Frequency effects in recognition and recall

Vencislav Popov (vencislav.popov@gmail.com)[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

Stimulus frequency, which is often evaluated using normative word frequency, is among the variables that have the most diverse and puzzling effects on memory. Word frequency can either facilitate or impair memory performance depending on the study and testing conditions. Understanding why and under what conditions frequency has positive or negative effects on performance is crucial for understanding basic properties about the human memory system. As a result, the study of word frequency has led to the development of multiple memory models. This chapter summarizes the current knowledge concerning word frequency effects on item recognition, associative recognition, free recall, cued recall, serial recall and source memory. We also discuss how word frequency interacts with manipulations concerning presentation rate, list-composition, age of the participants, memory load, midazolam injections, response deadlines and remember-know judgements. This review of frequency effects in memory identified four major classes of empirical findings, which can be further subdivided into a total of 21 key phenomena that any theory should account for. Based on these phenomena, we identify three high-level principles that characterize the diverse effects of frequency on memory – the probe dependency principle, the dual process principle, and the resource demands principle.

Keywords: *word frequency, mirror frequency paradox, list-composition paradox, episodic memory, working memory, recognition, recall*

Contents

Introduction

Of the many variables with which we can study human memory, normative word/stimulus frequency has proven to have some of the most diverse and puzzling effects. Word frequency can either facilitate or impair memory depending on the study and testing conditions. The most common example of this is the so called *word frequency paradox* (for an early review, see Gregg, 1976) – the fact that performance is better for high frequency (HF) words than low frequency words (LF) in recall tasks, while the opposite is true in recognition memory tasks. Hundreds of studies have contributed to uncovering word frequency effects across a variety of memory tasks, and researchers have identified many other crucial variables with which word frequency interacts. These diverse frequency effects have become a major challenge for models of human memory because most models can account either for positive or for negative word frequency effects, but not both. It is a tall order for models to account for all frequency effects, but this task has been made even more difficult by the lack of systematic reviews on the topic. In this chapter, we summarize the current knowledge concerning word frequency effects on item recognition, associative recognition, free recall, cued recall, serial recall and source memory. We also discuss how word frequency interacts with variables such as presentation rate, list-composition, age of the participants, memory load and remember-know judgements. Following recent efforts to identify benchmark findings concerning working memory (Oberauer et al., 2018) and contiguity effects (Healey, Long, & Kahana, 2018), we organize this chapter around major word frequency phenomena, which we then qualify with sub-phenomena.

I. Basic properties of the LF recognition advantage

Finding I.1: LF words have more hits and fewer false alarms than HF words

In a typical item recognition memory experiment, participants study a list of words and during test they have to discriminate words that were presented on the list ("old") from words that have not been presented on the list ("new"). Positive responses to old words are termed "Hits", and positive responses to new words are termed "False alarms". Word frequency affects item recognition memory negatively – LF words are better recognized compared to HF words. As shown on Figure 1, relative to HF words, LF words lead to both more hits and fewer false alarms, a pattern known as the *word frequency mirror effect* (Glanzer & Adams, 1985). This effect is one of the most robust findings in recognition memory, having been replicated by many independent investigators (e.g. Clark, 1992; Glanzer & Adams, 1985; Hintzman, 1994; Hockley, 1994, p. 194; MacLeod & Kampe, 1996; Malmberg & Murnane, 2002; Mandler, Goodman, & Wilkes-Gibbs, 1982; Reder et al., 2000; Schulman, 1967). The mirror effect also occurs with pictures, such that pictures rated as weakly familiar are recognized better than those rated as strongly familiar (Karlsen & Snodgrass, 2004). The mirror effect is not restricted to naturally occurring differences in frequency, but can also be demonstrated by experimentally manipulating the prior exposure frequency of novel stimuli, e.g., pseudowords (Reder, Angstadt, Cary, Erickson, & Ayers, 2002) or Chinese characters (Nelson & Shiffrin, 2013).

Finding I.2: LF words have more Remember and less Know judgements than HF words

Dual-process theories have posited that recognition judgements can occur based on two different processes termed *recollection* and *familiarity* (Joordens & Hockley, 2000; Reder et al., 2000; Reder, Paynter, Diana, Ngiam, & Dickison, 2007; Yonelinas, 2002). Recollection is thought to occur when participants recall specific contextual details associated with the study episode. In contrast, familiarity-based responses occur in the absence of such recollection when the probe seems so familiar that participants infer they must have seen it recently. Whether recognition is due to recollection or familiarity is typically assessed with the Remember/Know procedure. In the Remember/Know paradigm, when participants think a probe is old, they must provide an additional judgement – they must respond whether they explicitly *remember* seeing the probe on the list, or whether they just *know* it must have been on the list, because it seems very familiar to them (Tulving, 1985). When it comes to word frequency effects, Reder et al. (2000) found that old LF words receive more *remember* responses relative to old HF words (subsequently replicated by Hirshman, Fisher, Henthorn, Arndt, & Passannante, 2002; and Joordens & Hockley, 2000). In contrast, HF words show more *know* responses compared to LF words, both for hits and for false alarms (see Figure 2 and Figure 3). These results suggest that it is easier to retrieve the contextual details associated with the study episodes for LF probes.

Finding I.3: Recognition performance is a U-shaped function of word frequency

While most studies discussed above treat word frequency as a dichotomous variable (i.e., high vs low), word frequency is usually estimated on a continuous scale of "number of occurrences per million words". When we consider the entire range of the frequency scale, recognition hits show a U-shaped curve, such that very rare words (less than 1 occurrence per 4 million) are recognized worse than LF words (see Figure 1; Mandler et al., 1982; Rao & Proctor, 1984; Wixted, 1992; Zechmeister, Curt, & Sebastian, 1978) and in some studies even worse than HF words (Mandler et al., 1982). Rare words do not display a mirror effect – their recognition disadvantage is found in either lower hit rates (Mandler et al., 1982) or in higher false alarm rates (Wixted, 1992), although it is not clear what causes this inconsistency. Similar U-shaped results have been obtained in training studies that differentially familiarized previously unknown stimuli, specifically pseudowords (Reder et al., 2002) or three-digit numbers (Maddox & Estes, 1997). This U-shaped pattern appears contrary to results by Lohnas and Kahana (2013) who found that recognition is a monotonic decreasing function of word frequency. However, Lohnas and Kahana (2013)'s lowest bin of stimuli had an average frequency of 21 counts per million, which is significantly higher than the range typically considered for LF words. Thus, the recognition disadvantage for very rare words appears to be robust.

II. Factors that reduce or eliminate the LF recognition advantage

Research has revealed a number of manipulations that either reduce, eliminate or even reverse the LF advantage in recognition memory. This almost always occurs by reducing the hit rate advantage for LF words, without affecting their false alarm rate advantage (Hirshman & Arndt, 1997; Joordens & Hockley, 2000), suggesting that different mechanisms generate each component of the mirror frequency effect in recognition. These findings are summarized below.

Finding II.1: Faster presentation rates reduce the LF recognition advantage

Multiple word frequency recognition experiments have manipulated the presentation duration of individual words during study. A consistent finding from these experiments is that as the study time is reduced, the difference in hit rate between LF and HF gets smaller (Criss & McClelland, 2006; Hirshman et al., 2002; Malmberg & Nelson, 2003). The LF-HF differences disappear with study times as fast as 250 ms. per word, and even reverse when the study duration is as fast as 150 ms. per word (see Figure 4). This is not due to floor effects – when we consider the false alarm rates in these experiments, recognition performance remains well above chance even with the fastest presentation times (d'_{250ms} = 0.88 in Malmberg & Nelson, 2003 and d'_{150ms} = 0.58 in Criss & McClelland, 2006). In general, reducing study time tends to hurt performance for LF words more than it does for HF words (also see, Criss & Shiffrin, 2004), and this pattern also holds for other tasks such as source memory (Popov, So, & Reder, 2019). This decrease in hits, as a function of a decrease in study time for LF words, is manifest entirely in fewer *remember* responses (Figure 3) – the proportion of *know* responses to LF words does not change as a function of study time (Hirshman et al., 2002).

Finding II.2: Dividing attention at study reduces the LF recognition advantage

Another manipulation during study that can eliminate the hit rate advantage for LF words is dividing the participants' attention with a secondary task. In several experiments, Diana and Reder (2006) asked participants to study a list of words either in a single task condition or in a dual task condition. In the dual task condition, while studying the word list, participants had to also perform a serial addition task. They heard a different digit every four seconds, and immediately upon hearing the digit they had to mentally add it to the previous digit and report the result. Dividing attention not only reduced overall recognition performance, but it also eliminated the difference in hit rate between studied LF and HF words. In contrast, the false alarm advantage for novel LF words was not affected. As with the study duration manipulation, this reduction in the LF advantage is not due to floor effects ($d'_{HF} = 0.25$, $d'_{LF} = 0.48$ in the dual task condition). While dividing attention *at study* reduces the LF recognition advantage, dividing attention at test does not diminish the word frequency effect (Hintzman, Caulton, & Curran, 1994).

Finding II.3: Midazolam removes the LF hit rate advantage

The hit rate advantage for LF words can be eliminated if prior to study participants are injected with an amnesia inducing drug called midazolam (Hirshman et al., 2002). Midazolam is a benzodiazepine that creates temporary anterograde amnesia by preventing the storage of new associations in LTM (Ghoneim, 2004; Reder et al., 2006), but it does not impair pre-existing memory traces (Ghoneim, 2004) or their strengthening (Hirshman, Passannante, & Arndt, 2001). Hirshman and colleagues (2002) found that injecting participants with midazolam prior to study eliminated the LF hit rate advantage, but it had no effect on the false alarm difference between HF and LF words. As shown on Figure 5, in the midazolam condition, participants were more likely to respond "old" to HF words than to LF words, regardless of whether they were old or new, thus producing both more hits and more false alarms. This was due to the near elimination of *remember* responses, and as we reviewed earlier, HF words receive more *know* responses regardless if they are novel or studied. These effects have been replicated with another benzodiazepine, triazolam (Mintzer, 2003).

Finding II.4: LF recognition advantage is reduced in older adults

Balota and colleagues (Balota, Burgess, Cortese, & Adams, 2002) have shown that the age of participants also affects the mirror frequency effect in recognition (also see, Bowles & Poon, 1982). Specifically, as people get older, they find it more difficult to correctly recognize studied LF words; in contrast the hit rate for HF words does not change with age and false alarm rates increase slightly but for both frequency classes (see Figure 6). Thus, similar to the other manipulations reviewed above, ageing reduces the LF hit rate advantage, but has no effect on the *difference in* false alarm rates between HF and LF words.

Finding II.5: LF recognition advantage is reduced when fast responding is required

Balota et al. (2002) argued that as participants get older, they tend to rely more on baseline familiarity rather than on recollection. In a second experiment, they tested this idea by testing only young people in a response deadline paradigm in which participants are given a limited time to respond. Based on results by Jacoby (1999), they argued that a faster deadline (500 ms vs 1000 ms) would induce even young people to rely more on familiarity, and that the fast vs slow deadline results would mimic those in the old vs young comparison, which is what they found. These results are consistent with earlier data by Hintzman and colleagues (1994), who used a response-signal paradigm in which participants are given a signal to respond to the probe after variable amounts of time. The authors found that the hit rate LF advantage emerges only when sufficient time is given to respond (Hintzman et al., 1994). In contrast, the false alarm LF advantage is present even with the shortest response signal lags (Hintzman et al., 1994; also see, Joordens & Hockley, 2000).

