Research Article

ASSOCIATIVE PROCESSES IN FALSE RECALL AND FALSE RECOGNITION

Kerry J. Robinson and Henry L. Roediger, III

Washington University in St. Louis

Abstract—*Studying a list of words associated to a critical nonpresented word results in high rates of false recall and false recognition for that nonpresented item (Roediger & McDermott, 1995). Two experiments examined the effect of manipulating the number of associates presented on false recall and later false recognition of a nonpresented item: In Experiment 1, associate lists of varying lengths were studied; in Experiment 2, list length was held constant and the number of associates within the list was manipulated. In both experiments, the rate of critical intrusions in recall increased steadily with increasing number of associates studied. Most notably, the filler words used in Experiment 2 to equate the list lengths did not affect the rate of critical intrusions, although they did depress recall of studied words. False recall and false recognition appear to be tied to the total, not the mean, associative strength of items in the list.*

False memory for an event may be induced by the study of associated events. Deese (1959) demonstrated this phenomenon by having subjects study and recall lists of the 12 most common associates to a nonpresented item and then examining the probability with which the nonpresented item intruded during recall. For example, subjects studied the words *bed, rest, awake, tired, dream, wake, snooze, blanket, doze, slumber, snore,* and *nap*, all of which are strongly related to the nonstudied item *sleep*. Deese noted that for some of the lists, the probability of false recall of the nonpresented item (*sleep*) was very high, with intrusions occurring on 30% to 40% of subjects' recall protocols.

Roediger and McDermott (1995) revived the paradigm, first using six of Deese's (1959) 12-item study lists, and then expanding the materials to consist of twenty-four 15-item lists. For the 12-word lists, Roediger and McDermott reported a mean false recall probability of .40 (Experiment 1); use of the 15-word lists produced a mean false recall probability of .55 (Experiment 2). Rates of false recognition for the nonpresented items in some cases equaled the rates of correct recognition for studied words.

Other researchers have reported similar findings (McDermott, 1996; Payne, Elie, Blackwell, & Neuschatz, 1996; Read, 1996; Schacter, Verfaellie, & Pradere, 1996). Of note, several researchers have used Tulving's (1985) remember/know procedure, in which subjects are asked to indicate whether they actually *remember* the presentation of a nonpresented item in a study list or simply *know* that it occurred; these researchers have consistently found that subjects are highly likely to report remembering the occurrence of the critical items that were never presented (Roediger & McDermott, 1995; Payne et al., 1996; Read, 1996). Additionally, Payne et al. (1996, Experiment 3) demonstrated that subjects in this paradigm would willingly report whether a particular item had been presented in a male or a female voice when in fact that word had not been presented.

The mechanisms underlying the powerful false recollection observed in this paradigm remain the subject of debate. Underwood (1965) proposed that the study of a list item may produce an implicit associative response (IAR), or the unintentional conscious activation of a word that is strongly related to that studied item. As a consequence, nonpresented items that are strongly associated to study items may be encoded during study along with the actual list words. This account suggests that, during recall, the nonpresented associate items should behave similarly to the studied items. Roediger and McDermott (1995) reported that it is indeed the case that the nonpresented items are recalled with approximately the same probability as the studied items in the middle serial positions of a study list. In addition, they showed that in a recognition test, subjects judge critical nonstudied words to be old at rates comparable to the hit rates for studied items. Finally, they found that subjects readily report remembering (as opposed to knowing) that the nonstudied items were presented, lending support to the hypothesis that the nonstudied items were encoded during study along with the list words.

In his original discussion of this paradigm, Deese (1959) proposed a slightly different account. He suggested that "the probability of a particular word occurring as an intrusion in immediate free recall of a list of words may be predicted from the tendency for the intruding word to occur as a response in free association to the items on the list" (p. 21). Note that the associative strengths referred to by Deese represent the probabilities with which the critical item will be elicited by the various list items, or *backward* association strengths. These values are the reverse of the associative strengths used in constructing the study lists, which refer to the probabilities of the individual list words being elicited by the critical item (*forward* association strengths).

Deese (1959) concluded that "the probability of intrusion is very well predicted by the simple mean . . . of the associative tendencies for the intrusion to be elicited by the test words" (p. 21). According to Deese, this account "implies that in the process of recollection, words and concepts associated with remembered items will be added" (p. 21). In other words, Deese suggested that during recall, a nonstudied item will intrude to the extent that the average associative strength of the studied words to that specific intrusion is high. Indeed, he reported that the correlation between the average associative strength of the list items to the critical nonpresented word (or mean backward association strength of the list) and probability of intrusion of that word was +.87 across his lists.

