The false memory and the mirror effects: The role of familiarity and backward association in creating false recollections

David Anaki, Yifat Faran, Dorit Ben-Shalom, Avishai Henik

Abstract

The mirror effect refers to a phenomenon where the hit rate is higher for low frequency words while the false alarm rate is higher for high frequency distractors. Using a false memory paradigm (Roediger & McDermott, 1995), we examined whether false memory for non-presented lures would be influenced by the lure’s familiarity. The results revealed that false memory levels for low familiarity lures were higher than that for high familiarity lures, but only when the backward association strength between the presented list’s words and the lure was high. The veridical memory for the presented words also revealed greater accuracy for low familiarity words. In contrast, higher false alarms were observed for high frequency unrelated distractors. These results are discussed in light of current theories of the false memory effect, and it is suggested that they support an activation/monitoring account of the effect, according to which non-presented lures are activated during encoding.

Keywords: False memory; Familiarity; Backward association; Mirror effect
word was not presented in the list, it was falsely recalled at very high levels, reaching intrusion rates of 44% for some lures.

In spite of the robust false recall elicited by the Deese (1959) paradigm, subsequent memory research generally ignored his findings (see Bruce & Winograd, 1998, for an analysis of the topic from philosophical and sociological perspectives). The paradigm was revived by Roediger & McDermott (1995), who replicated the original study with a new set of stimuli. In addition, they modified the paradigm by introducing a recognition test following the lists’ presentation. Their results also showed a high occurrence of false memory. Specifically, the non-presented associates were recalled 40–55% of the time, a rate similar to presented words appearing in the middle of the list. In the recognition test, the results were more dramatic: The false alarm rate for the critical lures was almost equivalent to the hit rate for the words appearing in the lists.

Theoretical interpretations of false memory

Two basic accounts of false memory have emerged in the literature: The first, which we refer to as a memory-based approach (following Hirshman & Arndt, 1997), emphasizes encoding and representational factors as major determinants of the emergence of false memories. The second account, which we refer to as a decision-based account, claims that decision processes could influence the production of memories that did not occur. In the following section, we will elaborate on the former approach, while postponing the discussion on the latter until the General discussion.

Within the broad category of memory-based theories, diverging accounts of false memory effects can be pinpointed. One line of theories emphasizes the role of associative processes in creating false memories (e.g., Roediger, Balota, & Watson, 2001; Robinson & Roediger, 1998; Roediger, Watson, McDermott, & Gallo, 2001). According to this approach, the presentation of a word during the encoding phase activates its associates, including the critical lure, in the semantic memory network. The cumulative activation of the lure by the multiple words in the list causes participants, in a subsequent memory test, to experience source confusion and to misidentify the lure as a presented item (Johnson, Hashtroudi, & Lindsay, 1993). The activation of the lure during the encoding stages could be either conscious or unconscious. Underwood (1965), for example, argued that participants consciously generate the associated lure during the study phase. Others, however, have demonstrated that false memories can arise even when the list is presented at a rapid pace, preventing the conscious processing of both the words and the lure (Seamon, Luo, & Gallo, 1998; see also Thapar & McDermott, 2001 who showed false recall and recognition for lures from lists that were processed at a shallow level).

A second line of theories focuses on faulty encoding of the presented words as a major determinant in the production of false memory (Dodson & Schacter, 2001; Koutstaal & Schacter, 1997). According to fuzzy trace theory (Brainerd, Reyna, & Kneer, 1995; Reyna & Brainerd, 1995) memory is based on either a “ gist” trace, which preserves the meaning and interpretations of the experience, or a “verbatim” trace, which preserves the specific features of the experience. False memories are attributed to remembering the gist but not the verbatim representation of the presented words, while veridical memories are driven by an item-specific verbatim trace. This reliance on gist traces may result from a pattern separation failure (McClelland, McNaughton, & O’Reilly, 1995), which is the inability to form non-overlapping unique representations of the items in the list. As a result, participants have good memory for the common aspects of the words, but poor memory for the items themselves. Thus, the lures will be identified as part of the presented list because of its common attributes (see also Hunt & McDaniel, 1993 for a similar account).

The two approaches differ on the question of whether false memory for the non-presented lure is actually a genuine recollection of events occurring during study, or an illusory experience constructed during retrieval. According to the associative process account, the critical lure is activated (consciously or unconsciously) during the encoding stage and becomes an entity that is experienced during encoding and related to other events taking place at that time. In contrast, according to fuzzy trace theory and similar accounts, the failure to identify the critical lure as a non-presented item is not related to the activation status of the lure but to the encoding of the presented items. Thus, the activation of the lure is not a necessary condition for false memory.

Support for the claim positing that activation processes underlie the creation of false memories is provided by participants’ phenomenological experiences of these non-studied lures. Using the remember-know judgement paradigm (Tulving, 1985) numerous studies have found that participants can report re-experiencing the lures’ presentation during study at levels equal to those of items that were actually studied (e.g., Gallo, McDermott, Percer, & Roediger, 2001; Israel & Schacter, 1997; Roediger & McDermott, 1995). The fact that participants claim conscious recollection of the lure suggests that it was indeed activated during study, thus sharing similar features with the list’s items (but see Brainerd, Wright, Reyna, & Mojardin, 2001 for an alternative account from a fuzzy-trace theory perspective). However, one drawback of this line of reasoning is that conclusions are based on subjective evaluation of previous memory judgements and not on more objective data.
(Brainerd et al., 2001). Extracting evidence directly from the data may be a more reliable technique for observing whether memory for non-presented lures resembles veridical memory. One possible approach would be to investigate how the lure’s frequency influences false memory, since past research has found diverging effects of frequency on memory for presented and non-presented items.

The mirror effect in recognition memory

A common practice in memory research is to manipulate word frequency to influence memory. Studies have found that in a typical old–new recognition paradigm, the hit rate is higher for low frequency than high frequency targets. The false alarm rate, however, is higher for high frequency than for low frequency distractors (Chalmers & Humphreys, 1998; Glanzer, Adams, Iverson, & Kim, 1993; Glanzer, Kim, & Adams, 1998). This phenomenon is known as the word frequency effect or as the mirror effect because of the opposite influence of word frequency on hits and false alarms.

Several theoretical accounts have been offered to accommodate these findings. As in the field of false memory, they can be categorized as either decision-based or memory-based theories. Decision-based theories emphasize the role of decision processes in the emergence of the mirror effect (e.g., Glanzer et al., 1993; Greene, 1996; Hirshman, 1995). Greene, for example, suggests that the mirror effect arises from participants’ inclinations to equate the pattern of their responses to the different types of stimuli. Since old low frequency words generate more yes responses than high frequency words in a yes–no recognition test, due to their being more memorable, fewer yes responses will be produced for new low frequency than for new high frequency words.

Memory-based accounts, on the other hand, emphasize the unique encoding or representation of low frequency compared to high frequency words. The Source of Activation (SAC) model is one such account that was recently proposed to explain frequency effects in recognition memory (Reder et al., 2000). According to the SAC, when a word is presented at study it activates its concept node and event node, the latter containing contextual information relating to the specific encoding episode. During retrieval, activation spreads from the concept node to its associated event nodes. Concept nodes of high frequency words are more activated, since they are more frequently and recently seen, but also more event nodes are attached to them, so each of their many event nodes receives little activation from the concept node. In contrast, concept nodes of low frequency words are less activated, but fewer event nodes are attached to them so each event node receives greater activation. The mirror effect stems from the reliance of participants on the activation of the event nodes, which leads to greater accuracy for low frequency words. In contrast, reliance on the activation of the concept nodes of new words will result in greater false alarms for high frequency words, which are characterized by higher activation of the concept node.