Finding II.6: LF recognition advantage decreases as the proportion of LF words on a list increase

In nearly all memory tasks, regardless of whether frequency has a positive or a negative effect, *the magnitude* of the frequency effect often depends on the exact proportion of HF and LF words on each list. As we will review later, this *list composition effect* is well established for free recall, but some studies have found a list composition in recognition memory as well. Malmberg and Murnane (2002) presented participants with *mixed* lists composed of either 25% LF words or 75% LF words. The authors found that as the proportion of LF words on the list increases, the difference in recognition performance between LF and HF words becomes smaller. This was because it was more difficult to discriminate novel and studied LF words on lists with majority LF words; the discrimination of HF words was not affected by list composition. Dorfman & Glanzer (1988) also found that the HF-LF difference is reduced with lists consisting of a high proportion of LF words, and Clark & Burchett (1994) found that the HF-LF difference is slightly reduced in pure lists that contain only one frequency type per list (Clark & Burchett, 1994). In contrast to Malmberg and Murnane (2002), however, the latter two studies found that the list-composition effect was due to differences in recognition performance for HF words between lists, not LF words. It is unclear what is the cause of this discrepancy. Thus, while the overall list-composition effect is reliable, it is not clear whether the reduction of the HF-LF difference can be found in HF or LF words alone.

Finding II.7: HF recognition advantage in associative recognition

Thus far, we have only reviewed experiments in which participants study individual words followed by a discrimination task of those words from unstudied words; however, word frequency effects have also been investigated using associative recognition tasks. In those tasks, participants study a list of pairs (or even triplets) of words and at test, words are presented either in the same pair as during study, or are recombined to form new pairs. While all words are "old", participants must discriminate *intact* from *recombined* word pairs. In contrast to item recognition, the effect of word frequency is either eliminated (Hockley, 1994) or even reversed, such 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). This occurs even when item recognition for the same stimuli is lower for the HF words (Clark, 1992). The effect is also observed with experimental manipulations of frequency – for example, pre-exposure to LF words improves subsequent associative learning (Chalmers & Humphreys, 2003). Finally, the benefit for HF word pairs is removed with incidental encoding (Humphreys et al., 2010).

III. The HF recall advantage

We now turn to a discussion of the other side of the "word frequency paradox" – the fact that in recall tasks, HF words tend to be recalled more often than LF words. We will discuss results from free recall, cued recall and serial recall, as well as source memory tasks.

Finding III.1: Basic properties of the HF advantage in free, cued and serial recall

The *recall* advantage of HF words is well documented – it appears in free recall tasks, in which participants are asked to recall the list items in any order (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); in serial recall tasks, in which participants have to recall list items in their studied order (Hulme et al., 1997; Hulme, Stuart, Brown, & Morin, 2003; Miller & Roodenrys, 2012; Morin, Poirier, Fortin, & Hulme, 2006); in cued recall tasks, in which participants study a pair of words and at test are cued with one word from each pair and have to recall the other (Criss, Aue, & Smith, 2011; Madan, Glaholt, & Caplan, 2010; Reder, Liu, Keinath, & Popov, 2016); and in source memory tasks, in which participants have to recall some contextual detail associated with the studied word (DeWitt, Knight, Hicks, & Ball, 2012; Popov et al., 2019; Poppenk & Norman, 2012). In pure lists, the word frequency effect in free recall is parametric – recall probability increases as a monotonic function of word frequency (Deese, 1960; although mixed lists might show a U-shaped pattern, Lohnas & Kahana, 2013). The HF advantage is seen not only in terms of overall recall, but also in a faster learning rate when studying to a criterion (Sumby, 1963). Sumby (1963) asked participants to repeatedly study the same lists of HF or LF words until they were able to recall all words from each list. The number of study-test cycles (trials) that were required to learn a list to criterion varied parametrically with word frequency (Figure 7). The HF advantage is also seen for meaningless stimuli – serial recall is better for nonword syllables that occur more frequently in polysyllabic words (Nimmo & Roodenrys, 2002).

Finding III.2: The size of the HF recall advantage depends on the list-composition

As we noted earlier, one of the major variables that modifies the effect of word frequency is list composition. *Pure study lists* are defined as lists that contain only HF or only LF items;

mixed study lists are defined as lists that contain both LF and HF items. The HF advantage described above is present only in pure lists, and is reduced, absent or even reversed in lists of mixed frequency. Due to this reversal, this phenomenon has been termed *the mixed list paradox* (Gillund & Shiffrin, 1984; MacLeod & Kampe, 1996; Ozubko & Joordens, 2007; Popov & Reder, 2020; Popov et al., 2019; Ward et al., 2003; Watkins et al., 2000). Results show that the presence of LF items on the list hurts memory for HF items, and that the presence of HF items helps memory for LF items (i.e., performance for HF pure > HF mixed ~ LF mixed > LF pure). The mixed list paradox has been replicated multiple times and receives strong support from a meta-analysis of free recall studies (Ozubko & Joordens, 2007). While the mixed list paradox has been documented most reliably in free and serial recall, it is also seen in source memory performance – recalling the location at which a word was presented during study is easier for HF compared to LF words in pure lists, but not in mixed lists (Popov et al., 2019). Most prior source memory studies have only used mixed-lists and show a benefit for LF cues (Glanzer, Hilford, & Kim, 2004; Osth, Fox, McKague, Heathcote, & Dennis, 2018)

Furthermore, the list-composition effect is not limited only to the categorical pure vs mixed distinction – the word frequency effect in recall depends on the proportion of HF/LF words on the list. 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). In source memory, the ability to recall the spatial position of LF cues improves gradually as the proportion of HF words on the list increases from 0% to 40% to 60% (Popov et al., 2019). For comparison, both the free recall and the source memory list-composition results are illustrated on Figure 8.

Finding III.3: The size of the HF recall advantage depends on the order of HF and LF items and interacts with serial position

Serial recall studies, in which participants are typically given short lists and they have to recall items in the studied order, have shown that in mixed lists the magnitude of the HF advantage depends on the order of HF and LF in the list. As shown on Figure 9, serial recall is better when the first half of each mixed list contains HF words and the second half contains LF words. The performance on both types of mixed lists is intermediate between pure HF lists and pure LF lists (e.g., HH > HL > LH > LL, where the two letters reflect word frequency within the first and the second half of the lists; Miller & Roodenrys, 2012; Watkins, 1977). Furthermore, as can be seen from the left panel of Figure 8, the serial recall advantage for HF words increases with serial position, an effect that was first shown by Hulme and colleagues (Hulme et al., 1997) and has since been replicated multiple times (Hulme et al., 2003; Miller & Roodenrys, 2012).

Finding III.4: Speech rate, and the frequency and recency of rehearsals can account for part of the HF recall advantage

Ward and colleagues (Tan & Ward, 2000; Ward et al., 2003) showed that one contributor to the HF recall advantage is the fact that words in HF lists are rehearsed more often. Following a procedure introduced originally by Rundus (1971), they asked participants to rehearse items out-loud during the study session and recorded both the number of times each word was rehearsed and the proximity of the last repetition of a word to the end of the list. Results showed that words in HF lists were rehearsed both more often than words in LF lists, and also that the last rehearsal of HF words was much closer to the end of the list, making them more *functionally recent.* Nevertheless, these two factors only account for part of the HF recall advantage – HF words were still better recalled even after removing the effects of frequency and recency of rehearsals. Similarly, Hulme and colleagues demonstrated that even though speech rate is faster for HF words in serial recall, the effects of frequency remain even after accounting for the differences in speech rate between HF and LF words (Hulme et al., 1997).

Finding III.5: Greater temporal clustering of HF words in free recall

Another aspect of the HF recall advantage is that more or stronger associations are formed between successive words in HF lists than in LF lists. The strength of inter-item associations is usually reflected in contiguity effects in free recall (Healey et al., 2018; Kahana, 1996). Contiguity effects occur when after recalling one item, participants are more likely to recall the item/s that were originally presented immediately before or after the just recalled item in the study sequence. Such contiguity effects are stronger for words from HF lists (Figure 10) and these are not an artifact of the overall higher recall rates for HF words (Ward et al., 2003).

Finding III.6: Faster presentation rates increase the HF recall advantage

Just as with item recognition, the size of the frequency effect in free recall, cued recall and source memory tasks depends on the presentation rate during study, although this interaction is not as broadly studied and not as often replicated in recall tasks as it has been for recognition tasks. Data for free recall comes from Gregg and colleagues (Gregg et al., 1980) who used a continual distraction paradigm. In their task, each word was presented for the same duration, but the inter-stimulus-interval (ISI) between words was either 0 s., or it was 10 s. during which time participants had to count down from a three-digit number by threes until the next word was presented. The authors found that slowing down the presentation rate in this way decreased the HF recall advantage. The interpretation of these data is complicated by the fact that presentation rate is confounded with both attentional demands (the ISI is either filled with an attentionally demanding task or not), and with study-test lag (on average the lag between study and test for each word is larger in the 10s ISI condition). To our knowledge, no other study of free recall has manipulated both presentation rate and word frequency.

Similarly, there are no published data on the interaction between word frequency and presentation rate in cued recall tasks. However, unpublished data from our lab show that the HF cued recall advantage grows larger when speeding up the presentation rate and that this relationship is linear, at least within the 1.5-4.5 words per second range we tested (see Figure 11). This effect can also be observed in source memory tasks with much faster presentation rates (Popov et al., 2019). In two experiments, we presented participants with lists of 5 HF or LF words, where each word was shown on a different location on an invisible circle. Lists were presented at a rate of 0.5, 0.75 or 1 s. per word. Immediately after the last word was displayed in a list, one of the words was shown in the middle of the screen as a probe. Participants had to report that word's studied location by clicking on a circle. The degree of location recall error was lower for HF words compared to LF words, and that difference decreased with slower presentation times, disappearing completely when words were presented at a rate of 1 s. (Figure 12).

Finding III.7: The HF working memory advantage increases at higher memory loads

In addition to serial recall, effects of stimulus frequency appear in other working memory tasks, such as the N-back task, and with experimentally manipulated frequency, rather than with naturally occurring differences in frequency (Reder et al., 2016). Normative word frequency is confounded with multiple lexical factors (Cox et al, 2019; Maddox & Estes, 1997; Reder et al., 2002), and in order to demonstrate that stimulus frequency per se affects working memory, Reder et al (2016) familiarized participants with novel Chinese characters at different frequencies of exposure in a visual search task over multiple weeks. Participants subsequently performed an N-back task in which characters were presented one at a time, and participants had to respond whether the current character is the same as the one presented N trials ago (1, 2 or 3). As shown on Figure 13, characters trained at high frequencies led to better N-back performance than characters trained at lower frequencies. Furthermore, this frequency effect increased as the demands of the working memory task increased. Both of these effects were replicated in a subsequent study (Popov & Reder, 2020) that compared N-back performance for familiar animals, and for novel animals called Fribbles, that were trained at either low or high frequency (Figure 13, right).

Finding III.8: The effect of the cue frequency is smaller than that of the target frequency in cued recall

Finally, we will consider how word frequency affects cued recall. In cued recall tasks typically two words are studied together but, at test, one of the words is given as a cue to recall the other. This allows us to evaluate whether the frequency of the cue and the target have differential effects on performance. Two studies have explored this issue in depth by orthogonally manipulating the frequency of the two words in studied pairs (Criss et al., 2011; Madan et al., 2010). Both studies converge on similar results – while HF targets were recalled much better than LF targets, the frequency of the cue either had no effect (Madan et al., 2010) or had a positive frequency effect that was nevertheless much smaller in magnitude than that of the target (Figure 14). Popov and Reder (2020) have argued that with real words, frequency affects the cue less than the target, because a higher contextual fan of associations for HF words masks the frequency benefit. In support of this, a large effect of cue frequency was obtained in the Chinese character training study mentioned in the preceding section. In addition to the N-back task described above, participants also performed a cued-recall task in which they associated two different Chinese characters with an English word. Both of the characters were previously trained at either HF or LF. During test, both characters were presented, and participants had to recall the English word. Recall was higher for HF cues, and this benefit persisted on a delayed test a month later (Reder et al., 2016).

