J Exp Psychol Learn Mem Cogn. Author manuscript; available in PMC 2008 May 20.

Published in final edited form as:
J Exp Psychol Learn Mem Cogn. 2002 January ; 28(1): 138–152.

A Reexamination of Stimulus-Frequency Effects in Recognition: Two Mirrors for Low- and High-Frequency Pseudowords

Lynne M. Reder, **Paige Angstadt**, **Melanie Cary**, **Michael A. Erickson**, and **Michael S. Ayers**
Carnegie Mellon University

Abstract

The word-frequency mirror effect (more hits and fewer false alarms for low-frequency than for high-frequency words) has intrigued memory researchers, and multiple accounts have been offered to explain the result. In this study, participants were differentially familiarized to various pseudowords in a familiarization phase that spanned multiple weeks. Recognition tests given during the first week of familiarization replicated a result of W. T. Maddox and W. K. Estes (1997) that failed to show the classic word-frequency mirror effect for pseudowords; however, recognition tests given toward the end of training showed the classic mirror pattern. In addition, a stimulus-frequency mirror effect for "remember" vs. "know" judgments was obtained. These data are consistent with an account of the mirror effect that posits the involvement of dual processes for episodic recognition.

As researchers strive to develop more complete models of human memory, one benchmark used to measure the adequacy of the models has been the degree to which they can account for mirror effects observed in simple verbal learning paradigms. An example of a mirror effect is the word-frequency mirror effect (Glanzer & Adams, 1985, 1990; Glanzer, Adams, Iverson, & Kim, 1993), which refers to the phenomenon that the hit rate (correct recognition judgments for presented items) is higher for low-frequency words than for high-frequency words (e.g., Balota & Neely, 1980; Gorman, 1961; Kinsbourne & George, 1974; McCormack & Swenson, 1972; Schulman, 1967), whereas the false-alarm rate (spurious recognition judgments for items not studied) is higher for high-frequency words than for low-frequency words (e.g., Glanzer & Adams, 1985, 1990; Glanzer et al., 1993). When hits and false alarms are plotted with word frequency on the abscissa, the functions are mirror images—hence the name.

One class of models that has strived to account for mirror effects such as the word-frequency mirror effect is referred to as global memory models (e.g., Hilford, Glanzer, & Kim, 1997; Hintzman, 1994; Hirshman, 1995; Kim & Glanzer, 1993; Maddox & Estes, 1997; McClelland & Chappell, 1998). These models involve a single process and, in their basic form, assume that the memory strength or familiarity of an item is the factor that affects recognition memory. For these models to account for the word-frequency mirror effect, they often require the postulation of auxiliary assumptions such as differential stimulus salience and different response criteria for different word classes (e.g., Gillund & Shiffrin, 1984; Hintzman, 1988) or differential attention during encoding to words of different frequency classes (e.g., Glanzer & Adams, 1990; Glanzer, Adams, & Iverson, 1991).

Correspondence concerning this article should be addressed to Lynne M. Reder, Department of Psychology, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213. E-mail: reder@cmu.edu.

Lynne M. Reder, Melanie Cary, Michael A. Erickson, and Michael S. Ayers, Department of Psychology, Carnegie Mellon University; Paige Angstadt, Department of Computer Science, Carnegie Mellon University.

Paige Angstadt is now at Facset Research Systems, Greenwich, Connecticut. Michael S. Ayers is now at Canton Central School District, Canton, New York.

A competing account for these memory effects comes from dual-process models (e.g., Jacoby & Dallas, 1981; Mandler, 1980; Reder et al., 2000; Yonelinas, 1994). Reder et al. (2000) proposed a formal dual-process account for the word-frequency mirror effect. The account is embedded within the SAC (source of activation confusion) theory of memory. The SAC account of the word-frequency mirror effect also makes novel predictions concerning the pattern of “remember” and “know” judgments as a function of word frequency. The computer simulation based on the theory closely fitted individual subject “remember”–“know” data as well as “old”–“new” responses as a function of normative and experimental word frequency. SAC has also been used to account for other phenomena such as feeling of knowing, strategy selection, and the misinformation effect (Ayers & Reder, 1998; Reder & Schunn, 1996; Schunn, Reder, Nhouyvanisvong, Richards, & Stroffolino, 1997). It is also being extended to account for other mirror effects such as list length and list strength (Cary & Reder, 2001). One of the goals of our study was to test whether the SAC account of the memory effect would be confirmed using artificial stimuli, that is, pseudowords of different prefamiliarization.

The SAC Model: A Dual-Process Theory of Recognition

SAC incorporates the notion made by a number of memory theorists (e.g., Jacoby, 1991; Jacoby & Dallas, 1981; Mandler, 1980; Yonelinas, 1994) that recognition judgments can be based on one of two processes: familiarity or recollection (which are distinguished from the single-process views mentioned above). Recollection is conceived of as retrieving the trace that encoded the experience. In the case of a test of recognition memory, this trace would have encoded the fact that the word was studied in the experimental context (i.e., that it was on the study list). When a specific recollection of contextual information cannot be retrieved, the recognition judgment is based on familiarity. If the concept associated with the test item seems very familiar, a positive recognition can also be made on that basis.

SAC goes beyond earlier two-process theorizing by precisely specifying the mechanisms of these two processes and explaining what variables affect each of them. A formal, mathematical specification of the two-process recognition model can be found in Reder et al. (2000); however, here we outline the basic concepts. According to SAC, when an item is encoded in memory, either as part of an experiment or in a more natural setting, there are typically two consequences of the encoding. First, the strength of the node representing that item, the *concept node*, is increased. Strength is determined by prior history of exposures and translates into resting level of activation. The availability of a concept depends on its activation. Second, a node representing the encoding episode is built. This encoding *episode node* represents the knowledge about where and when the word was encoded. When the episode node is created, an association is formed between it and the concept node. The episode node is also associated with the node that represents the context, in this case, the list context. The list *context node* has contextual features that are processed at the same time as the list items. Figure 1 schematically illustrates how we represent the memory traces for a high- and low-frequency word studied on the first list in an experiment.[1]

According to SAC, words with greater normative frequency (i.e., a larger number of prior exposures) will have stronger conceptual representations in memory and have more contextual associations than will words of lower frequency. When a probe is presented during a recognition test, its concept node is activated and the total current activation of that node spreads to all associated nodes via the links that emanate from it. The amount of activation that spreads to a given receiving node is a function of the strength of the link from the sending to

---

[1] Just as the context node can be unpacked into its constituent or associated features, so too can the concept node. Associated with the concept node is lexical information such as its phonemic and orthographic information, semantic information such as related concepts and its component features, and contextual information such as previous and current encoding events.

the receiving node compared with the sum of the strengths of all of the competing links emanating from that sending node.

In other words, the amount of activation that spreads to any associated node from a node with many links will tend to be less than the amount spread to a node from one with only a few associated links.[2] Given that high-frequency words have been seen in more contexts and thus have more competing contextual associations, the amount of activation that can spread to the relevant episode node will be less from a high-frequency concept node than from a low-frequency concept node. According to SAC, recollection-based responses occur when sufficient activation arrives at the associated episode node. Therefore, we predicted fewer recollection-based judgments for high-frequency words, because of more contextual competition.

When the relevant episode node does not receive enough activation to pass threshold for recollection, a familiarity-based recognition may occur based on the strength of the concept node. According to SAC, high-frequency words will produce more familiarity-based responses, both when the word has been studied (hit) and when it has not (false alarm). There are two reasons for this. First, the base level of activation is greater for high-frequency words than for low-frequency words, making it easier to pass threshold when the item has not been recently studied (false alarms). Second, successful retrieval of the encoding episode is less likely for high-frequency words because of their greater contextual fan, leaving only the familiarity-based process for responding.

“Remember”–“Know” Responses

The predicted result that low-frequency words are more likely than high-frequency words to elicit recollection-based responses has been found several times in the literature (e.g., Gardiner & Java, 1990; Huron et al., 1995; Strack & Forster, 1995); however, the prediction that high-frequency words are more likely than low-frequency words to elicit familiarity-based responses was a novel prediction of SAC (Reder et al., 2000).

This conceptualization is testable using the “remember”–“know” procedure (Gardiner, 1988; Tulving, 1985), which requires participants to assess whether their “old” response was based on a recollection of experiencing the item on the study list or whether it was based on a feeling of familiarity. When participants feel that they recollect the study episode they respond “remember.” When they feel that the item is sufficiently familiar that it must have been on the study list but have no specific recollection of studying the item, they respond “know.” Although it might seem surprising that participants can actually discriminate between “remember” and “know” judgments, the procedure has been used many times with considerable success in terms of separating these two types of processes (e.g., Gardiner, 1988; Gardiner & Java, 1990; Gardiner & Parkin, 1990; Rajaram, 1993). Given the representational assumptions of SAC, participants are expected to give more “remember” responses and fewer “know” responses to lower frequency items than to higher frequency items for hits (i.e., a mirror pattern). In addition, SAC also predicts more “know” judgments for false alarms to high-frequency items than to low-frequency items. These predictions were confirmed with real words in Reder et al. (2000), and we anticipated that these results would generalize to artificial stimuli.

Given the straightforward predictions of SAC regarding “remember”–“know” responses, we decided to include the “remember”–“know” procedure in this study. That is, in addition to making “old”–“new” judgments, participants were also asked to make “remember”–“know”

---

[2] One might wonder whether the stronger base strength of a high-frequency concept node would cancel its disadvantage of higher fan. In the simulations of words of differing frequency, this did not happen. Consult Reder et al. (2000).

judgments on their “old” responses. The “remember”–“know” data should also provide insights into the relative frequency of recollection-based and familiarity-based “old” judgments as participants learn pseudowords presented with different degrees of frequency.

Direct Versus Indirect Memory Effects

On the one hand, the “remember”–“know” predictions follow transparently from SAC, and it is less clear how single-process models of memory would account for them. On the other hand, there is reason to doubt whether the pattern found by Reder et al. (2000) would replicate with pseudowords. Maddox and Estes (1997) explored the effects of frequency on recognition performance for pseudowords and failed to find the classic mirror pattern. They manipulated stimulus familiarity by varying the frequency of exposure to pseudowords in a familiarization phase that occurred prior to the recognition memory study–test phases. Instead of finding a mirror pattern for their pseudoword stimuli, they observed a concordant increase in both hit rate and false-alarm rate as a function of familiarization frequency.

Maddox and Estes’s (1997) concordant pattern of results was taken as support for global memory or single-process models, which tend to have difficulty explaining the mirror pattern. In their view, direct effects of stimulus frequency involve strengthening an item’s memory representation each time the item is encountered and will not produce the hit portion of the mirror effect (i.e., more hits for low-frequency than for high-frequency words). Indirect effects of stimulus frequency, as explained by Glanzer’s attention-likelihood model (Glanzer & Adams, 1990; Glanzer, Adams, & Iverson, 1991), occur when higher frequency items are allocated less attention and, thus, are encoded less well than lower frequency items. When that happens, low-frequency words will produce a higher hit rate than will high-frequency words. Maddox and Estes concluded that the reason mirror effects are seen with real words is that stimulus frequency can affect recognition performance both directly and indirectly and the mirror effect occurs only when there are also indirect effects.