Despite the large differences between decision-based and memory-based accounts of the mirror effect, they share a common notion that old and new words vary in their probability distributions across a memory strength dimension that represents their recent occurrence. Thus, the mirror effect can serve as a useful tool in comparing memory for presented and non-presented items (related as well as unrelated) and in delineating the processes underlying false memory. If indeed the critical lure is activated during the study phase, similar to the activation of a presented list word, then a word frequency effect should emerge. Greater old responses should be found for low frequency lures than for high frequency lures. However, if false memory arises as a result of pattern separation failure, originating at the encoding stage but manifesting itself during the retrieval stage when participant base their judgments on gist traces, there should be no difference in falsely remembering low versus high familiar lures. The predictions of this approach would be a lack of lure’s familiarity on false memory. Alternatively, high frequency lures would be falsely recognized more often than low frequency lures, just as more false alarms are found in recognition performance for new high frequency words.

The effect of the familiarity of the lure on memory was first investigated by Deese (1959) but has not been pursued since then. The reason might be that Deese did not find a significant correlation between the lure’s frequency and the magnitude of its false recall (r = .06). However, two methodological drawbacks in his study cannot allow us to discount potential frequency effects on false memory: First, the word frequency counts of the lures (taken from Thorndike & Lorge, 1944) were quite low. Thus, the lack of variability in frequency might have obscured any potential effect, as Deese himself admitted (p. 20). Second, recall was the only measure assessing memory performance in his study. Thus, the question of whether the lure’s frequency can influence false recognition is yet to be answered. The reason to examine whether the lure’s frequency influences false recognition stems also from findings that frequency effects in recall are influenced by factors that are not crucial in recognition, such as list structure (i.e., pure versus mixed lists that contain both high and low frequency words; Ward, Woodward, Stevens, & Stinson, 2003). The primary aim of the present study was, therefore, to use the mirror effect as a tool to investigate the
processes underlying false memory for non-presented lures. To this end, we manipulated the frequency scores of non-presented lures and probed their false memory with both recognition and recall.

Studies in which the mirror effect was found used word frequency as the independent variable (e.g., Glanzer & Adams, 1990; Greene, 1996; Kim & Glanzer, 1993). The frequency rating of a word is usually based on word frequency norms such as those of Kucera & Francis (1967) or Thorndike & Lorge (1944), which count the appearance of words in a large corpus sample of published literature. These norms are themselves an estimate of the number of times that a word appears in natural language. However, this reliance on word frequency norms is problematic for a number of reasons: First, occurrence counts of low frequency words may be unreliable since the majority of norms are based on published literature which is revised by editors inclined to use stylized low frequency words (Rudell, 1993). Second, since the norms themselves are samples, they may be subject to sampling errors. Some researchers have claimed that the size of the corpus may be a major determinant in the effective sampling of word use (Breland, 1996; Burgess & Livesay, 1995). This problem is especially acute in the existing Hebrew norms (Balgur, 1996), which are based on a corpus of only 160,000 words sampled from published literature designed for elementary school pupils.

In order to circumvent the potential problems that may arise with the use of normative frequency measures, we used a measure of experiential familiarity (Gernsbacher, 1984) in the present research, in addition to printed word frequency. This is a subjective rating in which participants were asked to rate how familiar they were with each word. The ability of this measure to approximate the frequency of occurrence of a word in natural language was found by Gernsbacher to reliably predict word recognition. In addition, the use of experiential familiarity measures instead of printed frequency measures has proved to be helpful in resolving inconsistencies in the word recognition literature where printed word frequency was manipulated along with a second variable. Furthermore, Gernsbacher found the converging validity of this measure to be quite high, yielding a correlation of .81 with printed frequency (see also Balota, Pilotti, & Cortese, 2001 for similar results). Recent studies by Chalmers and colleagues (Chalmers & Humphreys, 1998; Chalmers et al., 1997) have also shown that the mirror effect can be obtained with measures other than normative word frequency. Finally, we also performed a pilot study, which will be described more elaborately in the Method section of Experiment 1, to demonstrate that the mirror effect is also obtained when experiential familiarity is used to define high versus low frequency items.

**Backward association and false memory**

Backward association, which is the association strength between the items in the presented list and the critical lure, may be a relevant variable when examining the mirror effect in false memory. Deese (1959) was the first to hypothesize that the stronger the backward association, the more often participants would falsely recall the lure. Indeed, a strong correlation ($r = .87$) was found between the mean backward association and the likelihood of false recall of the lure. Roediger et al. (2001) report a correlation of .78 between the strength of the backward association and the probability of false memory of the lure. This factor accounted for 40% of the variance, exceeding by far any other factor. McEvoy, Nelson, & Komatsu (1999) have also addressed the influence of backward association strength on false recall and false recognition. They manipulated backward association as a between subjects variable and found more false alarms for non-presented lures, both in free recall and recognition tests, when the backward association was stronger.

The susceptibility of false memory to backward association strength can be easily interpreted by the first theoretical account of false memory described earlier. According to the first approach, which focuses on associative processes and lure activation, backward association strength determines the extent of lure activation (Roediger et al., 2001). According to the second approach, which emphasizes the pattern separation failure and the reliance on gist traces, as the list words are more related to the lure, the gist traces formed during encoding are more similar to the lure, leading to more false memories (McEvoy et al., 1999).

If frequency effects emerge for critical lures as well as for presented list items, the backward association strength of a list may modulate the effect. This claim stems from the fact that stronger backward association would lead to a greater activation of the lure during the encoding stage of the experiment and to the attribution of the critical lure to the presented list. Thus, these effects could be more pronounced for lures with strong backward association lists than for lures with weak backward association lists. Alternatively, the effects could appear only with the former type of lures.

**Experiment 1**

In Experiment 1, both the backward association strength between the lists’ words and the non-presented lures, and the familiarity of the lures, were manipulated orthogonally. As in previous studies (e.g., McEvoy et al., 1999), we expected to find more false memories for lures with strong backward association strength. However, our main interests were in the effect of familiarity on lure
recognition, and the interaction between familiarity and backward association. According to an associative activation account of false memory, familiarity effects are expected to emerge and low familiarity lures will be more falsely recognized than high familiarity lures. This effect may be further modulated by backward association strength. In contrast, fuzzy-trace theory and similar accounts would predict that familiarity effects would be absent in false recognition of lures, or that the pattern of these errors would be similar to distractors, namely, more false alarms to high than low familiarity words. Concerning recall, we did not have specific predictions since past research has not found consistent frequency effects, especially when mixed lists were used.

Method

Participants
Thirty-eight participants from the Ben-Gurion University participant pool took part in the experiment for course credit. All were native Hebrew speakers and had normal or corrected to normal vision and hearing.