IV. Sequential frequency effects during study

Thus far we have reviewed how frequency of a particular item affects memory for itself, in both recall and recognition tasks. A final key piece of the frequency puzzle concerns how memory for one item is affected by the frequency of items that were studied in close proximity to the tested item.

Finding IV.1: Better memory for items that are presented concurrently with a HF word

Recognition memory depends not only on the familiarity of the tested item, but also on the frequency of words that were studied simultaneously with it (Diana & Reder, 2006; Malmberg & Nelson, 2003; Reder et al., 2007). Malmberg and Nelson (2003) asked participants to study pairs of words that consisted of either two HF words, two LF words or one HF and one LF word. Even though participants studied pairs, during testing they only saw individual words, and they only had to respond whether each word was old or new. Consistent with results reviewed earlier, Malmberg and Nelson found a typical LF recognition advantage, when they considered the frequency of the target word. However, regardless of their own frequency, words were recognized better when they were studied alongside a HF word, rather than a LF word. This effect was removed when the study time was increased from 1.2 to 4 seconds per pair (Figure 15). These results were later replicated and extended by Diana & Reder (2006) both with word pairs and with picture-word pairs. Rather than fixing the time for study, Diana & Reder (2006) presented pictures with HF or LF words superimposed on them, and participants had to read out-loud the superimposed word. Each trial ended after the naming response. In a subsequent picture recognition test, participants had more hits for pictures that had HF words superimposed on them. Furthermore, the effect was due to a higher proportion of remember rather than know responses to pictures that had an HF word superimposed.

Finding IV.2: Better memory for items following a HF item during study

In addition to simultaneously studied words, the frequency of items presented previously in the study sequence also affects memory for a studied item. Consider Figure 16, which represents a sequence of study items. In a reanalysis of nine archival datasets, Popov and Reder (2020) recently discovered that memory for a specific item, X_k, where k represents the item's position on a study list, is affected by the frequency of the items that preceded it during study, i.e., X_{k-1}, X_{k-2}, etc. Specifically, memory performance was better across a variety of test types (item and associative recognition, cued and free recall) when the preceding X_{k-1} item was a HF word rather than a LF word (Cox, Hemmer, Aue, & Criss, 2018; Diana & Reder, 2006; Healey & Kahana, 2016; Ward et al., 2003), was a pseudoword trained at high rather than low frequencies (Reder et al., 2002), or was a word that was repeated multiple times during the preceding study sequence (Aue, Criss, & Novak, 2017; Buchler, Light, & Reder, 2008). This effect was a continuous function of the preceding word's frequency, and it occurred in both hit/recall rates and reactions times (see Figure 17 for Popov & Reder's, 2020, reanalysis of Cox et al.'s data across four different memory tests). Finally, the size of the sequential frequency effects depended on the frequency of the current item – preceding LF items have a bigger impairment on memory for the current item when the current item is LF as well (an example from the reanalysis of Ward et al., 2003, is shown on Figure 18). These effects, and the additional findings reviewed in the sections below, are summarized in Table 1.

Finding IV.3: Sequential study effects are cumulative and depend on study lag

Memory for the current item is affected not only by the frequency of the immediately preceding item, but also by the frequency of items further back in the study sequence. Popov and Reder's (2020) reanalysis revealed that these sequential effects accumulate, such that memory for an item is worse, if more of the preceding study items were LF rather than HF (an example from the reanalysis of Diana and Reder (2006) is shown on Figure 19, left). Further analyses revealed, that as the study lag between the tested item and the preceding item (i.e., the number of items that separate them in the study sequence) increases, the effect of the preceding item's frequency gets smaller and disappears at lags of 3 or 4 items (an example from the reanalyses of Diana & Reder, 2006 is shown on Figure 19, right).

V. High-level characteristics of frequency effects

What conclusions can we draw from the complex empirical landscape formed by these phenomena? A few high-level patterns emerge when we consider the commonalities and differences among the cases in which we observe positive or negative frequency effects.

- ***PROBE DEPENDENCY PRINCIPLE: a word's frequency affects memory performance differently depending on whether the item itself is used to probe memory***

If we consider the full patterns of results presented above, one thing stands out – whether an item's frequency would have a positive or a negative effect on retrieving its episodic memory trace depends on what cues are used to access the trace. An episodic memory trace is a binding of items and the context in which they were experienced. In free and serial recall, where memory is cued only with the experiential context (i.e., "recall all items experienced during the immediately preceding list only") and/or some positional cues, but not with the items themselves, frequency almost always has a strong positive effect (Finding III.1)[1]. At the other extreme, in item recognition, where memory is cued with the item itself, frequency has a negative effect on memory performance (Findings I.1-I.3). Thus, the *probe dependency principle* states that memory traces containing HF items are easier to access, unless the item itself is used to cue memory.

Further support for this probe dependency principle comes from the cued-recall results we reviewed above, which demonstrated that, just like in free recall, HF targets are easier to recall in cued recall tasks. At the same time, HF cues either provide no benefit over LF cues, or the difference between HF and LF cues is much smaller than the difference between HF and LF targets (Finding III.8). Similarly, in source memory, unless the encoding demands are high, LF cues serve better to retrieve contextual source information than HF cues (Findings III.2 and III.6). Finally, from the sequential effects we reviewed in Section IV, we can see that regardless of how memory is probed, if performance is conditioned based on the frequency of other items that were studied in close proximity, frequency effects are positive, even in item recognition (Findings IV.1 and IV.2). In summary, we suggest that the *word frequency paradox* should be conceptualized not in terms of the type of memory task (recall vs recognition) but rather in terms of whether the item whose frequency we are interested in is used to probe memory.

- ***DUAL PROCESS PRINCIPLE: Two different processes generate the hit rate and false alarm portions of the mirror effect***

Dual process theories of recognition memory state that recognition decisions can be made based on two qualitatively different processes – familiarity and recollection (Balota et al., 2002;

[1] As we noted, this effect disappears in mixed-lists. Even though some studies find that in mixed-lists the effect is reverse, the majority of mixed-list studies find no difference between HF and LF items in mixed-lists, and a meta-analysis of all mixed-list studies supports this conclusion (Ozubjo & Joorgens, 2007).

Joordens & Hockley, 2000; Reder et al., 2000; Yonelinas, 2002). The data we reviewed in Sections I and II are consistent with the need to posit two different processes for generating the hit rate and the false alarm rate parts of the mirror frequency effect. There are two reasons for this: First, the hit rate advantage for LF words is found primarily in recollection-based Remember responses, while the higher false alarm rate for HF words is due to more familiarity-based Know responses[2]. Second, virtually any manipulation that affects the size of the mirror frequency effect in recognition reduces/eliminates the *hit rate* advantage for LF words, but does not affect the greater false alarm rate for HF words. We reviewed many of these effects earlier but one of them is worth revisiting – the effect of midazolam and other benzodiazepines on the Mirror Frequency Effect. It has been demonstrated that these drugs block the ability to form new associations between items and context, thereby preventing the formation of novel memory traces (Reder et al., 2013). Midazolam was shown to eliminate the LF hit rate advantage, while leaving unaffected the higher FA rate for HF words (Finding II.3), thereby providing strong converging evidence for the dual process principle. Midazolam prevents the storage of new memory traces, which suggests that the higher false alarm rate for HF words can only be due to higher familiarity of the existing concept traces. Finally, the dual process principle is also supported by the different time courses of the hit rate and false alarm rate parts of the mirror frequency effect (Finding II.5). While the interpretation of some of these findings has been challenged (e.g., see Dunn, 2004, for criticisms of the Remember-Know procedure), it is difficult to reconcile the combined data with single-process models.

[2] Some theorists have argued against the view that remember and know judgements are based on different processes (Donaldson, 1996; Wixted & Stretch, 2004). However, Wixted has recently changed his mind and concurs that a single-process theory is not sufficient to account for the full pattern of remember/know responses in the literature (Wixted & Mickes, 2010).

- ***RESOURCE DEMANDS PRINCIPLE: the negative effects of frequency disappear or reverse under higher encoding demands, while positive effects of frequency get stronger***

Many of the variables with which word frequency interacts could be interpreted as affecting the difficulty of the memory formation process (Popov & Reder, 2020). Based on the data reviewed in Sections I-III, we can state the *resource demands principle*: the negative effects of frequency disappear or reverse under higher storage/resource demands, while the positive effects of frequency get larger. Here we will consider how each of the variables reviewed above contributes to this principle:

- Presentation rate – as the presentation rate speeds up, memory formation becomes more difficult and resource demanding. As a result, speeding up presentation rate reduces and reverses the LF recognition advantage (Finding II.1) and the source memory advantage (Finding III.6), and it also has the complementary effect of increasing the HF advantage in free and cued recall (Finding III.6).
- Dividing attention during encoding reduces the amount of available resources and as a result it eliminates the LF hit rate advantage (Finding II.2)
- Aging, which is also associated with a decline in cognitive resources, also reduces/eliminates the LF hit rate advantage (Finding II.4). Also see (Buchler & Reder, 2007; Buchler, Faunce, Light, Gottfredson, & Reder, 2011)
- Having to store and associate pairs of words, rather than a single word, has greater resource demands, because participants have to bind two words to a context, rather than a single word. This leads to an HF advantage in associative recognition (Finding II.7), despite the probe dependency principle. This is also why in cued recall the

frequency of the cue has a diminished positive but not a negative effect (Finding III.8), despite the probe dependency principle.

- Similarly, very rare words should be especially difficult to process and show worse recognition memory than LF words (Finding I.3).
- Increasing working memory demands directly by increasing the level of N-back tasks, increases the HF advantage in N-back memory tasks (Finding III.7).
- In serial-recall tasks, more items have been stored in memory with each subsequent serial position, and the HF recall advantage increases with serial position (Finding III.3)
- Finally, list-composition and sequential frequency effects could also be considered an example of this principle – adding items that are more difficult to process, such as LF words, to a study list, decreases memory for other items on the list (Findings II.6 and III.2), *specifically*, it hurts memory for the items that follow them on the study list (Findings III.3, IV.1, IV.2, and IV.3)

The resource demands principle suggests that LF and HF items differ in the amount of processing resources they require for storage – as the demands of the task increase, either due to manipulations of the stimuli (e.g., number of items, frequency of other items), the procedure (e.g., dividing attention, speeding up the presentation rate, increasing working memory load, increasing serial position), or due to individual differences (e.g. aging), the mnemonic benefit moves towards the direction of HF items.

VI. Theoretical accounts of frequency effects

Several different theories and models have been proposed to account for various subsets of the findings discussed above. We will briefly[3] discuss the major theoretical contenders and note to what degree each is consistent with the broad range of empirical data, as well as the three high-level principles outlined in Section V.

○ ***The Source of Activation Confusion (SAC) model***

The SAC model was among the early computational accounts of the Mirror Frequency Effect in item recognition (Reder et al., 2000) and in its revised version (Buchler et al, 2011; Diana & Reder, 2006; Reder et al., 2007; Popov & Reder, 2020) is currently the only model that can account for all of the findings reviewed above. It is also the model that predicted *a priori* the sequential frequency effects discussed in Section IV. The revised model implements a resource-depletion-and-recovery theory of memory based on the following assumptions:

- memory formation depletes a limited working memory resource that recovers gradually over time
- the amount of resources required to process an item is an inverse function of its current strength
- memory traces are less likely to be formed or are weaker when resources are insufficient
- HF words have stronger representations in memory compared to LF words
- HF words have more associations in memory compared to LF words, and thus are less effective in retrieving any specific one of them when used to probe memory

[3] For full description of each model, we refer the reader to the corresponding references in each section.

