order to make the text more readable, all modeling details are presented in **Appendix A**. The modeling code, data and analyses scripts are available at https://github.com/venpopov/prior-item-effects. Before we review evidence for each of the claims presented above, two qualifications are in order.

Normative word frequency. One of the main variables which we use as an indication of the existing strength of memory traces is normative word frequency. Normative word frequency is a quasi-experimental variable, such that one cannot randomly assign words to be of either low or high frequency. As a consequence, there has been a lot of doubt whether normative word frequency has independent effects on memory after controlling for confounded factors such as orthographic and semantic distinctiveness, concreteness, word length, etc. (Cox, Criss, Aue, & Hemmer, 2017; Maddox & Estes, 1997). Our model assumes that high frequency words have stronger memory traces, because the strength of a memory trace increases with repeated experience. Nevertheless, alternative explanations are always possible, and we will get around this issue in two ways. First, we discuss at length experiments that make alternative explanations unlikely. Second, we show that similar results hold when frequency is experimentally manipulated by differential training in the lab (Reder, Angstadt, Cary, Erickson, & Ayers, 2002; Reder et al., 2016; Shen, Popov, Delahay, & Reder, 2017; also see Nelson & Shiffrin, 2013).

Difference between number of chunks and chunk strength**.** A related concern is whether frequency, either natural or experimental, directly affects learning, or whether infrequent items are simply represented as multiple chunks. WM capacity is limited by people's ability to chunk pieces of information together into a single unit, allowing them to greatly expand the amount of information they can temporarily hold and manipulate. Numerous studies have shown that it is not the absolute amount of processed information that limits WM capacity, but it is rather the number of chunks into which this information can be compressed and organized (Miller, 1956; Simon, 1974; for recent reviews, see Cowan, 2001; or Gobet et al., 2001). Researchers may disagree about the specific number of units that constitute the limit of WM capacity (Cowan, 2001; Gobet & Clarkson, 2004; Luck & Vogel, 1997; Miller, 1956) or about the mechanism responsible (Cowan, 2001; Gobet et al., 2001; Oberauer et al., 2016); however, most theories of WM treat chunks as all or nothing. Current theorizing does not see chunks as varying in strength, and, as a consequence, they do not consider that variation in chunk strength would affect WM capacity. Some theorists have argued that exposure frequency in differential training studies does not have direct effects on memory, and that what we call "weak items" might reflect items that do not have a chunked representation. Thus, weak items might often lead to worse memory, because they have to be stored as several chunks[9]. We will show that contrary to this proposal, weak items perform worse in WM tasks even when there is evidence that they have been chunked.

A. Challenge 1: Stronger items consume less working memory resources

1. Natural word frequency and working memory

We begin our review with immediate serial recall and non-word repetition, tasks which require the temporary storage of a small number of items into STM. Research has consistently shown that HF words outperform LF words in pure lists (e.g. only HF or only LF, Caplan, Madan, & Bedwell, 2015; Hulme, 2003; Hulme et al., 1997; Watkins, 1977) and WM capacity appears greater for more familiar naturally occurring stimuli (Xie & Zhang, 2016, 2017)

Could these frequency/familiarity effects be explained by some other difference between HF and LF words, such as differences in 1) study strategies, 2) semantic associability, 3) speech rate or 4) frequency of rehearsals? The effect occurs with both intentional and with incidental encoding, suggesting that it is not due to different study strategies for LF and HF words (Morin, Poirier, Fortin, & Hulme, 2006). Immediate serial recall and non-word repetition are also better for non-word syllables that occur frequently in polysyllabic words, compared to those that do not (Nimmo & Roodenrys, 2002). Since these are meaningless stimuli, it is unlikely that the effect of frequency is due to greater semantic associability of HF items. Better immediate memory for HF words is also difficult to account for by other variables, such as faster speech rate, because the effect remains even when suppressing articulation (e.g. Tehan & Humphreys, 1988) or when accounting for speech rate (e.g. Hulme et al., 1997; Roodenrys, Hulme, & Brown, 1993). Finally, short-term recall of multi-digit numbers is higher if a HF pair of items is rehearsed during a delay interval between the encoding and retrieval of the digits (Humphreys et al., 2010). This result shows that holding a pair of LF words in WM hurts the subsequent recall of other items currently maintained in WM, which is incompatible with any of the alternative explanations discussed above. The fact that word-frequency/familiarity effects occur in all of those cases suggests that item strength per se might have a direct effect on the WM consumption. These effects are summarized in Table 2.

[9] We thank K. Oberauer, E. Awh & N. Cowan for pointing out this possibility to us

2. Experimental familiarization and working memory

The studies reviewed above involve stimuli whose familiarity is naturally occurring. With such quasi-experimental variables, there is always the possibility that some other property of the stimuli is responsible for these results. In two recent studies we gathered direct experimental support for the proposal that weaker chunks would exhaust limited WM resources to a greater degree than stronger chunks (Reder et al., 2016; Shen et al., 2017). Both studies were based on an extensive training procedure in which we familiarized participants with novel Chinese characters at different frequencies of exposure. The familiarization task contained hundreds of trials of visual search training over nine to twelve sessions during a three to four week period, three sessions per week. For each participant, sets of visually similar characters were randomly selected to be seen either at high or low frequency during the visual search task (20:1 ratio).

In the first study (Reder et al., 2016), we measured WM performance for high and low frequency characters at the end of training by using an N-back task. The N-back task required participants to respond whether the current stimulus was the same as the one presented N trials previous, where N could be 1, 2 or 3 in different trial blocks. Participants had to actively maintain the last *n* items in WM, bind each of them to a corresponding serial position, and rapidly update that binding on every trial (Owen, McMillan, Laird, & Bullmore, 2005). As such, this task was perfectly suited to explore whether the familiarity of an item affects the amount of WM resources necessary for encoding, binding, and manipulating the item. As predicted, the N-back performance was better for HF characters.

In the second study (Shen et al.,2017), we provided converging evidence that differential familiarity of stimuli also impacts performance in a different WM task. This study was modeled after an algebra problem-solving task that required participants to retain a digit span of varying length while solving an algebraic equation and then recall the digit span (Anderson, Reder, & Lebiere, 1996). In that experiment, maintaining longer digit spans hurt equation solution speed and accuracy and those effects were greater when the problems were more complex and when some of the digits had to be substituted into the equation to be solved. In the Shen et al. study we differentially familiarized Chinese characters as described above and then varied whether the math task involved more or less familiar characters on a given trial rather than varying the number of digits to hold in memory. Participants had to memorize two pairs of character-digit associations, then they had to solve an algebraic equation and, finally, they had to identify the characters and recall the corresponding digit associated with each character (e.g. 情 = 3 and 洋 = 7). The equations differed on two dimensions – whether they required one (e.g., $3x = 6$) or two steps (e.g., $3x – 2 = 7$) to be solved, and whether they required substitution of constants from the digit span (e.g., 情$x – 2 =$ 洋).

If LF characters indeed require more WM resources to be maintained and processed, then solving the equations should be more difficult when people hold pairs of LF characters associated with digits in WM. The effect should also increase as the equation becomes more complex, either by increasing the number of steps or by requiring variable substitution replicating the pattern shown by varying digit span in Anderson et al. (1996). This is exactly what we found. Importantly, it was more difficult to solve the equations even in the 2-step no-substitution condition, where participants did not have to use the characters they were holding in memory, making it unlikely that the results are due to interference or differential retrieval efficiency from LTM. Furthermore, we found that participants performed better in the recognition and recall task with HF characters, recognizing them more often and recalling the associated digits more accurately.

3. Discounting alternative explanations

Are low-frequency items chunks? The first potential criticism that we alluded to earlier is that results such as these might reflect difference in performance between chunked and non-chunked stimuli, and not differences in chunk strength *per se*. Some have suggested that these effects might occur because LF characters were not exposed often enough for people to develop unified representations of each character, and that as a result, multiple features had to be stored independently in WM for LF characters. While this is possible, at best it is only part of the story for three reasons. First, this explanation cannot account for the quasi-experimental results involving real words. Second, LF characters were exposed as targets on 24 trials over the course of 12 sessions and 4 weeks, and even 3 times more often as distractors on other trials. These multiple repetitions were spaced out and distributed practice leads to significant improvements in memory overall (e.g. Cepeda et al., 2006; Glenberg, 1976; Pavlik & Anderson, 2005). Given the overall frequency and spacing of trials with LF characters, it seems unlikely that they were not chunked by the end of training.

More importantly, while the Reder et al (2016) short report did not include this analysis, we have directly tested whether HF and LF characters reached chunk-like representations. A key property of chunks is that the number of features contained in that chunk is irrelevant for performance – arguably the definition of a chunk. Sets of visually similar characters[10] were randomly assigned to be in the HF or LF category, and the number of features present in each character naturally varied from set to set. If those previously unfamiliar characters did not develop unified memory representations with training, participants would have to store each target as a configuration of features on each visual search trial. In that case, we would expect to see performance in the visual search task to be negatively correlated with the number of features in a character. On the other hand, if these characters throughout training become chunked, we should see the effect of number of features to diminish with each training session.

To perform this analysis, we calculated the number of features present in each character based on existing orthographic vector representations (Xing & Li, 2004; Yang, McCandliss, Shu, & Zevin, 2009). These vector representations are based on an orthographic analysis of the characters and prior behavioral work (Xing et al, 2004). These representations have been used

[10] See the General Discussion for why visual similarity between targets and distractors in this training task is crucial for showing the effects of training frequency

successfully to model print-to-sound mappings in Chinese (Yang et al, 2009) with a connectionist model similar to those used in modeling English print-to-sound mappings (Harm & Seidenberg, 1999).

Figure 4 shows that during the first session, reaction times in the visual search task do depend on the number of basic features within a character. Importantly, the effect gets smaller with each training session for both frequency classes and while it disappears earlier for HF characters, by Session 9 both HF and LF characters show no effect of the number of features within a character. This is evidence that by the end of training both HF and LF were chunked successfully.

We also looked at whether the number of features within a character affect performance in the N-back task. Figure 5 shows that the number of features did not affect performance. Frequency itself contributed independently to lower performance for LF characters, as is evident from the lower intercept for LF characters.

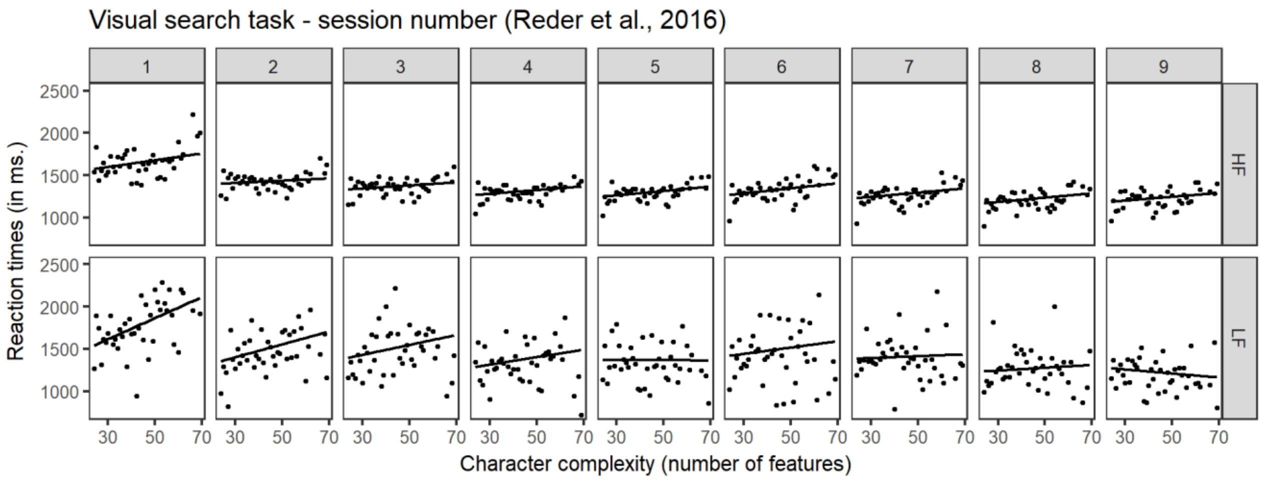

Figure 4. *Reaction times in the visual search task for HF (top row) and LF (bottom row) characters for sessions 1 through 9 (columns), depending on the Chinese character complexity, which was estimated as the number of features in its orthographic vector representations (Xing & Li, 2004; Yang et al., 2009). The data for LF characters is noisier around the slope lines because there are 20 times fewer observations per data point.*

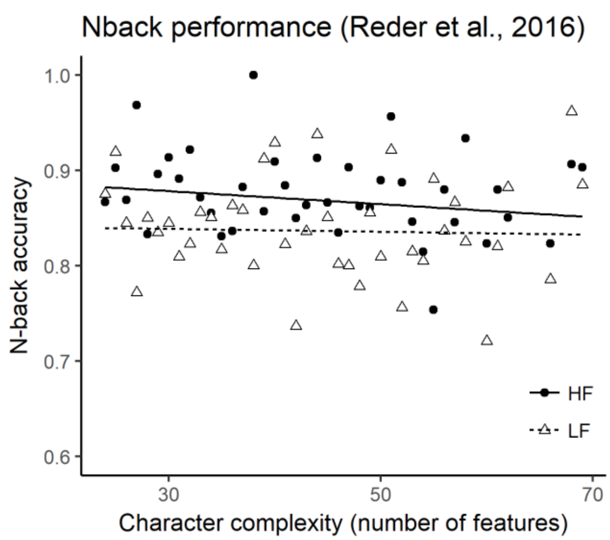

Figure 5. *N-back accuracy depending on the Chinese character complexity, which was estimated as the number of features in its orthographic vector representations (Xing & Li, 2004; Yang et al., 2009)*

Is it just the ease/speed of retrieving representations? A second potential criticism is that the results do not reflect difference in WM consumption, but rather the ease of encoding and/or retrieving the character representations from LTM[11]. The N-back task progresses quickly, and if it is more difficult to retrieve and store the representation of LF characters this might lead to a decrease in performance that is not due to different WM resource consumption.

To test this idea, we performed an additional analysis in which we derived a "processing efficiency" index for each character and each subject individually based on performance in the visual search task. We focused on trials where the target was absent and there were only three distractors, which were the trials containing only distractors from the same similarity class. On these trials, the participant had to do an exhaustive serial search, comparing the target against each distractor, and in the process retrieving the representation of each distractor in turn. On these trials, the reaction time is the sum of retrieval and comparison processes for each distractor. Thus, we used the average reaction time over all trials in which a distractor appeared as a proxy measure for the processing efficiency of that character. For example, if a distractor appeared on 3 trials in which the reaction times were 900, 1000, 1100 ms., then the efficiency index for that character was 1000 ms. This measure was calculated separately for each character and each subject. The measure was transformed into z-scores and reversed, such that larger values represented more efficient processing.

If our index of processing efficiency reflects differences in character processing, we should expect it to predict performance in the N-back task. Indeed, as can be seen from Figure 6, performance was higher for more efficiently processed characters, and the effect was strongest in the 1-back task, weaker in the 2-back, and non-existent in the 3-back. The important question, however, is whether there remains an effect of character frequency after accounting for the effect of processing efficiency. We repeated the mixed-effects regression analyses reported in Reder et al. (2016) with N-back level, character frequency and processing efficiency as predictors of accuracy. The effect of character frequency remained significant even after accounting for processing efficiency (Figure 7).

Interestingly, in Reder et al. (2016) we had predicted an interaction between character frequency and N-back level, such that the effect of frequency should increase as the demands on WM increase. The interaction term did not reach the needed level of significance; however, after accounting for the effect of processing efficiency of each character, we were able to observe that the character frequency had no effect in the 1-back task, had a small effect in the 2-back and had the largest effect in the 3-back, confirming our original prediction. Further, when regressing out encoding efficiency, the interaction of N-back level (difficulty) and character frequency was significant ($p < .01$). Thus, it seems that with low WM demands, the efficiency of retrieving and encoding item representations from LTM is a better predictor of performance; however, as the

[11] We thank K. Oberauer for suggesting this alternative explanation

WM demands increase, the strength of the representation determines WM performance to a greater degree.

The correction for encoding efficiency was necessary because characters differed naturally in complexity. We have since replicated the results of Reder et al (2016) with better controlled stimuli. Rather than using Chinese characters, we used artificial animals called Fribbles, which are three-dimensional combinations of various features and are available online courtesy of Michal Tarr (for a detailed introduction to Fribbles, see Barry, Griffith, De Rossi, & Hermans, 2014; Tarr, 2018; Williams, 1998). We familiarized participants with Fribbles at low and high frequency in a procedure identical to the one used by Reder et al (2016). After the training, N-back performance was lowest for low-frequency Fribbles, middle for high-frequency Fribbles, and highest for pictures of familiar animals. Importantly, as shown on Figure 8, the difference between the three familiarity conditions increased with N-back level, supporting the hypothesis that familiarity of stimuli interacts with working memory resources..

To summarize, we believe that the findings presented in Reder et al (2016), Shen et al (2017), the Fribbles replication of Reder et al (2016), and the additional analyses presented here provide strong converging evidence for the claim that HF items require less WM resources to be processed and maintained. The current analyses also demonstrated that lower performance in the N-back WM task was not due to failure to chunk the low frequency characters nor due to differences in the ease of encoding and retrieving the characters. Combined, these results demonstrate that WM performance is directly influenced by the *strength* of a chunk.

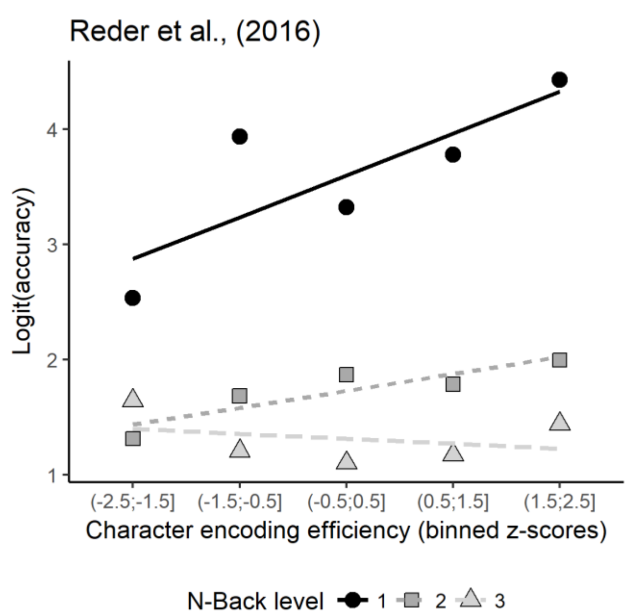

Figure 6. *Accuracy (in logit units) in the N-back task depending on N-back level and processing efficiency of each character. Efficiency was estimated based on performance in the visual search task (see text for more details).*

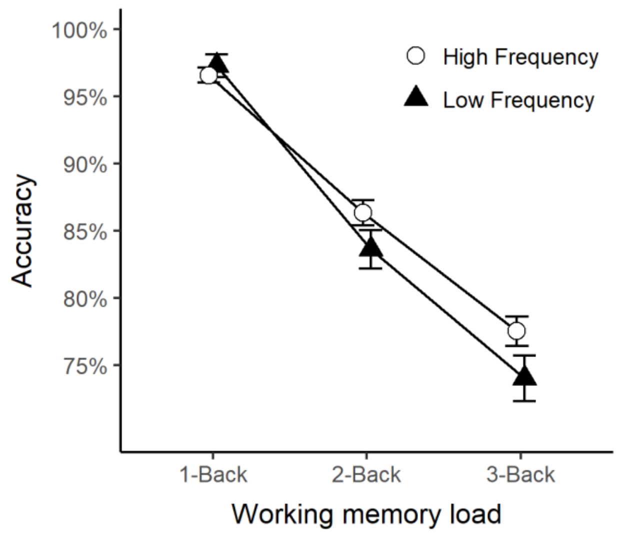

Figure 7. *Effect of character frequency on n-back performance after accounting for processing efficiency of each character. Efficiency was estimated based on performance in the visual search task (see text for more details).*

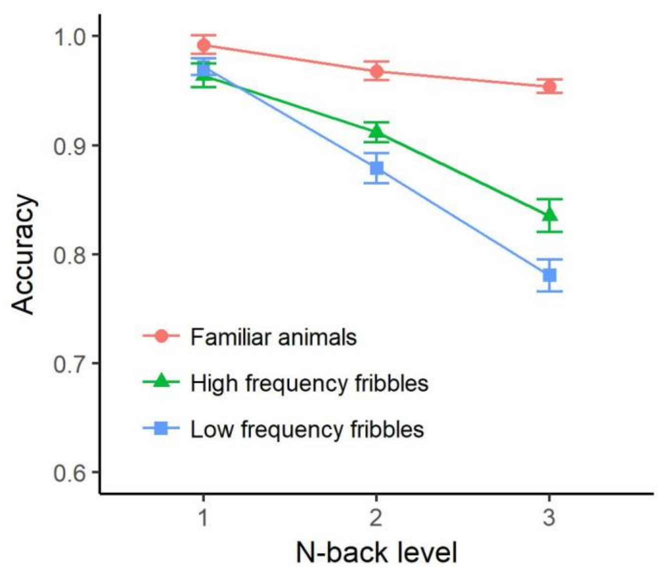

Figure 8. *N-back accuracy depending on stimulus type – familiar animals or Fribbles familiarized in preceding visual search task at either high or low frequency.*

Table 2 Findings consistent with the claim that stronger items consume less WM

Serial recall

- Serial recall of HF items better than LF (Hulme et al., 2003; Morin et al., 2006; Caplan et al., 2015).
- The effect occurs with both intentional and incidental encoding (Morin et al., 2006).
- Immediate serial recall and nonword repetition are better for nonword syllables that occur frequently in polysyllabic words (Nimmo & Roodenrys, 2002).
- The effect remains even when suppressing articulation (e.g. Tehan & Humphreys, 1988) or when accounting for speech rate differences (e.g. Hulme et al., 1997; Roodenrys et al., 1993).
- Short-term recall of multi-digit numbers is higher if HF words are rehearsed during a delay interval between the encoding and retrieval of the digits (Humphreys et al., 2010).