Materials
Lists were constructed from the Rubenstein, Anaki, & Henik (2003) word association norms (see Anaki & Henik, 2003, for a description of the norming procedure). For each prime word, we selected the 12 most frequent associates, which were not ambiguous in meaning. Then, we computed the backward association strength between each associate and its prime word. We first selected the associates that appeared as prime words in the Rubenstein et al. norms. Words that did not appear in the norms were tested separately in a new word association norming procedure, which was similar to the procedure for the original word association norms. Fifty participants took part in this norming procedure. Each associate was checked to see whether the prime word was given as a response in the word association procedure. Then, the proportion of participants that gave this response was calculated. Based on these proportions, a mean backward association was calculated for each list, which was the sum of the backward association proportion of the 12 associates to the prime, divided by the number of associates.

 Twenty lists were selected following this procedure. These lists were divided into two groups consisting of 10 lists each: One group had a low mean backward association and the other group had a high mean backward association. Half of the prime words in each group were words rated as low familiarity words and the other half were rated as high familiarity words. Familiarity ratings were collected in our laboratory in a way similar to the procedure described by Gernsbacher (1984). One hundred participants rated the words on a scale from 1 (very unfamiliar) to 7 (very familiar). Altogether there were five lists in each of the four conditions: weak backward association/low familiarity, weak backward association/high familiarity, strong backward association/low familiarity and strong backward association/high familiarity (see Appendix for the list of critical lures). The two strong backward association conditions differed from the two weak backward association conditions in their mean backward association score, $F(1,16) = 44.55$, $MSE = .000$, $p < .001$ (.16 vs. .03, respectively). A significant difference in familiarity ratings was found between the two high familiarity conditions and the two low familiarity conditions, $F(1,16) = 199.55$, $MSE = .196$, $p < .001$ (6.25 vs. 3.48, respectively). A similar difference was obtained when normed frequency scores (Balgur, 1968) were used, $F(1,16) = 4.56$, $MSE = 4913$, $p < .05$. Mean backward association and familiarity scores for each condition are presented in Table 1. The critical lures in the different conditions did not differ in their concreteness ratings, $F(1,16) < 1$, which were collected from 100 participants who rated the words on a scale from 1 (low concreteness) to 7 (high concreteness) following the instructions of Spreen & Schultz (1966). Finally, the lists did not differ in their mean forward association, $F(1,16) < 1$.

 In addition, we measured the familiarity and frequency ratings for the 240 list words (20 lists $\times$ 12 words in a list). Words from lists with high familiarity lures were more familiar than words from lists with low familiarity lures, $F(1,16) = 33.90$, $MSE = .02$, $p < .0001$ (6.23 vs. 5.47, respectively). Frequency ratings obtained for

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Mean lure familiarity and backward association strength scores for the different conditions in Experiment 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backward association</td>
<td>Strong</td>
</tr>
<tr>
<td>Weak</td>
<td>Backward association</td>
</tr>
<tr>
<td></td>
<td>M</td>
</tr>
<tr>
<td>Lure familiarity</td>
<td>3.32</td>
</tr>
<tr>
<td></td>
<td>6.12</td>
</tr>
<tr>
<td>Low</td>
<td>3.64</td>
</tr>
<tr>
<td>High</td>
<td>6.38</td>
</tr>
</tbody>
</table>
these words also corroborated the finding that words from lists with high familiarity lures were characterized by high frequency ratings compared to words comprising the lists of low familiarity lures, $F(1,16) = 6.68, MSE = 4539.90,$ $p < .05$.

For the recognition test, 40 distractors were selected. Half of them were high familiarity words (5.62) and the other half were low familiarity words (3.25), $n(38) = 10.22, SE = .23, p < .001$. They differed also in frequency $n(38) = 2.00, SE = 15.40, p = .05$, but not in concreteness.

Prior to conducting Experiment 1, we investigated whether a mirror effect can be obtained using subjective familiarity ratings of frequency. A total of 96 Hebrew nouns constituted the critical stimuli, 48 of which were high familiarity words ($M = 6.68$) and 48 low familiarity words ($M = 4.13$), $n(94) = 40.71, SE = .06, p < .001$. The high and low familiarity words differed also in their frequency norm ratings, $n(94) = 2.80, SE = 20.94, p < .01$, but not in their concreteness ratings (mean ratings of 6.00 and 5.98, respectively) and length (4.40 and 4.38 letters, respectively).

The 96 critical stimuli were divided into two lists (A and B) each containing 24 high familiarity and 24 low familiarity words. Half of the participants heard list A during the study phase, and received the words in list B in the recognition test as distractors. The other half heard list B in the study phase while list A words were presented as distractors in the recognition test.

Twenty-two undergraduate students at Ben-Gurion University of the Negev participated in the experiment for course credit. All participants were native Hebrew speakers. They were tested in groups, with each group consisting of 11 participants. They were informed that they would hear a list of words and a recognition test would be administered after completion of the list. The list was read by the experimenter at a rate of approximately 1 word every 2 s. A break of 3 min followed the presentation, during which the participants solved math problems. Following the break, a sheet containing the 96 critical stimuli was given to the participants, who were asked to indicate whether each word was “old” or “new.” The order of the words was randomized for every sheet.

Separate $t$ tests were performed on participants' ($t_1$) and items' ($t_2$) hit and false alarm proportions. The results clearly demonstrated the mirror effect: Low familiarity words were recognized more accurately (.76) than high familiarity words (.64), $t_{(21)} = 3.71, SE = .03, p < .001$, $t_{(21)} = 3.29, SE = .04, p < .001$. In contrast, the false alarm rate was higher for high familiarity words (.15) than for low familiar words (.11), $t_{(21)} = 2.87, SE = .01, p < .01$, $t_{(21)} = 1.94, SE = .02, p = .05$. Thus, it can be concluded that the mirror effect emerges even when subjective familiarity, rather than word frequency, is used as the independent variable.

**Procedure**

Each participant was tested individually in a quiet room and heard 20 lists presented on an audiotape by a female voice at a rate of approximately 1 word every 2 s. After each list, participants were asked to write down as many words as they could remember from the list. They were given 1 min to do so. The participants were asked not to guess and to write down the responses to each list on a different page in a notebook they were given from the experimenter. They were also asked not to add words to previous lists.

The 20 lists were randomly ordered so each type of list had the same chance of appearing at the beginning, middle or end of the study list. The resulting set of lists was further ordered in two different ways: forward (from list number 1 to list number 20), and backward (from list number 20 to list number 1), ensuring that on average, each of the 20 individual lists appeared in the middle of a set. Participants were randomly assigned to one of the two list orders.

After all 20 lists were presented and the participants performed a free recall test on each list, a recognition test was given. Stimuli were displayed on an Olivetti color monitor controlled by Micro Experimental Laboratory (MEL2) software (Schneider, 1988), implemented on an Olivetti M290-30 PC compatible computer. Twenty blocks of 6 words each were presented. Each trial began with the appearance of a word for which the participant had to decide, as accurately and quickly as possible, whether or not it had previously appeared in one of the lists. Words from each list appeared in the same list order as they were presented to the participant in the first stage of the experiment. Each recognition list contained three words from the original list: the 1st word, the 6th word and the 12th word of the list. In addition, there were two distractors that did not appear previously in any list and were not given as associates to any word on any list. The order of these words in each recognition list was randomized for each participant. The lure appeared at the end of each list (Roediger & McDermott, 1995; Experiment 1). Each word remained on the computer screen until the participant responded. The participants were asked to indicate their decision by pressing one of the two buttons: a red button (z key) for a word that did not appear in the any of the previous lists and a green button (/key) for words that did appear in a previous list. They were instructed to use the index fingers of both hands.