An alternative account of the Maddox and Estes’s (1997) findings is tested in our study. It is possible that Maddox and Estes did not obtain a mirror pattern because of the limited amount of familiarization to the novel stimuli. Items in their study received from 1 to 16 exposures, which is a relatively restricted range of fairly low frequencies. Although a mirror pattern is ubiquitously observed when recognition of low- and high-frequency words are compared, Wixted (1992) demonstrated that the mirror pattern does not obtain when recognition of very-low-frequency words (i.e., rare words) is compared to recognition of low-frequency words. Very-low-frequency words tend to be recognized less accurately (i.e., have a lower d') than low-frequency words (e.g., Mandler, Goodman, & Wilkes-Gibbs, 1982; Rao & Proctor, 1984; Wixted, 1992). Thus, whereas Maddox and Estes made an analogy between their different levels of stimulus familiarization and low-and high-frequency words, their manipulation may have been more similar to using levels of very-low-frequency and low-frequency words.

We manipulated stimulus frequency for pseudowords by presenting the items for study a different number of times. In general, the number of familiarization trials was far greater than the number that had been used in previous studies. Participants studied pseudowords for 20 familiarization sessions over 5 weeks, for a total of 10 to 360 exposures (cf. Maddox & Estes, 1997).

The results of a study by Gardiner and Java (1990) indicate that familiarity plays a larger role than recollection in the recognition of pseudowords but that for real words the pattern is reversed. Gardiner and Java’s participants gave more “know” responses than “remember” responses to pseudoword targets, but they gave more “remember” responses than “know” responses to word targets. (Their participants saw each pseudoword only once.) We conjectured that recollection-based processes may become more influential in recognition as novel stimuli become familiar. In a consistent vein, Joordens and Hockley (2000) argued that the mirror pattern occurs when recollection-based recognition dominates over familiarity-based recognition. Guttentag and Carroll (1997) also emphasized the role of recollection-based processes in producing the hit portion of the mirror pattern. Therefore we hypothesized that from early to late in our study, as experience with the pseudowords increased, the overall proportion of “remember” hits would increase and, as a result, the overall proportion of “know” hits would decrease.

In sum, we predicted that the pattern of hits and false alarms on the recognition tests early in the study would replicate the concordant pattern found by Maddox and Estes (1997), because early in the study our frequency manipulation was similar to theirs. However, we also predicted the standard recognition mirror effect of more hits and fewer false alarms for low-frequency items would emerge when the pseudowords received sufficient exposure. Likewise, we predicted that a stimulus-frequency mirror effect for “remember” versus “know” judgments also would emerge, such that low-frequency pseudowords would elicit more “remember” hits and fewer “know” hits than would high-frequency pseudowords.

Experiment

We manipulated familiarization frequency by varying the number of exposures to a given pseudoword during each acquisition cycle. We defined an *acquisition cycle* as two consecutive familiarization sessions that each contained three free-recall study lists. Free recall was not the major focus of this study; however, free-recall study and test cycles were critical to the experiment, because the free-recall study lists were used to manipulate familiarization of the pseudowords.

We also manipulated contextual fan by varying the number of study lists (per acquisition cycle) that contained a particular pseudoword. As discussed earlier, within SAC the number of associations emanating from a concept node influences the probability of that item eliciting a recollection-based recognition. Moreover, each exposure to an item in a new context leads to an additional association from the relevant concept node to a new episode node. This idea is similar to the notion that acquiring an additional fact about a concept leads to an additional conceptual association from the relevant concept node. Whereas the term *fan* has typically been used to refer to the number of facts associated with a particular concept (e.g., Anderson, 1974, 1976), the term is used more broadly here to apply to the number of contexts associated with a particular concept. Reder et al. (2000) used the idea of preexisting contextual fan in their account of the word-frequency mirror effect and to predict the word-frequency “remember”–“know” mirror effect.

Here, the term *contextual fan* refers to the relative number of familiarization lists on which an item was presented. Items in the *low contextual fan* conditions were presented on only one list per cycle, and items in the *high contextual fan* conditions were presented on all six lists per cycle. Conceivably pseudowords that were presented six times on one list per cycle would have as strong a base level of activation but less contextual fan than pseudowords that were presented one time on each of six distinct lists per cycle. When items have equivalent base levels of activation and different degrees of contextual fan, SAC predicts that the items with the greater contextual fan should produce worse recognition performance. Of course we realize that items presented only once per list might have been at an advantage for several reasons. First, there is a vast literature supporting the idea that distributed practice produces better retention (e.g., Glenberg, 1979; Greene, 1989; Reder, Charney, & Morgan, 1986). Second, there is evidence that recall is an excellent form of strengthening memory representations (Bjork, 1988). An item presented once per list for each of six lists was given six opportunities to be recalled (i.e., once after each list), whereas when an item's six exposures were massed onto one list, it was given only one opportunity to be recalled. Nevertheless, the contextual fan manipulation was included to provide a preliminary investigation of the role of this factor in recognition memory.

Method

Participants—Nineteen participants were recruited from Carnegie Mellon University by the use of electronic bulletin boards. The participants were paid for each hour of participation and received bonuses based on performance. Most of the earnings were withheld until participants had successfully completed the entire experiment. Pilot studies indicated that several weeks of studying pseudowords can be very boring, and many pilot-study participants lost their interest in trying hard. This problem was no doubt exacerbated by the list-discrimination interference inherent in the design. Therefore, in this experiment, participants were given the following reward structure to motivate performing well: They were paid 1 cent for every correct recall and fined 1 cent for every recall intrusion (i.e., recalling an item that was not on that familiarization list). Likewise on the recognition tests, participants were rewarded 1 cent for every hit and fined 1 cent for every false alarm. Given that the difficulty of the task and the extrinsic, monetary incentive might result in cheating, the experimenters carefully monitored participants' acquisition and test behavior.

Materials and design—Table 1 illustrates the 20 experimental sessions, which are described in more detail below. The sessions were scheduled at each participant's convenience with the constraints that there were exactly 4 sessions per week, no 2 sessions were on the same day, and each session was on a weekday. Each of the 10 acquisition cycles contained six familiarization lists with three familiarization lists presented in each of 2 sessions. The composition of items per list was varied for each cycle.

The stimuli were 80 different one-syllable pronounceable nonwords, which we refer to as *pseudowords.* Each pseudoword consisted of four letters, began with a consonant, and contained either one or two vowels (e.g., *bist, clow, nime, treg*). Items were randomly assigned to one of four familiarization conditions for each participant, and therefore any effects of materials would be pulled out of the analyses as part of the subject error term.

The four familiarization conditions were defined by two independent variables: familiarization frequency and contextual fan. Each row in Table 2 corresponds to one of these four conditions. The first column in Table 2 identifies the level of frequency (low, medium, or high) for each condition, and the second column identifies the level of contextual fan (low or high) for each condition. As illustrated in the last column, each pseudoword in the low-, medium-, and high-frequency conditions was presented for 1, 6, or 36 exposures per cycle, respectively. Over the course of the 10 acquisition cycles in the study, low-, medium-, and high-frequency pseudowords were presented for a total of 10, 60, and 360 familiarization exposures, respectively. As shown in the third and fourth columns of Table 2, pseudowords in the low contextual fan conditions were presented on only one familiarization list per cycle, whereas pseudowords in the high contextual fan conditions were presented on all six familiarization lists per cycle. The contextual fan manipulation was most relevant for the two medium-frequency conditions, because within the framework of SAC the medium-frequency pseudowords could have obtained equivalent base levels of activation while having different levels of contextual fan.[3]

---

[3] These two medium-frequency conditions were not ideally controlled in that the high-fan condition had spaced practice whereas the low-fan condition had massed practice. In addition, for any critical recognition test, the high-fan pseudowords would have been seen more recently than would the low-fan pseudowords, creating a confound of recency with fan as well as spacing with fan.

In sum, the four familiarization conditions constituted an incomplete 2 (contextual fan) × 3 (familiarization frequency) within-subject factorial design. The design was not a complete factorial because no items were studied only once per cycle on six lists, which is impossible, and no items were studied 36 times per cycle on only one list. The low-frequency/low-contextual-fan condition and the high-frequency/high-contextual-fan condition are often referred to in this article as the low-frequency condition and the high-frequency condition, respectively.

There were three types of memory measures: recall performance, recognition performance, and "remember"–"know" judgments. During the free-recall test that followed each familiarization list, participants were instructed to recall as many items as possible from the immediately preceding familiarization list. In addition to these recall tests, five critical recognition study and test lists occurred throughout the experiment. Their timing with respect to the 10 cycles is denoted in Table 1.

List construction—Each acquisition cycle consisted of six familiarization lists evenly divided across two sessions (see Table 1). As described above, the 20 medium-frequency/high-fan pseudowords were each presented once per list on all six lists. The 20 high-frequency/high-fan pseudowords were each presented six times per list on all six familiarization lists. The low-frequency/low-fan pseudowords were each presented once per cycle on one of the familiarization lists. Because these 20 items could not be evenly assigned to the six lists in a cycle, four lists contained 3 low-frequency items and two lists contained 4 low-frequency items. The 20 medium-frequency/low-fan pseudowords were each presented six times per cycle, with all 6 presentations of an item occurring on one familiarization list. Similar to low-frequency/low-fan items, each familiarization list contained the 6 trials for either 3 or 4 of the medium-frequency/low-fan pseudowords. For each acquisition cycle, pseudowords in the low-fan conditions were randomly assigned to lists based on these constraints. Each set of six familiarization lists consisted of 161, 162, or 167 trials, with two lists of each length per cycle. Hence there were 980 familiarization trials per cycle. The order of presentation of items on a list was randomly determined with the constraint that no pseudoword appeared twice consecutively. The assignment of pseudowords to lists was also randomized for each acquisition cycle, within the constraints imposed by the design.

As shown in Table 1, a recognition study list and a recognition test list were presented during Cycles 1, 2, 4, 6, and 10. Two concerns motivated the uneven spacing of recognition tests. One concern was that giving too many recognition tests would dilute the frequency manipulation, given that items from each condition were tested with equal probability on the recognition test. The countervailing concern was that giving too few recognition tests would not provide a reliable measure of how recognition performance shifts as familiarity changes. The placement of the recognition tests was intended to maximally satisfy these competing constraints. Each recognition study list consisted of half of the pseudowords randomly selected from each of the four familiarization conditions. The recognition test list consisted of all the pseudowords in the study. The presentation order of items within each study and test list was separately randomized.