Address correspondence to Kerry J. Robinson or Henry L. Roediger, III, Department of Psychology, Box 1125, Washington University in St. Louis, One Brookings Dr., St. Louis, MO 63130-4899; e-mail: kjrobins@artsci.wustl.edu or roediger@artsci.wustl.edu.

In evaluating this hypothesis, Deese (1959) used lists that consisted of exactly 12 study items. In two experiments reported here, we examined the pattern of false recall and false recognition that occurs when study lists of varying lengths are used. We tested recall and recognition of 3-, 6-, 9-, 12-, and 15-item lists of related words taken from the Roediger and McDermott (1995) materials. As already mentioned, the study lists were constructed specifically of items that are strongly associated to a critical nonpresented item. Because of the limited number of potential study items that are strongly related to a critical item, longer study lists will necessarily include some items that are less strongly associated to that item. Thus, a short list—consisting of, say, the six strongest associates of a critical item—will have a higher mean associative strength than a long list. If forward and backward association strengths are related, then the same pattern should be expected for the lists' mean backward association strengths, or the mean association strengths of the list words back to the critical item. Indeed, when the lists were compared, the mean backward association strength of the lists to the critical items was reliably greater for shorter than for longer lists, as discussed in the next section.

Consequently, if Deese's hypothesis were to be extended to study lists of varying lengths, it would predict that shorter lists would produce higher levels of false recall of a critical item than longer lists because of the greater average backward association strength for the shorter lists. It should be noted, however, that this strong form of the hypothesis was never explicitly stated by Deese (1959).

An analogous pattern would be expected if the mechanism were similar to Underwood's (1965) IAR proposal. According to this perspective, the strongly associated critical item should behave as an additional study list word. It has been well established that as length of a study list increases, the probability of correct recall of any specific studied item decreases systematically (Murdock, 1961). The total number of correct items recalled increases with list length, of course, but the probability of any single item being recalled drops as additional items are added to the study list. If highly associated nonstudied items behave as if they had appeared on the study list, then one would again expect the probability of false recall of a critical item to decrease with increasing list length. Thus, according to predictions drawn from both Deese's (1959) and Underwood's (1965) hypotheses, the rate of false recall of a critical nonpresented item should correlate negatively with the number of associates in a study list.

Other considerations suggest the opposite pattern: that a greater number of studied associates should lead to higher levels of false recall for the critical item. For example, false recall of a highly associated nonpresented item may be predicted not by the mean association strength of the list words, but by the total associative strength of these words. That is, the greater the total associative strength of the list—the more the words are associated to the critical nonpresented item—the greater the probability that the item will be falsely recalled and recognized. Thus, a longer list of associates should produce higher levels of false recall and false recognition of the critical item than should a shorter list, because a long list of associates will have a higher total association strength for that item. Not only does this prediction seem intuitive, but if list length is held constant, it is very similar to Deese's (1959) hypothesis implicating the average associative strength of the list.

There is some evidence to suggest that a larger number of studied associates may lead to increased false recognition for a critical word. Hall and Kozloff (1973) found that in a continuous recognition paradigm, false recognition for a nonpresented word was significantly higher when three associates had been previously presented than when a single associate had been presented three times. Further, Hintzman (1988, Experiment 1) reported that false recognition rates for a critical item increased linearly as the number of studied items from the same category as the critical item increased from one to three to five. The issue of how the number of associates studied affects false recall was not addressed by either set of researchers.

We conducted two experiments to examine the effect of the number of associates studied on false recall and false recognition of a related item. In Experiment 1, the lengths of the study lists were varied systematically (lengths of 3, 6, 9, 12, and 15 items) with the strongest associates to the critical item always included in the lists. We predicted that the probability of false recall and false recognition of the critical item would increase with list length, as the total associative strength of the longer lists was expected to exceed the total association strength of the shorter lists. Note that the proposals of Deese (1959) and Underwood (1965) would lead to the opposite prediction. In Experiment 2, the number of associates in the study lists was similarly manipulated, but this time unrelated filler words were included at the ends of the lists to equate all of the list lengths at 15 items. If the total associative strength of the list items predicts false recall and false recognition of the critical item, the inclusion of the unrelated filler words in the study list should have no effect on the probability of false memory for this item. However, if it is the mean associative strength of the list that predicts false recall and false recognition, then the probability of a false recollection occurring should decrease for longer lists that include more weakly associated items.