**Results and discussion**

The dependent variable was the proportion of words remembered (whether falsely or correctly). The results of the recognition test appear before the results of the free
recall test, although chronologically the free recall test preceded the recognition test. The order of this presentation was chosen because our main hypotheses, concerning the mirror effect and the effects of lure familiarity on false memory, involve the findings in the recognition test. Within each section the results for false memory are presented first, followed by the results for veridical memory. Analysis of variance (ANOVA) was conducted on the dependent variable with familiarity of the lure (high/low familiarity) and mean backward association (strong/weak backward association) as independent variables. Both independent variables were manipulated within subjects. Means of all critical conditions are presented in Table 2.

### Recognition

Both the familiarity of the lure and the backward association strength of the list affected the false recognition of the lures. Participants tended to falsely recognize the lures more often when the words in the list were strongly associated with the lure (.70) than when the backward association was weak (.56), \( F(1,37) = 16.82, MSE = .04, p < .0005 \). In addition, low familiarity lures were falsely remembered to a greater extent than high familiarity lures (.67 and .59, respectively; \( F(1,37) = 5.08, MSE = .04, p < .05 \)). Importantly, the influence of lures’ familiarity was more pronounced in the strong than in the weak backward association lists, as indicated by a significant backward association \( \times \) familiarity interaction, \( F(1,37) = 4.72, MSE = .03, p < .05 \). Whereas in the weak backward association condition false recognition for low and high familiarity lures was similar (.57 and .56 respectively, \( F < 1 \)), in the strong backward association condition low familiarity lures were falsely recognized more often (.76) than high familiarity lures (.63), \( F(1,37) = 7.19, MSE = .05, p < .01 \). That is, a familiarity effect, involving greater false recognition for low familiarity than high familiarity lures, was observed only in lists with strong backward association.

Analysis of the distractors in the recognition test revealed a different pattern of performance. Comparing the 20 high familiarity fillers with the 20 low familiarity fillers showed a higher false alarm rate (.07 vs. .04) for the high familiarity distractors than for the low familiarity distractors, \( t(38) = 2.21, SE = .01, p < .05 \).

Familiarity effects were also found in the recognition task for targets. Accuracy was higher for words in lists with high familiarity lures than for those from lists with low familiarity lures (.80 and .71, respectively; \( F(1,37) = 23.54, MSE = .01, p < .0001 \)). As mentioned earlier, the words in lists with high familiarity lures were also highly familiar and vice versa. Thus, the familiarity effect obtained here in veridical recognition resembles the frequency effect obtained in past studies, namely, higher hit rate for low frequency than for high frequency items. Moreover, the opposing influence of familiarity on false alarms to distractors corresponds to the mirror effect reported in the literature (e.g., Chalmers & Humphreys, 1998; Glanzer & Adams, 1990; Glanzer et al., 1998) and extends the phenomenon of the mirror effect to a new measure of word frequency, namely, its subjective experience.

The influence of lure’s familiarity on false recognition, in a pattern similar to old words, rather than to non-presented distractor words, is a new observation. Low familiarity lures were falsely recognized to a greater extent than high familiarity lures, just as low frequency list items were more accurately recognized than high frequency words presented in the lists. Spreading activation theories can explain this result as stemming from processes that are similar to those occurring with the recollection of low and high familiarity old words. Since both old words and critical lures are activated during the encoding phase, the memory strength of low familiar lures will be stronger than high familiar words, thus leading to greater false recollection of low familiarity words.

Fuzzy-trace theories, on the other hand, would encounter greater difficulty in accounting for the greater false recognition of the low familiarity lures. According to these theories, the underlying factor in the false memory of lures is the gist of the words presented in the list.

### Table 2

<table>
<thead>
<tr>
<th></th>
<th>Weak backward association</th>
<th></th>
<th>Strong backward association</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low familiarity</td>
<td>High familiarity</td>
<td>Low familiarity</td>
<td>High familiarity</td>
</tr>
<tr>
<td></td>
<td>( M )</td>
<td>( SE )</td>
<td>( M )</td>
<td>( SE )</td>
</tr>
<tr>
<td>Recognition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>List words</td>
<td>.84</td>
<td>.02</td>
<td>.75</td>
<td>.03</td>
</tr>
<tr>
<td>Lures</td>
<td>.57</td>
<td>.04</td>
<td>.56</td>
<td>.04</td>
</tr>
<tr>
<td>Free recall</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>List words</td>
<td>.69</td>
<td>.01</td>
<td>.67</td>
<td>.01</td>
</tr>
<tr>
<td>Lures</td>
<td>.15</td>
<td>.03</td>
<td>.12</td>
<td>.02</td>
</tr>
</tbody>
</table>
(Brainerd & Reyna, 2002; Brainerd et al., 2001). When stronger gist traces are formed and retrieval is based on them, a higher rate of false recognition is expected. However, factors that enhance gist formation, such as backward association, are usually attributed to list words’ characteristics and not to the non-presented lure. Thus, it is not clear how the familiarity of the lure can influence its false recognition.

The two-way ANOVA performed on the hit rate for the lists’ words yielded also a significant main effect for backward association, $F(1,37) = 14.11$, $MSE = .02$, $p < .001$. Participants recognized more words that appeared in lists with weak backward associations than words from lists with strong backward associations (.79 and .72, respectively). The greater recognition of low backward association lists is an unexpected finding because this factor is not assumed to influence recognition of the words that appeared in the lists. In addition, it stands, prima facie, in contrast to McEvoy et al. (1999, Experiment 1), who obtained greater recognition for strong rather than weak backward association lists. However, it should be noticed that a further analysis of their findings suggested that the higher recognition rate was not due to the strong backward association per se but rather to another factor, namely, the inter-item association of the words in their lists. Inter-item association is the degree of preexisting connections between the words in the lists and it is measured by the average probability of the list’s words yielding each other as associates. McEvoy et al. found that lists with strong backward association strength were actually also highly inter-connected, and attributed the greater veridical recognition to this interconnectivity. They corroborated this hypothesis in follow-up experiments (Experiments 2 and 3). Their explanation was that during encoding greater activation was accrued among items that were highly interconnected than between items that were less inter-connected.

Measurement of the inter-item association in the lists in the present experiment revealed that weak backward association lists were significantly higher in inter-item association than lists with the strong backward association, $F(1,16) = 7.19$, $MSE = .35$, $p < .01$ (1.85 vs. 1.13, respectively). Thus, we are inclined to conclude that the present results do not conflict with those of McEvoy and colleagues, but rather are compatible. As in McEvoy et al. study, we also found greater veridical recognition when the inter-item association was high. One point that remains unclear is the type of relationship between inter-item association and backward association. Although McEvoy et al. found greater inter-item association in lists with strong backward association strength, we found the opposite trend. This relationship could be spurious but we believe that negative relationship found in our study is more easily explained than their positive correlation: If the critical lure is a strong associate of the list’s words, then it is statistically less likely that other words from the list would also be produced as associates. Thus, strong backward association and inter-item association should be negatively correlated.\(^1\)

**Recall**

The false recall results revealed a similar pattern to the one observed in false recognition. Lures were recalled more often in lists with strong backward associations (.26) than in lists with weak backward associations (.13), $F(1,37) = 18.40$, $MSE = .03$, $p < .001$, and low familiarity lures (.24) intruded more often than high familiarity lures (.15), $F(1,37) = 11.74$, $MSE = .02$, $p < .001$. In addition, there was a significant interaction between mean backward association and lure familiarity, $F(1,37) = 6.35$, $MSE = .02$, $p < .01$. Further analysis of the interaction showed that in the weak backward association condition there was no difference in false recall between the low and high familiarity lures, $F < 1$ (false recall was .15 and .12, respectively). However, in the strong backward association condition there was a significant difference in false recall between the low and high familiarity lures, $F(1,37) = 12.86$, $MSE = .03$, $p < .001$. The proportion of low familiarity lure intrusions was .33 while the proportion of high familiarity lure intrusions was only .18.