N-back task

- N-back performance is better when the stimuli are Chinese characters *experimentally familiarized* at high frequency rather than low frequency (Reder et al, 2016).
- N-back advantage for HF characters remains even after accounting for encoding/processing efficiency.
- N-back advantage for HF characters increases with the difficulty of the task (N-back level).
- N-back performance highest for images of familiar animals, medium for images of HF experimentally familiarized novel animals (Fribbles), lowest for images of LF experimental familiarized Fribbles.

Solving algebraic equations and memory span

- Better accuracy in solving algebraic equations while holding in character-digit bindings in WM where the characters are experimentally familiarized at high rather than low frequency (Shen et al, 2017).
- The algebraic performance advantage for HF characters increases as the complexity of the equation increases (Shen et al, 2017).
- Higher subsequent recognition of the characters and associated digits when the characters were HF rather than LF (Shen et al, 2017).

B. Challenge 2: Item strength facilitates memory formation

The main claim we make in this section is that forming and storing novel memory traces is easier for stronger items. By memory formation, we refer to the process that binds the elements of an experience into a unified representation or a memory trace of a single episode. This includes both the items that are studied, as well as the experiential context. We believe that the evidence reviewed below supports our claim in three major ways – items with stronger existing memory traces 1) are easier to bind to an experiential context, 2) are easier to bind to one another, and 3) make it easier to form traces for other items experienced closely in time. These three classes of findings encompass the full spectrum of item strength effects at encoding (see Table 2, Table 3 and Table 4 for summaries), and while various explanations have been offered to account for some of these findings, our theory provides a single mechanism to deal with all of them.

1. Effects of word frequency on free recall

In free recall, performance is typically better for HF words in pure lists – study lists that contain only one type of frequency items (Balota & Neely, 1980; Deese, 1960; DeLosh & Mcdaniel, 1996; Gillund & Shiffrin, 1984; Gregg, Montgomery, & Castano, 1980; Sumby, 1963; Ward, Woodward, Stevens, & Stinson, 2003; Watkins, LeCompte, & Kim, 2000). The higher performance of HF words is seen both in terms of overall recall probability, as well as a faster learning rate when studying to a criterion (Sumby, 1963). In pure lists, the effect is parametric – recall is a monotonic function of word frequency (Deese, 1960; although mixed lists might show a U-pattern, Lohnas & Kahana, 2013).

Several explanations have been offered to explain these word-frequency effects. First, HF words are more likely to have preexisting semantic associations to one another, which can facilitate the formation of associations among HF words on the study list (e.g. Deese, 1960; Ozubko & Joordens, 2007; Sumby, 1963). Second, high-frequency words are considered to be more accessible and easier to generate, which facilitates their retrieval during free recall (e.g. Criss, Aue, & Smith, 2011; Madan, Glaholt, & Caplan, 2010). Third, because HF words are more accessible, participants might be rehearsing them more often during study compared to LF words, which strengthens their memory traces to a greater degree (Tan & Ward, 2000; Ward, Woodward, Stevens, & Stinson, 2003). Our model suggests another explanation – forming and storing episodic memory traces is easier for HF words because their existing traces in memory are stronger.

These mechanisms are not mutually exclusive and may all contribute to the observed HF advantage in free recall. Yet, the existing explanations listed above cannot account for the full pattern of word frequency effects. HF recall is higher even when accounting for the number and recency of rehearsals (e.g. Tan & Ward, 2000; Ward et al., 2003) or for the degree of the prior semantic associations among HF words. For example, Tan & Ward (2000) used an overt-rehearsal procedure, and showed that in pure lists HF items are rehearsed more often and later in the list. However, the frequency and recency of rehearsals does not account entirely for the free recall advantage of HF words. A significant effect of word frequency remains even after equating for the number and recency of rehearsals (Ward et al, 2003). Finally, the proposal that HF words are just easier to retrieve cannot explain why the effects disappear in mixed lists (e.g. Ozubko & Joordens, 2007; see Section III.C.1 for an in-depth discussion of pure vs mixed lists), or why they disappear when dividing attention during encoding (Gregg, Montgomery & Castano, 1980).

Our proposal is consistent with the findings discussed above. For example, if HF words are indeed easier to process because they require less resources, then we would predict that depleting resources *in between study items* with an attentionally demanding task would reduce or eliminate the difference between HF and LF items, because there would be insufficient resource to store either[12]. This is precisely what Gregg, Montgomery & Castano (1980) found – when participants had to count backwards from a random three-digit number by threes after studying each word, the recall of both HF and LF words decreased, but more importantly, the difference between them was also eliminated. Furthermore, if there are more resources available when studying HF words, people might form stronger associations between words studied closely in time, which is the case (Tulving & Patkau, 1962; Ward et al., 2003). These results suggest that part of the HF recall advantage comes not only from encoding individual items, but also from forming more and stronger bindings between adjacent items or items and the temporal context.

2. Effects of word frequency on item recognition

In contrast to serial and free recall, the most common finding in item recognition memory is that LF words are recognized better, showing more hits and fewer false alarms than HF words (Clark, 1992; Glanzer & Adams, 1990, 1990; Hockley, 1994; MacLeod & Kampe, 1996; Malmberg & Murnane, 2002; Reder et al., 2000; Schulman, 1967). At first blush it might seem that this effect is inconsistent with the proposal for an encoding advantage for HF words. This finding, known as the mirror frequency effect, was initially quite difficult to account for by the models of the time, but most current memory models can explain the LF advantage in item recognition (Glanzer & Adams, 1990; Hintzman, 1994; Joordens & Hockley, 2000; McClelland & Chappell, 1998; Reder et al., 2000; Shiffrin & Steyvers, 1997). SAC offers a dual-process explanation which has been extensively reviewed elsewhere (Diana et al., 2006; Reder et al., 2000, 2007). In essence, HF words show more false alarms because they have stronger concept nodes, which leads to a greater sense of familiarity for unstudied items. At the same time, HF words are experienced in a greater variety of contexts and situations which makes them much more difficult to retrieve any specific one of them. This context competition leads to the lower hit rate portion of the mirror frequency effect.

What is important for the current argument is that the contextual competition at retrieval masks the fact that HF words are stored more easily. This masking does not occur in free recall, because the contextual fan of the words plays no role – in free recall the only cue is the encoding context. Even though the encoding advantage of HF words is often out-weighed by its contextual

[12] In contrast, increasing demands on WM as part of the task itself, as in the N-back task we discussed earlier, would instead increase the difference

retrieval disadvantage in recognition tests, there are a number of cases in which the demands during encoding are great enough to reverse that trade-off (Reder et al., 2007).

In item recognition, the encoding disadvantage of less familiar words is evident with extremely rare words (e.g. Mandler, Goodman, & Wilkes-Gibbs, 1982; Schulman, 1976; Wixted, 1992; Zechmeister, Curt, & Sebastian, 1978). When a wide range of frequency is considered, recognition hits have a U-shape, where very rare words are recognized worse than HF words (Mandler et al, 1982; Zechmeister et al, 1978). Relatedly, recognition increases as a monotonic function of subjective familiarity of rare words (less than 1 in 4 million; Shulman, 1976). These findings appear to be in conflict with results by Lohnas & Kahana (2013) who found that recognition is a monotonic decreasing function of word frequency. However, Lohnas & Kahana (2003) tested words with frequency of more than 2 counts per million, and their lowest bin of stimuli had an average frequency of 21 counts per million. Thus, the recognition disadvantage for very rare words seems to be robust.

Just as we discussed in Section III.A.3 for WM, one criticism towards studies using very rare words is that their recognition might be worse not because they are less frequent per se, but rather because they are unknown to participants and might not have representations in LTM (Shiffrin & Steyvers, 1997). Without a preexisting representation in LTM, the task of remembering a word becomes more akin to associative learning. Associative learning is more difficult than item learning because it requires binding not only the item and the current context, but also the constituent parts of the unfamiliar item. Yet, similar U-shaped results have been found in training studies, in which participants were exposed to pseudowords (Reder et al, 2002) or to three-digit numbers (Maddox & Estes, 1997) at different experimental frequencies prior to a recognition memory task. In summary, when frequency of exposure is low enough, either quasi-experimentally or experimentally manipulated, its encoding disadvantage appears even in item recognition.

Finally, our theory predicts that the hit rate advantage of LF words in recognition should be lower or eliminated when there is a reduced encoding and binding ability, which will overcome their retrieval benefit. One group that consistently shows reduced WM capacity and episodic memory binding ability is older adults. As predicted by our account, the typical LF hit rate advantage gets monotonically smaller as age increases, while the recognition of HF words does not change (Balota, Burgess, Cortese, & Adams, 2002). While this effect might be attributed to increased familiarity of LF words with age, other attentional factors also eliminate the LF hit rate advantage – for example, reducing the study time (Malmberg & Nelson, 2003) or dividing attention during study (Diana & Reder, 2006)[13].

The fact that reducing study time reduces the LF hit rate advantage is the first modeling challenge we decided to tackle, because despite the theoretical account it was not clear whether a single recovery rate can reproduce the entire pattern of results, and because timing is critical when we consider that WM resources recover over time. We fit our model to Malmberg and Nelson's

[13] One could question whether the LF advantage disappears because of floor effects in all of these studies; however, when we consider the false alarm levels, recognition performance in all of those studies was well above chance levels, making this explanation unlikely.

(2003) findings by estimating a single WM recovery rate for all conditions (see Appendix A for modeling details). Figure 9 shows the empirical data and the fit of the SAC model. The excellent fit of the model reinforces the idea that the reduction of the LF hit rate advantage could be due to insufficient resources available as a result of the fast presentation rate.

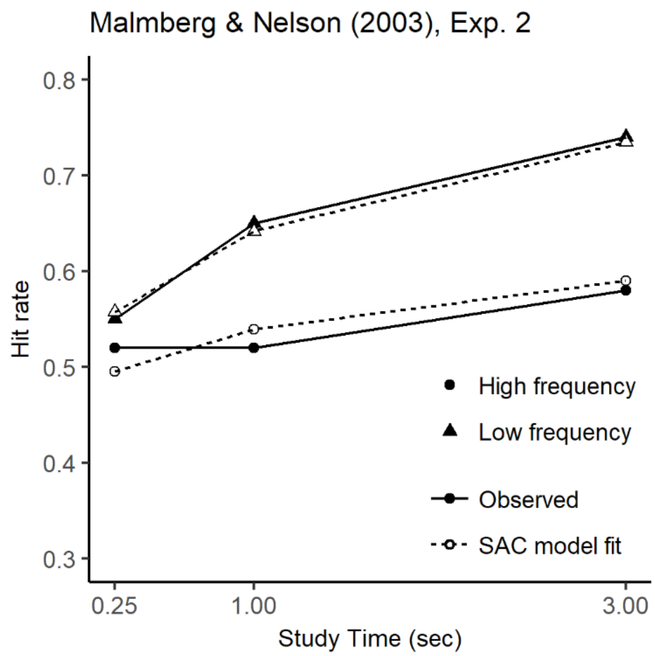

Figure 9. *Hit rate for item recognition of low and high frequency words depending on the trial duration during study. Data from Malmberg & Nelson (2003, Exp. 2) and SAC model fits.*

3. Effects of word frequency on associative recognition

The other case in which we observe beneficial effects of higher word frequency on recognition is with associative recognition wherein people study pairs of words and typically have to discriminate between intact (old words studied in the same pair), recombined (old words studied in different pairs) and novel pairs (new words). Associative learning is more demanding because it requires the encoding of each item, as well as binding the two items together; in our view, each of these operations draws on the same limited resource pool. Consistent with this idea, multiple experiments have demonstrated that pairs (Chalmers & Humphreys, 2003; Clark, 1992, Experiment 1; Clark & Shiffrin, 1992) or triplets (Clark, 1992, Experiment 2) of HF words are easier to recognize than pairs or triplets of LF words (although see Hockley, 1994 for a null effect). For example, Clark (1992) tested item recognition, pair recognition and free recall for the same items studied in triplets. While item recognition showed the typical mirror frequency result, free recall and associative recognition were better for triplets of HF words (replicated by Clark & Shiffrin, 1992, with word pairs as well). Figure 10 shows Clark's 1992 results and the fit of our computation model in which encoding LF words left fewer resources for encoding the association between them.

Our account is also supported by the boundary conditions of frequency effects in associative recognition. First, the HF advantage in associative recognition is present only when the words in the pair are not strongly semantically associated (Martin, 1964). When a strong prior association exists, participants only have to encode the episodic trace, but they do not have to bind the items to one another (presumably because they are already chunked as an association). As a result, there are sufficient resources to encode both low and high frequency word pairs. The HF advantage also disappears with incidental encoding (Humphreys et al., 2010). With incidental encoding instructions participants do not necessarily attempt to bind the two words together, and the demand on WM is smaller, which results in equivalent performance between HF and LF word pairs. Finally, like the familiarization studies we discussed earlier (e.g. Maddox & Estes, 1997; Reder et al., 2002), repeated exposure to LF words *prior* to the experiment facilitates subsequent associative learning (Chalmers & Humphreys, 2003), demonstrating a *direct* benefit of exposure frequency for association formation.

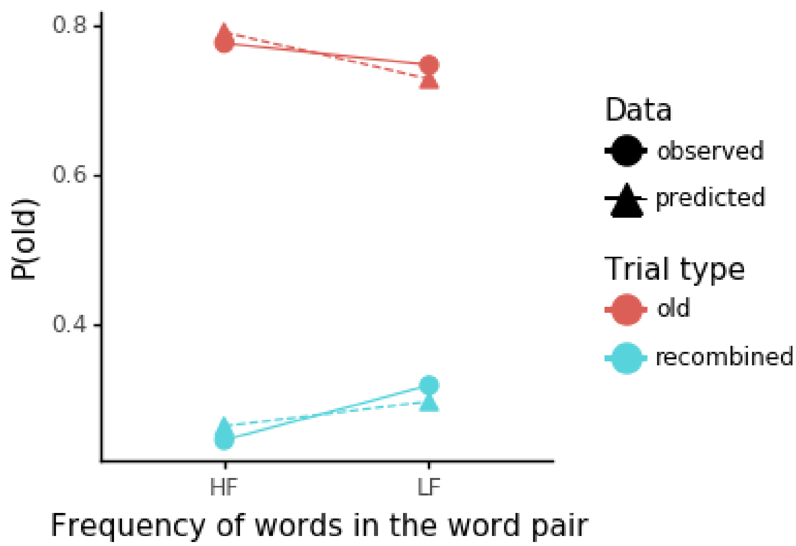

Figure 10. *Proportion of old decisions in associative recognition of word pairs containing HF of LF words in Clark (1992), and the fit of the computation model to the data.*

4. Effects of word frequency on cued-recall and source memory

The storage and association formation advantage of HF items is also demonstrated in studies involving source memory and cued-recall. In source memory tasks, participants have to retrieve the context in which they experienced a certain stimulus. Research has shown that it is easier to retrieve the study context for items that have preexisting representations (e.g., retrieving the scene in which famous and non-famous faces were seen, Reder et al., 2013) or for items rated as more familiar (DeWitt et al., 2012). According to our theory, the processing of an unknown face (Reder et al, 2013) or a less familiar item (DeWitt et al, 2012) consumes the available WM resources to build a long-term representation, so there are fewer resources available for forming a link between the item representation and the context in which they appear. Consistent with the position that these effects occur at encoding rather than retrieval, dividing attention during encoding reduces or removes the familiarity advantage; dividing attention during retrieval does not (DeWitt et al., 2012). Thus, item familiarity facilitates context association *formation* rather than its retrieval.

Two recent comprehensive studies on how word frequency affects cued-recall provide additional support for our position (Criss et al., 2011; Madan et al., 2010). Both studies evaluated separately the effect of cue and target frequency by manipulating them orthogonally. HF targets were better recalled than LF targets. When it comes to the cue, however, Madan et al (2010) found no effect of word frequency, while Criss et al (2011) found that while HF cues were more effective than LF cues, the effect was much smaller than that for the target. Criss et al. (2011) found those results surprising and argued that most memory models would predict a benefit for LF cues, similar to single item recognition, either because they better match their distinctive features in the memory trace (McClelland & Chappell, 1998; Shiffrin & Steyvers, 1997), or because they have fewer pre-experimental associations (Dennis & Humphreys, 2001; Reder et al., 2000). Our explanation for the weaker, but positive effect of cue frequency is that since LF cues take more resources to encode, there are fewer left for forming the cue-target association. Thus, as we discussed before in the item recognition section, the trade-off between lower contextual competition for LF words and their encoding disadvantage can mask the positive effects of the former, and the negative effects of the latter. Consistent with these claims, Criss et al (2011) found in Exp. 3 that when all words have high context diversity, the effect of the cue frequency is removed or reversed.

We modeled these data by fitting a single recovery rate to both experiments, and by estimating contextual fan from the SUBTLEX contextual diversity ratings for each word in the experiment. Figure 11 shows the fits of the SAC model to the data from Criss et al (2011), implementing these assumptions. The model captures the reduced benefit of HF cues compared to HF targets, and also the fact that the cue benefit disappears with high contextual fan, because little activation reaches all targets, regardless of cue.

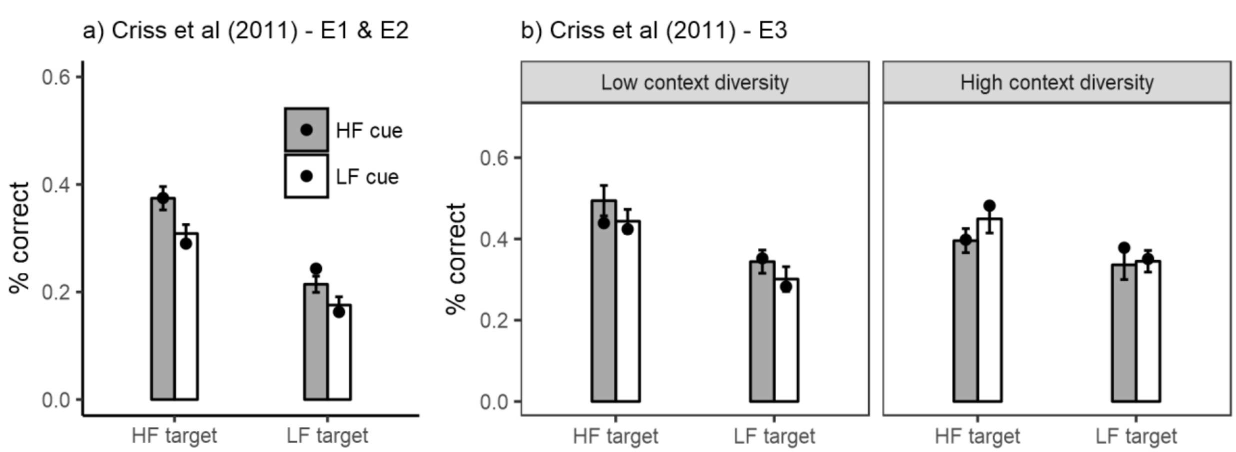

***Figure 11**. Cued recall probability as a function of word frequency of the cue and the target in Criss et al (2011, Exp. 1, 2 and 3), and fits of the SAC model (dots) – a) Exp. 1 and 2 combined; b) cued recall depending on whether the words had low or high context diversity (Exp. 3).*

5. Experimental familiarization and memory formation

Critically, as with most other studies we reviewed, both Madan et al (2010) and Criss et al. (2011) depended on normative word frequency. If the explanation based on the fan/frequency trade-off seems convoluted, consider the large-scale familiarization study with Chinese characters that we discussed earlier (Reder et al., 2016). In that study, in which there was no difference in contextual diversity between HF and LF characters, we also showed that cued-recall is better with novel cues that were trained at high rather than low-frequency, contrary to the findings of Criss et al (2011) and Madan et al. (2010). The cued-recall task was performed at the beginning of each week starting on Week 2. During the study session of the cued-recall task, participants saw two Chinese characters from the same frequency class paired together with an English word. Each character was presented in more than one triplet (two characters paired with a word) to force participants to pay attention to both characters. Each week, the study items were novel combinations of characters with a previously unpresented English word.