Procedure—At the beginning of each familiarization session, participants were told that they would see lists of pseudowords[4] on the computer screen and be asked to recall them. The instructions asked participants to read the pseudowords aloud and try to remember them. The instructions also reminded participants of the nature of the monetary reward structure.

---

[4] The pseudowords were referred to as "nonwords" to the participants.

Participants controlled the start of each list by pressing a key on the keyboard. The computer displayed the pseudowords individually for 2 s, followed by a blank screen for 250 ms.

After the participants saw one entire familiarization list, they recalled as many of the items from the preceding list as possible. The recall instructions told participants to recall each item only once, even if it was shown multiple times. Participants typed their responses using the computer keyboard, and the computer displayed each response as it was typed. Participants had unlimited time to recall items from the list. They indicated that they were done recalling by pressing a labeled key on the keyboard. Immediately after pressing this key, the computer displayed feedback on their performance. They were told the number of correct recalls, the number of intrusions, and the bonus earned in that session up to that point. Each familiarization session contained three iterations of this target-familiarization list and free-recall procedure, which took approximately 30 min. To help distinguish the contexts of the different familiarization lists, each of the three lists in a given acquisition session was presented in a different distinctive black font with a different background screen color. The same three fonts and colors were used in each session.

To minimize confusion with the familiarization lists, we conducted the recognition task in a different room from the one in which participants saw and recalled the familiarization items. The recognition task consisted of three phases: study, filler task, and test. Prior to the study list, instructions informed participants that they would see a list of items and would later be tested on them. Participants were also instructed to read each pseudoword aloud. The computer displayed each item for 2 s, followed by a blank screen for 250 ms. So participants could distinguish the recognition lists from the familiarization lists, the recognition study list and recognition test list stimuli were shown on a white background, rather than on a colored background, in a plain black font that was different from the fonts used for the familiarization lists.

After reading all 40 study items and prior to making recognition judgments, participants performed a number-based working memory filler task for 10 min. Immediately following the filler task, participants were required to distinguish study-list items from the other 40 pseudowords in the experiment. Items were presented individually and remained on the screen until all judgments regarding that item were recorded. Participants were instructed to read each item and decide whether they had seen it on the study list. The instructions emphasized that they were to make the judgment solely on the items from the study list, not on the items from any of the colored lists that they saw during the recall task. They were told to press the key labeled *new* if they did not recognize the item from the most recent study list and to press the key labeled *old* if they did recognize the item from that list.

When participants indicated that the item had been on the study list, they also were asked to judge whether they “remembered” seeing the item on the study list, or merely “knew” the item had been on the list. The “remember”–“know” instructions followed as closely as possible those used by Knowlton and Squire (1995, p. 701).

To establish that participants understood the task, the experimenter asked the participants to give one example of their own for each type of judgment. If participants were unable to generate examples, or if the experimenter felt that the examples did not clearly demonstrate understanding of the judgments, the experimenter clarified the instructions sufficiently for the participant to generate adequate examples. Participants generated examples for the first recognition test only. For subsequent recognition tests, participants received the same instructions for “remember” and “know” judgments but were not required to generate examples.

In our study, each recognition test trial proceeded as follows. First, a pseudoword appeared in the center of the screen, along with the word *new* below and to the left of the stimulus and the word *old* below and to the right of the stimulus. To record their “old”–“new” judgment participants pressed the key labeled *NEW* (the *D* key) or the key labeled *OLD* (the *K* key). When the participant judged the item to be *new*, the screen went blank for 1 s, followed by the presentation of the next item. Immediately following an “old” judgment, the words *new* and *old* disappeared and the letters *R* and *K* appeared on the screen, below and closer together on the screen than the *new* and *old* positions. To record their “remember”–“know” judgment, participants pressed the key labeled *R* (the *C* key) or the key labeled *K* (the *M* key). The layout of the keyboard positions matched the layout of the prompts on the screen and was arranged so that participants could keep their index and middle fingers poised over the four keys during the entire test list. Immediately following each recognition test, participants were informed of their hits, false alarms, and bonus earned for that test.

Results

Of the 19 individuals who participated in the study, 12 provided usable data and were included in all analyses. Of the 7 participants whose data are excluded, 3 discontinued the study because of personal schedule conflicts that prevented them from attending the required four sessions per week for 5 weeks. One participant was excluded from the study when it was discovered that he was writing down the pseudowords from the familiarization lists and using these notes to aid recall. The remaining 3 participants were excluded because it was clear from their data that they had made no attempt to respond appropriately to recognition test items (i.e., they tended to respond “new” to every item).

The first part of this section discusses recall performance as a function of acquisition cycle and familiarization condition to verify that participants actually learned the pseudowords. The second part reports overall recognition performance as a function of acquisition cycle and familiarization condition. In the remaining parts, the recognition data are discussed in greater detail. The organization of these data analyses also involves an initial focus on the critical comparisons of items in the low-frequency and high-frequency conditions, with a later section that discusses the data from the two medium-frequency conditions.

Recall—Of interest is whether participants actually learned the pseudowords in the experiment. Table 3 presents the mean proportion of correctly recalled target-list items and extraexperimental intrusions by week, as well as reporting recall by condition, also by week. A two-factor analysis of variance (ANOVA)—4 (Levels of Familiarization Condition) × 5 (Weeks of Study–Test)—was conducted on the proportion of correctly recalled target items. Participants reliably recalled more words each week over the course of the study, $F(4, 44) = 48.90$, $MSE = 0.03$, $p < .001$. They recalled less than 50% of the pseudowords during the first week, but they recalled 80% or more target list items during the last 3 weeks. These data provide evidence that participants were actually learning the pseudowords during the experiment. The intrusion data discussed below also support this conclusion.

The mean proportion of correctly recalled target list items also reliably varied as a function of familiarization condition, $F(3, 33) = 41.01$, $MSE = 0.02$, $p < .001$. Contrasts were conducted to examine the effects of frequency. Participants recalled items from the high-frequency condition ($M = .87$, $SEM = .02$) more accurately than they recalled items from the low-frequency condition ($M = .61$, $SEM = .04$), $F(1, 33) = 94.67$, $MSE = 0.02$, $p < .001$. Holding contextual fan constant, participants recalled medium-frequency/low-fan items ($M = .81$, $SEM = .03$) more accurately than they recalled low-frequency/low-fan items, $F(1, 33) = 54.85$, $MSE = 0.02$, $p < .001$. Likewise, participants recalled the high-frequency/high-fan items more accurately than they recalled the medium-frequency/high-fan items ($M = .67$, $SEM = .04$), F

(1, 33) = 58.50, *MSE* = 0.02, $p < .001$. Medium-frequency/low-fan items were recalled at a higher rate than medium-frequency/high-fan items. This suggests that performance was better, because in the former condition an item was presented six times on a single study list whereas in the latter condition an item was presented only once. Thus, it seems that the differential exposure to stimuli in the four familiarization conditions had the intended effect of making pseudowords differentially familiar to the participants.

The interaction of week and familiarization condition was also reliable, *F*(12, 132) = 14.52, *MSE* = 0.004, $p < .001$. The same pattern of recall accuracy across familiarization conditions occurred during each week of the study; however, the magnitude of the differences was largest during the earlier weeks when performance was further from ceiling.

Another way to examine participants' learning of the experimental stimuli is to look at changes, with practice, in the number of recall errors, or intrusions. Participants made two types of recall intrusions: extraexperimental and intraexperimental intrusions. Extraexperimental intrusions are items that were not presented in any part of the experiment. Intraexperimental intrusions are items that were presented in the experiment but were not presented on the immediately preceding familiarization list (i.e., the to-be-recalled list). As shown in Table 3, the mean number of extraexperimental intrusions per week decreased as a function of week, *F*(4, 44) = 8.62, *MSE* = 84.78, $p < .001$, indicating that participants made fewer extraexperimental intrusions with increasing exposure to the experimental materials. This pattern further indicates that participants successfully learned the pseudowords with increasing accuracy throughout the study.

Intraexperimental intrusions could be either low-frequency/low-fan items or medium-frequency/low-fan items. It was not possible to have intrusions from the two high-fan conditions, because high-fan items occurred on every familiarization list. A 2 (familiarization condition) × 5 (week of study) ANOVA conducted on the mean number of intraexperimental intrusions revealed a significant main effect of familiarization condition, *F*(1, 11) = 8.93, *MSE* = 140.50, $p < .05$. Medium-frequency/low-fan items intruded in recall more often (*M* = 8.37, *SEM* = 1.38) than did low-frequency/low-fan items (*M* = 1.90, *SEM* = 0.36). Week of the experiment did not reliably affect intraexperimental intrusions, nor did it significantly interact with item type, *F*(4, 44) = 1.52, *MSE* = 25.25 and *F*(4, 44) = 1.26, *MSE* = 11.32, respectively, *p*s > .05.

Overall recognition of pseudowords—Data from the five recognition tests were aggregated into three periods (early, middle, and late) to evaluate recognition performance as a function of level of learning. Data from the first two recognition tests, which occurred in Week 1, were combined to form the early period. These early period data were examined to determine whether the pattern of hits and false alarms early in the study replicated the concordant pattern found by Maddox and Estes (1997). Data from the last two recognition tests were combined to form the late period.[5] These late-period data were used to determine whether a stimulus-frequency mirror effect would obtain when participants had a relatively high degree of familiarization with the pseudowords. The middle period refers to the third recognition test, which occurred during Week 2.

Hit rate, false-alarm rate, and *d′* provided three dependent measures to analyze (using ANOVA) as a function of period in study (three levels) and familiarization condition (four levels). Focusing first on *d′*, recognition performance improved from early to late in the study, *F*(2,

---

[5] Even though the last two recognition tests were in Weeks 3 and 5, they were both included in the late period to provide the same number of observations as the early period. Further, the recall data indicate that participants had learned the pseudowords reasonably well by the end of Week 3.

22) = 4.74, *MSE* = 1.61, $p < .05$. The mean *d*′ for the early, middle, and late periods were 0.78, 1.35, and 1.55, respectively. Therefore, as participants gained more experience with the pseudowords, they got better at discriminating between recognition test items that had been presented on the recognition study list and those that had not.

Participants also differed in their ability to discriminate items as a function of familiarization condition. Over the entire experiment, there was a significant effect of condition on *d*′, *F*(3, 33) = 8.73, *MSE* = 0.94, $p < .001$. Low-frequency items were recognized best (*d*′ = 1.78), and high-frequency items were recognized worst (*d*′ = .67). The two medium-frequency conditions were intermediate, with the medium-frequency/low-fan items yielding better recognition (*d*′ = 1.43) than medium-frequency/high-fan items (*d*′ = 1.04). There was no significant interaction of period and familiarization condition on *d*′, *F*(6, 66) = 1.26, *MSE* = 0.54, $p > .05$.