The mean probability of veridical recall was .69. Veridical recall was equivalent across all conditions. The lack of influence of any of the dependent variables on veridical recall (as opposed to their role in veridical recognition) will be elaborated on below.

The similar pattern when comparing false recall and recognition of the lures may reflect the fact that analogous processes underlie both modes of false retrieval, as claimed by generate-recognize theories (Anderson & Bower, 1972, 1974; Balota & Neely, 1980) These models assume that recall involves two stages: (a) a generation process in long-term memory and (b) a recognition process performed on the items that have been retrieved. The generation of the lures, according to these theories, can occur in one of two ways: One possibility is that the accrued activation of the lure during encoding increases the probability of its recall.

---

1 We thank Colleen Kelley for this suggestion. In order to further assess the type of relationship between inter-item association and backward association strength, we computed the correlation between these two variables based on the word association norms used by McEvoy et al. (1999; Nelson, McEvoy, & Schreiber, 1998). The correlation, although small, was significant, $r = - .09$, $p < .0001$ ($n = 4095$). More importantly, the negative correlation strengthens the notion that a higher probability of supplying the lure as an associate (i.e., strong backward association) is coupled with a lower probability of producing other related associates (i.e., weaker inter-item association).
Another possibility (first suggested by Deese, 1959) is that the recall of closely related words during retrieval activates the critical lures. In any case, once the critical lure is activated and the second stage (recognition) is initiated, low familiarity lures are favored over high familiarity as candidates for an old judgment because of their greater memory trace strength. However, this account should be viewed with caution, as previous research (e.g., Deese, 1959; see also following discussion concerning veridical recall) has revealed inconsistent results regarding frequency effects in recall.

Although the current study is the first to investigate the influence of the lure’s familiarity on subsequent false recognition and recall, recent data further corroborate our findings. Roediger and colleagues (2001) conducted a multiple regression analysis in order to assess the potential contribution of various lists’ and lures’ characteristics to false memory. Their analysis was based on 55 lists obtained from different studies. Although backward association was found to be a strong predictor of false memory, no connection was found between the lure’s frequency and its false recall or recognition. However, our study points to the fact that familiarity effects are observed only when backward association is high. We therefore selected from the Roediger et al. (2001) study the lists that were similar to ours in both backward association and familiarity. Twenty two lists fulfilled these criteria, with a mean backward association score of .20 (range .10 to .43). Eleven lists consisted of low frequency lures (raw frequency of 55 based on Kučera & Francis, 1967) and 11 lists of high frequency lures (mean frequency of 290). The lists did not differ in their mean backward association score (.20 and .20 for low and high frequency lists, respectively; t(20) < 1). We then computed the amount of false recognition and false recall in these 2 groups of lists, based on data supplied by Roediger and colleagues. We found the results to be surprisingly similar to those from the present experiment. False recognition of the low familiarity lists was .75 while for the high familiarity lists it was .64, t(20) = 2.45, p < .05. The trend in false recall was similar, although it did not reach significance (.48 and .40 for low and high frequency lists, respectively, t(20) = 1.57, p > .13).

In contrast to veridical recognition, in which effects of both familiarity and inter-item association were found, in veridical recall no such effects were observed. Previous findings have shown that a robust frequency effect is indeed obtained in recall, as high frequency words are remembered better than low frequency words (e.g., Postman, 1970; Tan & Ward, 2000). However, this effect is limited mainly to pure lists, i.e., lists that are comprised of exclusively high or low frequency words. With mixed lists, the prevalent finding is the absence of a frequency effect (Gillund & Shiffrin, 1984; MacLeod & Kampe, 1996; Ward et al., 2003; Watkins, LeCompte, & Kim, 2000). In the current study, although low familiarity lists were on average less familiar than high familiarity lists, they cannot be considered pure lists because they contained words from both categories. Thus, the absence of a frequency effect in the veridical recall data would be expected. The correlation between the familiarity of the lists’ words and their recall was also non-significant.

The lack of an inter-item association effect in recall differs from its existence in recognition. Such a dissociation between the two tasks was not found by McEvoy et al. (1999, Experiment 2), where an enhanced memory performance for words with strong inter-item association was observed in both recall and recognition. However, several differences exist between their study and ours. One notable difference is that the strong and weak inter-item association lists in McEvoy et al. were characterized by weak backward association links between the lists’ words and their critical lures. In the present study, however, a negative correlation was found between inter-item association and backward association, such that strong backward association lists were low in connectivity and vice versa. Theoretically, two distinct factors may play a role in increasing veridical recall: The first is the inter-item association between the words in the list that causes either mutual activation during encoding or cuing during recall. The second factor is the average backward association of the list’s words to the critical lure, which serves as a common denominator that helps one group the words into a central theme. In the McEvoy et al. study, the second factor was deactivated by using only weak backward association lists. As a result, only the effect of inter-item association was observed. In the present study, both factors were active but may have cancelled out one another because lists with strong backward association had weak inter-item association strength, whereas lists with weak backward association had strong inter-item association strength. The opposing effects of the two factors may have resulted in the comparable accuracy in veridical recall. Nevertheless, further research manipulating systematically both backward and inter-item association is obviously required to evaluate these post hoc hypotheses.

Experiment 2

The aim of Experiment 2 was to replicate the results of the previous experiment while ruling out some alternative hypotheses. First and foremost was the possible influence of recall on the subsequent recognition test. Several studies (e.g., Roediger & McDermott, 1995; Roediger et al., 1997) have shown that recall can boost subsequent false recognition. Moreover, the pattern of false recognition found in Experiment 1 (i.e., more false recognition of low familiarity than high familiarity lures...
but only in the strong backward association condition) maps directly onto the pattern of prior false recall. This similarity suggests that the unique pattern of the recognition results we obtained may stem from the earlier recall and is not directly related to the familiarity of the lure. Therefore, in Experiment 2 half of the lists were followed by an immediate free recall test, whereas the other lists were not followed by recall. Following the presentation of the lists a recognition test was administered which included lures from all lists, recalled and not recalled alike. If the familiarity effect that was obtained in the recognition test is due to the recall performance then it will be found only for lures whose lists were recalled. If, however, the recognition familiarity effect is not dependent on prior recall, then it will be found in both types of lures.

Another factor that may have affected the results in Experiment 1, involves the correlations between the lures’ and the lists’ characteristics. Thus, for example, words from lists with low familiarity lures were less familiar than words from lists with high familiarity lures, and so potentially the differential false recognition of high and low familiarity lures could have resulted from the difference in the lists’ familiarity. Indeed, this possibility is unlikely because words from low familiarity lists were better recognized than words from high familiarity lists, and previous research has shown that veridical recognition is negatively correlated with false recognition (Roediger et al., 2001). Thus, low familiarity lures were expected to be less falsely recognized than high familiarity lures, a hypothesis that was not supported by our findings, which instead showed the opposite results. Nevertheless, in order to rule out possible confounding influences of the word lists’ characteristics we modified our lists and equated them on familiarity, inter-item association and other related variables. As the lures’ frequency effects were found in Experiment 1 only in strong backward association lists the critical lists generated in Experiment 2 were lists where the lures were strongly related to the lists’ words.