- Memory judgements can be based either on retrieving an episodic memory trace, resulting in recollection decisions, or on the strength of conceptual nodes, resulting in familiarity decisions

Computational modeling has revealed that these assumptions are sufficient to account for the entire pattern of frequency results presented in this chapter (for details on the model and simulations, see Popov & Reder, 2020). The first four assumptions listed above naturally account for the patterns associated with the resource demand principle, because all of the listed manipulations either increase the resource demand of the storage operations, decrease the overall amount of available resources, or prevent sufficient resource recovery.

Two factors in SAC give rise to the probe dependency principle. On the one hand, as a result of the assumptions outlined above, it is easier to associate HF words to others and to an experiential context, which leads to stronger episodic memory traces for HF items. On the other hand, the ease of access to a memory trace also depends on the number of other contextual associations to the probe. HF words, by definition, have been experienced more often than LF words which also means that they are associated with a greater number of contextual experiences. Research on fan effects in memory has demonstrated that cues with a greater fan/number of associates are less effective in retrieving any specific association than cues with a smaller number of associations (Anderson & Reder, 1999; Schneider & Anderson, 2012). Thus, because of the fact that HF words are associated with a greater number of memory traces they are less effective than LF words in retrieving any one of them when used to probe memory (Reder et al., 2000). Thus, even though HF items facilitate memory formation and lead to stronger memory traces, access to these traces is more difficult if the HF item is used to probe memory. This trade-off between an HF encoding advantage and LF recognition advantage results in the probe dependency principle outlined above (Reder et al., 2007). This contextual competition mechanism, which is also known under the term *contextual noise* (Dennis & Humphreys, 2001; Osth & Dennis, 2015), explains why old LF probes have more hits, and why the effect of a cue's frequency in cued recall is smaller than that of a target. It also explains why the LF hit rate advantage is due to more recollection-based remember responses in a *remember-know* paradigm (Finding I.2). It also partially explains why as people get older the LF hit rate advantage decreases – both HF and LF words increase in associated contextual experiences as people grow older, which reduces the contextual fan difference between them (Buchler & Reder, 2007).

Finally, the dual process principle is present in SAC because its representations include two types of nodes – semantic nodes that reflect knowledge about concepts, and episodic nodes that bind together semantic nodes and an experiential context to reflect an experienced episode. Memory decisions can be made based on either the strength of the semantic nodes, which causes a *familiarity* judgement or the episodic nodes, which causes a *recollection* judgement. Since HF are experienced more often they have stronger semantic nodes leading to more false alarms for new words, and more *know* judgements for old words. SAC posits that all of the manipulations discussed in Section II reduce the probability of forming episodic nodes, thus removing the LF recognition advantage without affecting the false alarm difference between LF and HF words.

A number of other mechanisms have also been proposed to explain some of these effects, and while none of them can currently account for the entire pattern of results presented so far, it is worth reviewing them as well. It should be noted that none of these mechanisms can currently account for the sequential frequency effects reviewed in section IV, which were only recently discovered (Popov & Reder, 2020). These alternative models are discussed below.

- ***The Attention-Likelihood Theory***

One of the first proposed mechanisms for explaining the mirror frequency effect in recognition memory was the Attention-likelihood theory (ALT; Glanzer & Adams, 1985; Glanzer, Adams, Iverson, & Kim, 1993). ALT posits that studying an item "marks" some of its features, and that the number of marked features increases when more attention is paid to a particular item. LF words are assumed to attract more attention and as a result have more features marked during study. Then, during test, when a probe is presented, a likelihood ratio is calculated based on how many features are marked, and how many features an old and a new item of this type is *expected* to have been marked. The theory assumes that participants know LF words are easier to recognize, which leads them to expect more marked features for them, and as a result of this expectation and the log-likelihood transformation, LF words produce more hits and fewer false alarms.

ALT presents a mathematically simple account of general mirror effects including other variables aside from frequency; nevertheless, many of its predictions concerning frequency have been disconfirmed (Hinztman et al., 1994; Hirshman & Arndt, 1997; Malmberg & Murnane, 2002; Stretch & Wixted, 1998). Notably, due to the way the likelihood ratio is calculated, ALT predicts that the hit rate and the false alarm portion of the mirror frequency effect should co-vary. As we have reviewed here, this is not the case (also see, Hirshman & Arndt, 1997). Further, ALT cannot explain why extremely low frequency words do not follow the same pattern as LF words, and while it does not deal with recall at all, only with recognition tasks, it is difficult to imagine how it might accommodate the HF recall advantage or many of its properties reviewed here. Finally, it cannot account for list-composition effects (Malmberg & Murnane, 2002).

- ***Differential feature overlap and the Retrieving Efficiently from Memory (REM) model***

Like ALT, the REM model (Shiffrin & Steyvers, 1997), depends on calculating an odds ratio, but in REM this is based on the number of features that overlap between the probe and the contents of memory (also see the Subjective likelihood model, SLiM, which provides a very similar account; McClelland & Chappell, 1998). LF words are assumed to have more distinctive features which they share with fewer other items in memory. For old LF words, when these distinctive features match an existing memory trace, they provide large diagnostic information due to their rarity, and produce a larger odds ratio compared to old HF words. When it comes to new words, they do not match fully any existing traces, but they partially activate other traces due to feature overlap. Since new HF words share more features with more traces in memory, they produce a bigger odds ratio and a greater sense of familiarity compared to new LF words, resulting in more false alarms.

The REM model has been successful in capturing a variety of recognition memory results, and it fares better compared to ALT with list-composition results, although it requires a modification for that purpose (for details, see Malmberg & Murnane, 2002). This modification is that LF words are encoded better on lists that contain a greater proportion of HF words. This explanation is in form similar to the one in SAC, although it is not clear within REM why list-composition would affect how well LF words are encoded. The differential resource depletion assumption in SAC provides a mechanistic version of this claim that justifies the difference in encoding strength among lists with different percentages of LF and HF words. Even with the modification introduced by Malmberg & Murnane (2002), REM in its current form would not be able accommodate the sequential study effects reviewed in section IV (as neither would SLiM be able to). Finally, REM and SLiM have mechanisms that allow them to account for the reversal of the frequency effects when the presentation rate speeds up (Criss & McClelland, 2006; Malmberg & Nelson, 2003), but it is not clear whether either can readily account for the effects of midazolam, divided attention or aging.

ALT, REM and SLiM are all examples of models that, prior to recognition, use a rescaling procedure to transform the familiarity distributions of old and new HF and LF items such that the transformed distributions have the correct order for producing a mirror effect. In addition to the issues we described above, the mechanistic plausibility of the various rescaling procedures has been questioned (Hintzman et al., 1994).

- ***The Word-Association-Space (WAS-FE) model***

As we noted previously, normative word frequency is confounded with multiple lexical factors (Cox et al, 2019; Maddox & Estes, 1997; Reder et al., 2002). In fact, one model of the mirror-frequency effect in recognition memory is based entirely on the way LF and HF are semantically structured in the lexicon (WAS-FE; Monaco, Abbott & Kahana, 2007). The ALT, REM and SLiM models discussed above all use random vectors to represent word inputs. However, it is unlikely that actual word representations are random, and prior work has shown that incorporating semantic structure into vector representations of words can account for many recognition memory phenomena (WAS; Steyvers et al., 2004). The original WAS model developed by Steyvers and colleagues (2004) is based on behavioral free association data which contains the probability with which a certain word is produced as an associate of a given cue word (Nelson et al., 2004). WAS uses a dimensionality reduction technique to develop word vectors of 400 dimensions each that optimally capture the semantic structure of the free-association dataset.

Following up on this work, Monaco and colleagues (2007) demonstrated that a popular Hopfield network model of recognition memory can produce a mirror frequency effect simply by feeding it with the WAS inputs for HF and LF words (this updated model is called WAS-FE). The idea is that since LF words have fewer meanings, they form tighter clusters in the semantic space, causing LF probes to generate less internal network energy. In contrast, HF words have many more meanings and are encoded more diffusely in the network, causing HF probes to produce more energy in the network. Network energy is conceived of as the inverse of familiarity, and thus LF words produce more familiarity than HF words in this model. This, by itself, does not produce a mirror effect, because both old and new LF words produce more familiarity than the corresponding class of HF words.

This problem is the same one faced by the ALT, REM and SLiM models. The difference is that the other models use a rescaling procedure to circumvent the problem while the WAS-FE model introduces a stimulus-dependent criterion shift. Specifically, the model uses two separate criteria for making an old decision for LF and HF words, enabling the mirror effect. Although this model shows that some part of the mirror-frequency effect in recognition can be due to semantic factors and not word frequency per se, the fact that it is a single-process model leads to the same issues we described for the ALT, REM and SLiM models. Furthermore, the stimulus-based criterion shift, especially when only applied to the judgement on old, but not new items, is controversial (Wixted & Stretch, 2000). As the authors themselves note, the fact that semantic factors contribute to the WFE, does not mean that other factors do not. Finally, this model would have difficulty accounting by itself for the training studies discussed earlier (Maddox & Estes, 1997; Reder et al., 2002).

- ***Redintegration***

Above we presented the main alternative explanations for the LF recognition advantage. What about the HF recall advantage? One of the most popular explanations states that memory traces degrade over time, either due to decay or interference. During retrieval, these degraded traces have to be reconstructed with the help of representations from semantic memory. Representations for HF items are thought to be more accessible and thus more likely to be redintegrated. While greater accessibility of HF words might be able to explain the main HF recall advantage, it has trouble accounting for many of the other patterns we have discussed. For example, in its base form, redintegration cannot account for list-composition effects. Hulme (2003) proposed a change to the redintegration hypothesis – that HF items are more likely to be semantically related and thus can prime each other and make themselves more available, which increases redintegration. However, Caplan et al. (2015) have shown that list-composition effects hold even when accounting for semantic similarity and other lexical factors. Redintegration and other accessibility-based explanations would struggle to account for the other findings that constitute the resource demands principle.

- ***The item-order hypothesis***

Finally, the idea that LF and HF items consume different amounts of resources, is not limited to the SAC model discussed above. According to the item-order hypothesis (Serra & Nairne, 1993), the HF recall advantage is due to the fact that LF items require more resources for processing, leaving fewer resources for encoding the order of items. Serra and Nairne (1993) argued that in free recall tasks participants store inter-item associations, which contain information about the order in which items appear during study, and then they use that order to guide recall despite the task being free recall. Thus, if there were not enough resources to encode order information, recall performance would be diminished in mixed-lists. The difference between the item-order hypothesis and the SAC model is that SAC does not limit the effect of resource depletion to preserving information about item order. Instead, when resources are insufficient, any memory formation operation suffers. Finally, the item-order hypothesis is not formally implemented in a computation model.

VII. Conclusion

The effect of normative word frequency has captivated researchers for decades and for good reasons – its study illuminates basic facts about the processes of memory formation and retrieval. The complex and often opposing effects of word frequency on these processes have proven to be a difficult challenge for modelers to this day. After reviewing a broad range of results across a variety of memory paradigms, we have identified three key principles, which any theory of the effects of frequency on memory should account for – the dual process principle, the probe dependency principle and the resource demands principle.

VIII. References

Anderson, J. R., & Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology General*, *128*, 186–197.

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.

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.

Bowles, N. L., & Poon, L. W. (1982). An Analysis of the Effect of Aging on Recognition Memory. *Journal of Gerontology*, *37*(2), 212–219.

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.