A large source of the difference in *d*′ across conditions was due to the differential proportion of false alarms across conditions. As shown in Table 4, there was a significant effect of familiarization condition on the false-alarm rate, *F*(3, 33) = 22.85, *MSE* = 0.05, $p < .001$. Low-frequency items produced the fewest false alarms, and high-frequency items produced the most false alarms. The medium-frequency conditions were intermediate, with the medium-frequency/low-fan items eliciting fewer false alarms than did the medium-frequency/high-fan items. As noted in Footnote 3, on average the high-fan items would have been seen more recently than would the low-fan items. Therefore, medium-frequency items that are high-fan may have seemed more familiar because they were seen more recently. Of course, more familiar (or higher frequency) real words were also, on average, more likely to have been seen recently. The medium-frequency/high-fan condition should have been seen more recently on average and, on the basis of what we know from studies of the spacing effect, the same number of presentations that are distributed in time will result in stronger memories. Therefore, we predicted that SAC might expect a higher base strength for the medium-frequency/high-fan condition. According to SAC, higher base-level activation would also make a concept more vulnerable to spurious false alarms.

The false-alarm rate decreased across the three periods of the study, *F*(2, 22) = 13.02, *MSE* = 0.02, $p < .05$. Familiarization condition interacted with period in the study, *F*(6, 66) = 3.37, *MSE* = 0.03, $p < .01$, such that the magnitude of the effect of familiarization condition decreased with period in the study, particularly from the early to middle period. These effects are discussed in more detail in the following sections.

With regard to hit rate, there was no reliable main effect of familiarization condition, *F*(3, 33) = 1.07, *MSE* = 0.02, $p > .10$. As shown in Table 4, the overall hit rate varied across the three periods of the experiment, *F*(2, 22) = 5.04, *MSE* = 0.02, $p < .05$. There was a reliable interaction on hit rate of familiarization condition and period of the study, *F*(6, 66) = 2.85, *MSE* = 0.02, $p < .05$. This interaction primarily reflects the finding that early in the experiment there were more hits for higher frequency and higher fan pseudowords, but that advantage disappeared as all pseudowords became more familiar. Late in the experiment there were more hits for low-frequency/low-fan items than for items in each of the other three conditions. These results are discussed in more detail in the following sections.

In summary, as the experiment progressed, participants improved at accurately discriminating studied from nonstudied recognition-test items. Pseudowords in the medium-frequency conditions were recognized at levels intermediate to the low-frequency and high-frequency conditions.

Recognition of pseudowords in the low- and high-frequency conditions—The SAC predictions for low- and high-frequency items differ as a function of amount of familiarization and are therefore analyzed as a function of early acquisition (the first two recognition tests during the first week of the study) versus late exposure (final two recognition tests during the last 3 weeks of the study). Figure 2 shows the pattern of hits and false alarms for items in the low-frequency and high-frequency conditions during these two periods. Data from the early period are displayed in Panel A, and data from the late period are displayed in Panel B. Early in the experiment our data replicate the concordant pattern found by Maddox and Estes (1997), in that high-frequency items produced significantly more hits, $F(1, 66) = 12.42$, $MSE = 0.02$, $p < .001$, and more false alarms, $F(1, 66) = 75.79$, $MSE = 0.03$, $p < .001$, than did low-frequency items.

In contrast, during the late period of the study, a mirror pattern occurred on the basis of stimulus frequency. The late-period interaction of familiarization frequency (low vs. high) and whether the item was presented on the recognition study list (hit vs. false alarm) was reliable, $F(1, 11) = 12.37$, $MSE = 0.02$, $p < .01$. It is important to note that low-frequency items produced more hits than did high-frequency items. This trend was marginally significant, $F(1, 66) = 2.72$, $MSE = 0.02$, $p = .10$.[6] Low-frequency items also produced fewer false alarms than did high-frequency items, $F(1, 66) = 8.89$, $MSE = 0.03$, $p < .01$.

Why did the pattern of data for hits change as the experiment progressed? It seems likely that the low-frequency items in Maddox and Estes's (1997) study and during the early period of our study more closely approximated very-low-frequency words (Rao& Proctor, 1984; Wixted, 1992), whereas the late period of the present study may have been a closer approximation of real-world differences between low- and high-frequency words (e.g., Glanzer & Adams, 1985, 1990). For the stimulus-frequency mirror effect to occur, the low-frequency items cannot be too low in frequency.

We further posit that the difference in hit patterns early and late in the study is due to the differential influence of recollection-based recognition. We propose that recollection-based processes become more influential in recognition as the representations for novel stimuli become unitized. Within the SAC framework, once a node is strong enough to support a higher-level structure, that constituent node is called a *chunk.* Before these pseudowords become chunks, episode nodes cannot be linked to them and can only be associated to the lower-level syllabic or letter chunks that constitute the pseudoword. Early in the present study, particularly for the low-frequency items, participants most likely had weak or incomplete nodes that were limited in their ability to support recollection. The "remember"–"know" data discussed below (see Figure 3) and the recall data are consistent with these ideas. For example, only 19% of the target low-frequency pseudowords were correctly recalled in the first week of the study, but 75% of them were correctly recalled in the last week of the study. These ideas are explored further in the *Discussion.*

According to the principles embodied in SAC, the "remember"–"know" data should exhibit a stimulus-frequency mirror effect, such that low-frequency pseudowords should elicit more "remember" hits and fewer "know" hits than do high-frequency pseudowords. Tables 5 and 6 present the "remember" and "know" data for hits and false alarms as a function of familiarization condition and period in the study (early vs. middle vs. late). When we averaged across all three periods of the study, a clear "remember"–"know" mirror effect was observed. This pattern is shown in Figure 4. Low-frequency pseudowords produced more "remember"

---

[6] Of course, we would have preferred that this marginally significant contrast be reliable, but we attribute the lack of significance to low power because of the small number of participants. If this experiment had not been so difficult to run, in terms of finding participants both willing to devote a large piece of a summer or semester to the task and motivated enough to perform well on a boring task over many weeks, and if it had not been so expensive, in terms of paying participants and a lab assistant to coordinate the efforts, we would have had a larger sample size to increase our power.

hits and fewer "know" hits than did high-frequency pseudowords, $F(1, 33) = 11.10$, $MSE = 0.04$, and $F(1, 33) = 15.23$, $MSE = 0.04$, respectively, ps < .01.

An additional prediction derived from SAC can also be evaluated with the "remember"–"know" data. Given the differential base levels of activation for high- and low-frequency stimuli, SAC predicts more "know" false alarms for higher frequency items. This prediction was supported: There were more "know" false alarms for high-frequency stimuli ($M = .41$) than for low-frequency stimuli ($M = .18$), $F(1, 33) = 9.26$, $MSE = 0.04$, $p < .01$.

Last, to examine whether the role of recollection-based recognition increased from early to late in the study, we looked at the proportion of "remember" hits and "know" hits during the early and late periods of the study. As shown in Figure 3, the proportion of "remember" hits increased and the proportion of "know" hits decreased from the early to late period. This interaction was reliable, $F(1, 11) = 6.40$, $MSE = 0.02$, $p < .05$.[7]

In summary, in addition to finding the stimulus-frequency mirror effect (i.e., more hits and fewer false alarms for low-frequency items) in the late period of this study, we also obtained the "remember"–"know" stimulus-frequency mirror effect for hits, supporting this novel prediction of SAC. Additionally, the data indicate that as participants became more experienced with the pseudowords, the role of recollection-based recognition increased and the role of familiarity-based recognition decreased.

Recognition of pseudowords in the medium-frequency conditions—Within our framework, the finding that the two medium-frequency conditions elicited different false-alarm rates indicates that items in these two conditions differed in their current levels of activation at test. That is, SAC predicts that false alarms should primarily be based on the activation of the concept node. As shown in Table 4, the ordering of these two conditions further indicates that the medium-frequency/high-fan items ($M = .51$) had higher levels of activation at test than their low-fan counterparts ($M = .34$). The "remember" and "know" data also support this idea. There was no reliable difference in the proportion of "remember" false alarms between the medium-frequency/high-fan items ($M = .16$) and the medium-frequency/low-fan items ($M = .12$; $F < 2$). In contrast, medium-frequency/high-fan items elicited significantly more "know" false alarms than did medium-frequency/low-fan items ($M = .37$ vs. $.23$), $F(1, 33) = 9.26$, $MSE = 0.04$, $p < .01$.

Given that the number of familiarization presentations was equal for the two medium-frequency conditions, there are two likely sources of the difference in concept node activation. First, the high contextual fan condition provided more distributed familiarization trials than did the low contextual fan condition, and distributed practice may result in greater strengthening of concept representations. Second, during the recognition test phase of the experiment the high-fan items may have had higher levels of residual concept node activation from the most recent familiarization list presentation than did the low-fan items. Because of the design of the study, high-fan items were presented on every familiarization list per cycle and, hence, were presented on the familiarization list immediately preceding the recognition test phase. In contrast, each low-fan item was presented on only one list per cycle and could have been presented up to six familiarization lists prior to the recognition test phase.[8]

---

[7] In our view, the most likely cause of the early period "remember" hits was sufficient activation at the episode node that was linked to the target item's constituent nodes. Early in our study, when the pseudoword was not yet a chunk, links could be made from the constituent chunks (e.g., syllables of spelling clusters) to an episode node; however, there were likely to be many fewer "remember" judgments from these constituents because the fan was quite high from the constituents. In contrast, during the later period, where there were more "remember" hits, the likely source of sufficient episode node activation came from the higher-level chunk representing the complete target item.

We intended that the so-called contextual fan manipulation would produce differences in the number of contextual associations for items such that items seen on multiple lists per cycle would have more contextual associations than items seen on only one list per cycle. According to SAC, less activation should spread to any given association when there are more competing associations. Thus, items that have higher contextual fan should produce fewer recollection-based hits than should items that have a lower contextual fan. This contextual fan manipulation appeared to have no impact on the proportion of "remember" hits during the early periods of the experiment; however, by the late period in the study, the medium-frequency/low-fan condition gave the suggestion of more "remember" hits ($M = .54$) than did the medium-frequency/high-fan condition ($M = .49$; consult Table 5).

Recall versus recognition stimulus-frequency mirror effect—A third type of mirror effect that is sometimes found with low- and high-frequency real words involves a comparison of recognition and recall performance. In many verbal learning studies, recall performance has been shown to differ as a function of normative word frequency, with participants recalling more high- than low-frequency words (e.g., Balota & Neely, 1980; Gregg, 1976; Mandler, Goodman, & Wilkes-Gibbs, 1982). However, that result is found consistently only with pure lists of a given frequency. When lists contain both low- and high-frequency words, this advantage tends to disappear (e.g., Gillund & Shiffrin, 1984; Gregg, Montgomery, & Castano, 1980) and has on rare occasions been found to reverse (Duncan, 1974; May & Tryk, 1970). Given this unclear pattern of recall advantage for a given word-frequency class, one can ask what pattern obtained with our pseudowords.