A final issue we aimed to clarify in Experiment 2 was the role of familiarity in false recall. Experiment 1 showed a familiarity effect in recall that was similar to the one obtained in recognition. Yet, despite the common patterns the sources of the familiarity effects may have been different for recall and recognition. More specifically, familiarity effects in recall could stem, according to the generate-recognize model (Anderson & Bower, 1972, 1974; Balota & Neely, 1980) from the recognition stage per se, thus leading to a recall pattern comparable to that found in recognition. Alternatively, familiarity effects could arise during the generation process. Although high familiarity words will be favored at this stage (e.g., Kintsch, 1970), the strong backward association of the lists and the specific nature of the words in the low familiarity lists may attenuate this advantage and may even reverse it. As a result, more low familiarity lures may be generated at this early stage than high familiarity lures. Inspection of the low familiarity lists that yielded high degrees of false recall revealed that some of the words recalled were synonyms of the lure (e.g., glen-valley) or specific instances of the lure (e.g., Jordan-valley). Thus, the familiarity effect in recall that we found in Experiment 1 could be the result of more low familiarity lures generated in the first stage of the recall process by their close semantically related associates. It should also be noted that these words did not appear in the recognition test, thus limiting the effect solely to recall. In the lists used in Experiment 2 these words were deleted, and some lists were replaced completely. We predicted that after the exclusion of specific items which were more likely to activate their strongly related lures, no familiarity effects would be found in recall.

Method

Participants

Forty-four participants from the Ben-Gurion University participant pool took part in the experiment for course credit. All of them were native Hebrew speakers and had normal or corrected to normal vision and hearing.

Materials

Ten critical lists were constructed using word association norms (Rubenstein et al., 2003), following the same procedure as in Experiment 1. The lists were shorter and included only 10 words in each list, in contrast to the 12 word lists in Experiment 1. Five lists were comprised of high familiarity lures ($M = 6.46$), whereas the non presented lures in the other five lists consisted of low familiarity lures ($M = 3.64$), $t(8) = 10.26$, $SE = .28$, $p < .0001$. The lures differed also in their frequency, $t(8) = 2.43$, $SE = 27.55$, $p < .05$ but were equated on concreteness, $t(8) = .93$. In addition, we attempted to equate the lists on other relevant attributes: The two groups of lists did not differ in their mean backward association strength, $t(8) = 1.16$, mean forward association strength, $t(8) = 1.22$, inter-item association, $t(8) = 1.75$, familiarity, $t(8) = 1.26$, concreteness, $t(8) = .54$, or frequency, $t(8) = 1.03$.

Forty distractors were selected for the recognition test, half of them high familiarity words and the other half low familiarity words, $t(38) = 32.15$, $SE = .08$, $p < .0001$. They differed also in frequency $t(38) = 2.91$, $SE = 28.69$, $p < .01$, but not in concreteness, $t(38) = 1.88$. The characteristics of the list words, lures, and distractors are presented in Table 3.

Ten control lists were added to the experiment so the approximate time to complete the experiment would match Experiment 1. They were developed in a way sim-
ilar to the critical lists, with each list containing the 10 most common associates to a non-presented lure. The results of these lists were not analyzed.

**Procedure**

The 10 critical lists were randomly interspersed among the 10 control lists, and the 20 lists were presented either in a forward (from list 1 to 20) or in a backward fashion (from list 20 to list 1). The study procedure was similar to the previous experiment, except that participants were asked to recall the words after only half of the lists, while after the other half they were asked to solve math problems. They were informed of the specific task to be performed only after the list was administered in order to prevent use of specific learning strategies (e.g., Balota & Neely, 1980). Since only five lists with critical lures were recalled during this stage, half of the participants were given two low familiarity lists and three high familiarity lists to be recalled, while the other half were given three low familiarity and two high familiarity lists. However, the recognition test that followed the presentation of the 20 lists included the lures of all critical lists. The structure and administration of the recognition test was the same as in Experiment 1.

**Results and discussion**

Analysis of variance (ANOVA) was conducted on the proportion of words remembered (whether falsely or correctly) with familiarity of the lure (high/low familiarity) and type of task administered after list presentation (recall/no recall [math problem solving]) as independent variables. Both independent variables were manipulated within subjects. Means of all critical conditions are presented in Table 4.

<table>
<thead>
<tr>
<th>Lists</th>
<th>Backward association</th>
<th>.14</th>
<th>.03</th>
<th>.19</th>
<th>.04</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Forward association</td>
<td>.07</td>
<td>.01</td>
<td>.06</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>Inter-item association</td>
<td>1.1</td>
<td>.33</td>
<td>.50</td>
<td>.12</td>
</tr>
<tr>
<td></td>
<td>Familiarity</td>
<td>4.52</td>
<td>.11</td>
<td>4.76</td>
<td>.17</td>
</tr>
<tr>
<td></td>
<td>Concreteness</td>
<td>5.17</td>
<td>.31</td>
<td>4.98</td>
<td>.15</td>
</tr>
</tbody>
</table>

Lures

| Familiarity | 3.64 | .26 | 6.46 | .10 |
| Concreteness | 5.86 | .30 | 6.17 | .16 |

Distractors

| Familiarity | 3.61 | .06 | 6.19 | .05 |
| Concreteness | 5.79 | .18 | 6.20 | .07 |

Note: The inter-item association value denotes the average number of list’s words produced by each associate.

Table 4

Mean proportion of recognition and free recall for list words and lures in Experiment 2 as a function of lure familiarity

<table>
<thead>
<tr>
<th>Recognition</th>
<th>Low familiarity</th>
<th>High familiarity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without recall</td>
<td></td>
<td></td>
</tr>
<tr>
<td>List words</td>
<td>.55</td>
<td>.03</td>
</tr>
<tr>
<td>Lures</td>
<td>.67</td>
<td>.05</td>
</tr>
<tr>
<td>With recall</td>
<td></td>
<td></td>
</tr>
<tr>
<td>List words</td>
<td>.73</td>
<td>.03</td>
</tr>
<tr>
<td>Lures</td>
<td>.75</td>
<td>.05</td>
</tr>
<tr>
<td>Free recall</td>
<td></td>
<td></td>
</tr>
<tr>
<td>List words</td>
<td>.66</td>
<td>.02</td>
</tr>
<tr>
<td>Lures</td>
<td>.22</td>
<td>.04</td>
</tr>
</tbody>
</table>

Recognition

The false recognition results of Experiment 2 produced a similar pattern as in Experiment 1. Low familiarity lures were more often falsely recognized than high familiarity lures (.71 and .58, respectively, $F(1,44) = 6.34$, $MSE = .13$, $p < .01$). As in Experiment 1, this familiarity effect was obtained in lists whose words were strongly associated to the non-presented lure.

Moreover, this pattern of results was not influenced by prior recall, discounting the hypothesis that the familiarity effect in false recognition in Experiment 1 stemmed from prior recall. Indeed, false recognition was higher for lures from lists that were recalled (.69) compared to lures from lists that were not (.59), $F(1,44) = 6.86$, $MSE = .05$, $p < .01$). However, the interaction between lure familiarity and prior recall was not significant, $F(1,44) < 1$, indicating that the familiarity effect (i.e., low familiarity false recognition - high familiarity false recognition) was comparable for lists that were recalled and lists that were not (.12 and .15, respectively).