EXPERIMENT 1

The purpose of Experiment 1 was to manipulate the number of associates in a study list and examine the effect of this variable on levels of false recall (and later false recognition) of the critical lure. Study lists were constructed such that the strongest associates of the critical item were usually presented first. Thus, the three-word list contained the three strongest associates, the six-word list the six strongest associates, and so on. Lists consisting of 3, 6, 9, 12, and 15 items were presented, and subjects were asked to recall each list following a 30-s delay. An overall recognition task including studied and critical items followed presentation and recall of all the lists.

Method

Subjects

Thirty-six Air Force recruits at Lackland Air Force Base in San Antonio, Texas, completed this study as a training requirement. Two subjects failed to follow instructions and were replaced with Rice University undergraduates, who received course credit for their participation.

Materials

The twenty-four 15-word study lists and accompanying critical targets introduced by Roediger and McDermott (1995) were modified for use in this experiment. The 24 lists were arbitrarily divided into six groups of 4 lists each for counterbalancing purposes, and the length of the lists in a group was rotated among groups (0, 3, 6, 9, 12, or 15 items). (In the "zero" list length condition, items from these lists were included on the final recognition test but were never studied.) Thus, the subjects experienced all of the lists in one group as consisting of 3 items, all the lists in another group as consisting of 6 items, and so on.

Each group of lists occurred in each length condition an equal number of times across subjects, and all subjects received the 24 study lists in the same random order (lists were not blocked by length). Subjects studied the first 3, 6, 9, 12, or 15 words from each list, as they were ordered in the appendix of Roediger and McDermott (1995). The mean backward association strengths of the lists to the nonstudied items were .29, .24, .21, .19, and .18 for lists of 3, 6, 9, 12, and 15 items, respectively, $F(4, 23) = 23.08$, $MSE = 0.002$, $p < .01$.

All words were presented in white on a black background. IBM-compatible computers were used for all instructions and stimuli presentation using Micro Experimental Laboratory software (Schneider, 1988).

Procedure

Subjects were seated in front of a computer in the testing room and given booklets and pencils to record their responses on the recall tests and sheets of scratch paper to complete distractor math problems. Initial instructions informed subjects that they were taking part in a memory experiment and that they should pay close attention to the presented words because they were going to be asked to recall them later. Subjects pressed the space bar to begin presentation of the first study list.

Words in a list were presented in a continuous sequence on the center of the computer screen for 2 s apiece. Following the presentation of the final word in the list, a set of four simple two-digit addition problems appeared, and subjects were instructed to solve the problems on the scratch paper. After 30 s, a short beep sounded, the screen turned white, and subjects were told that they had 1.5 min to recall as accurately as possible as many of the presented words as they could remember from the most recent list. Subjects were warned not to guess on the recall test, and recall responses were handwritten in the booklets. After 1.5 min, the screen turned black to signal completion of the test period, and subjects were instructed to stop writing, turn the page in their booklets, and press the space bar to begin presentation of the next study list. Casual observation revealed that most subjects completed recalling well before the end of the recall period. This procedure was repeated for 20 study lists (4 each of lengths 3, 6, 9, 12, and 15 items).

A final yes/no recognition test was administered to each subject following presentation and recall of all 20 study lists. The recognition test consisted of 40 studied items (2 from each of the 20 studied lists) and 40 nonstudied items (the 24 critical lures, 2 list items from each of the 4 nonstudied lists, and 8 unrelated filler words). In order for the tests to be identical regardless of which lists had been studied at each length, all of the studied items presented on the recognition test were chosen at random from among the first three words in each list. Recognition test items were presented individually on the computer screen, and each remained visible until the participant pressed either the "y" key (to indicate "yes, studied") or the "n" key (to indicate "no, not studied").

Results and Discussion

Recall

The purpose of this experiment was to examine the effect of the number of associates studied on the probability of falsely recalling and falsely recognizing a critical nonpresented word. Comparison of the two functions presented in Figure 1 reveals a clear distinction between the probabilities of veridical recall of the words on the lists and false recall of the critical items. Although the probability of veridical recall decreased with increasing list length, as expected, the probability of false recall increased.