Figure 5 presents the proportions of correct recall and recognition for the last week of the experiment, when the frequency manipulation was strongest. We focused on data from the last week, rather than the entire late period, to have a more comparable number of observations (i.e., two recall-test cycles and one recognition test). Had we included the entire late period, there would have been six complete recall-test cycles (i.e., sessions) and only two recognition tests. As it turns out, the descriptive and inferential statistics lead to the same conclusions for Week 5 alone compared with the entire late period. Here, correct recognition is the average proportion of correct responses, where a correct response is responding "old" to a studied item or responding "new" to a lure. A clear recall–recognition mirror pattern occurred as a function of stimulus frequency. Participants recalled more high-frequency target list items than low-frequency items, $F(1, 13) = 15.74$, $MSE = 0.01$, $p < .05$, but correctly recognized more low-frequency items than high-frequency items, $F(1, 11) = 8.57$, $MSE = 0.01$, $p < .05$.

There are three possible reasons for the recall pattern. First, high-frequency items may have yielded better recall because, relative to low-frequency items, they had accrued a higher base level of activation from many more previous exposures, making them more available to output. Second, high-frequency items may have had higher current (i.e., temporary) levels of activation at the time of recall because these items were presented multiple times on the familiarization list whereas low-frequency items were presented only once. Both of these accounts are consistent with SAC. A final possibility derives from the fact that high-frequency items were also high contextual fan items and were presented on every familiarization list. Therefore, participants may have learned to report the high-fan items on every recall list, meaning that the difference in recall for low- and high-frequency items might be an artifact of the experimental design.

---

[8] We also analyzed the recognition data as a function of recency of exposure to the item studied on only a single list. There were more hits and more false alarms when the item had been seen in the same session as compared with when it had been seen in the previous session, but these differences were not statistically reliable.

Data from the two medium-frequency conditions challenge the artifact account. If the recall advantage for high-frequency pseudowords were an artifact of the strategy to recall words that seemed to appear on every list, then one would expect recall to be better for items in the medium-frequency/high-fan condition than for items in the medium-frequency/low-fan condition, because participants could learn to report the former on every recall test. However, as shown in Table 3, participants actually recalled more target items from the medium-frequency/low-fan condition than from the medium-frequency/high-fan condition. This pattern is consistent with the notion that relative to one presentation, multiple presentations of an item on a familiarization list resulted in a higher current level of activation during recall, which led to a higher probability of correctly recalling the target item.

Regardless of whether the recall advantage for high-frequency pseudowords was due to an artifact, this mirror is not consistently observed with real words in mixed-frequency lists. How can SAC explain the occasional failure to find the high-frequency recall advantage in mixed lists of real words? We would argue that on mixed lists, sometimes the von Restorff effect (1933) of unusual items causes the differential allocation of encoding to low-frequency words in a mixed list (analogous to the explanation of Glanzer and Adams, 1990, for greater attention allocated to low-frequency words). That would also explain why on pure lists high-frequency words are recalled better than low-frequency words—differential allocation of attentional resources between different types of items cannot apply. Of course, participants might try harder on a low-frequency word list, but that seems less likely.

Why would one not expect the same von Restorff effect to apply in this study? We believe that all these pseudowords have been seen frequently enough in this experimental context over the course of the semester that nothing is deemed surprising. In this sense our experiment probably was not a close analogue of a study involving real words of differing frequencies.

Discussion

The results of this study demonstrate that with sufficient familiarization, a stimulus-frequency mirror effect will occur for pseudowords. With low levels of familiarization, the data replicated Maddox and Estes's (1997) concordant pattern of fewer hits and fewer false alarms for lower frequency items. However, with higher levels of familiarization, the data showed the typical stimulus-frequency mirror effect: more hits and fewer false alarms for lower frequency items. In addition, the proportion of "remember" hits increased from early to late in the study, indicating that the role of recollection-based recognition of targets increased as participants learned the pseudowords. In all periods of the study, the influence of familiarity was manifested as more "know" false alarms for high-frequency items than for low-frequency items.

Within the two medium-frequency conditions, the recognition data were also consistent with the hit and false-alarm trends. Additionally, the "remember"–"know" data for the two medium-frequency conditions were compatible with the idea that a high degree of contextual fan limits the proportion of activation that spreads to any one node, such as an episode node, thus affecting the likelihood of a recollection-based recognition. A dual-process theory of recognition, such as SAC, is consistent with these findings.

Dual-process theory of recognition—The results of our study appear to provide stronger support for the dual-process theory of recognition described earlier than for a single-process model. The traditional word-frequency mirror effect was obtained for pseudowords and can be explained by this class of models (e.g., Joordens & Hockley, 2000; Reder et al., 2000) without indirect attentional mechanisms. Within the framework of SAC, the hit portion of the stimulus-frequency mirror effect is caused by a higher rate of recollection-based hits to low-frequency items than to high-frequency items and the false-alarm portion is caused by a higher rate of familiarity-based false alarms to high-frequency items than to low-frequency items.

As described by Reder et al. (2000), recollection-based recognition may occur when sufficient activation arrives at the relevant episode node to pass threshold, allowing access to the episode node (see Figure 1). The amount of activation that spreads to the episode node from the concept node will typically be less for a concept with many associations than for a concept with few associations. Because a high-frequency concept has more contextual associations than a low-frequency concept, it is more likely that sufficient activation will spread to the relevant episode node for a low-frequency item. Consequently, the probability of an item eliciting a recollection-based recognition is greater for a low-frequency item than for a high-frequency item. The “remember”–“know” data support this assumption: there were more “remember” hits to low-frequency items than to high-frequency items.

If recollection fails (i.e., either there is not an episode node associated with the probe item or the episode node was not sufficiently activated), a familiarity-based recognition may occur when there is sufficient activation of the probe’s concept node to pass threshold. A concept with a high base-level activation is more likely to pass this threshold and elicit a familiarity-based recognition response than a concept with a low base-level activation. Therefore, as the strength of an item’s memory representation increases, the probability of that item eliciting a familiarity-based recognition increases. This is true both when the item has been studied (hit) and when it has not (false alarm). Accordingly, the probability of an item eliciting a familiarity-based hit or false alarm is greater for a high-frequency item than for a low-frequency item. Again, the “remember”–“know” data support these ideas. There were more “know” hits and more “know” false alarms to high-frequency items than to low-frequency items.

The “remember”–“know” data are also consistent with “remember”–“know” data from studies that used low- and high-frequency real words as stimuli. A “remember”–“know” mirror effect for hits was identified by Reder et al. (2000), who had predicted this effect on the basis of principles embodied in SAC. “Remember” hits show the stimulus-frequency pattern of more hits to low-frequency words than to high-frequency words, whereas the “know” hits show the opposing pattern of more “know” hits to high-frequency words than to low-frequency words (Joordens & Hockley, 2000; Reder et al., 2000). For the formal description and computational implementation of these ideas in SAC, see Reder et al. (2000).

A critical aspect of our results is the difference in hit patterns early and late in the study. As described above, we argue that the late hit pattern was caused by the differential rate of recollection-based recognition for low- and high-frequency items. We further conjectured that the early hit pattern was caused by limited access to episodic representations for items in both the low- and high-frequency conditions. The “remember”–“know” data support this account. From early to late in the study the proportion of “remember” hits increased and the proportion of “know” hits decreased, indicating that the role of recollection-based processes increased as participants learned the pseudowords.

Single-process models of “remember”–“know” judgments and the mirror effect —To understand the claims made in this article, one must recognize that there are two dimensions on which memory models of the mirror effect can be divided. The first dimension divides single- from dual-process accounts of memory. The second dimension divides models on the basis of whether they seek to model individual words or distributions of word classes. Models that are concerned with individual words include single-process models such as SAM (Gillund & Shiffrin, 1984), MINERVA 2 (Hintzman, 1988), ACT–R (Anderson & Lebiere, 1998), and TODAM (Murdock, 1982), as well as dual-process models such as SAC. Each of the models in this group seeks to explain memory behavior in terms of principles governing the storage and retrieval of individual memories, or traces. The other group defined by this second dimension is composed of models in which principles operate directly at the aggregate level of a word class. These include single-process variants of standard signal-detection theories of memory (Donaldson, 1996; Hirshman, 1998; Hirshman & Henzler, 1998; Hirshman & Master, 1997), as well as dual-process theories (Jacoby, 1991; Yonelinas, 1994; Yonelinas, 1999; Yonelinas & Jacoby, 1996). Herein we refer to the two groups of models along this dimension as *trace models* and *aggregate models*, respectively.

Our claim is not that no single-process model can account for our data. Rather, it is that our data are especially challenging for single-process trace models. We recognize that aggregate models of either the single- or the dual-process variety can provide excellent accounts of the data from this experiment. Such models, however, provide no clear principles to dictate the relative levels of familiarity for pseudowords or the placement of decision criteria in each of the four familiarization conditions across the three periods of the study (see Gardiner, Richardson-Klavehn, & Ramponi, 1998, for similar views). Herein lies the heart of our argument. On the basis of principles that operate at the level of individual memories or traces, SAC made qualitative predictions about the ordering of several dependent measures, including hits, false alarms, "remember" hits, and "know" false alarms, across the 12 principal conditions in this study. These predictions were novel and provided an explanatory framework to understand apparently contradictory results in the memory literature (e.g., Glanzer et al., 1993; Maddox & Estes, 1997). Thus, that this constellation of predictions was uniformly confirmed in this study is not a trivial matter.

Single-process aggregate models of memory do, however, offer the possibility of falsification. Because they propose that "remember" responses are generated using a criterion shift on the same distributions underlying "old"–"new" responses, they predict that the levels of d' for "old" and "remember" responses should be the same. Although this prediction was not the focus of this study, it should be noted that it too was confirmed. The d' values for "old" and "remember" responses were not significantly different ($F < 1$). Nevertheless, among the trace models of memory, the results of this experiment provide a clear challenge to single-process accounts.

Similar views—Recently, other researchers have come to similar conclusions about the roles of familiarity and recollection in recognition. First, Chalmers and Humphreys (1998) concluded that two types of memories are involved in recognition: generalized and episode-specific memories. Generalized memories are context-free memories that vary in strength and hence provide information about the familiarity of an item. Episode-specific memories are context specific. Second, like Reder et al. (2000), Joordens and Hockley (2000) argued that a dual-process theory of recognition, with familiarity and recollection as the processes, can account for several types of mirror effects, including the word-frequency mirror effect. They discussed familiarity and recollection as opposing influences on recognition, such that familiarity leads to more "old" responses to higher frequency items and recollection leads to more "old" responses to lower frequency items. Thus, they stated that the mirror pattern occurs when recollection-based recognition dominates over familiarity-based recognition. Consistent with these ideas, Greene (1999) proposed that familiarity-based processing produces a concordant hit–false-alarm pattern, rather than a mirror pattern. He also concluded that familiarity, as defined by frequency of exposure, is not sufficient to account for recognition performance. Taken together, it seems that recollection-based processes are reflected in participants' performance late in our study and in studies that have found the word-frequency mirror effect.