In contrast, the analysis of the distractors in the recognition test revealed a different pattern of performance, with higher false alarm rates for high familiarity lures (.17) than low familiarity lures (.07), $t(43) = 6.09$, $MSE = .01$, $p < .001$.

Analysis of veridical memory of list words revealed that participants recognized more words from lists that were previously recalled (.71) than from lists that were not recalled (.53), $F(1,44) = 44.95$, $MSE = .03$, $p < .0001$. No other effects were found to be statistically significant.

Recall

A familiarity effect was not found in the analysis of the recall results: False recall of low familiarity lures was similar to that of high familiarity lures (.22 and
whether lures are activated during the encoding phase, as claimed by some accounts of the false memory effect, or whether false memories are generated at retrieval due to their similarity to the list’s theme. In order to achieve this goal we manipulated the frequency of the lures. Due to the different patterns observed in the literature for low and high frequency old and new words, namely more hits for low frequency words and more false alarms to high frequency distractors, we sought to find out whether the lures' frequency effects in false memory would imitate old words or distractors.

The results suggest that the lure’s familiarity affected the likelihood of falsely remembering it as an item that appeared in the study list. In Experiment 1, false recognition was higher for low familiarity than high familiarity lures, but only in lists with strong backward association. This familiarity effect was replicated in Experiment 2. Moreover, the present study exhibited a pattern of frequency effects in lures that was more similar to the pattern found with hits for old items than with false alarms to distractors. While more low familiarity lures were recognized as old than high familiarity lures, the false alarm rate for distractors was greater for high familiarity than for low familiarity distractors.

The most traditional and prevalent account of the false memory effect is the activation/monitoring theory (a term designated by Roediger, Watson et al., 2001; see also Roediger, Balota et al., 2001), which is a hybrid model inspired by Underwood’s (1965) implicit associative response hypothesis, automatic spreading activation theory (Collins & Loftus, 1975; Posner & Snyder, 1975) and reality monitoring theory (e.g., Johnson et al., 1993). According to this account, the critical lure is activated during encoding by related associates, in a conscious or unconscious manner. During recognition, the partial activation of the lure induces participants to recognize it as old. In order to explain the pattern of results obtained in this research, the activation/monitoring theory incorporates memory-based approaches to the mirror effect that accentuate the unique memory trace of low frequency words. Thus, for example, the SAC model (Reder et al., 2000) assumes that two types of nodes, concept and event nodes, are activated when a word is studied, representing lexical-semantic and episodic-contextual information, respectively. Low and high frequency words differ in both their concept and event nodes. The baseline activation level of a high frequency concept node is stronger than that of a low frequency one due to past recurrent encounters with the word. In contrast, the event nodes for a low frequency word are more strongly activated during recognition, because fewer event nodes are associated with a low frequency concept node than with a high frequency concept node. During retrieval the stronger activation of the event node for a low frequency old word would facilitate the recollection of the word to a greater extent than a high frequency word, resulting in more hits for low frequency

General discussion

The aim of the present study was to investigate whether lures are activated during the encoding phase, 

\[ t(43) = 1.01 \]. The veridical memory results for both recognition and recall go against the possibility that the different levels of false memory could be attributed to lists’ characteristics, as the low and high familiarity lists were equated on many critical variables, to the maximum extent possible. This was reflected in the equivalent levels of veridical recognition and recall across the different conditions. The main difference between the Experiments 1 and 2 concerns the familiarity effects on false recall. Whereas in Experiment 1 low familiarity lures were recalled more often than high familiarity lures, no such effect was found in the present study. This null effect cannot be attributed to lack of statistical power as our power to detect a difference with a magnitude similar to Experiment 1, that is a Cohen’s $d$ of .54, was greater than .94 (Erdelder, Faul, & Buchner, 1996). Even if the true effect size was much smaller (for example, a Cohen’s $d$ of .35) our power to detect it would be greater than .75 ($p < .05$, one-tailed).

One basic change between the two experiments is the omission, in Experiment 2, of words that were strongly semantically related to the lure from the low familiarity lists. Because some lists were repeated in both experiments, we could compare false recall between the two experiments. Thus, for example, the intrusion rate of valley was .45 in Experiment 1, but it dropped dramatically to .06 in Experiment 2 after strongly semantically related words were deleted. The same was observed for the lure word lion, although the decrease was more moderate (from intrusion rate of .39 to .27). This decrease cannot be attributed solely to backward association strength, because the associative relationship was similar for low familiarity lists between experiments. In addition, the omission of these words did not affect the familiarity effects in false recognition. Thus, it appears that the familiarity effect in recall in Experiment 1 stemmed from the equivalent levels of veridical recognition and recall across the different conditions.

During retrieval the stronger activation of the event node for a low frequency old word would facilitate the recollection of the word to a greater extent than a high frequency word, resulting in more hits for low frequency

\[ r(43) = .21 \]. The fact that a familiarity effect was not obtained in false recall also supports the notion that the familiarity effect observed in recognition is independent of previous recall.

Also, the recall rate of the lists’ words did not differ between groups (.66 and .64 for low and high familiarity lists, respectively, $t(43) = 1.01$). The veridical memory results for both recognition and recall go against the possibility that the different levels of false memory could be attributed to lists’ characteristics, as the low and high familiarity lists were equated on many critical variables, to the maximum extent possible. This was reflected in the equivalent levels of veridical recognition and recall across the different conditions.

The main difference between the Experiments 1 and 2 concerns the familiarity effects on false recall. Whereas in Experiment 1 low familiarity lures were recalled more often than high familiarity lures, no such effect was found in the present study. This null effect cannot be attributed to lack of statistical power as our power to detect a difference with a magnitude similar to Experiment 1, that is a Cohen’s $d$ of .54, was greater than .94 (Erdelder, Faul, & Buchner, 1996). Even if the true effect size was much smaller (for example, a Cohen’s $d$ of .35) our power to detect it would be greater than .75 ($p < .05$, one-tailed).

One basic change between the two experiments is the omission, in Experiment 2, of words that were strongly semantically related to the lure from the low familiarity lists. Because some lists were repeated in both experiments, we could compare false recall between the two experiments. Thus, for example, the intrusion rate of valley was .45 in Experiment 1, but it dropped dramatically to .06 in Experiment 2 after strongly semantically related words were deleted. The same was observed for the lure word lion, although the decrease was more moderate (from intrusion rate of .39 to .27). This decrease cannot be attributed solely to backward association strength, because the associative relationship was similar for low familiarity lists between experiments. In addition, the omission of these words did not affect the familiarity effects in false recognition. Thus, it appears that the familiarity effect in recall in Experiment 1 stemmed from the process occurring solely in recall, perhaps in the generation stage, and was semantic in nature than frequency dependent. Thus, the results of Experiment 2, as well as previous research (Deese, 1959; Roediger et al., 2001) which did not find evidence for greater false recall of low familiarity lures, indicate that familiarity effects in false memory are restricted to recognition.