Two results from this study are consistent with the idea that item frequency benefits association formation. First, performance in the cued recall task increased each week, even though the triplet combinations were novel on each list. Importantly, pairs of HF characters were better cues for recalling the English words. The advantage of the HF characters was present even after 2-4 weeks post-training and was as large at this long delay as on the last of weekly tests. These results present strong experimental evidence that, rather than acting through confounded factors, frequency of exposure directly facilitates association formation. Finally, we argued before that the cue frequency had a small or no effect in Criss et al (2011) and Madan et al (2011) studies due to a trade-off of higher contextual competition with less WM demands. The results of Reder et al (2016) clearly support that explanation, because when there are no differences in contextual competition, higher cue frequency had a large, positive effect on cued-recall.

Table 3 Findings consistent with the claim that stronger items are easier to encode and bind

Free recall

- Single item free recall of HF is better than LF in pure lists (DeLosh & McDaniel, 1996; Sumby, 1963; Ward et al., 2003).
- Learning rate of HF is also faster, such that with repeated trials, HF word lists reach criterion more quickly (Sumby, 1963).
- Greater temporal clustering of HF words in free recall (Tulving & Patkau, 1962; Ward et al., 2003)
- Dividing attention during encoding eliminates the HF advantage in free recall (Gregg, Montgomery & Castano, 1980).
- The frequency advantage in free recall / serial recall is difficult to dilute even with repeated practice. People learned sequences of HF words more easily even after completing 8 trials of repeated free-recall testing on the same words (Sumby, 1963).

Item recognition

- Single item recognition of very low LF is worse than HF (with both natural frequency, Schulman, 1976, and pseudo-word familiarization, Reder et al., 2002).
- LF recognition advantage is reduced in older adults (Balota et al. 2002).
- LF recognition advantage eliminated with divided attention during encoding (Diana & Reder, 2006).
- LF item recognition advantage is eliminated with short study durations (Malmberg & Nelson, 2003).

Pair recognition

- Pair associate recognition of HF words is better than LF words (Clark, 1992; Clark & Shiffrin, 1992; Chalmers & Humphreys, 2003; although, Hockley 1994 finds no effect).
- HF pairs recognized better only when the words are not strongly associated (Martin, 1964).
- The benefit for HF word pairs is removed with incidental encoding (Humphreys et al., 2010).
- Preexposure to LF words improves subsequent associative learning (Chalmers & Humphreys, 2003).

Cued recall

- Cued recall is better for high-frequency cues and targets (Criss, Aue, & Smith, 2011; Madan et al., 2010), but only with pure lists (Clark & Burchett, 1994).
- Chinese characters ***experimentally familiarized*** at high frequency are easier to associate to one another as the compound cue and to the English language response term than those familiarized at low frequency (Reder et al, 2016).

Source memory

- Better source memory for more familiar items, but the effect disappears when attention is divided during item encoding (DeWitt, Knight, Hicks, & Ball, 2012).

C. Challenge 3: Strength of some items affects encoding of other items (list composition effects)

In this section we examine evidence for another aspect of our thesis – if the encoding of items depends on sharing a limited resource, and the amount of resources spent is an inverse function of the strength of the item being processed, then the strength of some items should affect success at storing subsequent items by influencing how much of these resources remain. We consider mainly list-composition effects, both at the global level (i.e., overall performance differences for pure and mixed frequency study lists) and at the local level (i.e., performance for specific items depending on the strength of immediately preceding or concurrently studied items). See Table 4 for a summary of these findings.

1. Pure vs mixed list paradoxes

A major variable that interacts with word frequency on memory performance is list composition. **Pure study lists** are those that contain only HF or only LF items, while **mixed study lists** are lists that contain both low and high frequency items in an either **random** or alternating **manner**.

The HF advantage in free recall holds only for pure lists, but is absent when the frequency of the stimuli are mixed within a list, a phenomenon often referred to as the mixed list paradox (e.g. Gillund & Shiffrin, 1984; MacLeod & Kampe, 1996; Ozubko & Joordens, 2007; Ward et al., 2003; Watkins et al., 2000). The mixed-list paradox is reliable, as evidenced by a recent meta-analysis (Ozubko & Joordens, 2007). The typical result is HF pure > HF mixed ~ LF mixed > LF pure (see Figure 12). Thus, the presence of LF items on the list hurts memory for HF items, and the presence of HF items on the list helps memory for LF items. As we noted earlier (section B.1), these results cannot be entirely explained by differential rehearsal, because accounting for the number and recency of rehearsals (Ward et al., 2003) or suppressing rehearsal altogether (Popov et al., 2018) does not remove list-composition effects. Our explanation is that less resources are depleted while storing HF items, which leaves more resources available for storing LF items in mixed rather than pure LF lists. We fit the SAC model to the mixed-vs-pure data of Ward et al (2003, Exp. 3), whose results can be seen in Figure 12. Estimating a single recovery rate for all four conditions provided an excellent fit to the data.

List-composition effects are also continuous – increasing the proportion of LF items on the list monotonically decreases performance – recall accuracy is highest in lists composed of 100% HF items, medium in lists composed of 75% HF items, and lowest in lists composed of 25% HF items (DeLosh & McDaniel, 1996). Similar results hold with single item recognition, where the discrimination between LF targets and foils improves as the proportion of HF words on the list increases (Malmberg & Murnane, 2002). Malberg and Murnane (2002) proposed that the probability of encoding the item features varied as a function of list composition. Our model, which proposes a very similar explanation also provides a mechanistic justification for why LF items would be encoded better in the presence of HF items. Malberg and Murnane (2002) did not offer a theoretical reason for why encoding probability should change as a function of list composition. In our model the reduced probability of storing LF items on mixed lists is due to the fact that storing HF words depletes less resources, and as a result, the greater the proportion of HF words on the list, the more resources remain to process the LF words.

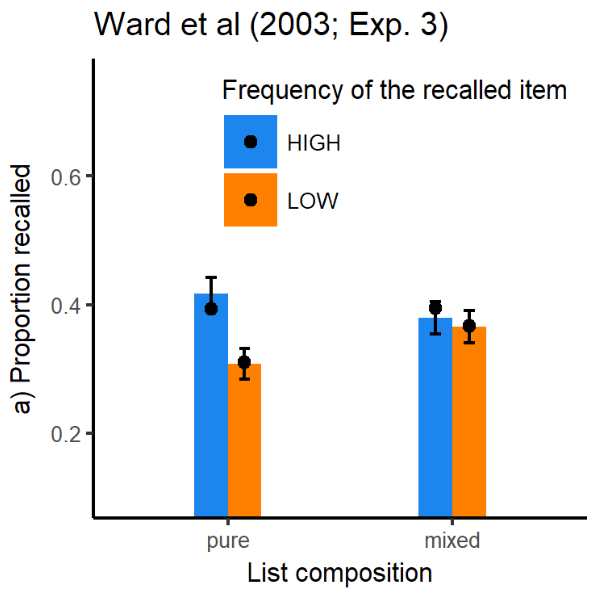

Figure 12. *Free recall proportion for high and low frequency words depending on whether they were studied in pure or mixed frequency lists. Data from Ward et al (2003, Exp. 3) and fits of the SAC model (dots).*

Our theory predicts not only such global list-composition effects, but it also expects that the precise order of LF and HF items on a list will impact performance (also see section IV). Support for this prediction comes from list-composition effects in immediate serial recall (Caplan et al., 2015; Hulme, 2003; Watkins, 1977). Watkins (1977) found that performance was better when the first half of each list contained HF words and the second half contained LF words, rather than the reverse (e.g., HH > HL > LH > LL, where the two letters reflect word frequency within the first and the second half of the lists). In addition, Hulme and colleagues have found that word frequency interacts with serial position in immediate serial recall (Hulme, 2003; Hulme et al., 1997) – the difference in performance increased monotonically with serial position and affected primarily later list positions (see Figure 13 for data and model fits).

Why would performance for mixed lists differ depending on the order of HF and LF items in the lists, if the overall proportion of frequency is the same? Why would frequency effects in pure lists increase with serial position? Within the current theory, WM resources are depleted with each encoding and they recover over time. When LF words are encountered first, they deplete more resources, and thus there are less left for encoding the second half of the list; the reverse is true when the HF items are encountered first. Similarly, each additional item compounds the effect of frequency due to spending more and more resources, leading to an interaction with serial position. Our simulations supported this explanation and the fits of the model to Hulme et al's (1997) results is shown in Figure 13.

A more in-depth exploration of the serial position curves in Hulme (2003) provides further support for the resource-depletion account. In addition to pure lists, Hulme had two mixed lists with alternating sequences that begin either with an LF item (LFHF) or with an HF item (HFLF). Figure 14 shows the result for both the pure and the alternating lists as a function of serial position. If we focus on the first item in each sequence, we can see that recall for the HF item in HFLF lists is equivalent to the recall of the first item in the pure HF lists, while recall for the LF item in LFHF lists is equivalent to that of the first item in pure LF lists. This makes sense, because resources have been depleted to the same degree on the first trial for pure HF lists and mixed HFLF lists; similarly for pure LF lists and LFHF lists. However, the situation changes when we look at HF items on the second trial. For items in the second serial position, performance is lower in the LFHF list than in the pure HF list, even though both test an HF word, because resources are depleted to a greater degree by the preceding LF item (vice versa for LF items on HFLF lists in the second position). Hulme also found similar results when the lists contained words and non-words, rather than LF and HF words (Figure 15). The model simulations (Figure 14 and Figure 15) showed the same pattern in the alternating lists, supporting our explanation[14]. In summary, the complex interaction of frequency and serial position in alternating lists supports the resource-depletion account.

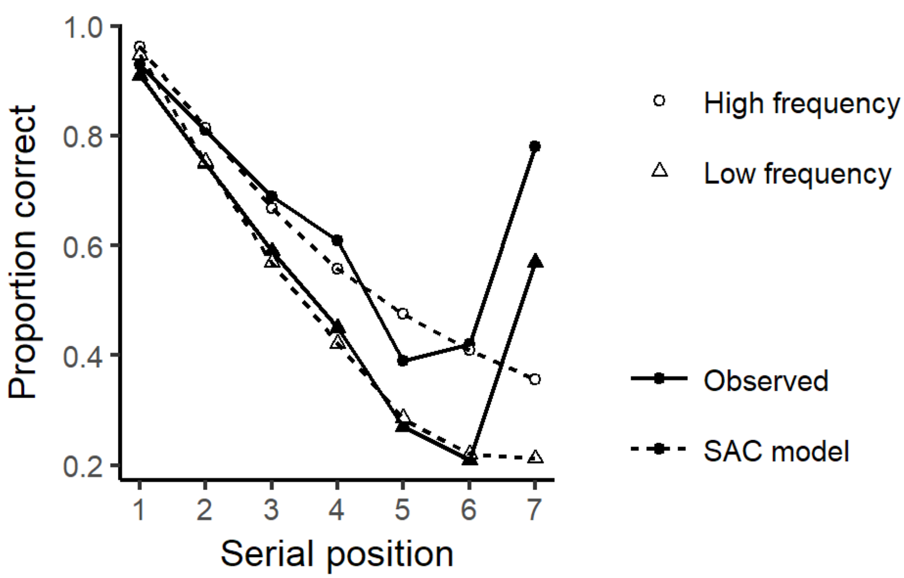

***Figure 13**. Word frequency interacts with serial position during immediate serial recall. Data from Hulme et al (1997) and fits of the SAC model.*

[14] The model fails to capture the one-item recency effect in serial-recall. The SAC model is primarily a model of episodic memory, and it was not developed to account for serial recall performance. However, we simulated the expected accuracy under the assumption that in serial recall the item is bound to a serial position cue, similarly to Anderson, Bothell, Lebiere & Matessa (1998). In contrast to Anderson et al, out model lacks a WM buffer component that was used in ACT-R to account for a one-item recency effects in serial recall. Our focus is on capturing the overall pattern associated with the interaction between word frequency, serial position, and list composition.

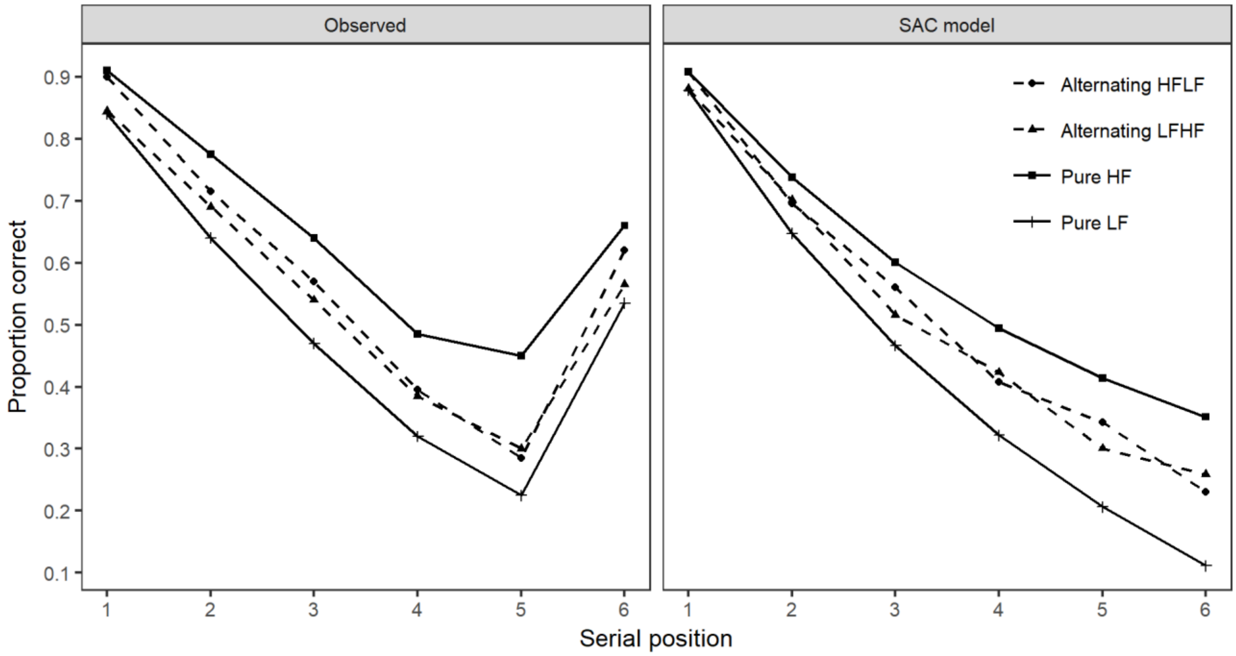

Figure 14. *WF effects in pure and alternating lists in immediate serial recall. Data from Hulme (2003; Averaged from Exp. 1 and Exp. 2) and SAC model fit.*

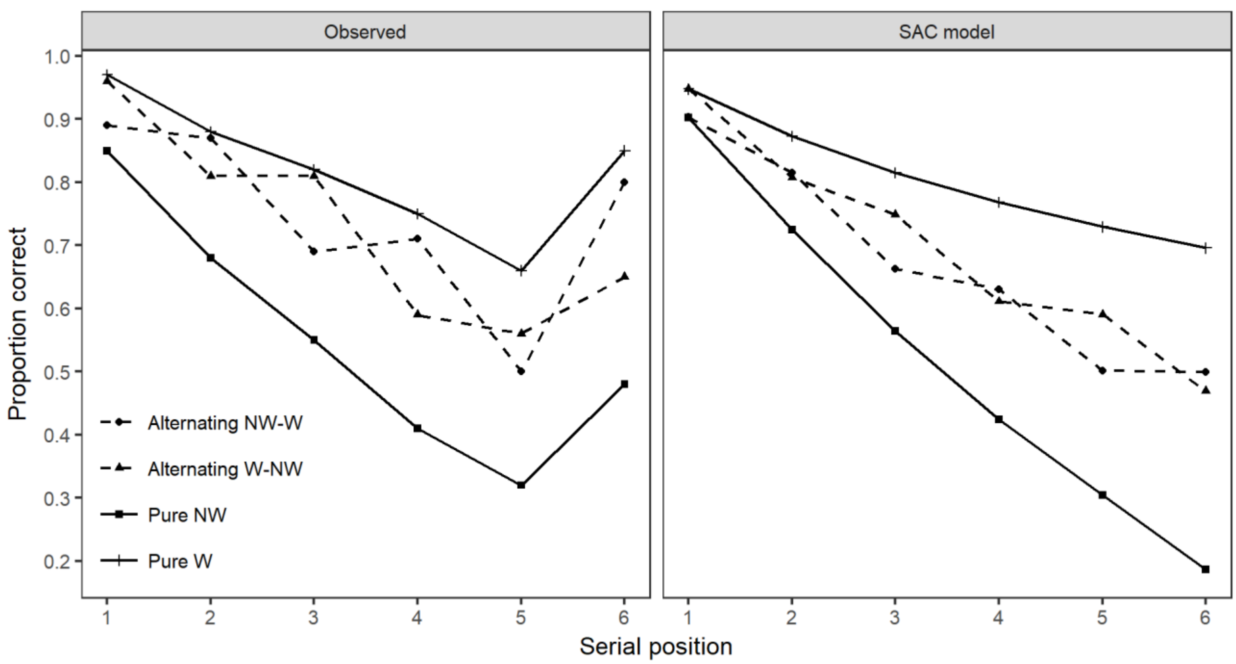

Figure 15 *Word vs non-word immediate serial recall in pure and alternating lists Data from Hulme (2003; Exp 3) and SAC model fit.*

2. Frequency of concurrently studied items

A direct test of the idea that LF items require more resources for encoding, leaving fewer resources for processing concurrent items, comes from several studies in which participants studied pairs of items, but received an item recognition test for individual items within each pair (Diana & Reder, 2006; Malmberg & Nelson, 2003).

Malmberg & Nelson (2003) asked participants to study pairs of two HF words, two LF words or mixed pairs of one HF and one LF word. When the study time was relatively short (1.2 sec), hit rates for both HF and LF words were lower when the other word in the pair was LF rather than HF. This effect is consistent with the idea that LF words require more resources for encoding, a conclusion also supported by the fact that LF targets were hurt more than HF targets in the presence of another LF word. Interestingly, the effect disappeared with longer study times (4.0 sec). Figure 16 shows these results. Malmberg and Nelson's (2003) interpretation was that LF words require more resources only during the initial encoding stages. While their interpretation is possible, our account postulates that resources recover over time and, as such, the 4 second condition may provide sufficient time for resources to recover to more easily store both items. To test this explanation, we fit the SAC model to the data by estimating only a single recovery rate for all conditions. The model fit is also shown in Figure 16.

One might argue, however, that it is unnecessary to invoke the concept of resources to explain these results: People recognize and name LF words more slowly than HF words, and they take more time to study LF words when encoding is self-paced (Rao & Proctor, 1984). It is possible then that in Malmberg & Nelson's (2003) study people merely spend more time encoding LF words, leaving them less time to encode the other word in the pair.

Diana & Reder (2006) discounted this explanation. In addition to replicating Malmberg & Nelson's (2003) results with word pairs, in a different experiment they asked participants to view pictures of objects with words superimposed on them. Participants had to name the word as quickly as possible but attend to both the word and the picture because their memory would be tested for both. Recognition memory for the pictures, which was tested independently of the words, was better for pictures studied with superimposed HF words. Importantly, there was no significant difference in naming times between LF and HF words (630 vs 607 ms.), which equated the study duration between the two frequency conditions. Since people allocate the same amount of time to pronounce LF and HF words, and the next trial began immediately after naming the word, it is not likely that these results can be explained by study time, *per se*. Furthermore, in the next section we will provide evidence that memory for an item is also influenced by the strength of the items that *precede* it during study, an effect that cannot be attributed to difference in study time for the current item.