Direct and indirect effects of stimulus frequency—As described earlier, Maddox and Estes (1997) suggested that normative word-frequency effects that typically occur result from both direct and indirect processes. Our findings, as well as the results of previous studies (Mandler, Goodman, & Wilkes-Gibbs, 1982; Rao & Proctor, 1984; Wixted, 1992), suggest that one should not expect to find a stimulus-frequency mirror pattern with very-low-frequency stimuli. The current findings do not entirely rule out the possibility that indirect attentional effects of frequency contribute to the stimulus-frequency mirror effect. Conceivably, toward the end of the experiment participants might have differentially allocated attention to low- and high-frequency pseudowords. However, it is not necessary to posit an indirect attentional mechanism to account for the obtained results, making a direct-effects account of both the early- and late-recognition performance more parsimonious.

Moreover, when variations in attentional allocation to different stimuli are possible, a mirror pattern is not necessarily obtained. Greene (1999, Experiment 1) found a concordant hit and false-alarm pattern for real words as a function of experimental familiarization frequency regardless of whether study-list items were presented for a controlled brief period of time or in a self-paced fashion. Chalmers and Humphreys (1998) used the three-phase paradigm to familiarize participants to novel words from zero to eight times. Even though they did not restrict rehearsal in their study, they failed to find a stimulus-frequency mirror effect. Their result provides further evidence that it is the use of very-low-frequency stimuli, not minimizing indirect effects, that causes the concordant pattern.

Unitization hypothesis: Sufficient concept strength to support higher-level structures—Given that recollection-based recognition processes increased from early to late in an experiment, it is worth commenting why we believe that this occurred. Over the course of the study, participants acquired memory representations (concept nodes) for the pseudowords that were constructed from constituent features such as phonemes and spelling clusters. In the early period, memory representations for the pseudowords were relatively weak nodes that merely reflected the association among the constituent elements of the pseudowords. According to the unitization hypothesis, these nodes were too weak to support an episodic link to them and hence were unable to support recollection-based recognition. Until then, any episodic links that were formed concerning the exposure were associated to the constituent chunks from which the (weak) concept node was constructed.

Each time a concept is encountered, the links from the constituents to the concept node are strengthened and the concept node itself is also strengthened. As the base-level activation of the concept node gets sufficiently high, it is automatically activated when encountered and it affords the ability to build an association from that chunk, such as an episodic trace, rather than building one from some of its constituents. For example, a pseudoword node such as *gret* might be built from constituents that are chunks such as *gr* and *et*. When the constituent chunks are experienced together in the same order often enough, the node that represents their co-occurrence becomes a chunk or “word” in this case. This notion of the requirement of unitization or minimal strength bears similarity to a number of other ideas (e.g., Hayes-Roth, 1977; Servan-Schreiber, 1991). For example, Chalmers and Humphreys (1998) proposed that it may be difficult for individuals to acquire adequate episodic memory for items that are meaningless or novel.

There are several reasons why it should be more difficult to retrieve an episodic trace from these constituents than from a node for the entire pseudoword. One problem with trying to retrieve from a feature is that recognition errors would be higher because the judgment is based on partial matching. Second, the fan off each constituent node is presumably very high, much greater than for even a high-frequency word. The difficulty of retrieving episodic nodes from high-fan items was discussed earlier in this article, shown in Reder et al. (2000), and illustrated in Reder, Donavos, and Erickson (1999, 2001).

Although evolution from concept node to chunk status depends on the amount of exposure to a pseudoword, an increase in the number of exposures to a concept can ultimately also work against recollection. At some point the fan for a frequently presented node could become sufficiently high that it would be difficult to access a specific episode node, as occurs with high-frequency words. In other words, if the concepts are too weak, episodes cannot be linked to them and are linked to constituent features. When concepts become much stronger, episode nodes can be linked directly to these nodes; however, whether the episode nodes can be subsequently accessed to enable a recollection depends on the degree of fan off the concept node.

Conclusion—A theoretical account of stimulus-frequency effects in recognition based only on the direct influence of frequency, per se, appears to be sufficient. SAC, a dual-process theory of recognition in which familiarity and recollection are the two processes, has been shown to account for both word- and pseudoword-frequency effects without positing indirect attentional mechanisms. This does not mean that indirect attentional effects play no role in memory—only that they are not necessary to account for stimulus-frequency effects.

Acknowledgements

This research was supported by National Institute of Mental Health Grants 1R01 MH52808 and 2R01 MH52808. We thank Jason Arndt for comments on a draft of this article.

References

Anderson JR. Retrieval of propositional information from long-term memory. Cognitive Psychology 1974;6:451–474.

Anderson, JR. Language, memory, and thought. Hillsdale, NJ: Erlbaum; 1976.

Anderson, JR.; Lebiere, C. The atomic components of thought. Mahwah, NJ: Erlbaum; 1998.

Ayers MS, Reder LM. A theoretical review of the misinformation effect: Predictions from an activation-based memory model. Psychonomic Bulletin & Review 1998;5:1–21.

Balota DA, Neely HH. Test-expectancy and word-frequency effects in recall and recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition 1980;6:576–587.

Bjork, RA. Retrieval practice and the maintenance of knowledge. In: Gruneberg, MM.; Morris, PE.; Sykes, RN., editors. Practical aspects of memory: Current research and issues: Vol. 1. Memory in everyday life. New York: Wiley; 1988. p. 396-401.

Cary M, Reder LM. A dual-process perspective of the list-strength and list-length mirror effects in recognition. 2001Manuscript submitted for publication, Carnegie Mellon University

Chalmers KA, Humphreys MS. Role of generalized and episode specific memories in the word frequency effect in recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition 1998;24:610–632.

Donaldson W. The role of decision processes in remembering and knowing. Memory & Cognition 1996;24:523–533.

Duncan CP. Retrieval of low-frequency words from mixed lists. Bulletin of the Psychonomic Society 1974;4:137–138.

Gardiner JM. Functional aspects of recollective experience. Memory & Cognition 1988;16:309–313.

Gardiner JM, Java RI. Recollective experience in word and nonword recognition. Memory & Cognition 1990;18:23–30.

Gardiner JM, Parkin AJ. Attention and recollective experience in recognition memory. Memory & Cognition 1990;18:579–583.

Gardiner JM, Richardson-Klavehn A, Ramponi C. Limitations of the signal detection model of the remember–know paradigm: A reply to Hirshman. Consciousness and Cognition 1998;7:285–288. [PubMed: 9690030]

Gillund G, Shiffrin RM. A retrieval model for both recognition and recall. Psychological Review 1984;91:1–67. [PubMed: 6571421]

Glanzer M, Adams JK. The mirror effect in recognition memory. Memory & Cognition 1985;13:8–20.

Glanzer M, Adams JK. The mirror effect in recognition memory: Data and theory. Journal of Experimental Psychology: Learning, Memory, & Cognition 1990;16:5–16.

Glanzer M, Adams JK, Iverson G. Forgetting and the mirror effect in recognition memory: Concentering of underlying distributions. Journal of Experimental Psychology: Learning, Memory, and Cognition 1991;17:81–93.

Glanzer M, Adams JK, Iverson GJ, Kim K. The regularities of recognition memory. Psychological Review 1993;100:546–567. [PubMed: 8356189]

Glenberg AM. Component-levels theory of the effects of spacing of repetitions on recall and recognition. Memory & Cognition 1979;7:95–112.

Gorman AM. Recognition memory for nouns as a function of abstractness and frequency. Journal of Experimental Psychology: Learning, Memory, and Cognition 1961;61:23–29.

Greene RL. Spacing effects in memory: Evidence for a two-process account. Journal of Experimental Psychology: Learning, Memory, and Cognition 1989;15:371–377.

Greene RL. The role of familiarity in recognition. Psychonomic Bulletin & Review 1999;6:309–312. [PubMed: 12199216]

Gregg, V. Word frequency, recognition, and recall. In: Brown, J., editor. Recall and recognition. New York: Wiley; 1976. p. 183-216.

Gregg VH, Montgomery DC, Castano D. Recall of common and uncommon words from pure and mixed lists. Journal of Verbal Learning and Verbal Behavior 1980;19:240–245.

Guttentag RE, Carroll D. Recollection-based recognition: Word frequency effects. Journal of Memory and Language 1997;37:502–516.

Hayes-Roth B. Evolution of cognitive structures and processes. Psychological Review 1977;84:260–278.

Hilford A, Glanzer M, Kim K. Encoding, repetition, and the mirror effect in recognition memory: Symmetry in motion. Memory & Cognition 1997;25:593–605.

Hintzman DL. Judgments of frequency and recognition memory in a multiple-trace memory model. Psychological Review 1988;95:528–551.

Hintzman DL. On explaining the mirror effect. Journal of Experimental Psychology: Learning, Memory, and Cognition 1994;20:201–205.

Hirshman E. Decision processes in recognition memory: Criterion shifts and the list-strength paradigm. Journal of Experimental Psychology: Learning, Memory, and Cognition 1995;21:302–313.

Hirshman E. On the utility of the signal detection model of the remember–know paradigm. Consciousness and Cognition: An International Journal 1998;7:103–107.

Hirshman E, Henzler A. The role of decision processes in conscious recollection. Psychological Science 1998;9:61–65.

Hirshman E, Master S. Modeling the conscious correlates of recognition memory: Reflections on the remember–know paradigm. Memory & Cognition 1997;25:345–351.

Huron C, Danion J, Giacomoni F, Grang D, Robert P, Rizzo L. Impairment of recognition memory with, but not without, conscious recollection in schizophrenia. American Journal of Psychiatry 1995;152:1737–1742. [PubMed: 8526239]

Jacoby LL. A process dissociation framework: Separating automatic from intentional uses of memory. Journal of Memory and Language 1991;30:513–541.

Jacoby LL, Dallas M. On the relationship between autobiographical memory and perceptual learning. Journal of Experimental Psychology: General 1981;110:306–340. [PubMed: 6457080]

Joordens S, Hockley WE. Recollection and familiarity through the looking glass: When old does not mirror new. Journal of Experimental Psychology: Learning, Memory, and Cognition 2000;26:1534–1555.

Kim K, Glanzer M. Speed versus accuracy instructions, study time, and the mirror effect. Journal of Experimental Psychology: Learning, Memory, and Cognition 1993;19:638–652.

Kinsbourne M, George J. The mechanism of the word-frequency effect on recognition memory. Journal of Verbal Learning and Verbal Behavior 1974;13:63–69.