The two functions in Figure 1 were analyzed by a 2 × 5 analysis of variance (ANOVA), which revealed a reliable interaction between probabilities of correct and false recall over the five list lengths, $F(4, 140) = 69.25$, $MSE = 0.02$, $p < .0001$. We then conducted separate one-way ANOVAs on the two functions. As anticipated, the probability of correctly recalling a presented item declined as list length increased, $F(4, 140) = 198.29$, $MSE = 10.38$, $p < .0001$. More interestingly, the probability of falsely recalling a critical item was found to increase systematically with increasing list length, $F(4, 140) = 12.16$, $MSE = 0.58$, $p < .0001$. As may be seen in Figure 1, the probability of false recall was .03 for a 3-word list, increasing to .31 for a 15-word list.

The rate of other, noncritical intrusions in recall was also examined in a one-way ANOVA, but the main effect of list length was not reliable for this variable, $F(4, 140) = 1.61$, $MSE = 0.44$, $p > .1$. The mean rates of these intrusions for list lengths of 3,

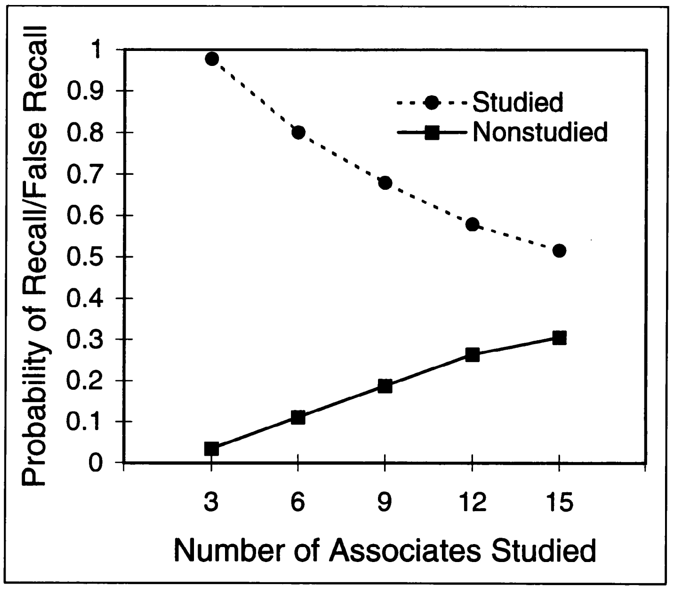

Fig. 1. Mean probabilities of correct and false recall in Experiment 1, as a function of the number of associates studied.

Table 1. *Proportion of "old" responses on the final recognition test in Experiments 1 and 2 as a function of the number of associates studied*

Item type	Number of associates 0	3	6	9	12	15
	Experiment 1					
Studied	.12	.82	.85	.89	.91	.87
Critical	.11	.37	.50	.58	.67	.69
Not recalled[a]	.11	.34	.45	.50	.60	.59
	Experiment 2					
Studied	.18	.83	.89	.86	.84	.86
Critical	.26	.49	.53	.64	.69	.69
Not recalled[a]	.26	.48	.45	.58	.60	.59

[a]Proportion of nonrecalled critical items that were later called "old" during the recognition test.

6, 9, 12, and 15 items were, respectively, .02, .09, .08, .10, and .11. The intrusion rate for 3-item lists is probably small because recall was almost perfect for these lists. Most intrusions appeared to be associates of one or more of the words on the study lists. Note that these numbers are mean intrusions, not proportions; that is, subjects averaged less than one intrusion (besides the critical nonpresented word) for every 10 lists they recalled. Therefore, they followed our instructions not to guess, and the increase in false recall of critical items in Figure 1 is not due to a general tendency to guess more with longer lists.

Recognition

Recognition data from Experiment 1 need to be interpreted with care, as the recognition test always followed recall tests for all of the study lists and consequently was contaminated by prior recall. These data are presented in Table 1. The probability of correct recognition for studied associates increased slightly but reliably with increasing list length, $F(4, 140) = 2.51$, $MSE = 1.18$, $p < .05$. Note that tested items were always drawn from the first three items on each list, which may explain why recognition of studied items did not decrease with list length as recall did; we did not test the entire list.