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

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

Clark, S. E., & Burchett, R. E. R. (1994). Word frequency and list composition effects in associative recognition and recall. *Memory & Cognition*, *22*(1), 55–62

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

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), 545–590

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

Criss, A. H., & McClelland, J. L. (2006). Differentiating the differentiation models: A comparison of the retrieving effectively from memory model (REM) and the subjective likelihood model (SLiM). *Journal of Memory and Language*, *55*(4), 447–460

Criss, A. H., & Shiffrin, R. M. (2004). Interactions between study task, study time, and the low-frequency hit rate advantage in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(4), 778.

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

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

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.

Donaldson, W. (1996). The role of decision processes in remembering and knowing. *Memory & Cognition*, *24*(4), 523–533.

Dorfman, D., & Glanzer, M. (1988). List composition effects in lexical decision and recognition memory. *Journal of Memory and Language*, *27*(6), 633–648.

Dunn, J. C. (2004). Remember-know: a matter of confidence. *Psychological review*, *111*(2), 524.

Ghoneim, M. M. (2004). Drugs and Human Memory (Part 1) Clinical, Theoretical, and Methodologic Issues. *The Journal of the American Society of Anesthesiologists*, *100*(4), 987–1002.

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

Glanzer, M., & Adams, J. K. (1985). The mirror effect in recognition memory. *Memory & Cognition*, *13*(1), 8–20

Glanzer, M., Adams, J. K., Iverson, G. J., & Kim, K. (1993). The regularities of recognition memory. *Psychological Review*, *100*(3), 546–567

Glanzer, M., Hilford, A., & Kim, K. (2004). Six Regularities of Source Recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1176–1195

Gregg, V. (1976). Word frequency, recognition and recall. In *Recall and recognition* (pp. x, 275–x, 275). Oxford, England: John Wiley & Sons.

Gregg, V., 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.

Healey, M. K., & Kahana, M. J. (2016). A four-component model of age-related memory change. *Psychological Review*, *123*(1), 23.

Healey, M. K., Long, N. M., & Kahana, M. J. (2018). Contiguity in episodic memory. *Psychonomic Bulletin & Review*, 1–22.

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

Hintzman, D. L., Caulton, D. A., & Curran, T. (1994). Retrieval constraints and the mirror effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(2), 275.

Hirshman, E., & Arndt, J. (1997). Discriminating alternative conceptions of false recognition: The cases of word concreteness and word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(6), 1306.

Hirshman, E., Fisher, J., Henthorn, T., Arndt, J., & Passannante, A. (2002). Midazolam amnesia and dual-process models of the word-frequency mirror effect. *Journal of Memory and Language*, *47*(4), 499–516

Hirshman, E., Passannante, A., & Arndt, J. (2001). Midazolam amnesia and conceptual processing in implicit memory. *Journal of Experimental Psychology: General*, *130*(3), 453.

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

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.

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

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

Joordens, S., & Hockley, W. E. (2000). Recollection and familiarity through the looking glass: When old does not mirror new. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *26*(6), 1534–1555.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, *24*(1), 103–109.

Karlsen, P. J., & Snodgrass, J. G. (2004). The word-frequency paradox for recall/recognition occurs for pictures. *Psychological Research-Psychologische Forschung*, *68*(4), 271–276

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

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

Miller, L. M., & Roodenrys, S. (2012). Serial recall, word frequency, and mixed lists: The influence of item arrangement. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(6), 1731–1740

Mintzer, M. Z. (2003). Triazolam-induced amnesia and the word-frequency effect in recognition memory: Support for a dual process account. *Journal of Memory and Language*, *48*(3), 596–602.

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.

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

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

Oberauer, K., Lewandowsky, S., Awh, E., Brown, G. D. A., Conway, A., Cowan, N., … Ward, G. (2018). Benchmarks for models of short-term and working memory. *Psychological Bulletin*, *144*(9), 885–958

Osth, A. F., & Dennis, S. (2015). Sources of Interference in Item and Associative Recognition Memory. *Psychological Review*

Osth, A. F., Fox, J., McKague, M., Heathcote, A., & Dennis, S. (2018). The list strength effect in source memory: Data and a global matching model. *Journal of Memory and Language*, *103*, 91–113.

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

Popov, V., & Reder, L. M. (2020). Frequency effects on memory: A resource-limited theory. *Psychological review*, *127*(1), 1-46

Popov, V., So, M., & Reder, L. (2019). Word frequency affects binding probability not memory precision. *PsyArXiv*. https://doi.org/10.31234/osf.io/deyjm

Poppenk, J., & Norman, K. A. (2012). Mechanisms supporting superior source memory for familiar items: A multi-voxel pattern analysis study. *Neuropsychologia*, *50*(13), 3015–3026

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., 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., 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)

Reder, L. M., Proctor, I., Anderson, J. R., Gyulai, F., Quinlan, J. J., & Oates, J. M. (2006). Midazolam does not inhibit association formation, just its storage and strengthening. *Psychopharmacology*, *188*(4), 462–471.

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.

Rundus, D. (1971). Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, *89*(1), 63.

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

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

Serra, M., & Nairne, J. S. (1993). Design controversies and the generation effect: Support for an item-order hypothesis. *Memory & Cognition*, *21*(1), 34–40.

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

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

Stretch, V., & Wixted, J. T. (1998). On the difference between strength-based and frequency-based mirror effects in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1379.

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.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology/Psychologie Canadienne*, *26*(1), 1–12

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), 186–210.

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.

Wixted, J. T., & Stretch, V. (2004). In defense of the signal detection interpretation of remember/know judgments. *Psychonomic Bulletin & Review*, *11*(4), 616–641.

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.

IX. Figures

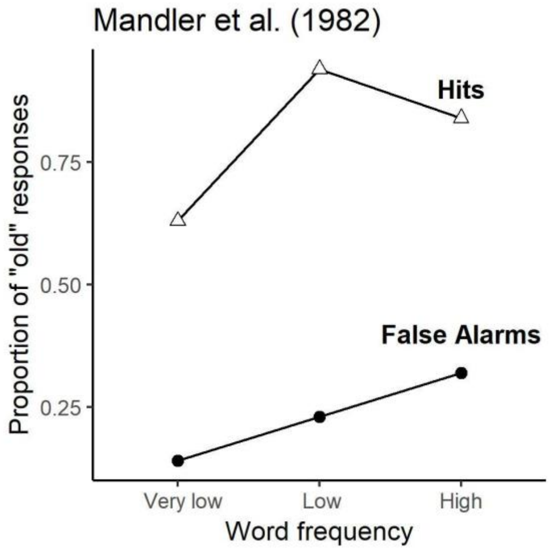

Figure 1. Proportion of Hits and False Alarms in an item recognition task as a function of word frequency. Data from Experiment 1 of Mandler et al (1982).

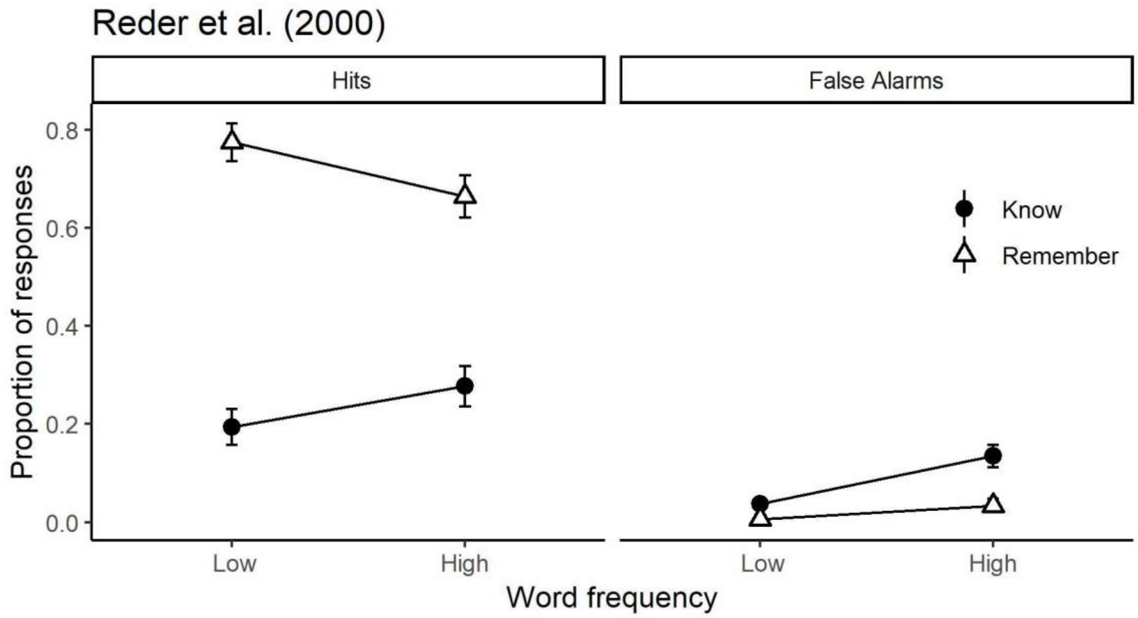

Figure 2. *Proportion of Remember and Know judgements for old and new words in an item recognition task as a function of word frequency. Data from Experiment 1 of Reder et al. (2000).*

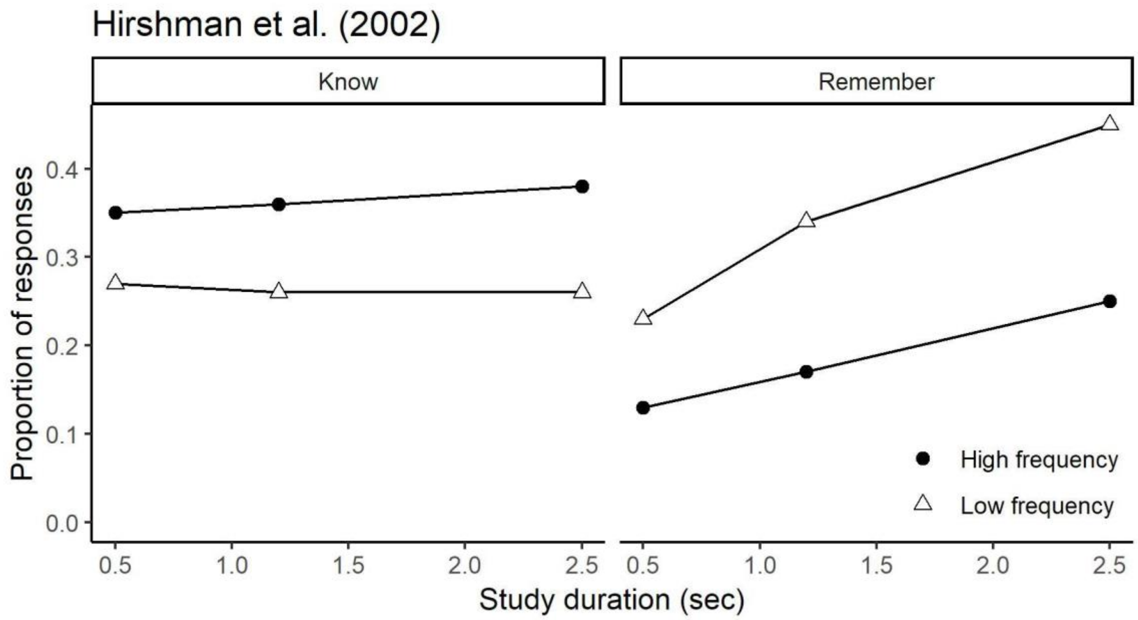

Figure 3. *Proportion of Remember and Know responses in an item recognition task as a function of word frequency and study duration. Data from Hirshman et al. (2002)*