Knowlton BJ, Squire LR. Remembering and knowing: Two different expressions of declarative memory. Journal of Experimental Psychology: Learning, Memory, and Cognition 1995;21:699–710.

Maddox WT, Estes WK. Direct and indirect stimulus-frequency effects in recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition 1997;23:539–559.

Mandler G. Recognizing: The judgment of previous occurrence. Psychological Review 1980;87:252–271.

Mandler G, Goodman GO, Wilkes-Gibbs DL. The word-frequency paradox in recognition. Memory & Cognition 1982;10:33–42.

May RB, Tryk HE. Word sequence, word frequency, and free recall. Canadian Journal of Psychology 1970;24:299–304.

McClelland JL, Chappell M. Familiarity breeds differentiation: A subjective-likelihood approach to the effects of experience in recognition memory. Psychological Review 1998;105:724–760. [PubMed: 9830377]

McCormack PD, Swenson AL. Recognition memory for common and rare words. Journal of Experimental Psychology 1972;95:72–77. [PubMed: 5070296]

Murdock BB. A theory for the storage and retrieval of item and associative information. Psychological Review 1982;89:609–626.

Rajaram S. Remembering and knowing: Two means of access to the personal past. Memory & Cognition 1993;21:89–102.

Rao KV, Proctor RW. Study-phase processing and the word frequency effect in recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition 1984;10:386–394.

Reder LM, Charney DH, Morgan KI. The role of elaborations in learning a skill from an instructional text. Memory & Cognition 1986;14:64–78.

Reder, LM.; Donavos, DK.; Erickson, MA. The role of perceptual similarity in explicit recognition: Evidence against multiple systems; 1999, November; Paper presented at the 40th Annual Meeting of the Psychonomic Society; Los Angeles, CA.

Reder LM, Donavos DK, Erickson MA. Perceptual match effects in direct tests of memory: The role of contextual fan. 2001Manuscript submitted for publication

Reder LM, Nhouyvanisvong A, Schunn CD, Ayers MS, Angstadt P, Hiraki K. 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 2000;26:294–320.

Reder, LM.; Schunn, CD. Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. In: Reder, LM., editor. Implicit memory and metacognition. Hillsdale, NJ: Erlbaum; 1996. p. 45-77.

Schulman AI. Word length and rarity in recognition memory. Psychonomic Science 1967;9:211–212.

Schunn CD, Reder LM, Nhouyvanisvong A, Richards DR, Stroffolino PJ. 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 1997;23:3–29.

Servan-Schreiber D. The competitive chunking theory: Models of perception, learning, and memory. 1991Unpublished doctoral dissertation, Carnegie Mellon University

Strack F, Forster J. Reporting recollective experiences: Direct access to memory systems? Psychological Science 1995;6:352–358.

Tulving E. Memory and consciousness. Canadian Psychologist 1985;26:1–22.

von Restorff H. Über die wirkung von bereichsbildungen im spurenfeld. Psychologische Forschung 1933;18:299–342.

Wixted JT. Subjective memorability and the mirror effect. Journal of Experimental Psychology: Learning, Memory, and Cognition 1992;18:681–690.

Yonelinas AP. Receiver-operating characteristics in recognition memory: Evidence for a dual-process model. Journal of Experimental Psychology: Learning, Memory, and Cognition 1994;20:1341–1354.

Yonelinas AP. The contribution of recollection and familiarity to recognition and source-memory judgments: A formal dual-process model and an analysis of receiver operating characteristics. Journal of Experimental Psychology: Learning, Memory, and Cognition 1999;25:1415–1434.

Yonelinas AP, Jacoby LL. Response bias and the process-dissociation procedure. Journal of Experimental Psychology: General 1996;125:422–434.

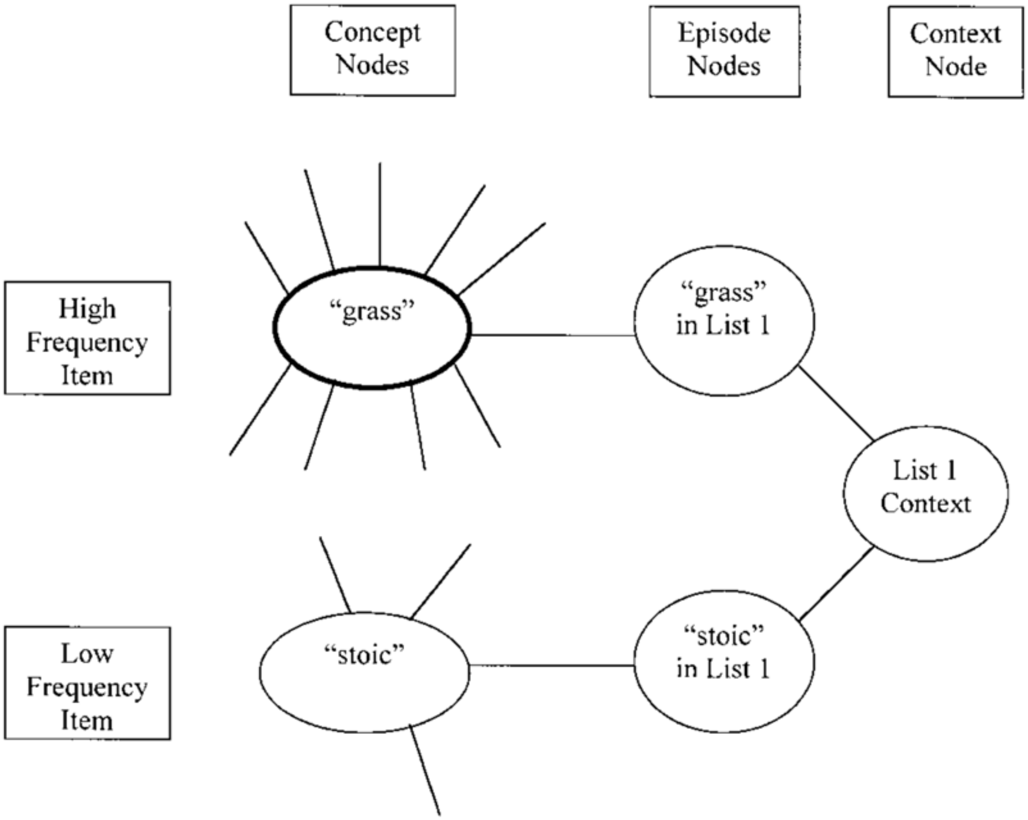

Figure 1.
Schematic illustration of memory representations for high- and low-frequency items. The higher base level of activation for the high-frequency item is denoted with a thicker oval.

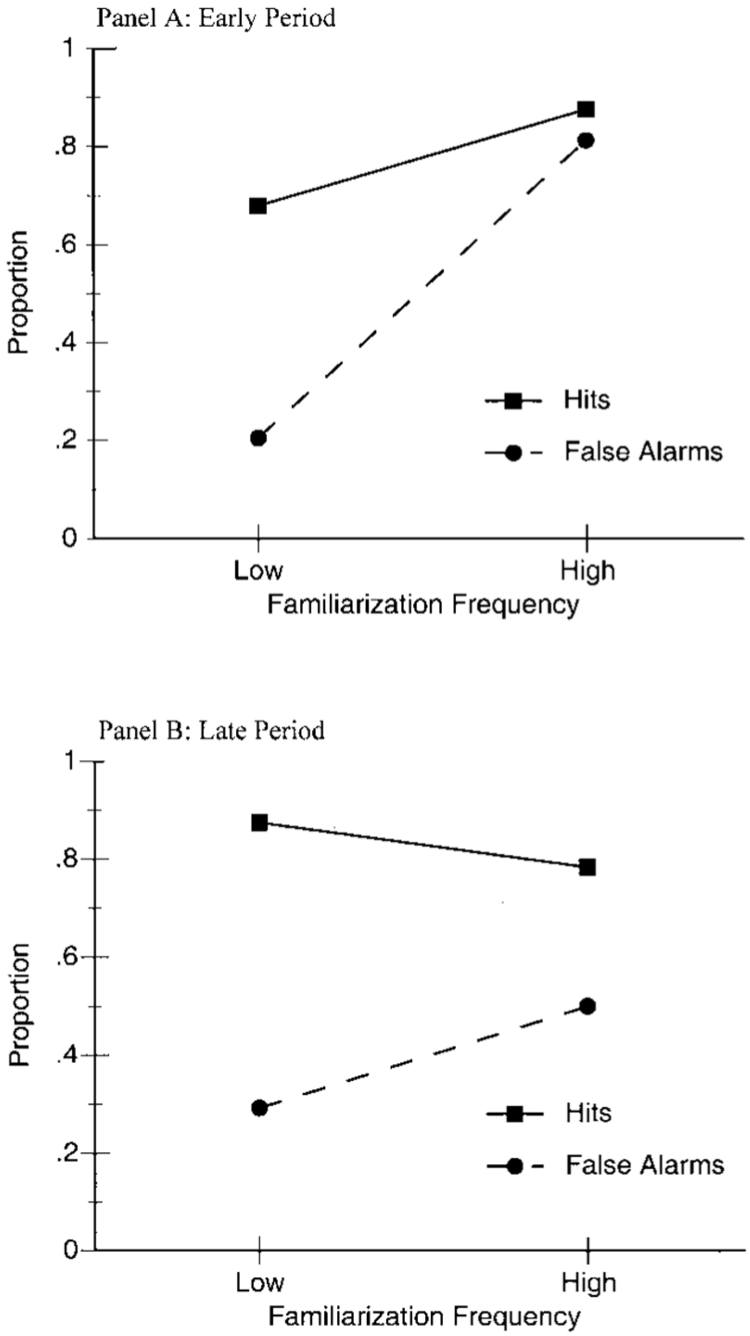

Figure 2.

Proportion of hits and false alarms by low- and high-familiarization frequency for the early (Panel A) and late (Panel B) periods.