False recognition of the critical items followed the same general pattern apparent in the recall data. As may be seen in Table 1, false recognition for the critical items increased with increasing list length, $F(5, 175) = 37.21$, $MSE = 0.74$, $p < .0001$. The pattern of false recognition coincided with that of recall, but this outcome may have been caused by the prior recall test. Thus, false recognition data for critical items that had not been falsely recalled are also presented in Table 1. Despite the fact that this set of data probably excludes critical items for which subjects had the "strongest" false memories, the overall pattern remains: The probability of false recognition increased reliably with list length, $F(4, 128) = 5.62$, $MSE = 0.06$, $p < .0001$. Thus, even when subjects did not falsely recall a critical item, the probability of false recognition for that item increased with the number of associates studied.

EXPERIMENT 2

Although the data from Experiment 1 suggest that false recall and false recognition are direct functions of the number of associated words in a study list, drawing this conclusion may be premature. In Experiment 1, number of associates was perfectly confounded with list length; thus, perhaps it was list length and not the number of related words that determined false recall and false recognition. The purpose of Experiment 2 was to remove this confounding by holding list length constant while varying the number of associated words within the list. We used the same lists from Experiment 1 as the initial portions of the lists in Experiment 2, but we added unrelated words to make all lists 15 items in length. Therefore, if the effect of number of associated items on false recollection is not simply due to list length, we expected the results of this experiment to replicate the pattern of results observed in Experiment 1.

Experiment 2 served another purpose as well. Because the exact lists used in Experiment 1 were used again in Experiment 2, the total associative strength of the study lists to the critical nonpresented items was the same across experiments. The mean associative strength of the lists declined in Experiment 2, however, as additional unrelated items were included in the lists. Thus, depending on whether false recollection is related to total or mean associative strength, levels of false recall in Experiment 2 would be expected either to equal those observed in Experiment 1 or to decline as a result of the inclusion of unrelated filler items. Further, because the same lists were used, and because subjects were randomly assigned to either Experiment 1 or Experiment 2, the two experiments may be directly compared.

Method

Subjects

Thirty-six Air Force recruits participated in this study to fulfill a requirement; 3 of these subjects failed to follow instructions and were later replaced by Rice University undergraduates.

Procedure

The procedure was identical to that of Experiment 1, except that unrelated words were used as fillers at the end of each of the shortened study lists to make lists of uniform length (15 items). Subjects in Experiments 1 and 2 were tested simultaneously.

Results and Discussion

Recall

The purpose of Experiment 2 was to disentangle the effect of list length from that of the number of associates in a study list. As shown in Figure 2, the results of Experiment 1 were generally replicated: As the number of associates in a study list increased, the probability of veridical recall of studied items decreased, but the probability of false recall of the critical item increased. The interaction between type of recall (correct or false) and number of associates studied was reliable, $F(4, 140) = 27.0$, $MSE = 0.02$, $p < .0001$.

The probability of correctly recalling a studied associate decreased as the number of such words in the list increased, $F(4, 140) = 330.16$, $MSE = 8.84$, $p < .0001$. Recall for studied

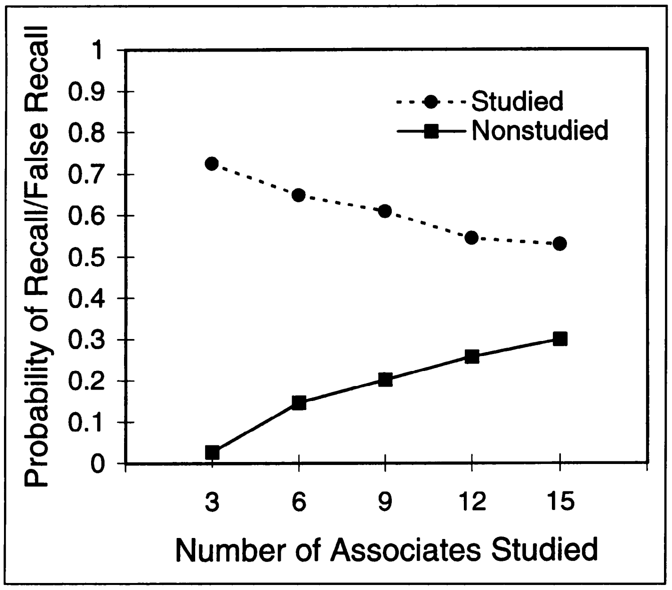

Fig. 2. Mean probabilities of correct and false recall in Experiment 2, as a function of the number of associated items in the study list.

words was poorer in this experiment than it was in Experiment 1 because the additional filler items reduced veridical recall, $F(1, 35) = 30.99$, $MSE = 0.01$, $p < .0001$. Additionally, the difference in correct recall between Experiments 1 and 2 was largest for lists with few associates (those that, in Experiment 2, contained a larger number of unrelated words), and then decreased systematically as number of associates increased. The interaction between experiment (1 or 2) and number of associates was significant, $F(3, 105) = 26.9$, $MSE = 0.01$, $p < .0001$.