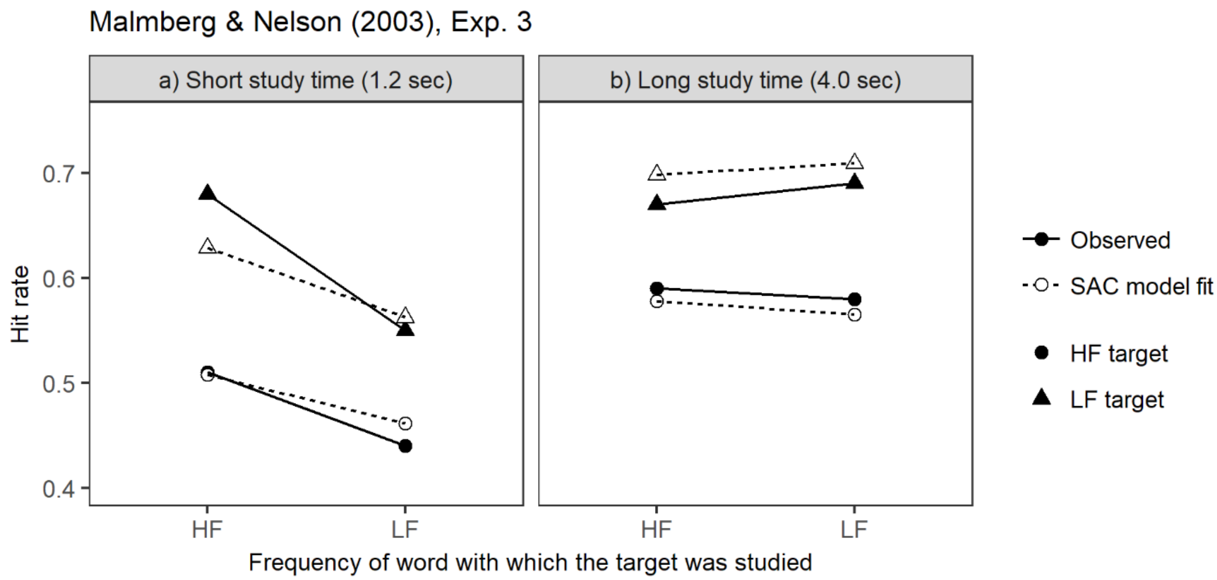

Figure 16*. Hit rate for low and high frequency targets, depending on the frequency of the other word in the study pair and the study duration. Data from Malmberg & Nelson (2003, Exp. 3) and the SAC model fits.*

Table 4 Findings consistent with the claim that strength of some items affects the encoding of other items

Pure vs mixed list paradoxes

- Free recall advantage for HF words only on pure lists, but not on mixed lists (MacLeod & Kampe, 1996; Watkins, LeCompte, & Kim, 2000; Ward et al., 2003; see Ozubko & Joordnes for a meta-analysis)
- Free recall of HF items improves as the proportion of HF items on the list increases (DeLosh & Mcdaniel, 1996).
- Single item recognition of LF improves as the proportion of HF words increases (Malmberg & Murnane, 2002).
- Serial recall of high-frequency words is worse, while low-frequency words is better, when embedded in mixed lists rather than pure lists (Watkins, 1977; Hulme et al., 2003; Caplan et al., 2015).
- Greater serial recall span when the first half of the list is HF and the second half LF, compared to the reverse (Watkins, 1977)

Frequency of concurrently studied items

- Single item recognition is better when the item was studied concurrently with HF rather than with LF item (Diana & Reder, 2006; Malmberg & Nelson, 2003).

Strength of immediately preceding items

- Preceding item frequency effects on learning of current item (Section IV)

IV. Analyses of preceding item strength: novel predictions

The viability of a model should be judged not only by its ability to fit existing data, but also by its ability to make novel predictions. If the unified explanation for list-composition effects that we presented is true, then we should expect that memory should be affected not only by the frequency of the concurrently studied items (Diana & Reder, 2006; Malmberg & Nelson, 2003), but also by the frequency of the immediately preceding items during study. We have posited that LF items deplete more resources, and that these resources are not returned immediately to the resource pool, but rather gradually recover over time. If this is the case, then memory should be worse for items that are preceded by LF items during study. In the remainder of the paper we will present *re-analyses* of multiple existing datasets from different labs that use a variety of memory paradigms, all of which support this prediction.

Most memory studies use a separate study-test cycle procedure in which items are first studied in a list before a memory test is presented. Our theory predicts that memory for a given item will also be a function of the strength/frequency of the items that immediately preceded it during the study session. Looking at Figure 17, which represents a sequence of study items, this means that memory for item X_k will be a function of how much of the resource was spent in memorizing the immediately preceding items X_{k-1}, X_{k-2}, X_{k-i} etc, where k denotes the position of the current item, and *i* denotes the lag or the temporal distance to the preceding items – for example, a lag of 2 means that the X_{k-2} item appeared two items ago during study.

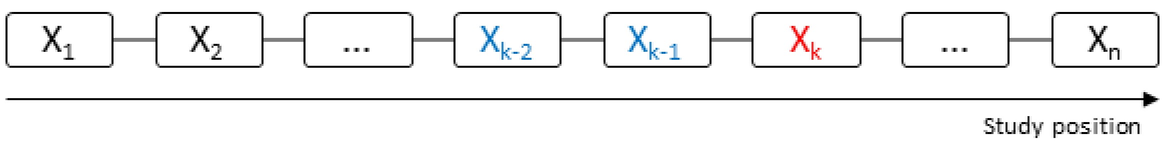

Figure 17. *Order of items during a study list*

Our theory makes several distinct predictions concerning the effect of the preceding items. As we explained before, memory performance for item X_k (denoted by $P(X_k)$) will be worse when the preceding item is weaker (e.g., LF vs. HF). This effect is also not discrete: $P(X_k)$ should be proportional to the strength of item X_{k-1}. These effects should also be cumulative such that $P(X_k)$ should be monotonically worse the more of the preceding items that are weak. This is because more resources will have been spent over time, reducing recovery. The effect of the preceding items X_{k-i} should further increase when the current item X_k is weaker, because the current item is in more need of resources. Finally, the effect of X_{k-1} should be stronger than the effect of X_{k-2} and in general the effect of the preceding items would decrease as the lag increases because more time would have passed, allowing for the recovery and subsequent depletion by other intervening items. These predictions, and the studies in which we found support for them are summarized in Table 5.

We do not expect to observe all of these effects in every study because the degree to which resources are depleted will depend on presentation rate, instructions and effort, nature of the material, power issues, etc. It will also depend on the study strategies used by the participants. For example, if for some reason participants decide to study primarily the HF items to maximize their performance, especially if they find the LF items too hard, then the effects are unlikely to appear, or at least be diminished. The discrete effect of prior item strength is the simplest prediction and likely will occur most often, especially when the current item is weak.

We present evidence that supports the predictions shown in Table 5 by reanalyzing eight published studies and by fitting a SAC model to each (see Appendix A for the modeling details). The most straightforward variable to analyze is word frequency, but the effects are likely to occur with any variable that affects how many resources are depleted in processing the preceding study items. Of the eight studies, four studies used word frequency as a factor (Cox et al, 2018; Diana & Reder, 2006; Reder et al, 2002; Ward et al, 2003), two studies manipulated the number of study repetition for each item (Buchler et al, 2008; Aue et al, 2017) and two directed-forgetting studies that manipulated whether each item should be remembered or forgotten (Marevic et al, 2017; Popov et al, 2018).[15]

Table 5. Predictions about the effects of prior item strength

#	Predictions	Studies							
		Diana *	**Ward**	**Buchler**	**Aue**	**Reder**	**Marevic**	**Popov**	**Cox**
1	**Discrete effect of prior item strength** *Example:* $P(X_k)$ is worse when X_{k-1} is LF	+	+	+	+	-	+	+	NA
2	**Continuous effect of prior item strength** *Example:* $P(X_k)$ is proportional to $freq(X_{k-1})$	NA	NA	+	NA	NA	NA	NA	+
3	**Cumulative effect of prior item strength** *Example:* $P(X_k)$ is worse when more of the preceding items are LF	+	-	+	-	+	+	+	NA
4	**Interaction between prior and current item strength** *Example:* The effect of $freq(X_{k-1})$ should be stronger when X_k is LF	+	+	+	+	-	-	-	NA
5	**Interaction between prior item strength and lag** *Example:* The effect of $freq(X_{k-i})$ should decrease as the lag i increases	+	-	+	-	-	+	+	+

Note. +: effects found in study, -: effect not found, NA: prediction could not be tested. Diana = Diana & Reder (2006), Ward = Ward et al (2003), Buchler = Buchler et al (2008), Aue = Aue et al (2017), Reder = Reder et al (2002), Marevic = Marevic et al (2017), Popov = Popov et al (2018), Cox = Cox et al (2018)

[15] We are grateful to all the researchers who shared their datasets with us.

1. Data analysis

For all results reported below we analyzed the accuracies and RTs via logistic and linear mixed-effects regression models (Baayen, Davidson, & Bates, 2008). We excluded incorrect responses from analyses of RTs. Random effects were determined through restricted likelihood ratio tests and all final models included varying intercepts for subjects and individual words/word pairs (i.e., different subjects and items differ in their overall accuracy and RT estimates). We inferred the significance of each effect based on likelihood ratio tests and AIC comparisons of the regression models that contained the effect in question with identical models that lacked this contrast. The modeling code, data and analyses scripts are available at https://github.com/venpopov/prior-item-effects.

2. Diana & Reder (2006) – worse memory for items that follow LF items during study

As we described previously, Diana & Reder (2006) performed an item recognition test for pictures that were studied with a HF or a LF word superimposed on it. The authors found that pictures paired with HF words were recognized better. Here we examined the probability of recognizing the picture as a function of whether the *preceding* item during study contained a HF or LF word (see Figure 18, left). There was no main effect of preceding word frequency (Δ AIC = 2, $\chi^2(1)$ = 0.31, p = .57), but there was a significant interaction between the frequency of the current word and the frequency of the preceding word during study (Δ AIC = -3, $\chi^2(1)$ = 5.20, p = .02). A picture was less likely to be recognized when, during the study list, it was preceded by a trial with a LF word, but only when the current word was also LF. This interaction pattern emerged in many of the other studies we reanalyzed as well. Furthermore, we looked at whether the effect was parametric – whether the number of preceding LF or HF words mattered. For each test trial, we calculated how many of the preceding study items were consecutively LF or HF. Picture recognition increased as the number of consecutive preceding LF words decreased from 4 to 1, and it increased as the number of consecutive preceding HF words increased from 1 to 4 (Figure 19; Δ AIC = -2, $\chi^2(1)$ = 3.62, p = .038). Finally, we looked at whether the effect of the prior item's frequency would decrease as the lag between it and the current item increased. The analyses shown in Figure 18, right panel, confirmed this expectation: The effect was biggest at lag 1 (odds ratio = 1.90, z = 2.216, p = .027), medium at lag 2 (odds ratio = 1.69, z = 1.78, p = .075), and smallest at lag 3 (odds ratio = 1.58, z = 1.57, p = .12). We could not test predictions 2 with this dataset, because word frequency did not vary continuously. Nevertheless, Predictions 1, 3, 4 and 5 were supported in the re-analysis of this study. Figure 18 and Figure 19 also show the fit of the SAC model, which provides further support for our explanation of these preceding item effects. The model captures well all effects we described by fitting a single recovery rate parameter and retrieval threshold and noise parameters.

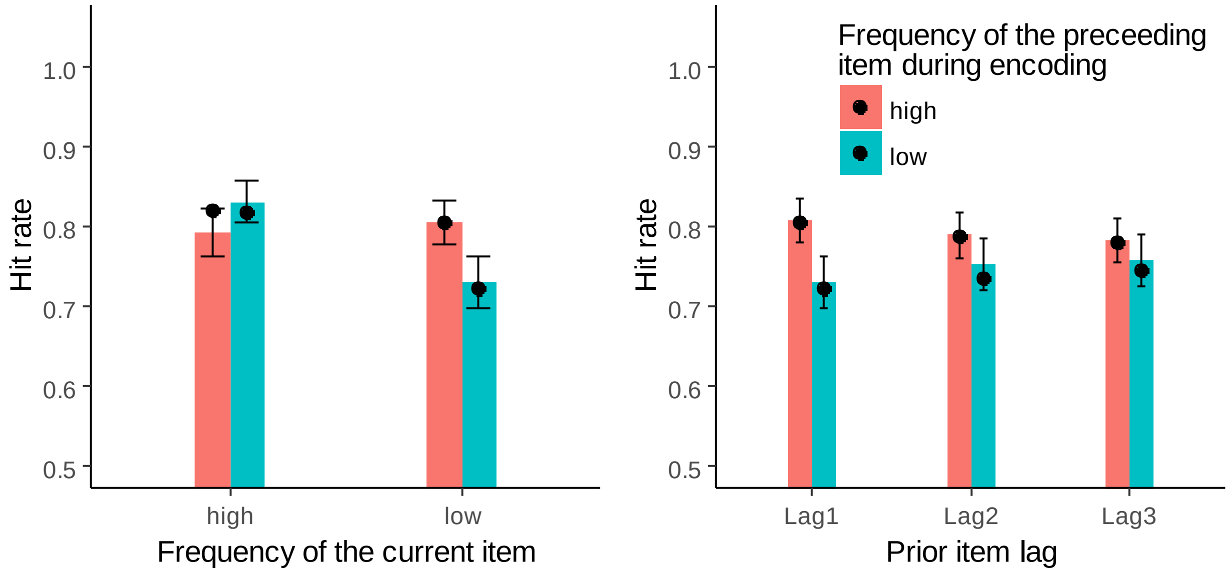

Figure 18. *Reanalyzed data from Diana & Reder (2006; bars) and SAC model fits (dots). Left – hit rate for pictures studied alongside a high or low frequency word, depending on whether they were preceded by by a high or low frequency word during study; Right – hit rate depending on whether they were preceded by a high or low frequency word during study at lag 1, 2 or 3;*

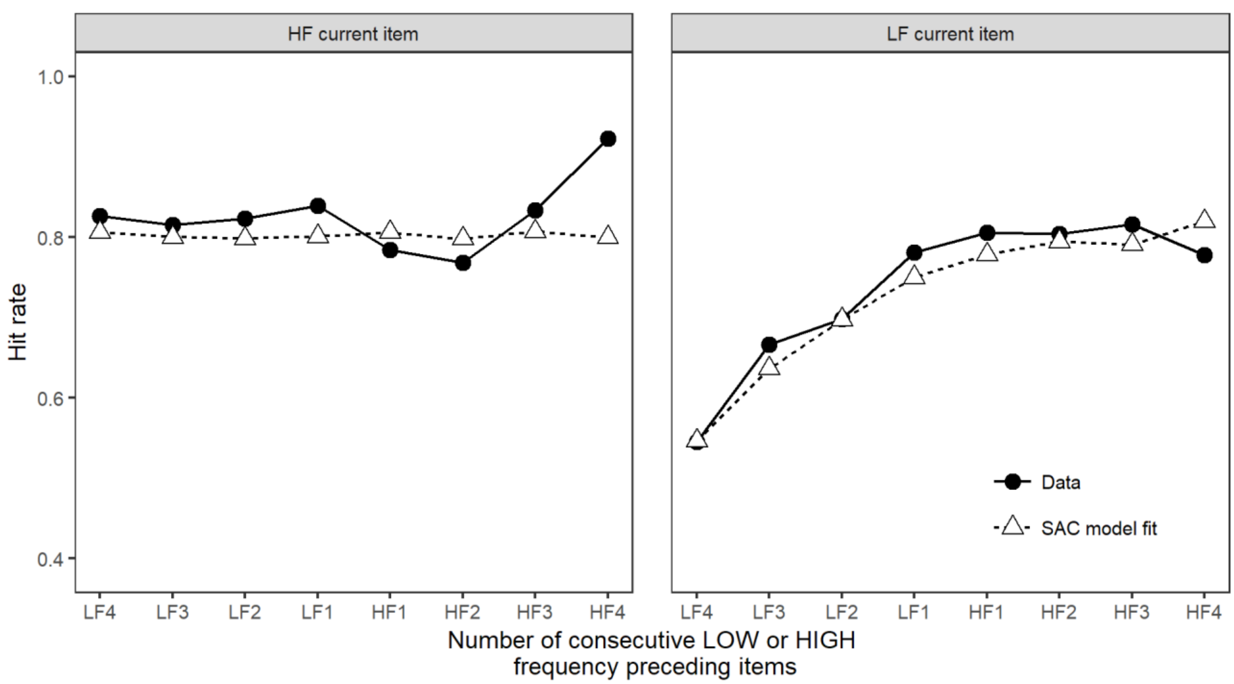

Figure 19. *Reanalyzed data from Diana & Reder (2006) and SAC model fits. hit rate for pictures as a function of how many previous pictures during the study list had low or high freqency words superimposed on them.*

3. Ward et al (2003) – rehearsal borrowing cannot explain the effect of preceding frequency

Can the prior item effects we found in Diana & Reder (2006) be explained by differences in rehearsal between item with HF and LF words? It might be the case that because LF words are more difficult to remember, participants continue to rehearse them when the subsequent item appears, which limits their ability to process and store the subsequent item. To address this question, we tested whether we would find the same effects in Ward et al (2003), who performed mixed-list free recall experiments in which words of LF and HF were intermixed and found that the memory impairment for LF words disappears in the mixed-lists. They also asked participants to rehearse words out loud, and they recorded the number of times each word was rehearsed. This allows us to test whether rehearsal borrowing can explain the effect of preceding item frequency. The reanalyzed results for Ward's Experiment 3 are shown in Figure 20a. Replicating our findings in the reanalysis of Diana & Reder (2006), the frequency of the items preceding the current item during the study list had a main effect on recall for the current item – words preceded by LF words during study were less likely to be recalled than words preceded by HF words (Δ AIC = -8, $\chi^2(1) = 9.56$, $p = .002$). As with the Diana & Reder (2006) analysis above, we also found an interaction between the frequency of the current item and that of the preceding item – the effect of preceding item frequency was present only when the current item was an LF word (Δ AIC = -3, $\chi^2(1) = 5.07$, $p = .024$). These results could not be accounted for by differential rehearsal – there were no differences in the number of rehearsal depending on the frequency of the prior study item ($p > .3$, Figure 20b). Thus, we found evidence for Predictions 1 and 4 with this dataset, and we established that rehearsal borrowing is not a likely explanation of the results. Figure 20 also shows the fit of the SAC model to the preceding item frequency effects (see Figure 12 for the mixed-vs-pure model fit of the same data).

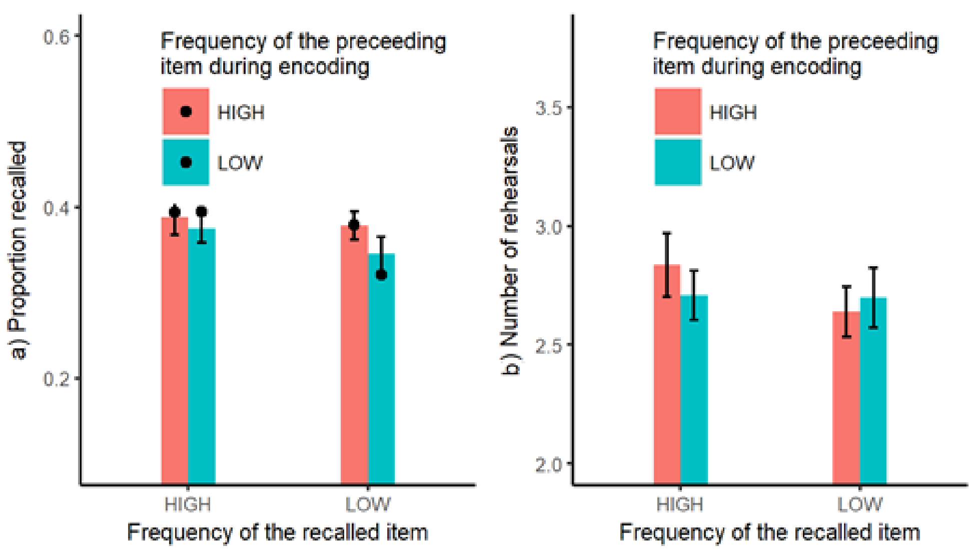

Figure 20. *Reanalysis of Ward et al (2003) Dots show the fit of the SAC model. a) proportion recall of high or low frequency words depending on whether they were preceded by by a high or low frequency word during study; b) number of rehearsal of high or low frequency words, depending on whether they were preceded by by a high or low frequency word during study*

4. Cox et al (2018) – continuous effect of frequency on memory for the following study item

Is the effect of preceding item frequency continuous? We explored this question by reanalyzing data from Cox et al (2018), who asked participants to perform five different memory tasks all using the same stimuli: item recognition, associative recognition, cued recall, free recall and lexical decision. We focus on the first four tasks, because the lexical decision task did not involve study lists. In all lists, participants studied words pairs without knowing the nature of the subsequent test. After studying a given list, they were given one of the four test described above. Word frequency was not a factor in this experiment, but the frequency of the words still varied, allowing us to test prediction #2, specifically that memory for the current item X_k will be affected continuously by the frequency of the prior word. The study items in this experiment were word pairs so we calculated the average frequency for a pair by using log SUBTLEX frequency (Brysbaert & New, 2009). The log frequency followed a normal distribution. Figure 22 shows the hit rate and the RTs for each memory task as a function of the preceding study item's frequency (binned in 20 frequency groups with equal number of observations). Performance in all tasks improved when the preceding item during study was of higher frequency – hit rates increased as a function of prior item frequency, $\Delta AIC = -5$, $\chi 2$ (1) = 6.63, $p = .01$. There was no significant interaction between prior item frequency and task type, $\Delta AIC = 5$, $\chi 2$ (3) = 0.99, $p = .80$. Finally, we found that the effect of prior item frequency decreased with the lag between the current item and the prior item during study, which was also predicted by our model (Figure 23a, b, respectively) – the effect was biggest at lag 1 (odds ratio for 1 log unit increase in frequency = 1.08, $z = 2.58$, $p = .009$), medium at lag 2 (odds ratio for 1 log unit increase in frequency = 1.03, $z = 0.86$, $p = .388$), and smallest at lag 3 (odds ratio for 1 log unit increase in frequency = 0.99, $z = -0.51$, $p = .61$). Thus, we found support for predictions 1, 2 and 5 by showing that the effect of prior item frequency is continuous and that its impact diminishes as the study lag between the current and the prior item increases.