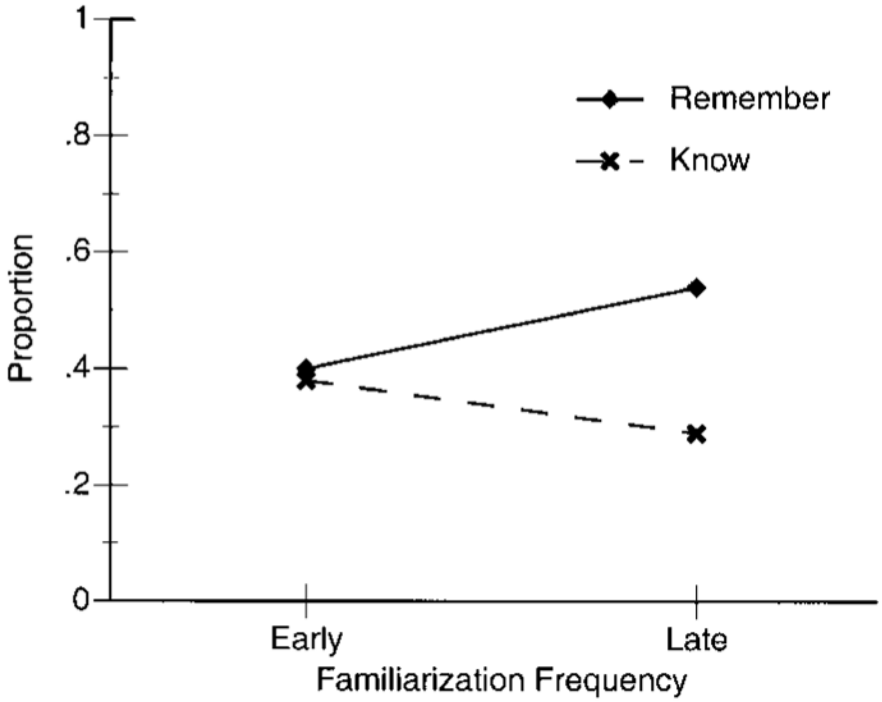

Figure 3.
Proportion of “remember” hits and “know” hits for the early and late periods.

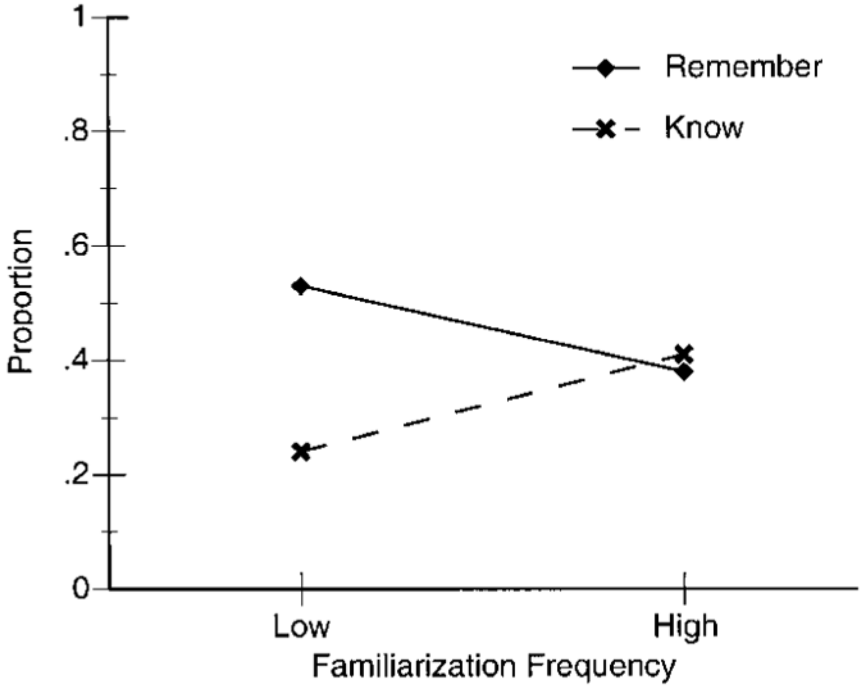

Figure 4.
Proportion of “remember” hits and “know” hits for the low-and high-familiarization frequency conditions.

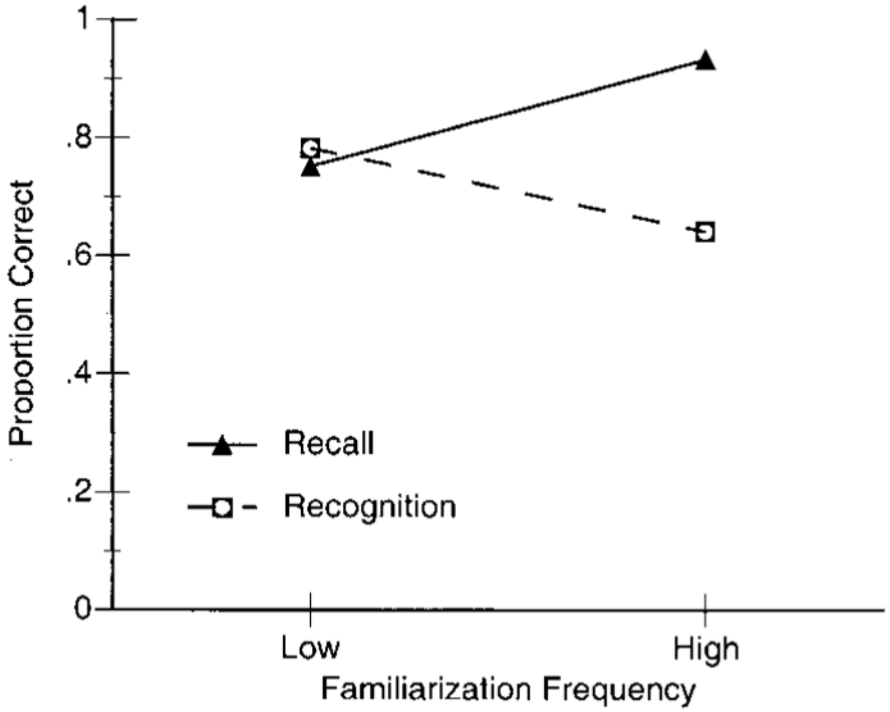

Figure 5.
Proportion of correct recall and recognition of low- and high-frequency items for Week 5.

Table 1

Overview of Experimental Sessions

Week and acquisition cycle	Session
Week 1	
1	1, 2[a]
2	3, 4[a]
Week 2	
3	5, 6
4	7, 8[a]
Week 3	
5	9, 10
6	11, 12[a]
Week 4	
7	13, 14
8	15, 16
Week 5	
9	17, 18
10	19, 20[a]

Note. There were three familiarization lists per session.

[a] Sessions in which recognition study and test lists were presented.

Table 2

Four Familiarization Conditions

		Familiarization		
Frequency	**Contextual fan**	**Lists per cycle**	**Exposures per list**	**Exposures per cycle**
Low	Low	1	1	1
Medium	Low	1	6	6
Medium	High	6	1	6
High	High	6	6	36

Table 3

Mean Proportion Correct Recall for Target Items by Familiarization Condition and Week and Mean Number of Extraexperimental Intrusions by Week

Dependent measure and familiarization condition	**Week**				
	1	**2**	**3**	**4**	**5**
Correct recall					
Low frequency/low fan	.19 (.03)	.61 (.07)	.71 (.07)	.77 (.08)	.75 (.08)
Medium frequency/low fan	.57 (.06)	.85 (.04)	.86 (.04)	.85 (.05)	.92 (.05)
Medium frequency/high fan	.35 (.05)	.64 (.06)	.74 (.07)	.78 (.28)	.80 (.08)
High frequency/high fan	.72 (.03)	.88 (.04)	.91 (.03)	.92 (.04)	.93 (.04)
M	.46 (.04)	.75 (.03)	.80 (.03)	.83 (.03)	.85 (.03)
Extraexperimental intrusions	22.75 (5.55)	12.17 (4.11)	6.67 (1.79)	4.25 (1.14)	4.33 (1.03)

Note. Standard error of the means are shown in parentheses.

Table 4

False Alarms and Hit Rate by Familiarization Condition and Period

Dependent measure and familiarization condition	**Period**			
	Early	**Middle**	**Late**	***M***
False alarms				
Low frequency/low fan	.20 (.02)	.29 (.07)	.29 (.04)	.26 (.03)
Medium frequency/low fan	.40 (.06)	.30 (.07)	.30 (.05)	.34 (.04)
Medium frequency/high fan	.64 (.07)	.46 (.05)	.44 (.07)	.51 (.04)
High frequency/high fan	.81 (.08)	.63 (.07)	.50 (.08)	.65 (.05)
M	.52 (.05)	.42 (.04)	.38 (.03)	
Hit rate				
Low frequency/low fan	.68 (.06)	.74 (.07)	.88 (.04)	.77 (.04)
Medium frequency/low fan	.73 (.05)	.65 (.08)	.81 (.05)	.73 (.04)
Medium frequency/high fan	.80 (.04)	.76 (.07)	.81 (.06)	.79 (.03)
High frequency/high fan	.88 (.06)	.71 (.07)	.78 (.08)	.79 (.04)
M	.77 (.03)	.72 (.04)	.82 (.03)	

Note. Standard error of the means are shown in parentheses.

Table 5

“Remember” False Alarms and Hit Rate by Familiarization Condition and Period

Dependent measure and familiarization condition	**Period**			
	Early	**Middle**	**Late**	***M***
“Remember” false alarms				
Low frequency/low fan	.08 (.02)	.09 (.04)	.05 (.01)	.07 (.01)
Medium frequency/low fan	.14 (.02)	.08 (.03)	.11 (.02)	.12 (.02)
Medium frequency/high fan	.21 (.05)	.11 (.04)	.13 (.03)	.16 (.04)
High frequency/high fan	.33 (.09)	.23 (.06)	.15 (.05)	.24 (.06)
M	.19 (.04)	.13 (.03)	.11 (.03)	
“Remember” hit rate				
Low frequency/low fan	.43 (.05)	.53 (.08)	.63 (.07)	.53 (.06)
Medium frequency/low fan	.32 (.04)	.38 (.08)	.54 (.07)	.42 (.05)
Medium frequency/high fan	.31 (.06)	.39 (.07)	.49 (.06)	.40 (.05)
High frequency/high fan	.37 (.09)	.32 (.07)	.45 (.07)	.39 (.07)
M	.36 (.04)	.41 (.07)	.53 (.06)	

Note. Standard error of the means are shown in parentheses.

Table 6

"Know" False Alarms and Hit Rate by Familiarization Condition and Period

Dependent measure and familiarization condition	**Period**			***M***
	Early	**Middle**	**Late**	
"Know" false alarms				
Low frequency/low fan	.12 (.03)	.20 (.05)	.24 (.04)	.18 (.02)
Medium frequency/low fan	.27 (.06)	.22 (.07)	.20 (.04)	.23 (.03)
Medium frequency/high fan	.43 (.07)	.35 (.06)	.31 (.06)	.37 (.06)
High frequency/high fan	.49 (.09)	.40 (.08)	.35 (.07)	.41 (.06)
M	.33 (.05)	.29 (.05)	.27 (.03)	
"Know" hit rate				
Low frequency/low fan	.25 (.04)	.21 (.05)	.25 (.05)	.24 (.03)
Medium frequency/low fan	.41 (.05)	.27 (.07)	.27 (.05)	.33 (.04)
Medium frequency/high fan	.49 (.05)	.37 (.06)	.32 (.06)	.40 (.05)
High frequency/high fan	.50 (.09)	.39 (.08)	.33 (.07)	.41 (.07)
M	.41 (.04)	.31 (.04)	.29 (.05)	

Note. Standard error of measurement values are shown in parentheses.