The probability of falsely recalling the critical item increased systematically with the number of associates presented for study, $F(4, 140) = 13.16$, $MSE = 0.49$, $p < .0001$. As may be seen in Figure 2, the probability of false recall increased from .03 when 3 associates had been studied to .30 when 15 associates had appeared in the study list.

The rate of other, noncritical intrusions was not reliably affected by the number of associates, $F(4, 140) = 1.09$, $MSE = 0.67$, $p > .3$. The intrusion rates for lists with 3, 6, 9, 12, and 15 associated items were .13, .13, .13, .20, and .17, respectively. Again, keep in mind that these intrusion rates are means, not proportions. Although somewhat higher than in Experiment 1, the rates are still relatively low, indicating that subjects did not guess wildly during recall.

Recognition

Once again, the preceding recall test may have influenced the recognition results. As can be seen in Table 1, the probabilities of falsely recognizing a nonstudied critical item followed the same pattern observed in the recognition results of Experiment 1: The probability of false recognition of a critical item increased with the number of associates on the study list, $F(5, 175) = 19.39$, $MSE = 0.82$, $p < .0001$. False recognition for critical items that had not previously been falsely recalled was also examined, and the probabilities are presented in Table 1. The probability of false recognition again increased with the number of associates studied, $F(4, 136) = 2.91$, $MSE = 0.07$, $p < .05$.

The probability of calling a studied item "old" did not change systematically across list conditions, $F(4, 140) = 1.47$, $MSE = 0.85$, $p > .2$. This finding differs from the trend observed in Experiment 1, in which there was a slight but reliable increase in correct recognition as the number of associates studied increased. However, this finding is not unexpected: Just as the trend for correct recall was less pronounced in this experiment than it was in Experiment 1, so was the trend for correct recognition diminished by the inclusion of the filler words.

GENERAL DISCUSSION

The two experiments reported here revealed several new findings about associative processes in false recall and false recognition. First, the occurrence of false recall (in the form of the critical nonpresented item) was directly related to the number of associated words presented in the lists. Second, the increase in false recall as a function of the number of associated words was unaffected by the presence of filler words. Presented in Figure 3 are the false recall results from both experiments plotted together. Although filler items depressed veridical recall of list items, they left false recall unaffected. Third, false recognition also increased in direct proportion with the number of associates in the list in both experiments. Although this increase in false recognition may have been induced by the prior recall test, we suspect that the confounding effects of recall contributed little to the increase. When we conditionalized recognition scores to include only items that had not been recalled, the same increase was apparent in both experiments. In addition, our recognition findings replicate the work of other researchers who have shown a direct relation between the number of studied associates and false recognition (Hall & Kozloff, 1973; Hintzman, 1988; Shiffrin,

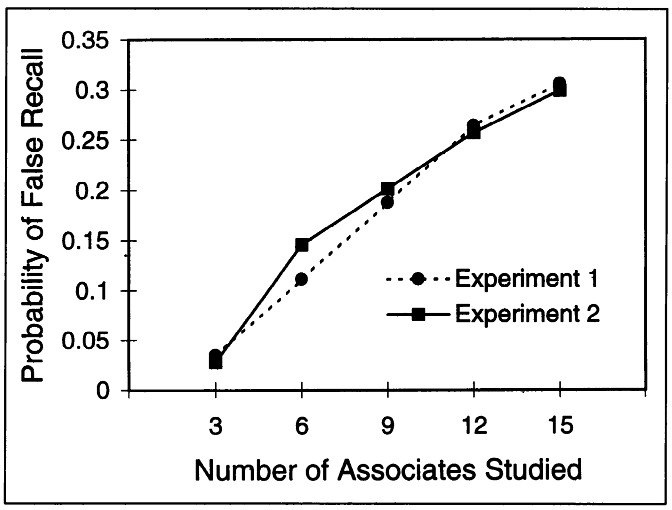

Fig. 3. Mean probabilities of false recall in Experiments 1 and 2 as a function of the number of associated items studied.