We fit the SAC model to Cox et al's data. Since participants did not know what the nature of the test would be while they studied the word pairs, we fit a single model to all tasks. We fit only a single recovery rate, retrieval threshold and retrieval noise parameters for the whole model. As can be seen from Figure 21, the model fit the overall hit rates well across the four memory tasks. The different levels of performance were due to differences in the episodic node activation levels – the episodic node receives activation only from the context nodes during free recall, from the context node and the cue during cued recall, from the context node and the two associative cues during associative recognition. In single item recognition both the episode and the semantic node's activation contribute to performance, which results in the highest level of activation. Figure 22 and Figure 23 show the fits to the preceding item frequency effects. Figure 22 shows that the model captures well the slope of the frequency effects. The intercept for each task is not fit perfectly, which is due to the fact that we used a single retrieval threshold for all tasks and instead depended on inherent differences in activation levels between retrieval tasks to capture the overall hit rates. Figure 23 shows that the model also captures the fact that the preceding item effects decrease with lag, providing further support for the current theory.

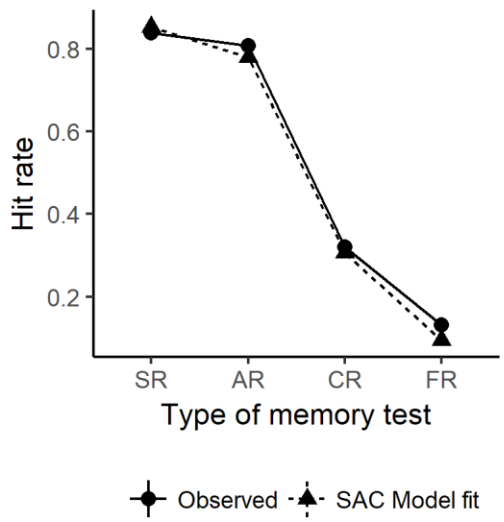

Figure 21. *Overall results in Cox et al (2018) and the SAC model fit to the four different test types. SR – single item recognition; AR – associative recognition; CR – cued recall; FR – free recall.*

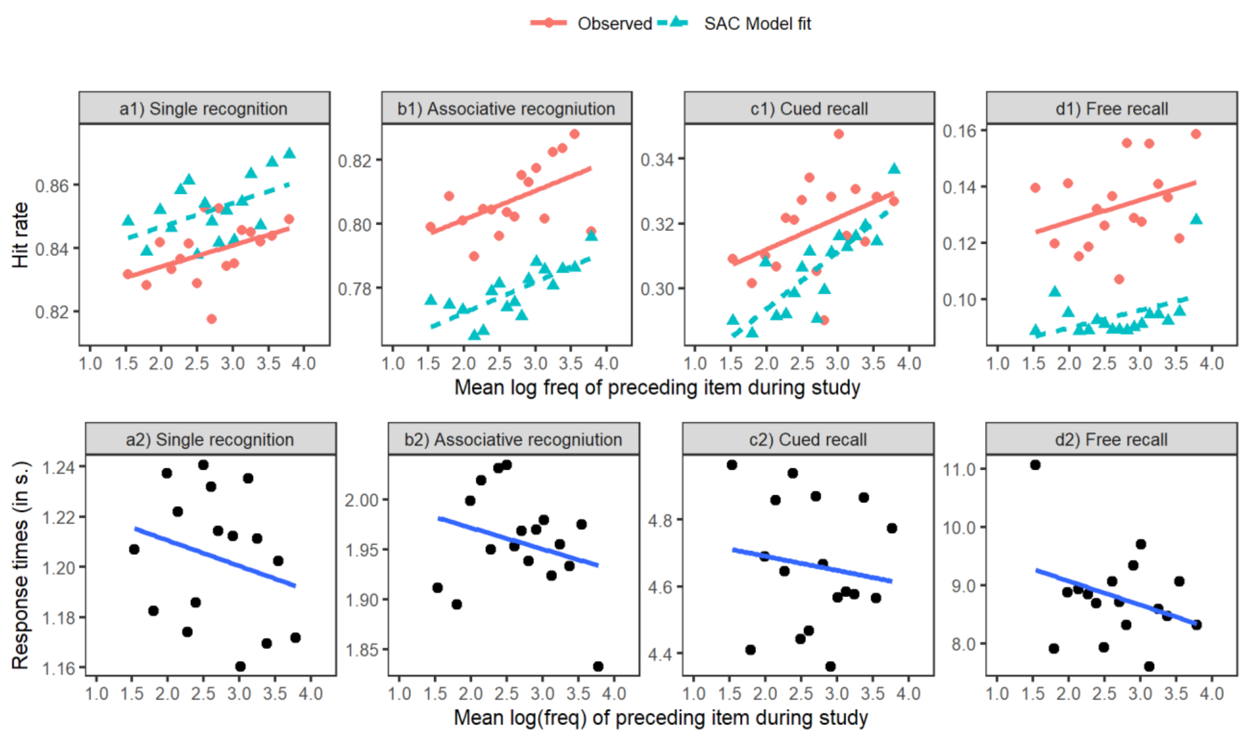

Figure 22. *Reanalysis of Cox et al (2018) and SAC model fits – Hit rates (top panels) and RTs for correct responses (bottom panels) as a function of the mean word frequency of the word pair that preceded the current pair during study and task type - a) Single recognition, b) Associative recognition, c) Cued recall, d) Free recall. Frequencies were binned into 20 bins of equal size and points represent the mean in each bin.*

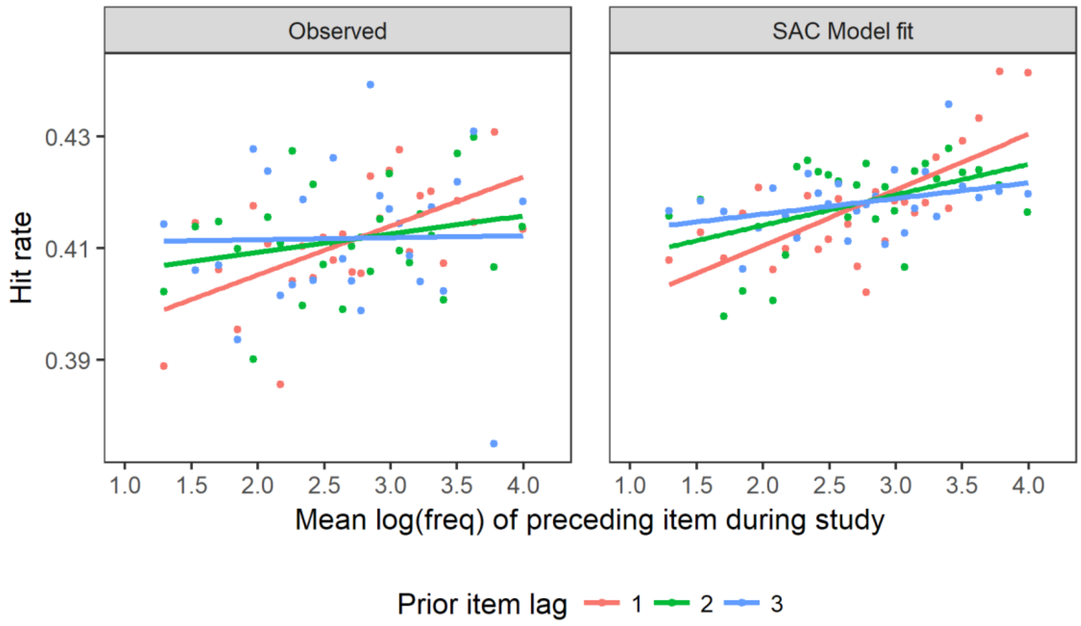

Figure 23. *Reanalysis of Cox et al (2018) and SAC model fit – Hit rate over all tasks depending on the frequency of the items that preceded the current item during study and their lag (e.g., how many trials earlier they occurred). Left – empirical data; Right – simulated data from the SAC model.*

5. Buchler et al (2008) – study repetitions of an item affect following items the same way as their normative frequency

The next two studies that we re-analyzed involved manipulations of item strength by varying the number of repetitions during study. According to our model, when some items are repeated multiple times on a study list their strength will be higher with each repetition and each subsequent encoding should require less resources. That means that there should be more resources available to store other items that follow those items that had been repeated multiple times.

Buchler et al (2008) performed an associative recognition experiment, in which they manipulated how many times a specific word pair was repeated (either 1 or 5) and whether words were repeated always with same word, or with 5 different words. We focus only on those pairs that stayed intact, either 1 or 5 times. This means that during study a trial could be preceded by a word pair that has appeared 1, 2, 3, 4 or 5 times already. First, we collapsed the study trials into two groups – we examined whether a word pair X_{k-1} appeared for the first time or had been repeated multiple times. Figure 24 show the results of this analysis. Hits for studied pairs were higher and false alarms to recombined pairs were lower when the target had been studied right after a repeated pair, Δ AIC = -2, $\chi2(1) = 3.92$, $p = .047$. This effect was continuous, such that hits were higher when the preceding trial during study contained a pair that was repeated more (Figure 25, left panel). Importantly, the effect of the prior item strength interacted with the current pairs strength: the advantage of following a strong pair was bigger when the current item was weak (studied only once), Δ AIC = -14, $\chi2(1) = 16.35$, $p < .001$. As we found in the previous analyses, the effect of a prior item's strength decreased as the lag between study positions increased – the effect was biggest at lag 1 (odds ratio = 1.68, z = 4.134, p < .001), medium at lag 2 (odds ratio = 1.11, z = 2.914, p = .004), and smallest at lag 3 (odds ratio = 0.94, z = 1.465, p = .143). Finally, performance was a function of how many of the preceding items during study were weak or strong (Figure 25, right). In summary, we found support for predictions 1, 2, 3, 4 and 5 with manipulated frequency just as we had with normative (natural) word frequency. The figures also show the fit of the SAC model, which accounted well for all of the effects.

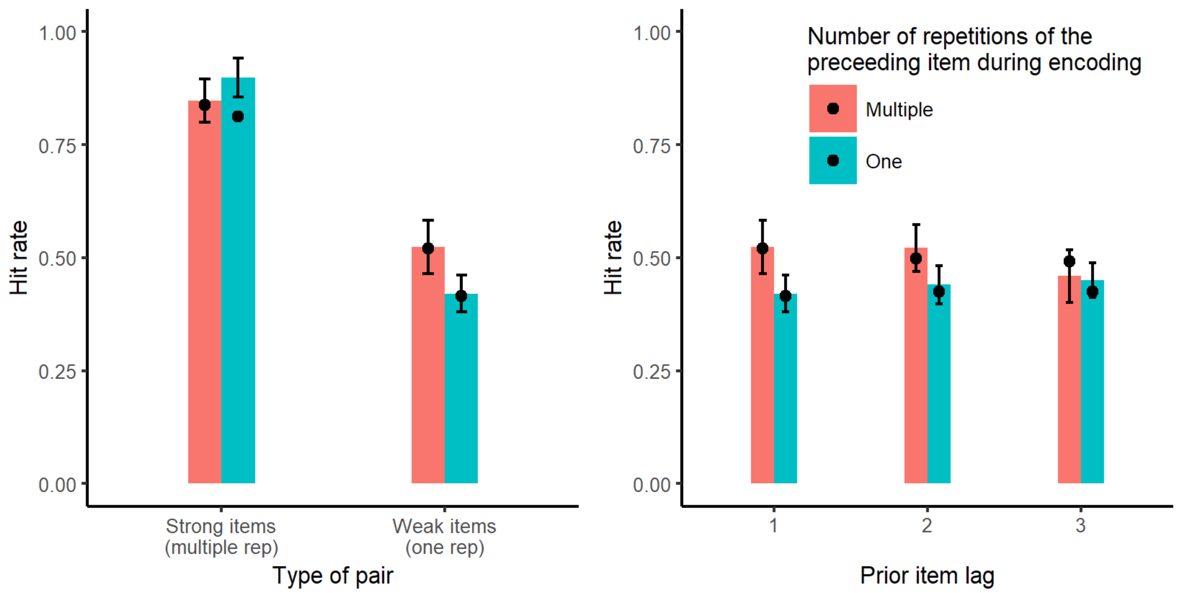

Figure 24. *Reanalysis of Buchler et al (2008) and SAC model fits (dots). Left panel – Hits for repeated and non-repeated items depending on whether they were preceded during study by a repeated or a non-repeated pair. Right panel - hit rate depending on whether the items were preceded by a repeated or a non-repeated pair during study at lag 1, 2 or 3.*

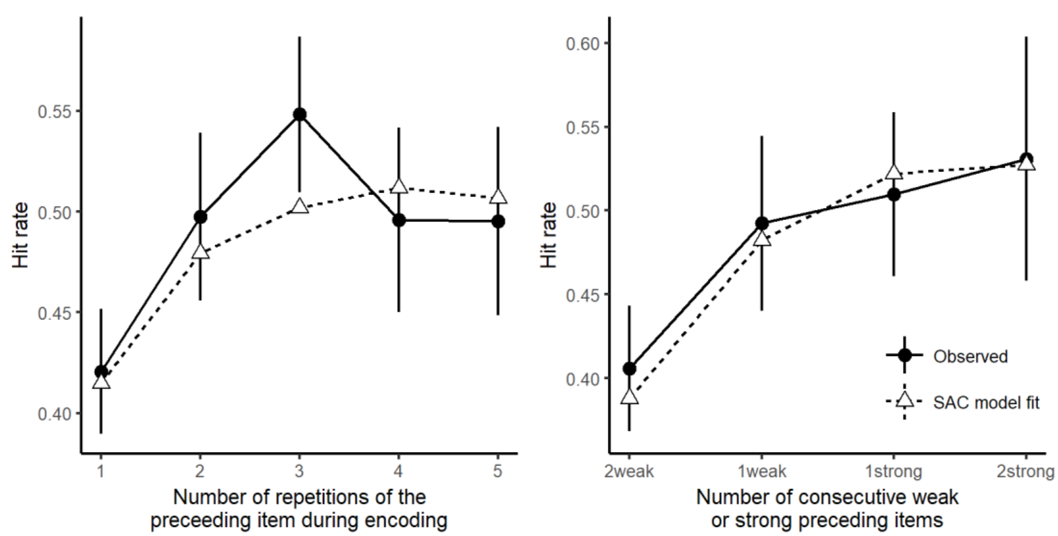

Figure 25. *Reanalysis of Buchler et al (2008) and SAC model fits. Left panel – hits for the current item depending on how many times the preceding item was repeated during study. Right panel – hits as a function of how many trials during the study list had weak (one repetition) or strong (multiple repetition) pairs preceding the test pair*

6. Aue et al (2017) – is repeating one word of a pair sufficient to facilitate memory for following study item?

Does the whole pair need to be repeated, in order for the following item to be better remembered? Our model would predict that even repeating only one word of the pair across pairs would cause less resource depletion, leaving more resources for the subsequent item (though the effect would be weaker). We explored this question by reanalyzing data from three experiments concerned with proactive facilitation (Aue, Criss, & Novak, 2017). In each of their cued-recall experiments participants studied two lists of word pairs and some of the pairs on the second list shared a word with pairs in the first list. The authors found that cued recall for List 2 was better for those pairs that share words with List 1 compared to pairs with cues unique to list 2. The authors argued, similar to our position, that when a cue was repeated from List 1 to List 2, it was more familiar, making it easier to associate to a new target, counteracting the effects of interference.

As in Buchler et al (2008), we looked at performance for the current pair depending on whether it was preceded during study by a pair that was presented for the first time (new), or whether it was preceded by a pair that contained a previously seen cue word from List 1. We combined the data from Experiments 1, 2 and 4 (Exp. 3 tested List 1 performance and was thus irrelevant). Figure 26 shows the results and the corresponding SAC model fit. Cued recall was higher when the pair was preceded during study by a pair with a repeated cue, but only when the current item was new, Δ AIC = -3, $\chi2(1) = 4.74$, p = .029. Likewise, confidence ratings for correct responses, which are a more continuous measure of memory strength, were also higher, Δ AIC = -2, $\chi2(1) = 3.99$, p = .046. Thus, we replicated the main effect we found in Buchler et al (2008). It is likely that the effects are weaker, because in this study only one of the words was repeated, not the entire pair, and it was repeated only one time, not up to five times as in Buchler et al (2008).

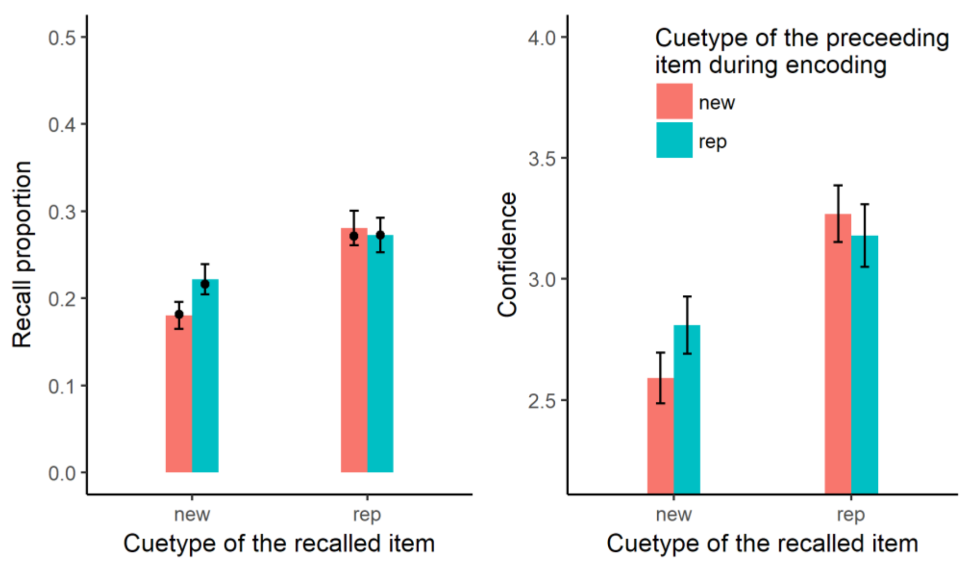

***Figure 26**. Reanalysis of Aue et al (2017) and SAC model fit (dots). Cued-recall probability (left) and confidence ratings (right) for pairs with a new or a repeated cue, depending on whether they were preceded during study by a pair with a new or a repeated cue.*

7. Reder et al (2002) – extending the predictions to experimentally manipulated frequency

Throughout this text, we brought up the benefits of making inferences from training studies that experimentally manipulate stimulus familiarity prior to the task of interest, rather than from studies that use normative word frequency (a quasi-experimental variable). Would the effects of the preceding item familiarity show up in such training studies as well? We reanalyzed data from Reder et al (2002), who trained people to learn pseudowords over a period of five weeks, and some pseudowords were exposed 6 times more often than others in free recall learning tasks. Once each week, throughout the training, people performed single item recognition tasks, where they studied a list containing some of the pseudowords, and then they had to discriminate the items studied on the list from others they were learning in the familiarity training generally. We performed the same analyses as in the other tasks, this time comparing memory for items that were preceded by pseudowords pre-trained at high or low frequency. There was evidence of a speed-accuracy trade-off (Reed, 1973), such that the responses to the conditions we expected to be worse were less accurate but faster. To finesse that interpretation problem, we report analyses done on inverse efficiency scores (RTs divided by average accuracy, separately for each condition; Townsend & Ashby, 1983). The main effect of the immediately preceding item was not significant (Δ AIC = 0, $\chi2(1) = 2.00$, $p = .15$), and there was no interaction with the current item frequency (Δ AIC = 2, $\chi2(1) = 0.36$, $p = .54$). However, as shown in Figure 27, there was a cumulative effect of the number of prior LF or HF items – pseudoword recognition increased as the number of consecutive preceding LF decreased from 3 to 1, and it increased as the number of consecutive preceding HF words increased from 1 to 3, (Δ AIC = -9, $\chi^2(1) = 10.85$, $p < .001$). Thus, even though the immediately preceding item effect was not significant, we found support for prediction 3 just as in Diana & Reder (2006).