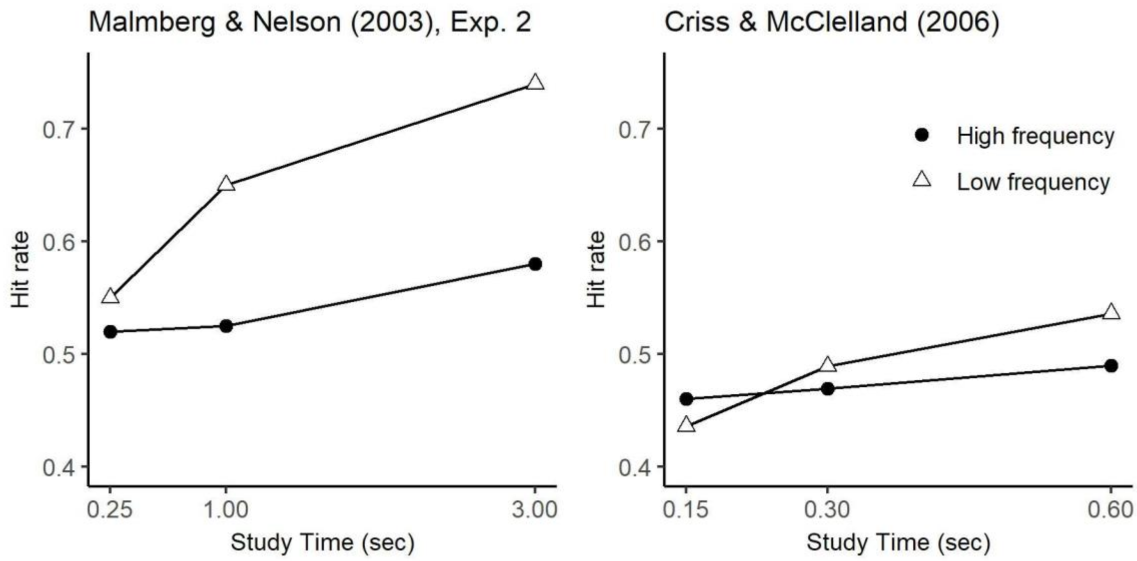

Figure 4. Hit rates for high and low frequency words in an item recognition task depending on the study duration. Left – data from Experiment 2 of Malmberg & Nelson (2003); right – data from Criss & McClelland (2006).

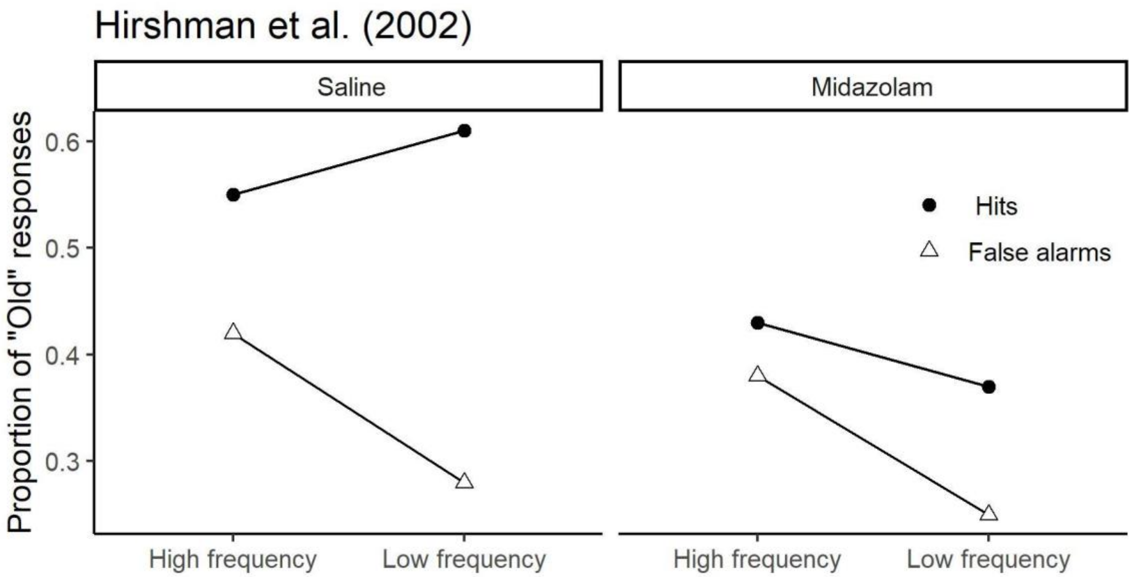

Figure 5. *Proportion of Hits and False alarms in an item recognition task as a function of word frequency and drug condition. Data from Hirshman et al. (2002).*

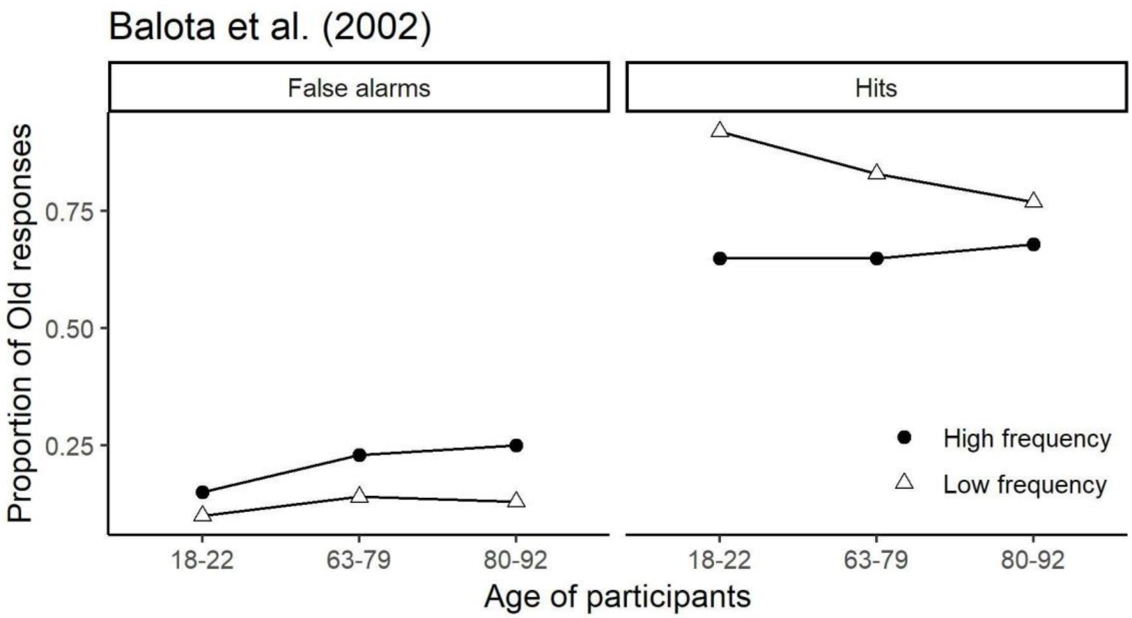

Figure 6. *Proportion of Hits and False alarms in an item recognition task as a function of word frequency and age of participants. Data from Balota et al. (2002).*

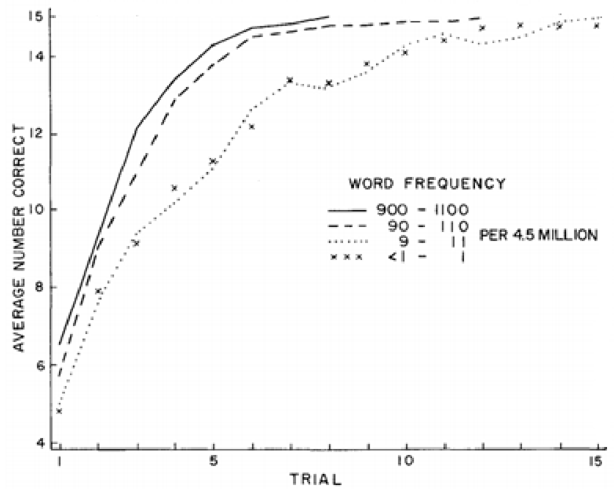

Figure 7. *Number of recalled words from lists of 15 words as a function of the number of study attempts per list ("trials"). Participants had to learn each list to criterion of perfect recall. Figure reproduced from Sumby (1963) with permission. Copyright 1963 by Elsevier*

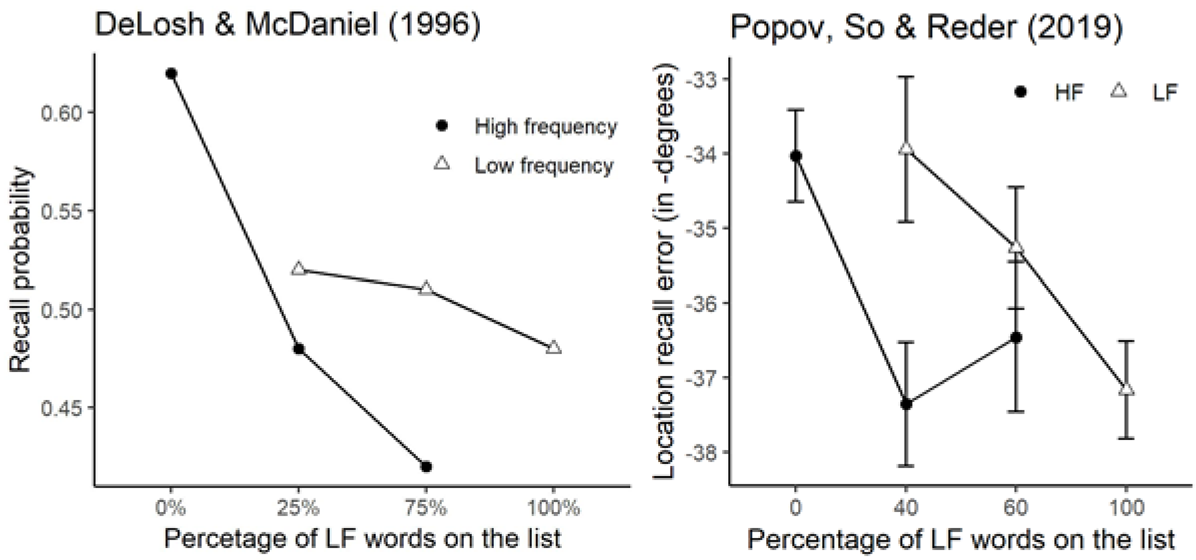

Figure 8. *List composition effects in free recall (left) and source memory (right) – performance as a function of word frequency and the percentage of LF words on the study list. Left panel shows data from DeLosh & McDaniel (1996). Right panel shows data from Popov et al (2019). The y-axis in the right panel shows the inverted (negative) location recall error in – degrees so that higher numbers represent less error and better performance.*