Huber, & Marinelli, 1995), although our experiments extended the range of this relation to include more associates than in previous experiments.

Perhaps the most striking finding in the present data is the similarity between the false recall results observed in Experiments 1 and 2. As may be seen in Figure 3, the inclusion of unrelated filler items in the study lists of Experiment 2 had no apparent effect on false recall of the critical items. Indeed, a 2 (experiment) × 5 (number of associates) ANOVA showed no effect of experiment, $F(1, 70) = 0.844$, $MSE = 1.14$, $p > .8$, and no interaction between the variables, $F(4, 280) = 0.19$, $MSE = 0.53$, $p > .9$. The main effect of number of associates studied, however, was highly reliable, $F(4, 280) = 25.05$, $MSE = 0.53$, $p < .0001$.

Given the data from our experiments, it is clear that false recall is not simply a function of the mean associative strength of the list words to the nonpresented target, as Deese (1959) proposed. If this were the case, then false recall levels in Experiment 2 (in which filler words were included) would have been greatly reduced, as the mean associative strengths of these lists are diluted by the presence of the unrelated words. A more likely interpretation based on these data is that the rate of false recall is a function of the total associative strength of the list words: highest for lists including many associates, and completely unaffected by unrelated filler items in the lists. However, this hypothesis awaits a more direct test than provided here.

The finding that the probability of veridical recall of a studied associate decreased when a larger number of associates was studied (observed in both experiments) is typical of list-length effects in recall. However, the probability of false recall of the critical items clearly did not follow this pattern; on the contrary, the data show a strong effect in the opposite direction. This difference is particularly interesting in light of the prediction we drew from Underwood's (1965) IAR concept, viz., that the highly associated nonpresented items, implicitly activated during study, should reveal the same patterns of recall as do studied words. Clearly, veridical recall drops as a function of the number of studied associated items, whereas false recall increases.

An alternative interpretation of Underwood's (1965) IAR hypothesis, however, could explain the reported findings. Underwood implied that an IAR for a nonstudied word may occur when a single highly associated word is studied. Each of the study lists used in these experiments included multiple items that were highly associated to a single nonpresented word, the critical item. It follows, therefore, that when a list of associates was studied, an IAR of the critical item might have occurred multiple times. Although the IAR probably did not occur in response to every studied associate, a safe assumption is that it occurred more frequently in the lists containing many associates than in those containing few associates. Thus, the increased probabilities of false recall and false recognition in the lists with many associates could be explained by this assumption: Just as veridical recall generally increases with the number of repetitions of a studied item, so false recall might rise as a function of "implicit repetitions."

Another possible cause for the higher rate of false recall observed with the lists containing many associates could be differences occurring during retrieval. The number of associates recalled is larger for a studied list containing many associates than for a short list. Thus, in conditions in which the study lists contained large numbers of associates, it could be the larger number of associates produced during recall that mediated the observed higher levels of false recall, rather than or in addition to the larger number of words studied. Indeed, Cramer (1970) presented evidence suggesting that IARs may occur during both study and retrieval tasks.

False recall and false recognition have also been explained by fuzzy trace theory (e.g., Brainerd, Reyna, & Brandse, 1996; Reyna & Brainerd, 1995). Payne et al. (1996) and Schacter et al. (1996) have applied the theory to the false recall and false recognition observed in the Roediger-McDermott (1995) procedure used here. In brief, fuzzy trace theory maintains that subjects may "develop two separate representations during encoding, a verbatim memory trace and a 'gist' representation that stores semantic content. An admirable feature of this theoretical framework is that it provides an account of both how people can accurately recall study items (by accessing the verbatim representation of these items) as well as how systematic errors can occur (by consulting the gist representation of these items)" (Payne et al., 1996, pp. 280–281). Therefore, a subject may attempt to recall a study list based on the general theme, or gist, encoded during study and consequently intrude items that are representative of that gist. Fuzzy trace theory could account for the direct relation between the number of associates studied and false recall and false recognition in our experiments if we assume that a stronger gist representation is formed when many associates are studied, or similarly, that the probability of forming a useful gist representation is greater when many associates are studied. Either way, false recall levels would increase systematically as the number of studied associates increased.