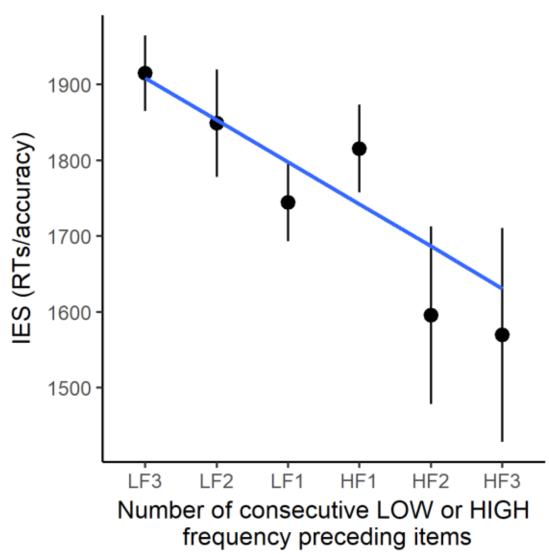

Figure 27. *Reanalysis of Reder et al (2002). Memory performance in inverse efficiency (IES = RTs/accuracy) as a function of how many trials during the study list had low or high freqency words preceding the test word*

8. Marevic et al (2017) – better memory after instructions to forget the preceding study item.

While the focus in thiss paper has mostly been on differences in strength due to word-frequency and experimental familiarization, the theory predicts that similar preceding item effects should occur with any variable that affects how strongly each item is processed and how much resources are depleted. To test this idea, we re-analyzed the results from a Directed Forgetting (DF) study (Marevic, Arnold, & Rummel, 2017). In the item-based version of the DF paradigm (for reviews, see Benjamin, 2006; Bjork, 1972; Epstein et al., 1972; MacLeod, 1998), participants view a sequence of items, each of which is followed either by a “forget” or by a “remember” cue. Participants are instructed that they will be tested only on remember items and that they should forget the items followed by a “forget” cue. The typical result shows that both recognition and recall memory is lower for the to-be-forgotten items, compared to the to-be-remembered items. The novel prediction from our theory is that memory will also be affected by whether the preceding item was given a remember or forget instruction: Specifically, items following a “remember” cue should be at a disadvantage because more WM resources will be consumed trying to learn the previous item compared to when the previous item was to be forgotten.

We tested this prediction with a reanalysis of a recent published study on item-based DF (Marevic, Arnold, & Rummel, 2017) and a follow up dual-task study (Popov, Marevic, Rummel & Reder, 2018). The full analysis and results are described in Popov et al (2018), and we will briefly summarize the results. In Marevic et al’s (2017) study, participants saw word pairs and each pair was followed either by a "to-be-remembered" (TBR) cue or by a "to-be-forgotten" (TBF) cue. The study list was followed by a free recall test of all pairs and then by a cued-recall test of all pairs. The analysis of preceding study item showed that cued-recall and free-recall paired-associate memory for items presented on study trial *k* was better when study trial *k-1* was TBF, rather than TBR. There was also a cumulative effect – when more of the previous trials (k-3, k-2, k-1) were to-be-remembered, memory for the item on trial k was worse. Finally, the effect of preceding item cue type decreased as the study lag between the current item and the prior item increased (e.g., the item on trial *k-3* had a weaker effect than that on trial *k-2*, which in turn had a weaker effect than the item on trial *k-1*). In summary, we found support for predictions 1, 3 and 5.

9. Popov et al (2018) – could the directed forgetting results be explained by rehearsal or attentional borrowing?

It is possible that when people study an item, they continuously rehearse or reactivate the memory traces for the preceding items (Barrouillet & Camos, 2001; Camos, Lagner, & Barrouillet, 2009; Loaiza, Duperreault, Rhodes, & McCabe, 2015; McFarlane & Humphreys, 2012), and that this rehearsal or attentional borrowing is greater the more previous items had to be remembered. Popov et al (2018) discounted those explanations with a dual-task paradigm, in which participants performed the paired-associate learning task in four conditions - a control single-task condition equivalent to Marevic et al (2017), a rehearsal suppression dual task condition, a divided attention dual task condition and a combined rehearsal suppression plus divided attention dual task condition. All of the effects we described were replicated, and they were present in all conditions – while divided attention and rehearsal suppression hurt overall memory performance, they did not remove the effect of the prior item cue type (see Popov et al, 2018, for more details). In summary, neither rehearsal nor attentional borrowing can account for the detrimental effect of having to remember the prior items, and the overall pattern is consistent with the claim that TBR items deplete more resources, which leads to less resources available for processing the subsequent item.

10. Discussion

We presented evidence from eight studies that support a novel prediction that memory for a given item is affected by the strength of the preceding items during the study phase. Specifically, memory for item X_k studied on trial *k* was better when the preceding item during study, X_{k-1}, was stronger and thus required less resources to be processed. This occurred when the preceding item was a high rather than a low frequency word or word pair (Diana & Reder, 2006; Ward et al, 2003; Cox et al, 2018), was a pseudo-word that was experienced more frequently in training prior to the study list (Reder et al, 2002), was a word or word pair that was repeated on multiple trials (Buchler et al, 2008) or lists (Aue et al, 2017), or was a word pair that was supposed to be forgotten, rather than remembered (Marevic et al, 2017; Popov et al, 2018). This effect was a parametric function of word frequency (Cox et al, 2018), it accumulated when more of the previous items during study were weaker (Diana & Reder, 2006; Buchler et al, 2008; Reder et al, 2002; Marevic et al, 2017; Popov et al, 2018), it decreased when the lag between the current and the prior item increased (Cox et al, 2018; Diana & Reder, 2006; Buchler et al, 2008; Marevic et al, 2017; Popov et al, 2018), and it was bigger when the current item was weaker itself (Buchler et al, 2008; Diana & Reder, 2006; Ward et al, 2003). These results are summarized in Table 5.

Importantly, these results cannot be explained by other mechanisms such as more rehearsing of the weaker preceding items during the processing of the current item. In the analysis of Ward et al (2003), who made participants rehearse out loud, we showed that the number of rehearsals did not differ as a function of the preceding item frequency, and the effect of the preceding item remained significant even after account for rehearsal frequency and recency in the regression model. In Popov et al (2018), we further showed that the effect remains even when rehearsal is suppressed or when attention is divided. In summary, we believe that these results provide strong support for the claim that memory formation depletes a limited pool of resources as a function of the current strength of items, and that these resources are not returned immediately, but rather recover over time.

V. General Discussion

We have presented a resource-depletion theory to account for many phenomena concerned with the effects of pre-existing item strength on learning and WM (summarized in Tables 1-4). We demonstrated that it is easier to store new episodic traces involving items with stronger current traces in LTM, such as high frequency words or frequently exposed novel stimuli. Such items are also easier to maintain in WM, and we discounted multiple alternative explanations for these results. Furthermore, we showed that the strength of one item interacts with surrounding items and provided evidence from pure-vs-mixed list comparisons and by reanalyzing 8 existing datasets that showed memory for one item depends on the strength of the items that immediately preceded it during study. The current theory accounts for these results by positing a limited resource that is used in processing, storage and maintenance, a resource that recovers slowly over time. This theory explains the encoding advantage for stronger items as a result of the fact that the encoding, binding and retrieval of stimuli deplete WM resources inversely proportional to the strength of the stimulus representation in long-term memory. Finally, a computational implementation of the theory provided good fits to many of the patterns discussed in the paper. In the remainder of this discussion we will consider some difficulties in providing evidence for the current proposal, describe a number of additional puzzling phenomena that the theory might explain, and we will conclude by discussing the concept of resources and its usefulness in explaining memory performance.

A. Partial matching's role in working memory resource depletion

A question that begs for an answer in the presentation of this theory is *why hasn't this theory been 'discovered' until now?* The simple yet obscure answer is that it is difficult to demonstrate evidence for this hypothesis experimentally. Quasi-experimental evidence consistent with the theory has existed for some time (reviewed here and also in Reder et al, 2007) and more new quasi-experimental evidence continues to support this view (Xie & Zhang, 2016, 2017b). However, experimental evidence was lacking until recently (Reder et al, 2016, Shen et al, 2017). The reason it is so difficult to demonstrate that working memory resource consumption is inversely related to the familiarity of the stimuli being processed is that people *naturally adapt* to limitations of working memory resources by using *partial matching* to finesse these limitations.

Partial matching is a psychological process that operates regularly in cognitive processing and especially when our working memory resources are exhausted. People can only attend to so much and they focus selectively, tending not to notice the rest. As a result, people can accept as a match information that only partially overlaps with the contents of memory. This phenomenon has been demonstrated in a variety of paradigms. It underlies the Moses Illusion where people respond with "Two" to the question "How many animals of each kind did Moses take in the Ark?", failing to notice that it was Noah, not Moses, who took the animals on the ark (Kamas, Reder, & Ayers, 1996; Reder & Kusbit, 1991). It is also a factor in Change Blindness where subjects fail to notice the change in the person they had just looked at while offering assistance (Simons & Levin, 1997) or in Feeling of Knowing in arithmetic tasks where the impression of knowing the answer occurs even when the operator of the math problem is swapped (Reder & Ritter, 1992; Schunn et al. 1997) as well as in *game-show* experiments where participants give spurious feelings of knowing the answers to trivia questions when terms in the questions had been primed together earlier (Reder, 1987). Partial matching is also observed when people accept foils that overlap in content with studied items (Diana, Peterson & Reder, 2004; Popov, Hristova, & Anders, 2017; Zhang, Walsh, & Anderson, 2017). All of these phenomena have the property that there tends to be too much information for a person to focus on all of it simultaneously. This is especially true when some of the information is not easily chunked into higher level units that are already well-known, which would reduce the WM load.

So *how* does partial matching make it difficult to demonstrate our theoretical position? In the training experiments we chose to use Chinese characters because they are relatively complex and were previously unknown to our participants. We expected that characters given much more exposure would perform better in our tasks than characters with less exposure but we failed to anticipate some of the heuristics people use to adapt to the challenges of unfamiliar items. For example, when characters are paired together and assigned to an English word for a cued-recall task, participants might only attend to one of the two characters. To counter this possibility, we forced them to attend to both by using every character in two different pairs. That correction was insufficient because participants could also partial match at the feature level within a character as well, finding features in characters that are already strong chunks. Figure 28 illustrates how someone might encode personally strong chunks such as the symbol π, the number *4* or the letter *k* in those unfamiliar stimuli instead of the entire characters.

Figure 28. *Examples of three Chinese characters and potential chunks (highlighted in grey) that might be encoded (π, 4 and k) and used for partial matching.*

When people only encode parts of a novel stimulus that involve previously known chunks, rather than (in this case) the entire Chinese character, there are fewer demands on the WM resource. However, this short-cut means that a complete representation of the character is not created and defeats the goal of controlling the familiarity of the stimulus. Nickerson and Adam's (1979) study showed that people who had used pennies every day of their lives for decades failed to discriminate the correct representation of a penny from plausible foils. Feigenbaum and Simon (1984) argued that people only encode those features required for discriminating among categories and the size and color of a penny are sufficient to distinguish it from other US coins. With this insight, we first tried to force discrimination by using ideographs with highly similar distractors. With promising results (Reder, 2011), we returned to natural stimuli (Chinese characters). We created visually similar sets of distractors for the visual search training materials.[16] When fine-grained discrimination between characters was forced in this way, the predicted effects finally emerged.

There are two strong reasons to believe that forcing discrimination is the key to finding an effect of frequency. One is that an otherwise identical version of the published Reder et al (2016) experiment contained a bug such that the assignment of distractor characters to the visual search task was random, rather than visually similar, and again there was no effect. The other strong support came from a post-hoc analysis of the Reder et al (2016) results in which we calculated the similarity between target and distractors for each set of character using vector representations described earlier in this paper (Xing & Li, 2004; Yang et al, 2009). We found that the more similar the character to its distractor set, the worse (slower and less accurate) the performance on the initial visual search trials; however, greater similarity facilitated learning and eventually lead to better performance for these targets in visual search, associative learning and N-back working memory performance (Popov & Reder, in prep). In summary, experimental manipulations of frequency require discrimination training with highly similar foils in order for the benefits of exposure frequency to emerge.

B. Accounting for additional phenomena

The theory put forward in this article provides plausible explanations for a few other surprising phenomena. In the following section we discuss how this theory can account for correlations between expertise and memory performance, WM capacity and long-term memory, age and memory decline, neuroscientific data for repetition priming and primacy effects.

1. Better WM memory in experts, ***even for random configurations of stimuli,*** *in their domain of expertise*

One classic result in the memory literature is that experts are much better at memorizing material within their domain of expertise, compared to non-experts (Chase & Simon, 1973; Chiesi, Spilich, & Voss, 1979) While initial studies showed that masters and novices do not differ in memorizing ability for random chess positions (Chase & Simon, 1973), later analyses revealed that there is a robust, albeit smaller, advantage of expertise even for random chess positions (Gobet & Simon,

[16] These were selected by Xiaonan Liu, a native Chinese speaker and Reder's grad student at the time

1996). A recent meta-analysis showed that this advantage for random configurations occurs in many other domains such as music, sports or computer programming (Sala & Gobet, 2017).

One explanation for these findings is that experts likely recognize smaller subsets of chunks even within randomly structured materials, and that these chunks allow them to memorize more of the information (see Gobet, 1998, for a theoretical review; Sala & Gobet, 2017). While this is certainly a possibility, an additional contributor might be the experts' greater familiarity with individual items from their domain of expertise, e.g., chess pieces and board positions, musical notes, programming commands, or climbing moves. An expert might exhaust less WM resources in binding these more familiar items into a random configuration compared to a novice with the same overall WM capacity. We demonstrated experimentally that this is the case *within-subjects* when they are trained to recognize Chinese characters at low and high-frequency (Reder et al, 2016). The same principle could be underlying, at least partially, the memory advantage for experts even with random configurations of materials.

2. Better LTM for people with more WM capacity

Given the exposition so far, perhaps it is not surprising that people with higher WM capacity also perform better on long-term episodic memory tests (Anguera et al., 2012; Anguera, Reuter-Lorenz, Willingham, & Seidler, 2010; Marevic, Arnold, & Rummel, 2017; Unsworth, Brewer, & Spillers, 2009; Unsworth & Spillers, 2010). Multinomial modeling has revealed that WM capacity correlates with encoding success, rather than with retrieval probability (Marevich et al, 2017). This correlation is one of the key puzzles that a theory of WM must explain, and Oberauer et al (2016) noted that no current resource-based account of WM can do so. The theory presented in this paper provides a natural account of for this observation – the creation and binding of novel episode nodes depletes a limited resource, and the strength of these nodes depends on how much resources are available. People with less WM capacity, either due to less total capacity, or due to slower recovery rate, would form weaker chunks, especially when the demands are greater.

3. Binding problems in old age

The current theory also provides an account of the correlation between WM capacity and learning performance evident in the effects of aging on both WM and episodic learning (Buchler & Reder, 2007; also see Buchler et al, 2011). There is overwhelming evidence that WM capacity decreases at a steady rate from about 20 years of age onwards (Brockmole & Logie, 2013; Salthouse & Babcock, 1991), and that this decline is greater for bindings than for item information (Cowan, Naveh-Benjamin, Kilb, & Saults, 2006; Peterson & Naveh-Benjamin, 2016, 2017). A large array of varied findings from item and associative recognition also indicates that the underlying cause of most age-related memory impairments is a decreased ability to form novel bindings (Ahmad, Fernandes, & Hockley, 2015; Buchler et al., 2011; Chalfonte & Johnson, 1996; Light, Patterson, Chung, & Healy, 2004). Our model suggests a causal link between the correlated decline in WM capacity and binding ability with age – forming new bindings requires WM resources and the decrease in WM capacity is responsible for the reduced ability to establish and store such bindings (Buchler et al, 2011). We propose that these decrements in WM and episodic learning are not simply comorbid; rather, the reduction in WM capacity causes episodic learning problems.

What makes us think that WM decline is related to the binding deficit? First, neuroimaging evidence is consistent with this claim. Activity in prefrontal areas involved in WM is reduced in old adults when binding is required, e.g., during the encoding of unfamiliar faces (Grady et al., 1995) and word pairs (Anderson et al., 2000; Cabeza et al., 1997), but not during encoding single items (Grady et al, 1995). Second, divided attention during encoding, which reduces the amount of resource available for memory operations, reduces or sometimes eliminates the age difference (Anderson et al., 2000; Craik & Byrd, 1982; Jennings & Jacoby, 1993), and it also leads to a reduction in prefrontal WM-related activity during encoding (Anderson et al., 2000; Iidaka, Anderson, Kapur, Cabeza, & Craik, 2000; Shallice et al., 1994). Finally, while the binding deficit in LTM for older adults has been studied for some time, recent studies are finding that, similar to LTM episodic memory, older adults' WM impairments are greater when they have to maintain item-item or item-context bindings (Peterson & Naveh-Benjamin, 2016, 2017). This overall pattern of results is consistent with the proposed connection between WM decline and reduced binding ability in old age.

4. Children, WM and second language acquisition

Another puzzling observation for which this theory can provide a plausible explanation is that children typically acquire a previously unfamiliar second language far more easily than their parents when they move to a new country. Children become more fluent and learn at a faster pace than their parents. A similar phenomenon that seems counter-intuitive is that younger children are much more adept at figuring out how to use "new-fangled" gadgets such as new computer devices and software compared to adults for whom these gadgets are very difficult to master. Adults often marvel at these feats of young children and consider their superior performance both surprising and somewhat amusing.

One reason that these phenomena have been viewed as humorous or counter-intuitive is based on the belief that children have less WM than adults (Brockmole & Logie, 2013; Nelson Cowan, Ricker, Clark, Hinrichs, & Glass, 2015). On the other hand, the theory put forward in this paper suggests that children might have as much or more WM resources than adults. Elsewhere it is cogently argued that children's performance on WM tasks is inferior to that of adults (Brockmole & Logie, 2013; Cowan et al., 2015) and that it improves gradually with age until young adulthood (Gathercole, Pickering, Ambridge & Wearing, 2004); however, if one accepts the proposition that WM consumption is affected by the number and strength of chunks that must be processed, it is clear why researchers would argue that children have less WM resources. The current theory, however, suggests it is possible that WM resources do not increase from childhood to adulthood but rather the size of the chunks and the strength of the chunks increase from childhood to adulthood such that less resources are required to perform the same task as children get older. In essence, we are proposing that the evidence used to argue for children having less WM is confounded with the notion we have pushed that WM consumption interacts with familiarity of chunks, not just number of chunks.

While the claim that children have at least as many WM resources is certainly controversial, a generally accepted related idea is that people can more easily process information when it is bound into higher level chunks, thereby having few chunks to maintain (Miller, 1956; Simon, 1974). It seems unlikely that anyone would dispute that as children mature, they form more higher-level chunks. The novel theoretical assertion put forward here is two-fold: first, it is easier to form higher level chunks when the lower chunks are stronger (Reder, et al., 2016); and second, higher level chunks will get stronger with more experience/exposure and thereby consume less of a limited WM resource pool. To put it a different way, we propose that there is a knowledge/WM trade-off: Young children have more WM but fewer high level chunks and their chunks are weaker from less cumulative exposure.

There is a growing body of evidence that is consistent with our theoretical position although the theoretical interpretations are different. One series of studies by Chi (1978) demonstrated that expert children chess players have a better memory for meaningful chess positions than adult chess experts, despite a worse WM span measured with neutral stimuli. Chi's interpretation was not that WM resources are consumed as an inverse function of familiarity but rather that WM is not involved in these tasks at all. Although her interpretation fits her results as well as our theoretical position, her interpretation does not fit the other patterns reviewed in this paper. Chi's interpretation is also inconsistent with a later replication of this phenomenon, which demonstrated in addition that the benefit in expert children's immediate recall memory was even greater for random configurations of chess stimuli (Schneider et al., 1993). Random configurations do not match existing chunks in memory, so the benefit cannot be explained by retrieving information from LTM; rather, greater familiarity with the individual pieces, similarly to high frequency words, would make it easier to bind them to each other, which results in the observed advantage in immediate memory for children.