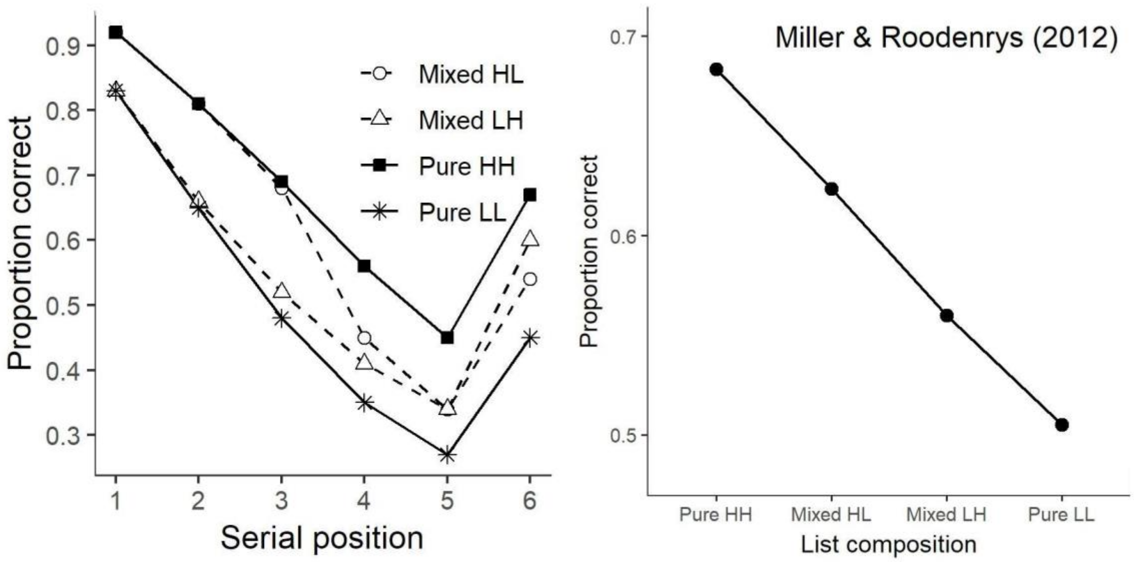

Figure 9. *Serial recall as a function of list composition. Both panels show the same data, but the left panel splits performance by serial position and frequency. Data from Miller & Roodenrys (2012).*

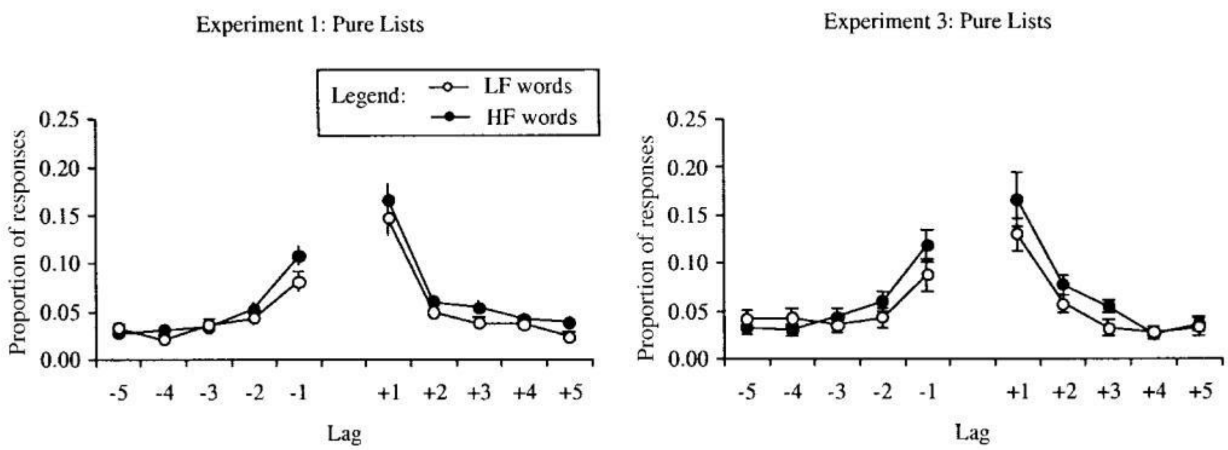

Figure 10. *Proportion of words recalled at each study lag from the previously recalled word. For example, a lag of +2 means that the currently recalled word was studied 2 trials after the previously recalled word. Figure adapted from Ward et al. (2003). Copyright 2003 by the American Psychological Association.*

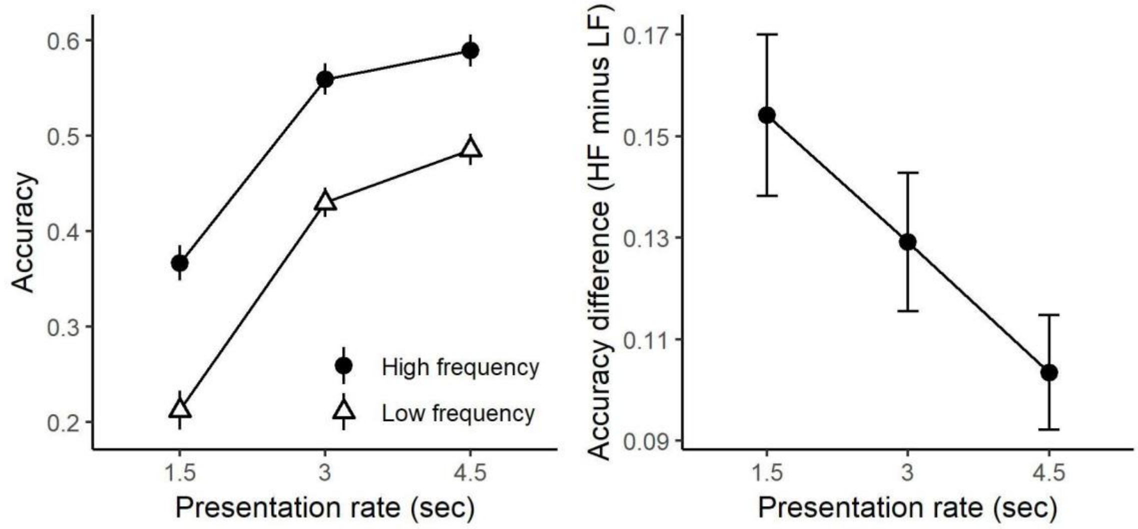

Figure 11. Left – cued recall accuracy as a function of word frequency and presentation rate. Right panel shows the accuracy difference between the HF and LF as a function of presentation rate. Data from an unpublished experiment by Popov, So & Reder. In this experiment, 24 participants studied 30 lists of word pairs. Each list contained 12 pairs that were either only HF or LF, and all words within a list were presented for the same duration, 1.5, 3 or 4.5 s. Immediately after each study list, participants were given one word from each pair, and they had to type the associated word. The words in each list and each pair were drawn at random for each participant, and the order of lists and conditions was also randomized.

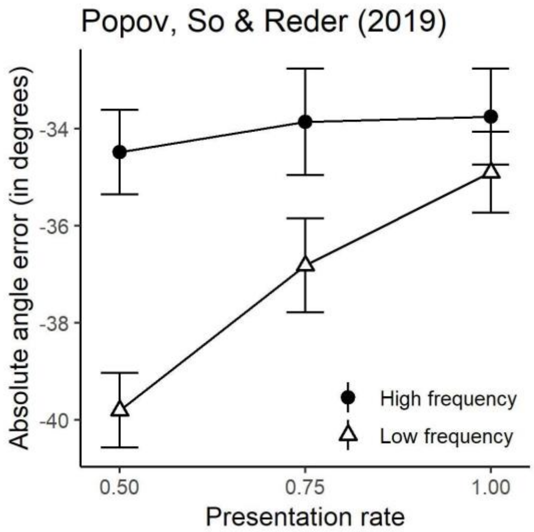

Figure 12. *Source memory performance as a function of word frequency and presentation rate. The y-axis shows the inverted (negative) location recall error in –degrees so that higher numbers represent less error and better performance. Data from Popov et al. (2019)*

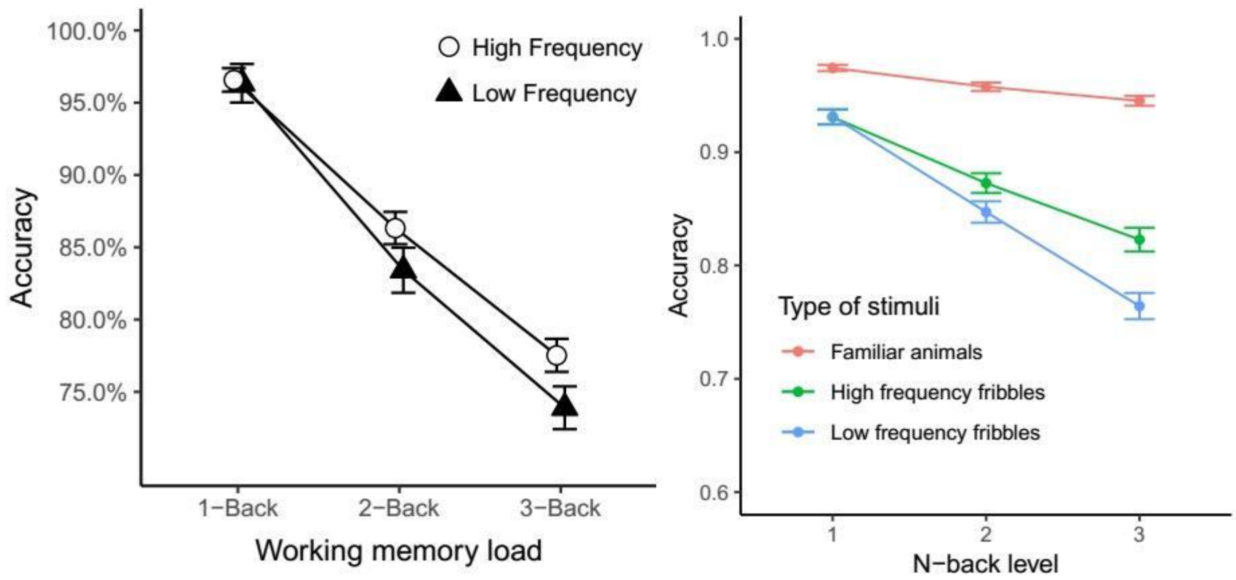

Figure 13. *N-back accuracy as a function of n-back level (working memory load) and stimulus training frequency. Data from Reder et al. (2016). Figures reproduced with permission from Popov & Reder (2020). Copyright 2019 by the American Psychological Association*

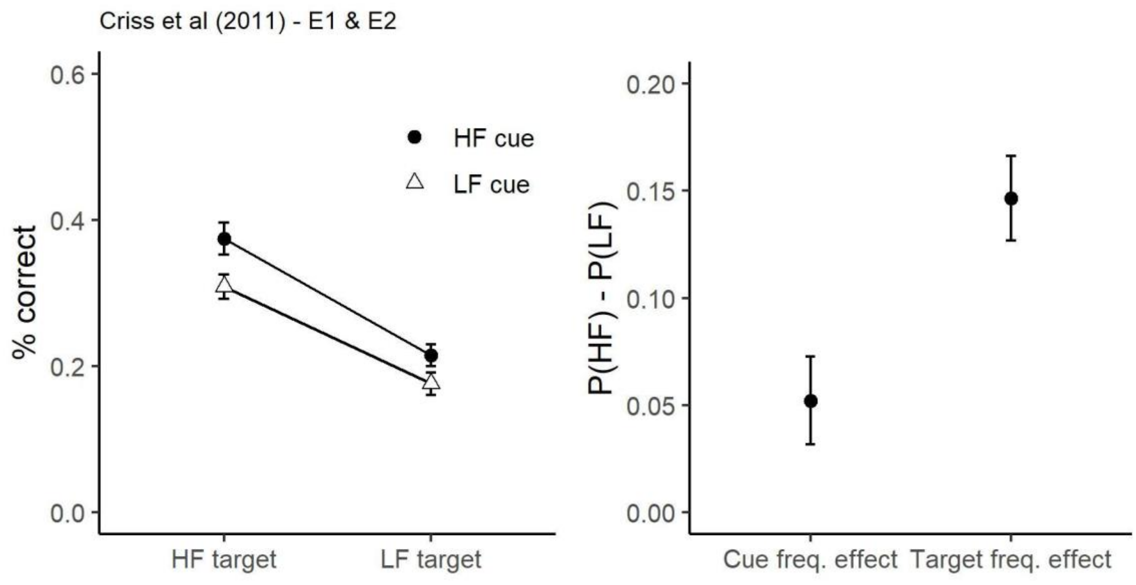

Figure 14. *Left – cued recall performance depending on the word frequency of the cue and the target. Right panel shows the size effect of the word frequency of the cue and the target separately. Data from Criss et al (2011).*