The results suggest that the lure’s familiarity affected the likelihood of falsely remembering it as an item that appeared in the study list. In Experiment 1, false recognition was higher for low familiarity than high familiarity lures, but only in lists with strong backward association. This familiarity effect was replicated in Experiment 2. Moreover, the present study exhibited a pattern of frequency effects in lures that was more similar to the pattern found with hits for old items than with false alarms to distractors. While more low familiarity lures were recognized as old than high familiarity lures, the false alarm rate for distractors was greater for high familiarity than for low familiarity distractors.

The most traditional and prevalent account of the false memory effect is the activation/monitoring theory (a term designated by Roediger, Watson et al., 2001; see also Roediger, Balota et al., 2001), which is a hybrid model inspired by Underwood’s (1965) implicit associative response hypothesis, automatic spreading activation theory (Collins & Loftus, 1975; Posner & Snyder, 1975) and reality monitoring theory (e.g., Johnson et al., 1993). According to this account, the critical lure is activated during encoding by related associates, in a conscious or unconscious manner. During recognition, the partial activation of the lure induces participants to recognize it as old. In order to explain the pattern of results obtained in this research, the activation/monitoring theory incorporates memory-based approaches to the mirror effect that accentuate the unique memory trace of low frequency words. Thus, for example, the SAC model (Reder et al., 2000) assumes that two types of nodes, concept and event nodes, are activated when a word is studied, representing lexical-semantic and episodic-contextual information, respectively. Low and high frequency words differ in both their concept and event nodes. The baseline activation level of a high frequency concept node is stronger than that of a low frequency one due to past recurrent encounters with the word. In contrast, the event nodes for a low frequency word are more strongly activated during recognition, because fewer event nodes are associated with a low frequency concept node than with a high frequency concept node. During retrieval the stronger activation of the event node for a low frequency old word would facilitate the recollection of the word to a greater extent than a high frequency word, resulting in more hits for low frequency
words. On the other hand, the stronger baseline activation of the concept node for high frequency words would result in falsely identifying high frequency new distractors as old, leading to higher false alarm rates than with low frequency distractors.

If concept and event nodes are also activated for the non-presented lures by summation of activation of strongly related associates, then, during recognition, low familiarity lures will be more likely to be judged as old than high familiarity lures, because event nodes of the former would be activated to a greater extent than the latter. This effect will be limited only to lures from lists with strong backward association, as the probability of activating the lure, and consequently its event node, is positively correlated with the strength of connections between the lure and its associates.\(^2\)

The second memory-based approach to false memory distinguishes between two different types of information that are processed during study (Arndt & Reder, 2003; Brainerd & Reyna, 2002; Brainerd et al., 2001; Hunt & McDaniel, 1993; Payne, Elie, Blackwell, & Neuschatz, 1996). The first is gist (or relational) information, which is the semantic similarities among study items. The second is verbatim (or item-specific) information, which includes the unique and distinctive aspect of each item. The failure to correctly reject the lures as new items stems from the difficulty in recollecting specific characteristics of the studied items and reliance on the common aspects of the list’s words.

The enhancement of false recognition of low familiarity lures in lists with strong backward association could be easily accounted for by this approach. The claim that might be advanced is that lists whose items are strongly associated with the non-presented lures share common characteristics from which relational information is derived. However, it would be harder to provide a plausible explanation to the familiarity effect obtained for the lures in the high backward association condition. As the source of the false recognition effect is identified in the reliance on the preserved themes of the lists, the familiarity of the lure seems irrelevant and does not seem to play a role in the identification of commonalities among list items. Thus, the most probable prediction of this approach is that the strong effect of the relational information, produced by the strong backward association, would nullify any frequency effects. Alternatively, lures’ performance should be similar to distractors’ performance, namely, more false recognition for high than low frequency items.

One possible explanation that can be offered by this latter approach is that low familiarity lures are more similar to the gist formed by the words in their lists than high familiarity words, which may have more distinctive characteristics that differentiate them from the list’s gist. However, according to this claim, greater false recognition would have been expected for low familiarity lures in Experiment 1, regardless of the strength of backward association. Instead, low familiarity lures were falsely recognized to a greater extent only when the backward association was high. This result, as well as the failure of lure frequency to contribute to false recognition in the Roediger et al. (2001) multiple regression analysis, indicates that low frequency lures are not necessarily more similar to the gist of their lists.

The same type of difficulty would be encountered by a third theoretical approach that was proposed to explain the false memory phenomenon (Miller & Wolford, 1999). According to this view, false memories are the consequence of criterion shifts initiated by the participants. During the study of the lists, participants consciously realize the interrelationship among the list items and categorize them into a unitized topic. Later, during the test, when a related item appears, participants tend to judge that it appeared previously due to its strong semantic links to the list. Thus, participants set a lower criterion for related lures (see also Jacoby, Kelley, & Dywan, 1989).

As with the previous approach, the criterion-shift accounts of false memory can explain the increase in false recognition of lures from strong backward association lists compared to weak backward association lists. This advantage is determined by the different criteria applied to the different types of lists (the problematical issues associated with this procedure are addressed critically by Wixted & Stretch, 2000). Unfortunately, it confronts difficulties when attempting to account for the frequency effects found within the strong backward association lists.

Despite the apparent superior ability of the activation/monitoring model in explaining familiarity effects in false recognition, we do not advocate it as the sole mechanism underlying the phenomenon. The list-learning paradigm, first introduced by Deese (1959) and recently resurrected by Roediger & McDermott (1995), yields complex effects that probably encompass several processes occurring both at encoding and at retrieval. Some findings, such as forgetting rates (Payne et al., 1996) and the effects of pictorial presentation (Israel & Schacter, 1997; Koutstaal & Schacter, 1997), give less support to the activation/monitoring approach and fit better with the gist memory-based approach. In our results we can point to findings which may be explained more easily by the latter approach. We refer to the lack
of a familiarity effect for weak backward association lists in Experiment 1, despite a sizeable level of false recognition. We admit that it is not likely that this level of false memory can be attributed to activation of the lure itself. One possible mechanism that may underlie false memory in this condition is the over reliance on schematic or gist information which emphasizes the common aspects of the words in the lists and allows the lure to be judged as old. Yet, this admission does not undermine the importance of the familiarity effect found in the present study in the lists with strong backward association. As detailed above fuzzy trace theories encounter greater difficulty than activation theories in accounting for the influence of familiarity on false recognition. More importantly, the procedure introduced in this study allowed us to show empirically that a lure may indeed be activated during study if the words presented in the list are strongly associated to it. We therefore believe that this approach may prove to be a promising technique in elucidating and dissociating the processes underlying true and false memories.

Appendix. Critical lures of Experiments 1 and 2

Weak backward association/low familiarity: feather, tank, hell, fuel, necklace.
Weak backward association/high familiarity: father, week, penny, family, mistake.
Strong backward association/low familiarity: ant, valley, lion, wheat (fox in Experiment 2), forest.
Strong backward association/high familiarity: house, country (money in Experiment 2), table, car, telephone.

References


---

An anonymous reviewer suggested that the high rate of false recognition in the weak backward association lists may indeed be attributed to the lure's activation during study. Yet, no familiarity effects were observed in false recognition in these lists, compared to strong backward association lists, because the lures were not activated to the extent of forming an explicit (conscious) representation as the lures in the strong backward association lists. This account is consistent with the findings of the present study, which emphasize the importance of activation process in false memory. Yet, we suspect that the very weak backward association strength of the lists (i.e., .03) was unlikely to activate even unconsciously the lure to such a degree that it would yield such high rates of false recognition.