The development of other processes can also lead erroneously to the impression that children have less WM. For example, word repetition and counting speed increase linearly with age through childhood, and research has shown that once these factors are controlled for, six-year-olds perform just as well as adults on word span and counting span tasks (Case, Kurland, & Goldberg, 1982) Another contributing factor is the way in which attentional processes change with development. Research on developmental reversals has suggested that, rather than having less attentional resources, younger children distribute their attention differently – globally to all features of the stimuli, rather than focusing on task-relevant information only (Deng & Sloutsky, 2016; Plebanek & Sloutsky, 2017) . As a result, children exhibit better processing and memory for task-irrelevant information in change-detection and visual search tasks (Plebaneck & Sloutsky, 2017) and they learn accidental category features better than adults (Deng & Sloutsky, 2016). As we discussed before, our model predicts that when resources are spread among a greater number of features, each feature is encoded proportionally to the resources attributed to it. Given these results, it is not surprising that children generally perform worse than adults in most tasks that test memory for task-relevant stimuli.

If, WM capacity does not increase from childhood to adulthood, then how would we explain the fact that developmental maturation of the prefrontal cortex correlates with the increase in WM performance (Diamond, 2002; Scherf, Sweeney & Luna, 2006)? As we noted, lower performance on WM tasks in young children might not reflect a smaller WM capacity, but a diminished ability to allocate attention to task-relevant stimuli. The studies on developmental reversals that we discuss have shown that while children might perform worse in change-detection and other WM tasks for task-relevant stimuli, they actually outperform adults if memory is tested for stimuli that they were supposed to ignore. So it might be the case that they distribute their capacity more broadly, resulting in what seems like reduced capacity when only task-relevant information is tested. Given that parts of the prefrontal cortex (e.g. DLPFC) have been shown to be involved in attention allocation, and specifically the ability to inhibit distractors (Diamond, 2002), the development of those areas might reflect the same thing that we argue from the behavioral data - development of the ability to direct attention, rather than the development of overall capacity.

Nevertheless, given the controversial nature of this claim, more research is needed to support it fully. As Cowan et al (2015) note, it is difficult to pinpoint the mechanisms responsible for developmental maturation because multiple cognitive processes are involved in every task and these develop in concert. In contrast to our proposal, Cowan et al (2015) argue that knowledge development cannot account for age differences in WM performance. They tested first, third and fifth grade children and college students on a change detection task with English letters and unfamiliar symbols. The results showed that performance increased linearly with age, even when the stimuli were unfamiliar to both groups. However, the same caveat about other processes applies to this result as well. For example, the array of three items to be remembered was presented for only 750 ms. Given that the speed of both voluntary and reflexive eye-movements increases dramatically until 25 years of age (Fischer, Biscaldi, & Gezeck, 1997), it is possible that the difference in performance is due to the fact that younger children could not encode as many items in the same time period. Thus, it is hard to accept this result as unambiguous evidence against the proposal put forward in this section.

This brings us to our original question: why is second language learning so much easier for children than adults? Some of the explanations offered are straightforward: children have more time on their hands to absorb the new language; however, total time on task does not explain their better learning (Johnson & Newport, 1989). Another has been the notion of a “critical period” for fluent language acquisition (Johnson & Newport, 1989; Lenneberg, 1967). While we do not discount these assumptions entirely, our explanation for this phenomenon is not based on either, but rather follows from the other axioms we have put forward. Children tend to process novel languages from the bottom-up – they tend to hear the phonemes and language structure before they feel compelled to speak in the new language. Given that they have more WM and they are not trying (initially) to map the sounds onto structures in their LTM, they can build up lower level chunks that are based on phonemes and syntactic structures that are different from their first language. Young children, with a lot of exposure to speech, without the burden on WM of trying to produce complex thoughts, can strengthen the low level phonemic and syntactic structures.

In contrast, adults tend to hear much more complex sentences that they try to parse and map onto their native tongue. Given our contention that adults do not have a larger WM pool, by parsing the new language, they are exhausting their resources and are much more prone to partial match the input with existing linguistic structures. Adults have larger and stronger chunks that make partial matching easier to do. Given the greater complexity and demands of the input they receive, partial matching is of greater necessity.

It is important to note that the veracity of the theory presented here does not rest on the claim that children might have just as much WM resources as adults do. We find this to be an interesting speculation that is suggested by the theory and that, as we reviewed above, is not inconsistent with current data. A weaker version of this claim, which suggests that the developmental increase in WM capacity has been overstated with current measurement methods, is possible as well.

C. Epilogue: The concept of resources as an explanation

At this point it might be good to ask the following question – how useful is the concept of resources in explaining the rich pattern of results we have reviewed? One could argue that resource-based explanations are circular – worse performance is attributed to having less resources, which on the face of it does not seem to be a satisfactory explanation. While we do agree that often resource-based verbal theories can suffer from circularity, we believe that this is not the case here. We have presented a mechanistic model with a precise formulation of resources, a model which allows us to make quantifiable predictions that go well beyond the generic less-resources-worse-performance claim. Global and local list-composition effects are one such example. The current theory posits that the resource does not recover immediately after use, but rather recovers slowly over time, which led to the prediction that word frequency and word repetition effects should be affected by the specific sequence of trials. At the global level, word frequency effects are attenuated when mixed in a single list, because the greater depletion of WM resources by LF words leaves less resources for processing the high frequency words and vice versa. At the local level, we predicted that the more low-frequency items are presented in a row, the worse the memory for the subsequent item would be.

The precise mechanistic formulation of a resource is not trivial, and involved making decisions about the underlying mathematical details. As we noted when introducing the model details, we had to decide what happens when you have to remember a display of several different items or features, and the remaining resources are not sufficient to encode them all. One possibility is that the system allocates the default proportion of resources to as many items as possible and fails to encode the remaining items. The choice we made for the model is that all k stimuli share the resources proportionally to their default cost. Whether the resource is spread amongst all items or whether a fixed amount is allocated for a limited number of items is currently under debate in the literature (Donkin, Nosofsky, Gold, & Shiffrin, 2013; Van den Berg, Awh, & Ma, 2014; W. Zhang & Luck, 2008). Slot-based theories posit that WM can actively maintain a limited number of distinct representations by allocating them to discrete units/slots (Donkin et al., 2013; Zhang & Luck, 2008). Others have argued, as we do here, and in Shen et al. (2017), that the resource can instead be flexibly allocated. In our view, the fact that WM performance varies as a function of item strength favors continuous resource theories.

Another aspect that would require further work is specifying the form of the resource recovery function. In the current work we chose a linear recovery rate, which worked well for the studies modeled here; nevertheless, there is no particular reason to choose a linear recovery rate. Within most simulations used in the paper, the recovery rate parameter is of such size that WM recovers fully within several seconds. Yet, we all experience growing fatigue over the course of the day and WM and LTM performance varies as a function of time of day (for a review, see Schmidt, Collette, Cajochen, & Peigneux, 2007). Thus, it is possible that there is a component of resource recovery that operates on a longer-time scale than the one presented here, just as there are multiple fuel sources and corresponding recovery rates for muscle contraction, which is the analogy we used in the introduction to this paper. Additional work is necessary to identify such components and to provide more data for constraining the possible functional forms of the resource recovery rate/s.

Finally, the current theory provides a mechanistic account for why additional learning benefits weaker memory traces more. As we noted in the theory description section, when items are restudied, stronger representations in the model are strengthened to a lesser degree than weaker representations, due to the delta learning function. On the surface, the specific function of the model is descriptive and not an explanation of why stronger items are strengthened less. The explanation for why this function is suitable comes from a rational perspective – every strengthening depletes resources proportional to the amount of strengthening. Since resources are limited, it is most optimal to spend less for processing stronger items, because they have a lower probability of being forgotten, and to reserve those resources for processes that need them more. We argue that, similarly to partial matching, this learning function has arisen as an adaptation to the challenge that there is typically much more information demanding our attention than our limited resources would allow us to process.

VI. Acknowledgements

We are grateful to Klaus Oberauer, Ed Awh, Nelson Cowan & Jason Sandor for fruitful discussions concerning many of the phenomena reported, and for suggesting some plausible alternative explanations. We thank John Anderson, Marc Coutanche, Markus Ostarek and Matthew So for commenting on previous drafts of the paper. We are grateful to the researchers who shared their data for the novel reanalyses reported in section IV.

VII. References

Ahmad, F. N., Fernandes, M., & Hockley, W. E. (2015). Improving associative memory in older adults with unitization. *Aging, Neuropsychology, and Cognition*, *22*(4), 452–472.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. L. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060.

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, *38*(4), 341–380.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, *79*(2), 97–123.

Anderson, J. R., Matessa, M., & Lebiere, C. (1997). ACT-R: A theory of higher level cognition and its relation to visual attention. *Human-Computer Interaction*, *12*(4), 439–462.

Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, *128*(2), 186.

Anderson, J. R., Reder, L. M., & Lebiere, C. (1996). Working memory: Activation limitations on retrieval. *Cognitive Psychology*, *30*(3), 221–256.

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

Anderson, N. D., Ebert, P. L., Jennings, J. M., Grady, C. L., Cabeza, R., & Graham, S. J. (2008). Recollection- and familiarity-based memory in healthy aging and amnestic mild cognitive impairment. *Neuropsychology*, *22*(2), 177–187.

Anderson, N. D., Iidaka, T., Cabeza, R., Kapur, S., McIntosh, A. R., & Craik, F. I. (2000). The effects of divided attention on encoding-and retrieval-related brain activity: A PET study of younger and older adults. *Journal of Cognitive Neuroscience*, *12*(5), 775–792.

Anguera, J. A., Bernard, J. A., Jaeggi, S. M., Buschkuehl, M., Benson, B. L., Jennett, S., … Seidler, R. D. (2012). The effects of working memory resource depletion and training on sensorimotor adaptation. *Behavioural Brain Research*, *228*(1), 107–115.

Anguera, J. A., Reuter-Lorenz, P. A., Willingham, D. T., & Seidler, R. D. (2010). Contributions of spatial working memory to visuomotor learning. *Journal of Cognitive Neuroscience*, *22*(9), 1917–1930.

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.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412.

Bahrick, H. P. (1979). Maintenance of knowledge: Questions about memory we forgot to ask. *Journal of Experimental Psychology: General*, *108*(3), 296.

Balota, D. A., Burgess, G. C., Cortese, M. J., & Adams, D. R. (2002). The word-frequency mirror effect in young, old, and early-stage Alzheimer's disease: Evidence for two processes in episodic recognition performance. *Journal of Memory and Language*, *46*(1), 199–226.

Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(5), 576.

Barrouillet, P., & Camos, V. (2001). Developmental increase in working memory span: Resource sharing or temporal decay? *Journal of Memory and Language*, *45*(1), 1–20.

Barry, T. J., Griffith, J. W., De Rossi, S., & Hermans, D. (2014). Meet the Fribbles: novel stimuli for use within behavioural research. *Frontiers in Psychology*, *5*.

Benjamin, A. S. (2006). The effects of list-method directed forgetting on recognition memory. *Psychonomic Bulletin & Review*, *13*(5), 831–836.

Bjork, R. A. (1972). Theoretical implications of directed forgetting. *Coding Processes in Human Memory*, 217–235.

Blumenfeld, R. S., Parks, C. M., Yonelinas, A. P., & Ranganath, C. (2010). Putting the Pieces Together: The Role of Dorsolateral Prefrontal Cortex in Relational Memory Encoding. *Journal of Cognitive Neuroscience*, *23*(1), 257–265.

Brockmole, J. R., & Logie, R. H. (2013). Age-Related Change in Visual Working Memory: A Study of 55,753 Participants Aged 8–75. *Frontiers in Psychology*, *4*.

Broek, G. S. E. van den, Segers, E., Takashima, A., & Verhoeven, L. (2014). Do testing effects change over time? Insights from immediate and delayed retrieval speed. *Memory*, *22*(7), 803–812.

Buchler, N. E. G., & Reder, L. M. (2007). Modeling age-related memory deficits: A two-parameter solution. *Psychology and Aging*, *22*(1), 104–121.

Buchler, N. G., Faunce, P., Light, L. L., Gottfredson, N., & Reder, L. M. (2011). Effects of Repetition on Associative Recognition in Young and Older Adults: Item and Associative Strengthening. *Psychology and Aging*, *26*(1), 111–126.

Buchler, N. G., Light, L. L., & Reder, L. M. (2008). Memory for items and associations: Distinct representations and processes in associative recognition. *Journal of Memory and Language*, *59*(2), 183–199.

Cabeza, R., Mangels, J., Nyberg, L., Habib, R., Houle, S., McIntosh, A. R., & Tulving, E. (1997). Brain regions differentially involved in remembering what and when: a PET study. *Neuron*, *19*(4), 863–870.

Camos, V., Lagner, P., & Barrouillet, P. (2009). Two maintenance mechanisms of verbal information in working memory. *Journal of Memory and Language*, *61*(3), 457–469.

Caplan, J. B., Madan, C. R., & Bedwell, D. J. (2015). Item-properties may influence item-item associations in serial recall. *Psychonomic Bulletin & Review*, *22*(2), 483–491.

Carrier, M., & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, *20*(6), 633–642.

Cary, M., & Reder, L. M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *Journal of Memory and Language*, *49*(2), 231–248.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*(3), 386–404.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354–380.

Chalfonte, B. l, & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition*, *24*(4), 403–416. https://doi.org/10.3758/BF03200930

Chalmers, K. A., & Humphreys, M. S. (2003). Experimental manipulation of prior experience: Effects on item and associative recognition. *Memory*, *11*(3), 233–246.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*(1), 55–81.

Chi, M. T. (1978). Knowledge structures and memory development. *Children's Thinking: What Develops*, *1*, 75–96.

Chiesi, H. L., Spilich, G. J., & Voss, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, *18*(3), 257–273.

Clark, S. E. (1992). Word frequency effects in associative and item recognition. *Memory & Cognition*, *20*(3), 231–243.

Clark, S. E., & Shiffrin, R. M. (1992). Cuing effects and associative information in recognition memory. *Memory & Cognition*, *20*(5), 580–598.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*(1), 87–114; discussion 114-185.

Cowan, N., Naveh-Benjamin, M., Kilb, A., & Saults, J. S. (2006). Life-span development of visual working memory: When is feature binding difficult? *Developmental Psychology*, *42*(6), 1089–1102.

Cowan, N., Ricker, T. J., Clark, K. M., Hinrichs, G. A., & Glass, B. A. (2015). Knowledge cannot explain the developmental growth of working memory capacity. *Developmental Science*, *18*(1), 132–145.

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., & Byrd, M. (1982). Aging and cognitive deficits. In *Aging and cognitive processes* (pp. 191–211). Springer.

Criss, A. H., Aue, W. R., & Smith, L. (2011). The effects of word frequency and context variability in cued recall. *Journal of Memory and Language*, *64*(2), 119–132.

Daily, L. Z., Lovett, M. C., & Reder, L. M. (2001). Modeling individual differences in working memory performance: A source activation account. *Cognitive Science*, *25*(3), 315–353.

Deese, J. (1960). Frequency of usage and number of words in free recall: The role of association. *Psychological Reports*, *7*, 337–344.

DeLosh, E., & Mcdaniel, M. (1996). The role of order information in free recall: Application to the word-frequency effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1136–1146

Deng, W. S., & Sloutsky, V. M. (2016). Selective attention, diffused attention, and the development of categorization. *Cognitive Psychology*, *91*, 24–62.

Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, *108*(2), 452.

DeWitt, M. R., Knight, J. B., Hicks, J. L., & Ball, B. H. (2012). The effects of prior knowledge on the encoding of episodic contextual details. *Psychonomic Bulletin & Review*, *19*(2), 251–257.

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: Cognitive functions, anatomy, and biochemistry. *Principles of frontal lobe function*, 466-503.

Diana, R. A., & Reder, L. M. (2005). The list strength effect: A contextual competition account. *Memory & Cognition*, *33*(7), 1289–1302.

Diana, R. A., & Reder, L. M. (2006). The low-frequency encoding disadvantage: Word frequency affects processing demands. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 805.

Diana, R. A., Reder, L. M., Arndt, J., & Park, H. (2006). Models of recognition: A review of arguments in favor of a dual-process account. *Psychonomic Bulletin & Review*, *13*(1), 1–21.

Donkin, C., Nosofsky, R. M., Gold, J. M., & Shiffrin, R. M. (2013). Discrete-slots models of visual working-memory response times. *Psychological Review*, *120*(4), 873–902.

Epstein, W., Massaro, D. W., & Wilder, L. (1972). Selective search in directed forgetting. *Journal of Experimental Psychology*, *94*(1), 18.

Feigenbaum, E. A., & Simon, H. A. (1984). EPAM-like Models of Recognition and Learning*. *Cognitive Science*, *8*(4), 305–336.

Fischer, B., Biscaldi, M., & Gezeck, S. (1997). On the development of voluntary and reflexive components in human saccade generation. *Brain Research*, *754*(1), 285–297.

Fiser, J., & Aslin, R. N. (2005). Encoding multielement scenes: statistical learning of visual feature hierarchies. *Journal of Experimental Psychology: General*, *134*(4), 521.

Gathercole, S. E., Pickering, S. J., Ambridge, B., & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Developmental psychology*, *40*(2), 177.

Geldorp, B. van, Parra, M. A., & Kessels, R. P. C. (2015). Cognitive and neuropsychological underpinnings of relational and conjunctive working memory binding across age. *Memory*, *23*(8), 1112–1122.

Gilbert, A. C., Boucher, V. J., & Jemel, B. (2014). Perceptual chunking and its effect on memory in speech processing: ERP and behavioral evidence. *Frontiers in Psychology*, *5*, 220.

Gilchrist, A. L. (2015). How should we measure chunks? A continuing issue in chunking research and a way forward. *Cognition*, 1456.

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

Glanzer, M., & Adams, J. K. (1990). The mirror effect in recognition memory: data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(1), 5.

Glenberg, A. M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior*, *15*(1), 1–16.

Gobet, F. (1998). Expert memory: a comparison of four theories. *Cognition*, *66*(2), 115–152.

Gobet, F., & Clarkson, G. (2004). Chunks in expert memory: evidence for the magical number four ... or is it two? *Memory (Hove, England)*, *12*(6), 732–747.

Gobet, F., Lane, P. C., Croker, S., Cheng, P. C., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, *5*(6), 236–243.

Gobet, F., & Simon, H. A. (1996). Recall of rapidly presented random chess positions is a function of skill. *Psychonomic Bulletin & Review*, *3*(2), 159–163.

Grady, C. L., McIntosh, A. R., Horwitz, B., Maisog, J. M., Ungerleider, L. G., Mentis, M. J., … Haxby, J. V. (1995). Age-related reductions in human recognition memory due to impaired encoding. *Science*, *269*(5221), 218–221.

Gregg, V. H., Montgomery, D. C., & Castano, D. (1980). Recall of common and uncommon words from pure and mixed lists. *Journal of Verbal Learning and Verbal Behavior*, *19*(2), 240–245.

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

Hintzman, D. L. (1994). On explaining the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(1), 201–205.

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.

Hockley, W. E. (1994). Reflections of the mirror effect for item and associative recognition. *Memory & Cognition*, *22*(6), 713–722.

Hulme, C. (2003). High- and low-frequency words are recalled equally well in alternating lists: Evidence for associative effects in serial recall. *Journal of Memory and Language*, *49*(4), 500–518.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology-Learning Memory and Cognition*, *23*(5), 1217–1232.

Humphreys, M. S., Maguire, A. M., McFarlane, K. A., Burt, J. S., Bolland, S. W., Murray, K. L., & Dunn, R. (2010). Using maintenance rehearsal to explore recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(1), 147–159.

Iidaka, T., Anderson, N. D., Kapur, S., Cabeza, R., & Craik, F. I. (2000). The effect of divided attention on encoding and retrieval in episodic memory revealed by positron emission tomography. *Journal of Cognitive Neuroscience*, *12*(2), 267–280.

Jennings, J. M., & Jacoby, L. L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology and Aging*, *8*(2), 283.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99.

Kamas, E. N., Reder, I. M., & Ayers, M. S. (1996). Partial matching in the Moses illusion: Response bias not sensitivity. *Memory & Cognition*, *24*(6), 687–699.

Lenneberg, E. H. (1967). The Biological Foundations of Language. *Hospital Practice*, *2*(12), 59–67.