One potential difficulty for the fuzzy trace account of false recall and false recognition is the fact that subjects seem to experience these false memories as tapping quite specific knowledge. The subjects claim to remember the actual occurrence of the items in the lists, are willing to attribute serial positions to these items, and also make voice attributions for nonstudied items to the same extent as they do for studied items (see Payne et al., 1996; Read, 1996; Roediger & McDermott, 1995). One might expect that false memories based on gist (and not on verbatim traces, which we take to be representations that carry specific information) would not lead to such attributions. Gist representations might lead to familiarity, but why should they lead to subjects making *remember* judgments and attributing serial position and modality to the critical nonstudied items and—even more perplexingly—at the same levels as for studied items? From the perspective of fuzzy trace theory, studied words should have specific and general (gist) information associated with them, whereas critical nonstudied words should have only the latter. But should not studied words then enjoy an advantage in recall, in recognition, and in various metamemory judgments because of this extra (specific) information?

Nonetheless, in terms of the present results, both the IAR and gist hypotheses potentially explain the increase in false recall and false recognition. Future research will be required to distinguish the various hypotheses that may account for false memories in this paradigm (see Roediger & McDermott, 1995, for other possibilities). The present results add to current knowledge of this memory illusion by showing that it is not mean associative strength of items in the list, but rather total associative strength, that predicts false recall and false recognition.

The memory illusion studied here is one of many under current scrutiny by cognitive psychologists (see Roediger, 1996, and Schacter, 1995, for overviews). In general, one key to false recollections is demonstrated here: The greater the number of related events that have been encoded, the more easily people are led to recollect that some other related event also occurred, even though this related event never happened.

Acknowledgments—The research reported here was supported by Grant F49620-92-J-0437 from the Air Force Office of Scientific Research. We thank Kathleen B. McDermott for her continued assistance and helpful comments.

REFERENCES

Brainerd, C.J., Reyna, V.F., & Brandse, E. (1996). Are children's false memories more persistent than their true memories? *Psychological Science, 6,* 359–364.

Cramer, P. (1970). Semantic generalization: IAR locus and instructions. *Journal of Experimental Psychology, 83,* 266–273.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology, 58,* 17–22.

Hall, J.W., & Kozloff, E.E. (1973). False recognitions of associates of converging versus repeated words. *American Journal of Psychology, 86,* 133–139.

Hintzman, D.L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95,* 528–551.

McDermott, K.B. (1996). The persistence of false memories in list recall. *Journal of Memory and Language, 35,* 212–230.

Murdock, B.B., Jr. (1961). The retention of individual items. *Journal of Experimental Psychology, 62,* 618–625.

Payne, D.G., Elie, C.J., Blackwell, J.M., & Neuschatz, J.S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language, 35,* 261–285.

Read, J.D. (1996). From a passing thought to a false memory in 2 minutes: Confusing real and illusory events. *Psychonomic Bulletin and Review, 3,* 105–111.

Reyna, V.F., & Brainerd, C.J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7,* 1–75.

Roediger, H.L., III. (1996). Memory illusions. *Journal of Memory and Language, 35,* 76–100.

Roediger, H.L., III, & McDermott, K.B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803–814.

Schacter, D.L. (1995). Memory distortion: History and current status. In D.L. Schacter, J.T. Coyle, G.D. Fischbach, M.-M. Mesulam, & L.E. Sullivan (Eds.), *Memory distortion: How minds, brains and societies reconstruct the past* (pp. 1–43). Cambridge, MA: Harvard University Press.

Schacter, D.L., Verfaellie, M., & Pradere, D. (1996). The neuropsychology of memory illusions: False recall and recognition in amnesic patients. *Journal of Memory and Language, 35,* 319–334.

Schneider, W. (1988). Micro Experimental Laboratory: An integrated system for IBM PC compatibles. *Behavior Research Methods, Instruments and Computers, 20,* 206–217.

Shiffrin, R.M., Huber, D.E., & Marinelli, K. (1995). Effects of category length and strength on familiarity in recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 267–287.

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist, 26,* 1–12.

Underwood, B.J. (1965). False recognition produced by implicit verbal responses. *Journal of Experimental Psychology, 70,* 122–129.

(Received 7/3/96; Accepted 11/20/96)