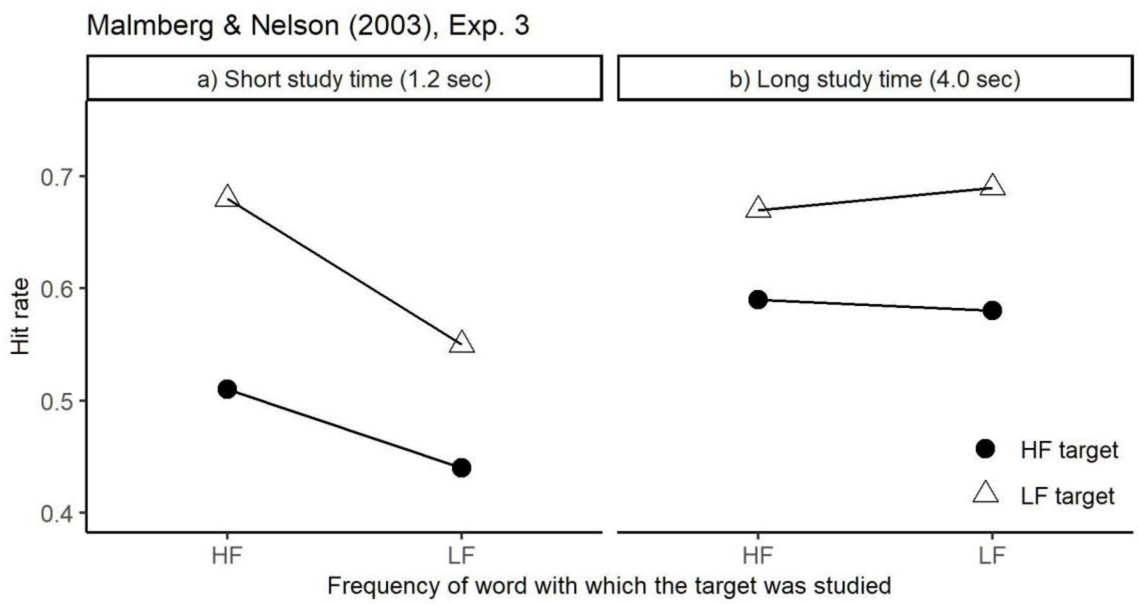

Figure 15. *Hit rate for LF and HF targets, depending on the frequency of the other word in the study pair and the study duration. Data from Malmberg & Nelson (2003, Exp. 3).*

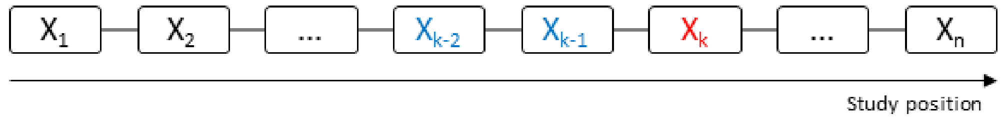

Figure 16. Order of items during a study list

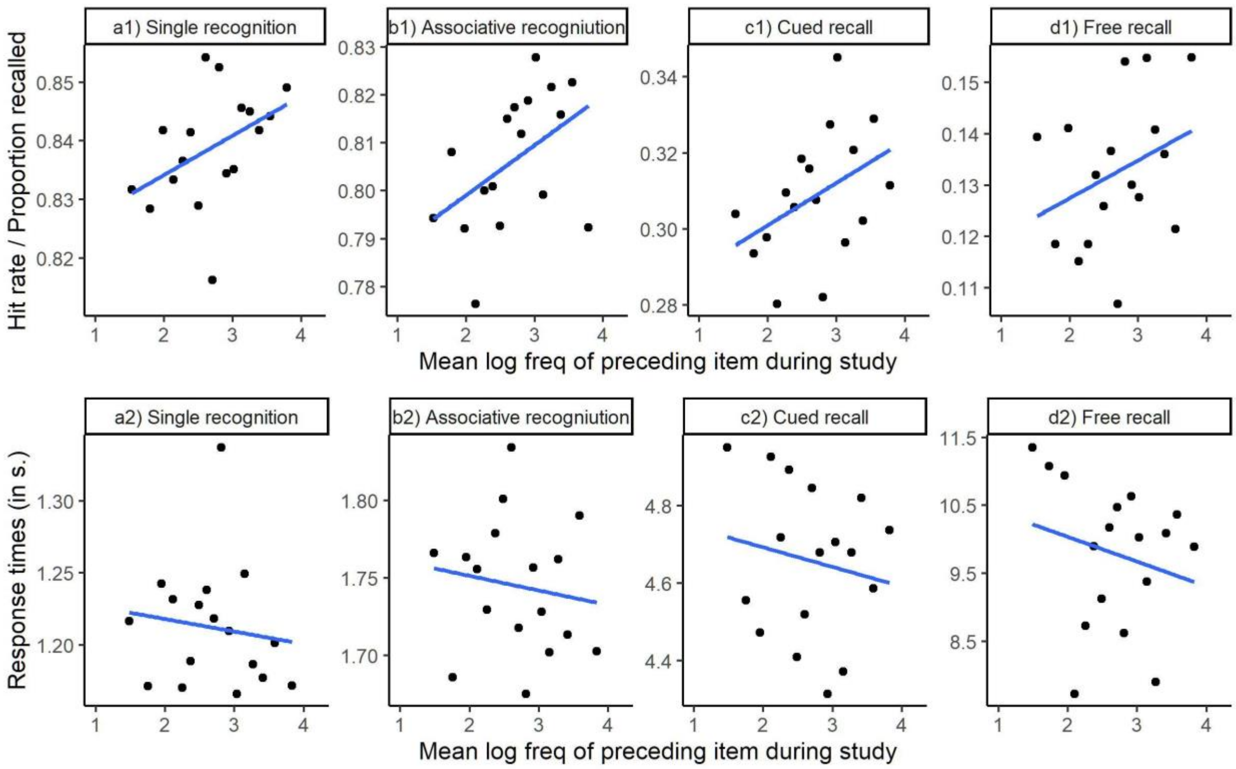

Figure 17. Reanalysis of Cox et al (2018) by Popov & Reder (2019). Hit rates or probability of recall (top panels) and RTs for correct responses (bottom panels) as a function of the mean word frequency of the word pair that preceded the current pair during study and task type - a) Single recognition, 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. Lines show the best fitting regression line to the data.

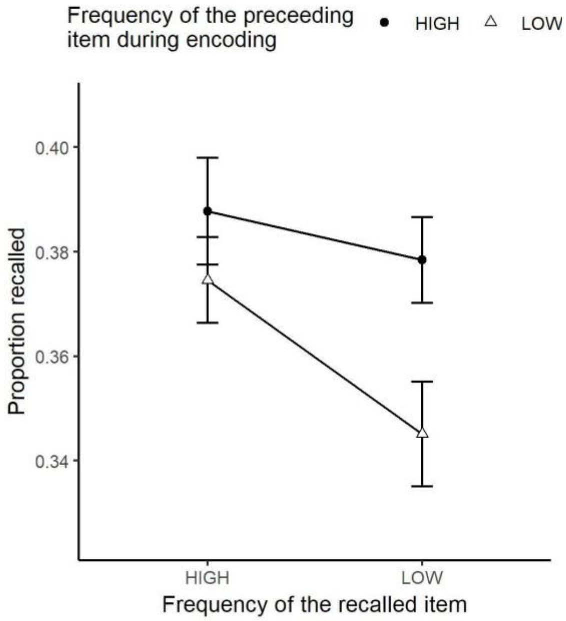

Figure 18. Popov & Reder's (2019) of data by Ward et al. (2003). Proportion free recalls of HF or LF words depending on whether they were preceded by a HF or LF word during study.

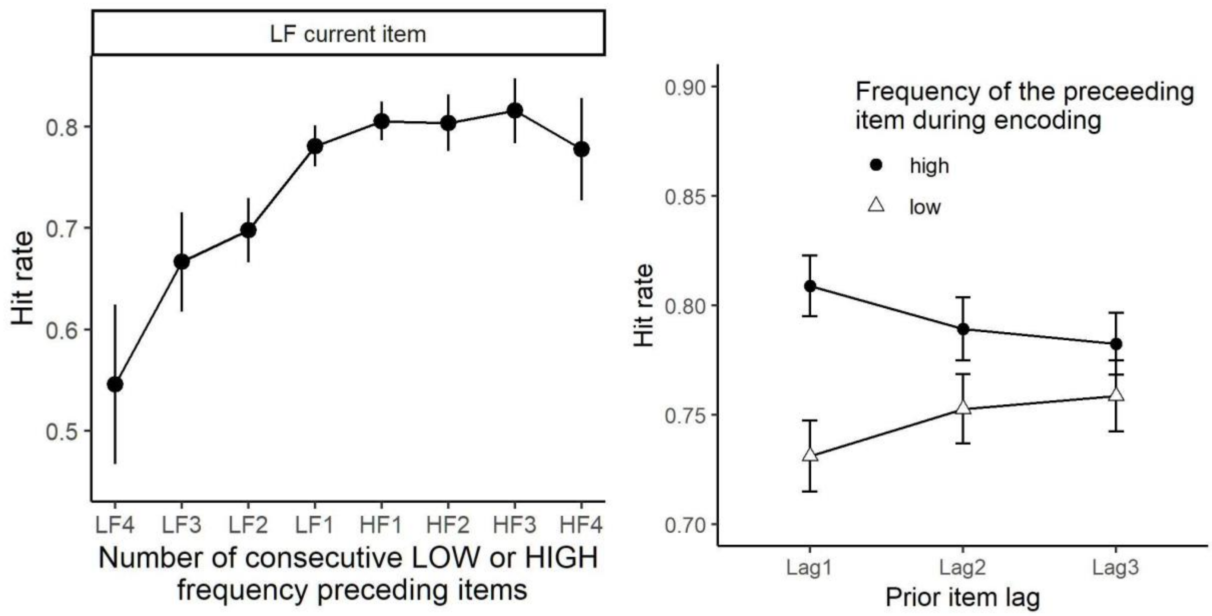

Figure 19. Popov & Reder's (2019) reanalysis of data from Diana & Reder (2006). Left panel – hit rate for pictures as a function of how many previous pictures during the study list had LF or HF words superimposed on them. Right panel - the effect of the preceding study item frequency decreases with study lag to the tested item.

Table 1. Findings about the effects of preceding study item frequency/strength

#	Predictions	Studies									
		Diana*	**Ward**	**Buchler**	**Aue**	**Reder**	**Marevic**	**Popov**	**Cox**	**PEERS**	**Overall**
1	**Discrete effect of prior item strength** *Example:* $P(X_k)$ is worse when X_{k-1} is LF	0	+	+	+	0	+	+	NA	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	+	0	+	0	+	+	+	NA	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	+	+	+	+	0	0	NA	NA	NA	+
5	**Interaction between prior item strength and lag** *Example:* The effect of $freq(X_{k-i})$ should decrease as the lag *i* increases	+	0	+	0	0	+	+	+	+	+

Note. +: effect found in study; 0: null effect; -: effect found in opposite of predicted direction; 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. (in press), Cox = Cox et al. (2018), PEERS = Penn Electrophysiology of Encoding and Retrieval Study (Healey & Kahana, 2016), Overall = Estimate from a meta-analytic mixed-effects regression. Table reproduced with permission from Popov & Reder (2019). Copyright 2019 by the American Psychological Association.

X. Tables