Light, L. L., Patterson, M. M., Chung, C., & Healy, M. R. (2004). Effects of repetition and response deadline on associative recognition in young and older adults. *Memory & Cognition*, *32*(7), 1182–1193. https://doi.org/10.3758/BF03196891

Liu, X. L., Liang, P., Li, K., & Reder, L. M. (2014). Uncovering the Neural Mechanisms Underlying Learning from Tests: e92025. *PLoS One*, *9*(3).

Loaiza, V. M., Duperreault, K. A., Rhodes, M. G., & McCabe, D. P. (2015). Long-term semantic representations moderate the effect of attentional refreshing on episodic memory. *Psychonomic Bulletin & Review*, *22*(1), 274–280.

Lohnas, L. J., & Kahana, M. J. (2013). Parametric effects of word frequency effect in memory for mixed frequency lists. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *39*(6), 1943–1946.

Lovett, M. C., Daily, L. Z., & Reder, L. M. (2000). A source activation theory of working memory: Cross-task prediction of performance in ACT-R. *Cognitive Systems Research*, *1*(2), 99–118.

Lovett, M.C., Reder, L.M., & Lebiere, C. (1999). Modeling working memory in a unified architecture: An ACT-R perspective. In Miyake, A. and Shah, P. (Eds). *Models of Working Memory*. Cambridge University Press, 135-182.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.

MacLeod, C. M. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional Forgetting: Interdisciplinary approaches* (pp. 993–995). Mahwah, NJ: Lawrence Erlbaum Associates.

MacLeod, C. M., & Kampe, K. E. (1996). Word frequency effects on recall, recognition, and word fragment completion tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(1), 132.

Madan, C. R., Glaholt, M. G., & Caplan, J. B. (2010). The influence of item properties on association-memory. *Journal of Memory and Language*, *63*(1), 46–63.

Maddox, W. T., & Estes, W. K. (1997). Direct and indirect stimulus-frequency effects in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(3), 539.

Malmberg, K. J., & Murnane, K. (2002). List composition and the word-frequency effect for recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(4), 616–630.

Malmberg, K. J., & Nelson, T. O. (2003). The word frequency effect for recognition memory and the elevated-attention hypothesis. *Memory & Cognition*, *31*(1), 35–43.

Mandler, G., Goodman, G. O., & Wilkes-Gibbs, D. L. (1982). The word-frequency paradox in recognition. *Memory & Cognition*, *10*(1), 33–42.

Marevic, I., Arnold, N. R., & Rummel, J. (2017). Item-Method Directed Forgetting and Working Memory Capacity: A Hierarchical Multinomial Modeling Approach. *The Quarterly Journal of Experimental Psychology*, 1–34.

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), 724–760.

McFarlane, K. A., & Humphreys, M. S. (2012). Maintenance rehearsal: The key to the role attention plays in storage and forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 1001–1018.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97.

Morin, C., Poirier, M., Fortin, C., & Hulme, C. (2006). Word frequency and the mixed-list paradox in immediate and delayed serial recall. *Psychonomic Bulletin & Review*, *13*(4), 724–729.

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

Nelson, A. B., & Shiffrin, R. M. (2013). The co-evolution of knowledge and event memory. *Psychological Review*, *120*(2), 356–394.

Nickerson, R. S., & Adams, M. J. (1979). Long-term memory for a common object. *Cognitive Psychology*, *11*(3), 287–307.

Nimmo, L. M., & Roodenrys, S. (2002). Syllable frequency effects on phonological short-term memory tasks. *Applied Psycholinguistics*, *23*(04), 643–659.

Oberauer, K., Farrell, S., Jarrold, C., & Lewandowsky, S. (2016). What limits working memory capacity? *Psychological Bulletin*, *142*(7), 758–799.

Osth, A. F., & Dennis, S. (2015). Sources of interference in item and associative recognition memory. *Psychological review*, *122*(2), 260.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: a meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping*, *25*(1), 46–59.

Ozubko, J. D., & Joordens, S. (2007). The mixed truth about frequency effects on free recall: Effects of study list composition. *Psychonomic Bulletin & Review*, *14*(5), 871–876.

Palmer, S. E. (1977). Hierarchical structure in perceptual representation. *Cognitive Psychology*, *9*(4), 441–474.

Pavlik, P. I., & Anderson, J. R. (2005). Practice and Forgetting Effects on Vocabulary Memory: An Activation-Based Model of the Spacing Effect. *Cognitive Science*, *29*(4), 559–586.

Perruchet, P., & Vinter, A. (1998). PARSER: A model for word segmentation. *Journal of Memory and Language*, *39*(2), 246–263.

Peterson, D. J., & Naveh-Benjamin, M. (2016). The role of aging in intra-item and item-context binding processes in visual working memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *42*(11), 1713–1730.

Peterson, D. J., & Naveh-Benjamin, M. (2017). The role of attention in item-item binding in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(9), 1403.

Plebanek, D. J., & Sloutsky, V. M. (2017). Costs of Selective Attention: When Children Notice What Adults Miss. *Psychological Science*, *28*(6), 723–732.

Popov, V., Hristova, P., & Anders, R. (2017). The relational luring effect: Retrieval of relational information during associative recognition. *Journal of Experimental Psychology: General*, *146*(5), 722.

Popov, V., Marevic, I., Rummel, J., & Reder, L. (2018, June 6). Forgetting is a Feature, not a Bug: Intentionally Forgetting Some Things Helps Us Remember Others by Freeing up Working Memory Resources. *PsyArXiv. https://doi.org/10.17605/osf.io/ysjzu R*etrieved from psyarxiv.com/ysjzu

Rao, K. V., & Proctor, R. W. (1984). Study-phase processing and the word frequency effect in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(3), 386.

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology*, *19*(1), 90–138.

Reder, L. M. (2011). Familiarity of elements affects knowledge formation. *Talk presented at Tenth Annual Summer Interdisciplinary Conference, Spain.*

Reder, L. M., Angstadt, P., Cary, M., Erickson, M. A., & Ayers, M. S. (2002). A reexamination of stimulus-frequency effects in recognition: Two mirrors for low- and high-frequency pseudowords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(1), 138–152.

Reder, L. M., & Kusbit, G. W. (1991). Locus of the Moses illusion: Imperfect encoding, retrieval, or match? *Journal of Memory and Language*, *30*(4), 385–406.

Reder, L. M., Liu, X. L., Keinath, A., & Popov, V. (2016). Building knowledge requires bricks, not sand: The critical role of familiar constituents in learning. *Psychonomic Bulletin & Review*, *23*(1), 271–277.

Reder, L. M., Nhouyvanisvong, A., Schunn, C. D., Ayers, M. S., Angstadt, P., & Hiraki, K. (2000). A mechanistic account of the mirror effect for word frequency: A computational model of remember–know judgments in a continuous recognition paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(2), 294.

Reder, L. M., Paynter, C., Diana, R. A., Ngiam, J., & Dickison, D. (2007). Experience is a Double-Edged Sword: A Computational Model of The Encoding/Retrieval Trade-Off With Familiarity. In *Psychology of Learning and Motivation* (Vol. 48, pp. 271–312). Elsevier.

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(3), 435.

Reder, L. M., & Schunn, C. D. (1996). Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. Retrieved from http://d-scholarship.pitt.edu/22852/1/licence.txt

Reder, L. M., Victoria, L. W., Manelis, A., Oates, J. M., Dutcher, J. M., Bates, J. T., … Gyulai, F. (2013). Why It's Easier to Remember Seeing a Face We Already Know Than One We Don't Preexisting Memory Representations Facilitate Memory Formation. *Psychological Science*, *24*(3), 363–372.

Reed, A. V. (1973). Speed-Accuracy Trade-Off in Recognition Memory. *Science*, *181*(4099), 574–576.

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. *Classical Conditioning II: Current Research and Theory*, *2*, 64–99.

Rhodes, M. G., Castel, A. D., & Jacoby, L. L. (2008). Associative recognition of face pairs by younger and older adults: The role of familiarity-based processing. *Psychology and Aging*, *23*(2), 239–249.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255.

Roodenrys, S., Hulme, C., & Brown, G. (1993). The development of short-term memory span: Separable effects of speech rate and long-term memory. *Journal of Experimental Child Psychology*, *56*(3), 431–442.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: A meta-analytic review of the testing effect. *Psychological Bulletin*, *140*(6), 1432–1463.

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review*, *103*(4), 734–760.

Sala, G., & Gobet, F. (2017). Experts' memory superiority for domain-specific random material generalizes across fields of expertise: A meta-analysis. *Memory & Cognition*, *45*(2), 183–193.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology*, *27*(5), 763–776. https://doi.org/10.1037/0012-1649.27.5.763

Scherf, K. S., Sweeney, J. A., & Luna, B. (2006). Brain basis of developmental change in visuospatial working memory. *Journal of cognitive neuroscience*, *18*(7), 1045-1058.

Schmidt, C., Collette, F., Cajochen, C., & Peigneux, P. (2007). A time to think: Circadian rhythms in human cognition. *Cognitive Neuropsychology*, *24*(7), 755–789.

Schneider, D. W., & Anderson, J. R. (2012). Modeling fan effects on the time course of associative recognition. *Cognitive Psychology*, *64*(3), 127–160.

Schneider, W., Gruber, H., Gold, A., & Opwis, K. (1993). Chess Expertise and Memory for Chess Positions in Children and Adults. *Journal of Experimental Child Psychology*, *56*(3), 328–349.

Schulman, A. I. (1967). Word length and rarity in recognition memory. *Psychonomic Science*, *9*(4), 211–212.

Schulman, A. I. (1976). Memory for rare words previously rated for familiarity. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(3), 301.

Schunn, C. D., Reder, L. M., Nhouyvanisvong, A., Richards, D. R., & Stroffolino, P. J. (1997). To calculate or not to calculate: A source activation confusion model of problem familiarity's role in strategy selection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(1), 3.

Shallice, T., Fletcher, P., Frith, C. D., Grasby, P., Frackowiak, R. S. J., & Dolan, R. J. (1994). Brain regions associated with acquisition and retrieval of verbal episodic memory. *Nature*, *368*(6472), 633–635

Shen, Z., Popov, V., Delahay, A. B., & Reder, L. M. (2017). Item strength affects working memory capacity. *Memory & Cognition*.

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

Simon, H. A. (1974). How big is a chunk? *Science*, *183*(4124), 482–488.

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in cognitive sciences*, *1*(7), 261-267.

Sumby, W. H. (1963). Word frequency and serial position effects. *Journal of Verbal Learning and Verbal Behavior*, *1*(6), 443–450.

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(6), 1589.

Tarr, M. (2018). *Novel Objects (Online)* (Accessed June 6, 2018). Available online at: http://wiki.cnbc.cmu.edu/Novel_Objects

Tehan, G., & Humphreys, M. S. (1988). Articulatory loop explanations of memory span and pronunciation rate correspondences: a cautionary note. *Bulletin of the Psychonomic Society*, *26*, 293–296.

Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. CUP Archive.

Tulving, E., & Patkau, J. E. (1962). Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. *Canadian Journal of Psychology/Revue Canadienne de Psychologie, 16*(2), 83.

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2009). There's more to the working memory capacity—fluid intelligence relationship than just secondary memory. *Psychonomic Bulletin & Review*, *16*(5), 931–937.

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, *62*(4), 392–406.

Van den Berg, R., Awh, E., & Ma, W. J. (2014). Factorial comparison of working memory models. *Psychological Review*, *121*(1), 124.

Ward, G., Woodward, G., Stevens, A., & Stinson, C. (2003). Using overt rehearsals to explain word frequency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*(2), 186.

Watkins, M. J. (1977). The intricacy of memory span. *Memory & Cognition*, *5*(5), 529–534.

Watkins, M. J., LeCompte, D. C., & Kim, K. (2000). Role of study strategy in recall of mixed lists of common and rare words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(1), 239.

Wixted, J. T. (1992). Subjective memorability and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(4), 681.

Wixted, J. T., & Mickes, L. (2010). A continuous dual-process model of remember/know judgments. *Psychological Review*, *117*(4), 1025–1054. https://doi.org/10.1037/a0020874

Xie, W., & Zhang, W. (2016). Familiarity increases the number of remembered Pokémon in visual short-term memory. *Memory & Cognition*. https://doi.org/10.3758/s13421-016-0679-7

Xie, W., & Zhang, W. (2017). Familiarity Speeds Up Visual Short-Term Memory Consolidation. *Journal of Experimental Psychology: Human Perception and Performance*.

Xing, H. B., & Li, P. (2004). The acquisition of Chinese characters: Corpus analyses and connectionist simulations. *Journal of Cognitive Science*, *5*(1), 1–49.

Yang, J., McCandliss, B. D., Shu, H., & Zevin, J. D. (2009). Simulating Language-specific and Language-general Effects in a Statistical Learning Model of Chinese Reading. *Journal of Memory and Language*, *61*(2), 238–257.

Yonelinas, A. P. (2002). The Nature of Recollection and Familiarity: A Review of 30 Years of Research. *Journal of Memory and Language*, *46*(3), 441–517.

Zechmeister, E. B., Curt, C., & Sebastian, J. A. (1978). Errors in a recognition memory task are a U-shaped function of word frequency. *Bulletin of the Psychonomic Society*, *11*(6), 371–373.

Zhang, Q., Walsh, M. M., & Anderson, J. R. (2017). The Effects of Probe Similarity on Retrieval and Comparison Processes in Associative Recognition. *Journal of Cognitive Neuroscience*, *29*(2), 352–367.

Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235.

VIII. Appendixes

A. Appendix A. Modeling details

The SAC model is a process model that takes a sequence of trials that is given to a participant and performs each operation on a trial-by-trial basis. This results in an episodic and semantic activation value for each test trial, which are converted into response probabilities according to section II.A.5. Memory Retrieval. We fit the model to each study by generating such response probabilities for each trial and each participant, then summarizing the response probabilities over all subjects and separately for each condition of interest. The parameters of the model were optimized to reduce the root mean squared error (RMSE) between the summarized model predictions and the observed data. The optimization was performed using the downhill simplex algorithm as implemented in Python's Scipy library. We kept most of the parameters fixed across experiments. The decay rates for the base-level strength and current activation of the nodes and the strength of the links were imported from the previous version of SAC (Reder et al. 2000). The default delta learning rate and the parameters that convert word frequency into prior base-level strength were estimated anew since the learning function differs from the one implemented in the previous model versions. To estimate it, we fit the SAC model to Reder et al's (2000) data, who performed a continuous single item recognition experiment, in which they manipulated the number of repetitions for each word and the words' normative frequency. Figure A1 shows the proportion of Remember and Know responses as a function of presentation number and frequency, as well as the SAC model fit. The prior base-level strength for the lowest frequency words was estimated to be 0.2, and that for the highest frequency words was 0.4, with intermediate frequencies linearly extrapolated between those two values. The optimal learning rate was 0.8. These values were used as defaults in all subsequent simulations. Table A1 describes all model parameters and Table A2 summarized the parameter estimates for the 14 simulations presented in the paper. The modeling code, data and analyses scripts are available at https://github.com/venpopov/prior-item-effects.

Table A1 Description of SAC parameters

Model parameter	Description
d_n	Power decay rate for node base-level strength
d_l	Power decay rate for link strength
y	Exponential decay rate for current activation
δ	Learning rate for base-level strength
W	Total WM resource capacity
w_r	WM recovery rate
$\boldsymbol{\Theta}_{epi}$	Retrieval threshold for episodic nodes
$\boldsymbol{\sigma}_{epi}$	Standard deviation of the noise added to episodic activation
$\boldsymbol{\Theta}_{sem}$	Retrieval threshold for semantic nodes
$\boldsymbol{\sigma}_{sem}$	Standard deviation of the noise added to semantic activation

Table A2 Parameter estimates for SAC models

#	Study	dn	dl	γ	δ	W	w_r	Θ_{epi}	σ_{epi}	Θ_{sem}	σ_{sem}	rmse
1	Reder et al (2000), Exp. 1	-0.18	-0.12	0.2	**0.8**	**3.000**	**0.750**	**0.398**	**0.369**	**0.506**	**0.139**	0.034
2	Clark (1992)	-0.18	-0.12	0.2	0.8	3.000	**0.663**	**0.412**	**0.210**	**0.391**	**0.180**	0.027
3	Malmberg & Nelson (2003), Exp. 2	-0.18	-0.12	0.2	0.8	3.000	**0.864**	**0.101**	**0.263**	-	-	0.014
4	Criss et al (2011), Exp. 1 & 2	-0.18	-0.12	0.2	0.8	3.000	**0.584**	**0.333**	**0.304**	-	-	0.016
5	Criss et al (2011), Exp. 3	-0.18	-0.12	0.2	0.8	3.000	0.584	**0.332**	**0.207**	-	-	0.030
6	Ward et al (2003), Exp. 3	-0.18	-0.12	0.2	0.8	**2.800**	**0.450**	**0.527**	**0.356**	-	-	0.013
7	Hulme et al (1997), Exp. 2	-0.18	-0.12	0.2	0.8	**5.871**	**0.215**	**0.383**	**0.539**	-	-	0.031
8	Hulme et al (2003), Exp. 1 & 2	-0.18	-0.12	0.2	0.8	**3.837**	**0.330**	**0.572**	**0.469**	-	-	0.026
9	Hulme et al (2003), Exp. 3	-0.18	-0.12	0.2	0.8	**6.832**	**0.827**	**0.463**	**0.390**	-	-	0.036
10	Malmberg & Nelson (2003), Exp. 3	-0.18	-0.12	0.2	0.8	**5.000**	**0.966**	**0.270**	**0.054**	-	-	0.025
11	Diana & Reder (2006), Exp. 1	-0.18	-0.12	0.2	0.8	3.000	**1.114**	**0.091**	**0.020**	-	-	0.035
12	Cox et al (2018)	-0.18	-0.12	0.2	0.8	3.000	**0.970**	**0.143**	**0.067**	-	-	0.024
13	Buchler et al (2008), Exp. 1	-0.18	-0.12	0.2	0.8	3.000	**0.850**	**0.134**	**0.358**	-	-	0.026
14	Aue et al (2017), Exp. 1,2 & 4	-0.18	-0.12	0.2	0.8	3.000	**0.520**	**0.332**	**0.219**	-	-	0.007

Note: **bold-underlined** parameters were free to vary in estimating the model. The remaining parameters were fixed. Dn, dl, and γ were imported from the prior version of SAC

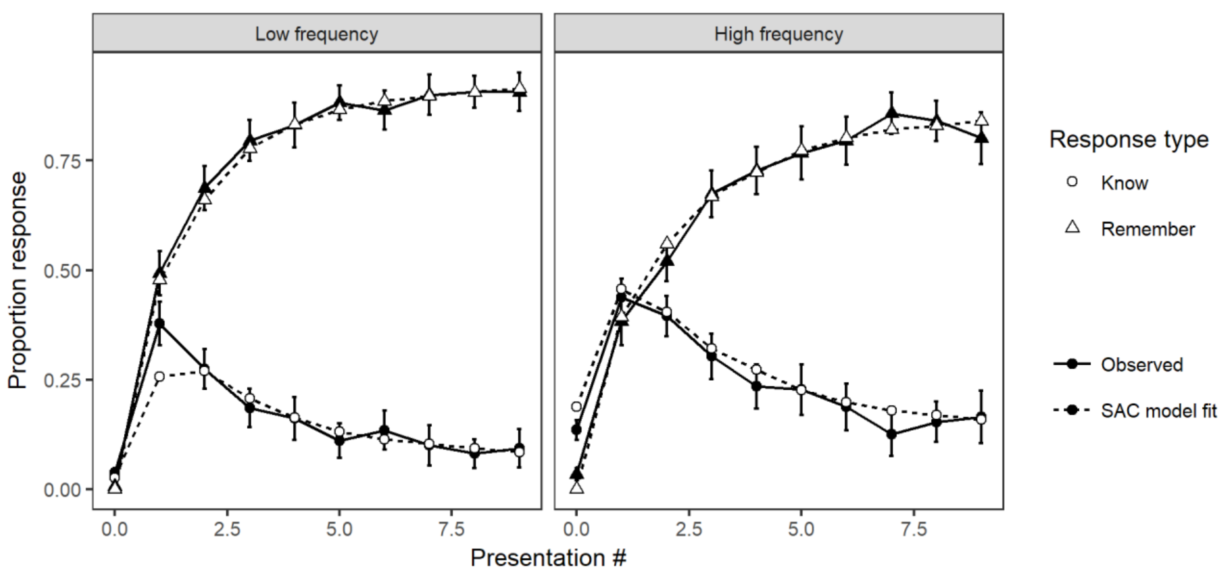

Figure A1. *Proportion of Remember and Know responses in single item recognition in Reder et al (2000) for low and high frequency words as a function of presentation number of each word.*